XIII CONGRESSO BRASILEIRO DE AGROINFORMÁTICA

10 A 12 DE NOVEMBRO DE 2021



ANAIS



XIII Congresso Brasileiro de Agroinformática

Anais

Ana Paula Lüdtke Ferreira

(Org.)

Promoção

Organização

Patrocínio

Apoio

Congresso Brasileiro de Agroinformática (13. : 2021: Bagé, RS)
SBIAGRO 2021 [recurso eletrônico] / Ana Paula Lüdtke Ferreira (Org.) Bagé: Universidade Federal do Pampa (UNIPAMPA), 2021.

407 p.:il.

Modo de acesso: <https://eventos.unipampa.edu.br/sbiagro/anais/>
ISSN: 2177-9724
ISBN: 978-65-00-34526-1

Instituições promotoras: Universidade Federal do Pampa, EMBRAPA Pecuária Sul, Sociedade Brasileira de Agroinformática.

1. Agropecuária de Precisão. 2. Agricultura Digital. 3. Agro 4.0. 4. Computação Aplicada. I. Ferreira, Ana Paula (coord.). II. Universidade Federal do Pampa. III. EMBRAPA Pecuária Sul. IV. Sociedade Brasileira de Agroinformática. V. Título. VI. Anais do SBIAGRO 2021. VII. Anais do Congresso Brasileiro de Agroinformática.

CDU: 004

APRESENTAÇÃO

A Associação Brasileira de Agroinformática (SBIAgro), sociedade sem fins lucrativos, foi criada no dia 18 de setembro de 1996 por cerca de 30 pessoas que participaram da assembleia de fundação realizada em Viçosa (MG). A SBIAgro nasceu como resultado do I Seminário Internacional de Informatização da Agropecuária (Agrosoft 95), realizado de 25 a 27 de outubro de 1995, em Juiz de Fora (MG), sendo uma das proposições discutidas no evento. O objetivo da SBIAgro é conglomerar a comunidade de agroinformática, promovendo a conscientização, estudos e disseminação de tecnologias no setor. Dentre as diversas realizações da SBIAgro desde a sua fundação, destaca-se o seu congresso bienal, oportunidade para reunir a comunidade sempre crescente de agroinformática – agora repaginada pela denominação de agricultura digital.

O XIII Congresso Brasileiro de Agroinformática ocorreu entre os dias 10 e 12 de novembro de 2021, com organização da Universidade Federal do Pampa e EMBRAPA Pecuária Sul e contando com a colaboração de pesquisadores de diversas outras instituições de ensino e pesquisa do país, com o patrocínio da *Ripple Impact – University Blockchain Research Initiative* e do *Center for Artificial Intelligence* da USP e com apoio do Instituto de Estudos Avançados da USP, Sociedade Brasileira de Computação, SEBRAE, Celeiro AgroHub e TecnoPUC.

No congresso apresentamos, discutimos e disseminamos as tendências e soluções em agricultura digital, não apenas em pesquisa e desenvolvimento, pois procuramos promover uma aproximação e integração entre PD&I trazendo empresas, microempresas e *startups*, visando a articulação de parcerias para inovação aberta, fomentando o ecossistema de inovação da agricultura digital. Neste ano, o congresso cobriu o lançamento do manifesto da Rede GO FAIR para a Agricultura, palestras e painéis com temas em *blockchain*, mineração de dados e inteligência artificial aplicados aos sistemas produtivos agropecuários, políticas e iniciativas governamentais para a agricultura digital, bem como desafios e perspectivas sobre o tema. Como tem ocorrido desde a sua primeira edição, nas nove sessões técnicas foram apresentados trabalhos de excelência, que contribuem para o avanço de PD&I na área. O sub-evento CONECT@, cujo objetivo é proporcionar um espaço de interação direta entre empresas e PD&I para inovação aberta ou outras parcerias, foi realizado no formato painéis de experiências em inovação aberta, acessíveis a todos os congressistas e transmitido em formato aberto e gratuito para todos os interessados.

As palestras e painéis envolveram 42 pesquisadores, gestores e empresários, que engrandeceram as discussões realizadas. O comitê de programa foi formado por quase uma centena de pesquisadores doutores de todas as regiões do país. As revisões de artigos foram rigorosas, cuidadosas e formativas. A coordenação do evento agradece a todas e todos que participaram, cedendo seu tempo para que tivéssemos a possibilidade de realizar um encontro com a qualidade que tivemos.

O SBIAGRO 2021 teve primeira edição *on line*, desafio provocado pela pandemia da COVID-19 e vencido pela força do compromisso, valores e união da nossa comunidade.

Maria Fernanda Moura
Presidente da SBIAgro 2020/2021

Ana Paula Lüdtke Ferreira
Presidente do SBIAGRO 2021

COMITÊS

Organização Geral
Ana Paula Lüdtke Ferreira (UNIPAMPA)
Antônio Mauro Saraiva (USP)
Maria Angélica de Andrade Leite (EMBRAPA Informática Agropecuária)
Maria Fernanda Moura (EMBRAPA Informática Agropecuária)
Naylor Bastiani Perez (EMBRAPA Pecuária Sul)

Coordenação do Conect@
Naylor Bastiani Perez (EMBRAPA Pecuária Sul)

Coordenação do Concurso de Iniciação Científica
Érico Marcelo Hoff do Amaral (UNIPAMPA)

Coordenação do Concurso de Teses e Dissertações
Maria Fernanda Moura (EMBRAPA Informática Agropecuária)
Sandro da Silva Camargo (UNIPAMPA)

Coordenação do Comitê de Programa
Ana Paula Lüdtke Ferreira (UNIPAMPA)

Comitê de Programa
Adauto Mancini (EMBRAPA)
Alceu Britto Jr. (PUCPR)
Alexandro Schäfer (UNIPAMPA)
Ana Maria Heuminski de Ávila (CEPAGRI)
Antônio Mauro Saraiva (USP)
Bruno Nogueira (UFMS)
Bruno Albertini (USP)
Camilo Carromeu (EMBRAPA/UFMS)
Carlos Amaral Hölbig (UPF)
Carlos Kamienski (UFABC)
Carlos Michel Betemps (UNIPAMPA)
Carolina Bremm (DDPA/SEAPDR)
Celia Regina Grego (EMBRAPA)
Celina Takemura (EMBRAPA)
Claudio Bazzi (UTFPR)
Daniel Victoria (EMBRAPA)
Debora Drucker (EMBRAPA)
Eduardo Speranza (EMBRAPA)
Eduardo Fávero Caires (UEPG)
Eduardo Augusto Bezerra (UFSC)
Emanuelle Baldo Gaspar (EMBRAPA)

Émerson Medeiros Del Ponte (UFV)
Érico Hoff Amaral (UNIPAMPA)
Ewerton Andrade (UNIR)
Fábio Livi Ramos (UNIPAMPA)
Fábio Prass (FP2 Tecnologia)
Francisco Arnold (UNICAMP)
Gustavo Trentin (EMBRAPA)
Hana Karina Salles Rubinsztejn (UFMS)
Heloisa Camargo (UFSCAR)
Irineu Sotoma (UFMS)
Jayme Barbedo (EMBRAPA)
Jaziel Lobo (IFS)
Jésus Franco Bueno (UFMT)
João Antunes (EMBRAPA)
João Silva (EMBRAPA)
João Camargo Neto (EMBRAPA)
Josenalde Oliveira (UFRN)
Juliana Saragiotto Silva (IFMT)
Juliana Yassitepe (EMBRAPA)
Julio Saraçol Domingues Júnior (UNIPAMPA)
Júlio César Dalla Mora Esquerdo (EMBRAPA)
Kleber Xavier Sampaio de Souza (EMBRAPA)
Leandro dos Santos Coelho (PUCPR)
Leonardo Bidese de Pinho (UNIPAMPA)
Leonardo Queirós (EMBRAPA)
Lubia Vinhas (INPE)
Luciana Romani (EMBRAPA)
Luiz Camolesi Jr. (UNICAMP)
Márcia Cristina Silveira (EMBRAPA)
Maria Fernanda Moura (EMBRAPA)
Marilton Sanchotene Aguiar (UFPEL)
Naylor Bastiani Perez (EMBRAPA)
Nuno Escudeiro (ISEP)
Raul Teruel dos Santos (UFMT)
Renato Azevedo (UFSM)
Robert Domingues (EMBRAPA)
Rodrigo Bezerra de Araújo Gallis (UFU)
Ronilson de Souza Santos (UFPA)
Rosane Falate (UEPG)
Sandro da Silva Camargo (UNIPAMPA)
Sergio Manuel Serra da Cruz (UFRRJ)
Silvia Massruhá (EMBRAPA)
Sonia Ternes (EMBRAPA)
Tercia Torres (EMBRAPA)
Thiago Santos (EMBRAPA)

Vilmar Rodrigues Moreira (PUC-PR)
Willingthon Pavan (UPF)

Organização Local
Ana Paula Lüdtke Ferreira
Érico Marcelo Hoff do Amaral
Leonardo Bidese de Pinho
Julio Saraçol Domingues Júnior
Sandro da Silva Camargo
Alex Sandro Ernandes Valério
Aline Pons Alves Lisboa
Anderson dos Santos da Rosa
Beatriz Assumpção Tergolino
Eugênio Teixeira Schreiber Pierazzoli
Felipe Bedinotto Fava
Fiama Jimenez Ochoa
Francine Moreira Ferreira
George Bezerra da Rocha Silva
Jairo Silveira Genro Neto
Juliane Pires Martins
Maiara Sousa Gomes
Mikael Adolfo Maria
Nathália da Silva Fialho
Rui Antônio da Rocha Lorensi

PROGRAMAÇÃO

Palestra Convidada

FUTURO DO USO DA INTELIGÊNCIA ARTIFICIAL NA PRODUÇÃO ANIMAL

Profª Drª Irenilza A. Nääs



Irenilza de Alencar Nääs é Engenheira graduada pela UNICAMP (1974), Mestre em Agricultura, pela California Polytechnic State University (Cal-Poly, 1975) e PhD em Agricultural Engineering pela Michigan State University (1980). Bolsista de Produtividade em Pesquisa do CNPq - Nível 1A, hoje é professora titular na Universidade Paulista (UNIP) desenvolvendo pesquisas em redes de produção de alimentos. É também professora colaboradora na Universidade Estadual de Campinas, na Universidade Federal da Grande Dourados e na Florida University, EUA. Orientadora de mestrado e doutorado, tem mais de trezentos artigos publicados em periódicos especializados, vários capítulos de livros e livros publicados, além de produtos tecnológicos e softwares patenteados, em temas relacionados à aplicação dos conceitos de engenharia na indústria de produção animal. Foi presidente da Associação Brasileira de Engenharia Agrícola-SBEA, da Asociación Latino Americana y del Caribe de Ingenieria Agricola - ALIA, da International Comission of Agricultural Engineering - CIGR e da FACTA-Fundação de Ciência e Tecnologia Avícola. Irenilza é Fellow da American Society of Agricultural an Biological Engineers-ASABE e membro da Scientific Research Honor Society-Sigma Xi.

Palestra Convidada

MINERAÇÃO DE DADOS NA PECUÁRIA: INTELIGÊNCIA ARTIFICIAL E INTELIGÊNCIA HUMANA TRABALHANDO CONJUNTAMENTE

Prof. Dr. Guilherme J. M. Rosa



Guilherme Rosa é Zootecnista e Mestre em Zootecnia pela Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, Botucatu (1994), Doutor em Agronomia (Estatística e Experimentação Agronômica) pela Universidade de São Paulo (ESALQ/Piracicaba, 1999), e pós-doutor em Estatística Genética pela Universidade de Wisconsin (UW) - Madison, EUA (2000). Foi professor de Bioestatística na UNESP por 10 anos, professor de Estatística

Genética na Michigan State University (EUA) por 4 anos e atualmente é professor de Estatística Genética e Genômica na Universidade de Wisconsin, Madison/WI, também nos EUA. Orientador de doutorado, com mais de 200 artigos publicados em periódicos internacionais, tem experiência na área de delineamento de experimentos e análise de dados, bem como em genética quantitativa e melhoramento genético em diferentes espécies animais. Desenvolve pesquisa principalmente com modelos de efeitos mistos e inferência Bayesiana, métodos de Monte Carlo, estudos de expressão gênica, mapeamento de genes e seleção assistida por marcadores.

Palestra Convidada

A REDE GO-FAIR BRASIL: FAIR DIGITAL OBJECTS E A INTERNET PARA MÁQUINAS

***Introdução temática**: Luana Sales (IBICT, Coordenação Escritório Brasileiro Go-Fair), Maria Luiza Campos (UFRJ, Coordenação de Assuntos Internacionais Go-Fair Brasil), Carla Macario e Debora Drucker (EMBRAPA Agricultura Digital, Coordenação da Rede de Implementação Temática Go-Fair Agro)*

***Palestrante**: Prof. Dr. Luiz Olavo Bonino da Silva Santos (University of Twente, Leiden University Medical Center e Go-Fair International Support and Coordination Office)*



Luiz Olavo Bonino da Silva Santos é Mestre em Ciência da Computação pela UFES e Doutor em Ciência da Computação pela University of Teente. Atualmente é professor associado da University of Twente e do Leiden University Medical Center. Atua como International Technology Coordinator na GO FAIR International Support and Coordination Office.

Painel

TECNOLOGIAS DE BLOCKCHAIN NO AGRO

Painelistas:



Marcos Antônio Simplício Jr.
University Blockchain Research Initiative / Amazonia 4.0
Escola Politécnica da Universidade de São Paulo (EPUSP)



Eric Luque
Ecotrace
https://ecotrace.info



Inacio Henrique Yano
EMBRAPA Agricultura Digital



Gladstone Arantes Jr.
BNDES



Wilian França Costa
Universidade Presbiteriana Mackenzie



Luciano Ermilivitch

Mediador:



Bruno de Carvalho Albertini
Escola Politécnica da Universidade de São Paulo (EPUSP)

Painel

POLÍTICAS E INICIATIVAS GOVERNAMENTAIS PARA A AGRICULTURA DIGITAL

***Painelistas**:*



Daniel Trento do Nascimento
Coordenador Geral de Articulação para Inovação
Departamento de Apoio à Inovação para Agropecuária
Secretaria de Inovação, Desenvolvimento Rural e Irrigação
Ministério da Agricultura, Pecuária e Abastecimento (MAPA)



Eliana C. Emediato de Azambuja
Coordenadora-Geral de Transformação Digital
Departamento de Ciência, Tecnologia e Inovação
Secretaria de Empreendedorismo e Inovação
Ministério da Ciência, Tecnologia e Inovações (MCTIC)



Pedro Lucas Araújo
Diretor de Investimentos e Inovação
Departamento de Aprimoramento do Ambiente de Investimentos em Telecomunicações
Secretaria de Telecomunicações
Ministério das Comunicações (MCOM)

***Mediadora**:*

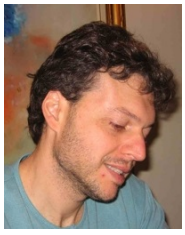


Sílvia Maria Fonseca Silveira Massruhá
Chefe geral da EMBRAPA Agricultura Digital

Painel

DESAFIOS E PERSPECTIVAS DA INTELIGÊNCIA ARTIFICIAL NA AGRICULTURA DIGITAL

***Painelistas**:*



Alexandre Cláudio Botazzo Delbem
Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP)



Daniel Gonzalez
A De Agro
https://www.adeagro.com.br/



Daniel Jardim Issa
A De Agro
https://www.adeagro.com.br/



Jayme Garcia Arnal Barbedo
EMBRAPA Agricultura Digital



Marcus Adonai Castro da Silva
BIOME4ALL Interpretação de Dados Genéticos
https://biome4all.com.br/



Rodrigo Yoiti Tsukahara
Fundação ABC
https://fundacaoabc.org/

Mediador:



Fernando Santos Osório
Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP)

Painel Conect@ Inovação Aberta

EXPERIÊNCIAS DA EMBRAPA EM INOVAÇÃO ABERTA

Painelistas:



Paulo do Carmo Martins
Ideas for Milk
EMBRAPA Gado de Leite



Luciana Alvim Santos Romani
TechStart Agro Digital
EMBRAPA Agricultura Digital



Carina Ferreira Gomes Rufino
Open Innovation Soja
EMBRAPA Soja

Mediador:



Vinícius Milléo Kuromoto
EMBRAPA Agricultura Digital

Painel Conect@ Inovação Aberta

EXPERIÊNCIAS DE FOMENTO EM INOVAÇÃO ABERTA

Painelistas:



Luís Gustavo Henrique Peles
Associação Nacional de Entidades Promotoras de Empreendimentos Inovadores (Anprotec)



Luis Humberto Villwock
Celeiro AgroHub
VENTIUR/Aglab/TecnoPUC



Gustavo Schneck Moreira
Gestor de Projetos de Inovação
SEBRAE-RS

***Mediador**:*



Naylor Bastiani Perez
EMBRAPA Pecuária Sul

Painel Conect@ Inovação Aberta

EXPERIÊNCIAS DE *START-UPS* EM INOVAÇÃO ABERTA

***Painelistas**:*



Ulf Bogdawa
Skydrones (https://skydrones.com.br/)



Volmir Pavlak Pagno
Leigado (https://leigado.com.br/pt)



Adriano Pereira de Freitas
OvinoPro (https://www.ovinopro.com.br/)



Erli Salgueiro
AcertoFácil (https://acertofacil.com.br/)

***Mediadores**:*



Tiago Rolim Marques
Embrapa Pecuária Sul



Cibele Bolzan Scherer
SEBRAE-RS

PREMIAÇÕES

Melhores artigos

1º lugar
CanControl: plataforma para diagnóstico do Cancro Europeu da Macieira
Wilson C. Branco Neto (IFSC), Leonardo Araújo (Epagri), Felipe A. M. Ferreira Pinto (Epagri), Rodrigo A. Machado (IFSC), Yan F. B. Ribeiro (IFSC), Wilson F. Cordova Junior (IFSC), Kevin Mondadori Mattos (IFSC)

2º lugar
Experimental evaluation of Data Augmentation heuristics for plant identification systems based on Deep Learning
Luciano Araújo Dourado Filho (UEFS), Rodrigo Tripodi Calumby (UEFS)

3º lugar
Multi-objective methods for crop insurance premiums: framework proposal and a case study in sugarcane
Roberto F. Silva (USP São Carlos), Marcos R. Benso (USP São Carlos), Gabriela C. Gesualdo (USP São Carlos), Eduardo M. Mendiondo (USP São Carlos), Antônio M. Saraiva (USP), Patrícia A. A. Marques (ESALQ/USP), Alexandre C. B. Delbem (USP São Carlos)

Melhores Dissertações de Mestrado

1º lugar
Estimating and tuning adaptive action plans for the control of smart interconnected poultry houses
Autor: Darlan Felipe Klotz
Orientador: Marcelo Teixeira, Coorientador: Dalcimar Casanova
Programa de Pós-graducação em Engenharia Elétrica, Universidade Tecnológica Federal do Paraná (UTFPR)

2º lugar
Mineração de dados em rede social para avaliação de tendências de consumo do queijo artesanal no Brasil
Autor: Thallys da Silva Nogueira
Orientadora: Priscila Vanessa Zabala Capriles Goliatt
Programa de Pós-Graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora (UFJF)

Melhores trabalhos de Iniciação Científica

1º lugar
***Deep Features* para sistemas de identificação de plantas**
Luciano Araújo Dourado Filho, Rodrigo Calumby
Universidade Estadual de Feira de Santana (UEFS)

2º lugar
Estimativa da Cercosporiose do café usando clima em Varginha - Minas Gerais
Camila Cabral, Rafael Fausto de Lima, Lucas Aparecido, Fábio Corsini, Geraldo Júnior
Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas (IFSULDEMINAS), Campus Muzambinho

3º lugar
Redes Neurais na Simulação de Condições Climáticas Potenciais ao Cultivo de Bambu no Brasil
João Lorençone, Pedro Lorençone, Lucas Aparecido, Guilherme Torsoni
Instituto Federal de Mato Grosso do Sul (IFMS)

ARTIGOS COMPLETOS

1 **A methodology for detection and localization of fruits in apples orchards from aerial images**
Thiago T. Santos, Luciano Gebler 1

2 **AgroDigital-NE: um mapeamento da produção científica sobre tecnologias de informação e comunicação aplicadas ao agronegócio no Nordeste brasileiro**
Laura Emmanuela A. dos S. Santana, Leonardo Rodrigues de Lima Teixeira, Alixandre Santana, Josenalde Barbosa de Oliveira 10

3 **Análise bibliométrica da produção científica da EMBRAPA na Web of Science**
Alexandre D. Alves 19

4 **Avaliação de agitabilidade bovina baseada em dados de acelerômetro**
Tulon Pereira da Fontoura 29

5 **BrangusSelection: um algoritmo para seleção ótima de acasalamentos com índice de seleção customizável**
Bruno Ernesto Techera da Motta, Ana Paula Lüdtke Ferreira, Marcos Jun-Iti Yokoo . 35

6 **CanControl: plataforma para diagnóstico do Cancro Europeu da Macieira**
Wilson C. Branco Neto, Leonardo Araújo, Felipe A. M. Ferreira Pinto, Rodrigo A. Machado, Yan F. B. Ribeiro, Wilson F. Cordova Junior, Kevin Mondadori Mattos . . 44

7 **Controle de temperatura e de umidade de uma composteira utilizando o ESP32**
Luan D. Rodrigues, Rodrigo M. Pereira, David Willy F. Palmera, Guilherme Henrique A. Silva, Marco Antonio F. Leite, Juliana S. Silva 53

8 **Desempenho da classificação supervisionada com a inclusão da banda termal em áreas selecionadas do bioma Pampa – Brasil**
Patricia Michele Pereira Trindade, Daniela Wancura Barbieri Peixoto, Tatiana Mora Kuplich, Claudio Aparecido de Almeida 62

9 **Desenvolvimentos em inteligência artificial na avicultura de frangos de corte**
Lucas Gabriel Galdino da Silva, Francisco Gomes Cordeiro Neto, Josenalde Oliveira, Laura Santana 71

10 **Diferenciação e caracterização de municípios produtores de bovinos de corte no bioma Mata Atlântica segundo o Censo Agropecuário 2017**
Maria do Carmo Ramos Fasiaben, Stanley Robson de Medeiros Oliveira, André Steffens Moraes, Maxwell Merçon Tezolin Barros Almeida, Octávio Costa de Oliveira, Gabriela dos Santos Eusébio 80

11 **DSSAT-Lite: a web-based application for running crop models and analyzing results from DSSAT-CSM**
Jonas de Abreu Resenes, Gerrit Hoogenboom, Alexandre Lazaretti Zanatta, Willingthon Pavan, Carlos Amaral Hölbig 90

12 **Effects of agro-sensor time series approximation on plant stress detection: an experimental study**
Marcos A. de Oliveira Jr., Gregory Sedrez, Anderson Monteiro, Fernando Emilio Puntel, Gerson Geraldo H. Cavalheiro 99

13 **Estimating and tuning adaptive action plans for the control of smart interconnected poultry houses**
Darlan F. Klotz, Dalcimar Casanova, Marcelo Teixeira 108

14 **Estruturação de um *Data Warehouse* para repartição territorial de desmatamentos no bioma Amazônia e o Cadastro Ambiental Rural**
Renato da Silva de Deus, Mariela Almeida, Carlos Alberto de Carvalho 116

15 **Estudo Comparativo entre Máquina de Vetores de Suporte e Rede Neural Convolucional na Classificação do inseto *Diaphorina citri***
Mário Lúcio G. Q. Pierre Júnior, José Leonardo dos Santos Melo, Marcelo Pedreira de Miranda, Michele Fúlvia Angelo 126

16 **Experimental evaluation of Data Augmentation heuristics for plant identification systems based on Deep Learning**
Luciano Araújo Dourado Filho, Rodrigo Tripodi Calumby 136

17 **Explorando técnicas de aprendizado de máquina para aprimoramento da previsão de geadas no sul e sudeste do Brasil**
José Roberto Motta Garcia 144

18 **H-Pasture: uma aplicação móvel para estimativa da disponibilidade de forragem e ajuste de carga em pastagens sob lotação contínua ou rotativa**
Wendell Gasparoni, Ana Paula Lüdtke Ferreira, Teresa Cristina Moraes Genro, Márcia Cristina Teixeira Silveira 154

19 **Implantação da Rede Temática GO-FAIR Agro Brasil: Primeiros Passos**
Debora P. Drucker, Sérgio Manuel Serra da Cruz, Antonio M. Saraiva, Juliana M. Fortaleza, Patrícia R. B. Bertin, Victor P. M. Simão, Milena A. Telles, Alessandra R. da Silva, Paulo S. S. Santos, Carla G. N. Macario 164

20 **IrrigaFeijão – Software para Manejo da Irrigação do Feijoeiro**
Pedro M. Silveira, Marcelo G. Narciso, Alexandre B. Heinemann, Silvando C. Silva, Fábio F. Noleto, Luiz F. Stone 172

21 **Mineração de dados em rede social para avaliação de tendências de consumo do queijo artesanal no Brasil**
Thallys da Silva Nogueira, Kennya Beatriz Siqueira, Priscila Vanessa Zabala Capriles Goliatt 179

22 **MIP^2 - um sistema de informação especializado em manejo integrado de pragas**
Rodolfo Chagas Marinho Nascimento, Monica Simão Giponi, Wellington Moreira de Oliveira, Vânia Maria Xavier 188

23 **Modelagem de uma RSSF para manejo de pragas em pomares usando o simulador Eboracum**
Felipe Frantz Zanini, Weslen Schiavon de Souza, Lisane Brisolara de Brisolara, Paulo R. Ferreira Jr. 197

24 **Modelos de aprendizagem de máquina para predição da presença de desoxinivalenol em grãos de trigo por meio de análises multiespectrais**
Eduarda Zanini, Deividi Felipe Zaions, Alexandre Lazaretti Zanatta, Carlos Amaral Hölbig, Willingthon Pavan 207

25 **Monitoramento de plantas em casas de vegetação para assimilação de dados**
Monique Pires Gravina de Oliveira, Luiz Henrique Antunes Rodrigues 215

26 **Multi-objective methods for crop insurance premiums: framework proposal and a case study in sugarcane**
Roberto F. Silva, Marcos R. Benso, Gabriela C. Gesualdo, Eduardo M. Mendiondo, Antônio M. Saraiva, Patrícia A. A. Marques, Alexandre C. B. Delbem 225

27 **Ontologias para modelos de negócio na pecuária de corte: uma revisão da literatura**
Matheus Collares Machado, Ana Paula Lüdtke Ferreira 234

28 **Perfil tecnológico e de automação das unidades armazenadoras de grãos do Estado do Paraná**
Fernando Carvalho, Luiz Antonio Rodrigues 245

29 **Processamento de Linguagem Natural para consultas de invasores na cultura da soja**
Carolinne Roque e Faria, Cinthyan Renata Sachs C. de Barbosa 254

30 **Proposta de um modelo de processo para aplicação de métodos de interpolação**
Gustavo E. Kida, Raul T. Santos 264

31 ***Semantic Sensor Network Ontology* como modelo de referência em soluções de monitoramento agrícola**
Julio C. Nardi, Vitor N. Alves, Carlos A. R. Junior, Fabiano B. Ruy, Robson Prucoli Posse 273

32 **SementeSuite: desenvolvimento de uma aplicação móvel para determinação do teor de água em sementes**
Érik Alves Almeida, Humberto B. da Cunha Júnior, Cosmo Rufino de Lima, Fábio José Marques, Pablo Radamés C. de França 283

33 **Simulação de Sistemas de Produção de Novilho Precoce: uma Aplicação Prática**
Roberto A. F. Abreu, Thiago Basso, Camilo Carromeu, Rodrigo da Costa Gomes, Fernando R. T. Dias, Debora Maria Barroso Paiva 292

34 **Sistemas de suporte à decisão nas fazendas inteligentes: um mapeamento sistemático**
Jonas S. Gomes, José Maria N. David, Regina Braga, Bryan Barbosa, Wagner Arbex, Wneiton Luiz Gomes, Leonardo M. Gravina Fonseca 301

35 **SQUADRO-i: Uma plataforma para o monitoramento da sustentabilidade de sistemas produtivos agropecuários**
Kimberlly da Rosa Moreles, Vinicius do Nascimento Lampert, Érico Marcelo Hoff Amaral 311

36 **Towards an IoT-Based Architecture for Monitoring and Automated Decision-Making in an Aviary Environment**
Vinícius C. Lopes, Roberto Felício de Oliveira, Valdemar Vicente Graciano Neto . . 320
37 **Um estudo comparativo entre as arquiteturas de redes neurais profundas AlexNet e YOLO aplicadas ao problema de visão computacional em ambientes agrícolas**
Danielly Omori Antunes de Oliveira, Josué Pereira de Castro 329
38 **Uma plataforma de dados geoespaciais para caracterização e combate à desertificação da Caatinga**
Jaudete Daltio, André Rodrigo Farias, Lucíola Alves Magalhães, Daniela Maciel Pinto 339
39 **Uso de *workflows* ETLH para integrar *datasets* pedológicos: estudo para adequação aos princípios FAIR**
Sabrina Santos Cruz de Oliveira, Emerson de Barros Duarte, Élton Carneiro Marinho, Sérgio Manuel Serra da Cruz 348

TRABALHOS DE INICIAÇÃO CIENTÍFICA

1 **Avaliação da eficiência do classificador Iso Cluster para o monitoramento de desmatamento na Amazônia Legal**
Adriana M. Iora, Andressa Boni, Igor Vani, Sandro L. B. Fensterseifer 358
2 **Classificação Climática de Thornthwaite (1948) Para a Região Centro-Oeste: uma abordagem inédita para a maior região produtora de grãos do Brasil**
Rafael F. Lima, Lucas E. O. Aparecido, Guilherme B. Torsoni, Gabriel H. O. Souza . 362
3 **Deep Features para sistemas de identificação de plantas**
Luciano A. D. Filho, Rodrigo T. Calumby 366
4 **Estimativa da Cercosporiose do café usando clima em Varginha, Minas Gerais**
Camila D. Cabral, Rafael F. Lima, Lucas Eduardo de O. Aparecido, Fábio dos S. Corsini, Geraldo G. de O. Júnior 370
5 **Impacto das mudanças climáticas no período 2040-2080 sobre a aptidão do cultivo de *Coffea arabica* no Brasil**
João Antonio Lorençone, Lucas E. de O. Aparecido, Pedro Antonio Lorençone, Guilherme Botega Torsoni 374
6 **PDAgro: Uma proposta de protocolo para *compliance* à LGPD**
Samara B. Marques, Aline P. A. Lisboa, Érico M. H. do Amaral, Vinícius do N. Lampert 378
7 **Redes neurais na simulação de condições climáticas potenciais ao cultivo de bambu no Brasil**
João Antonio Lorençone, Lucas E. de O. Aparecido, Pedro Antonio Lorençone, Guilherme Botega Torsoni 382



A methodology for detection and localization of fruits in apples orchards from aerial images

Thiago T. Santos[1], **Luciano Gebler**[2]

[1]Embrapa Informática Agropecuária
Caixa Postal 6041 – CEP 13083-886 – Campinas, SP – Brazil

[2]Embrapa Uva e Vinho
Caixa Postal 177 – CEP 95200-000 – Vacaria, RS – Brazil

{thiago.santos,luciano.gebler}@embrapa.br

Abstract. *Computer vision methods based on convolutional neural networks (CNNs) have presented promising results on image-based fruit detection at ground-level for different crops. However, the integration of the detections found in different images, allowing accurate fruit counting and yield prediction, have received less attention. This work presents a methodology for automated fruit counting employing aerial-images. It includes algorithms based on multiple view geometry to perform fruits tracking, not just avoiding double counting but also locating the fruits in the 3-D space. Preliminary assessments show correlations above 0.8 between fruit counting and true yield for apples. The annotated dataset employed on CNN training is publicly available.*

1. Introduction

Crop monitoring is essential for anomaly detection, yield prediction and risk assessment in agriculture, basing the farmer's interventions. A continuous data collection during the fruits' growth cycle would allow an accurate modeling of its development, identifying anomalies and bottlenecks. Recently, convolutional neural networks [LeCun et al. 2015] have been employed for ground-level, image-based detection for different fruits [Sa et al. 2016], as apples [Häni et al. 2020] and mangoes [Bargoti and Underwood 2017]. However, just a few works [Liu et al. 2019, Häni et al. 2020, Santos et al. 2020] have addressed the *data association* problem in fruit counting: how to properly integrate the detections found in multiple images for accurate, row-level fruit tracking.

The present work describes a methodology for detecting and locating apples in orchards from aerial images sequences. This methodology allows not only the detection of

fruits in the images, but also their association between images, identifying apples already observed previously, an essential requirement for fruit counting. The identified apples are properly mapped in the three-dimensional space, enabling the analysis of the variability in the field. The present methodology was able to produce promising results from aerial images of about 1 cm per pixel, thus being an alternative for autonomous monitoring of entire plots in orchards by unmanned aerial vehicles (UAVs).

2. Materials and methods

The data employed in the development of the methodology came from a plot located at the Embrapa's Temperate Climate Fruit Growing Experimental Station at Vacaria-RS (28°30'58.2"S, 50°52'52.2"W). The plot, seen in Figure 1 (a), is composed of 10 rows of apple trees, of which the 8 inner rows contain the plants of interest (the first and last rows are border ones). The rows contain plants of the varieties *Fuji* (west facing) and *Gala* (east facing). The images were taken during December 13, 2018. For aerial shots, an UAV (DJI Phantom 4 Pro) performed a 12 m height flight over the orchard's rows, capturing imagery data in the form of a 4K resolution video (3840×2160 pixels). The camera tilt is not nadir, allowing a more extensive view of the canopy if compared to a top/nadir one. The terms *frame* and *image* will be employed interchangeably in this text.

2.1. Methodology

The methodology consists of three steps. The first one is apple detection performed on each image, using a deep convolutional neural network [LeCun et al. 2015]. The second step estimates the camera position and orientation at each frame, using the structure-from-motion framework from multiple view computer vision [Hartley and Zisserman 2003, Schönberger and Frahm 2016]. The last step, the main contribution in this work, uses projective geometry and directed graphs to represent multiple alternative associations between fruits observed in different frames. Each *path* in the graph represents an association hypothesis, determining the location of the same fruit in different images, and a greedy algorithm is used to choose the paths.

2.1.1. Apple detection

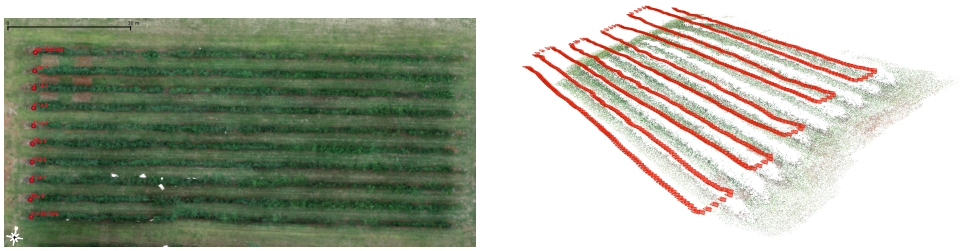
To this task, we have built an annotated dataset, formed by random selected 256×256 pixels samples from the frames extracted from the UAV video sequences. The dataset was split in training and test subsets for supervised machine learning, as shown in Table 1. This dataset is publicly available[1].

Table 1. Dataset for image-based apple detection training and evaluation.

	Number of images	Number annotated apples
Training	1025	2204
Test	114	267

For apple detection, we have employed a Faster R-CNN network [Ren et al. 2017], using a ResNet-50 backbone [He et al. 2016]. We employed the

[1]Available at `https://doi.org/10.5281/zenodo.5586329`.



(a) (b)

Figure 1. The orchard. (a) The plot presenting 10 rows. (b) Structure-from-motion performed by COLMAP - the UAV pose at capture time for each image is shown by the red frustums.

implementation available in PyTorch [Paszke et al. 2019] (see the `torchvision` library). The details of the training process, including data augmentations techniques, optimizer, batch sizes, number of epochs and hyperparameters can be seen in the publicly available code[2] and, due to text size restrictions, they will not be described here.

2.1.2. Relative camera pose estimation

To estimate the camera position at the time of capture for each video frame, we have employed the Structure-from-Motion (SfM) system COLMAP [Schönberger and Frahm 2016]. A SfM system estimates the *projection matrix* $\mathtt{P}_i$, a 3×4 matrix, for each image i: for each three-dimensional point $\mathbf{X} = (X, Y, Z, 1)^\intercal$ in the field, its 2-D projection $\mathbf{x}_i = (x_i, y_i, 1)^\intercal$ on the image plane of frame i can be computed[3] by the product

$$\mathbf{x}_i = \mathtt{P}_i \mathbf{X}. \tag{1}$$

The matrices $\mathtt{P}_i$ also allow the computation of the relative position $\mathbf{C}_i$ between cameras in the 3-D space by the property $\mathtt{P}_i \mathbf{C}_i = 0$. Figure 1 (b) illustrates the position of the UAV camera at the time of each frame capture in the flight over the plot.

2.1.3. Data association: tracking apples in the frames sequence

Projections matrices $\mathtt{P}_i$ and $\mathtt{P}_j$ allow the computation of the *fundamental matrix* $\mathtt{F}_{\mathtt{i,j}}$ [Hartley and Zisserman 2003]. Suppose that a point $\mathbf{X}$ in the 3-D space is mapped to the 2-D points $\mathbf{x}_i$ and $\mathbf{x}_j$ on the i-th and the j-th frames of the video sequence, respectively. The fundamental matrix maps $\mathbf{x}_i$ on frame i to a *epipolar line* $\mathbf{l}_{i,j}$ on frame j that contains $\mathbf{x}_j$. In our apple tracking problem, we have $\mathbf{x}_i^{(m)}$, the centroid of the m-th apple detected by the neural network on frame i. We can employ the fundamental matrix linking frames i and j to aid us in choosing the most suitable detections to correspond to $\mathbf{x}_i^{(m)}$, as seen in Figure 2.

[2]Available at `https://github.com/thsant/add256-fastercnn`.

[3]Points $\mathbf{X}$ and $\mathbf{x}_i$ are in *homogenous coordinates*, what explains the 1 in their last dimension.

Consider the centroids of the N apples detected by the neural network on the j-th frame, $\mathbf{x}_j^{(n)}, n = 1..N$. The detection corresponding to apple $\mathbf{x}_i^{(m)}$ should be close[4] to the line $\mathbf{l}_{i,j}^{(m)}$ in frame j, given by

$$\mathbf{l}_{i,j}^{(m)} = \mathtt{F_{i,j}} \cdot \mathbf{x}_i^{(m)}. \tag{2}$$

The fundamental matrix can be computed from the projection matrices by

$$\mathtt{F_{i,j}} = [\mathbf{e}_j]_\times \mathtt{P}_j \mathtt{P}_i^+, \tag{3}$$

where $\mathtt{P}_i^+$ is the pseudo-inverse of $\mathtt{P}_i$, and $\mathbf{e}_j = \mathtt{P}_j \mathbf{C}_i$ is the *epipole*, with $\mathtt{P}_i \mathbf{C}_i = 0$, i.e., $\mathbf{C}_i$ is projection center for the camera in frame i [Hartley and Zisserman 2003].



Figure 2. Epipolar restriction. Detected apples are shown as magenta 'x' markers. The point corresponding to the apple marked in red on Frame 350 defines the red epipolar line seen in Frame 351. The same apple should be observed near this line, limiting the number of options for apple tracking.

Our proposed apple tracking algorithm employs a *graph*, G, to represent multiple fruit associations hypothesis. Each *node* $v_i^{(m)} \in G$ corresponds to the centroid $\mathbf{x}_i^{(m)}$ of the m-th apple detected on a frame i. We add an *edge* $v_i^{(m)} \rightarrow v_j^{(n)}$ iif

$$\text{dist}(\mathbf{x}_j^{(n)}, \mathbf{l}_{i,j}^{(m)}) = \frac{\mathbf{x}_j^{(n)} \cdot \mathbf{l}_{i,j}^{(m)}}{\sqrt{a^2 + b^2}} \leq \tau_{\text{epipolar}}, \tag{4}$$

being $\mathbf{l}_{i,j}^{(m)} = (a, b, c)^\intercal = \mathtt{F_{i,j}} \cdot \mathbf{x}_i^{(m)}$. In other words, we are testing if the distance between the point and the epipolar line is below a threshold τ_{epipolar}. This procedure is performed by the lines 4–9 in Algorithm 1, FRUITASSOCIATION. So, *an edge in G represents a possible association between two detections in different frames*. As seen in line 5, for each frame i, the following k frames are evaluated for associations, what provides robustness to momentaneous misdetections of a fruit by the neural network.

A sequence of edges $v_i^{(m)} \rightarrow v_j^{(n)} \rightarrow \ldots \rightarrow v_k^{(o)}$ is a *path*. *Each path represents a possible association hypothesis for a fruit detected in frame i and the fruits detected in the following frames.* Lines 10-16 in Algorithm 1 implement a path selection process, employing a second algorithm, FRUITESTIMATION3D (Algorithm 2).

Algorithm 2 starts performing a depth-first search (DFS) from node $v_i^{(m)}$, getting all possible paths starting at $v_i^{(m)}$. An algorithm based on *random sample consensus*

[4]Ideally, in a noisy-free, perfect detection scenario, $\mathbf{x}_j^{(n)} \in \mathbf{l}_{i,j}^{(m)}$, i.e., $\mathbf{x}_j^{(n)} \cdot \mathbf{l}_{i,j}^{(m)} = 0$.

Data: The detected apples' centroids $\mathbf{x}_i^{(m)}$ for each frame i, $i = 1..F$
Result: A set of 3-D points (apples centers) $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_L\}$ and their tracks

begin
  for $i \leftarrow 1$ **to** F **do**
    foreach *detected apple* $\mathbf{x}_i^{(m)}$ *in* i **do** Add node $v_i^{(m)}$ to G
  for $i \leftarrow 1$ **to** F **do**
    for $j \leftarrow i + 1$ **to** $\min(i + k, F)$ **do**
      for *all* $\mathbf{x}_i^{(m)}$ *and* $\mathbf{x}_j^{(n)}$ **do**
        if $\mathrm{dist}(\mathbf{x}_j^{(n)}, \mathbf{l}_{i,j}^{(m)}) \leq \tau_{\text{epipolar}}$ **then**
          Add the edge $v_i^{(m)} \rightarrow v_j^{(n)}$ to G
  $\mathcal{X} \leftarrow \emptyset$
  for $i \leftarrow 1$ **to** F **do**
    for *each* $v_i^{(m)}$ **do**
      $\mathbf{X}, \mathcal{I}, r_{\mathcal{I}} \leftarrow$ FRUITESTIMATION3D(G, $v_i^{(m)}$)
      if $\mathbf{X} \neq$ *NIL* **then**
        Add $\mathbf{X}$ to $\mathcal{X}$
        Remove from G all edges $v_i^{(m)} \rightarrow v_j^{(n)}$ such that $v_i^{(m)}, v_j^{(n)} \in \mathcal{I}$
  return $\mathcal{X}$

Algorithm 1: FRUITASSOCIATION.

(RANSAC) [Fischler and Bolles 1981] is employed to estimate the tridimensional point $\mathbf{X}$ corresponding to a path (an apple's 3-D position in space). At each iteration, the TRIANGULATIONRANSAC algorithm pick three 2-D points, $\mathbf{x}_i^{(m)}$, $\mathbf{x}_j^{(n)}$ and $\mathbf{x}_k^{(o)}$ (corresponding to nodes $v_i^{(m)}$, $v_j^{(n)}$ and $v_k^{(o)}$ in a path T) and estimates the 3-D point $\mathbf{X}$. The estimation of X is performed by a least-squares minimization algorithm [Hartley and Sturm 1997]. Next, $\mathbf{X}$ is projected on each frame i in the path, defining the points $\hat{\mathbf{x}}_i^{(m)} = \mathtt{P}_i\mathbf{X}$ and their corresponding *geometrical errors*, i.e., the Euclidean distance between $\hat{\mathbf{x}}_i^{(m)}$ and $\mathbf{x}_i^{(m)}$. Nodes in the path T whose geometrical error is below the threshold τ_{geom} are considered *inliers*. At each iteration, the RANSAC procedure keeps the point $\mathbf{X}$ that delivered the largest number of inliers. Algorithm 2 looks for the path presenting the largest rate of inliers $\mathcal{I}$, keeping the longest path presenting the largest inlier rate. In other words, the inlier ratio acts as a quality measure for the inter-frame association hypothesis regarding fruit $\mathbf{x}_i^{(m)}$, represented by a path starting from $v_i^{(m)}$. Once a path is selected, the algorithm remove its edges from G (line 16 in Algorithm 1), avoiding those associations to be employed again. However, the nodes are preserved in the graph, allowing *fruits occlusions* to be considered: fruits that occlude each other can create crossing paths in G, i.e., paths sharing nodes.

Data: An association graph G and a initial node $v_i^{(m)}$.
Result: The 3-D apple position $\mathbf{X}$, a set of inliers nodes $\mathcal{I}$, and the inlier ratio $r_{\mathcal{I}}$.

1 **begin**
2 $\quad\mathcal{T} \leftarrow \text{DFS}(G, v_i^{(m)})$
3 $\quad\mathbf{X} \leftarrow \text{NIL}$
4 $\quad\mathcal{I} \leftarrow \emptyset$
5 $\quad r_{\mathcal{I}} \leftarrow 0$
6 $\quad$**for** *each track* $T = \langle v_i^{(m)} \rightarrow v_j^{(n)} \rightarrow \ldots \rangle \in \mathcal{T}$*, from the longest to the shortest* **do**
7 $\quad\quad\mathbf{X}_T, \mathcal{I}_T \leftarrow \text{TRIANGULATIONRANSAC}(T)$
8 $\quad\quad$**if** $\mathcal{I}_T \neq \emptyset$ **then**
9 $\quad\quad\quad r_T \leftarrow \frac{\|\mathcal{I}_T\|}{\|T\|}$
10 $\quad\quad\quad$**if** $r_T > r_{\mathcal{I}}$ **then**
11 $\quad\quad\quad\quad r_{\mathcal{I}} \leftarrow r_T$
12 $\quad\quad\quad\quad\mathbf{X} \leftarrow \mathbf{X}_T$
13 $\quad\quad\quad\quad\mathcal{I} \leftarrow \mathcal{I}_T$
14 $\quad$**return** $\mathbf{X}, \mathcal{I}, r_{\mathcal{I}}$

Algorithm 2: FRUITESTIMATION3D.

3. Results and discussion

Figure 3 displays the tracks determined by Algorithm 2 for two different apples. Each line in the figure corresponds to an apple's track (only the *inliers*). Note how the look of the fruit and its surroundings varies slightly as the pose (the UAV position) changes from frame to frame. Each track determines the three-dimensional position of an apple:

all *inliers* are used in the final estimation of the fruit's position $\mathbf{X}$ in the 3-D space, again by employing the least-squares algorithm [Hartley and Sturm 1997]. Figure 4 displays a total of 9,237 apples found in the plot. Fruits were automatically divided into the ten rows of the field by K-means clustering.

Disregarding two rows out of the UAV's field of view, caused by imprecision in the vehicle positioning system[5], the observed linear correlation between the counted apples in each row and the row's yield was 0.11 for *Fuji* and 0.80 for *Gala*, considering six rows. However, one of the rows (row 8) looks like a severe outlier: considering just the other five rows, linear correlation is 0.93 for *Fuji* and 0.88 for *Gala*. Although promising, the results should be viewed with caution, given that few rows were evaluated, at a single plot. More extensive experiments are yet necessary for a full characterization of uncertainty in yield prediction and the proportion of the fruits that is visible in imagery. It should also be noted that the images were captured in December and the harvest was carried out in February of the following year, which indicates that the methodology has the potential to provide yield estimated in early stages. Indeed, the presented methodology can be employed as a component of a more sophisticated yield prediction system.

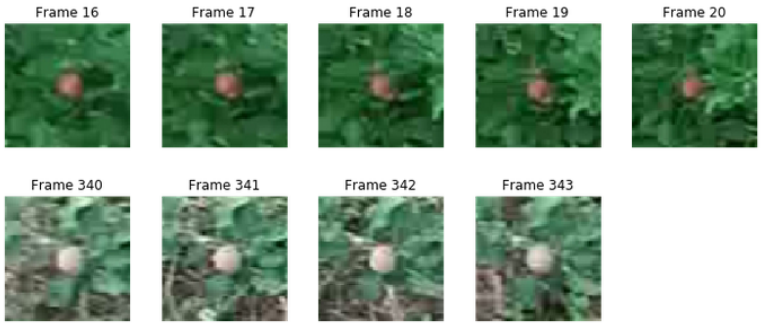


Figure 3. Two inter-frame fruit association examples found by Algorithm 1. Each row corresponds to an apple, observed in a few frames. Note the inter-frame variations, caused by UAV's pose changes during recording.

4. Conclusions

Fruit detection and tracking can be, in short term, applied to yield prediction and crop monitoring. In the long term, precise detection and 3-D localization can be employed on harvesting by autonomous agents. Detection and tracking allow autonomous agents to estimate their position relative to the fruit, so that accurate handling planning can be performed by the machine. The three-dimensional localization can also characterize the spatial variability of the fruits in the plots, helping on growing management according to precision agriculture practices.

The presented methodology is not restricted to aerial images: the same algorithms could be adapted to images obtained by ground vehicles with embedded cameras. Autonomous aerial vehicles with precise positioning control, such as devices equipped with

[5]Precise flights, able to keep the plants in the UAV's field of view, can be performed by vehicles presenting a precise position control, as a Real-Time Kinematic (RTK) Global Navigation Satellite System (GNSS). Unfortunately, the vehicle used in this work presented an ordinary GNSS system, without the positioning corrections provided by RTK.

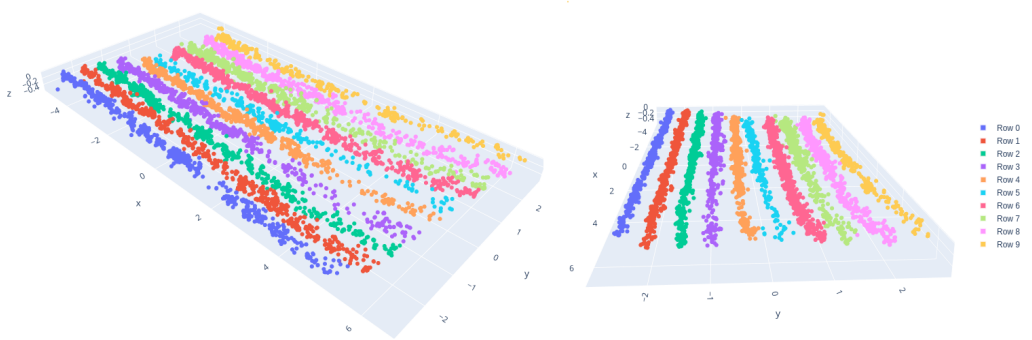


Figure 4. Fruit automatic localization. Each point represents the three-dimensional location determined for an apple. Colors represent different lines in the field, automatically identified using the K-means algorithm.

RTK GNSS, could be used as a row-scanning system able to perform automated field monitoring. New experiments, with a greater variability of plants, management regimes and plant architectures, should be carried out to validate and adapt the methodology for operation in different scenarios, and provide a better characterization of the estimation errors in yield prediction.

Acknowledgments

This work was supported by Brazilian Agricultural Research Corporation (Embrapa) under grant 01.14.09.001.05.04 and by FAPESP under grant (2017/19282-7).

References

Bargoti, S. and Underwood, J. (2017). Deep fruit detection in orchards. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3626–3633. IEEE.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition.

Hartley, R. I. and Sturm, P. (1997). Triangulation. *Computer vision and image understanding*, 68(2):146–157.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Häni, N., Roy, P., and Isler, V. (2020). A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics*, 37(2):263–282.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Liu, X., Chen, S. W., Liu, C., Shivakumar, S. S., Das, J., Taylor, C. J., Underwood, J., and Kumar, V. (2019). Monocular Camera Based Fruit Counting and Mapping With Semantic Data Association. *IEEE Robotics and Automation Letters*, 4(3):2296–2303.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. *Sensors*, 16(8).

Santos, T. T., de Souza, L. L., dos Santos, A. A., and Avila, S. (2020). Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Computers and Electronics in Agriculture*, 170:105247.

Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.



AgroDigital-NE: Um mapeamento da produção científica sobre tecnologias de informação e comunicação aplicadas ao agronegócio no Nordeste brasileiro

Laura Emmanuella A. dos S. Santana[1], Leonardo Rodrigues de Lima Teixeira[1], Alixandre Santana[2], Josenalde Barbosa de Oliveira[1,3]

[1]Escola Agrícola de Jundiaí – Universidade Federal do Rio Grande do Norte (UFRN)
Macaíba – RN – Brasil

[2]Universidade Federal do Agreste de Pernambuco (UFAPE)
Garanhuns – PE – Brasil

[3]Instituto de Engenharia de Sistemas, Computadores, Pesquisa e Desenvolvimento do Brasil (INESC P&D Brasil)
Santos - SP - Brasil
{laura.santana,leonardo.teixeira, josenalde.oliveira}@ufrn.br,
alixandre.santana@ufape.edu.br

Abstract. *Technological innovation in rural areas and the consequent advance of digital agriculture demand research in computing applied to this sector. Due to its continental size, in Brazil, these technologies and related research need to be aligned with regional issues such as economy, climate, and production factors. In this sense, this paper presents a survey of the scientific work produced by authors from the northeast of Brazil, which encompassed computing technologies and their respective targeted agribusiness areas. In addition, we also listed teaching and research institutions involved in the subject. By making an inventory of such information, we expect to foster collaboration networks between researchers, institutions, and the productive sector in the northeast region of Brazil.*

Resumo. *A inovação tecnológica no campo e consequente avanço da agricultura digital demanda pesquisa em computação aplicada ao setor. Visto que cada região tem especificidades ambientais, econômicas e de insumos em relação à produção agrária, são necessárias pesquisas específicas a essas demandas. Este artigo apresenta o resultado de levantamento da produção científica sobre TICs aplicadas ao agronegócio na Região Nordeste do Brasil, com o objetivo de mapear as áreas de aplicação e as técnicas usadas, bem como as instituições de ensino e pesquisa que têm se dedicado ao tema. Ao inventariar*

tais informações, acreditamos contribuir para a formação futura de redes de colaboração entre pesquisadores, instituições e o setor produtivo.

1. Introdução

O setor agrícola influencia de forma significativa a economia do Brasil. Segundo dados do Ministério da Agricultura, Pecuária e Abastecimento (BRASIL, 2021a), as safras dos últimos dois anos têm alcançado sucessivamente novos recordes de produção de grãos, chegando finalmente ao número de 262,13 milhões de toneladas de grãos. O Brasil fornece alimentos para mais de 150 países e a produção de origem animal e vegetal ultrapassa 400 produtos provenientes da agricultura (EMBRAPA, 2018).

A tecnologia empregada no campo foi determinante para que a agricultura brasileira alcançasse o patamar atual (MASSRUHÁ; LEITE, 2017), permitindo a modernização da prática agropecuária para melhorar o desempenho qualitativo e quantitativo. O uso de sensores e drones agregados a bases de dados analisadas com algoritmos inteligentes para o tratamento e gerenciamento de dados agrícolas, transformando-os em informação útil ao agricultor ou à pesquisa agropecuária tem feito parte da chamada agricultura 4.0 ou agricultura digital (RIBEIRO; MARINHO; ESPINOSA, 2018; QUEIROZ et al., 2020).

No entanto, é possível observar que as principais indústrias de máquinas e implementos agrícolas, bem como os principais grupos de pesquisa em agricultura de precisão e TICs aplicadas ao agronegócio (agroinformática) encontram-se na Região Centro-Sul do país, como por exemplo, Embrapa Informática Agropecuária e programas de mestrado em tecnologias computacionais para a agricultura — na Universidade Federal Tecnológica do Paraná (UTFPR), na Universidade Federal de Santa Maria (UFSM), na Universidade Federal do Pampa (UNIPAMPA) e na Universidade Federal do Mato Grosso do Sul (UFMS); um programa com mestrado e doutorado na Universidade Estadual de Ponta Grossa (UEPG) (CAMARGO, 2019) e um programa de especialização em Agronegócio Digital na Universidade Federal do Paraná (UFPR).

A agricultura praticada no Nordeste brasileiro (NE) é muito variada, seja com relação às culturas plantadas, seja com relação a aspectos como o nível de tecnologia empregada na produção agrícola (CASTRO, 2013). No estudo de Castro citado, o autor mostra que a tecnologia empregada na produção regional é, em muitos casos, defasada com relação àquela empregada em atividades congêneres desenvolvidas no restante do país. Contudo este cenário está em constante transformação e evolução, e estudo do Ministério da Agricultura, Pecuária e Abastecimento (BRASIL, 2021b) indica em dez anos uma produção de 36 milhões de toneladas de grãos com 9,3 milhões de hectares plantados na região denominada MATOPIBA, com três estados da Região NE. Outro cinturão agrícola de destaque é a viticultura irrigada no Vale do São Francisco (ARRUDA, 2017) com várias vinícolas instaladas na região de Petrolina/PE e Juazeiro/BA, bem como a exportação recorde de melão com destaque para países asiáticos, no estado do Rio Grande do Norte. A finalização próxima da transposição das águas do Rio São Francisco agregará ainda melhor infraestrutura ao desenvolvimento agrícola do NE. Uma produção e divulgação científica que reflita este avanço é, contudo, função da organização de grupos de pesquisa nas instituições de ensino e principalmente programas de pós-graduação stricto e lato sensu.

Em participação recente no principal evento científico da área de Informática aplicada à Agricultura no Brasil, Congresso Brasileiro de Agroinformática (SBIAGRO, 2019) foi observada a baixa participação de pesquisadores da Região Nordeste, com apenas 7 trabalhos apresentados, 4 dos quais apresentados pela representação da Universidade Federal do Rio Grande do Norte (UFRN). Essa observação também foi apresentada no trabalho de Camargo et al. (2017), onde os autores informam que "Artigos produzidos por autores principais das Regiões Norte e Nordeste foram raros", referindo-se a trabalhos publicados no SBIAgro 2015. Foi possível ainda observar a diferença dos problemas existentes na Região NE (como as questões edafoclimáticas, escassez hídrica e o próprio acesso a tecnologias mais avançadas por indisponibilidades financeiras) em relação às regiões Centro-Sul, com maior presença no evento. Este fato motivou o grupo de pesquisa “Modelagem, Identificação e Controle de Sistemas Agrários” a investigar as iniciativas de pesquisadores do NE em soluções de TICs para o agronegócio.

O pressuposto do presente trabalho é que, num primeiro momento, é preciso realizar um levantamento das expertises e pessoas de interesse nas instituições de pesquisa da Região NE, para que de posse dessas informações, pesquisadores, governos e instituições interessadas no tema possam se conhecer e possivelmente se articular em prol de eventuais colaborações de pesquisa e/ou desenvolvimento, delineando assim o atual ecossistema de pesquisa em agroinformática do NE.

Nesse sentido, o objetivo deste estudo é mapear a produção científica sobre TICs aplicadas ao agronegócio na Região Nordeste do Brasil, com base nas publicações dos principais congressos da área no país. O estudo pretende ainda identificar e categorizar os temas pesquisados (aplicações e técnicas) e as instituições participantes das pesquisas.

2. Material e Métodos

Esta pesquisa se caracteriza como descritivo-exploratória (WOHLIN e AURUM, 2014) e tem como método de pesquisa a revisão bibliográfica que procura responder a seguinte questão de pesquisa principal: *Quais são as principais tecnologias computacionais aplicadas no contexto do agronegócio na Região do Nordeste brasileiro sendo estudadas e desenvolvidas por pesquisadores desta região, bem como, quais as áreas de aplicações mais exploradas?*

Para o desenvolvimento desta pesquisa foi definido o método de coleta de dados não automático em fontes selecionadas. Primeiro foi realizado o levantamento dos trabalhos publicados por pesquisadores de instituições nordestinas nos anais disponíveis dos principais congressos nacionais de tecnologia aplicada ao agronegócio nos últimos 15 anos. Neste levantamento, foi registrado o título do trabalho, pesquisadores, instituições e palavras-chaves dos trabalhos. Os congressos nacionais considerados como estratégicos para este levantamento por suas abrangências, históricos e públicos-alvo foram: Congresso Brasileiro de Agroinformática (SBIAgro) e Congresso Brasileiro de Agricultura de Precisão (ConBAP), organizados respectivamente pela Associação Brasileira de Agroinformática, estabelecida no ano de 1996, e a Associação Brasileira de Agricultura de Precisão, estabelecida no ano de 2016. De acordo com o tempo determinado de 15 anos, os anais levantados do SBIAgro foram: 2005, 2007, 2009, 2011, 2013, 2015, 2017 e 2019. Já para o ConBAP, apenas os anais de 2014 (CONGRESSO..., 2014), 2016 e 2018 (ASSOCIAÇÃO..., 2020) estavam disponíveis na Internet. Segundo

os organizadores do evento em seu site, os anais dos anos anteriores foram publicados apenas através de CDs, aos quais não tivemos acesso.

Os critérios de inclusão para construção de dataset com o levantamento feito foram: trabalhos com a) pelo menos um dos autores com afiliação a uma instituição de pesquisa sediada no NE (ex. UFRN, Embrapa Semiárido etc); b) contendo uma tecnologia computacional aplicada (ex. aprendizado de máquina) a um problema dentro do contexto do agronegócio (ex. detecção de pragas em folhas). Como resultado, foram identificados 73 artigos de um total de 1066 artigos analisados para compor o dataset final. A partir deste dataset, foram extraídas as tecnologias computacionais e respectivas áreas de aplicação no agronegócio e demais metadados de interesse. Na sequência, aplicamos uma análise de categorias (WOHLIN e AURUM, 2014) sobre os construtos identificados durante a extração de dados.

3. Resultados e Discussão

A partir do levantamento feito, pôde-se confirmar a observação inicial deste estudo sobre a baixa participação de pesquisadores de instituições do Nordeste brasileiro nos eventos de agroinformática levantados, uma vez que dos 776 trabalhos analisados no SBIAgro, apenas 50 foram publicados por autores de instituições do NE, equivalente a 6,44%. Já no ConBAP, dos 290 trabalhos levantados, apenas 23 são de instituições do NE, o que equivale a 7,93%.

A Figura 1 mostra o quantitativo de trabalhos de instituições do NE nas edições do SBIAgro e ConBAP.

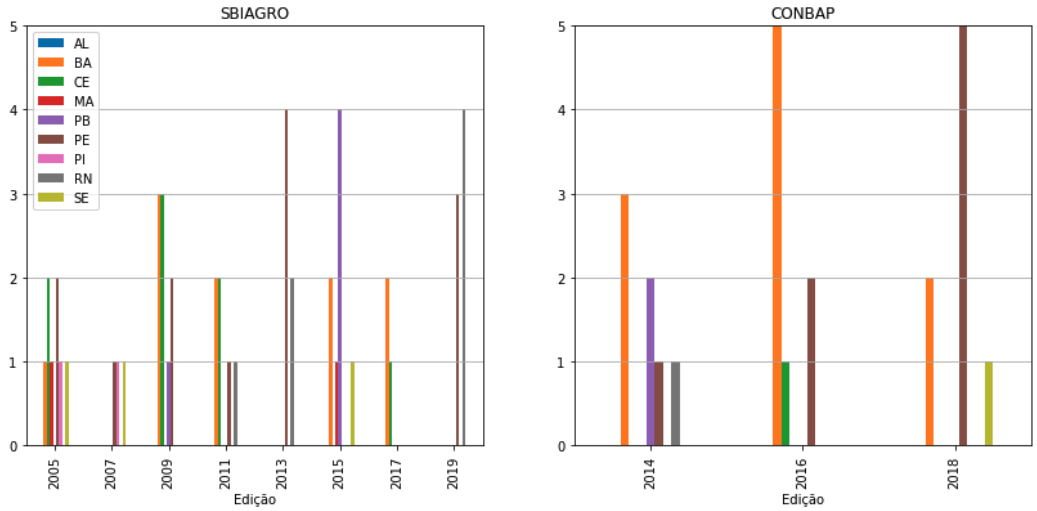


Figura 1. Trabalhos de pesquisadores do NE nas edições do SBIAGRO e do ConBAP

De acordo com esse levantamento, observa-se que os estados de Pernambuco e Bahia são os que têm mais publicações, 21 e 20 respectivamente, do total de 73, seguidos por Ceará (9), Rio Grande do Norte (8) e Paraíba (7). Os estados com menor participação são Sergipe (4), Maranhão (2) e Piauí (2). Alagoas não aparece nesse levantamento. Uma outra visualização é apresentada na Figura 2, onde o mapa de calor indica o acumulado de trabalhos de cada estado nos dois eventos analisados.

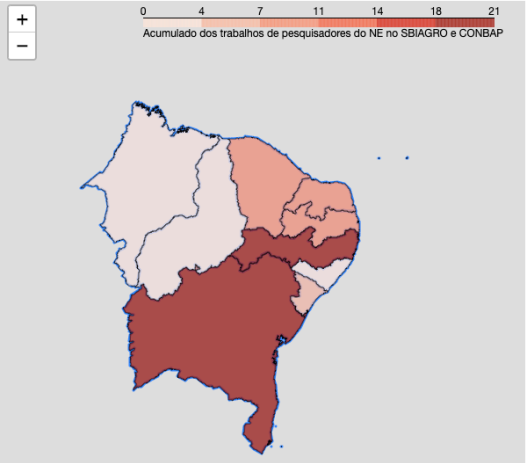


Figura 2. Acumulado dos trabalhos de pesquisadores do NE no SBIAGRO e ConBAP

A Tabela 1 mostra quais são as instituições publicadoras nos eventos em estudo e as respectivas quantidades de publicações de cada uma. Pode-se observar que o estado com mais instituições com publicação nos eventos levantados é a Bahia com 7 instituições, seguido do estado de Pernambuco com 5.

Tabela 1. Instituições e quantitativo de publicações no SBIAgro e ConBAP no período analisado

UF	INSTITUIÇÃO
BA	IFBA (2), UESC (11), Embrapa Mandioca e Fruticultura (1), UFRB (1), UNEB (1), IF Baiano (1), UESB (2) e UNIVASF (1)
CE	Embrapa Caprinos e Ovinos (1), UFC (6), UECE (1) e Faculdade Integrada do Ceará (1)
MA	IFMA (1) e UFMA (1)
PB	UFPB (5) e UFCG (3)
PE	UFAPE (3), UFRPE (7), Embrapa Semiárido (3), UPE (2) e UNIVASF (7)
PI	Embrapa Meio Norte (1) e UFPI (1)
RN	UFRN (8), IFRN (1) e UFERSA (1)
SE	UFS (1) e Embrapa Tabuleiros Costeiros (3)

Em relação às áreas de aplicação e técnicas adotadas nos trabalhos levantados, obtidas a partir do resumo e palavras-chaves de cada trabalho, as Figuras 3 e 4 apresentam o compilado.

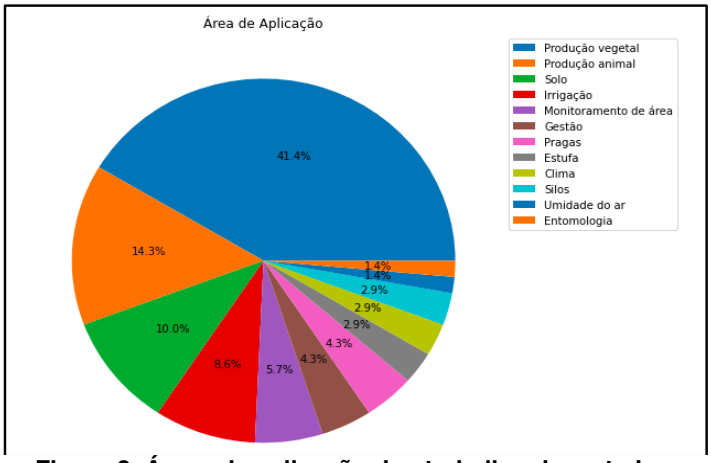


Figura 3. Áreas de aplicação dos trabalhos levantados

Pode-se observar pela Figura 3 que a área de aplicação mais presente nos trabalhos é a produção vegetal, seguida da produção animal e solo. Irrigação aparece em 4º lugar. A produção vegetal como 1ª área de aplicação tem sintonia com o fato de o NE ser forte produtor na área de fruticultura, conforme discutido na Seção 1.

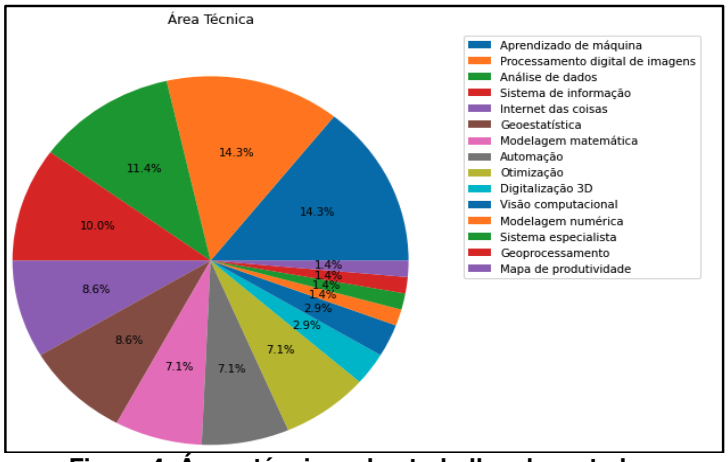


Figura 4. Áreas técnicas dos trabalhos levantados

A Figura 4 mostra que as técnicas mais usadas nos trabalhos levantados são aprendizado de máquina, processamento digital de imagens e análise de dados, seguido de sistemas de informação.

O resultado compilado e detalhado deste levantamento foi organizado em um dataset e disponibilizado em repositório público (SANTANA, 2021), juntamente com código de análise de dados. A partir desses dados, é possível identificar, por exemplo, onde e quais pesquisadores trabalham com visão computacional e fruticultura, bem como outros cruzamentos de dados entre tecnologias computacionais e área de aplicação, que por questões de espaço não foram aqui descritas.

4. Conclusão

A pesquisa aqui apresentada teve por objetivo mapear a produção científica sobre TICs aplicadas ao agronegócio na Região NE do Brasil, com base nas publicações dos

principais congressos da área no país. Além disso, o estudo buscou identificar e categorizar os temas pesquisados (aplicações e técnicas) e as instituições participantes das pesquisas.

De acordo com os resultados apresentados, pôde-se confirmar a observação inicialmente levantada sobre a baixa participação de pesquisadores de instituições do NE nos eventos em análise quando comparada com as demais regiões do país, sabidamente a Região Centro-Sul. Dentre as publicações encontradas, observou-se maior representação de instituições da BA e PE. Em relação aos temas dessas publicações, o levantamento mostrou maior percentual de publicações sobre produção vegetal e técnicas de processamento digital de imagens e aprendizado de máquina.

Algumas inferências em relação ao levantamento feito indicam que BA e PE possuem grupos de pesquisa relacionados de alguma maneira. Esse resultado é coerente com a presença relevante da BA no MATOPIBA, tendo MA e PI menor expressão no levantamento feito. Dos 8 estados com produção nos congressos indicados no período de estudo, 5 publicaram em parceria com unidades Embrapa e com um número baixo de publicações conjuntas. Tal observação aponta para uma baixa interação entre as universidades e institutos federais e as unidades da Embrapa presentes na região.

É importante considerar que os pesquisadores do NE podem ter escolhido outras formas de dar publicidade aos seus trabalhos, para além do ConBAP e SBIAgro. Por exemplo, podem optar por outros locais de publicação ou focam na geração de produtos ou soluções e patentes. Por fim, é importante também observar que a baixa taxa de publicação nos eventos levantados pode estar relacionada com o fato de a região não possuir programas de pós-graduação diretamente relacionados, apesar de existir graduação com foco na aplicação das TICs no agronegócio, como o curso de Análise e Desenvolvimento de Sistemas da UFRN (SANTANA e OLIVEIRA, 2021).

Como trabalhos futuros, pretende-se identificar junto a governos estaduais as respectivas vocações regionais econômicas no contexto do agronegócio, procurando relacioná-las com as expertises já mapeadas no ecossistema de agroinformática neste trabalho ou ainda, mapear tecnologias computacionais utilizadas em outros ecossistemas que possam ser transferidas e aplicadas na Região Nordeste. Pesquisadores e demais interessados são encorajados a entrar em contato com os autores do presente trabalho para discutir ações que promovam a pesquisa de computação aplicada ao agronegócio no contexto do Nordeste brasileiro.

Referências

ARRUDA, A. C. A fruticultura irrigada do Vale do submédio do São Francisco e a modernização agrícola do Sertão. **Boletim Estudos e Pesquisas, n. 02** - SEBRAE/PE, 2017.

ASSOCIAÇÃO BRASILEIRA DE AGRICULTURA DE PRECISÃO. Publicações. 2020. Disponível em: <https://asbraap.org/index.php?page=anais>. Acesso em: 13 ago. 2021.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Safra de grãos deve chegar a 262,13 milhões de toneladas. Brasília, 2021a. Disponível em: <https://www.gov.br/pt-br/noticias/agricultura-e-pecuaria/2021/06/safra-de-graos-deve-chegar-a-262-13-milhoes-de-toneladas>. Acesso em: 11 jul. 2021.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Nota n. 15/2021 CGAPI/DCI/SPA/MAPA: Projeções do Agronegócio: projeções de longo prazo Brasil 2020-21 a 2030-31.2021. Brasília, 2021b. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/producao-de-graos-crescera-27-nos-proximos-dez-anos-chegando-333-milhoes-de-toneladas/NotadasProjeesdoAgronegcio20202021a20302031.pdf>. Acesso em 13 ago. 2021.

CAMARGO, A. M. Entrevista com Maria Fernanda Moura. TECCOGS – Revista Digital de Tecnologias Cognitivas, n. 20, jul./dez. 2019. Disponível em: <https://revistas.pucsp.br/index.php/teccogs/issue/viewIssue/TECCOGS%20n.%2020/pdf_1>. Acesso em: 11 jul. 2021.

CAMARGO, S. S. et al. Um panorama da colaboração científica na pesquisa agroinformática brasileira. In: XI Congresso Brasileiro de Agroinformática, 2017, Campinas/SP. Anais Eletrônicos… p. 527-536.

CASTRO, C. N. A agricultura no Nordeste brasileiro: oportunidades e limitações ao desenvolvimento. IPEA Boletim regional, urbano e ambiental, n. 8. jul./dez. 2013. Disponível em: <https://ipea.gov.br/portal/images/stories/PDFs/boletim_regional/140423_boletimregional8_cap8.pdf >. Acesso em: 11 jul. 2021.

CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO. Edição Atual. 2014. Disponível em: <http://conbap.sbea.org.br/crbst_2.html>. Acesso em: 13 ago. 2021.

EMBRAPA. Visão 2030: O futuro da agricultura brasileira. – Brasília: Embrapa, 2018. Disponível em: <https://www.embrapa.br/visao/o-futuro-da-agricultura-brasileira>. Acesso em: 11 jul. 2021.

MASSRUHÁ, S. M. F. S.; LEITE, M. A. A. Agro 4.0 – rumo à agricultura digital. JC na Escola Ciência, Tecnologia e Sociedade: Mobilizar o Conhecimento para Alimentar o Brasil. 2 ed. São Paulo: Centro Paula Souza, 2017. Disponível em: <https://www.alice.cnptia.embrapa.br/bitstream/doc/1073150/1/PLAgro4.0JCnaEscola.pdf >. Acesso em: 11 jul. 2021.

QUEIROZ, D. M.; VALENTE, D. S. M.; PINTO, F. A. C.; BORÉM, A. (eds.) Agricultura Digital. Viçosa: UFV, 2020.

RIBEIRO, J. G.; MARINHO, D. Y.; ESPINOSA, J. W. M. Agricultura 4.0: Desafios à produção de alimentos e inovações tecnológicas. In: Simpósio de Engenharia de Produção, 2018. Catalão/GO. Disponível em: <https://sienpro.catalao.ufg.br/p/25913-anais-do-ii-sienpro-2018-issn-2594-410x>. Acesso em: 11 jul. 2021.

SANTANA, L. E. A. S. AgroDigitalNE, 2021. Disponível em: < https://github.com/lauraemmanuella/AgroDigitalNE>. Acesso em: 13 ago. 2021.

SANTANA, L. E. A. S.; OLIVEIRA, J. B. Agricultura 4.0 e o desenvolvimento de pesquisas de computação aplicada às ciências agrárias. SBC Horizontes, março. 2021. ISSN 2175-9235. Disponível em: <http://horizontes.sbc.org.br/index.php/2021/03/Agricultura-4-0-e-o-desenvolvimento-de-pesquisas-de-computacao-aplicada-as-ciencias-agrarias/>. Acesso em: 11 jul. 2021.

SBIAGRO 2019. XII Congresso Brasileiro de Agroinformática. Indaiatuba/SP, 2019. Anais Eletrônicos… Disponível em:

<https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1125706/1/PC-Anais-SBIAGRO-2019.pdf> Acesso em: 13 ago. 2021.

WOHLIN, C; AURUM, A. Towards a decision-making structure for selecting a research design in empirical software engineering. Empir Software Eng. 2014. DOI 10.1007/s10664-014-9319-7.



Análise bibliométrica da produção científica da EMBRAPA na Web of Science

Alexandre D. Alves

Centro de Matemática, Computação e Cognição – Universidade Federal do ABC (UFABC)
Av. dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP – Brasil

alexandre.donizeti@ufabc.edu.br

Abstract. *This article aims to present a bibliometric analysis of the scientific production of authors linked to EMBRAPA considering the articles published in journals indexed in the Web of Science from 1974 to 2021. During this period, 27,089 articles were published in collaboration with co-authors from 140 different countries, being the main collaboration with co-authors from the United States.*

Resumo. *Este artigo tem como objetivo apresentar uma análise bibliométrica da produção científica dos autores vinculados à EMBRAPA considerando os artigos publicados em periódicos indexados na Web of Science no período de 1974 a 2021. Nesse período foram publicados 27.089 artigos em colaboração com coautores de 140 países diferentes, sendo a principal colaboração com coautores dos Estados Unidos.*

1. Introdução

A Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) é uma instituição pública de pesquisa vinculada ao Ministério da Agricultura, Pecuária e Abastecimento. Foi criada em 26 de abril de 1973 e é uma empresa de inovação tecnológica focada na geração de conhecimento e tecnologia para a agricultura e pecuária brasileira. Ao longo desses anos, a agropecuária brasileira se tornou, com certeza, uma das mais eficientes e sustentáveis do planeta.

A produção científica da EMBRAPA ocupa lugar de destaque entre as instituições do País considerando os artigos publicados em periódicos indexados na Web of Science (WoS). A WoS é uma base de dados reconhecida mundialmente por toda a comunidade científica pela sua amplitude. Nessa base, a EMBRAPA ocupa atualmente a oitava posição com 27.089 (3,42%) artigos, ficando atrás apenas das seguintes instituições: USP (176.573; 22,27%), UNESP (61.451; 7,75%), UNICAMP

(60.826; 7,67%), UFRJ (53.919; 6,80%), UFRGS (44.294; 5,59%), UFMG (41.097; 5,18%) e UNIFESP (31.456; 3,97%).

Na revisão da literatura realizada de forma sistemática e sintética, constatou-se que há poucos trabalhos que analisam a produção científica da EMBRAPA na WoS (Penteado Filho e Avila, 2009a; Penteado Filho e Avila, 2009b; Penteado Filho e Avila, 2009c; Penteado Filho e Avila, 2010). Porém, esses trabalhos consideraram a produção científica no período de 1977 a 2006. Há também um trabalho recente que considerou a produção científica da EMBRAPA na base Scopus (Irizaga e Vanz, 2021).

O objetivo deste trabalho é analisar a produção científica dos autores vinculados à EMBRAPA a partir das informações sobre os artigos publicados em periódicos indexados na WoS no período de 1974 a 2021. Este estudo considera indicadores bibliométricos baseados em coautoria e citações, incluindo distribuição geográfica, análise temporal dos periódicos e palavras-chaves mais usadas pelos autores vinculados à EMBRAPA.

O mapeamento da produção científica de uma instituição permite contar um pouco de sua própria história, assim como pode indicar o perfil de colaboração científica nas publicações dos pesquisadores dessa instituição. Conhecer esse perfil permite definir estratégias de incentivo para a instituição.

Este texto está organizado da seguinte maneira: na Seção 2 é descrito como foram coletados os dados de interesse. Também são destacadas as ferramentas usadas na elaboração deste trabalho. Na Seção 3 são apresentados os principais resultados, discutindo alguns pontos importantes. Finalmente, na Seção 4 são apresentadas as considerações finais.

2. Materiais e Métodos

Inicialmente, foi realizada uma busca avançada por artigos da organização consolidada "*Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA)*" publicados em periódicos indexados na WoS que engloba as diversas formas que uma instituição foi informada naquela base. Para este trabalho foram considerados apenas os seguintes índices de citação: *Science Citation Index Expanded* (SCI-EXPANDED), *Social Sciences Citation Index* (SSCI) e *Arts & Humanities Citation Index* (A&HCI). Assim, foi possível obter a produção científica de autores vinculados à EMBRAPA a partir de informações das publicações (artigos) indexadas na WoS.

O total de artigos encontrados foi 27.089. A partir disso, foi possível baixar os registros desses artigos. Em seguida, foi feita a extração e os dados foram armazenados em um banco de dados. Todo esse processo, brevemente aqui descrito, foi realizado no dia 3 de setembro de 2021. As seguintes informações foram obtidas para este trabalho: artigos (id, título, ano, periódico, total de autores, total de citações, afiliações e palavras-chave). O acesso a essas informações foi feito através do Portal de Periódicos da CAPES.

O mapa apresentado neste trabalho foi gerado usando a ferramenta GPS Visualizer (acessível em http://www.gpsvisualizer.com). Os mapas de visualização foram gerados usando a ferramenta VOSviewer (Van Eck e Waltman, 2010), que permite a construção e visualização de redes bibliométricas. O acesso ao fator de

impacto (FI) de 2020 no *Journal Citation Reports*® (JCR®) também foi feito através do Portal de Periódicos da CAPES.

3. Resultados e Discussões

De acordo com os registros da WoS, o primeiro artigo da EMBRAPA foi o de Turnipseed et al. (1974), no periódico "Journal of Economic Entomology", e o mais citado é o artigo de Aljanabi e Martinez (1997), publicado no periódico "Nucleic Acids Research", com 1.752 citações. Desde o primeiro artigo registrado na WoS até 03/09/2021, os autores vinculados à EMBRAPA publicaram 27.089 artigos em periódicos indexados nesta base.

Na Figura 1 apresenta-se o número de artigos no período de 1974 a 2021 e a média de autores por ano. Nos últimos 20 anos o número de artigos aumentou quase cinco vezes, enquanto a média de autores dobrou, chegando a 8,03 autores por artigo em 2020. Ao longo dos anos a média de autores foi 6,04. Com relação a autoria dentre os 27.089 artigos, 455 (1,68%) foram publicados por um único autor, enquanto 26.634 (98,32%) foram em coautoria. Do total de artigos em coautoria, 1.789 (6,60%) artigos foram publicados por dois autores, 3.058 (11,29%) por três autores e 21.787 (80,43%) por 4 ou mais autores. O artigo com mais autores (729) foi publicado por Kattge et al. (2020), no periódico "Global Change Biology".

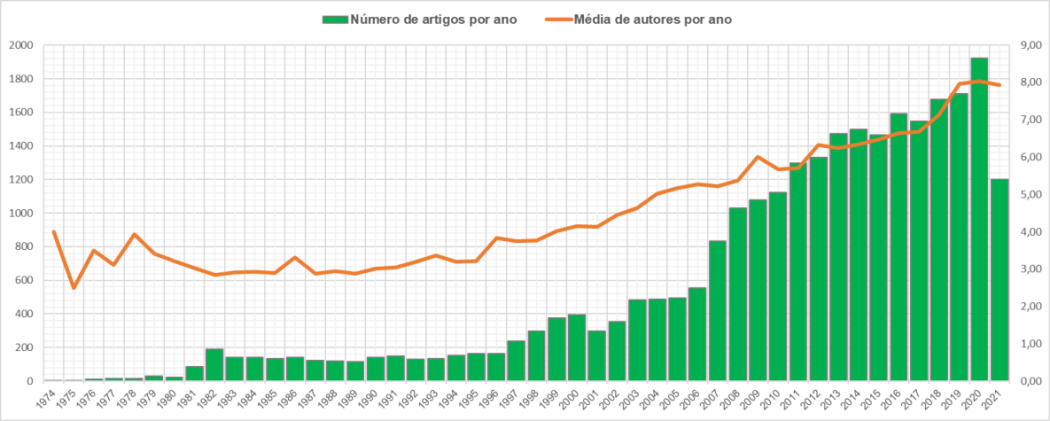


Figura 1. Número de artigos publicados pelos autores vinculados à EMBRAPA em periódicos indexados na WoS e média de autores por ano

Do total de 27.089 artigos publicados em periódicos, 3.634 (13,42%) ainda não foram citados. Porém, 3.121 (85,88%) desses artigos foram publicados a partir de 2020. Os outros 23.455 (86,58%) artigos foram citados por 396.764 artigos, com uma média de 16,92 citações por artigo. Desse total, 446 foram citados mais de 100 vezes e 4 artigos foram citados mais de 1.000 vezes.

Para publicar esses 27.089 artigos, os autores vinculados à EMBRAPA colaboraram com autores de outros 140 países. É importante destacar que um determinado artigo pode ser contabilizado por mais de um país. Isso se deve ao fato de que a WoS faz a contagem contabilizando todas as afiliações informadas nos artigos (Almeida e Guimarães, 2013). A Figura 2 apresenta a distribuição geográfica dos coautores que publicaram com autores vinculados à EMBRAPA. São destacados na cor

"cinza" os países (44; 31,43%) em que houve menos de cinco colaborações. As principais colaborações foram com os Estados Unidos (2.698), França (776) e Inglaterra (556).



Figura 2. Distribuição geográfica dos coautores que publicaram artigos com autores vinculados à EMBRAPA em periódicos indexados na WoS no período de 1974 a 2021. Os 140 países identificados estão destacados por cores de acordo com o número de documentos em colaboração. Os países com menos de 5 colaborações estão destacados na cor cinza

Essas colaborações também são destacadas no mapa de visualização da rede de coautorias por país de origem ilustrado na Figura 3. 75 (53,19%) países com pelo menos 10 publicações no período de 1974 a 2021 foram incluídos no mapa de visualização e estão agrupados em 6 clusters. A espessura de uma relação entre quaisquer dois países é um indicativo da importância da coautoria (ou seja, colaboração). É possível observar a posição de destaque do Brasil, representado no cluster azul, juntamente com outros 7 países das Américas e a Espanha. Neste mesmo cluster também é curioso observar que Honduras se relaciona apenas com o Brasil. O maior cluster, com 49 países, é representado na cor vermelha, sendo a Austrália o país com o maior número de artigos.

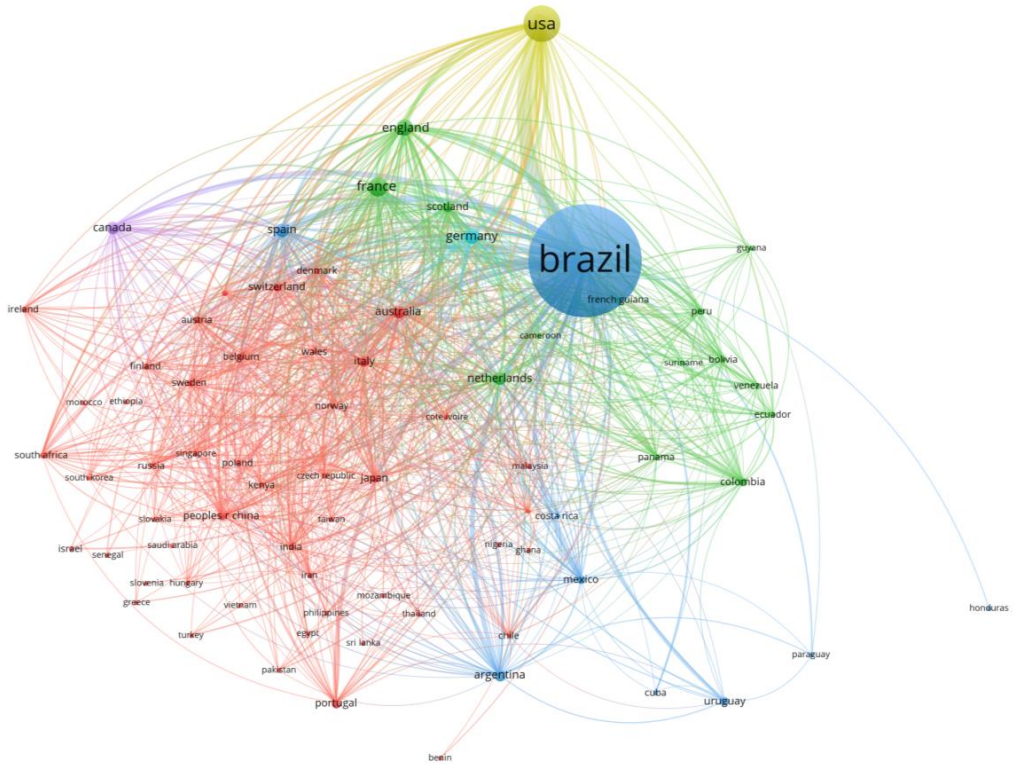


Figura 3. Mapa de visualização da rede de coautorias por país de origem de acordo com a afiliação informada nos artigos publicados por autores vinculados à EMBRAPA em periódicos indexados na WoS no período de 1974 a 2021. A relação entre os países é determinada pelo número de documentos em colaboração. Entre os 141 países identificados, 75 tinham pelo menos 10 documentos e estão agrupados em 6 clusters

A Tabela 1 apresenta a distribuição dos 5 periódicos mais usados por período pelos autores vinculados à EMBRAPA para publicação de artigos em periódicos indexados na WoS. Percebe-se que em todos os períodos há maior divulgação de pesquisas em periódicos brasileiros. É interessante também observar que no período de 2011 a 2020, o periódico "PLOS ONE" passou a ser um dos mais usados, embora o periódico tenha sido criado em 2009. O periódico "Pesquisa Agropecuária Brasileira" foi o mais usado em quase todos os períodos analisados. Porém, no ano de 2021 (ainda incompleto) não conste na relação dos 5 periódicos mais usados. Nota-se também que nos últimos dois anos há uma predileção pelo periódico "Scientific Reports", em detrimento ao periódico "PLOS ONE". Talvez o FI tenha influenciado nessa mudança.

O periódico "Semina: Ciências Agrárias" foi o mais usado nos últimos 2 anos e também está entre os mais usados no período de 2011 a 2020. O periódico "Ciência Rural" também foi destaque nesses mesmos períodos. O número de periódicos distintos usados no período de 2011 a 2020 aumentou significativamente (71,80%) em relação à década anterior. Isso também impactou na porcentagem do periódico mais usado, sendo quase 3 vezes menor. É curioso que só afetou a porcentagem do periódico mais usado.

Tabela 1. Distribuição dos periódicos indexados na WoS mais usados pelos autores vinculados à EMBRAPA

ISSN	Periódico	Artigos (%)	FI 2020
1974 a 2021 (27.089 artigos) [2.005 periódicos]			
0100-204X	Pesquisa Agropecuária Brasileira	3.324 (12,27)	1,088
0103-8478	Ciência Rural	615 (2,27)	0,803
0100-2945	Revista Brasileira de Fruticultura	517 (1,91)	0,912
0100-0683	Revista Brasileira de Ciência do Solo	472 (1,74)	1,683
1676-546X	Semina: Ciências Agrárias	429 (1,58)	0,564
2001 a 2010 (6.737 artigos) [922 periódicos]			
0100-204X	Pesquisa Agropecuária Brasileira	896 (13,30)	1,088
1413-7054	Ciência e Agrotecnologia	252 (3,74)	1,390
0100-0683	Revista Brasileira de Ciência do Solo	201 (2,98)	1,683
0100-2945	Revista Brasileira de Fruticultura	182 (2,70)	0,912
1806-9290	Revista Brasileira de Zootecnia	180 (2,67)	0,967
2011 a 2020 (15.513 artigos) [1.584 periódicos]			
0100-204X	Pesquisa Agropecuária Brasileira	713 (4,60)	1,088
0103-8478	Ciência Rural	436 (2,81)	0,803
1676-546X	Semina: Ciências Agrárias	373 (2,40)	0,564
0100-2945	Revista Brasileira de Fruticultura	318 (2,05)	0,912
0100-0683	Revista Brasileira de Ciência do Solo	270 (1,74)	1,683
1932-6203	PLOS ONE	270 (1,74)	3,240
2020 (1.921 artigos) [653 periódicos]			
1676-546X	Semina: Ciências Agrárias	40 (2,08)	0,564
0100-204X	Pesquisa Agropecuária Brasileira	38 (1,98)	1,088
1932-6203	PLOS ONE	32 (1,67)	3,240
2045-2322	Scientific Reports	29 (1,51)	4,379
0103-8478	Ciência Rural	25 (1,30)	0,803
2021 (1.200 artigos) [490 periódicos]			
1676-546X	Semina: Ciências Agrárias	38 (3,17)	0,564
2045-2322	Scientific Reports	22 (1,83)	4,379
1678-992X	Scientia Agricola	20 (1,67)	2,137
0103-8478	Ciência Rural	20 (1,67)	0,803
0100-2945	Revista Brasileira de Fruticultura	17 (1,42)	0,912
1519-566X	Neotropical Entomology	17 (1,42)	1,434

A Figura 4 apresenta o mapa de visualização da rede de coautorias de acordo com as instituições informadas nas afiliações. Entre as 11.711 instituições identificadas, 131 tinham pelo menos 100 documentos e estão agrupadas em 7 clusters. A EMBRAPA e a USP ocupam posição de destaque no mapa de visualização devido ao número de documentos. Porém, a USP (130) e a Universidade Federal de Viçosa (129) se destacam em relação ao número de relacionamentos. A EMBRAPA, com 110 relacionamentos de colaboração, é prejudicada pois pode-se notar que há diferentes designações para a instituição representadas no mapa. Essa falta de padronização das afiliações dos autores vinculados à EMBRAPA na WoS inclusive já foi estudada por Penteado Filho e Fonseca Júnior (2017). Diante disso, a instituição publicou uma resolução normativa interna regulamentando a afiliação de seus pesquisadores em publicações nacionais e internacionais.

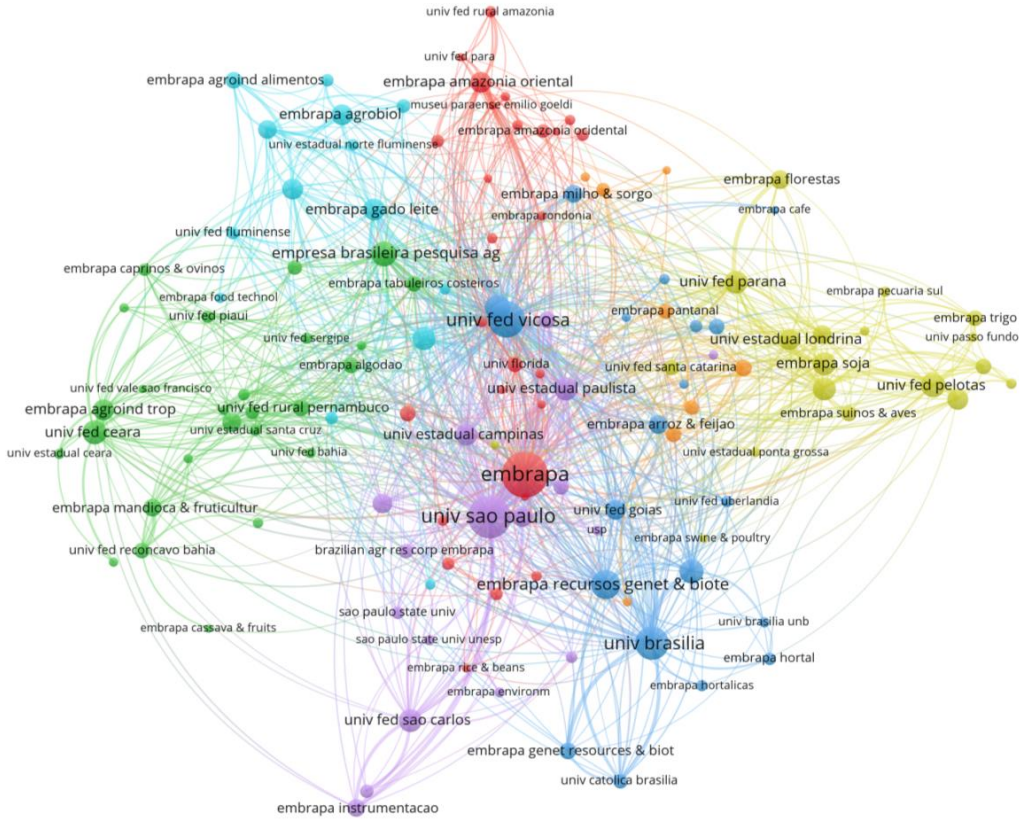


Figura 4. Mapa de visualização da rede de coautorias por instituição de origem de acordo com a afiliação informada nos artigos publicados por autores vinculados à EMBRAPA em periódicos indexados na WoS no período de 1974 a 2021. A relação entre as instituições é determinada pelo número de documentos em colaboração. Entre as 11.711 instituições identificadas, 131 tinham pelo menos 100 documentos e estão agrupadas em 7 clusters

Também foi realizada uma análise das palavras-chave mais usadas pelos autores vinculados à EMBRAPA, conforme ilustrado na Figura 5. No total de 27.089 artigos considerados neste trabalho foram identificadas 39.897 palavras-chave distintas, sendo que 175 foram usadas em pelo menos 50 documentos e estão agrupadas em 7 clusters.

As palavras-chave mais usadas foram "*glycine max*" (458), "*biological control*" (439) e "*Brazil*" (414). É interessante que essas palavras-chave também foram as mais usadas em 2020. Em 2021, as palavras-chave mais usadas por enquanto foram "*biological control*", "*phenolic compounds*" e "*sustainability*". No mapa de visualização, a palavra-chave "*Brazil*" se relaciona com 98 outras palavras-chave, enquanto "*glycine max*" e "*biological control*" se relacionam com 93 e 50 outras palavras-chave, respectivamente. É curioso que a palavra-chave "*chitosan*" não se relaciona com nenhuma outra.

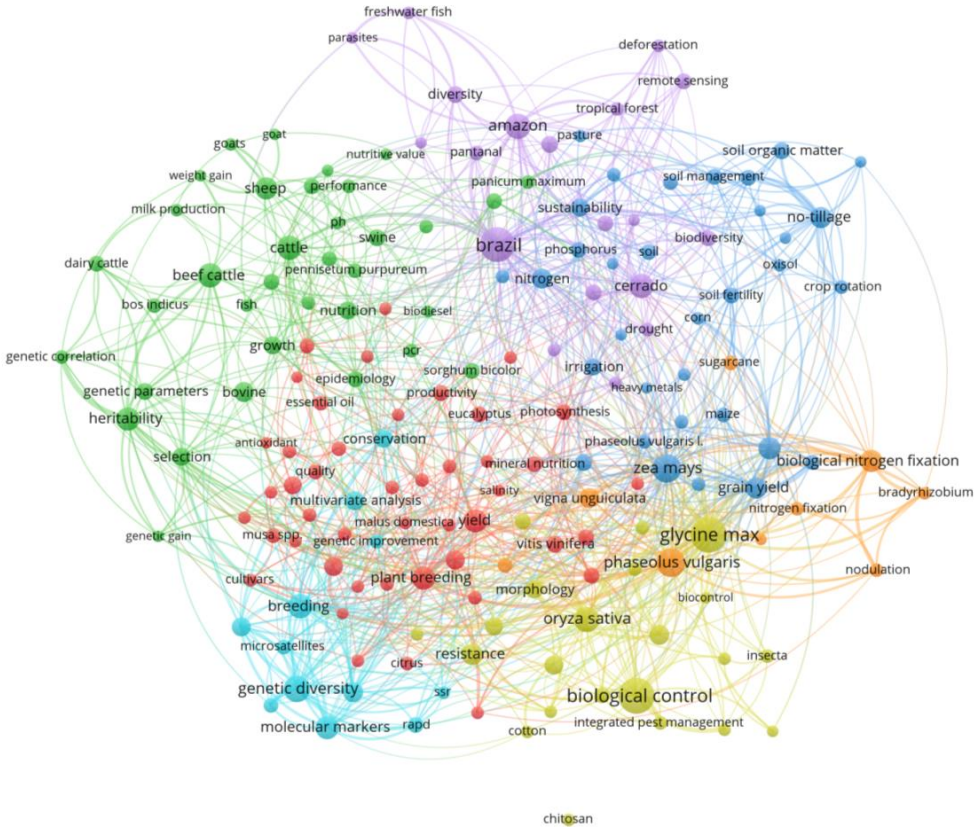


Figura 5. Mapa de visualização da rede de co-ocorrência de palavras-chave usadas em artigos publicados pelos autores vinculados à EMBRAPA em periódicos indexados na WoS no período de 1974 a 2021. A relação entre as palavras-chave é determinada pelo número de documentos em que ocorrem em conjunto. Entre as 39.897 palavras-chave identificadas, 175 foram usadas em pelo menos 50 documentos e estão agrupadas em 7 clusters

4. Considerações Finais

Neste trabalho foi apresentada uma análise bibliométrica da produção científica dos autores vinculados à EMBRAPA considerando os artigos publicados em periódicos indexados na WoS no período de 1974 a 2021. A produção científica da EMBRAPA aumentou significativamente (130,27%) no período de 2011 a 2020 em relação ao período anterior de 2001 a 2010. No período de 1974 a 2021 foram publicados 27.089 artigos e mais de 85% dos artigos já foram citados, sendo que 446 desses artigos foram citados mais de 100 vezes e 4 foram citados mais de 1.000 vezes. Nesse período, a EMBRAPA colaborou com pesquisadores de 140 países diferentes, o que certamente

contribuiu para aumentar o reconhecimento da instituição no cenário mundial. A média de autores foi de 6,04 autores por artigo. Entretanto, essa média vem aumentando, já sendo superior a 8 autores por artigo. Os artigos publicados pelos autores vinculados à EMBRAPA foram em sua maioria (98,32%) em coautoria, sendo que mais de 80% foram publicados por 4 ou mais autores. A principal colaboração ocorreu com coautores dos Estados Unidos. Embora fique evidente uma predileção por periódicos brasileiros, isso vem mudando um pouco nos últimos anos. Recentemente, nota-se que periódicos internacionais e multidisciplinares como "PLOS ONE" e "Scientific Reports" têm sido bastante considerados pelos autores vinculados à EMBRAPA. Também é importante destacar a 8ª posição que a EMBRAPA ocupa entre as instituições que mais publicaram artigos em periódicos indexados na WoS considerando a produção científica brasileira.

A análise da produção científica de uma instituição permite contar parte de sua própria história, assim como pode indicar o perfil de colaboração científica nas publicações dos seus pesquisadores, suas competências, fraquezas, sua inserção e visibilidade nacional e internacional. Um melhor conhecimento da instituição permite definir estratégias de incentivo, corrigir deficiências e acelerar melhorias.

Referências

Aljanabi, S. M. e Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research, 25(22), 4692–4693.

Almeida, E. C. E. e Guimarães, J. A. (2013). Brazil's growing production of scientific articles—how are we doing with review articles and other qualitative indicators?, Scientometrics, 97(2), 287–315.

Irizaga, K. R. D. F. e Vanz, S. A. D. S. (2021). A produção científica por unidade federativa brasileira em ciências agrárias na Scopus: uma análise bibliométrica. AtoZ: novas práticas em informação e conhecimento. Curitiba: Universidade Federal do Paraná, Programa de Pós-Graduação em Gestão da Informação. Vol. 10, n. 1 (jan./abr. 2021), p. 1-14.

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., ... e Cuntz, M. (2020). TRY plant trait database–enhanced coverage and open access. Global Change Biology, 26(1), 119-188.

Penteado Filho, R. D. C. e Avila, A. F. D. (2009a). A Scientometric Analysis of Embrapa's Brazil Web of Science Articles from 1977 to 2006. In Proceedings of ISSI (pp. 786-797).

Penteado Filho, R. D. C. e Avila, A. F. D. (2009b). Embrapa Brasil: Análise bibliométrica dos artigos na Web of Science (1977–2006) 2009. Texto para Discussão (Brasília. 1998), v. 36, p. 7–116.

Penteado Filho, R. D. C. e Avila, A. F. D. (2009c). Estudo das citações dos artigos da Embrapa na Web of Science de 1977 a 2006. Texto para Discussão (Brasília. 1998), v. *37*, p. 7–131.

Penteado Filho, R. D. C. e Avila, A. F. D. (2010). Reflections on a Scientometrical Analysis of Embrapa Brazil Web of Science Articles from 1977 to 2006. Collnet Journal of Scientometrics and Information Management, *4*(1), 9-20.

Penteado Filho, R. D. C. e Fonseca Júnior, W. C. D. (2017). O problema da padronização das afiliações de autores na base de dados Web of Science: o caso Embrapa e sua solução. Em Questão; v. 23, Edição Especial 5 EBBC, 2017; 74-93.

Turnipseed, S. G., Heinrichs, E. A., Da Silva, R. F. P. e Todd, J. W. (1974). Response of soybean insects to foliar applications of a chitin synthesis inhibitor TH 6040. Journal of Economic Entomology, 67(6), 760-762.

Van Eck, N. J. e Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.



Avaliação de agitabilidade bovina baseada em dados de acelerômetro

Tulon Pereira da Fontoura[1]

[1]PPGCAP – Universidade Federal do Pampa (Unipampa)
Maria Anunciação Gomes de Godoy, 1650 – 96413-172 – Bagé – RS – Brazil

tulonfontoura@unipampa.edu.br

Abstract. *In beef cattle raising, the more docile temperament is related to ease of handling and better meat quality, and the composite score is a widely used method for the evaluation of temperament. In the research on which this article is concerned, a tool for collecting data on the movement of cattle in a containment environment was developed, on which calculations representing the temperament when compared to the composite score were proposed and evaluated. For the collection tool, an accelerometer sensor capable of monitoring movements in three dimensions was used. After analyzing the correlations between the proposed calculations and the composite score, it was concluded that the calculation involving the sum of the standard deviations of the three movement axes is indicated for the assessment of temperament.*

Resumo. *Na pecuária de bovinos de corte, o temperamento mais dócil está ligado à facilidade de manejo e melhor qualidade da carne, e o escore composto de balança é método amplamente utilizado para a avaliação do temperamento. Na pesquisa sobre a qual versa este artigo, desenvolveu-se uma ferramenta de coleta de dados de movimentação de bovinos em ambiente de contenção, sobre a qual foram propostos e avaliados cálculos que representassem o temperamento quando comparado ao escore composto de balança. Para a ferramenta de coleta, utilizou-se um sensor acelerômetro capaz de monitorar movimentações em três dimensões. Após a análise das correlações entre os cálculos propostos e o escore composto de balança, concluiu-se que o cálculo envolvendo o somatório dos desvios padrão dos três eixos de movimentação é indicado para a avaliação de temperamento.*

1. Introdução

Conforme há o aumento populacional do planeta, é necessário que mais alimento seja produzido, assim, novos métodos de produção de alimentos são criados. Quando trata-se

de produção animal, principalmente bovinos de corte, o controle da evolução dos animais torna-se muito importante. Uma das características relevantes para esse processo produtivo é o temperamento dos bovinos. Animais com temperamento mais agressivo tendem a prejudicar benfeitorias e tratadores[Probst et al. 2012], ter menor ganho de peso e produzir carne de pior qualidade[Piovesan 1998], sendo menos desejáveis. Em oposição, os bovinos mais calmos e tranquilos são mais fáceis de serem manejados, menos danosos às benfeitorias e produzir carne mais macia[Melo Menezes et al. 2017]. A avaliação do temperamento bovino pode ser feita utilizando o método desenvolvido por [Piovesan 1998], chamado de escore composto de balança, no qual um observador monitora, por um período determinado de tempo, as movimentações do animal quando este está restrito ao local de pesagem e, a seguir, atribuindo-lhe uma nota de 1 a 5.

O objetivo da pesquisa que aqui se apresenta foi desenvolver uma forma alternativa de realizar a avaliação do temperamento de bovinos, de maneira não dependente de um observador subjetivo. Para atingir este objetivo foi necessário desenvolver uma ferramenta experimental de coleta de dados de movimentação, elaborar e propor cálculos a serem aplicados sobre os dados coletados e comparar os resultados com outro método de avaliação do temperamento já utilizados na atualidade. Por último, recomenda um entre os vários cálculos propostos para ser utilizado na avaliação do temperamento com base em dados de acelerômetro.

A justificativa da pesquisa se deu pela necessidade de desenvolver uma forma objetiva de atribuir o temperamento de forma a causar menos interferência nos processos de manejo já adotados na propriedade e reduzindo erros de anotação e afastando a subjetividade do observador.

Após esta introdução, será apresentada a ferramenta experimental de coleta criada para a coleta de dados de movimentação, passando então a apresentação dos cálculos propostos, em seguida, apresenta o método de avaliação de temperamento com o qual os cálculos propostos serão correlacionados. Por fim, serão apresentados os resultados obtidos e a conclusão.

2. Ferramenta de coleta

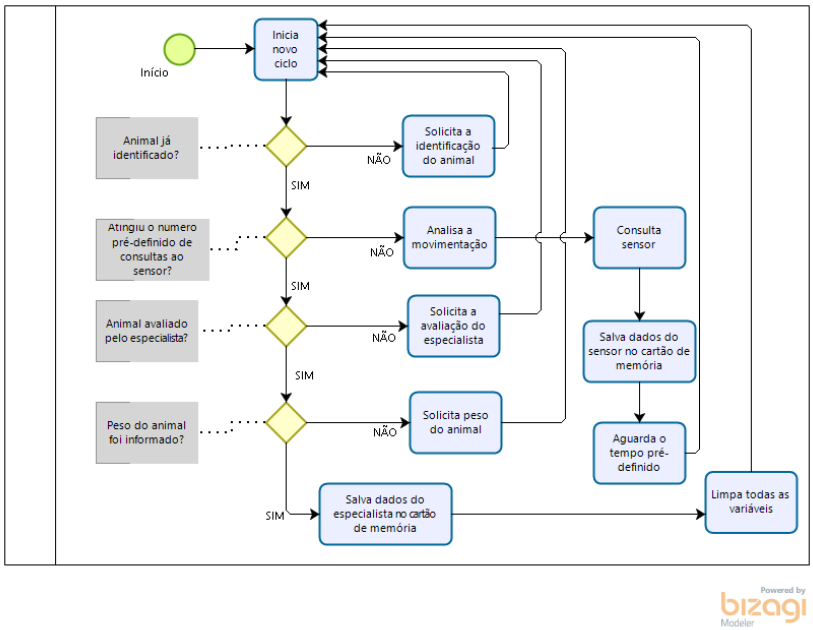
A ferramenta experimental de coleta foi desenvolvida utilizando a plataforma de prototipação chamada Arduíno Uno a fim de avaliar o conceito proposto. O Arduíno Uno utiliza um microcontrolador de oito Bits do modelo ATmega328p e trabalha, por padrão, com frequência de 16MHz[1] para controlar todos os demais componentes da ferramenta. Ao Arduíno Uno foram conectados um acelerômetro tri-axial, um *display* de cristal líquido, um teclado e um cartão de memória. Foi utilizado o sensor acelerômetro do modelo MPU6050 GY-521, escolhido devido a características como frequência de operação, sensibilidade e tipo de comunicação, aliado ainda à facilidade de encontrá-lo e seu custo reduzido. Este modelo de acelerômetro é capaz de informar dados de aceleração nos eixos X (movimentação lateral), Y (movimentação longitudinal) e Z (movimentação vertical), além de outros dados não utilizados nesta pesquisa. O acelerômetro foi instalado dentro de uma caixa estanque, para isolá-lo de poeira, urina e fezes dos animais, uma vez que foi afixado à parte inferior da balança onde o animal fica contido e com movimentações limitadas. O *display* de cristal líquido foi utilizado em conjunto com o

[1] 16MHz corresponde a até 16 milhões de ciclos de processamento por segundo

teclado para que assim fosse possível ao operador visualizar mensagens da ferramenta de coleta e inserir os dados dos animais em avaliação. O cartão de memória armazenou as informações dos animais e também dos dados de movimentação dos mesmos.

O Arduino Uno trabalha com a arquitetura de *loop* infinito, ou seja, utiliza uma lógica cíclica de programação: sempre que chega ao fim do ciclo ele volta ao seu início, e, por este motivo, a programação deve ser realizada utilizando vários ciclos aninhados (um ciclo dentro de outro ciclo), cada um deles com seu critério de saída. O primeiro ciclo é o que aguarda o operador inserir a identificação do animal. O operador deve inserir o brinco do animal e confirmar pressionando <enter>, que deve ser pressionado apenas com o animal fechado na balança. Em seguida, inicia-se o ciclo de coleta de dados do acelerômetro, no qual são captados e armazenados os dados de movimentação do acelerômetro. Este ciclo se repete por 200 vezes aguardando um intervalo de 200 milissegundos entre cada ciclo. A seguir, inicia-se o ciclo no qual o operador deve informar a avaliação pelo método do escore composto de balança, conforme proposto por [Piovesan 1998]. Por último, inicia-se o ciclo no qual o operador deve informar o peso do animal auferido pela balança.

Figura 1. Representação do fluxo de processamento



Fonte: [Fontoura 2020]

Os dados coletados são armazenados em arquivo de texto no formato CSV[2]. O formato foi escolhido devido a sua simplicidade e portabilidade e é preenchido linha a linha em cada ciclo de coleta de dados de acelerômetro. Ao final do registro da coleta de dados de movimentação, é inserida uma linha contendo o escore composto de balança e o peso informados. A cada animal pertence um bloco de 201 linhas do arquivo final. Para

[2]*Comma Separated Value*

esta pesquisa, foram coletados e analisados dados de movimentação de 165 animais da raça Brangus com idades em torno de 18 meses.

3. Cálculos propostos

A cada animal pertence um conjunto de 200 linhas de registro de aceleração instantânea coletados pelo acelerômetro e armazenados pelo Arduíno. Desta forma, é necessário que estes dados sejam processados para que seja obtido apenas um valor numérico que corresponda a toda a movimentação do animal. Foram propostos 29 cálculos que envolvem os eixos de maneira independente, o conjunto dos eixos X e Y para representar as movimentações no plano horizontal e cálculos envolvendo os eixos X, Y e Z representando a movimentação tridimensional.

Quanto às fórmulas matemáticas utilizadas, dispõe-se de desvios padrão, médias, médias dos módulos, raízes quadradas e logaritmos. Para o eixo X, que representa a movimentação lateral, foram utilizados os cálculos de desvio padrão, média, média dos módulos somatório das raízes e somatório dos logaritmos. Para o eixo Y, que representa a movimentação longitudinal, e para o eixo Z que representa a movimentação vertical, foram replicadas as fórmulas do eixo X. Para os conjuntos de eixos X e Y, que representa a movimentação no plano horizontal, e para os eixos X, Y e Z, que representa a movimentação tridimensional foram utilizadas as fórmulas de somatório dos desvio padrão dos eixos, somatório das médias dos eixos, somatórios das médias dos módulos dos eixos, somatórios das raízes quadradas da soma dos quadrados dos eixos, desvio padrão das raízes da soma dos quadrados dos eixos, somatório das raízes quadradas dos eixos e o somatório dos logaritmos dos eixos.

4. Apresentação de resultados

Antes de apresentar-se os resultados, é necessário rever, mesmo que brevemente, duas noções essenciais para a compreensão da análise realizada no âmbito deste estudo: o Escore composto de balança e a Correlação de Pearson.

No método descrito por [Piovesan 1998], nomeado Escore composto de balança, o observador monitora a movimentação do animal na balança por um período de tempo pré-determinado, após este período, o observador classifica a movimentação do animal de acordo com as características descritas a seguir: 1 - calmo, nenhum movimento, nenhuma respiração audível; 2 - inquieto, alternando a posição das patas; 3 - se contorcendo, tremendo, movimentando ocasionalmente a balança, respiração ocasionalmente audível; 4 - movimentos contínuos e vigorosos, movimentando a balança, respiração audível; 5 - movimentos vigorosos e contínuos, movimentando a balança, virando-se ou lutando violentamente, respiração audível. Animais classificados com valores menores indicam temperamento mais calmo e dócil [Melo Menezes et al. 2017]. Para esta pesquisa o tempo de observação do animal coincide com o tempo que é realizado o monitoramento da movimentação pela ferramenta de coleta, de forma simultânea.

Já a Correlação de Pearson mede o grau de correlação entre duas variáveis e varia de -1 a 1 onde o sinal negativo indica a existência de correlação inversa entre as variáveis. [Doria Filho 1999] classifica os coeficientes de correlação da seguinte forma: inexistente = 0; fraca quando varia de -0,5 a 0,5; média quando varia entre 0,5 e 0,75 ou entre -0,5 e -0,75; forte quando varia entre 0,75 e 1 ou -0,75 e -1; perfeita quando é igual a 1 ou -1.

Tabela 1. Cálculos propostos

	Cálculo utilizado	Fórmula
Eixo X	Desvio padrão do eixo X (SDx)	σ^X
	Média do eixo X (Mx)	$\bar{X}$
	Média dos módulos do eixo X (MMx)	$\overline{\lvert X\rvert}$
	Soma das Raízes do eixo X ($\sum Rx$)	$\sum\sqrt{X}$
	Soma dos Logarítmos do eixo X ($\sum \log x$)	$\sum \log X$
Eixo Y	Desvio padrão do eixo Y (SDy)	σ^Y
	Média do eixo Y (My)	$\bar{Y}$
	Média dos Módulos do eixo Y(MMy)	$\overline{\lvert Y\rvert}$
	Soma das Raízes do eixo Y($\sum Ry$)	$\sum\sqrt{Y}$
	Soma dos Logarítmos do eixo Y ($\sum \log y$)	$\sum \log Y$
Eixo Z	Desvio Padrão do eixo Z (SDz)	σ^Z
	Média do eixo Z (Mz)	$\bar{Z}$
	Média dos módulos do eixo Z (MMz)	$\overline{\lvert Z\rvert}$
	Soma das Raízes do eixo Z($\sum Rz$)	$\sum\sqrt{Z}$
	Soma dos Logarítmos do eixo Z ($\sum \log z$)	$\sum \log Z$
Eixos X e Y	Soma dos desvio padrão dos eixos X e Y ($\sum SDxy$)	$\sigma^X + \sigma^Y$
	Soma das médias dos eixos X e Y ($\sum Mxy$)	$\bar{X} + \bar{Y}$
	Soma das médias dos módulos dos eixos X e Y ($\sum MMxy$)	$\overline{\lvert X\rvert} + \overline{\lvert Y\rvert}$
	Soma das raízes dos quadrados dos eixos X e Y ($\sum RQxy$)	$\lvert(\sqrt{X^2+Y^2})\rvert$
	Desvio padrão das raízes dos quadrados dos eixos X e Y ($SDRQxy$)	$\sigma(\sqrt{X^2+Y^2})$
	Soma das Raízes dos eixos X e Y ($\sum Rxy$)	$\sum\sqrt{X} + \sum\sqrt{Y}$
	Soma dos Logarítmos dos eixos X e Y ($\sum \log xy$)	$\sum \log X + \sum logY$
Eixos X, Y e Z	Soma dos desvios padrão dos eixos X, Y e Z ($\sum SDxyz$)	$\sigma^X + \sigma^Y + \sigma^Z$
	Soma das médias dos eixos X, Y e Z ($\sum Mxyz$)	$\bar{X} + \bar{Y} + \bar{Z}$
	Soma das médias dos módulos dos eixos X, Y e Z ($\sum MMxyz$)	$\overline{\lvert X\rvert} + \overline{\lvert Y\rvert} + \overline{\lvert Z\rvert}$
	Soma das raízes dos quadrados dos eixos X, Y e Z ($\sum RQxyz$)	$\lvert(\sqrt{X^2+Y^2+Z^2})\rvert$
	Desvio padrão das raízes dos quadrados dos eixos X, Y e Z ($SDRQxyz$)	$\sigma(\sqrt{X^2+Y^2+Z^2})$
	Soma das Raízes dos eixos X, Y e Z ($\sum Rxyz$)	$\sum\sqrt{X} + \sum\sqrt{Y} + \sum\sqrt{Z}$
	Soma dos Logaritmos dos eixos X, Y e Z ($\sum \log xyz$)	$\sum \log X + \sum \log Y + \sum \log Z$

Fonte: [Fontoura 2020]

Tendo realizado os cálculos propostos foi possível estabelecer as correlações, utilizando-se a correlação de Pearson, entre os resultados dos cálculos propostos e o escore composto de balança de cada animal. As correlações encontradas entre o escore composto de balança e os resultados obtidos estão apresentados na tabela 2.

Ao analisar a tabela 2 é possível identificar que os cálculos propostos que utilizam o desvio padrão para serem calculados apresentam correlações superiores a 0,7, o que, de acordo com [Doria Filho 1999] são consideradas de média a forte. As correlações do escore composto de balança com cálculo de média dos módulos apresenta correlações médias. As com cálculo de médias apresenta correlação baixa. Os demais cálculos apresentam correlação fraca. Os outros cálculos apresentam correlações ainda menores.

5. Conclusões

A ferramenta de coleta foi desenvolvida com o objetivo único de coleta de dados de movimentação e por este motivo retornará à fase de desenvolvimento para que possa apre-

Tabela 2. Correlações entre escore composto de balança e resultados obtidos

Eixo X		Eixo Y		Eixo Z		Eixos X e Y		Eixos X, Y e Z	
SDx	0,72	SDy	0,77	SDz	0,71	$\sum SDxy$	0,75	$\sum SDxyz$	0,75
Mx	0,09	My	0,06	Mz	0,34	$\sum Mxy$	0,10	$\sum Mxyz$	0,34
MMx	0,61	MMy	0,71	MMz	0,01	$\sum MMxy$	0,68	$\sum MMxyz$	0,60
$\sum Rx$	0,10	$\sum Ry$	0,56	$\sum Rz$	-0,09	$\sum Rxy$	0,34	$\sum Rxyz$	0,04
						$\sum RQxy$	-0,01	$\sum RQxyz$	-0,11
						$\sum SDRQxy$	0,75	$SDRQxyz$	0,72
$\sum logx$	-0,09	$\sum logy$	0,10	$\sum logz$	-0,08	$\sum logxy$	-0,00	$\sum logxyz$	-0,04

Fonte: Autor (2020)

sentar o resultado do cálculo sugerido ao usuário. Também foi constatada a necessidade de incluir a verificação de conectividade dos sensores e outros dispositivos conectados, para que seja garantido que haja a correta realização da coleta de dados de movimentação. Bem como deve ser incluído um leitor RFID a fim de otimizar o processo de identificação dos animais e evitar erros de digitação ou leitura do brinco do animal. Conforme observado nos resultados apresentados neste artigo, as correlações mais fortes foram encontradas nos cálculos que utilizam o desvio padrão para sua realização. Indo além, sugere-se que seja utilizado o cálculo do somatório dos desvios padrão dos eixos X, Y e Z tendo desta forma um resultado que demonstra a movimentação tridimensional do animal.

Referências

Doria Filho, U. (1999). *Introdução à bioestatística*. Elsevier, São Paulo, SP.

Fontoura, T. (2020). Temperamento bovino: avaliação de agitabilidade baseada em dados de acelerômetro coletados em ambiente de contenção para manejo. *Bagé,RS: Universidade Federal do Pampa*.

Melo Menezes, L. d., Cardoso, F. F., e Silveira, I. D. B. (2017). Temperamento em bovinos de corte: características genéticas, metodologias de mensuração e desempenho. *REDVET. Revista Electrónica de Veterinaria*, 18(9):1–19.

Piovesan, U. (1998). Análise de fatores genéticos e ambientais na reatividade de quatro raças de bovinos de corte ao manejo. *Jaboticabal: Universidade Estadual Paulista*.

Probst, J. K., Neff, A. S., Leiber, F., Kreuzer, M., e Hillmann, E. (2012). Gentle touching in early life reduces avoidance distance and slaughter stress in beef cattle. *Applied Animal Behaviour Science*, 139(1-2):42–49.



BrangusSelection: um algoritmo para seleção ótima de acasalamentos com índice de seleção customizável

Bruno Ernesto Techera da Motta[1], **Ana Paula Lüdtke Ferreira**[1], **Marcos Jun-Iti Yokoo**[2]

[1]Universidade Federal do Pampa - UNIPAMPA

[2]EMBRAPA Pecuária Sul

{brunomotta.aluno, anaferreira}@unipampa.edu.br
marcos.yokoo@embrapa.br

Abstract. *This work presents the BrangusSelection algorithm, which produces an optimal mating scheme, maximizing the expected value of the herd's next generation. Tests with the algorithm show a polynomial runtime in the number of animals received as input, which indicates that the problem is not computationally intractable. The algorithm depends on a selection index built from the animals' selection objectives/criteria and can operate on different animal species and selection indexes.*

Resumo. *Este trabalho apresenta o algoritmo BrangusSelection, que produz um esquema ótimo de acasalamentos, maximizando o valor esperado da próxima geração do rebanho. Os testes com o algoritmo mostram um tempo de execução polinomial na quantidade de animais recebida como entrada, o que indica que o problema não é intratável computacionalmente. O algoritmo depende da existência de um índice de seleção, construído a partir dos objetivos e critérios de seleção dos animais e pode operar sobre qualquer espécie de animal e índice de seleção.*

1. Introdução

O crescimento da população demanda aumento da produção alimentos sem, contudo, o aumento do uso de recursos naturais, que já estão no seu limite [FAO 2009]. Uma das estratégias para o aumento da produção é melhorar o potencial genético dos rebanhos. O melhoramento genético animal consiste no direcionamento da genética das novas gerações para o aumento da saúde e das caraterísticas dos animais que se traduzam em ganhos para o sistema de produção. As principais estratégias são os sistemas de seleção e de acasalamentos [Eler 2017a, Cardoso 2009]. A seleção escolhe os animais que se tornarão

pais da próxima geração e o sistema de acasalamento determina quem são os pares de animais que devem reproduzir. Os pais passarão sua superioridade genética para a prole, minimizando características indesejáveis e perpetuando as vantajosas [Nieto et al. 2013].

Um *objetivo de seleção* é uma característica desejável que se quer obter no futuro rebanho, que pode compreender características genéticas ou econômicas [Simões et al. 2020]. Os *critérios de seleção* são as características pelas quais os animais são avaliados para que os objetivos de seleção sejam atingidos [Abreu et al. 2013]. Critérios de seleção podem ser combinados, gerando um *índice de seleção* que expressa, em um único valor, o mérito de cada animal [Eler 2017b]. Índices de seleção são utilizados pelos diversos programas de melhoramento genético, visto que facilitam a decisão de seleção dos reprodutores [Cardoso 2009]. Dados os animais selecionados para reprodução, o *problema de maximização de escolha de acasalamentos* visa encontrar o melhor esquema de pareamento de machos e matrizes, objetivando o maior índice possível da prole.

Os trabalhos relacionados encontrados na literatura fazem uso majoritário de algoritmos genéticos, meta-heurística apropriada para problemas para os quais não se conhece algoritmo eficiente. Cita-se Fontoura et al. (2020), que busca a otimização de acasalamentos com base nas *diferenças esperadas de progênie* (DEP); Carvalho et al. (2016) apresenta um sistema com o objetivo de maximizar o ganho de peso e minimizar a endogamia do rebanho caprino em estudo e, para isso, faz uso de algoritmos genéticos também com uso de DEP; Barreto et al. (2014) otimizou esquemas de acasalamento em ovinos, também com abordagem genética com valores estimados por DEP-BLUP e também com o objetivo de minimizar a consanguinidade; Carvalheiro et al. (2010) usa algoritmos genéticos para seleção de animais para reprodução também visando a diminuição da coancestralidade da nova geração do rebanho. Nenhum dos trabalhos citados investiga a complexidade do problema de encontrar o melhor esquema possível de acasalamentos, ainda que alguns afirmem que o problema é computacionalmente intratável, sem referências ou prova para essa afirmação. O argumento de que o espaço de estados tem um tamanho exponencial no número de animais envolvidos e que considerá-lo por inteiro é inviável é tautológica, mas não informa sobre a real complexidade do problema. Um dos objetivos deste trabalho foi fazer essa investigação, na busca de um algoritmo ótimo com tempo de execução polinomial no número de animais envolvidos.

Neste texto, é apresentado o algoritmo BrangusSelection, que usa uma estratégia que combina programação linear e o método *branch and bound* para encontrar a solução ótima do problema e estabelecer diferentes cenários para investigar a eficiência com a qual o problema é resolvido. O índice bioeconômico desenvolvido em [Simões et al. 2020], com adição de variáveis correspondentes ao temperamento, será usado para apresentação dos resultados. No entanto, o algoritmo proposto é genérico e pode ser executado com qualquer índice.

O restante do texto está assim organizado: a Seção 2 apresenta a caracterização do problema, o modelo do algoritmo e o ferramental tecnológico usado para a construção da solução; a Seção 3 discute os resultados obtidos; a Seção 4 apresenta as considerações finais e uma discussão sobre a complexidade do problema de otimização dos acasalamentos.

2. Caracterização e modelo do problema

O problema de maximização de escolha de acasalamentos procura otimizar as opções de acasalamento de acordo com o índice de seleção escolhido, objetivando o maior índice possível da prole, calculado como a média dos índices dos progenitores. O processo de seleção dos animais que vão acasalar é feito previamente e está fora do escopo deste trabalho. O índice de seleção consiste em um conjunto de características (desejadas ou não) que podem ocorrer nos animais, com um valor de ponderação sobre essas características. Formalmente, um *índice de seleção* é um par $\mathcal{I} = (T, w)$ em que T é o conjunto de características consideradas no índice e $w : T \to \mathbb{R}$ é a função de ponderação do índice que designa o peso relativo de cada característica. Não há restrições quanto aos valores que a ponderação das características e nem os valores precisam estar normalizados. Valores negativos são aceitáveis, visto que eles representam penalidades, no sentido de que um valor alto para a característica correspondente torna o valor do índice do animal menor.

O objetivo do índice de seleção é estimar a contribuição individual de cada animal para a próxima geração do rebanho, que é a soma do valor de suas características ponderadas pela importância relativa de cada uma delas. Se A é um conjunto de animais, $\mathcal{I} = (T, w)$ um índice de seleção e $\nu : T \times A \to \mathbb{R}$ a função que retorna o valor de uma característica para cada animal, a *contribuição individual* $\iota^{\mathcal{I}}$ de um animal $a \in A$ é

$$\iota^{\mathcal{I}}(a) = \sum_{k \in T} \nu(k, a) \cdot w(k) \tag{1}$$

quando o índice de seleção é entendido a partir do contexto, escreve-se $\iota(a)$ em vez de $\iota^{\mathcal{I}}(a)$.

O valor de um acasalamento é a média da contribuição individual dos progenitores, a menos que o grau de endogamia $r : A \times A \to \mathbb{R}^+$ supere o limite máximo l definido pelo usuário; nesse caso, é atribuído à contribuição o valor $-\infty$ (menos infinito) para excluir a possibilidade dessa opção de acasalamento compor a solução final. O algoritmo pode ser transformado para adicionar punições menos drásticas; aqui, assume-se que se trata de uma restrição. Logo, pares com nível de endogamia acima do aceitável não podem mesmo fazer parte da solução sem que a restrição seja violada. Formalmente, a contribuição do acasalamento do par $(m \in M, f \in F)$ é dado por

$$\pi(m, f) = \begin{cases} (\iota^{\mathcal{I}}(m) + \iota^{\mathcal{I}}(f))/2 & r(m, f) \leq l \\ -\infty & r(m, f) > l \end{cases} \tag{2}$$

onde $\iota^{\mathcal{I}}$ é a contribuição individual de cada animal, dado um índice de seleção $\mathcal{I}$. O problema de maximização de seleção de acasalamento é apresentado a seguir.

Problema 1 (Problema de maximização de seleção de acasalamentos)

Entrada*: uma matriz de contribuição $\Pi_{|M|\times|F|}$ em que cada $\pi_{ij} = \pi(i, j)$ (equação 2) e uma função $t : M \to \mathbb{N}$ que indica o limite máximo de utilização de cada macho no processo de cruzamento.*

Saída*: uma matriz binária $\mathcal{B}^*$, com dimensão $|M| \times |F|$ onde cada elemento b^*_{ij} de $\mathcal{B}^*$ é*

$$b^*_{ij} = \begin{cases} 1 & \textit{se o macho } i \textit{ e a fêmea } j \textit{ devem acasalar} \\ 0 & \textit{caso contrário} \end{cases}$$

respeitando as seguintes restrições:

$\sum_{i=1}^{|M|} b_{ij}^* = 1$, *para cada fêmea* j
$\sum_{j=1}^{|F|} b_{ij}^* \leq t_i$, *para cada macho* i *e limite de reprodução individual* t_i

e de modo que, para qualquer outra matriz binária B *com as mesmas características, temos que*

$$\sum_{i=1}^{|M|}\sum_{j=1}^{|F|} \pi_{ij} b_{ij} \leq \sum_{i=1}^{|M|}\sum_{j=1}^{|F|} \pi_{ij} b_{ij}^*$$

Resumidamente, a entrada do Problema 1 é uma matriz Π com dimensão $|M|\times|F|$ e a saída é uma matriz binária $\mathcal{B}^*$ com dimensão $|M| \times |F|$. Cada elemento da matriz binária $\mathcal{B}^*$ deverá ser 1, indicando que o par deve acasalar, ou 0, indicando que o par em questão não faz parte da solução do problema. Em relação às restrições, cada fêmea j deve ser coberta por um único touro ($\sum_{i\in M} b_{ij}^* = 1$) e o número de vezes que um macho i é usado não excede sua utilização máxima ($\sum_{j\in F} b_{ij}^* \leq t(i)$).

O problema da maximização pode ser caracterizado como um problema de otimização em virtude de sua saída ser a melhor solução dentre um conjunto de soluções possíveis, no que é denominado espaço de estados das soluções [Cormen et al. 2002]. Como a saída do problema é uma função inteira com domínio em F e contradomínio em M, pode-se modelar o problema como uma instância de programação linear inteira [Hillier e Lieberman 2013], cuja função objetivo pode ser expressa como

$$\text{MAX} \sum_{i=1}^{|M|}\sum_{j=1}^{|F|} \pi_{ij} c_{ij} \tag{3}$$

estando seu valor sujeito às seguintes restrições:

$$\begin{array}{l} \forall j \in F \ \sum_{i=1}^{|M|} c_{ij} = 1 \\ \forall i \in M \ \sum_{j=1}^{|F|} c_{ij} \leq t(i) \end{array} \tag{4}$$

Note-se que a função objetivo é uma função linear, visto que o valor da contribuição de cada par $(i, j) \in M \times F$ é um valor constante. As variáveis do problema são os valores c_{ij}, que podem tomar valores no conjunto $\{0, 1\}$. O modelo construído, dados os conjuntos de machos M e de fêmeas F, gera uma função objetivo com $|F|.|M|$ termos e um conjunto de $|F| + |M|$ restrições.

A estratégia de resolução do problema de programação inteira por meio da combinação das técnicas de programação linear e *branch and bound* consiste no relaxamento do problema de programação inteira em um problema de programação linear em que as variáveis podem ter valores reais. O método (polinomial) simplex [Hillier e Lieberman 2013] pode ser usado para resolver o problema linear, que contém todas as soluções inteiras do problema. O algoritmo *branch and bound* investiga as soluções inteiras dentro do espaço de soluções, eliminando aquelas que não podem retornar um valor maior do que a melhor solução encontrada até o momento. Essa estratégia permite que o espaço de soluções não precise ser todo investigado. Problemas NP-completos

ainda geram uma árvore de tamanho proporcional a uma função exponencial no número de variáveis, mesmo que nem todo o espaço seja investigado. A análise da solução construída e dos resultados da técnica apontaram para a existência de um algoritmo de complexidade polinomial, como será discutido ao final do trabalho.

3. Resultados e discussão

O algoritmo foi desenvolvido e testado em um computador com processador Intel Core i7-6500U de 2.50GHz; 12GB de memória RAM; Armazenamento SSD de 500GB. O sistema operacional usado foi a distribuição Linux Ubuntu 20.04 LTS 64-bit. Para a implementação do algoritmo foi utilizada a linguagem de programação R [Ihaka e Gentleman 1996], versão 4.0.5. A IDE usada foi o RStudio versão 1.4. Os pacotes R lpSolve [Berkelaar et al. 2020], lpsymphony [Kim 2020] e Rcpp [Eddelbuettel e Balamuta 2018] foram usados solução do modelo de programação linear, *branch and bound* e integração com C++.

A base de dados dos animais utilizada neste trabalho é oriunda da EMBRAPA, em formato de planilha eletrônica, transformada em arquivo *.csv* e importada para o ambiente RStudio. A base de dados contém dados referentes a 1366 bovinos da raça *Brangus*, sendo 511 machos e 855 fêmeas. Os dados correspondem aos valores de DEP aprimorada pela Genômica (DEPG) das características dos animais que compõem o índice de seleção bioeconômico construído em [Simões et al. 2020], acrescido da variável de temperamento. O algoritmo pode, contudo, ser executado para qualquer índice, visto que trabalha somente sobre a matriz de contribuição. A complexidade do cálculo dos valores do índice para cada animal é constante, dado um índice fixo, visto que o valor de uma característica de um animal pode ser buscada em tempo constante em uma tabela indexada. Ambas as tabelas de DEPG e consanguinidade são indexadas. Após esse cálculo, que leva à construção da matriz de contribuição, o algoritmo só executa sobre a matriz de contribuições.

Os testes objetivaram verificar como a função do tempo de execução cresce na medida que o tamanho da entrada (número de machos e fêmeas) aumenta. Para a execução dos testes, aumentou-se gradualmente o número de animais medindo-se o tempo de execução do algoritmo. A matriz de contribuição foi construída de forma que os machos e fêmeas estejam ordenados por índice de contribuição. Essa ordenação fez com que o algoritmo executasse mais rapidamente (da ordem de minutos na configuração dada acima). As consequências desse fato serão discutidas na conclusão deste trabalho.

Para o teste, foram configuradas 9 instâncias, que são determinadas pelo número de machos, número de fêmeas e uso máximo dos touros. Cada instância foi executada 10 vezes, utilizando em cada execução uma amostra diferente da base de dados. O tempo médio, variância e desvio padrão de cada instância executada foram calculados. A Tabela 1 mostra os resultados do teste. As colunas indicam o número da instância, número de machos, número de fêmeas, uso máximo de cada touro, tempo médio de execução e suas variância e desvio padrão, respectivamente. Nesse experimento, fixou-se o número de fêmeas (total das fêmeas da base de dados) e aumentou-se o número de machos progressivamente. O valor do uso máximo de cada touro foi definido como o menor número possível para cobrir o total de vacas. Na instância 9, foram utilizados praticamente todos os animais da base de dados, na qual o BrangusSelection utilizou em média 583,11 segundos (aproximadamente 10 minutos) para encontrar a solução ótima.

Tabela 1. Teste do BrangusSelection com incremento dos machos

n	$\|M\|$	$\|F\|$	max	tempo médio	σ^2	σ
1	100	855	9	29,6 s	0,08	0,29
2	150	855	6	59,9 s	0,19	0,43
3	200	855	5	102,7 s	0,44	0,67
4	250	855	4	157,1 s	1,43	1,19
5	300	855	3	220,0 s	5,13	2,26
6	350	855	3	291,7 s	7,19	2,68
7	400	855	3	381,6 s	4,90	2,21
8	450	855	3	465,9 s	19,49	4,41
9	500	855	3	583,11 s	3,19	1,78

Um teste similar, para observar como o tempo de execução varia ao aumentar o número de fêmeas com um número fixo de machos, foi realizado. O comportamento foi equivalente aos dados apresentados na Tabela 1. A Figura 1a) apresenta a relação do tempo de execução e número de animais da Tabela 1 e a Figura 1b) mostra a relação do tempo de execução e número de animais correspondendo ao aumento do número de fêmeas. Em ambos os gráficos é possível observar a similaridade na curvatura dos tempos de execução em relação ao tamanho das entradas em ambas variações do teste. Da mesma forma, pode-se verificar que tanto os dados da tabela quanto a forma do gráfico apontam para uma função polinomial.

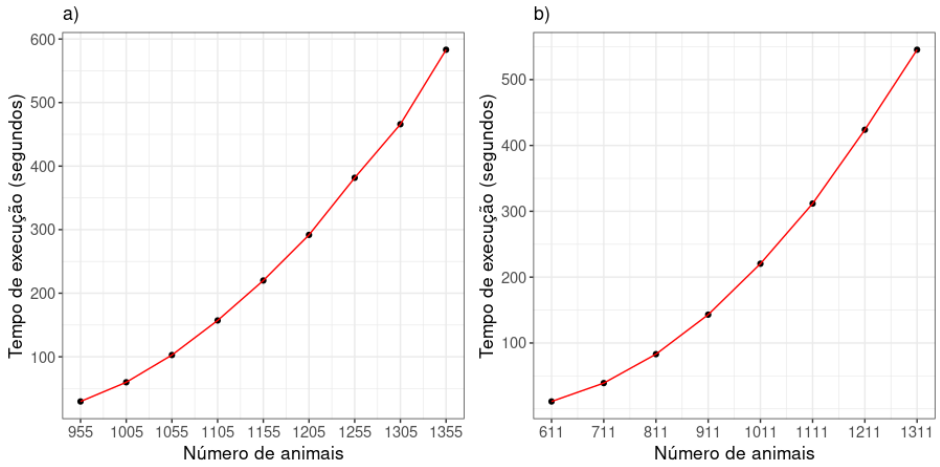


Figura 1. Gráficos dos resultados dos testes realizados com o BrangusSelection.

O formato da função do tempo de execução indicar uma complexidade polinomial tem implicações importantes. A complexidade do problema da programação linear é $O(nm)$ em que n é o número de variáveis e m é o número de restrições do problema [Daguspta et al. 2006]. No caso do BrangusSelection, visto que o número de variáveis do problema é $|M| \cdot |F|$ e o número de restrições é $|M| + |F|$ (cf. Seção 2), a complexidade é da ordem de $O(|M| \cdot |F| \cdot (|M| + |F|))$, ou seja, quadrática no número de animais. Se o algoritmo *branch and bound* não está influindo no aumento da complexidade, é uma pista importante de que o problema pode ter algoritmo ótimo mais eficiente que o resolva,

dispensando abordagens com meta-heurísticas que não garantem uma solução ótima ao final de sua execução.

A saída do algoritmo é uma tabela com a recomendação dos acasalamentos, já sumarizada a partir da matriz binária de saída do algoritmo (cf. definição do Problema 1). No arquivo de entrada, cada animal é identificado com um código, que é usado para apresentação do resultado, também em arquivo com formato csv. A Tabela 2 exibe um exemplo de recomendação de acasalamentos, obtida a partir da saída do programa. O valor da função objetivo, que determina o valor médio esperado da próxima geração, também é apresentado.

Tabela 2. Exemplo de recomendação de acasalamentos

Touro	**Vacas**			
L102	K482	K477	K500	I252
L840	L343	M76	L879	L363
M122	L372	M313	M212	M29
M175	M134	K407	K369	K331
M243	K554	M53	K418	M121
Valor função objetivo: 278,2408				

4. Conclusão

Este trabalho apresentou o algoritmo BrangusSelection, com o objetivo de produzir uma estratégia ótima de acasalamentos entre animais selecionados para reprodução. O algoritmo depende da existência de um índice de seleção, mas funciona independente de que índice é esse, que pode variar em relação às espécies, raças e objetivos de seleção. O exemplo usado envolveu gado bovino da raça Brangus, mas o algoritmo é genérico e funciona para qualquer espécie de animais.

A utilização do algoritmo depende a existência de duas bases de dados: dos valores individuais dos animais para cada caraterística de interesse (ou os valores das contribuições individuais dos animais, já computados) e da matriz de coancestralidade (ou de consanguinidade) existente entre os animais selecionados para acasalamento. Programas de melhoramento animal mantém esses dados, frequentemente no formato de planilhas eletrônicas e, com menos frequência, em bancos de dados. De toda forma, a informação está disponível e pode ser transformada em arquivos csv, para entrada no algoritmo. Os demais parâmetros – número máximo de vezes que cada macho pode ser utilizado no processo de acasalamento à frente e índice máximo de consanguinidade aceitável, que pode ser zero – são passados diretamente ao algoritmo.

O BrangusSelection está correntemente codificado na linguagem de programação R, com pacotes para otimizar os laços iterativos por meio da inserção de código em C++, usando pacotes específicos para o algoritmo simplex e o método *branch and bound*. Todo o ferramental tecnológico usado é livre e pode ser utilizado imediatamente. Estamos em processo de disponibilização de um pacote R para esse fim específico e, no futuro, vislumbra-se a incorporação do algoritmo a uma API (*application programming interface*) para disponibilização do código como serviço.

Os resultados referentes à ordem de complexidade do problema são outra contribuição importante deste trabalho. Algoritmos genéticos e outras meta-heurísticas são valiosas quando o problema a ser resolvido é computacionalmente intratável. O indicativo da execução deste trabalho, com resultado ótimo produzido em um tempo da ordem de minutos para uma base com cerca de 1500 animais é compatível com uma função polinomial quadrática. Isso quer dizer que o problema da maximização de seleção de acasalamentos é computacionalmente tratável. A prova de que o problema pertence à classe de complexidade P já foi construída [Ferreira et al. 2021]. De fato, o problema pode ser resolvido em tempo $O(|F| \cdot |M|)$.

Como trabalhos futuros, prevê-se a análise da complexidade de outros sistemas de acasalamentos. Particularmente, de acasalamentos compensatórios, em que o objetivo, juntamente com a solução máxima da contribuição, é produzir um rebanho mais homogêneo, com menos animais excepcionais mas também com menos animais com deficiências em relação ao objetivo de seleção. Este trabalho também já está em andamento. Estratégias para diminuir a consanguinidade do rebanho a longo prazo começam a ser estudadas também.

Referências

Abreu, U. G. P. d., Sonohata, M. M., e Lopes, P. S. (2013). Definição de pesos econômicos e de índices de seleção para sistemas de produção. In Rosa, A. d. N., Martins, E. N., Menezes, G. R. d. O., e Silva, L. O. C. d., editors, *Melhoramento Genético Aplicado em Gado de Corte. Programa Geneplus-Embrapa*, volume 1, chapter 11. Embrapa Gado de Corte.

Barreto Neto, A. D. et al. (2014). Estrutura populacional e otimização de esquemas de acasalamento em ovinos com uso de algoritmos evolucionários.

Berkelaar, M. et al. (2020). lpsolve: Interface to 'lp_solve' v. 5.5 to solve linear/integer programs. R package version 5.6.15.

Cardoso, F. (2009). Ferramentas e estratégias para o melhoramento genético de bovinos de corte. *Embrapa Pecuária Sul-Documentos (INFOTECA-E)*.

Carvalheiro, R., Queiroz, S. A. d., e Kinghorn, B. (2010). Optimum contribution selection using differential evolution. *Revista Brasileira de Zootecnia*, 39(7):1429–1436.

Carvalho, T., Santos, N., Lira, W., Oliveira, P. A., Neto, P. S., Lindenberg, J., e Rabêlo, R. (2016). Um sistema de informação para melhoramento genético de caprinos e ovinos. In *Anais Principais do XII Simpósio Brasileiro de Sistemas de Informação*, pages 100–107. SBC.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., e Stein, C. (2002). Algoritmos: teoria e prática. *Editora Campus*, 2:296.

Daguspta, S., Papadimitriou, C. H., e Vazirani, U. V. (2006). *Algorithms*. McGraw-Hill Higher Education, 1 edition.

Eddelbuettel, D. e Balamuta, J. J. (2018). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *The American Statistician*, 72(1):28–36.

Eler, J. P. (2017a). *Teorias e métodos em melhoramento genético animal: bases do melhoramento genético animal*. Faculdade de Zootecnia e Engenharia de Alimentos da USP.

Eler, J. P. (2017b). *Teorias e métodos em melhoramento genético animal: seleção*. Faculdade de Zootecnia e Engenharia de Alimentos da USP.

FAO (2009). Food and agriculture organization of the united nations (2009) how to feed the world 2050. Discussion paper prepared for Expert Forum: 12–13 October.

Ferreira, A. P. L., Yokoo, M. J.-I., e Motta, B. E. T. (2021). On the problem of optimal mating in animal breeding. In *Anales de CLEI 2021*.

Fontoura, D. C. N., Camargo, S. S., Jr, R. A. A. T., Carvalho, H. G., e Cardoso, F. F. (2020). Optimizing mate selection: a genetic algorithm approach. In *Proceedings of the ICAR Conference*, number 24 in ICAR Technical Series, Prague, CZ.

Hillier, F. S. e Lieberman, G. J. (2013). *Introdução à pesquisa operacional*. McGraw Hill Brasil.

Ihaka, R. e Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3):299–314.

Kim, V. (2020). lpsymphony: Symphony integer linear programming solver in r. http://R-Forge.R-project.org/projects/rsymphony, https://projects.coin-or.org/SYMPHONY, http://www.coin-or.org/download/source/SYMPHONY/.

Nieto, L. M., Alencar, M. M. d., e Rosa, A. d. N. (2013). Critérios de seleção. In Rosa, A. d. N., Martins, E. N., Menezes, G. R. d. O., e Silva, L. O. C. d., editors, *Melhoramento Genético Aplicado em Gado de Corte. Programa Geneplus-Embrapa*, volume 1, chapter 10. Embrapa Gado de Corte.

Simões, M. R. S., Leal, J. J. B., Minho, A. P., Gomes, C. C., MacNeil, M. D., Costa, R. F., Junqueira, V. S., Schmidt, P. I., Cardoso, F. F., Boligon, A. A., e Yokoo, M. J. (2020). Breeding objectives of brangus cattle in brazil. *Journal of Animal Breeding and Genetics*, 137(2):177–188.

CanControl: plataforma para diagnóstico do Cancro Europeu da Macieira

Wilson C. Branco Neto[1], Leonardo Araújo[2], Felipe A. M. Ferreira Pinto[2], Rodrigo A. Machado[3], Yan F. B. Ribeiro[3], Wilson F. Cordova Junior[3], Kevin Mondadori Mattos[3]

[1]Depto. de Ensino, Pesquisa e Extensão – Instituto Federal de Santa Catarina (IFSC). Lages – SC – Brasil.

[2]Estação Experimental de São Joaquim - Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri). São Joaquim – SC – Brasil.

[3]Curso de Ciência da Computação – Instituto Federal de Santa Catarina (IFSC). Lages – SC – Brasil.

wilson.castello@ifsc.edu.br,
{leonardoaraujo, felipepinto}@epagri.sc.gov.br,
{digao1297, yanburigoribeiro, wilsonfernandes382}@gmail.com,
kevin_mondadori@hotmail.com

Abstract. *This paper presents a platform created to publish information about Apple Canker and facilitate the diagnosis and monitoring of infected plants. It is made up of an API, a web system and a mobile application, which enable farmers to register suspicious plants and Epagri phytopathologists to remotely diagnose them. The application has been available since June of 2021 and has already been downloaded 110 times, 25 producers have registered themselves in the system and registered 23 suspicious plants*

Resumo. *Este artigo apresenta uma plataforma criada para divulgar informações sobre o Cancro Europeu da Macieira e facilitar o diagnóstico e monitoramento das plantas infectadas. Ela é composta por uma API, um sistema web e um aplicativo móvel, que possibilitam que os produtores cadastrem plantas suspeitas e os Fitopatologistas da Epagri realizem o diagnóstico de forma remota. O aplicativo está disponível desde junho de 2021 e já conta com 110 downloads, 25 produtores realizaram seu cadastro no sistema e registraram 23 plantas suspeitas.*

1. Introdução

O Cancro Europeu da Macieira (CEM), causado por *Neonectria ditissima,* é uma importante doença que atinge pomares em várias regiões do mundo [Weber 2014]. No Brasil, a presença da doença foi constatada pela primeira vez em 2002 em um viveiro no Rio Grande de Sul e, aproximadamente dez anos depois, iniciaram os registros da doença em diversos pomares comerciais do Rio Grande do Sul e de Santa Catarina [Alves e Czermainski 2015].

Segundo Lazzarotto e Alves (2015), "nas condições ambientais brasileiras, a doença tem se manifestado de maneira bastante agressiva", acarretando prejuízos em curto, médio e longo prazo. Os mesmos autores afirmam que o CEM pode inviabilizar a produção de maçãs nas regiões atingidas e projetam os resultados econômicos de quatro cenários distintos em relação ao nível de infecção e às medidas de controle adotadas:

- Cenário 1 - sem incidência: a lucratividade de um pomar é de 21,91% e sua vida útil de 20 anos;
- Cenário 2 - incidência baixa: a lucratividade cai para 17,33% e a vida útil permanece em 20 anos;
- Cenário 3 - incidência média a alta com medidas de controle: a vida útil permanece em 20 anos, mas a lucratividade cai para 8,86% em função do alto custo para controle da doença, bem como da redução de produtividade;
- Cenário 4 - incidência média a alta sem medidas de controle: além da lucratividade cair para 5,57%, a vida útil do pomar é de apenas 12 anos.

Devido à gravidade da situação, o Ministério da Agricultura Pecuária e Abastecimento publicou a instrução normativa – IN 20, de 20/06/2013 [Brasil 2013], que foi atualizada pela portaria 319, de 26/05/2021. Esta portaria, em seu artigo 1º, define como objetivo "Instituir o Programa Nacional de Prevenção e Controle do Cancro Europeu - PNCE, cujo agente etiológico é o fungo *N. ditissima"* [Brasil 2021].

A portaria estabelece, entre outros pontos, as medidas que devem ser adotadas pelos produtores e entidades envolvidas. Merecem destaque dois artigos. O artigo 13º determina que em Unidades de Produção (UP) com incidência de até 5% de plantas infectadas, as plantas com sintomas deverão ser arrancadas e incineradas, e as plantas vizinhas marcadas e vistoriadas mensalmente. O artigo 14º estabelece que nas UP com incidência superior a 5% de plantas infectadas, deverão ser realizados procedimentos para prevenção e controle do CEM, tais como a eliminação e incineração de partes da planta ou a limpeza e tratamento do mesmo [Brasil 2013].

Além das normativas federais, o Governo do Estado de Santa Catarina sancionou a lei 17.825/2019 [Santa Catarina 2019]. Dentre outros aspectos, esta lei determina como deve ser feita a fiscalização das propriedades e as penalidades previstas, como demonstra o "Art. 22. Sem prejuízo das responsabilidades penal, ambiental e civil cabíveis, fica o infrator sujeito às seguintes penalidades, aplicadas de maneira isolada ou cumulativa: I - Advertência; II - Multa; e III - Suspensão ou cancelamento de autorizações, registros, inscrições, credenciamentos, cadastros, habilitações, certificados ou documentos de trânsito.

Com a obrigatoriedade da fiscalização e as penalidades impostas por esta lei, pretende-se evitar que o CEM se dissemine em Santa Catarina como aconteceu no Rio

Grande do Sul, pois a situação da incidência da doença nos pomares difere nos dois principais estados produtores de maçã no Brasil.

No Rio Grande do Sul, segundo a Secretaria de Estado de Agricultura, Pecuária e Irrigação (SEAPI), 51% das propriedades já apresentam plantas infectadas. Um levantamento realizado em 277 propriedades, entre 2015 e 2016, mostrou que aproximadamente 60% destas já possuíam incidência superior a 1% do total de plantas, o que desobrigava os produtores a arrancarem e incinerarem as plantas, pois este era o percentual definido pela IN 20/2013 vigente na época [Araujo *et al*. 2019]. A incidência da doença no estado de Santa Catarina é menor. Conforme dados da CIDASC (Companhia Integrada de Desenvolvimento Agrícola de Santa Catarina), apresentados por Araujo *et al.* (2019), aproximadamente 10% das unidades produtivas já tiveram registro de ocorrência do CEM, destas, 93% possuíam menos de 1% de incidência da doença nos pomares.

Em função destas diferenças, a Epagri (Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina) e a CIDASC trabalham na tentativa de erradicação ou de manutenção dos baixos índices de incidência da doença no estado. Para que estes objetivos sejam alcançados, é fundamental realizar a diagnose de forma correta e o mais cedo possível. No entanto, no processo de identificação do CEM nos pomares, muitos técnicos e fruticultores podem ficar em dúvida se a lesão encontrada em uma macieira é causada por *N. ditíssima* ou não, devido ao processo de diagnose adotado.

Atualmente, o diagnóstico é realizado de forma empírica pelos produtores e responsáveis técnicos, os quais, em caso de dúvida, coletam material e enviam para análise na Estação Experimental de São Joaquim (Epagri-EESJ). A falta de conhecimento para um diagnóstico preciso a campo e a necessidade de deslocamento para análise do material suspeito pode levar a diagnósticos incorretos ou, até mesmo, a sua não efetivação. Por isto, o desenvolvimento de tecnologias que viabilizem um diagnóstico mais rápido e fácil é importante porque possibilita que a planta seja erradicada antes que ocorra a disseminação da doença para outras macieiras saudáveis.

Este artigo apresenta uma plataforma desenvolvida em um projeto realizado em parceria entre o Instituto Federal de Santa Catarina (IFSC) e a Epagri, com o objetivo de auxiliar os produtores e responsáveis técnicos a realizar o diagnóstico correto, assim como difundir o conhecimento existente sobre a doença.

2. Materiais e Métodos

Para atingir os objetivos propostos, foram desenvolvidos três sistemas, como pode ser visto na Figura 1, para atender dois perfis de usuários: 1) Produtores/ Responsáveis Técnicos; 2) Fitopatologistas da Epagri-EESJ. Os produtores e responsáveis técnicos utilizam o aplicativo móvel para efetuar o seu cadastro no sistema e registrar plantas com suspeita de infecção. Os fitopatologistas da Epagri-EESJ, por meio do sistema web, analisam os registros feitos e emitem o diagnóstico. Além das funcionalidades relacionadas ao diagnóstico das plantas, tanto o aplicativo móvel quanto o sistema web apresentam um conjunto de informações em formato de textos, fotos e vídeos que visam difundir o conhecimento sobre a doença.

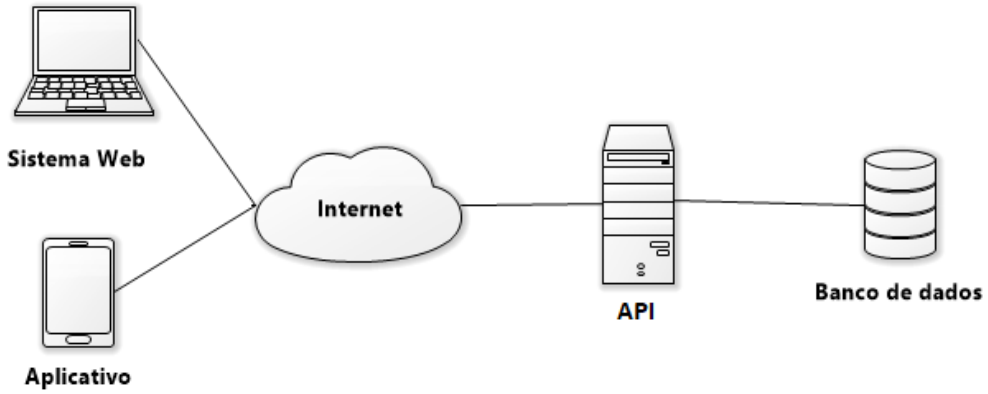


Figura 1. Arquitetura da plataforma

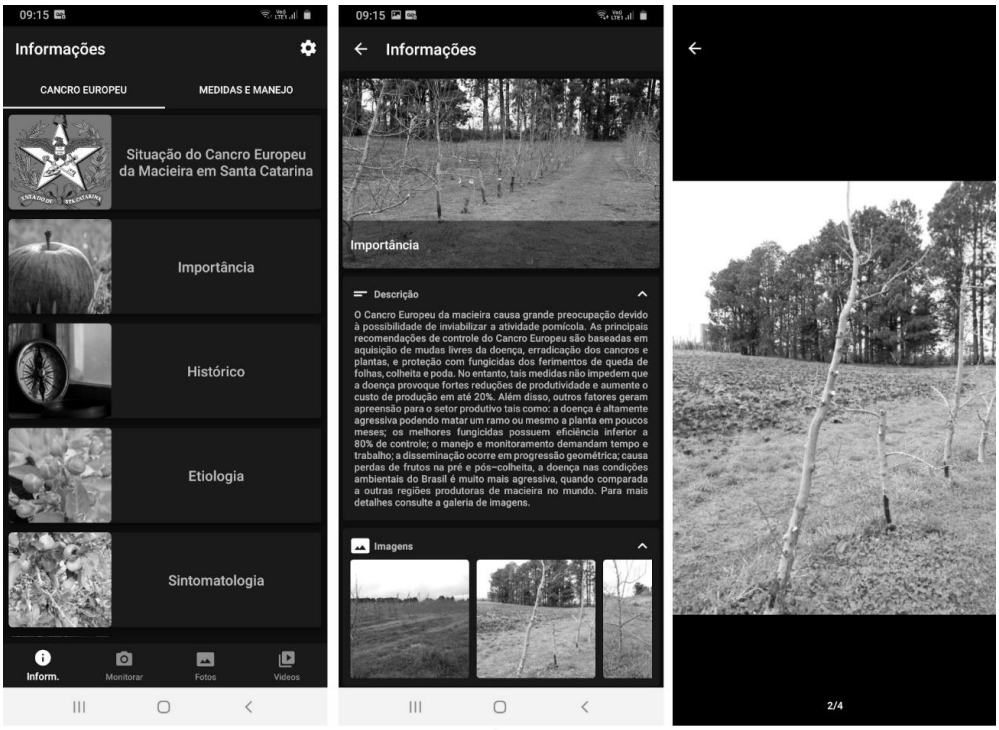
O banco de dados que armazena as informações dos produtores e responsáveis técnicos cadastradas, bem como os monitoramentos realizados por eles, foi desenvolvido em PostgreSQL. A API foi construída em JavaScript, sendo executada sobre o NodeJS. Na API estão implementados todos os métodos de acesso ao banco de dados, a lógica de negócio e os aspectos de segurança. A versão disponível do aplicativo móvel foi desenvolvida em Kotlin e, por isto, pode ser executada apenas em dispositivos com o sistema Android. Uma nova versão do aplicativo móvel está sendo desenvolvida em Flutter para atender os usuários que possuem o sistema IOS. O sistema web foi construído usando o framework Vue.js.

2.1 Aplicativo móvel

As funcionalidades do aplicativo móvel estão organizadas em quatro seções: *Informações*, *Fotos*, *Vídeos* e *Monitorar.* A seção de *Informações* apresenta textos explicativos sobre a doença, sintomas, disseminação, estratégias de controle etc., produzidos pelos fitopatologistas da Epagri-EESJ. A Figura 2 apresenta a interface desta seção com os tópicos informativos sobre a doença (2a); ao clicar sobre um deles, uma página é aberta com o texto e imagens (2b); e ao clicar sobre uma das imagens é possível ampliá-la para uma análise detalhada (2c).

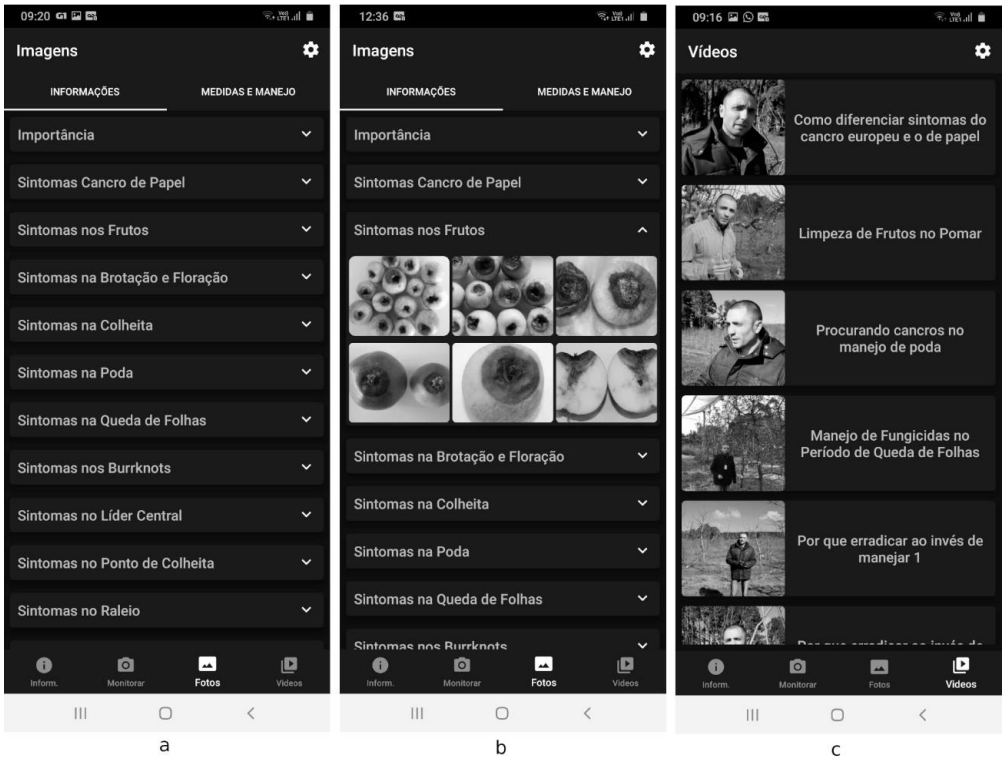
O aplicativo também possui uma seção com imagens que apresentam lesões em diferentes partes da planta causadas pelo CEM e, também, por outras doenças com sintomas semelhantes para que o usuário consiga diferenciá-las. A Figura 3 apresenta a interface com a lista de categorias de imagens (3a); ao clicar sobre uma das categorias são apresentadas miniaturas das imagens daquela categoria (3b); por fim, ao clicar sobre uma das imagens é possível visualizá-la ampliada, bem como aplicar zoom a ela para analisá-la em detalhes de forma semelhante ao apresentado na Figura 2(c).

Além dos textos e imagens, o aplicativo apresenta vídeos que explicam diferentes aspectos da doença e demonstram como realizar o diagnóstico a partir de sintomas nas diferentes partes da planta e em diferentes estádios fenológicos. A Figura 3(c) apresenta parte da lista com os vídeos disponíveis. De forma semelhante à seção de informações, a obtenção e seleção das imagens e a produção dos vídeos foram realizadas pelos fitopatologistas da Epagri-EESJ.



a b c

Figura 2. Interface do aplicativo móvel – Informações



a b c

Figura 3. Interface do aplicativo móvel – Imagens e Vídeos

Outra funcionalidade do aplicativo móvel é a seção de monitoramento apresentada na Figura 4. Ela permite que o usuário cadastre os monitoramentos realizados, preenchendo os dados solicitados e registrando as imagens da planta (4a); que ele acompanhe e visualize o resultado do diagnóstico emitido pelos pesquisadores da Epagri (4b); e que ele localize a planta infectada por meio de suas coordenadas geográficas (4c) para arrancá-la e incinerá-la caso o diagnóstico seja positivo.

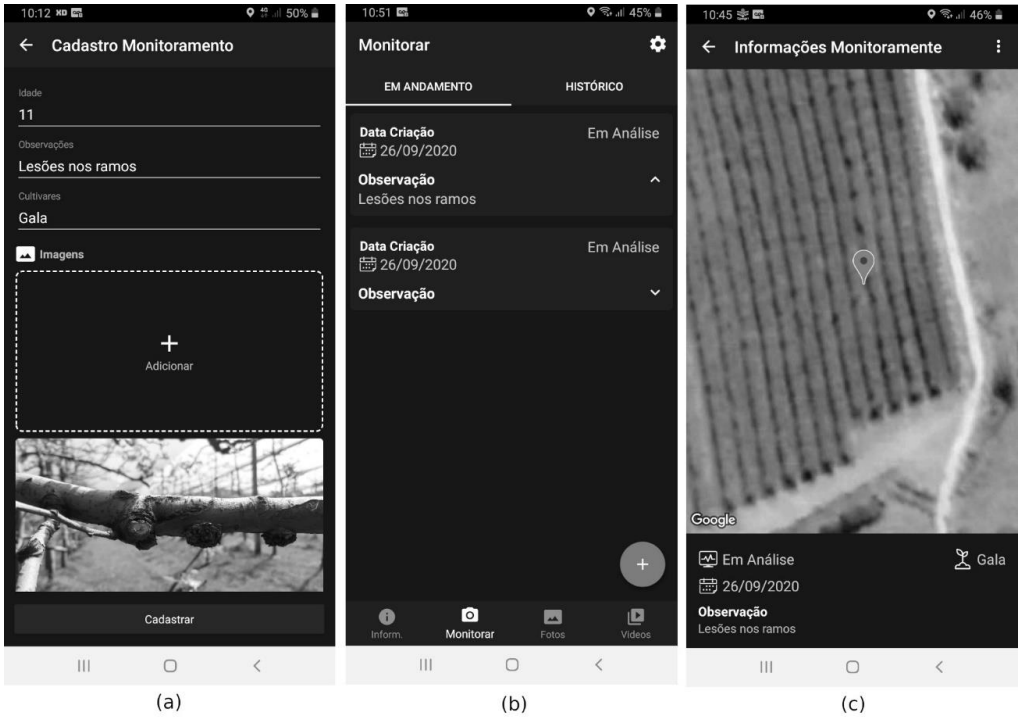


Figura 4. Interface do aplicativo móvel – Monitoramento

Apesar do aumento da cobertura da Internet nas áreas rurais do Brasil, ainda existem muitas regiões que não possuem conexão à rede. Por isto, o aplicativo móvel foi desenvolvido para funcionar *offline*. O produtor tem acesso a todos os textos explicativos e fotos, e pode fazer o seu cadastro e o registro de plantas suspeitas mesmo desconectado da Internet. Ao retornar a um ponto com acesso à Internet, o sistema conecta-se à API e sincroniza os dados registrados automaticamente. A única função do sistema que não funciona *offline* é a de vídeos, pois eles estão armazenados no Youtube e não no aplicativo para não sobrecarregar o dispositivo dos usuários.

Também com o intuito de melhorar a usabilidade do sistema, foi criado um mecanismo de notificação para avisar o usuário quando os fitopatologistas emitem o diagnóstico do monitoramento registrado por ele, evitando que ele precise ficar entrando no aplicativo para verificar se a resposta já foi enviada ou não.

2.2 Sistema Web

O sistema web foi desenvolvido para que os fitopatologistas possam analisar os monitoramentos enviados pelos produtores. Além disto, ele apresenta as mesmas

informações sobre o CEM para os produtores e responsáveis técnicos que prefiram usá-lo para visualizar os textos, fotos e vídeos. Ele também possibilita que os fitopatologistas da Epagri, mediante login e senha, visualizem um mapa georreferenciado dos casos registrados, para auxiliar as instituições responsáveis pelo monitoramento e fiscalização da doença.

A interface inicial do site apresenta um menu que leva às seções de *Informações*, *Imagens*, *Vídeos* e *Arquivos*. As três primeiras possuem as mesmas informações disponíveis no aplicativo e a última contempla documentos tais como legislações, artigos científicos, boletins epidemiológicos, dentre outros, que não foram incorporados ao aplicativo para reduzir seu tamanho.

Por meio da opção *Login*, os pesquisadores da Epagri têm acesso à parte interna do site que apresenta os monitoramentos realizados pelos diferentes produtores e responsáveis técnicos. Ao clicar sobre um dos monitoramentos da lista, o sistema apresenta as informações detalhadas e as imagens para que o diagnóstico possa ser emitido, por meio da interface apresentada na Figura 5.

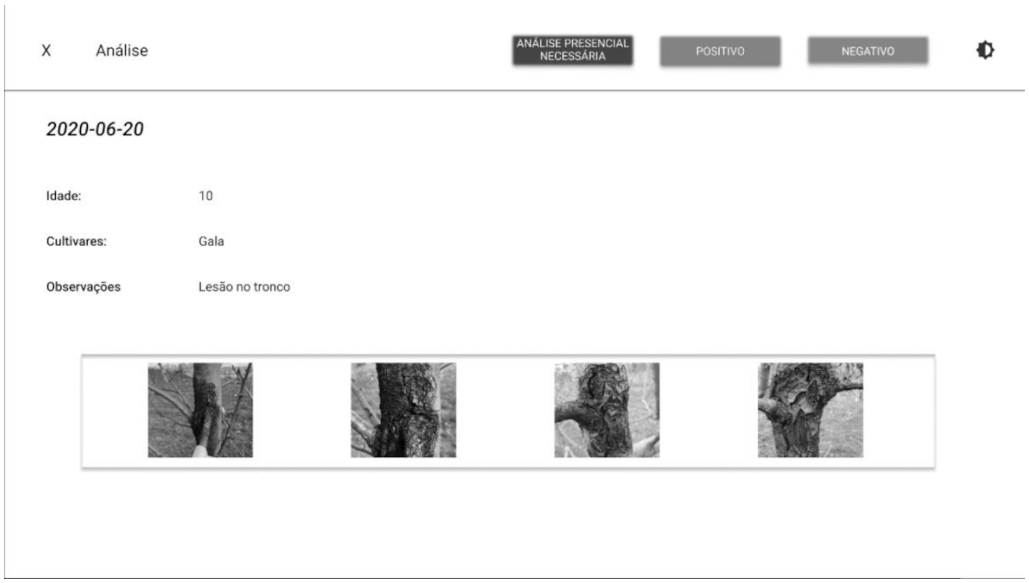


Figura 5. Interface para análise e realização do diagnóstico

O resultado do diagnóstico pode ser:

- *Positivo*, quando é possível identificar, por meio das fotos cadastradas pelos produtores e responsáveis técnicos, que se trata do CEM;
- *Negativo*, quando é possível descartar a doença, com base nas imagens enviadas;
- *Análise Presencial Necessária*, quando não é possível concluir o diagnóstico a partir das fotos. Neste caso, o produtor deve manter a prática atual e enviar o material para análise junto à Clínica Fitopatológica da Epagri-EESJ.

3. Resultados e Discussão

O aplicativo Android foi disponibilizado na Google Play[1] no dia 02 de junho de 2021 e até o momento atingiu a marca de 110 *downloads*. Dos usuários que instalaram o aplicativo, 25 já realizaram o seu cadastro e 23 monitoramentos foram registrados. Destes monitoramentos, 9 foram diagnosticados como *Positivo,* 10 como *Negativo* e 4 como *Análise Presencial Necessária*, o que demonstra a capacidade da plataforma desenvolvida para reduzir o tempo para realização da maioria dos diagnósticos e evitar o deslocamento físico até a Epagri-EESJ (19 de 23 casos). O sistema web está temporariamente disponível em um servidor do IFSC[2], mas uma parceria recém firmada entre IFSC, Epagri e ABPM (Associação Brasileira dos Produtores de Maçã), garantiu recursos para que ele seja transferido para um servidor privado, com domínio próprio a ser registrado em breve.

Estes números devem crescer nos próximos meses, pois os fitopatologistas da Epagri-EESJ estão elaborando material e planejando ações de divulgação junto aos produtores e responsáveis técnicos pelas propriedades. A publicação da versão IOS, prevista para setembro de 2021, também deve contribuir para o aumento dos números, pois atingirá um grupo de usuários que não pode se cadastrar e registrar monitoramentos por não ter um dispositivo compatível com o sistema Android.

4. Conclusões

Este artigo apresentou os sistemas desenvolvidos para auxiliar na divulgação de informações sobre o CEM, seu diagnóstico e monitoramento. Os textos, fotos e vídeos disponíveis no aplicativo móvel e no site devem elevar o nível de conhecimento dos produtores e responsáveis técnicos, possibilitando que eles consigam diagnosticar corretamente a doença para tomar as medidas definidas nas normatizações relacionadas.

Além disto, a funcionalidade de monitoramento é capaz de aproximar os produtores dos fitopatologistas da Epagri, que são profissionais com maior conhecimento sobre o assunto e capacidade de diagnóstico, nos casos em que a dúvida persista. Desta forma, pretende-se aumentar a quantidade de diagnósticos corretos realizados, ao evitar a análise presencial do material, contribuindo para auxiliar o produtor na tomada de decisão, facilitar o manejo, reduzir a disseminação da doença e a consequente inviabilização da produção de maçãs no sul do Brasil.

Como trabalhos futuros, além da conclusão do aplicativo para iOS e realização das ações de divulgação da plataforma, pretende-se iniciar estudos na área de Redes Neurais para criar um módulo que seja capaz de analisar automaticamente as imagens enviadas e realizar o diagnóstico em parte delas, sem a necessidade de intervenção por parte dos fitopatologistas.

Referências

Alves, S. A. M. e Czermainski, A. B. (2015), "Síntese descritiva de eventos relacionados ao cancro europeu das pomáceas no Brasil". Nota Técnica Embrapa, Bento Gonçalves.

[1] https://play.google.com/store/apps/details?id=br.edu.ifsc.cancontrol

[2] http://www.cancroeuropeu.com.br

Araujo, L., Pinto, F. A. M. F & Vieira, J. S. (2019), "Situação do Cancro Europeu no Brasil." *In:* Cancro Europeu no Brasil. Organizado por Alves, S. A. M. e Czermainski, A. B (org). p. 33-41. Brasília: Embrapa.

Brasil. (2013), Ministério da Agricultura, Pecuária e Abastecimento. Instrução normativa n. 20, de 20 de jun. de 2013.

Brasil. (2021), Ministério da Agricultura, Pecuária e Abastecimento. Portaria n. 319, de 26 de maio de 2021.

Lazzarotto, J. J. e Alves , S. A. M. (2015), "Prejuízos Econômicos e Financeiros Associados ao Cancro Europeu em Sistemas de Produção de Maçã de Vacaria, RS". Comunicado Técnico EMBRAPA, Bento Gonçalves, v. 169, mar. 2015.

Santa Catarina. (2019), Lei estadual nº 17.825 de 12 de dezembro de 2019. Dispõe sobre a defesa sanitária vegetal no estado e estabelece outras providências. Diário Oficial do Estado de Santa Catarina, Florianópolis, SC, 16 Dez. 2019.

Weber, R. W. S. (2014), "Biology and control of the apple canker fungus Neonectria ditissima (syn. N. galligena) from a Northwestern European perspective". In: Erwerbs-Obstbau, v. 56, n. 3, p. 95–107.



Controle de temperatura e de umidade de uma composteira utilizando o ESP32

Luan D. Rodrigues[1], Rodrigo M. Pereira[1], David Willy F. Palmera[1], Guilherme Henrique A. Silva[1], Marco Antonio F. Leite[1], Juliana S. Silva[2]

[1]Departamento de Área de Eletroeletrônica, Campus Cuiabá – Cel. Octayde Jorge da Silva, Instituto Federal do Mato Grosso, Cuiabá, Mato Grosso, Brasil.

[2]Departamento de Área de Computação, Campus Cuiabá – Cel. Octayde Jorge da Silva, Instituto Federal do Mato Grosso, Cuiabá, Mato Grosso, Brasil.

luan_ds.rs@hotmail.com, rodrigom3317@gmail.com, david_willyfla@gmail.com, guilhermehenriquee9@gmail.com, marcofranzin1999@gmail.com, juliana.silva@cba.ifmt.edu.br

Abstract. *The Organic Agriculture has become applicable to several crops, due to its low environmental impact, since it does not use chemical inputs and genetically modified organisms in cultivation, as occurs, in general, in conventional agriculture, which is widely spread in Brazil. However, it is noteworthy that there are difficulties in the practice of organic agriculture, since the microorganisms responsible for the decomposition process of organic matter need a suitable environment for their proliferation. Therefore, the objective of this article is to present the creation of a composter prototype, with automatic temperature and humidity control. Thus, for the development of the prototype, the following software were used: (i) Arduino IDE; (ii) Fritzing; and (iii) Inventor. The hardware components were: (i) ESP32 microcontroller; (ii) Humidity sensor; (iii) Stepper motor; and (iv) Driver A4988. With the combination of hardware and software resources, it was possible to build the prototype. The results were satisfactory, because the composter structure was controlled through the analysis of information measured by sensors, using ESP32.*

Resumo. *A Agricultura Orgânica vem tornando-se aplicável em diversos cultivos, em razão do seu baixo impacto ambiental, pois em seu cultivo não há a utilização de insumos químicos e organismos geneticamente modificados, como ocorre, em geral, na agricultura convencional que é amplamente difundida no Brasil. No entanto, vale ressaltar que existem dificuldades para a prática da agricultura orgânica, visto que os microrganismos responsáveis pelo processo de decomposição da matéria orgânica necessitam de um meio propício à sua proliferação. Portanto, o objetivo deste artigo é apresentar a*

criação de um protótipo de uma composteira, com controle automático de temperatura e umidade. Assim, para o desenvolvimento do protótipo, foram utilizados os seguintes software: (i) Arduino IDE; (ii) Fritzing; e (iii) Inventor. Os componentes de hardware foram: (i) Microcontrolador ESP32; (ii) Sensor de umidade; (iii) Motor de passo; e (iv) Driver A4988. Com a junção dos recursos de hardware e software foi possível confeccionar o protótipo. Os resultados alcançados foram satisfatórios, pois a estrutura da composteira foi controlada por meio da análise das informações medidas pelos sensores, utilizando o ESP32.

1. Introdução

Entre os séculos XVIII e XIX, ocorreu a revolução agrícola, trazendo como elementos basilares técnicas modernas, como o uso de fertilizantes, agrotóxicos, organismos geneticamente modificados e a monocultura – que contribuíram para o aumento da produção e a redução de custos [Bin 2004]. No entanto, isso ocasionou uma série de impactos ambientais, tais como, degradação do solo, contaminação da água, intoxicação de agricultores e aparecimento de novas pragas. Levando em consideração todos esses problemas, surgiu o conceito de uma agricultura mais sustentável, sendo denominada de Agricultura Orgânica [São Paulo 2011].

Foi justamente nesse contexto que, no decorrer da década de 1940, o pesquisador inglês Albert Howard, durante pesquisas em viagens à Índia, examinou a utilização de meios orgânicos na agricultura, a partir do uso de práticas de adubação orgânica. De forma paralela, outros pesquisadores também realizaram estudos sobre a implementação da agricultura biológica, com escasso uso de insumos agroquímicos [Ormond *et al.* 2014]. Com isso, a Agricultura Orgânica nasceu como uma forma de manejo voltada a mitigar os impactos ao meio ambiente, pois se utiliza de insumos de origem orgânica, gerados por meio de técnicas como a compostagem; prevenindo, assim, o uso de agrotóxicos e organismos geneticamente modificados [Alves, Santos e Azevedo 2012].

A compostagem é um método de decomposição da matéria orgânica de forma controlada. Apesar dos benefícios advindos do seu uso, como baixo custo e menor impacto ambiental, é importante lembrar que consiste em um método de decomposição da matéria difícil de ser realizado, uma vez que há a necessidade da intervenção de microrganismos decompositores, que dependem de um meio favorável à sua sobrevivência. Por conseguinte, diversos fatores precisam ser controlados para que a compostagem gere um insumo desejável – como a temperatura, a umidade, a aeração, o pH do solo, a granulometria e a relação Carbono e Nitrogênio (C/N) [Valente *et al*. 2009]. Alguns desses fatores (como a temperatura e a umidade) podem ter um controle automático, o que pode contribuir para a redução no tempo de trabalho do produtor rural. Caso o controle ocorra de forma adequada, o processo da compostagem tende a ser otimizado.

Diante desta conjuntura, surge uma oportunidade para se realizar diversas pesquisas a respeito dessa temática. Sendo assim, o objetivo deste artigo é apresentar a criação de um protótipo de uma composteira, com controle automático de temperatura e umidade, utilizando o ESP32.

Desta forma, o artigo está dividido em cinco seções, incluindo esta introdução. A seção 2 traz os conceitos teóricos basilares para a compreensão acerca dos temas trabalhados no decorrer da pesquisa. A seção 3 apresenta os materiais e métodos utilizados no desenvolvimento do protótipo. Na sequência, a seção 4 aborda o processo de construção da composteira. Por fim, a seção 5 explicita as principais vantagens e dificuldades para a elaboração deste protótipo.

2. Revisão Bibliográfica

Esta seção tem como objetivo descrever, de forma sucinta, os termos fundamentais para a compreensão desta pesquisa, como Agricultura Orgânica, Compostagem e Dispositivos Computacionais.

2.1. Agricultura Orgânica

A Agricultura Orgânica consiste na produção de alimentos utilizando os recursos de matéria orgânica presentes no solo, sem a adição de fertilizantes químicos artificiais, que são necessários para o desenvolvimento do cultivo agrícola. A presença desse material no solo é resultado do trabalho dos microrganismos já estabelecidos nos compostos biodegradáveis [Ormond *et al.* 2014].

Essa prática auxilia na diminuição da degradação ao meio ambiente, visto que não há a utilização de agrotóxicos ou a alteração nas características genéticas do cultivável [Alves, Santos e Azevedo 2012]. No entanto, quando se utiliza de agrotóxicos, observam-se diversos impactos negativos (tais como: erosão, intoxicação do solo e poluição da água), tanto para a saúde dos agricultores, quanto para os que consomem esses produtos [Souza 2005]. Em função dessa realidade, nota-se uma preocupação crescente com o meio ambiente, a saúde e o bem estar humano – fatores que têm contribuído para o incremento do mercado de produtos orgânicos no Brasil [Barbosa e Sousa 2012].

É importante mencionar, ainda, que existem diversas estratégias para a produção de matéria orgânica, a ser utilizada nas plantações, dentre as quais a compostagem é um exemplo – como descrito a seguir.

2.2. Compostagem

A compostagem consiste em um processo de decomposição orgânica, gerado a partir de ações de um conjunto diversificado de microrganismos (de forma aeróbica ou anaeróbica), que são realizadas em uma série de etapas, de forma sequencial, que são divididas em dois grupos, sendo eles: termofílica e mesofílico. No meio deste processo, várias ações favorecem condições termofílicas, gerando calor, que é dissipado em todo o composto [Oliveira, Sartori e Garcez 2008].

Assim, o objetivo principal da compostagem é gerar um produto que possa ser utilizado como adubo orgânico, que tem um alto valor químico-físico-biológico [Pires e Ferrão 2017]. Um exemplo de uma composteira manual, frequentemente utilizada em residências, onde o indivíduo deve realizar o manuseio do composto com as próprias mãos, sendo disposto na Figura 1.



Figura 1. Composteira em balde [Lobe 2019]

Para que o processo de compostagem seja eficiente, o composto precisa estar propício à vida, ou seja, deve estar bem aerado, possibilitando a presença de oxigênio.

Desta forma, as temperaturas, dentro de uma composteira, devem passar dos 50°C (para eliminar agentes patológicos), mas não exceder à 70°C – pois, assim, a vida dos microrganismos será afetada. Ademais, a umidade no meio do composto deve estar entre 40% e 60%, pois, abaixo desse valor, a atividade microbiana pode ser inibida e, acima, gera condição de anaerobiose (ações geradas por organismos que não usam oxigênio) [Valente *et al.* 2009].

Além disso, para facilitar o monitoramento e o controle das variáveis umidade e temperatura, pode-se utilizar de dispositivos computacionais de apoio, como sensores e microcontroladores – como descrito a seguir.

2.3. Dispositivos Computacionais

Microcontrolador é um dispositivo computacional [Oliveira Júnior e Duarte 2010], constituído pela junção de diversos componentes, em um único *chip*: um processador, portas de entrada/saída, memória de acesso aleatório (*Random Access Memory* – RAM) e memória somente de leitura (*Read Only Memory* – ROM) [Silva, Targa e Cezar, 2019].

Existem diversas aplicações dos microcontroladores, destacando-se a Automação Residencial e a Agricultura de Precisão [Maia 2012]. Para realização dessas atividades, são desenvolvidas programações específicas, utilizando-se, em geral, as linguagens *Python* ou *C*.

Dentre os diversos modelos de microcontroladores disponíveis no mercado e de ampla utilização em pesquisas estão: o *Arduino*[1] e o *ESP32*[2] [Pereira 2020]. Para desenvolver o projeto em questão, utilizou-se o *ESP32*, por possuir uma maior conectividade com a rede externa (internet).

Ademais, para realizar o monitoramento de diversas variáveis do ambiente (como umidade, temperatura, índice pluviométrico, incidência solar, acidez do solo, entre outros), torna-se necessário utilizar alguns sensores.

Os sensores são equipamentos que podem receber estímulos do ambiente, sendo sensíveis à luminosidade, energia cinética e térmica. Com isso, têm a capacidade de transformar esses estímulos em valores de grandezas conhecidas (tais como: temperatura, pressão, velocidade, corrente elétrica, aceleração e posição), em um sinal de saída

[1] Informações adicionais sobre a placa microcontrolada Arduino estão disponíveis no site do fabricante: https://www.arduino.cc

[2] Detalhes a respeito da placa microcontrolada ESP32 podem ser consultadas no *Datasheet*: https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf.

[Wendling 2010]. Existem dois tipos de sensores, o analógico e o digital, sendo separados pelo tipo de sinal de saída [Thomazini e Albuquerque 2011] – ambos foram utilizados no desenvolvimento do protótipo.

Visto que alguns conceitos essenciais para o trabalho foram expostos, a seção seguinte dedica-se à apresentação do material e métodos utilizados na criação do protótipo.

3. Material e Metódos

O presente trabalho tem como objetivo a elaboração de um protótipo para realizar o controle de temperatura e umidade de uma composteira utilizando o microcontrolador ESP32.

Para tanto, o desenvolvimento deste trabalho compreendeu às seguintes etapas: (i) Revisão da literatura; (ii) Identificação dos problemas enfrentados para o pleno funcionamento de uma composteira; (iii) Desenvolvimento do protótipo de uma composteira, utilizando recursos de *hardware* e *software*; e (iv) Teste e análise dos resultados.

Na etapa (i) houve a busca por artigos na área, com o propósito de apresentar conceitos básicos para facilitar o entendimento do tema. Na sequência, a etapa (ii) consistiu no mapeamento dos recursos de *hardware* e *software* necessários para o desenvolvimento do protótipo, analisando o custo-benefício gerado, para o proprietário, com a implementação do equipamento.

Com a escolha dos equipamentos que seriam utilizados, a etapa (iii) compreendeu a criação do protótipo para o controle de temperatura e umidade de uma composteira, utilizando o microcontrolador ESP32. Assim, foi realizado a programação e a simulação do protótipo, utilizando-se os recursos de *software* e de *hardware*, dispostos, respectivamente, nos Quadros 1 e 2.

Quadro 1. *Softwares* utilizados

Software	**Definição**	**Figura**
Arduino IDE	Ambiente de desenvolvimento integrado, utilizado para a implementação do código de programação. A versão adotada foi a 1.8.9.	
Fritzing	Ambiente utilizado para realizar a simulação dos circuitos necessários do protótipo; versão empregada - 0.9.3b.64.	
Inventor	Possibilitou a criação da estrutura virtual do protótipo; versão utilizada – *Autodesk Inventor* 2019.	

Fonte: elaborado pelos autores

Quadro 2. Componentes de *Hardware utilizados*

Hardware	**Definição**	**Figura**
ESP32	Consiste em um microcontrolador, que foi utilizado para receber o código do compilador, para realizar o acionamento e a leitura dos dados da composteira.	
Sensor de Umidade	Sensor de umidade capacitivo analógico, utilizado para a medição da umidade do solo.	
Motor de Passo	Motor com precisão de posicionamento.	
A4988	*Driver* utilizado para controlar a rotação do motor de passos.	

Fonte: elaborado pelos autores

A última etapa desta pesquisa (iv), consistiu na elaboração de testes e na análise dos resultados, por meio dos quais foi possível identificar os principais benefícios da implementação do protótipo.

Após a apresentação das etapas e dos recursos de *software* e *hardware* utilizados nesta pesquisa, a próxima seção tem por objetivo descrever o processo de desenvolvimento do protótipo da composteira.

4. Protótipo da composteira automatizada

As principais etapas a serem considerados para a criação de uma composteira, com controle automático de umidade e temperatura, estão descritas a seguir.

4.1. Etapa 1: Simulação dos Circuitos

Esta etapa consistiu na simulação dos circuitos necessários para a construção do protótipo, por meio do programa *Fritzing*. Na simulação realizada (Figura 2), onde foram conectados todos os dispositivos utilizados para montagem do protótipo, juntamente com a alimentação necessária. Com isso, foi possível estabelecer os testes dos circuitos de monitoramento de umidade, temperatura e rotação do recipiente da composteira.

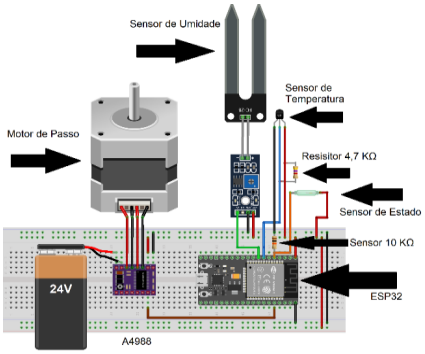


Figura 2. Montagem Virtual do Circuito utilizado no protótipo da composteira

4.2. Etapa 2: Desenvolvimento da programação

Esta etapa compreendeu a criação do código fonte responsável pelo controle das variáveis temperatura e umidade, bem como da rotação do motor. A programação deste código fonte foi realizada no *Software Arduino IDE*, utilizando a linguagem *C*.

Como requisito para o desenvolvimento deste código fonte, foi definido que a rotação do balde da composteira seria constituída por um motor, que foi iniciado, automaticamente, sempre que os valores dos parâmetros de umidade e temperatura excedessem um valor estipulado no código fonte. Ademais, as variáveis de temperatura e umidade foram mensuradas, periodicamente, por meio do sensor *DS18B20* e do Módulo Sensor de Umidade para Solos.

É importante registrar que a versão final do código desenvolvido e a esquemática para a criação do projeto estão disponíveis para a visualização no *site* do *GitHub*, na seguinte *url*: https://github.com/Luan-DR/Controle_Automatico_de_Uma_Composteira.

4.3. Etapa 3: Estrutura da composteira

Na terceira etapa, houve a elaboração da modelagem de uma estrutura da composteira, por meio do *software Inventor*, o que possibilitou a visualização do modelo 3D (Figura 3), para a construção da composteira.

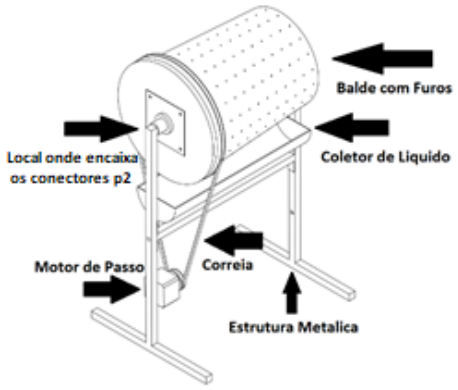


Figura 3. Modelo 3D da Composteira confeccionado no software Inventor

Uma vez que foram explicitados os passos utilizados para a montagem do protótipo, a próxima seção se dedica a abordar o funcionamento do equipamento desenvolvido.

5. Resultados e Discussão

Como resultado deste trabalho, foi possível construir o protótipo final da composteira, como demonstrado na Figura 4. Neste protótipo, o sensor de temperatura *DS18B20* (disposto na Figura 4a) realizou sua comunicação por meio do protocolo *One Wire*, que, para sua utilização, necessitou das seguintes bibliotecas: *OneWire.h* e *DallasTemperature*. Estas bibliotecas têm a finalidade de interpretar os valores obtidos pelo sensor de temperatura.

Além disso, o sensor de umidade (também disposto na Figura 4a) realizou a leitura de dados analógicos, cujas tensões variaram entre 0V e 3,3V. Assim, para que o usuário conseguisse manuseá-lo foi necessário realizar uma interpolação dos valores dos sinais, que relacionou o valor analógico, aferido no sensor, com a porcentagem de água presente no composto.

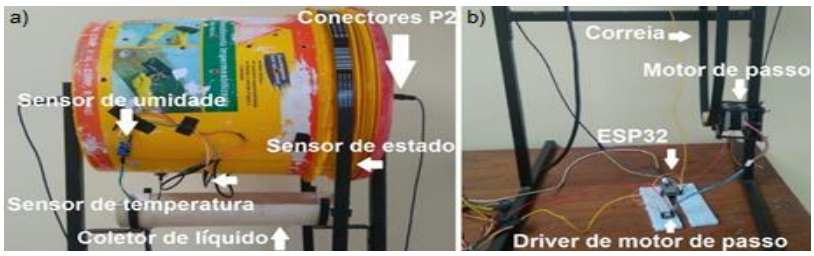


Figura 4. Protótipo final da composteira

Já a utilização dos conectores P2 (Figura 4a), justificou-se por ser uma forma simples para o transporte de dados dos sensores para o microcontrolador e, assim,

possibilitou a rotação em torno de um eixo, mesmo com o *ESP32* estando fora do balde (Figura 4b), o que facilitou sua alimentação e o controle do motor.

É importante salientar, também, que a estrutura da composteira necessita ser de um material resistente, para suportar o balde repleto de composto orgânico; por esse motivo, a estrutura do protótipo foi construída de ferro, que foi movimentada por um motor de passo, de 8 a 32V, alimentado por corrente contínua (como pode ser visualizado na Figura 4b).

Por fim, com relação à literatura disponível, faz-se necessário registrar que existem trabalhos semelhantes a este, como o de [Silva 2019], que se realizou um monitoramento de temperatura, umidade e oxigênio; porém, o trabalho não se aplicou um microcontrolador. Ademais, para o armazenamento do composto, o autor utilizou leiras estáticas (método de compostagem feito no solo), enquanto, no projeto aqui proposto, o monitoramento foi realizado em um balde, possibilitando, assim, a implantação desse protótipo em pequenas residências.

6. Conclusão

Por meio do desenvolvimento deste trabalho, pode-se concluir que a implementação do protótipo ocorreu de maneira satisfatória, uma vez que, foi possível realizar o controle automatizado de um composto orgânico por meio dos dados mensurados pelos sensores de umidade e de temperatura. Face ao exposto, salienta-se a importância da utilização de um protótipo como este para a produção de matéria orgânica, como viabilidade para a aplicação nos cultiváveis.

Na elaboração do trabalho surgiram algumas dificuldades, sendo, a primeira, a montagem da estrutura metálica da composteira, em função de sua capacidade de armazenamento que era de 20L. A outra consistiu no posicionamento correto dos sensores, visando a melhor veracidade dos dados coletados.

Além disso, tem-se a perspectiva da realização de trabalhos futuros, a partir deste protótipo, onde será desenvolvido um aplicativo e um *website*, possibilitando uma interação com o usuário final. Com isso, existe a necessidade de um servidor próprio para armazenar os dados, para que os resultados do monitoramento da composteira sejam mostrados ao usuário por meio de gráficos e relatórios.

Referências

Alves, A. C. O., Santos, A. L. S. e Azevedo, R. M. C. (2012) “Agricultura orgânica no Brasil: sua trajetória para a certificação compulsória”. *Revista Brasileira de Agroecologia.* [S. I.], vol. 7, n. 2, pp 19-27, abr. 2012.

Barbosa, W. F. e Sousa, E. P. (2012) “Agricultura orgânica no Brasil: características e desafios”. *Revista Economia & Tecnologia*, [S.I.], v. 8, n. 4, p. 67-74, 2012.

Bin, A. (2004) Agricultura e meio ambiente: contexto e iniciativas da pesquisa pública - Campinas, SP, 2004.

Lobe, L. M. K. (2019) Benefícios de um minhocário ou composteira - Meu santo composto. Disponível em: http://www.sustentaacoes.com/2019/02/beneficios-de-um-minhocario-ou-composteira-meu-santo-composto-1.html. Acesso em: 17 out 2020.

Maia, G. (2012) Acionamento Remoto de Portões Elétricos Via Celular Através de Microcontrolador. Orientador: Profa. M. C. Maria Marony Sousa Farias. 1 f. Trabalho de Conclusão de Curso (Engenharia de Computação) - Centro Universitário de Brasília, Brasília.

Oliveira, E. C. A., Sartori, R. H. e Garcez, T. B. Compostagem. (2008) 19 p. Universidade de São Paulo – Escola Superior de Agricultura Luiz Quieroz, Piracicaba, 2008. Disponível em: https://www.agencia.cnptia.embrapa.br/Repositorio/Compostagem_000fhc8nfqz02wyiv80efhb2adn37yaw.pdf. Acesso em: 13 out. 2020.

Oliveira Júnior, M. e Duarte, R. O. (2010) Introdução ao Projeto com Microcontroladores e Programação de Periféricos. Universidade de Minas Gerais. (Apostila de Introdução ao Projeto com Microcontroladores e Programação de Periféricos).

Ormond, J.G.P., Paula, S. R. L., Faveret Filho, P. S. C. e Rocha, L. T. M. (2014) Agricultura Orgânica: Quando o Passado é Futuro. Biblioteca Digital do BNDES, BNDES - Biblioteca Digital, 17 out. 2014. Disponível em: https://web.bndes.gov.br/bib/jspui/handle/1408/2479. Acesso em: 10 nov. 2020.

Pereira, C. L. O. (2020) Programação de Microcontroladores para a Comunicação com Aplicações Móveis Usando Bluetooth. 44 f. Trabalho de Conclusão de Curso (Engenharia da Computação) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Pires, I. C. G e Ferrão, G. E. (2017) "Compostagem no Brasil sob a perspectiva da legislação ambiental". *Revista Trópica: Ciências Agrárias e Biológica*, [S. I.], v. 9, n. 1, pp. 1-18.

São Paulo (Estado). (2011) Secretaria do Meio Ambiente / Coordenadoria de Biodiversidade e Recursos Naturais. Agricultura sustentável. Kamiyama, Araci. São Paulo: SMA. 75p., 21 x 29,7cm (Cadernos de Educação Ambiental, 13).

Silva, J. R. (2019) Automação do Processo de Compostagem: uso de sensores para monitoramento e controle de parâmetros de um processo sustentável. 2019. 75 p. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Goiás, Catalão.

Silva, M. C., Targa, (2019) M. S. e Cezar, V. R. S. Uso de Microcontrolador Arduino para a Determinação da Permeabilidade do Solo [S. l.], p. 1-28.

Souza, J. (2005) Agricultura Orgânica: Tecnologias para a Produção de Alimentos Saudáveis. -Instituto Capixaba de Pesquisa. v. 2, 257 p.

Thomazini, D. e Albuquerque, P. U. B. (2011) Introdução. In: Sensores Industriais: Fundamentos e Aplicações. 4. ed. rev. São Paulo: Érica.

Valente, B. S. *et al* .(2009) "Fatores Que Afetam O Desenvolvimento Da Compostagem De Resíduos Orgânicos". *Archivos de Zootecnia*, [S. I.], vol. 58, pp. 59-85, maio.

Wendling, M. (2010) Sensores. Apostila da Universidade Estadual Paulista. 19 p. Disponível em: https://www.feg.unesp.br/Home/PaginasPessoais/ProfMarceloWendling/4---sensores-v2.0.pdf. Acesso em: 25 maio 2021.



Desempenho da classificação supervisionada com a inclusão da banda termal em áreas selecionadas do bioma Pampa – Brasil

Supervised classification performance with the inclusion of the thermal band in selected areas of the Pampa biome - Brazil

Patricia Michele Pereira Trindade[1], Daniela Wancura Barbieri Peixoto[1], Tatiana Mora Kuplich[1], Claudio Aparecido de Almeida[2]

[1]Instituto Nacional de Pesquisas Espaciais (INPE) - Coordenação Espacial do Sul (COESU), Santa Maria – RS

[2] Divisão de Projetos Estratégicos 1 (DIPE1), São José dos Campos - SP, Brasil

{tatiana.kuplich@inpe.br; daniwbarbieri@gmail.com; patriciatrindade.sr@gmail.com; claudio.almeida@inpe.br}

Abstract. *The Pampa biome is the focus of profound changes in recent decades and grasslands have been converted to rice, soybean and forestry crops. In this article, it was investigated whether the addition of the thermal band in Maxver supervised classification with Landsat images increased the accuracy of vegetation cover mappings in the Pampa Biome. The results pointed out to a moderate increase in accuracy of the classifications with the use of the thermal band. Visual interpretation of images is also facilitated with temperature data. For both approaches (digital and manual), it is important to consider seasonal images, which include vegetation in different stages and vigor conditions.*

Resumo. *O bioma Pampa é foco de profundas alterações nas últimas décadas e os campos nativos têm sido convertidos para culturas de arroz, soja e silvicultura. Neste artigo foi investigada se a adição de banda do termal em classificação supervisionada Maxver com imagens Landsat aumentaria a acurácia dos mapeamentos de cobertura vegetal no Bioma Pampa. Os resultados apontam para moderado aumento na acurácia das classificações com a utilização da banda do termal. A interpretação visual das imagens também é facilitada com dados de temperatura. Para ambas abordagens*

(digital e manual), é importante imagens temporais de épocas distintas, que contemplem coberturas vegetais em diferentes estágios e condições de vigor.

1. Introdução

O bioma Pampa ocupa 63% do território do Rio Grande do Sul e é uma das áreas de campos temperados mais importante do mundo (MMA, 2019). O Pampa é um dos biomas brasileiros que teve sua cobertura vegetal, primariamente campestre, bastante alterada (Embrapa, 2016), e parte destas alterações são devidas à expansão das áreas de cultivo agrícola e silvicultural (Nabinger et al. 2009). Na metade sul do RS a principal ocupação do bioma sempre foi a pecuária extensiva, porém a partir do século XX ocorreu a implantação da agricultura intensiva, destacando a produção de arroz irrigado nas áreas de várzeas. No início do século XXI começou a elevada expansão da silvicultura e cultivo de soja (MMA, 2019).

Neste cenário de conversões de cobertura da terra, dados de sensoriamento remoto se revelam essenciais para a detecção e mapeamento dos campos e de seu estado de conservação, como no Programa de Monitoramento Ambiental dos Biomas Brasileiros, do Ministério do Meio Ambiente, executado por equipes do INPE e FUNCATE (Fundação de Ciência, Aplicações e Tecnologia Espaciais), cujos dados estão disponíveis na Plataforma Terrabrasilis - http://terrabrasilis.dpi.inpe.br/ (Assis et al., 2019, Almeida, 2020).

A utilização de bandas na faixa do termal em associação às bandas ópticas tem sido apontada como favorável à discriminação de coberturas da terra (Rodriguez-Galiano et al. 2012, Sun e Schulz 2015, Eisavi et al. 2015, Zhao et al. 2019). O aumento da acurácia de classificações automáticas, mesmo supervisionadas (com treinamento), em mapeamentos de coberturas vegetais, poderá contribuir para o monitoramento da conservação e manejo dos biomas brasileiros.

Este artigo teve como objetivo avaliar a adição da banda termal (banda 10) do sensor TIRS (Thermal Infrared Sensor) às bandas ópticas OLI (Operational Land Imager) do satélite Landsat 8 na classificação de campos e outras coberturas vegetais do bioma Pampa, com a abordagem de Máxima Verossimilhança (Maxver).

2. Metodologia

Para a realização das etapas metodológicas, selecionou-se nove células de 25 x 25 km no bioma Pampa (Figura 1). As células foram distribuídas nas regiões leste, centro e oeste nas órbitas ponto 221/81, 223/81 e 224/81 respectivamente.

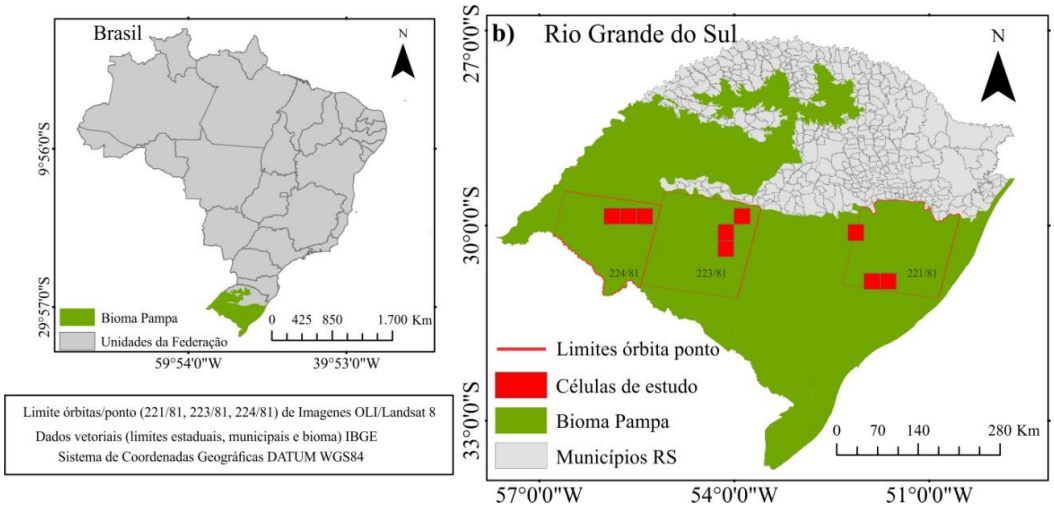


Figura 1. – Localização e limites do Bioma Pampa, das células classificadas e das órbitas ponto das imagens Landsat utilizadas no estudo.

2.1 Materiais

Foram utilizadas imagens dos sensores OLI e TIRS do satélite Landsat 8, correspondentes às datas 15/02/2018, 26/08/2018 e 16/12/2018 (órbita ponto 223/81); 15/02/2018, 26/08/2018 e 16/12/2018 (órbita ponto 224/81); e 13/01/2017, 25/08/2017 e 15/12/2017 (órbita ponto 221/81). As imagens foram adquiridas no Serviço Geológico dos Estados Unidos (USGS), e foram consideradas as bandas do espectro óptico (B2, B3, B4, B5, B6 e B7) em reflectância de superfície (SR), resolução espacial de 30 metros e a banda termal em temperatura de brilho (Tb), com a resolução espacial reamostrada para 30 metros.

2.2 Métodos

Os procedimentos metodológicos seguiram três etapas: Etapa 1 – pré-processamento das imagens; Etapa 2 – coleta de amostras de treinamento; Etapa 3 – classificação supervisionada e análise de acurácia.

2.2.1 Etapa 1

As imagens termais em Temperatura de Brilho (Tb) passaram pelo processo de correção de emissividade (Equação 1) com auxílio do *software* QGIS e metodologia implementada por Ndossi e Avdan (2016). As imagens de emissividade consideram os valores de NDVI (Tabela 1), conforme metodologia utilizada em diferentes pesquisas (Van De Griend & Owe, 1993; Valor & Caselles, 1996; Sobrino, Jiménez-Muñoz, & Paolini, 2004; Zhang, Wang, & Li, 2006).

$$TS=Tb/(1+(\lambda Tb/\alpha)*\ln\varepsilon) \qquad \text{Equação 1}$$

TS – Temperatura de Superfície, Tb = temperatura de brilho registrada pelo sensor em Kelvin; λ= comprimento de onda médio do infravermelho termal; α = hc/K = 1.438 x 10-2 m K (K= constante de Stefan-Boltzman (1.38 x 10-23J/K); h = constante de Planck (6.28 x 10-34J/s); c = velocidade da luz (2.998 x 108m/s)).

Tabela 1. Relação NDVI e Emissividade

Valores NDVI	Valores Emissividade
NDVI < -0.185	0.995
-0.185 ≤ NDVI < 0.157	0.985
0.157 ≤ NDVI ≤ 0.727	1.009 + 0.04 x ln(NDVI)
NDVI > 0.727	0.990

Fonte: adaptado Ndossi e Avdan (2016).

2.2.2 Etapa 2

A coleta de amostras de treinamento abordou as classes área de cultivo (AC), campo (CA) e vegetação arbórea (VA). Todas as coberturas que não se enquadravam nessas classes foram atribuídos à classe "outros" (não computada nas análises). Uma máscara de corpos hídricos disponível na Plataforma TerraBrasilis foi aplicada, estas áreas não foram consideradas na classificação. No total, foram coletados 3000 pixels de treinamento (1000 pixels para cada classe).

2.2.3 Etapa 3

Neste estudo foram realizadas classificações supervisionadas com o algoritmo paramétrico Máxima Verossimilhança – Maxver, um dos algoritmos mais utilizados em pesquisas de sensoriamento remote (Silveira et al., 2020). Para fins de comparação, duas classificações foram realizadas: classificação 1 – somente as bandas SR e classificação 2 – combinação das bandas SR e bandas Termal.

A análise da acurácia das classificações foi realizada a partir de 450 pontos distribuídos aleatoriamente nas células de estudo, sendo cada ponto interpretado por diferentes intérpretes para garantir maior confiabilidade nos pontos de referência. A interpretação contou com composições coloridas das imagens Landsat 8 para diferentes estações do ano e imagens de alta resolução disponíveis no *Google Earth* com datas próximas das imagens utilizadas para as classificações. Os pontos também foram conferidos com os dados de desmatamento 2016/2018 e disponibilizados na plataforma TerraBrasilis do INPE (Assis et al., 2019, Almeida, 2020).

A partir das imagens classificadas e os pontos de referência, foram calculadas as matrizes de erros para cada classificação, e assim calculadas a Exatidão Global (EG), Exatidão do Produtor (EP) e Exatidão do Usuário (EU). Desta forma, foi possível verificar o desempenho do classificador e a contribuição da banda termal nas classificações.

3.Resultados e Discussão

A figura 2 representa as características das classes selecionadas no bioma Pampa. Na imagem termal de dezembro verificou-se que as classes estudadas apresentaram diferenças de TS (Figura 2 a), as áreas de cultivo com os valores de TS mais elevados e a vegetação arbórea com a menor média de TS.

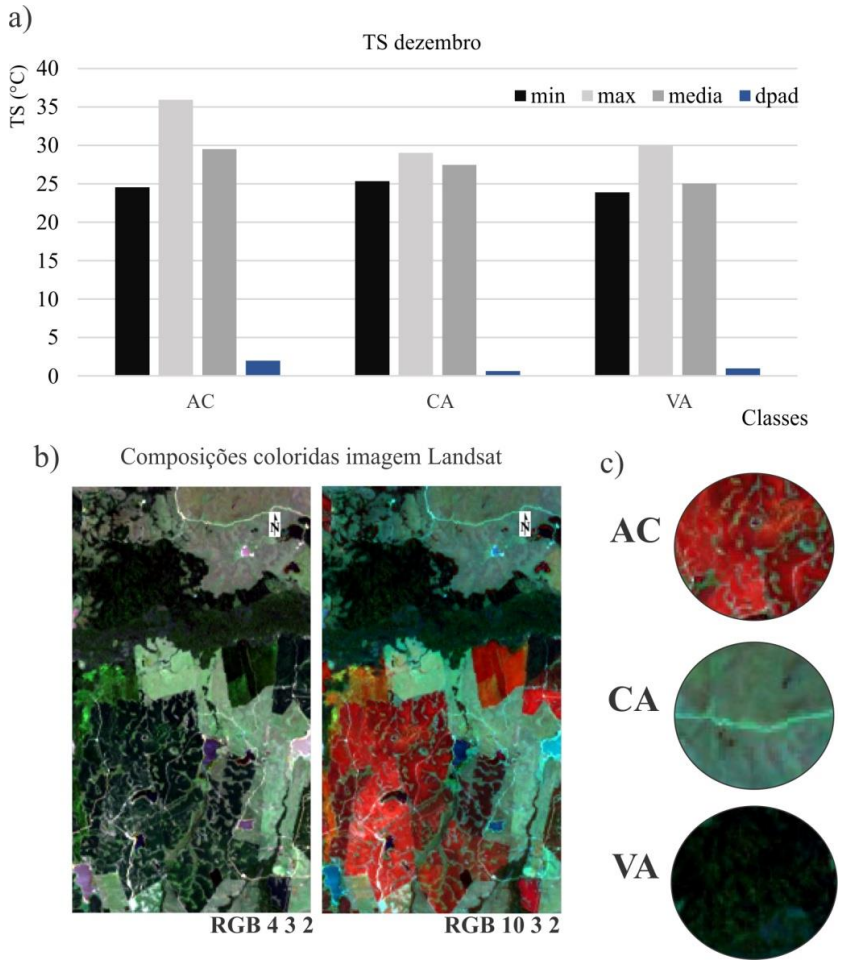


Figura 2 Características das classes área de cultivo (AC), campo (CA) e vegetação arbórea (VA): a) estatísticas da TS para cada classe na imagem 16/12/2018; b) composições coloridas identificando cada classe; c) destaque para as classes AC, CA e VA.

A adição da banda termal na composição colorida falsa cor RGB 10 3 2 (Figura 2 b) possibilitou maior diferenciação entre as classes (Figura 2 c), e maior destaque para a classe AC. Na composição colorida cor verdadeira RGB 432 da imagem de fevereiro, as classes área de cultivo e vegetação arbórea não foram facilmente diferenciadas, o que destaca a contribuição da banda termal e de diferente estação do ano para análise qualitativa das imagens. O padrão de cobertura das áreas agrícolas muda frequentemente (Blaschke, et al., 2000), portanto, em época de alto vigor vegetativo (fevereiro) a temperatura da superfície é inferior à temperatura da época de solo exposto (dezembro), contribuindo para o maior destaque quando utilizada a banda termal na composição de bandas. Eisavi et al. (2015) destacam que a utilização de imagens multitemporais e termais tem potencial de melhorar a eficiência da classificação da cobertura da terra.

Um exemplo dos mapas produzidos neste estudo é apresentado na Figura 3, sendo possível verificar as diferenças das áreas classificadas. Na classificação 1 foram utilizadas

somente as bandas em reflectância de superfície (SR), e a classificação 2 abordou as bandas SR e TS.

O desempenho das classificações foi avaliado conforme o cálculo da Exatidão Global (EG), Exatidão do Produtor (EP) e Exatidão do Usuário (ES). A EG das classificações teve variação de 80% (classificação 1) a 83% (classificação 2), ou um acréscimo de 3% na acurácia da classificação com a inclusão da banda termal.

Ao utilizar imagens Landsat 8 de diferentes estações do ano, Zhao et al. 2019 também avaliaram a classificação de imagens com a inclusão da banda termal. Os autores concluíram que a adição das imagens termais aumentou a acurácia dos mapas de cobertura da terra, apresentando um incremento de 1% nos mapas classificados.

Ao analisar os padrões de cobertura do solo, Sun e Schulz (2015) também destacaram a eficácia das bandas termais. Os autores constataram que ao combinar a banda termal (Landsat 8) às bandas do visível e infarvermelho próximo, ocorreu um aumento da acurácia da classificação de 3% a 6%.

O estudo realizado por Eisavi et al. (2015) apontou as seguintes exatidões globais para diferentes cenários: ao considerar apenas bandas ópticas EG de 86,48%; apenas banda termal EG de 82,26%; e a combinação de bandas ópticas com termal apresentou uma EG de 90,63%. Assim, pode-se confirmar que a banda termal sozinha não tem a eficácia de aumentar a EG da classificação, mas a combinação com as demais bandas do espectro pode contribuir para o aumento da acurácia nas classificações de cobertura da terra.

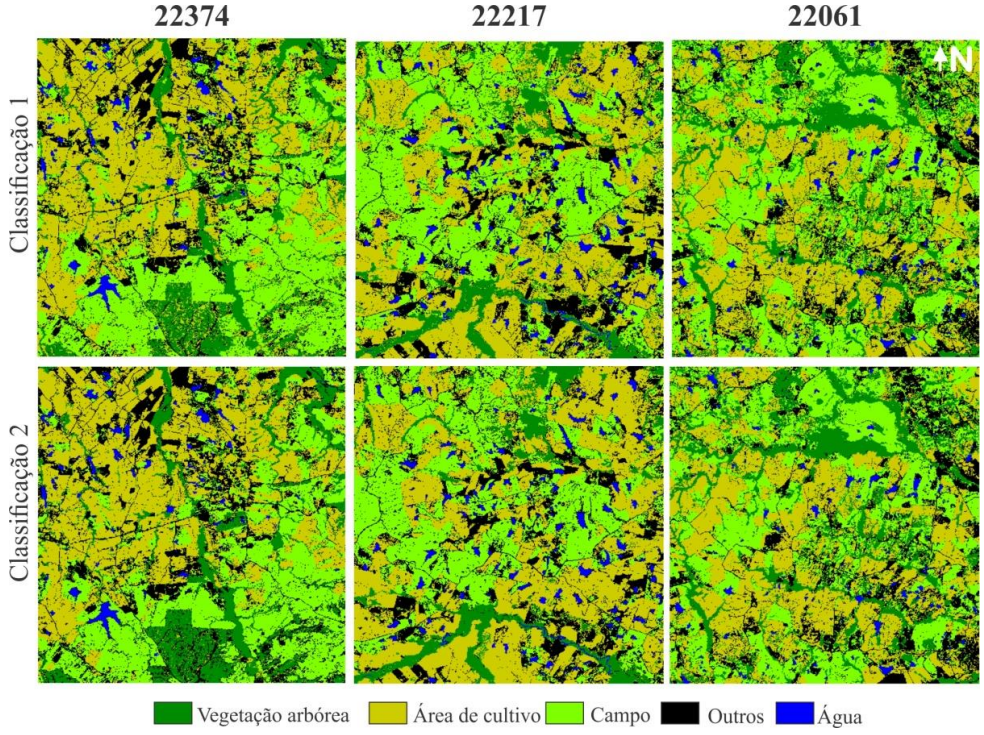


Figura 3. Exemplo de células classificadas com banda termal (Classificação 2) e sem a banda termal(Classificação 1).

A tabela 2 mostra o desempenho da classificação para cada classe a partir da EP e EU. Foi possível verificar que a média de EP variou de 79,67% (classificação 1) a 83,43% (classificação 2) e a EU de 79,32 % a 82,43%. A inclusão da banda termal na classificação garantiu um aumento da EP e EU para as três classes analisadas. Ao analisar diferentes cenários de classificação com a banda 10 Landsat 8, Eisavi et al. (2015) verificaram uma variação de EP média de 77,48% para classificação sem a banda termal e 80,12% para classificação com a inclusão da banda termal. Os valores de EU encontrados foram de 77,14% (sem termal) e 78,61 (com termal).

Tabela 2 – Exatidões do produtor e usuário para as classes área de cultivo (AC), campo (CA) e vegetação arbórea (VA) nas classificações

	Exatidão do produtor (%)				Exatidão do usuário (%)			
	AC	**CA**	**VA**	**Média**	**AC**	**CA**	**VA**	**Média**
Classificação 1	77,83	84,07	77,10	79,67	84,46	76,41	77,10	79,32
Classificação 2	79,68	88,50	81,54	83,24	89,26	78,33	79,70	82,43

*Classificação 1 –somente bandas SR; Classificação 2 – bandas SR e Termal.

A classe campo apresentou a maior EP (88,5%), ou seja, 88,5% das áreas de campo foram corretamente classificadas com a inclusão da banda termal, enquanto a classificação 1 (somente bandas SR) apresentou um EP de 84,07% para a classe supracitada. A classe VA apresentou um aumento de 4,44% na EP e 2,6% na EU. As áreas de cultivo apresentaram um incremento de 1,85% na EP e 4,8% na EU. Resultados semelhantes foram encontrados por Zhao et al. (2019), com incremento na EP 1% e EU de até 2% para áreas de cultivo quando considerada a banda termal na classificação.

6. Considerações finais

A realização do trabalho permitiu destacar as seguintes afirmações:

- A inclusão da banda termal aumentou a acurácia dos mapas classificados, com incremento de 3%, valor dentro dos resultados encontrados em diferentes pesquisas internacionais.

- A adição da banda termal em composições coloridas falsa-cor possibilitou a melhor diferenciação entre as classes. Então, a banda termal também contribui de forma qualitativa para a interpretação e classificação de imagens.

- A combinação de bandas de diferentes estações do ano possibilita a melhoria da classificação, principalmente, para a interpretação das imagens.

- Estão sendo realizados outros processamentos com diferentes classificadores para refinar as classificações da cobertura da terra para o bioma Pampa, considerando imagens de diferentes datas em séries temporais mais extensas.

Agradecimentos

Este artigo é um dos resultados do Projeto CNPq 444418/2018-0 (Monitoramento dos biomas brasileiros por satélite - Construção de novas capacidades), que disponibilizou bolsas para as duas primeiras autoras.

Referências

Almeida, C.A.; Valeriano, D.M; Maurano, L.; Vinhas, L.; Fonseca, L.M.G.; Silva, D.; Santos, C.P.F; Martins, F. S. R. V.; Lara, F.C.B.; Maia, J.S.; Profeta, E.R.; Santos, L.O.; Santos, F.C.O; Ribeiro, V. 'Deforestation monitoring in different Brazilian Biomes: challenges and lessons.'In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-3/W2-2020, 2020, Santiago, Chile.

Assis, L. F. F. G, Ferreira, K. R., Vinhas, L., Maurano, L., Almeida, C., Carvalho, A., Camargo, C. 2019. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. ISPRS International Journal of Geo-Information, 8(11), 513. doi: 10.3390/ijgi8110513

Blaschke, T.; Lang, S.; Lorup, E.; Strobl, J.; Zeil,P.; "Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications". In: Cremers, A. B. & Greve, K. (Hrsg). (2000)

Eisavi, V., Homayouni, S., Yazdi, A. M., & Alimohammadi, A. (2015). Land cover mapping based on random forest classification of multitemporal spectral and thermal images. Environmental Monitoring and Assessment, 187(5), 1–14. https://doi.org/10.1007/s10661-015-4489-3

Embrapa. Nosso Pampa desconhecido. Porto Alegre: Fundação Zoobotânica do Rio Grande do Sul, 2016. In: Chomenko, l.; Bencke, G. A. (Org.) Capítulo em livro técnico (INFOTECA-E). 2016.

MMA (Ministério do Meio Ambiente). "*Pampa*". Folder Pampa. (2019) Accessed 15 Abril 2020. https://www.mma.gov.br/biomas/pampa

Nabinger, C.; Ferreira, E.T.; Freitas, A.K.; Carvalho, P.C.F.; Sant'anna, D.M. Produção animal em campo nativo: aplicações de resultados de pesquisa. In: Pillar, V.P.; Müller, S.C.; Castilhos, Z.M.S.; Jacques, A.V.A. (Org.). Campos sulinos: conservação e uso sustentável da biodiversidade. Brasília: Ministério do Meio Ambiente - MMA, p. 175-198. 2009

Ndossi, M. I., And Avdan, U. 2016. Application of open source coding technologies in the production of Land Surface Temperature (LST) maps from Landsat: A PyQGIS plugin. Remote Sensing, 8 (5), 1 - 31. doi: 10.3390/rs8050413

Rodríguez-Galiano, V. F., Ghimire, B., Pardo-Igúzquiza, E., Chica-Olmo, M., & Congalton, R. G. "Incorporating the downscaled Landsat TM thermal band in land-cover classification using random forest". *Photogrammetric Engineering and Remote Sensing*, **78**(2), 129–137 (2012). doi: 10.14358/PERS.78.2.129

SILVEIRA, A.H de M.; SILVA, F.M,; HADAD, R. M.; LIBÓRIO, M.P. Alicações, preferências e comparações entre métodos de classificação supervisionada: o caso de Natal/RN. Ra'e Ga. Curitiba, v.47, n.1, p. 120-135.

Sobrino, J. A., Jiménez-Muñoz, J. C., And Paolini, L. 2004. Land surface temperature retrieval from LANDSAT TM 5. Remote Sensing of Environment, 90(4), 434–440. doi: 10.1016/j.rse.2004.02.003

Sun, L.; Schulz, K. The Improvement of Land Cover Classification by Thermal Remote Sensing. Remote Sensing, 7, 8368 – 8390. doi:10.3390/rs70708368

Valor, E., And Caselles, V. 1996. Mapping land surface emissivity from NDVI: Application to European, African, and South American areas. Remote Sensing of Environment, 57(3), 167–184. doi: 10.1016/0034-4257(96)00039-9

Van De Griend, A. A., and Owen, M. 1993. On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces. International Journal of Remote Sensing, 14(6), 1119–1131. doi: 10.1080/01431169308904400

Zhang, J., Wang, Y., And Li, Y. 2006. A C++ program for retrieving land surface temperature from the data of Landsat TM/ETM+ band6. Computers and Geosciences, 32(10), 1796–1805. doi: 10.1016/j.cageo.2006.05.001

Zhao, J., Yu, L., Xu, Y., Ren, H., Huang, X., & Gong, P. (2019). Exploring the addition of Landsat 8 thermal band in land-cover mapping. *International Journal of Remote Sensing*, *40*(12), 4544–4559. https://doi.org/10.1080/01431161.2019.1569281.



Desenvolvimentos em inteligência artificial na avicultura de frangos de corte

Lucas Gabriel Galdino da Silva[1,2], Francisco Gomes Cordeiro Neto[1], Josenalde Oliveira[1,2], Laura Santana[1,2]

[1]Escola Agrícola de Jundiaí – Universidade Federal do Rio Grande do Norte (UFRN)
Macaíba – RN – Brasil

[2]Instituto de Engenharia de Sistemas, Computadores, Pesquisa e Desenvolvimento do Brasil (INESC P&D Brasil)
Santos - SP - Brasil

{lucas.gabriel.silva.706, francisco.neto.009}@ufrn.edu.br,
{josenalde.oliveira, laura.santana}@ufrn.br

Abstract. *The Brazilian poultry industry has been gaining prominence as a major producer and occupies the top position in the world ranking. In this review, based on an exploratory survey, there are presented research outcomes with Artificial Intelligence (AI) throughout the production chain in broiler poultry farming, from behavior analysis for disease identification and slaughter weight prediction, to quality control. With a period beginning with the international release of Industry 4.0, there are described 26 papers which demonstrate the growing adoption of AI and its divisions.*

Resumo. *A indústria avícola brasileira vem ganhando destaque como grande produtor e ocupa o topo de colocações no ranking mundial. Na presente revisão, a partir de um levantamento exploratório, são apresentados resultados de pesquisas que utilizem Inteligência Artificial (IA) ao longo da cadeia de produção de frangos de corte, desde a análise de comportamentos para identificação de doenças e predição de peso de abate, ao controle de qualidade. Fazendo o período de estudo coincidir com o início internacional da Indústria 4.0, são relatados 26 trabalhos que apontam para a adoção crescente da IA e suas divisões.*

1. Introdução

A cadeia produtiva de frangos de corte é uma das cadeias mais importantes no agronegócio brasileiro (BASSI; SILVA, 2017), o que é corroborado ao analisar-se os dados de produção e exportação desse setor. De acordo com relatório 2021 da Associação Brasileira de Proteína Animal (ASSOCIAÇÃO BRASILEIRA DE PROTEÍNA ANIMAL, 2021), em 2020 o Brasil produziu 13,845 milhões de toneladas de carne de frango e exportou 4,231 milhões de toneladas, o que torna o país o maior exportador mundial do produto, sendo que dessa exportação, 66,99% é do produto tipo cortes, com frango inteiro, embutidos, salgados e industrializados complementando este percentual. No estudo feito por Bassi e Silva (2018), os autores afirmam que os grandes avanços da pesquisa no desenvolvimento de novas tecnologias no setor da avicultura ajudaram a posicionar o Brasil em termos de produção e exportação mundial. Nesse contexto, é interessante observar que a agroindústria, assim como a indústria de modo geral, vem passando por um processo de revolução tecnológica conhecido como Indústria 4.0, onde as chamadas tecnologias habilitadoras 4.0, dentre elas a Inteligência Artificial (IA), permitem não apenas otimizar e monitorar processos, mas também auxiliar na tomada de decisão, estando usualmente integrada às camadas de modelagem supervisionada e não supervisionada em algoritmos de ciência de dados.

No caso do agronegócio, tem sido comum adotar genericamente o termo Agricultura 4.0 e nesse caso, tais tecnologias também fomentam a agricultura de precisão, na qual busca-se adotar "estratégias de manejo mais eficientes, considerando a variabilidade espacial e temporal da unidade produtiva, permitindo o uso adequado dos insumos a fim de evitar desperdícios" (SANTANA; OLIVEIRA, 2021). É comum a adoção dos termos inteligente ou *smart* para diferenciar da prática convencional. Na produção animal, a zootecnia de precisão tem se ocupado de questões "para supervisão e análise em tempo real da saúde, nutrição, comportamento e ambiência, visando os índices produtivos, reprodutivos e bem-estar animal de forma automática, contínua, não invasiva e sem gerar estresse ao animal" (ROCHA; SILVA, 2020). O tema tem, inclusive, motivado a ampliação de chamadas públicas, como o Programa Agro 4.0 da Agência Brasileira de Desenvolvimento Industrial (ABDI), cujo 1º edital selecionou e premiou 14 projetos pilotos de adoção e de difusão de tecnologias 4.0 no agronegócio (BRASIL, 2020a). Na mesma linha seguiu a Financiadora de Estudos e Projetos (FINEP), com chamada de subvenção econômica às empresas em tecnologias 4.0, incluindo o setor agropecuário (BRASIL, 2020b).

Devido ao volume, variedade e velocidade de dados disponíveis para geração de conhecimento, a IA tem sido muito explorada na atualidade em soluções para cidades, saúde e indústria. Não seria diferente para o setor agropecuário, onde a IA encontra adequação natural à complexidade de mapeamento das relações biológicas e de ambiência. Diante do exposto, este artigo tem como objetivo fazer um levantamento exploratório inicial sobre como a IA tem sido adotada na avicultura de frango de corte no período que configura o início do conceito Indústria 4.0, destacando as principais aplicações e identificação de demandas para o avanço da área a partir desta tecnologia habilitadora.

2. Material e Métodos

Com base na metodologia PRISMA para organização de revisões bibliográficas (MOTHER et al., 2009), para a etapa de identificação, foi utilizada a busca geral por assunto no Portal de Periódicos CAPES, com acesso federado via Cafe/RNP, com a string ("machine learning" OR "artificial intelligence") AND (broiler OR poultry OR chicken OR aviculture). Dado que a busca contempla por padrão os campos Título, Resumo, Abstract, Palavras-Chave, o uso da string com termos em língua inglesa também contempla as principais bases indexadoras (Springer, IEEE, Science Direct) e periódicos nacionais. De 4.176 trabalhos, são apresentados o total de 26 estudos, após aplicação dos critérios de inclusão: período de 2016-2021, avicultura de frango de corte, trabalhos completos indexados e revisados por pares. Adicionalmente, os critérios de exclusão foram aplicados: avicultura de postura, ênfase em Internet das Coisas (*Internet of Things*, IoT), não frango, e com conteúdo não diretamente relacionado, além da remoção de duplicados. A busca com termos em português: ("inteligência artificial" OR "aprendizagem de máquina") AND (frango OR avicultura) não retornou resultados.

Embora trabalhos com IoT apresentem intersecções e possam contemplar etapas adicionais de análise e tratamento dos dados, revisões específicas para o aspecto de instrumentação podem ser encontradas em Oliveira et al. (2019) e Astill et al. (2020). O período dos últimos cinco anos foi selecionado de modo a coincidir com a promoção do termo Indústria 4.0 a partir do trabalho de Schwab (2016) e sua relação com IA.

3. Resultados e Discussão

Dos 26 estudos listados, usando o critério de afiliação do primeiro autor e instituição de condução da pesquisa, o Brasil encontra-se em 2. lugar, com 3 trabalhos, atrás apenas da China, com 6 trabalhos. Os demais artigos distribuem-se pela Ásia, América do Norte e Europa. A representação brasileira inclui os estados de São Paulo e Paraná (RIBEIRO et al., 2019; BARBON et al., 2018; PEREZ et al., 2018), com instituições como Unicamp, USP, UEL, UFPR, UTFPR e PUC-PR.

Da gestão e manejo das condições de ambiência em aviários para aprimoramentos no bem-estar animal (RIBEIRO et al., 2019), seleção genética de matrizes, produtividade e detecção antecipada de enfermidades, ao processamento agroindustrial em suas etapas de controle de qualidade, o melhor conhecimento do processo por meio da análise automatizada dos dados de produção baseada em IA, maximiza a assertividade da tomada de decisão e o cumprimento de exigências sanitárias para os mercados mais exigentes. Estudos críticos como em Rowe et al. (2019) discutem, contudo, que o enfoque ainda permanece em ganhos de produtividade, não necessariamente considerando o aspecto de bem-estar animal. No caso de frangos de corte, com ambientes de alta densidade de população e ciclo curto (em média 45 dias), este fato demanda maior atenção pelos agentes envolvidos e tem sido alvo de investigação com base em análise de imagens (OKINDA et al, 2020). É digno de nota que alguns autores chamam à atenção para a consolidação de estratégias existentes antes da adoção de tecnologias disruptivas associadas ao 4.0, caracterizando o conceito intermediário de Indústria 3.5 (WANG et al, 2021).

Neste sentido, de modo a contemplar a diversidade de aplicações, os trabalhos selecionados foram classificados em duas categorias (i) diagnóstico de doenças/bioinformática veterinária (ii) monitoramento de produção e processamento.

Este trabalho portanto amplia e atualiza o estudo de Milosevic et al. (2019). Dada a atualidade do tema de bioinformática, este tópico foi adicionado à categoria (i) sob o enfoque veterinário. O termo 'processamento' na categoria (ii) busca contemplar o conceito *farm-to-fork*, incluindo o controle de qualidade dos produtos.

A Tabela 1 abaixo resume as áreas de aplicação e tarefas baseadas em técnicas de IA, descritos nas subseções a seguir.

Tabela 1: quadro resumo de aplicações, tarefas e trabalhos relacionados

Autores	Aplicação	Tarefa
A. Banakar et al., 2016	Diagóstico de doenças	Mineração de dados
A. Mortensen et al., 2016	Aprimoramento de pesagem	Regressão
M. Kamruzzaman et al., 2016	Monitoramento de produção	Regressão
S. Amraei et al., 2018	Predição de peso	Regressão
S. Barbon et al., 2018	Monitoramento de produção	Classificação
Y. Fajar et al., 2018	Monitoramento de produção	Mineração de dados
N. Perez et al., 2018	Monitoramento de produção	Mineração de dados
X. Zhuang et al., 2018	Alerta/Diagnóstico de doenças	Classificação
E. Rowe et al., 2019	Ganhos de produtividade no bem-estar animal	Mineração de dados
B. Milosevic et al., 2019	Predição de crescimento e saúde	Classificação
R. Ribeiro et al., 2019	Aprimoramentos da gestão e manejo de ambiência	Mineração de dados
A. Polewko et al., 2020	Marcadores genéticos	Regressão
C. Okinda et al., 2020	Monitoramento de produção	Classificação
G. Kollenburg et al., 2020	Monitoramento de produção	Regressão
T.L. Passarafo et al., 2020	Combinação de marcadores genéticos e rastreabilidade	Regressão
J. Astill et al., 2020	Diagnóstico de doenças	Regressão
K. Cuan et al., 2020	Diagnóstico de doenças	Sons / Regressão
R. Yusof et al., 2020	Monitoramento de produção	Mineração de dados
I. Nyalala et al., 2021	Predição de peso e volume de aves	Mineração de dados
D. Seo et al., 2021	Marcadores genéticos e rastreabilidade	Regressão
C. Wang et al., 2021	Módulos para previsões e monitoramento de peso	Mineração de dados
S. Yousefinaghani et al., 2021	Alerta/Diagnóstico de doenças	Classificação
D. Dourou et al., 2021	Monitoramento de produção	Classificação
J. You et al., 2021	Módulos de monitoramento de peso	Classificação
C. Fang et al., 2021	Diagnóstico de doenças	Classificação

3.1. Diagnóstico de doenças/bioinformática veterinária

A identificação precoce de condições de gatilho para doenças comuns em aviários a partir de variáveis de ambiência, bem como o monitoramento de comportamento são problemas de pesquisa que fomentam a relação academia-indústria, sendo tais estudos frutos de parcerias ou realizados em instalações próprias de institutos de pesquisa e universidades, reproduzindo as condições reais do setor produtivo em escala adequada. Banakar et al. (2016) propõem um modelo de estudo para o diagnóstico inteligente de doenças aviárias com base em sinais sonoros dos frangos, utilizando um dispositivo para diagnosticar doenças aviárias, como a Influenza Aviária, a Doença de Newcastle (ND) e Vírus da Bronquite (BV), usando métodos de mineração de dados e teoria da evidência Dempster-Shafer (D-S). Além da vocalização, Astill et al. (2018) revisam estudos de modelos preditivos para detecção de gripe aviária que consideram uma variedade de fontes, com análises de imagens, vídeo e robótica móvel, para inspeção de animais mortos ou doentes, além da limpeza no ambiente. Na mesma linha, Cuan et al. (2020) focam no vírus da gripe aviária H9N2. O método proposto detecta sons complexos emitidos pelas aves baseado nas diferenças espectrais, obtendo um espectrograma para entrada na rede neural convolucional. Numa abordagem diferenciada e baseada em série histórica de dados, Yousefinaghani et al. (2021) definem um *framework* e *dataset* com escala espaço-temporal para predição de ocorrência e disseminação em área geográfica, com utilização de *features* associadas à gripe aviária.

Com o avanço da tecnologia de sensores, equipamentos de detecção, e consequente injeção de dados nos sistemas inteligentes, a antecipação de ações e disparo de alarmes resultante da operação pode ser um diferencial inovador. Por exemplo, Zhuang et al. (2018) apresentam um sistema de alerta baseado em imagens de câmeras e extração de *features*, para classificar a pose/comportamento de aves saudáveis e doentes. O modelo de Máquina de Vetores de Suporte (SVM) com hiperparâmetros e a função de kernel POLY se destacou, atingindo uma precisão de 99,46% nas amostras de teste. De modo similar, Fang et al. (2021) comparam *frameworks* de redes neurais profundas para estimação não invasiva de pose e identificação de diferentes movimentos das aves e associação às enfermidades, com destaque para o classificador Naive Bayes.

A bioinformática enquanto área multidisciplinar com enfoque na determinação de genes, proteínas, e suas relações, encontra atual e relevante aplicação na área veterinária, com especial atenção à avicultura. Visto o volume de dados e relações estatísticas inerentes, técnicas de IA e aprendizagem de máquina têm sido empregadas. A seleção e combinação ótima de marcadores genéticos, para fins de qualidade e tolerância à contaminantes das matrizes (PASSAFARO et al., 2020; POLEWKO-KLIM et al., 2020) e posterior rastreabilidade de origem (SEO et al., 2021) são exemplos de aplicações em crescimento. Destaca-se a parceria de Passafaro et al. (2020) com a empresa Cobb-Vantress, valorizando a interação academia-indústria.

3.2. Sistemas de monitoramento de produção

O acompanhamento do ganho de peso diário é determinante para atender aos padrões da indústria, garantindo a remuneração adequada dos agentes envolvidos e qualidade final dos produtos. Existem sistemas de pesagem automática baseados em balanças, imagens ou uma combinação de ambos. Uma ampla revisão sobre técnicas baseadas em imagens, não apenas para frango de corte, pode ser encontrada em Nyalala et al. (2021).

Mortensen et al. (2016), a partir de *dataset* rotulado com 13.000 imagens para

treinamento de rede neural Bayesiana, utilizaram câmera 3D para capturar o frango sobre uma balança, obtendo 12 descritores de peso diferentes que permitem estimar o peso individual. Amraei et al. (2018) propõem um sistema equivalente, com *dataset* de 2440 imagens e combinação da Transformada de Hough (HT) generalizada e método de Chan-Vese para segmentação. Os 6 descritores obtidos são utilizados para construção de modelo de regressão para predição de peso. Diferentemente, Wang et al. (2021) propõem uma estrutura baseada em dados, incluindo dois módulos para monitoramento e previsão de peso, por meio de análise de *big data* coletados da nuvem da Amazon (AWS), incluindo sinais eletrônicos coletados das granjas por meio de vários sensores e dispositivos. O módulo de monitoramento de peso é construído por um modelo que integra uma mistura gaussiana, reamostragem de *bootstrapping* e técnica de média ponderada para estimar o peso atual de frangos de corte vivos na granja. Dada a complexidade em filtrar dados obtidos de balanças automáticas e maximizar sua confiabilidade, You et al. (2021) comparam os algoritmos K-Nearest Neighbors (KNN), Random Forest, SVM e redes neurais para classificação de outliers, obtendo melhores resultados com o Random Forest.

Seja no processamento das carnes no frigorífico (YUSOF et al., 2020), na identificação da qualidade de carne em aves saudáveis para exigências sanitárias regionais ou culturais (FAJAR, 2018), ou na classificação das partes do frango (PEREZ et al., 2018), o uso combinado de câmeras hiperespectrais, para aquisição de dados com IA é comum em vários trabalhos. Kollenburg et al. (2020) investigaram a autenticidade de filés de frango com base em imagens no espectro Infravermelho Próximo (NIR) e técnicas de estatística multivariada, enquanto Kamruzzaman et al. (2016) utilizaram regressão por mínimos quadrados parciais para detecção de adulteração em carne moída de frango. Análise semelhante foi realizada por Barbon et al. (2018) porém com utilização de árvore de decisão. Mais recentemente, a presença de contaminantes no fígado do frango foi investigada por Dourou et al. (2021), com opção pela combinação de árvores de decisão (Extra Trees) para extração de características e SVM para classificação.

4. Conclusões

A presente revisão apresentou resultados de pesquisa bibliográfica no período 2016-2021 de sistemas baseados em IA aplicados à avicultura de frangos de corte. Dada a amplitude dos conteúdos dos trabalhos, foi adotada a classificação nas categorias de diagnóstico de doenças/bioinformática veterinária e sistemas de monitoramento de produção, incluindo aspectos de controle de qualidade e processamento agroindustrial.

O diagnóstico antecipado e com assertividade, com indicação de previsão ou probabilidade da ocorrência de doenças em frangos de corte é de grande interesse prático, com abordagens baseadas em monitoramento de comportamento e correlações com a ambiência do aviário. Adicionalmente é notório o avanço nas pesquisas de bioinformática para principalmente o melhoramento genético para minimizar suscetibilidade a doenças. De modo complementar, os sistemas de monitoramento de produção visam otimizar a predição de peso de abate, com estudos de técnicas que possam filtrar medidas de peso indevidas ou incoerentes. Destaca-se o número expressivo de trabalhos na etapa de controle de qualidade da carne de frango, com utilização de câmeras hiperespectrais no espectro NIR em combinação com aprendizagem de máquina.

A presença da IA é realidade na avicultura industrial de frangos de corte, dado o contexto internacional da pesquisa. Embora ainda caracterizada como um elemento diferenciador, espera-se que a incorporação da IA venha a ser natural no futuro próximo, dada a gama de ferramentas e plataformas disponíveis que abstraem a complexidade de seus algoritmos e facilitam a adoção. Aliada a outras tecnologias habilitadoras 4.0, a saber, IoT, *big data* e computação em nuvem, tem demonstrado contribuir para a produtividade, com amplo campo de pesquisa e problemas em aberto na cadeia de produção. Dado o estágio de desenvolvimento das grandes empresas avícolas nacionais e os recordes de exportação, a adoção das tecnologias 4.0, em particular a IA, pode estar inserida nos processos frigoríficos e de logística, necessitando de maior adoção pelos avicultores, principalmente para manejo de ambiência e acompanhamento de ganho de peso. Embora os resultados de pesquisa científica com os critérios adotados neste trabalho não apontem para publicações resultantes da relação indústria-academia nacional, fatores como sigilo por propriedade intelectual e priorização por publicação em eventos serão considerados em trabalhos futuros.

Referências

AMRAEI, S.; MEHDIZADEH, S. A.; NAAS, I. A. Development of a transfer function for weight prediction of live broiler chicken using machine vision. **Engenharia Agrícola**, v. 38, p. 776-782, 2018.

ASSOCIAÇÃO BRASILEIRA DE PROTEÍNA ANIMAL. **Relatório anual 2021.** 2021. Disponível em: <https://abpa-br.org/relatorios/>. Acesso em: 18/ago. 2021.

ASTILL, J. et al.. Smart poultry management: smart sensors, big data, and the internet of things. **Computers and Electronics in Agriculture**, v. 170, 105291, 2020.

ASTILL, J. et al. Detecting and predicting emerging disease in poultry with the implementation of new technologies and big data. **Frontiers in Veterinary Science**, v. 5, 263, 2018.

BANAKAR, A.; SADEGHI, M.; SHUSHTARI, A. An intelligent device for diagnosing avian diseases: Newcastle, infectious bronchitis, avian influenza. **Computers and Electronics in Agriculture**, v. 127, p. 744-753, 2016.

BARBON, S. et al. Machine Learning Applied to Near-Infrared Spectra for Chicken Meat Classification. **Journal of Spectroscopy**, v. 2018, 2018.

BASSI, N. S. S.; SILVA, C. L. Oportunidades e desafios para a cadeia produtiva de frangos de corte brasileira. **Avicultura Industrial**, Itu, ed. 1271, ano 109, n. 10, p.16-21, 2017.

BASSI, N. S. S.; SILVA, C. L. Pesquisa e Desenvolvimento na Cadeia Produtiva de Frangos de Corte no Brasil. **Revista de Economia e Sociologia Rural [online]**, v. 56, n. 3, p. 467-482, 2018.

BRASIL. Agência Brasileira de Desenvolvimento Industrial. **ABDI divulga selecionados do Edital Agro 4.0**. 2020a. Disponível em: <https://www.abdi.com.br/postagem/abdi-divulga-selecionados-do-edital-agro-4-0>. Acesso em: 18 ago. 2021.

BRASIL. Financiadora de Estudos e Projetos. **Seleção pública MCTI/FINEP/FNDCT. Subvenção econômica à inovação - tecnologias 4.0**, 2020b. Disponível em: <http://www.finep.gov.br/chamadas-publicas/chamadapublica/643>. Acesso em: 22 ago. 2021.

CUAN, K. et al. Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network. **Computers and Electronics in Agriculture**, v. 178, id. 105688, 2020.

DOUROU, D. et al. Rapid microbial quality assessment of chicken liver inoculated or not with salmonella using FTIR spectroscopy and machine learning. **Frontiers in Microbiology**, v. 11, 2021.

FAJAR, Y. Artificial neural network for healthy chicken meat identification. **Int. Journal of Artificial Intelligence**, v. 7, n. 1, p. 63-70, 2018.

FANG, C. et al.. Pose estimation and behavior classification of broiler chickens based on deep neural networks. **Computers and Electronics in Agriculture**, v. 180, id. 105863, 2021.

KAMRUZZAMAN, M., MAKINO, Y., OSHITA, S. Rapid and non-destructive detection of chicken adulteration in minced beef using visible near-infrared hyperspectral imaging and machine learning. **Journal of food engineering**, v. 170, p. 8-15, 2016.

KOLLENBURG, G. et al. Dataset of the application of handheld NIR and machine learning for chicken fillet authenticity study. Data in Brief, v. 29, 2020.

MILOSEVIC, B. et al. Machine learning application in growth and health prediction of broiler chickens. **World's Poultry Science Journal**, v. 75, p. 401-410, 2019.

MORTENSEN, A. K.; LISOUSKI, P.; AHRENDT, P. Weight prediction of broiler chickens using 3D computer vision. **Computers and Electronics in Agriculture**, v. 123, p. 319-326, 2016.

MOTHER, D. et al. Preferred reporting items for systematic reviews and meta-analysis: the PRISMA statement. **PLoS Med**, 6, e1000097, 2009.

NYALALA, I. et al. Weight and volume estimation of poultry and products based on camera vision systems: a review. **Poultry science**, v. 100, n. 5, id. 101072, 2021.

OKINDA, C. et al. A review on computer vision systems in monitoring of poultry: a welfare perspective. **Artificial Intelligence in Agriculture**, v. 4, p. 184-208, 2020.

OLIVEIRA, F. D.; SANTANA, A.; PEREIRA, L. F. A. **Impactos da IoT na Avicultura:** um Mapeamento Sistemático. In: XII Congresso Brasileiro de Agroinformática, Indaiatuba, p. 418-427, 2019.

PASSAFARO, T. L. et al. Would large dataset sample size unveil the potential of deep neural networks for improved genome-enabled prediction of complex traits? the case of broiler weight prediction. **BMC Genomics**, v. 21, 771, 2020.

PEREZ, N. et al. Classification of Chicken Parts Using a Portable Near-Infrared (NIR) Spectrophotometer and Machine Learning. **Applied spectroscopy**, v. 72, n. 12, p.1774-1780, 2018.

POLEWKO-KLIM, A. et al. Sensitivity analysis based on random forest machine learning algorithm identifies candidate genes for regulation of innane and adaptive immune response of chicken. **Poultry Science**, v. 99, n. 12, p. 6341-6354, 2020.

RIBEIRO, R. et al. Generating action plans for poultry management using artificial neural networks. **Computers and Electronics in Agriculture**, v. 161, p. 131-140, 2019.

ROCHA, G. B.; SILVA, I. J. O. **A dinâmica da Zootecnia de Precisão e tecnologias de suporte atuais.** 2020. Disponível em: <https://www.aviculturaindustrial.com.br/imprensa/a-dinamica-da-zootecnia-de-precisao-e-tecnologias-de-suporte-atuais/20201218-085415-w971>. Acesso em: 18 ago. 2021.

ROWE, E.; DAWKINS, M.S.; GEBHARDT-HENRICH, S.G. A systematic review of precision livestock farming in the poultry sector: is technology focussed on improving bird welfare? **Animals**, v. 9, 614. doi:10.3390/ani9090614, 2019.

SANTANA, L.; OLIVEIRA, J. **Agricultura 4.0 e o desenvolvimento de pesquisas de computação aplicada às ciências agrárias.** SBC Horizontes, marco. 2021. ISSN 2175-9235. Disponível em: <http://horizontes.sbc.org.br/index.php/2021/03/Agricultura-4-0-e-o-desenvolvimento-de-pesquisas-de-computacao-aplicada-as-ciencias-agrarias/>. Acesso em: 18 ago. 2021.

SCHWAB, K. **The fourth industrial revolution**. World Economic Forum, 2016.

SEO, D. et al. Identification of target chicken populations by machine learning models using minimum number of SNPs. **Animals**, v. 11, n. 241, 2021.

WANG, C.; CHEN, Y.; CHIEN, C. Industry 3.5 to empower smart production for poultry farming and an empirical study for broiler live weight prediction. **Computers & Industrial Engineering**, v. 151, id. 106931, 2021.

YOUSEFINAGHANI, S. et al. A framework for the risk prediction of avian influenza occurrence: An Indonesian case study. **PLoS ONE,** doi: https://doi.org/10.1371/journal.pone.0245116, 2021.

YUSOF, R. et al. Image segmentation and verification based on machine learning for vision inspection of chicken slaughtering. **Journal of Physics**, v. 1147, n. 1, 2020.

YOU, J. et al. A supervised machine learning method to detect anomalous real-time broiler breeder body weight data recorded by a precision feeding system. **Computers and Electronics in Agriculture**, v. 185, 106171, 2021.

ZHUANG, X. et al. Development of an early warning algorithm to detect sick broilers. **Computers and Electronics in Agriculture**, v. 144, p. 102-113, 2018.



Diferenciação e caracterização de municípios produtores de bovinos de corte no bioma Mata Atlântica segundo o Censo Agropecuário 2017

Maria do Carmo Ramos Fasiaben[1], Stanley Robson de Medeiros Oliveira[1], André Steffens Moraes[2], Maxwell Merçon Tezolin Barros Almeida[3], Octávio Costa de Oliveira[3], Gabriela dos Santos Eusébio[4]

[1]Embrapa Agricultura Digital, Campinas, SP.

[2]Emprapa Soja, Londrina, PR.

[3]IBGE, COAGRO, Rio de Janeiro, RJ.

[4]Instituto de Economia/Unicamp, Campinas, SP

{maria.fasiaben; stanley.oliveira; andre.moraes}@embrapa.br, {maxwell.almeida; octavio.oliveira}@ibge.gov.br, gabeusebio@gmail.com

Abstract. *The objective of this work is to typify beef cattle producing municipalities in the* Mata Atlântica *biome based on data from the 2017 Agricultural Census. Special census tabulations were needed to aggregate agricultural establishments by municipality, with emphasis on the technical characteristics of livestock production systems. Four groups of municipalities were identified according to the biome's production patterns: two high-tech groups, one of them with emphasis on confinement; an intermediate technology group and a low technology group. The mapping of the different production patterns throughout the territory can support research, technology transfer and development actions that are more appropriate to the different situations of producers.*

Resumo. *O objetivo deste trabalho é tipificar os municípios produtores de bovinos de corte do bioma Mata Atlântica a partir de dados do Censo Agropecuário 2017. Tabulações especiais do censo foram necessárias para agregar os estabelecimentos agropecuários por município, com ênfase nas características técnicas dos sistemas de produção pecuária. Identificaram-se quatro grupos de municípios segundo os padrões de produção do bioma: dois grupos de alta tecnologia, um deles com destaque para confinamento; um grupo de tecnologia intermediária e outro de baixa tecnologia. O mapeamento dos distintos padrões de produção ao longo do território pode apoiar ações de pesquisa, de transferência de tecnologia e de desenvolvimento mais apropriadas às diferentes situações dos produtores.*

1. Introdução

O bioma Mata Atlântica se estende de norte a sul do país desde o Estado do Rio Grande do Sul até o Piauí ao longo da costa brasileira, e se expande para o oeste, na região Sudeste, até as fronteiras com a Argentina e o Paraguai. Aí vive cerca de 70% da população brasileira e se desenvolve 80% da atividade econômica do país [FUNDAÇÃO SOS MATA ATLÂNTICA, 2021].

Ocupando originalmente 15% do território nacional, atualmente o remanescente de vegetação nativa acima de três hectares deste bioma é de 12,4% em relação ao original, resultado de um intenso processo de exploração [FUNDAÇÃO SOS MATA ATLÂNTICA; INPE, 2021].

Embora em alguns dos 17 estados com área no bioma o desmatamento esteja se aproximando de zero (tecnicamente, quando a área desmatada é menor que 100 ha no período avaliado), a tendência em 10 dos estados é de perda acentuada da vegetação nativa. Os fragmentos maiores do que 100 ha representam somente 8,5% da área de floresta original do bioma [PINTO et al. 2021].

Apesar desse alto grau de devastação, a Mata Atlântica é um dos biomas terrestres mais ricos em espécies animais e vegetais do planeta, sendo considerado pela Unesco um dos *hotspots* mundiais em biodiversidade e um dos biomas brasileiros prioritários para conservação [UNESCO, 2021]. Sua rica variedade biológica se deve aos diferentes ecossistemas existentes no bioma e suas características climáticas e geológicas [INCT-BioNat, s.d; EMBRAPA, 2018]. Por sua extensão ao longo do território nacional, esse bioma ocupa condições diversas de solo, relevo e clima, abrigando ampla variedade de cultivos: grãos, algodão, café, cana-de- açúcar, silvicultura, fruticultura e outros. As pecuárias de leite e de corte podem ser encontradas praticamente em toda sua extensão [FASIABEN et al., 2018].

Apesar de o Brasil ser o segundo maior produtor e o maior exportador mundial de carne bovina, a principal característica no desenvolvimento dessa atividade no país é a heterogeneidade nos sistemas de produção e nos mecanismos de gestão e de comercialização do gado. Sistemas de produção baseados na produção extensiva, pequena intensidade tecnológica e padrões precários de gestão e de comercialização trazem ao setor um forte impacto ambiental [CARVALHO & ZEN, 2017].

Além do aumento da produção para suprir uma demanda crescente, a bovinocultura de corte brasileira terá que se adaptar ao novo cenário nacional e internacional que demanda produtos oriundos de sistemas sustentáveis, ademais da concorrência por área com outras atividades agrícolas [Diagnóstico..., 2016]. O aumento da produtividade dentro de padrões sustentáveis é um imperativo.

No bioma Mata Atlântica, a produção de bovinos cuja principal finalidade é corte representa uma parcela importante da produção nacional. Os dados do Censo Agropecuário 2017 mostram que, se considerada a totalidade dos estabelecimentos agropecuários produtores de bovinos de corte do país (aqueles com uma ou mais cabeças com essa finalidade), a Mata Atlântica detém perto de um terço dos estabelecimentos produtores, 20% do rebanho e 16% da área de pastagens. Tendo-se em conta os estabelecimentos produtores com mais de 50 cabeças de bovinos de corte, de maior expressão comercial, nesse bioma se encontram cerca de 26% dos estabelecimentos, 18% do rebanho e 14% da área de pastagens destinadas a essa finalidade no país,

ficando, nesse último caso, atrás somente do Cerrado e da Amazônia nesses quesitos (IBGE, 2020).

Fatores como a extensão territorial do bioma, sua diferenciação edafoclimática e a diversidade socioeconômica dos produtores implicam em uma grande variedade de sistemas de produção na pecuária de corte (FASIABEN et al., 2018). Sistemas de produção com maior variedade funcional, incluindo sistemas de produção pecuários, tendem a ser mais sustentáveis (MARCONDELLI et al., 2018), o que é bastante positivo em um bioma que requer atenção para a conservação de seus recursos naturais.

Utilizando dados oriundos do Censo Agropecuário 2017 do IBGE, o objetivo deste trabalho foi elaborar uma tipificação dos municípios do bioma Mata Atlântica segundo características da criaçao de bovinos de corte, com enfoque especial na tecnologia empregada pelos produtores. Espera-se que os resultados sejam úteis para o planejamento de ações de pesquisa agropecuária, transferência de tecnologia e desenvolvimento rural, colaborando para a conservação do bioma.

2. Material e métodos

Foi obtida uma tabulação especial do Censo Agropecuário 2017 com variáveis agrupadas por municípios e filtrada por estabelecimentos agropecuários com mais de 50 cabeças de bovinos de corte (IBGE, 2020). Variáveis quantitativas (exceto número de estabelecimentos) com menos de três informantes por município foram desidentificadas, por motivo de sigilo estatístico. As variáveis para a tipificação são listadas a seguir:

1. Número de cabeças de bovinos de corte
2. Taxa de lotação (cabeças/ha de pastagem)
3. Área de pastagem/área em processo produtivo
4. Área de lavoura/área em processo produtivo
5. Área de silvicultura/área em processo produtivo
6. Área de Sistemas Agroflorestais/área em processo produtivo
7. Número de fêmeas de mais de 2 anos/número total de cabeças de bovinos de corte
8. Valor da produção advindo de bovinos de corte/valor da produção agropecuária
9. Receita agropecuária/receita familiar
10. Porcentagem de estabelecimentos que fazem suplementação alimentar
11. Porcentagem de estabelecimentos que fazem confinamento
12. Porcentagem de estabelecimentos que recebem orientação técnica
13. Porcentagem de estabelecimentos que usam adubos químicos
14. Porcentagem de estabelecimentos que usam corretivos do solo
15. Porcentagem de estabelecimentos que têm tratores
16. Porcentagem de estabelecimentos familiares
17. Valor da venda de bovinos de corte para abate/valor de venda de bovinos

Na etapa de modelagem, adotou-se o algoritmo *Expectation-Maximization* (EM) para se estimar o número de grupos (clusters) [DEMPSTER et al., 1977]. A razão da escolha desse algoritmo baseia-se nos critérios: a) EM estende o paradigma usado no k-means, associando uma distribuição de probabilidade a cada instância, que indica a probabilidade dessa instância pertencer a uma única população; b) EM tem a habilidade de lidar com valores faltantes e de estimar o número de clusters por meio de um

conjunto de parâmetros que descreve a distribuição de probabilidade de cada cluster. O usuário também pode especificar *a priori* quantos clusters gerar.

O algoritmo calcula os estimadores de máxima verossimilhança para problemas onde existem dados incompletos entre os dados observados ou quando os estados das variáveis não foram observados, dados não observados. Cada iteração do algoritmo EM envolve dois passos que são: *expectation* (passo E) e *maximization* (passo M). O Passo E desse algoritmo consiste em obter uma estimativa da esperança (*expectation*) dos dados faltosos para completar a amostra de dados incompleta, partindo-se de um valor inicial empírico. No Passo M, com os dados completados, realiza-se aprendizagem das probabilidades, baseado nas frequências dos estados das variáveis na amostra, de modo a calcular a função de maximização da esperança encontrada. Os Passos E e M fazem parte de um processo iterativo, em que as novas probabilidades, calculadas na fase M, serão utilizadas para realizar a inferência na fase E.

A validação cruzada realizada para determinar o número de clusters é feita nas seguintes etapas: a) o número de clusters é definido como 1; b) o conjunto de treinamento é dividido aleatoriamente em 10 partições (*folds*); c) EM é processado 10 vezes usando as 10 partições da maneira usual de validação cruzada; d) a probabilidade da máxima verossimilhança é calculada e a média dos 10 resultados é obtida; e) se a probabilidade de máxima verossimilhança aumentar, o número de clusters é aumentado em 1 e o programa continua na etapa 2. O número de partições é fixada em 10, desde que o número de instâncias (observações) no conjunto de treinamento não seja menor que 10. Se for esse o caso, o número de partições é definido igual ao número de instâncias.

Em particular, foi utilizada a versão do algoritmo EM disponível no *software* Weka, versão 3.8.5 [FRANK et al., 2006]. Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ele contém ferramentas para preparação de dados, classificação, regressão, *clustering*, mineração de regras de associação e visualização.

Uma vez configurados os agrupamentos a partir do conjunto de variáveis, outras informações foram empregadas para melhor caracterizá-los, conforme se descreve nos resultados a seguir.

3. Resultados

O Bioma Mata Atlântica apresentou 91.554 estabelecimentos agropecuários de bovinos de corte com mais de 50 cabeças, detendo um total de 22,1 milhões de cabeças e 15,6 milhões de hectares de áreas de pastagens. Esses estabelecimentos estavam presentes em 2.405 municípios, que foram classificados em quatro grupos (clusters) segundo a metodologia estatística descrita (Tabela 1).

Em primeiro lugar, observa-se a presença de áreas de lavouras que chegam a ser expressivas nos diferentes clusters, denotando a diversificação como característica geral dos sistemas de produção de bovinos de corte praticados no bioma Mata Atlântica.

Os Clusters com maior nível tecnológico (Clusters 2 e 3) são os menos frequentes. Se somados, representam 12% dos estabelecimentos, que se distribuem em 28% dos municípios. Entretanto, esses dois clusters detêm somente 2% e 3% do efetivo de bovinos de corte e 1% e 4% da área de pastagem do bioma, respectivamente (Tabela

1). O Cluster 2 é o que apresenta maior taxa de lotação (média de 2,45 cabeças/ha), seguido do Cluster 3 (2,3 cabeças/ha). Ambos se diferenciam, principalmente, pela maior presença de estabelecimentos que fazem confinamento no Cluster 2. O Cluster 2 e o Cluster 3 são também os que apresentam maior diversificação: a proporção de área destinadas a lavouras (em relação às áreas em processo produtivo) é, respectivamente, de 33% e 56%. O maior percentual de ocupação com florestas plantadas se observa no Cluster 2, sendo aí muito superior à média do bioma (Tabela 2). Observando as variáveis que refletem o nível tecnológico (Tabela 3), vemos que esses dois clusters se apresentam bem acima da média da Mata Atlântica em relação aos quesitos confinamento, suplementação alimentar, uso de corretivos do solo, presença de tratores, orientação técnica e acesso à Internet. Foge da regra o fato de o Cluster 2 ter apresentado menor média de emprego de adubo químico em relação à média do bioma. Já no caso do Cluster 3 o uso de adubação química é bem superior à média geral do bioma. A maior presença de lavouras poderia explicar o maior uso de adubos, uma vez que o censo de 2017 não coleta a informação do destino dos insumos. Uma melhor localização em termos de solos e a possibilidade de rotações entre lavouras e pastagens também podem ser fatores a explicar as maiores taxas de lotação nos Clusters 2 e 3.

O Cluster 1 se apresenta como o agrupamento de municípios que agrega estabelecimentos com nível intermediário de tecnologia, quando se considera o bioma Mata Atlântica. Ele representa 34% dos estabelecimentos e 31% dos municípios analisados. Esse cluster detém 30% do efetivo de bovinos de corte e 26% da área de pastagens destinada a esse tipo de rebanho do bioma (Tabela 1). Das terras em processo produtivo nos estabelecimentos do Cluster 1, 72% são ocupadas com pastagens, 21% com lavouras, 3% com florestas plantadas e 4% com sistemas agroflorestais. Nesse último caso, essa é a maior média entre todos os clusters (Tabela 2). A média da taxa de lotação desse cluster é de 1,64 cabeças/ha. O conjunto de dados da Tabela 3 mostra que esse cluster apresenta valores intermediários nas diferentes variáveis tecnológicas, com exceção do uso de adubos químicos e corretivos, conforme mencionado anteriormente.

O Cluster 4 é o que predomina em termos de estabelecimentos de bovinos de corte (62% do total do bioma) e municípios (41% do total) e é o que se apresenta com menor nível tecnológico no bioma. Ele concentra 62% das cabeças e 69% da área de pastagens destinadas a essa finalidade na Mata Atlântica (Tabela 1). O uso das terras em processo produtivo aponta para 92% com pastagens (superior à média do bioma), 5% de áreas de lavouras (inferior à média do bioma) e 1% tanto para as áreas de florestas plantadas quanto para sistemas agroflorestais (também inferiores à média). Na Tabela 3 pode-se verificar que os valores de todas as variáveis indicativas do nível tecnológico apresentam-se abaixo da média do bioma. Chama a atenção a porcentagem de estabelecimentos agropecuários que recebe algum tipo de orientação técnica neste agrupamento, que é de apenas 40%. A Internet chega a somente 39% dos estabelecimentos agropecuários desse grupo.

Tabela 1. Ocorrências dos clusters de bovinos de corte em estabelecimentos com mais de 50 cabeças no Bioma Mata Atlântica, agregados por municípios, segundo variáveis selecionadas

Variável / Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	M. Atlântica
Número de ocorrências (municípios)	750	311	356	988	2.405
Porcentagem de ocorrências em relação ao total do Bioma	31%	13%	15%	41%	100%
Número de estabelecimentos produtores de bovinos de corte com mais de 50 cabeças (unid)	31.402	3.154	7.813	49.185	91.554
Porcentagem de estabelecimentos em relação ao total do Bioma	34%	3%	9%	54%	100%
Efetivo de bovinos de corte no cluster (cabeças)	6.593.997	418.293	1.428.583	13.631.252	22.072.125
Porcentagem de cabeças do cluster em relação ao total do Bioma	30%	2%	6%	62%	100%
Área de pastagem do cluster (ha)	4.027.140	170.717	621.049	10.742.422	15.561.328
Porcentagem da área de pastagem em relação ao total do Bioma	26%	1%	4%	69%	100%

Tabela 2. Características quanto ao uso da terra de estabelecimentos produtores de bovinos de corte com mais de 50 cabeças no Bioma Mata Atlântica, agregados por municípios, segundo clusters e total Mata Atlântica

Variável / Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	M. Atlântica
Área total (média) dos estabelecimentos agropecuários (ha)	229	142	257	290	261
Área média em processo produtivo dos estabelecimentos (ha)	179	103	205	237	210
Área média de pastagem dos estabelecimentos (ha)	128	54	79	218	170
Porcentagem da área ocupada com pastagens nos estabelecimentos	72%	53%	39%	92%	81%
Porcentagem da área ocupada com lavouras nos estabelecimentos	21%	33%	56%	5%	15%
Porcentagem da área ocupada com florestas plantadas nos estabelecimentos	3%	10%	2%	1%	2%
Porcentagem da área ocupada com sistemas agroflorestais nos estabelecimentos	4%	1%	1%	1%	2%

Tabela 3. Aspectos gerais da tecnologia empregada nos estabelecimentos produtores de bovinos de corte com mais de 50 cabeças no Bioma Mata Atlântica, agregados por municípios, segundo clusters e total Mata Atlântica

Variável / Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	M. Atlântica
Tamanho médio do rebanho por estabelecimento (cabeças)	210	133	183	277	241
Porcentagem de vacas com 2 anos ou mais em relação ao rebanho total	34%	27%	31%	28%	30%
Taxa de lotação (cabeça/ha)	1,64	2,45	2,30	1,27	1,42
Porcentagem de estabelecimentos que tem acesso a internet	46%	60%	56%	39%	44%
Porcentagem de estabelecimentos que tem trator	61%	71%	77%	40%	51%
Porcentagem de estabelecimentos que faz confinamento	18%	45%	29%	10%	16%
Porcentagem de estabelecimentos que usa ração, grãos e subprodutos agroindustriais	66%	100%	79%	57%	63%
Porcentagem de estabelecimentos que usa adubos químicos	36%	23%	44%	20%	28%
Porcentagem de estabelecimentos que usa corretivos do solo	38%	34%	50%	26%	33%
Porcentagem de estabelecimentos que recebe orientação técnica	57%	67%	77%	40%	50%

Tratando das características socioeconômicas dos estabelecimentos agregados por município (Tabela 4), observa-se que o Cluster 2 concentra o maior percentual de estabelecimentos da agricultura familiar (LEI-11326 de 24-07-2017), com o valor de 51%. Para o público objeto desta análise, a média de estabelecimentos classificados como de agricultura familiar é de 33% e os demais clusters se aproximam desse valor.

Tabela 4. Características socioeconômicas dos estabelecimentos produtores de bovinos de corte com mais de 50 cabeças no Bioma Mata Atlântica, agregados por municípios, segundo clusters e total Mata Atlântica

Variável / Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	M. Atlântica
Porcentagem de estabelecimentos familiares	34%	51%	33%	30%	33%
Porcentagem do valor da produção advinda dos bovinos em relação ao valor total da produção	39%	34%	19%	71%	47%
Porcentagem da receita familiar que advém da agropecuária	87%	80%	94%	65%	78%
Porcentagem de outras receitas (aposentadoria, trabalho fora, etc) em relação à receita total do estabelecimento	10%	15%	4%	33%	20%
Porcentagem de estabelecimentos que tem energia elétrica	93%	95%	94%	90%	91%

A porcentagem do valor da produção agropecuária que advém da produção de bovinos é maior no Cluster 4, da ordem de 71%, lembrando que esse é o cluster que mostrou menor grau de diversificação da produção. Já para os Clusters 2 e 3 (os mais diversificados) esse percentual cai para 34% e 19%, respectivamente, enquanto no Cluster 1 é de 39%. A maior parte da receita familiar provém da atividade agropecuária em todos os clusters, mas essa participação é menor no Cluster 4 - o de menor nível tecnológico e menor percentual de estabelecimentos familiares -, sendo aí da ordem de 65%. No Cluster 4 é também onde se observa a maior participação de outras receitas (aposentadoria, trabalho fora, etc) na formação da receita total do estabelecimento, com uma média de 33% (Tabela 4).

Por fim, a Tabela 4 nos aponta que a energia elétrica ainda não é uma realidade para 10% dos estabelecimentos do Cluster 4. A média de estabelecimentos com acesso à energia elétrica entre os produtores de bovinos de corte com mais de 50 cabeças no bioma Mata Atlântica é de 91%.

As Figuras 1 a 4 mostram a localização espacial dos quatro clusters da Mata Atlântica. Verifica-se que que o Cluster 2 e o Cluster 3, de maior nível tecnológico, concentram-se nos estados do Sul e Sudeste do país. O Cluster 1 (nível tecnológico intermediário) se distribui predominantemente desde o estado de Santa Catarina até o Sudeste da Bahia. Já o nível tecnológico baixo (Cluster 4) se distribui ao longo de toda a extensão do bioma, mas com baixa frequência nos estados de Santa Catarina e Rio Grande do Sul. A Figura 5 reúne a localização espacial dos quatro clusters. Análises adicionais com emprego de geoestatística são interessantes para explicar o efeito da localização espacial sobre a produção e produtividade dos diferentes agrupamentos.

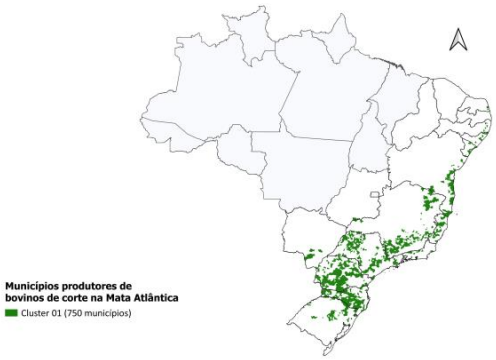


Figura 1. Localização espacial do Cluster 1

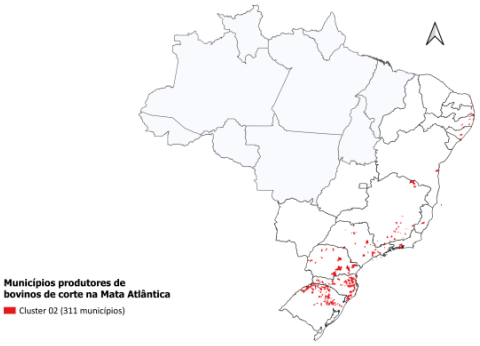


Figura 2. Localização espacial do Cluster 2

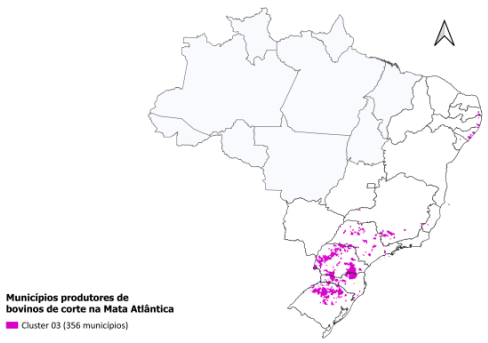


Figura 3. Localização espacial do Cluster 3

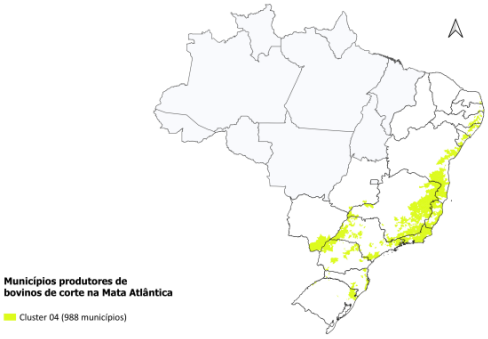


Figura 4. Localização espacial do Cluster 4

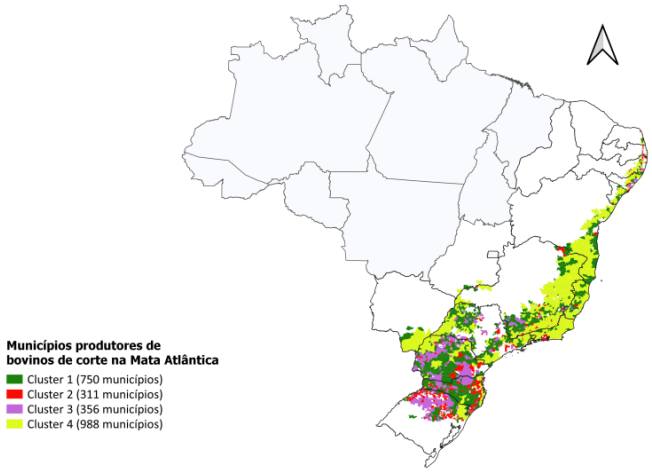


Figura 5. Localização espacial dos 4 clusters de municípios com produtores de bovinos de corte com mais de 50 cabeças no bioma Mata Atlântica.

4. Conclusões

Estudos como este são relevantes por mostrarem a realidade da produção praticada ao longo do país, podendo apontar as áreas prioritárias para ações de pesquisa, transferência de tecnologia e desenvolvimento, colaborando para a conservação dos recursos naturais do bioma.

Analisando os estabelecimentos agropecuários com mais de 50 cabeças de bovinos de corte do bioma Mata Atlântica, agrupados por municípios, pode-se diferenciar e caracterizar os sistemas de produção predominantes, permitindo elaborar o mapa da sua distribuição ao longo do território nacional, com enfoque principal no seu nível tecnológico e acesso a serviços como assistência técnica, internet e energia elétrica. Por esse mapa, podem ser melhor definidas áreas prioritárias para atuação institucional com a finalidade de melhorar a produção e a produtividade da pecuária regional.

Entre os estudos futuros, considera-se importante estabelecer comparações entre os dados dos dois últimos censos agropecuários (de 2006 e 2017) de modo a compreender a evolução dos sistemas de produção praticados no país nesse espaço de onze anos.

Referências

Carvalho, T. B. de; Zen, S. de (2017). A cadeia de Pecuária de Corte no Brasil: evolução e tendências. Revista iPecege v.3, n.1, p. 85-99. DOI: 10.22167/r.ipecege.2017.1.85. Disponível em: <https://revista.ipecege.org.br/Revista/article/view/109>. Acesso em: 22 out. 2021.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. "Journal of the Royal Statistical Society". Series B (Methodological), Blackwell, v. 39, n. 1, p. 1-38.

Frank, E.; Hall, M. A.; Witten, I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.

DIAGNÓSTICO da pecuária de corte em Minas Gerais. Belo Horizonte: Sistema FAEMG, 2016. 147 p. Disponível em: <http://www.faemg.org.br/Content/uploads/publicacoes/arquivos/Xcui1592321508129.pdf>. Acesso em: 22 out. 2021.

EMBRAPA. "Rede ILPF. BIOMA: MATA ATLÂNTICA". [S.l.] Disponível em: <https://www.embrapa.br/web/rede-ilpf/bioma/mata-atlantica>. Acesso em: 19 ago. 2021.

Fasiaben, M. do C. R., Oliveira, Almeida, Abreu. (2018) "Diferenciação e caracterização de municípios produtores de bovinos de corte no bioma Mata Atlântica". In: CONGRESSO DA SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 56., 2018, Campinas. Transformações recentes na agropecuária brasileira: desafios em gestão, inovação, sustentabilidade e inclusão social: anais. Brasília, DF: SOBER, 2018. Não paginado.

Fundação SOS Mata Atlântica, INPE. (2021) "Atlas dos remanescentes florestais da Mata Atlântica, período 2019-2020". Relatório Técnico. São Paulo. 73p.

Fundação SOS Mata Atlântica. (2021) "Desmatamento da Mata Atlântica cresce em dez estados". 2021. Disponível em: <https://www.sosma.org.br/noticias/desmatamento-da-mata-atlantica-cresce-em-dez-estados/> Acesso em: 19 ago. 2021.

IBGE. (2020) "Tabulação especial do Censo Agropecuário 2017 de estabelecimentos de bovinos de corte com mais de 50 cabeças agregados por município". Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, Coordenação de Agropecuária.

INCT-BioNat. "Mata Atlântica". [S.l.]. Disponível em: <http://inct-bionat.iq.unesp.br/biomas/mata-atlantica/>. Acesso em: 19 ago. 2021.

Marcondelli, A. C. B., Nicodemo, M. L. F., Barioni Júnior. (2018) "Diversidade funcional em sistemas de produção pecuários no bioma Mata Atlântica". In: SIMPÓSIO BRASILEIRO DE AGROPECUÁRIA SUSTENTÁVEL, 9.; CONGRESSO INTERNACIONAL DE AGROPECUÁRIA SUSTENTÁVEL, 6., 2018, Viçosa, MG. Ciência, tecnologia e inovação para o desenvolvimento agropecuário sustentável: anais...Viçosa, MG: Biblioteca Central, Campus UFV, 2018. 142-146 SIMBRAS.

Pinto, L. F. G., Hirota, M., Rosa, M. (2021) "Atlas da Mata Atlântica e Mapbiomas se complementam no monitoramento do bioma". Disponível em: <https://www.sosma.org.br/artigos/atlas-da-mata-atlantica-e-mapbiomas-se-

complementam-no-monitoramento-do-bioma/>. Acesso em: 19 ago. 2021.

UNESCO. “World Heritage List”. (2021) Disponível em: <http://whc.unesco.org/en/list/≥. Acesso em: 19 ago. 2021.



DSSAT-Lite: a web-based application for running crop models and analyzing results from DSSAT-CSM

Jonas de Abreu Resenes[1], Gerrit Hoogenboom[2], Alexandre Lazaretti Zanatta[1], Willingthon Pavan[1,3], Carlos Amaral Holbig[1]

[1]Graduate Program in Applied Computing (PPGCA)
University of Passo Fundo (UPF)
Campus 1 – 99.052-900 – Passo Fundo – RS – Brazil

[2]University of Florida
Gainesville – FL – United States of America

[3]International Fertilizer Development Center (IFDC)
Muscle Shoals – AL – United States of America

jabreu.ar@gmail.com, gerrit@ufl.edu

zanatta@upf.br, wpavan@ifdc.org, holbig@upf.br

Abstract. *DSSAT is a set of tools that facilitate the creation and management of experiment, soil, and weather files. Its tools were built using programming languages like Visual Basic and Delphi, which are difficult to run remotely or in the cloud. In this work, we present DSSAT-Lite, a web-based application designed to facilitate the simulation of DSSAT-CSM models over HTTP. It provides simulation and visualization tools and integration with highly optimized third-party software for processing DSSAT files. Abstraction layers allow for a standardized methodology for reading multiple DSSAT files into a beautifully formatted data structure. These features allowed DSSAT-Lite to run remote simulations, specifically using mobile devices and web browsers.*

Resumo. *DSSAT é um conjunto de ferramentas que facilita a criação e gerenciamento de arquivos de experimentos, solo e clima. Suas ferramentas foram construídas utilizando linguagens de programação como Visual Basic e Delphi, que se tornam difíceis de serem executadas remotamente ou na nuvem. Neste trabalho, apresentamos o DSSAT-Lite, uma aplicação baseada na web, projetada para facilitar a simulação de modelos DSSAT-CSM sobre HTTP. Ele*

fornece ferramentas de simulação e visualização e integração com software de terceiros altamente otimizado para o processamento de arquivos DSSAT. As camadas de abstração permitem uma metodologia padronizada para ler vários arquivos DSSAT em uma estrutura de dados muito bem formatada. Estas características permitiram ao DSSAT-Lite executar simulações remotas, especificamente usando dispositivos móveis e navegadores web.

1. Introduction

With each technological advance and with the spread of the use of computer programs, new computational challenges arise. The replacement of the command line interface with the graphical interface facilitated the increasing use of programs by users. However, this transformation caused another problem for software developers: code reuse across different platforms and integration with different programming languages.

In recent years, several innovations in web technology have allowed designers and software engineers to develop responsive web applications [Shahzad 2017]. The web-based approach also reduces the complexity of software distribution, which is a challenge for any development team. However, many tools created before the 2000s use technologies that today are hard to maintain and integrate with modern computational tools.

The tool that drives this work is the Decision Support System for Agrotechnology Transfer (DSSAT), a software package that integrates soil effects, genetic coefficients, climate data, and management options. DSSAT is a collection of databases, crop models, and different applications that are operated by software to help the user select and compare different options and predict the final results [Hoogenboom et al. 2019a].

The DSSAT development was done using different programming languages and with little attention to the structure of the software. It has been redesigned to facilitate the incorporation of new scientific advances, applications, documentation, and maintenance. The principle of this remodeling is a modular structure, called **CSM** (cropping system model), in which the components are separated along scientific discipline lines and structured to allow easy replacement or addition of new modules in the DSSAT-CSM [Jones et al. 2003]. Figure 1, shows that DSSAT-CSM is made up of four different components: a) a database collection, used to organize and store a minimum set of input data required to run crop models; (b) a series of crop models; c) applications to analyze and display the simulated experiments; d) a user interface [Verhagen et al. 2001].

Each DSSAT module has six operational steps, as shown in Figure 1, which are responsible for structuring and organizing a simulation, such as: initialization, crop installation, rate calculations, integration, daily outputs and summary of outputs. The *Main Program* controls when each of these steps is triggered and when each module should perform a task.

The DSSAT functions have been developed primarily to support the use of crop simulation models in decision making [Jones et al. 1998]. The following crop models are currently accessible in the DSSAT environment:

- CERES for cereal models, which include corn, wheat, rice, barley and sorghum..
- CROPGRO de models for legumes such as soybeans and peanuts.
- CROPSIM for root crops such as cassava and potato.

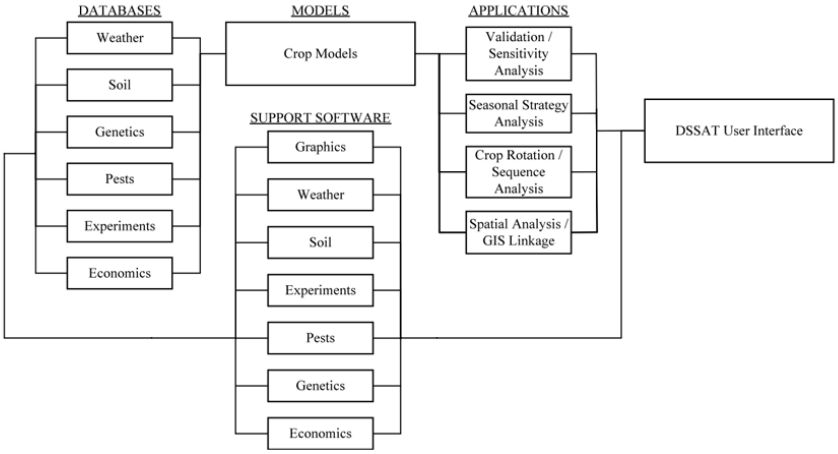


Figure 1. Diagram of database, application, and support software components and their use with crop models for applications in DSSAT [Jones et al. 2003].

- CROPGRO allows users to modify values in a model culture file without changing any code.

One of DSSAT's strengths is the ability to analyze many management strategies. Several years of simulation are run using climate data generated to simulate the effects of the climate in the future [Jones et al. 2003]. In addition, DSSAT is suitable for long-term studies, which assists in the evaluation of efficient crop management strategies and optimize production [Faria and Bowen 2003].

Among the limitations of the DSSAT, the restriction to homogeneous field-scale analyzes stands out [Jones et al. 2003]. Also, the fact that crop models are written using the FORTRAN programming language makes it difficult to integrate with new components. The DSSAT user interface, Figure 1, is supported on Windows only and uses technologies in the end of life support.

The DSSAT models are executed through the command-line interpreter that requires a particular experience in computer science. Most of the operations in DSSAT are made over a command-line interpreter, which requires a piece of good computational knowledge. Since the DSSAT community is interdisciplinary, not only computer science or others are related, the use of command-line becomes a challenge for most users. Many times, it is a challenge for researchers in the computer science area as well.

There is one alternative for the command-line, which is the DSSAT Shell program (Figure 2). It provides a working environment where various stand-alone tools and applications are seamlessly integrated with the DSSAT crop models. Within the shell, the user can launch applications for creating and modifying data files, running the crop models, and analyzing the results [Hoogenboom et al. 2019b]. However, the current DSSAT Shell is a desktop application that runs on Windows Operation System only, and it was written using Delphi and many tools in Visual Basic language. Like any other desktop application, it requires installing in a local machine, and simulation and data visualization are accessed offline.

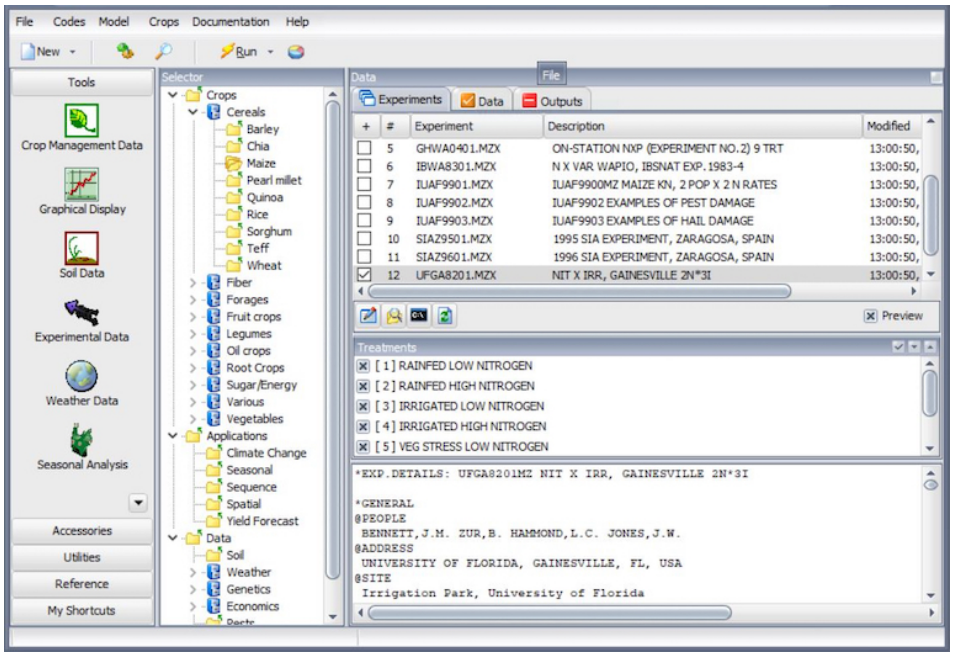


Figure 2. Screenshot from the main DSSAT-Shell window, available under DSSAT version 4.7.5 installation package.

This study intent to help the DSSAT community which contains approximately $14,000$ researchers, educators, consultants, extension agents, growers and decision makers in over 179 countries worldwide. This work aims to build a solution that allows the DSSAT community to run and visualize DSSAT crop models over HTTP, to run simulations and visualize results from anywhere in the globe, through a mobile device or via browser using World Wide Web.

2. Software description

Some functional requirements and functionalities have guided our implementation. The main functionalities from DSSAT Shell that DSSAT-Lite will support are 1) **Responsive**: DSSAT-Lite pages should render well on a variety of devices and window or screen sizes. It also must support mobile devices; 2) **Parallelism**: it must support multiple simulations at once. Simulations should not impact each other; 3) **Performance**: slow sites have a negative impact for the users. DSSAT should be able to run simulations and visualize result quickly; and 4) **User-friendly**: DSSAT-Lite must provide an user-friendly interface to the users; even users that are starting with DSSAT should be able to find the path to run a simulation and to visualize results easily.

3. Technical Implementation

Many frameworks and libraries were used during DSSAT-Lite development. These frameworks are **jDSSAT** [Resenes et al. 2019], **Express.js**, **Angular**, **Bootstrap** and **Jasmine**.

The DSSAT-Lite application design is based on an architecture that follows the Client/Server approach via Representational State Transfer (REST) [Fielding and Taylor 2000]. All the components of the architecture are

shown in Figure 3. The client, a web application, handles the DSSAT-Lite user interface with responses to user events such as selecting experiments for a given crop and running DSSAT model simulation. The server-side provides the data for the user's events, which runs on a remote computer. It has a DSSAT-CSM version installed that will run the DSSAT models when requested from the client. The server-side is a NodeJs REST API that is accessible through HTTP from the client. Since the DSSAT folder structure is shared in the remote computer, every user has a temporary folder for experiments and simulation outputs. There is a service installed on the remote computer to clean up the temporary folder every seven days.

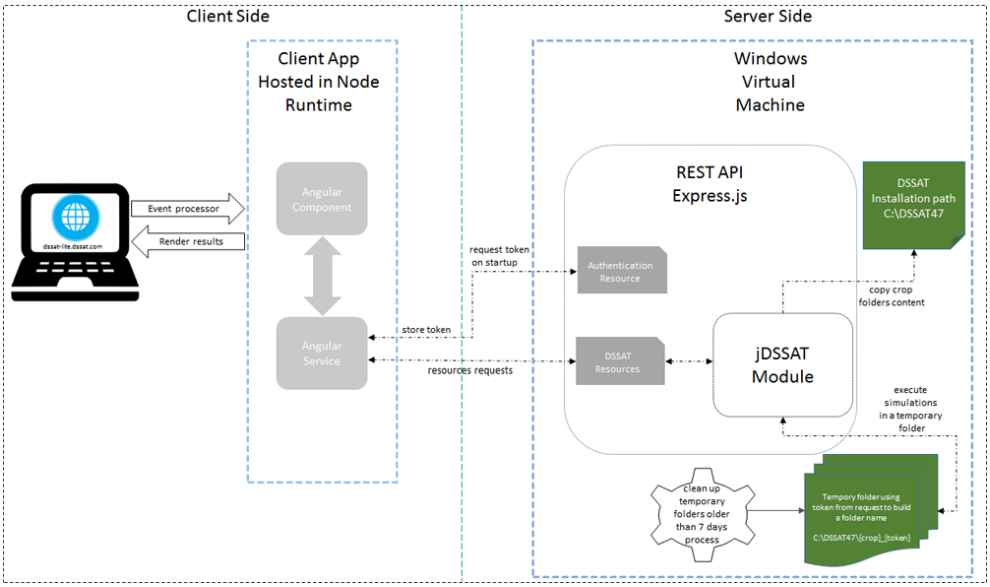


Figure 3. The DSSAT-Lite follows client-server architecture.

Once the user accesses the DSSAT-Lite Uniform Resource Locator (URL), a request is made to the Middleware REST API, specifically to the authentication service, in order to generate a token. We keep the token in the browser local storage. This token is used later to create a temporary crop folder to avoid concurrence issues. The simulation from user *A* won't impact a simulation from user *B*. Both users will have different temporary folders.

After the token is generated and stored, we can now use the DSSAT resource available in the Middleware API. The DSSAT resource has a different path in the URL. For instance, to retrieve experiments for a given crop, the URL [1] path will contain the crop name. The HTTP method, in this case, will be a GET. The result of this request will be a JSON Array [Crockford 2006], and it will be handle by the client-side before displaying it to the user.

DSSAT-Lite allows any user to run a DSSAT-CSM simulation model remotely. In this case, a temporary folder with the results is created. For example, the user *A* accesses the DSSAT-Lite URL; at this moment, a token is generated by the middleware

[1] https://fqdn/api/dssat/treatments:crop

API (for instance *9f5IEN7BWP*). The user then will select what: the crop, experiments, and treatments to be simulated, in this order, and push *Run Model* button in the DSSAT-Lite page, and a request is made to the Middleware API to run the simulation.

In the DSSAT-Lite interface (Figure 4), there are four menus on the left: Simulation, Graph Builder, Version Info, and Contact. The simulation contains information about the experiments and treatments. Graph Builder is used to plotting graphs and visualize simulation results. The other two menus are version info, including the current release note and contacts used to report issues. At the top of the Simulation page, a dropdown with the available crops to be selected. This means that the users can choose different crops to run a simulation. Next, the button Run Simulation is used to run a simulation in the remote server after selecting Experiments and treatments.

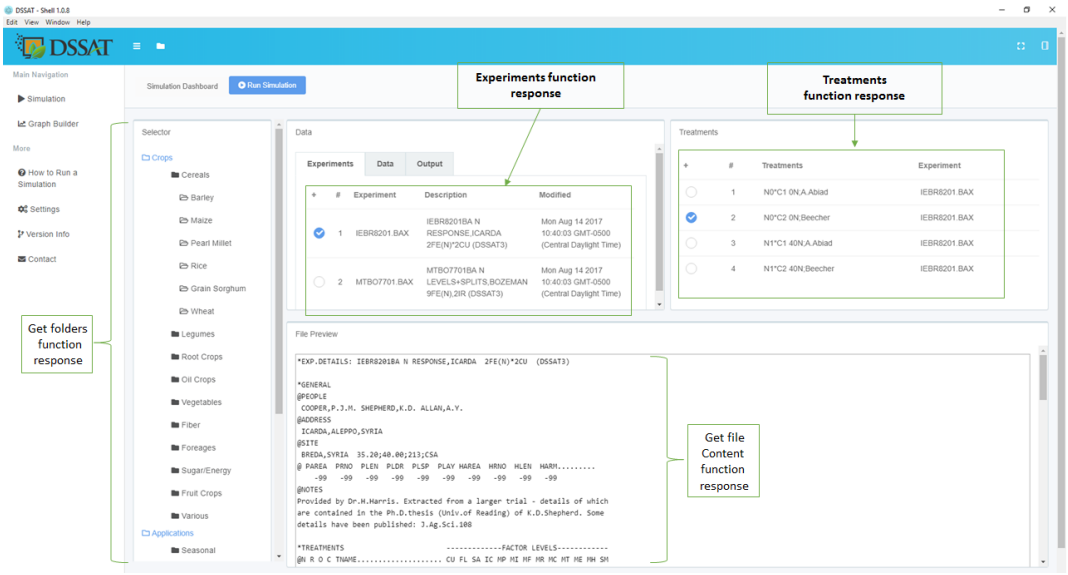


Figure 4. DSSAT-Lite simulation page. The data card displays all experiments based on the crop selected, Maize, in this case. The treatments displayed on the right card are loaded based on the experiment file the user selected. There is also a File Preview card of the experiment file

4. Illustrative example

DSSAT-Lite has been implemented over HTTP with the integration of **ploty.js** library for graphical results. Furthermore, this solution enables multiple simulations simultaneously using APIs.

Our example, depicted in Figure 4, represents the Maize crop selection by the DSSAT-Lite user. Whenever an event occurs in the Web interface, a request is made to the server where the DSSAT-Lite backend is hosted. Once the server receives the request, it looks at experiments files available from the Maize crop folder. The treatments available for selection are based on the experiment selected. After treatments and experiments are selected, the user can now run a simulation. A graphical representation of running simulation steps is shown in Figure 5.

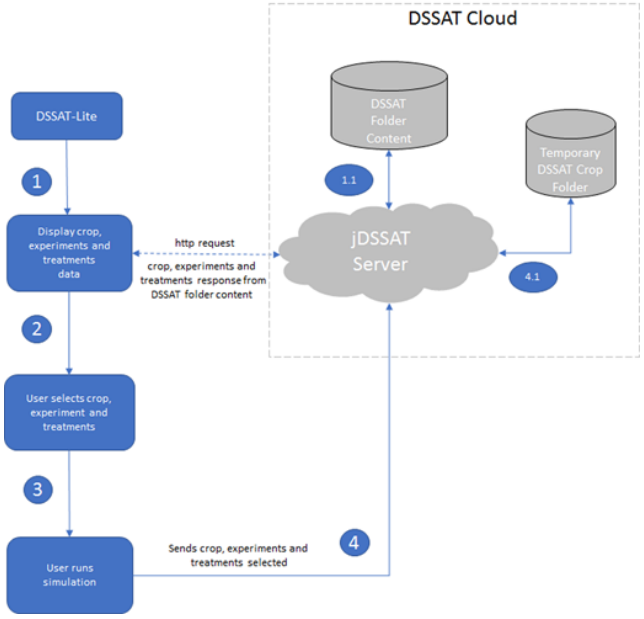


Figure 5. Flow diagram showing the steps to run a simulation. Step 1 is for the DSSAT-Lite to load the crops, experiments ad treatments using jDSSAT server API (step 1.1). Step 2 represents DSSAT-Lite displaying the data. Step 3 represents experiment and treatments selection by the user. Finally, the user runs a simulation. The jDSSAT server processes the simulation request and creates a temporary folder to save the simulation results and retrieve them later (steps 4 and 4.1).

After a simulation is completed, the user can visualize the results. Crop simulation models are process-based and provide very detailed outputs of the simulated crop and soil processes [Yang et al. 2014]. However, analyzing the outputs is a challenge, and the existing tools such Easy Grapher [Yang et al. 2014] do not run through the web. The DSSAT-Lite is responsive, and it automatically resizes, hides, shrinks, or enlarges its content to make its user interface fit in all devices (desktops, tablets, and phones). Figure 7, shows how the DSSAT-Lite layout fits in a mobile device. The DSSAT-Lite provides a page to visualize simulation results, shown in Figure 6.

5. Final Considerations

The philosophy of abstracting complexity from software execution should enable the evolution of DSSAT as software. DSSAT-Lite combines APIs and third-party tools to allow the DSSAT community to run experiments remotely. In the current DSSAT-Lite version, the users meet limited usage features compared with the current DSSAT-Shell. However, running applications through the web opens up the possibility to run multiple simulations at once. The APIs used by DSSAT-Lite is versatile and extensible. These APIs can be used to build other solutions, such as the R package, to run simulations using R language, leveraging its power of data visualization.

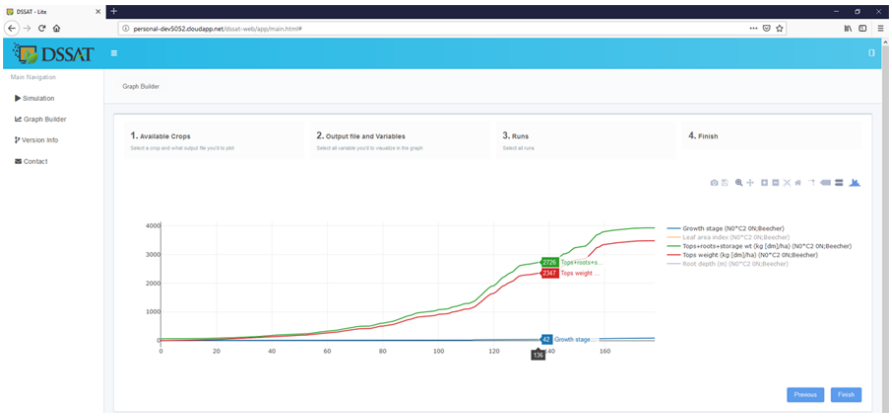


Figure 6. Graph builder page. It displays the values of the variables from a simulation. It has options to download the graph, zoom in and out and show/hide variables.

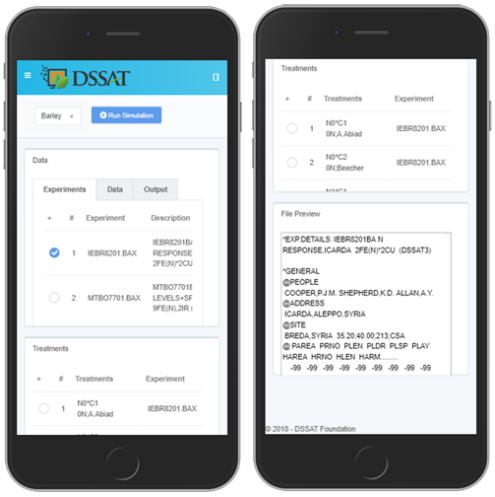


Figure 7. DSSAT-Lite running in a mobile device.

References

Crockford, D. (2006). The application/json media type for javascript object notation (json). *RFC*, 4627:1–10. https://doi.org/10.17487/RFC4627.

Faria, R. T. d. and Bowen, W. T. (2003). Evaluation of dssat soil-water balance module under cropped and bare soil conditions. *Brazilian Archives of Biology and Technology*, 46(4):489–498.

Fielding, R. T. and Taylor, R. N. (2000). *Architectural styles and the design of network-based software architectures*, volume 7. University of California, Irvine Doctoral dissertation.

Hoogenboom, G., Porter, C., Boote, K., Shelia, V., Wilkens, P., Singh, U., White, J., Asseng, S., Lizaso, J., Moreno, L., Pavan, W., Ogoshi, R., Hunt, L., Tsuji, G., , and Jones, J. (2019a). The dssat crop modeling ecosystem. In Boote, K., editor, *Advances in Crop*

Modeling for a Sustainable Agriculture, pages 173–216. Burleigh Dodds Science Publishing. http://dx.doi.org/10.19103/AS.2019.0061.10.

Hoogenboom, G., Porter, C., Boote, K., Shelia, V., Wilkens, P., Singh, U., White, J., Asseng, S., Lizaso, J., Moreno, L., Pavan, W., Ogoshi, R., Hunt, L., Tsuji, G., and Jones, J. (2019b). The dssat crop modeling ecosystem. In Boote, K., editor, *Advances in Crop Modeling for a Sustainable Agriculture*, pages 173–216. Burleigh Dodds Science Publishing. http://dx.doi.org/10.19103/AS.2019.0061.10.

Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., and Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3):235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Jones, J. W., Tsuji, G. Y., Hoogenboom, G., Hunt, L. A., Thornton, P. K., Wilkens, P. W., Imamura, D. T., Bowen, W. T., and Singh, U. (1998). Decision support system for agrotechnology transfer: DSSAT v3. In Tsuji, G. Y., Hoogenboom, G., and Thornton, P. K., editors, *Understanding Options for Agricultural Production*, pages 157–177. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-017-3624-4_8.

Resenes, J. A., Pavan, W., Hölbig, C. A., Fernandes, J. M. C., Shelia, V., Porter, C., and Hoogenboom, G. (2019). jDSSAT: A javascript module for dssat-csm integration. *SoftwareX*, 10:100271. https://doi.org/10.1016/j.softx.2019.100271.

Shahzad, F. (2017). Modern and responsive mobile-enabled web applications. *Procedia Computer Science*, 110:410 – 415. 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) / 12th International Conference on Future Networks and Communications (FNC 2017) / Affiliated Workshops. https://doi.org/10.1016/j.procs.2017.06.105.

Verhagen, A., Conijn, J., and Schapendonk, A. (2001). Quickscan of simulation models. Technical report, Plant Research International BV. Scientific report, Nota 130. Available at: https://edepot.wur.nl/26948.

Yang, J., Drury, C., Yang, J., Li, Z., and Hoogenboom, G. (2014). Easygrapher: software for data visualization and statistical evaluation of dssat cropping system model and the canb model. *International Journal of Computer Theory and Engineering*, 6(3):210. https://doi.org/10.7763/IJCTE.2014.V6.864.



Effects of agro-sensor time series approximation on plant stress detection: an experimental study

Marcos A. de Oliveira Junior[1,2], Gregory Sedrez[1], Anderson Monteiro[1,2], Fernando Emilio Puntel[1], Gerson Geraldo H. Cavalheiro[1]

[1]Programa de Pós-Graduação em Computação – Universidade Federal de Pelotas
Pelotas – RS – Brazil

[2]Instituto Federal de Educação, Ciência e Tecnologia Farroupilha
Santa Maria – RS – Brazil

{marcos.oliveira,gdbsedrez,anderson.monteiro,fepuntel,
gerson.cavalheiro}@inf.ufpel.edu.br

Abstract. *This paper describes an experimental study on the effect of reducing time series collected from IoT electrical agro-sensors through approximation techniques, in time series classification tasks, for plant stress detection. From large sets of real data, stored in time series format, experiments were carried out to analyze: (i) performance of mathematical methods to reduce the dimensionality of time series - PAA, SAX and MCB; and (ii) Whether the application of these techniques influences the performance of time series classification models for plant stress detection, using machine learning algorithms KNN, SVM and ANN. Both in terms of data volume reduction and time series classification, the experiment showed significant improvements in terms of compression rate and accuracy, with the best result found in the use of PAA+SAX techniques for reduction and SVM for classification.*

1. Introduction

The usage of sensors and small boards offering processing capacity with low-energy requirements is allowing the growing of many applications in the IoT domain. The large number of devices in such applications produces lots of data that need to be stored, processed and transmitted between the application components. The time series format is often used in these applications to represent a set of data collected by a device, such as sensors, over a period of time [Blalock et al. 2018]. One characteristic of data collected by sensors is large volume and/or dimensionality. From this, algorithms emerge in order to synthesize time series of data, in order to facilitate the communication, processing and storage of these data [Krawczak and Szkatuła 2014].

Agriculture is one of the many areas that benefit from these techniques, as sensors are installed in plantations, grain cycles, and agricultural machinery [Zhang et al. 2020]. However, in most cases, sensors and stations installed in agricultural environments do not have wide availability of resources, such as storage space, or even limited hardware for data transmission. One of the important applications in the area of agriculture is related to activities that involve the application of water resources. The concern with the sustainability and finite availability of this resource implies solutions that address this issue, proposing an efficient irrigation control, since both scarcity or excess of water can cause problems for plantations. Still, in relation to the application of agricultural inputs, there is a big economic issue related, also, to an intelligent application of the products. To collect this information on water stress and obtain a better basis for decision-making, many solutions provide for the installation of sensors in different regions of the field and/or even on the plants for data collection and transmission. In many cases, data collection and transmission is performed every few seconds, generating a large amount of data. Due to the large amount of data generated, there are currently several data reduction techniques applied to agriculture, which use mathematical tools, such as time series approximation, to reduce the volume of data to be manipulated.

With a range of available techniques (reduction, compression, indexing, approximation, among others) it is extremely important to evaluate these in real applications. Because of this, this article aims to evaluate methods of approximation of time series, from real data, collected in an experiment of electrophysiology of bean plants [Toledo 2019]. For this, mathematical techniques were implemented to approximate time series known in the literature, in order to reduce the initial dataset in different ways and analyze the impact of each reduction format. Based on the concepts presented above, an experiment and a discussion of the results were developed, with the purpose of measuring and analyzing the impact of the reduction of time series, in machine learning tasks, specifically in model training and classification of time series.

The paper is organized as follows: Session 2 presents works related to the approximation of time series; Session 3 presents characteristics of the used dataset and the description of the experiment performed; Session 4 presents the discussion of the obtained results; and Session 5 presents the conclusion and future research possibilities.

2. Literature Review/Related Work

Time series can be generated from any measurement, in a time interval, where often the amount of data is immense. The approximation of time series is extremely important, both to reduce the volume of stored data and to reduce the amount of data transmitted and processed. In several applications, such as the Financial Market [Xin et al. 2019] and the Health area [Tobore et al. 2020], time series are widely used in solutions that generate a large volume of data in short periods of time. Because of this, data approximation techniques are valuable tools to obtain satisfactory system performance.

However, unlike the areas mentioned, where, generally, information collection, transmission and processing are carried out on computers with wide availability of resources, in agriculture it is important to use time series approximation algorithms to improve sensors scattered in crops, with scarce computational resources. Solutions with an agricultural bias, such as [Al-Qurabat and Kadhum Idrees 2020], use time se-

ries approximation algorithms for data reduction, storage and transmission, and also with low sensor energy consumption and avoiding redundant data. In these solutions for agricultural sensors, it is also observed the use of clustering algorithms when receiving the data, seeking better performance and preservation of sensors and stations in [Al-Qurabat and Idrees 2019] crops.

Currently, there are different algorithms for approximation of time series that can be used in the most diverse areas. Thus, in this work, an evaluation of some traditional approximation algorithms for time series is proposed, in combination with classification algorithms, in order to identify the effect of approximation on classification tasks. Analyzing different aspects of the behavior of the algorithms, a comparative analysis between three time series approximation strategies, combined with three classification algorithms, in an experiment carried out with real data from the electrophysiology sensing of common bean plants is presented below.

3. Materials and Methods

In this section, the characteristics of the data set used are presented, as well as a brief detailing about the time series approximation techniques implemented.

3.1. Experiment dataset

In our case study, we used the data collected in an experiment at the Plant Cognition and Electrophysiology Laboratory (LACEV), at the Federal University of Pelotas. The experiment detailed in [Toledo 2019] presents an electrophysiological analysis of bean plants, of the species *Phaseolus vulgaris L.*, cultivar BRS Expedito, through a benchtop experiment. In an environment with controlled conditions (such as humidity, temperature, and lighting), bean plants were cultivated and information on their electrical activity was collected through sensor electrodes connected to the plant stems. In order to analyze the variations in electrical activity, stimuli, such as controlled changes on soil humidity and salinity, were applied and the electromas (set of electrical activities in plants) were analyzed before and after the stimuli inductions.

The collected data were stored in time series format, with an average sampling rate of 450,000 samples every 2 hours of measurement. Measurements were taken for 2 hours before application of a stimulus and 2 hours after application. Thus, for each plant/sensor, two time series of 450,000 points were generated, one referring to the state of the plant before application and one after. For classification purposes, a value of 0 was assigned to the time series collected before the stimulus, to indicate absence of stress, while a value of 1 was assigned to the time series collected after the application of a stimulus, representing stress. For this study, measurements taken before and after the application of stimuli in 37 sensors were made available, with a total of 74 files in text format (*.txt*), containing a time series per file, of 450,000 lines, on average. Each line containing the value of a floating-point number. Each file, with these settings, has around 6 MB, with the total size of the data set available for the experiment being 441.83 MB.

3.2. Time Series Approximation

Given the large volume of information, for the training of machine learning models, even in the work of [Toledo 2019] it was necessary to reduce the data. In the aforementioned

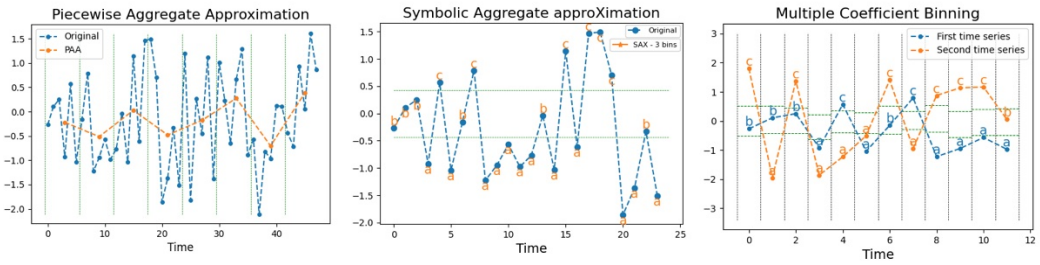


Figure 1. Illustration of approximations with *PAA*, *SAX* and *MCB*.

study, the authors propose to reduce the dimensionality of a data set by segmenting all time series into intervals of fixed size of 30,000 points. Each interval is represented by a tuple containing three information: the maximum and the minimum value annotated in the interval, and the average of all values. In this experiment, for performance analysis purposes, the same interval size was used.

In order to explore other possibilities for reducing the volume of data, in the present work, some alternatives for the representation of time series were investigated. The objective here is to analyze the impact of these alternatives on model training and compare them with the technique used initially, in [Toledo 2019]. Thus, three approximation algorithms were chosen and implemented, as detailed in the following.

- ***Piecewise Aggregate Approximation - PAA***: it consists in the creation of an alternative representation of a time series, from the series segmentation into intervals and in the replacement of this interval by its average value [Keogh et al. 2002]. The objective of the technique is to decrease the number of points and reduce the noise of a time series, preserving the trend. The implementation of this technique can be parameterized by configuring the size of the interval to be reduced or the number of points desired in the output. For this experiment, with an interval of 30,000 points, the time series of 450,000 points were then reduced to a time series of 15 points.
- ***Symbolic Aggregate approXimation - SAX***: uses a continuous data categorization, aiming to reduce noise and capture the trend of a time series [Lin et al. 2007]. This method independently transforms each time series (in this experiment, a sequence of floating-point numbers) into a sequence of symbols. The parameters of this technique are the number of *bins*, the strategy to determine the width of the bins and the alphabet of characters. The *bins* are the number of classes into which data will be categorized. Graphically, they correspond to the horizontal segments, as shown in the figure. The strategy can be: *uniform* (all bins of the same width); *quantile*, (all bins with the same number of points); or *normal* (bins defined according to a standard normal distribution). In this experiment, the standard parameters of the technique were used, with 4 bins, the characters *{1, 2, 3, 4}* and the distribution strategy *quantile*.
- ***Multiple Coefficient Binning - MCB***: it also uses a continuous data categorization, transforming it into strings, in order to reduce noise and capture the trend of a time series [Schäfer and Högqvist 2012]. However, this method takes into account only the values within the range to be categorized, storing each time period inde-

pendently, unlike *SAX*, which classifies the ranges according to the distribution of the entire series. This algorithm, like the previous one, also receives as input the number of *bins*, the strategy to define the width of the *bins* and the alphabet. In this experiment, the same parameters of the *SAX* technique were used.

Figure 1 graphically illustrates the application of approximation techniques: on the left, an example of the application of the PAA technique, where the original time series (blue) is reduced to a time series formed by the average value of each interval (orange); in the center, an example of the *SAX* algorithm, in which bins (horizontal lines) determine the categorization of time series values (blue) for new symbols (orange); and on the right, an example of the *MCB* technique, which determines the bins according to the values of each interval, categorizing the two presented time series (blue and orange).

Since the objectives of the above algorithms are different, the first seeks to reduce the sampling rate, while the others aim at categorizing the data, in this experiment these techniques were tested in a combined way. In order to improve the data reduction, the *SAX* and *MCB* techniques were used as a complement to *PAA*, that is, the input of these techniques was the set of time series already reduced with the previous technique.

3.2.1. Algorithms Implementation

The implementation of the time series approximation algorithms, detailed above, was done in the *Python* language, in the *Google Colab* tool environment. The *PYTS* [Faouzi and Janati 2020] library was used, dedicated to time series classification, which provides pre-processing and utility tools, as well as implementations of several time series classification algorithms. Data from the 74 files, detailed above, were used as input, and a single file was obtained as output, for each algorithm, with the information reduced. The techniques were applied sequentially, and the *PAA* algorithm was the first to be executed, for later the *SAX* and *MCB* techniques use the output of the first, with input for their executions.

Model training was also implemented in the same programming environment, adopting the open-source library *Scikit-learn* [Pedregosa et al. 2011] to have access to tools for predictive data analysis. The objective of the training stage was to develop models capable of identifying water stress patterns (0 without stress and 1 with stress), from the data set, in order to enable the classification of future time series (new time series). The machine learning algorithms implemented were: *K-Nearest Neighbors (KNN)*, which uses the Euclidean distance between points to classify data according to their K nearest neighbors; *Support Vector Machine (SVM)*, an algorithm that searches, through the distances between points of the same set, a linear classifier that better classifies these data; and *Artificial Neural Network (ANN)*, a mathematical model inspired by neural structure. All chosen algorithms are traditional and widely used in artificial intelligence applications. For model training and performance verification, the data sets were separated into training (80%) and test (20%) sets. For the *KNN* algorithm, the default parameters were used, with *n=3*. The *SVM* algorithm was also implemented with the default parameters. The neural network, *ANN*, was implemented with two layers, 3 and 5 nodes, with the activation function *relu*. It should be noted that, once the parameters for an algorithm were defined, they were used for training and classification in all data sets.

Approximation	Compressed Size (KB)	Compression Ratio	Space Saving
[Toledo 2019]	36,78	12013/1	99,991675%
PAA	19,90	22201/1	99,995495%
PAA+SAX	2,42	182201/1	99,999451%
PAA+MCB	2,42	182201/1	99,999451%

Table 1. Performance of time series approximations to reduce data volume.

4. Results and Discussion

The analysis on the data volume aspect is presented through the metrics *Compression Ratio* and *Space Savings* [Lelewer and Hirschberg 1987]. Both were chosen because they address the issue of data size, so important for agricultural applications, as they are usually run in environments with low availability of storage resources. As for the evaluation of the machine learning models, the metric *Accuracy* was used, based on the measurements of the hits of the classifications carried out by the models. Table 1 presents the results.

4.1. Data Size Reduction

The main objective, when implementing the time series reduction algorithm, is to reduce the volume of data and, consequently, the requirements involved in storing information. *Compression Ratio* is a metric that makes it possible to analyze this aspect and is obtained by dividing the original file size by the compressed file size. Complementarily, the *Space Saving* metric, indicates, in percentage, the amount of space reduced by the compression. It should be noted that the values obtained in the experiment are fixed for the data sizes and for the length of the interval used. The table presents the results found for compression ratios and saved space, in relation to the original data size, 441.83 MB, for the techniques implemented here, compared to the technique used in the [Toledo 2019] study.

Regarding the reduction in data volume, it is possible to observe that the implemented techniques achieve an excellent performance, compared to the original data volume and also to the interval mathematics technique. The *PAA* algorithm allows a great reduction in the sampling rate and, when applied in combination with the *SAX* or *MCB*, due to the categorization of the data, it achieves a very high Space Saving rate, of almost 100% as shown in the table. Although the improvements, compared to [Toledo 2019], may seem small in percentage levels, these results imply a significant reduction of space in absolute values (compressed size): more than 90%.

With such a reduction, there is a great potential for the application of these alternatives in edge devices. The level of reduction provided by the algorithms means that with the same space used to store 4 hours of measurement, it would be possible to save information uninterruptedly of approximately 10 years, using only *PAA*, and 85 years, using *PAA* in combination with *SAX* or *MCB*. This high storage gain makes the use of these techniques highly viable in agricultural applications, in scenarios of scarcity of online technological resources and/or limited hardware, such as devices like the *Arduino Uno*, a simple and cheap model, but with limited memory.

4.2. Time Series Classification

Decreasing the volume of data can be achieved in several ways, such as decreasing the sampling rate. However, in the scope of electrophysiology of bean plants, as detailed by

Approximation	KNN	SVM	ANN
[Toledo 2019]	***66,35%***	52,51%	52,00%
PAA	46,66%	47,82%	46,66%
PAA+SAX	60,86%	***95,65%***	***53,33%***
PAA+MCB	39,13%	60,86%	51,02%

Table 2. Performance of time series classification algorithms.

[Toledo 2019], electrical variations can occur at very small intervals (milliseconds), so a high sampling rate is important. Therefore, a reduction strategy that does not impact data quality is necessary. Thus, the main scientific contribution brought by this work is precisely to demonstrate possibilities of data reduction, without compromising the quality of the stored data, that is, without making the use of data unfeasible.

Likewise, compared to the interval mathematics technique, the reduction techniques investigated in this work present significant improvements in terms of performance in model training. Table 2 presents the results obtained, in relation to the accuracy, regarding the performance of the implemented time series classification algorithms. The best accuracies for each algorithm are highlighted. For the *KNN* and *SVM* algorithms, the value of a single execution was considered, since both consider the distances between the points for classification, therefore, there is no variation between one execution and another. As for the training of the *ANN* algorithm, for each data set the algorithm was executed (training and testing) 10 epochs, considering the average accuracy for analysis purposes. From these results, it is observed that, while for the approximation strategies *PAA* and *PAA+MCB* present results, on average, inferior to the interval mathematics, the alternative *PAA+SAX* presented very superior results, especially when combined with the *SVM* algorithm.

5. Conclusion and Future Work

This article presented an experimental study on time series reduction and its impact on time series classification tasks. From real data, collected by [Toledo 2019], a first step was carried out to reduce the volume of data, where three time series approximation techniques were implemented: *Piecewise Aggregate Approximation - PAA*, *Symbolic Aggregate approXimation - SAX* and *Multiple Coefficient Binning - MCB*. The reduced time series were then used as input for the training of artificial intelligence models. For training, the algorithms *KNN, SVM* and *ANN* were used. The results of the experiment were analyzed from the *Compression Rate* and *Space Saving* metrics, for the compression techniques, and the *Accuracy*, for the classification models.

The main result of the work is the high compression rate achieved with the implemented techniques, validated together with time series classification algorithms with high accuracy rate. It has been shown that some implemented techniques outperform interval mathematics, originally used in [Toledo 2019]. The best result was identified in the compression technique that combines the *PAA+SAX* methods, which allows a space-saving of 99.999451%, used in conjunction with the *SVM* classification algorithm, which achieved an accuracy of 95%. This experiment explored simple mathematical alternatives for the reduction of time series, in order to identify possibilities for use in edge devices, with low need for computational requirements and the results found, therefore, validate the use of

these techniques for agricultural solutions.

Given the experimental nature of the study, the initial results obtained are considered successful and indicate interesting alternatives to be explored. As future works, it is intended to extend the analysis to other data reduction techniques. Also, explore the use of artificial intelligence techniques in the data compression stage, as well as the use of additional data in time series classification tasks.

Acknowledgement

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

Al-Qurabat, A. K. M. and Idrees, A. K. (2019). Two level data aggregation protocol for prolonging lifetime of periodic sensor networks. *Wireless Networks*, 25(6):3623–3641.

Al-Qurabat, A. K. M. and Kadhum Idrees, A. (2020). Data gathering and aggregation with selective transmission technique to optimize the lifetime of internet of things networks. *International Journal of Communication Systems*, 33(11):e4408.

Blalock, D., Madden, S., and Guttag, J. (2018). Sprintz: Time series compression for the internet of things. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–23.

Faouzi, J. and Janati, H. (2020). pyts: A python package for time series classification. *Journal of Machine Learning Research*, 21(46):1–6.

Keogh, E., Chakrabarti, K., Pazzani, M., and Mehrotra, S. (2002). Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3.

Krawczak, M. and Szkatuła, G. (2014). An approach to dimensionality reduction in time series. *Information Sciences*, 260:15–36.

Lelewer, D. A. and Hirschberg, D. S. (1987). Data compression. *ACM Comput. Surv.*, 19(3):261–296.

Lin, J., Keogh, E., Wei, L., and Lonardi, S. (2007). Experiencing sax: A novel symbolic representation of time series. *Data Min. Knowl. Discov.*, 15:107–144.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Schäfer, P. and Högqvist, M. (2012). Sfa: A symbolic fourier approximation and index for similarity search in high dimensional datasets. pages 516 – 527.

Tobore, I., Kandwal, A., Li, J., Yan, Y., Omisore, O. M., Enitan, E., Sinan, L., Yuhang, L., Wang, L., and Nie, Z. (2020). Towards adequate prediction of prediabetes using spatiotemporal ecg and eeg feature analysis and weight-based multi-model approach. *Knowledge-Based Systems*, 209:106464.

Toledo, G. R. A. (2019). *Caracterização eletrofisiológica do feijão (*Phaseolus vulgaris *L.) cv. BRS-Expedito sob diferentes disponibilidades hídricas*. PhD thesis, UFPel, Pelotas.

Xin, B., Peng, W., Kwon, Y., and Liu, Y. (2019). Modeling, discretization, and hyperchaos detection of conformable derivative approach to a financial system with market confidence and ethics risk. *Advances in Difference Equations*, 2019(1):1–14.

Zhang, X., Cao, Z., and Dong, W. (2020). Overview of edge computing in the agricultural internet of things: Key technologies, applications, challenges. *IEEE Access*, 8:141748–141761.



Estimating and tuning adaptive action plans for the control of smart interconnected poultry houses

Darlan F Klotz[1], **Dalcimar Casanova**[1], **Marcelo Teixeira**[1]

[1]Graduate Program in Electrical Engineering
Federal University of Technology Paraná
Pato Branco – PR – Brazil

darlan.klotz@ifc.edu.br, {dalcimar,mtex}@utfpr.edu.br

Abstract. *In poultry farming, the systematic choice, update, and implementation of periodic (t) action plans define the feed conversion rate ($FCR^{\langle t \rangle}$), which is an acceptable measure for successful production. Appropriate action plans provide tailored resources for broilers, allowing them to grow within the so-called comfort zone, without waste or lack of resources. Although the implementation of an action plan is automatic, its configuration depends on a specialist, which tends to be inefficient and result in variable $FCR^{\langle t \rangle}$. In this project, the specialist's perception is reproduced, to some extent, by computational intelligence. By combining deep learning and genetic algorithm techniques, we show how action plans can adapt their performance over the time, based on previous well succeeded plans. We also implement a network infrastructure to replicate our method over distributed poultry houses, for their smart, interconnected, and adaptive control. A supervision system is provided as interface to users. Experiments using real data suggest an improvement of 5% on the performance of the most productive specialist, staying close to the optimal $FCR^{\langle t \rangle}$.*

1. Introduction

Poultry farming is the agro-industrial sector focused on producing broiler meat for consumption. Worldwide, the activity has attracted attention due to its economic and social role, aligned with the continuously increasing world demand for broilers. Only in 2019 the USA was responsible for producing 19.941 million tons of meat, followed by Brazil and China with 13.245 and 13.750 million tons, respectively (ABPA 2021).

Technically, broilers raise within *poultry houses* (PHs), that are automatically controlled by electronic devices. The daily recipe, prescribing the parameters for the automatic controller, is called an *action plan* (Acres 2020). The quality of an action plan reflects the general welfare for broilers, allowing them to grow up under appropriate conditions of feeding, housing, and health. In this paper, in particular, an action plan is

exploited in terms of its environment-related variables. We seek for combined indoor climate conditions that allow broilers to grow up under *thermal comfort*, achieved within the so-called *comfort zone* (Botreau et al. 2007). The comfort zone directly affects productivity, which is acceptably measured by periodic ($t \in \mathbb{N}$) *feed conversion rates* ($FCR^{\langle t \rangle}$) estimations. The more accurate the action plan, the lower the accumulated $FCR^{\langle t \rangle}$, and the greater it tends to be the production.

Advances on improving $FCR^{\langle t \rangle}$ focus on health, nutrition, genetics, thermal comfort (Henriksen et al. 2016; Ribeiro et al. 2019), etc. A remaining barrier is, however, the lack of expert systems to help humans with the choice, update, and implementation of daily action plans. This leads specialists to use their own experience to set up action plans, which tends to be complex, inefficient, and prone to errors, besides to cause conflicting opinions. As a consequence of unsystematic interventions, each PH may return a distinct $FCR^{\langle t \rangle}$, which makes unclear the overall performance estimation.

The literature provides alternatives to estimate intelligent action plans. For example, (Ribeiro et al. 2019) uses a bio-inspired model based on *Artificial Neural Networks* (ANNs) to suggest action plans according to climatic variations. However, this approach is not integrated with the process controller, so that it cannot reconfigure its actions dynamically according to the ANN suggestion. Furthermore, (Ribeiro et al. 2019) does not consider the dynamic adaptation of its estimations over the time, which is an important feature when scenarios of production are variable.

In this project, the specialist perception is reproduced, to some extent, by computational intelligence. Differently from (Ribeiro et al. 2019), we use a more suitable memory-aware model based on *Long short-term memory* (LSTM) networks, instead of ANNs. We also show how estimated action plans can be combined across algorithms that artificially emulate genetic mutations. This results in improved action plans that are capable of adapting themselves over the time based on previously successful action plans. By replicating our method over a set of PHs, we solidify a new foundation to support smart, cooperative, distributed, and interconnected poultry farming.

Experiments conducted using real data indicate that our approach estimates an $FCR^{\langle 40 \rangle}$ of 1.5610 by the end of a flock. In comparison, the $FCR^{\langle 40 \rangle}$ resulting from the progressive application of empiric action plans, chosen by the specialist who had the best performance among all others in the evaluated group, under the same setup, was of 1.640, therefore 0.079 or 5% worse than ours. Our method also shows proximity with the best performance achievable by *Ross* broilers strain, which is $FCR^{\langle 40 \rangle} = 1.558$ (Ross 2019).

2. An intelligent solution for poultry management

A usual measure to quantify the efficiency in transforming feed into animal protein, is based on periodic samples of *Feed Conversion Rate*, formalised as:

$$FCR^{\langle t \rangle} = \frac{FC^{\langle t \rangle}}{LB^{\langle t \rangle} \cdot MW^{\langle t \rangle}}\,, \tag{1}$$

where t is an integer associated with the day of observation; $FC^{\langle t \rangle}$ is the amount of *feed consumption* until day t; $LB^{\langle t \rangle}$ is the number of *living birds* at day t; and $MW^{\langle t \rangle}$ is the *mean weigh* at day t.

An effective action plan is achieved when the $FCR^{\langle t \rangle}$ is improved, i.e., when it approximates to zero. The more accurate the action plan, the lower the conversion rate, and the greater the production tends to be. However, obtaining a satisfactory $FCR^{\langle t \rangle}$, at the end of a flock (usually $t = 40$), requires enormous ability from the specialist to conduct the daily management. It depends, basically, on how the daily parameters are informed to the controller so that, in combination, they could result in a reasonable accumulated $FCR^{\langle 40 \rangle}$ for the entire flock.

In general, poultry farmers handle more than one PH within a shared infrastructure, called a *poultry condominium* (PC), shown in Fig. 1 and formalised as a finite set $\mathrm{Cond} = \{PH_1, \cdots, PH_n\}$ of poultry houses PH_i, for $1 \leq i \leq n$.

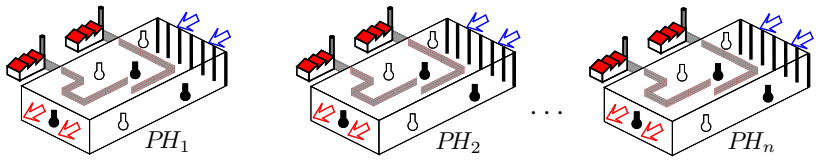


Figure 1. Structure poultry house condominium $\mathrm{Cond} = \{PH_1, \cdots, PH_n\}$.

Each PH has automated feed and drinking systems working continuously from 6 AM until 10 PM. It also includes electronic devices to inform the daily *feed consumption* ($FC^{\langle t \rangle}$), and the *mean weight* ($MW^{\langle t \rangle}$) captured by automatic scales. Each flock takes about 40 days to be produced ($t = 1, \cdots, 40$), and the final weight (i.e, $MW^{\langle 40 \rangle}$) is expected to be of approximately $2.8\ Kg$ per bird. Temperature ($Min(T)^{\langle t \rangle}$, $Avg(T)^{\langle t \rangle}$, and $Max(T)^{\langle t \rangle}$), and relative humidity ($Min(H)^{\langle t \rangle}$, $Avg(H)^{\langle t \rangle}$, and $Max(H)^{\langle t \rangle}$), inside the PH are collect by sensors and stored into a database, which is part of the control system infrastructure. The specialist also records the *mortality* on the day t ($M^{\langle t \rangle}$), which alternatively can be exposed as the number of *living broilers* by:

$$LB^{\langle t \rangle} = |F| - \sum_{t=1}^{p} M^{\langle t \rangle}\,, \tag{2}$$

i.e., as the difference between the *initial amount of birds* in a flock ($|F|$), and the accumulated mortality recorded by the specialist during a period p.

2.1. Dataset acquisition

For the purpose of this paper, the real dataset was obtained by observing a PC composed of three PHs, which were used for collecting daily data sampling. The three PHs contained only male, *Ross* strain broilers.The data acquisition was conducted between August 2019 and March 2020, alternating therefore different seasonal features. A total of 12 flocks were observed (i.e., 12 samples), the same as used in related literature (Johansen et al. 2017), which has shown to be sufficient for the intended analysis.

Table 1 shows the variables collected and used in the experiments that follow. Only input and output variables have been formalised so far, while *normalised variables* will be introduced in section Section 2.2. *Input variables* correspond to the action plan for the day t, while *output variables* are those influenced by this action plan. The selection algorithm *Greedy Stepwise* (Gevrey et al. 2003) has been applied to confirm the relevance of the input variables over the output.

Input variables	Unit	**Notation**
day	$t = 1, \cdots, 40$	t
minimum temperature	[°C]	$Min(T)^{\langle t \rangle}$
mean temperature	[°C]	$Avg(T)^{\langle t \rangle}$
maximum temperature	[°C]	$Max(T)^{\langle t \rangle}$
minimum humidity	[%]	$Min(H)^{\langle t \rangle}$
mean humidity	[%]	$Avg(H)^{\langle t \rangle}$
maximum humidity	[%]	$Max(H)^{\langle t \rangle}$
Output variables		**Notation**
mean daily weigh	$[g]$	$MW^{\langle t \rangle}$
daily feed consumption	$[Kg]$	$FC^{\langle t \rangle}$
number of living birds	[birds]	$LB^{\langle t \rangle}$
daily mortality	[birds]	$M^{\langle t \rangle}$
Output variables normalized		**Notation**
daily feed consumption per bird	$[Kg/bird]$	$FC\mathrm{b}^{\langle t \rangle}$
number of living birds per area	$[bird/m^2]$	$LB\mathrm{a}^{\langle t \rangle}$
daily mortality per area	$[bird/m^2]$	$M\mathrm{a}^{\langle t \rangle}$

Table 1. Variables to be considered in our experiments.

A daily measurement was collected for each of the 7 input variables in Table 1, totalling 280 records (7×40 days of the flock) in one sample. For the 3 output variables, the daily measurement for $MW^{\langle t \rangle}$, $FC^{\langle t \rangle}$, and $LB^{\langle t \rangle}$ corresponds to 120 records (3×40), also for one sample. Therefore, by observing 3 PHs, each one providing 4 samples, we obtained the total of 12 samples, corresponding to an input data matrix of 12×280, while the corresponding output data form a matrix sized 12×120.

2.2. Data normalisation

The 3 observed PHs occupy two different areas. A_{PH_1} is $150\ m$ long by $16\ m$ wide, totalling $2400\ m^2$, and it can host about 34800 birds. A_{PH_2} is $150\ m$ long by $12\ m$ wide, with a floor area of $1800\ m^2$ that accommodates 26500 birds. Thus, $M^{\langle t \rangle}$ and $LB^{\langle t \rangle}$ can be normalised by the area A_{PH_i}, such that:

$$M\mathrm{a}^{\langle t \rangle} = \frac{M^{\langle t \rangle}}{A_{PH_i}}; \quad \text{and} \quad LB\mathrm{a}^{\langle t \rangle} = \frac{LB^{\langle t \rangle}}{A_{PH_i}}. \tag{3}$$

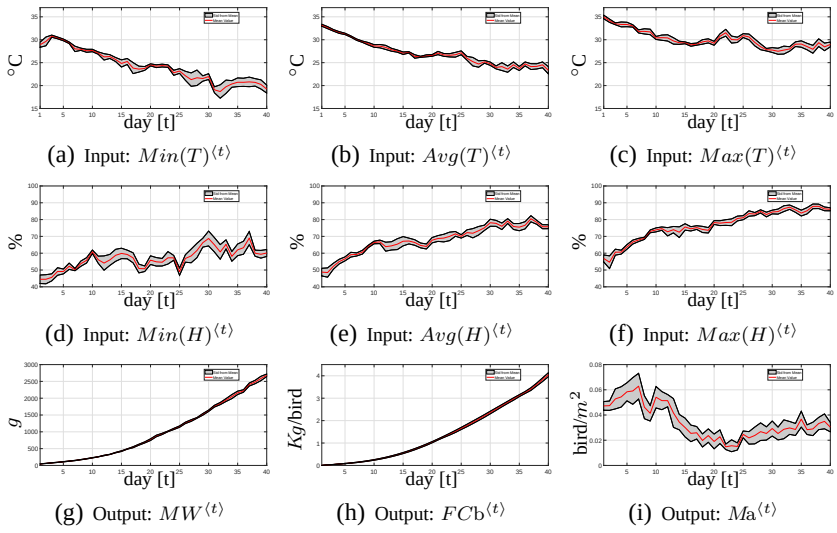
Also $FC^{\langle t \rangle}$ can be normalised by $LB^{\langle t \rangle}$, given the *feed consumption per bird*:

$$FC\mathrm{b}^{\langle t \rangle} = \frac{FC^{\langle t \rangle}}{LB^{\langle t \rangle}}. \tag{4}$$

In this way, $FCR^{\langle t \rangle}$ can result from Eq.1, additionally pondering: (i) the number of living birds per area A_{PH_i} ($LB\mathrm{a}^{\langle t \rangle}$); and (ii) the daily feed consumption by the number of living birds ($FC\mathrm{b}^{\langle t \rangle}$). Eq.5 formalises this idea:

$$FCR^{\langle t \rangle} = 1000\, \frac{FC\mathrm{b}^{\langle t \rangle}}{LB\mathrm{a}^{\langle t \rangle}\ MW^{\langle t \rangle}}\ LB\mathrm{a}^{\langle t \rangle}, \tag{5}$$

where 1000 compensates the weight conversion from Kg to g. As *Mean* (μ) and *standard deviation* (σ) are essential measures, they are detailed for the 12 flocks in Fig. 2.



(a) Input: $Min(T)^{\langle t \rangle}$ (b) Input: $Avg(T)^{\langle t \rangle}$ (c) Input: $Max(T)^{\langle t \rangle}$

(d) Input: $Min(H)^{\langle t \rangle}$ (e) Input: $Avg(H)^{\langle t \rangle}$ (f) Input: $Max(H)^{\langle t \rangle}$

(g) Output: $MW^{\langle t \rangle}$ (h) Output: $FCb^{\langle t \rangle}$ (i) Output: $Ma^{\langle t \rangle}$

Figure 2. Statistical view of the collected variables.

2.3. Proposed LSTM Neural Network

Now, we feed the LSTM neural network shown in Fig. 3 with: the day under analysis (t); the measured temperature ($Min(T)^{\langle t \rangle}$, $Avg(T)^{\langle t \rangle}$, and $Max(T)^{\langle t \rangle}$) and humidity ($Min(H)^{\langle t \rangle}$, $Avg(H)^{\langle t \rangle}$, and $Max(H)^{\langle t \rangle}$); and the previous day mean weigh ($MW^{\langle t-1 \rangle}$), feed consumption per bird ($FCb^{\langle t-1 \rangle}$), and number of living birds per area ($LBa^{\langle t-1 \rangle}$).

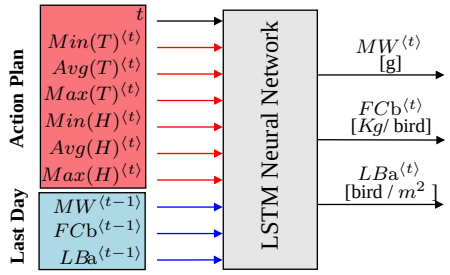


Figure 3. Proposed LSTM network architecture.

From now forward, inputs and outputs are referred to as vectors: $\mathbf{x}^{\langle t \rangle} = [t, Min(T)^{\langle t \rangle}, \cdots, Max(H)^{\langle t \rangle}]$; $\mathbf{h}^{\langle t-1 \rangle} = [MW^{\langle t-1 \rangle}, FCb^{\langle t-1 \rangle}, LBa^{\langle t-1 \rangle}]$; $\mathbf{h}^{\langle t \rangle} = [MW^{\langle t \rangle}, FCb^{\langle t \rangle}, LBa^{\langle t \rangle}]$. Vector $\mathbf{x}^{\langle t \rangle}$ carries the input action plan for day t; $\mathbf{h}^{\langle t-1 \rangle}$ is the input vector with the previous days outputs; and $\mathbf{h}^{\langle t \rangle}$ is the output vector for day t. By concatenating $\mathbf{x}^{\langle t \rangle}$ and $\mathbf{h}^{\langle t-1 \rangle}$ (denoted $\mathbf{x}^{\langle t \rangle} \frown \mathbf{h}^{\langle t-1 \rangle}$), and by applying it as the neural network input, we obtain a prediction for $\mathbf{h}^{\langle t \rangle}$. The data pre-processing step and associated hyperparameters settings for training the LSTM model can be found in (Klotz et al. 2020).

2.4. LSTM Implementation

The proposed LSTM-based architecture can be implemented as in Algorithm 1, and then it is taken as a fitness function of a GA in Algorithm 2.

After 1400 iterations the proposed algorithm starts to stabilise. This leads search to stop by the timeout criterion with no changes in the fitness function. The $FCR^{\langle 40 \rangle}$

Algorithm 1: LSTM implementation.

Procedure: LSTMCalc($wm, \mathbf{max}, \mathbf{min}, \mathbf{te}$);
Input: wm, $\mathbf{max}$, $\mathbf{min}$, $\mathbf{te}$;
Output: $\mathbf{h}^{\langle w_S \rangle}$;
begin
  $\mathbf{te}^{'} = \Lambda(\mathbf{te},\ \mathbf{max},\ \mathbf{min})$;
  for *t=1 to 1* **do**
    $\mathbf{x'}^{\langle t \rangle} = \mathbf{te}^{'}[1, 7]^T$;
    $\mathbf{h'}^{\langle t-1 \rangle} = \mathbf{te}^{'}[8, 10]^T$;
    $\mathbf{h'}^{\langle t \rangle} = \phi(\mathbf{x'}^{\langle t \rangle}, \mathbf{h'}^{\langle t-1 \rangle}, wm)$;
    Update the LSTM *state*;
  end
  for *t=2 to* w_S **do**
    $\mathbf{x'}^{\langle t \rangle} = \mathbf{te}^{'}[7t-3, 7t+3]^T$;
    $\mathbf{h'}^{\langle t \rangle} = \phi(\mathbf{x'}^{\langle t \rangle}, \mathbf{h'}^{\langle t-1 \rangle}, wm)$;
    Update the LSTM *state*;
  end
  $return\ (\mathbf{h}^{\langle w_S \rangle} = \Lambda^{-1}(\Lambda(\mathbf{h'}^{\langle w_S \rangle}, \mathbf{max}, \mathbf{min})))$;
end

Algorithm 2: GA and LSTM Network Integration

Input: $wm_w, \mathbf{max}_w, \mathbf{min}_w,\ w = 1, \cdots, 6$;
Output: **FAP**, **i_c**;
begin
  $G_{Sw=6} = 38$;
  $G_{Sw=5,4,\cdots,1} = 52$;
  set $P_S = 200$, $\beta = 0.6$;
  for w = *6 to 1* **do**
    set $\mathbf{A} = [2\ G_{Sw}$ x $G_{Sw}]$;
    set $\mathbf{x} = [G_{Sw}$ x $1]$;
    set $\mathbf{b} = [2\ G_{Sw}$ x $1]$;
    set $\mathbf{P} = [P_S$ x $G_{Sw}]$;
    $\mathbf{P} = P_I(\mathbf{A}, \mathbf{x}, \mathbf{b})$;
    while *Some stopping criterion is not met* **do**
      for *i=1 to* P_S **do**
        $\mathbf{h}_w^{\langle w_S \rangle} = \text{LSTMCalc}(wm_w, \mathbf{max}_w, \mathbf{min}_w, \mathbf{P}(i,\ :)^T$;
        if w == *6* **then**
          $ff = FCR_w^{\langle w_S \rangle}$;
        else
          $ff = \mu \left| \mathbf{h}_w^{\langle w_S \rangle} - \mathbf{h}_{w+1}^{\langle 1-1 \rangle} \right|$
        end
      end
      for *j=1 to* P_S **do**
        $\mathbf{p_1}, \mathbf{p_2} = SU(ff)$;
        $\mathbf{child} = HC(\mathbf{p_1},\ \mathbf{p_2}, \beta\ )$;
        $\mathbf{G} = AM(\mathbf{child})$;
        $\mathbf{P}(j,\ :) = \mathbf{G}$;
      end
    end
    $\mathbf{h}_w^{\langle 1-1 \rangle} = \mathbf{G}(8 : 10)$;
    $\mathbf{FAP}(w,\ :) = \mathbf{G}$;
  end
  i_c $= \mathbf{h}_w^{\langle 1-1 \rangle}$;
  $return\ (\mathbf{i_c},\ \mathbf{FAP})$;
end

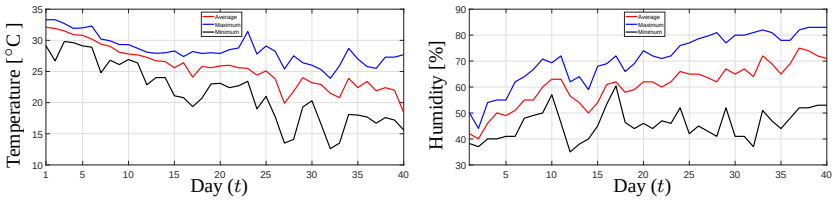
upon stabilisation is 1.5602. However, this value is exclusively linked to wm_6, i.e., as the Algorithm 2 searches from wm_6 to wm_1, the value of 1.5602 is guaranteed if and only if the progressive error between the difference of $\mathbf{h}_w^{\langle w_S \rangle}$ (estimated by the wm_w model) and

$\mathbf{h}_{w+1}^{\langle 1-1 \rangle}$ (of the model wm_{w+1}) is null (see Algorithm 2, line 19). The following Table 2 shows step-by-step these errors.

w	$\lvert \mathbf{h}_{w}^{\langle w_S \rangle} - \mathbf{h}_{w+1}^{\langle 1-1 \rangle} \rvert$					
	Error			**Relative Error (%)**		
	$MW^{\langle t \rangle}$	$FC\mathrm{b}^{\langle t \rangle}$	$LB\mathrm{a}^{\langle t \rangle}$	$MW^{\langle t \rangle}$	$FC\mathrm{b}^{\langle t \rangle}$	$LB\mathrm{a}^{\langle t \rangle}$
5	17.4840	0.0066	0.1448	0.8335	0.2230	1.0269
4	32.7932	0.0011	0.0092	2.3440	0.1960	0.0639
3	15.7567	0.0070	0.1289	1.7667	0.7056	0.9059
2	0.0000	0.0080	0.0765	0	2.2396	0.5358
1	1.6142	0.0015	0.1461	1.0338	1.4681	1.0176

Table 2. Errors and relative erros comparison.

We present in Fig. 4 the action plan $\mathbf{FAP}$ returned by the Algorithm 2 after it iterates recurrently through the 6 weeks, leading to $FCR^{\langle E \rangle} = 1.5602$. The abscissa axis shows the input t of the LSTM network (see Fig.3). The ordinate axis shows the inputs $Min(T)^{\langle t \rangle}$, $Avg(T)^{\langle t \rangle}$, and $Max(T)^{\langle t \rangle}$ (Fig. 4(a)); and the inputs $Min(H)^{\langle t \rangle}$, $Avg(H)^{\langle t \rangle}$, and $Max(H)^{\langle t \rangle}$ (Fig. 4(b)) that were discovered by the GA to reach $FCR^{\langle E \rangle} = 1.5602$.



(a) Temperature (b) Humidity

Figure 4. Action plan suggested by our approach.

Now, we apply the Algorithm 1 progressively from $wm_w = 1$ to $wm_w = 6$, having $\mathbf{FAP}$ and $\mathbf{i_c}$ as input for each respective model. Upon termination of Algorithm 1, we capture the 3 outputs from the last day of wm_6 (i.e., $\mathbf{h}_{w=6}^{\langle w_S \rangle}$) and apply Eq.5 to obtain $FCR^{\langle R \rangle}$. It is expected a very close relationship between $FCR^{\langle R \rangle}$ and $FCR^{\langle E \rangle}$, considering the statistical correctness of Table 2. For the particular dataset, $FCR^{\langle R \rangle} = 1.5610$ was found for the 40-day flock, therefore very close to $FCR^{\langle E \rangle} = 1.5602$.

2.4.1. Quantitative analysis

In order to assess how competitive our estimations are, we compare the model performance ($FCR^{\langle R \rangle}$) against: (i) the optimal performance (denote $FCR^{\langle O \rangle}$) theoretically expected for the *Ross* broiler strain; (ii) the performance resulting from real managements applied by human specialists; (iii) the performance resulting from 1000 other candidate, artificially created, action plans.

The optimal performance for *Ross* broiler strain after 40 days is $FCR^{\langle O \rangle} = 1.558$ (Ross 2019). Therefore, the $FCR^{\langle R \rangle} = 1.5610$, resulting from our method, suggests an approximation only 0.192% worse than $FCR^{\langle O \rangle}$. The best specialist (denote $FCR^{\langle X \rangle}$), among all others, returned $FCR^{\langle X \rangle} = 1.640$, therefore 0.0790 (or 5%) worse than $FCR^{\langle R \rangle} = 1.5610$. Although the difference seems to be minor, our model allows to

save $0.0790\ Kg$ (or $79\ g$) of feed per bird, for each $1\ Kg$ of produced meat. For a slaughtering weight of $2.8\ Kg$, each broiler requires $221.20\ g$ less feed to be produced. For a PH with about 34000 broilers, this means $7520,8\ Kg$ of feed saving during a single flock.

We further construct 1000 more synthetic action plans to simulate different specialists handling a flock. The best possible performance found was $FCR^{\langle S \rangle} = 1.5614$, therefore still worse than $FCR^{\langle R \rangle} = 1.5610$.

3. Conclusion

In this paper, deep learning and genetic algorithms are combined to estimate action plans for poultry farming. We exploit a production model that interconnects multiple poultry houses and allows us to both, collect distributed data, and propagate optimised action plans throughout a network infrastructure, commanded by a central supervision system.

We show that the specialist perception about the process can be captured and reproduced, to some extent, by using a bio-inspired method, and this may increase productivity. A software solution is also provided and fits with any number of poultry houses, including one, when this is a particular case.

Ongoing researches focus on technology transfer and tests in real scale. Other results are expected from the observation of a more complete set of events related to poultry business, such as *cost*, *profit*, and *market demand*. Environment forecasts and energy consumption are also subjects to be approached in future extensions.

References

ABPA (2021). Brazilian association of animal protein. EMBRAPA.

Acres, A. (2020). Broiler management handbook.

Botreau, R., Bracke, M., Perny, P., Butterworth, A., Capdeville, J., Van Reenen, C., and Veissier, I. (2007). Aggregation of measures to produce an overall assessment of animal welfare. *Animal*, 1(8):1188–1197.

Gevrey, M., Dimopoulos, I., and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3):249–264.

Henriksen, S., Bilde, T., and Riber, A. B. (2016). Effects of post-hatch brooding temperature on broiler behavior, welfare, and growth. *Poultry Science*, 95(10):2235–2243.

Johansen, S. V., Bendtsen, J. D., Martin, R., Mogensen, J., et al. (2017). Data driven broiler weight forecasting using dynamic neural network models. *IFAC-PapersOnLine*, 50(1):5398–5403.

Klotz, D. F., Ribeiro, R., Enembreck, F., Denardin, G., Barbosa, M., Casanova, D., and Teixeira, M. (2020). Estimating action plans for smart poultry houses.

Ribeiro, R., Casanova, D., Teixeira, M., Wirth, A., Gomes, H. M., Borges, A. P., and Enembreck, F. (2019). Generating action plans for poultry management using artificial neural networks. *Computers and Electronics in Agriculture*, 161:131–140.

Ross (2019). 308 and 308 FF performance objectives.



Estruturação de um *Data Warehouse* para repartição territorial de desmatamentos no bioma Amazônia e o Cadastro Ambiental Rural

Renato da Silva de Deus[1] , Mariela Almeida[1], Carlos Alberto de Carvalho[1]
[1]Embrapa Territorial
Av. Soldado Passarinho, 303, CEP 13070-115 - Campinas - SP

carlos-alberto.carvalho@embrapa.br,
{mariela.almeida,renato.deus}@colaborador.embrapa.br

Abstract. *Due to the large amount of spatial data on the Amazon biome and the need to understand the relationships between rural properties and deforestation, this article defines the proposal of a geospatial Data Warehouse structure in order to organizing this spatial information and relating deforestation phenomena in the Amazon Biome with the territorial dimension occupied by rural properties in the Rural Environmental Registry (CAR). This work presents the conceptual methods for sanitizing CAR data used to develop scripts for preparing and loading data into the Data Warehouse, creating an adequate structure for statistical calculations and data presentation.*

Resumo. *Devido a grande quantidade de dados espaciais sobre o bioma Amazônia e a necessidade de entender os relacionamentos entre os imóveis rurais e desmatamentos, este artigo define a proposta de uma estrutura de Data Warehouse geoespacial com vistas a organizar estas informações espaciais e relacionar os fenômenos de desmatamento no Bioma Amazônia com a dimensão territorial ocupada pelos imóveis rurais do Cadastro Ambiental Rural (CAR). Apresenta os métodos conceituais de sanitização dos dados do CAR utilizados para o desenvolvimento dos scripts de preparação e carga dos dados no Data Warehouse, criando uma estrutura adequada para cálculos estatísticos e apresentação de dados.*

1. Introdução

Este artigo está inserido no projeto *"Inteligência Territorial Estratégica aplicada à identificação, quantificação e qualificação de processos de desmatamento no bioma Amazônia e ao apoio e orientação de políticas públicas e privadas na região*

amazônica", parceria entre a Embrapa Territorial e o Centro Gestor e Operacional do Sistema de Proteção da Amazônia voltado para realizar a qualificação numérica, cartográfica, iconográfica e temática dos processos de desmatamento no bioma Amazônia com vistas a apoiar e orientar políticas públicas no bioma com base em inteligência territorial. Este projeto maior nasceu do contexto complexo inerente ao bioma Amazônia, onde diversas dimensões humanas, territoriais, econômicas, agrárias e ambientais se relacionam, da necessidade de proteção, preservação, regeneração e harmonia entre o desenvolvimento dos diversos grupos humanos, sustentabilidade ambiental e econômica e dos processos de desmatamento que ocorrem na região (EMBRAPA, 2019).

Para apreender esse contexto complexo, é fundamental qualificar os processos de desmatamento, pois eles prosseguem combinando tempo e espaço de forma variada, em toda região, com diversos atores e interesses cuja identificação é necessária à compreensão do fenômeno, tanto para orientar nas definições de estratégias e mecanismos de acompanhamento, fiscalização e controle como apoiar a formulação de políticas públicas diferenciadas com base no contexto territorial da região. Diante dessa complexidade de situações, faz-se necessária a organização dessas informações de forma a realizar análises de contexto territorial. Este artigo apresenta uma proposta de estruturação e análise dessas informações sobre o bioma Amazônia.

A organização do artigo se apresenta da seguinte forma: uma seção teórica que aborda as definições e conceitos sobre o Cadastro Ambiental Rural (CAR) e *Data Warehouse* utilizados para a definição dos métodos. Em materiais e métodos são detalhadas, as fontes de dados, áreas de estudos e os métodos para determinar os relacionamentos entre a repartição territorial de interesse e os desmatamentos, bem como os métodos para a construção e estruturação do *Data Warehouse* utilizando um Sistema Gerenciador de Banco de Dados Relacional com extensão Geoespacial. Na seção de resultados são apresentadas demonstrações parciais dos métodos empregados que serviram de base para preparação e carga na estrutura do *Data Warehouse* também exposto nesta seção.

2. Referencial Teórico

O CAR foi criado para integrar as informações ambientais das propriedades e posses rurais, de forma a compor uma base de dados para controle, monitoramento, planejamento ambiental e econômico e combate ao desmatamento (BRASIL, 2012). A base de dados que abriga o CAR é o Sistema de Cadastro Ambiental Rural - SICAR (SICAR, 2021). Ela é única e atualizada através da inserção do cadastro de um imóvel rural no sistema. Os imóveis rurais podem ser classificados por tamanho (pequeno, médio e grande) a depender da sua quantidade de módulos fiscais. Esta classificação garante ao imóvel rural condições diferenciadas (benefícios, tributação etc.) segundo a sua função social.

O INCRA define módulo fiscal como uma forma de medida em hectares definida para cada município. Essa definição leva em consideração a renda e o tipo da exploração

predominante; outros tipos de exploração secundárias, que sejam significativas para o município; e o conceito de propriedade familiar. A classificação dos imóveis rurais fica então da seguinte maneira, de acordo com a lei n° 8.629, de 25 de fevereiro de 1993: pequeno imóvel (≤ 4 módulos fiscais), médio imóvel (> 4 e ≤ 15 módulos fiscais) e grande imóvel (> 15 módulos fiscais) (INCRA, 2020b).

Para a organização das informações, utilizou-se de referencial teórico sobre a construção de *Data Warehouse* (DW), conforme apresentado a seguir.

A Oracle resume o objetivo e algumas características de um *Data Warehouse* da seguinte maneira, "[...] é um tipo de sistema de gerenciamento de dados projetado para ativar e fornecer suporte às atividades de *business intelligence* (BI), especialmente a análise avançada. Os *data warehouses* destinam-se exclusivamente a realizar consultas e análises avançadas e geralmente contêm grandes quantidades de dados históricos. Os dados em um *data warehouse* geralmente são derivados de uma ampla variedade de fontes, como arquivos de *log* de aplicativos e aplicativos de transações."(ORACLE, 2021). O cientista da computação William H. Inmon define *Data Warehouse* como "[...] um depósito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões gerenciais"(ASTERA, 2020).

Existem diversas abordagens para a construção de um *Data Warehouse*, este artigo apresenta a definida por Ralph Kimball onde a relação entre os conceitos de fatos e dimensões se dão da seguinte maneira: Kimball apresenta a tabela fato como: "Uma tabela de fatos contém as medidas numéricas produzidas por um evento de medição operacional no mundo real. Na granularidade mais baixa, uma linha da tabela de fatos corresponde a um evento de medição e vice-versa."(KIMBALL, 2013). Já as dimensões, como cita Astera, "O esquema em estrela é o elemento fundamental do modelo de data warehouse dimensional. Nesse esquema em estrela, uma tabela de fatos é delimitada por várias dimensões. A modelagem dimensional Kimball permite que os usuários construam vários esquemas em estrela para atender a várias necessidades de relatórios. A vantagem do esquema em estrela é que as consultas de pequenas tabelas dimensionais são executadas instantaneamente"(ASTERA, 2020).

Outro processo utilizado para construir um *Data Warehouse* é o ETL (do inglês: Extrair, Transformar e Carregar). A *Spotlessdata*, definiu esse processo como: "O processo de obtenção de dados de seus vários conjuntos de dados e outras fontes para mesclar em um conjunto de dados perfeito em seu data warehouse é conhecido como extrair, transformar, carregar (ETL) e é usado ao atualizar partes de sua plataforma ou transferir de um legado para uma nova plataforma. Os dados são extraídos de várias fontes, transformados em dados de qualidade consistente antes de serem carregados no data warehouse" (SPOTLESSDATA, 2017).

3. Materiais e Métodos

A seguir são apresentados os materiais utilizados, em fontes de dados e área de estudo, e os métodos e procedimentos aplicados sobre esses conjuntos de dados para a definição dos relacionamentos entre o fenômeno de desmatamento e a dimensão territorial dos Imóveis cadastrados no CAR e da estrutura do Data Warehouse. Em linhas gerais, foram utilizados os conceitos pesquisados e expostos acima que mostram que é útil nessa

situação de grande conjunto de dados (desmatamentos, imóveis rurais do CAR, áreas públicas e bioma)

3.1. Fontes de Dados Utilizadas

Os dados vetoriais utilizados vieram das seguintes fontes: 1) Sistema Nacional de Cadastro Ambiental Rural (SICAR); 2) Instituto de Colonização e Reforma Agrária (INCRA); 3) Instituto Brasileiro de Geografia e Estatística (IBGE); 4) Ministério do Meio Ambiente (MMA); e 5) Instituto Nacional de Pesquisa Espaciais (INPE).

Os limites geográficos dos imóveis rurais do CAR foram obtidos a partir de (SICAR, 2021). Por também serem consideradas como imóveis rurais, foram obtidas as unidades de conservação de reservas extrativistas (RESEX) e de desenvolvimento sustentável (RDS) a partir de (MMA, 2020). As informações sobre os assentamentos agrários foram adquiridas no Acervo Fundiário do INCRA (2020a), uma vez que também são considerados como imóveis rurais.

Os limites territoriais foram obtidos na base de dados do IBGE (2020). Para a delimitação do bioma Amazônia utilizou-se a versão disponível em IBGE (2004). Os polígonos de desmatamento foram obtidos junto ao Projeto Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite - PRODES (INPE, 2020). Estes polígonos são gerados a partir de imagens dos satélites e possuem área mínima de 6,25 hectares.

3.2. Área de estudo

O recorte territorial é o Bioma Amazônia. De acordo com IBGE (2019), bioma é uma área de grande extensão geográfica, ocupando 49,5% do território brasileiro, determinada pela vegetação dominante, com condições geoclimáticas semelhantes, onde a fauna e flora vivem de maneira associada. O clima é identificado como quente e úmido, a temperatura média é de 25°C e com chuvas bem distribuídas ao longo do ano. A fitofisionomia predominante é a florestal com vegetação caracterizada como Floresta Ombrófila Densa (Figura 1). A população estimada para o ano de 2017 neste bioma foi de 21 milhões de habitantes, o que corresponde a 10,5% da população do Brasil. O Índice Firjam de Desenvolvimento Municipal (IFDM) de 2016 atribui um valor de 0,568, o que significa desenvolvimento regular (IBGE, 2019).



Figura 1. Mapa de localização do Bioma Amazônia. Fonte: IBGE (2004)

3.3. Métodos e procedimentos

3.3.1. Métodos de definição do relacionamento entre o fenômeno de desmatamento e a dimensão territorial dos Imóveis cadastrados no CAR

Este trabalho envolve 3 etapas sucessivas. A primeira etapa foi realizar uma série de operações de geoprocessamento nos dados brutos do SICAR de 2020, relativas à retirada de possíveis duplicatas de imóveis rurais e separação em 3 categorias de tamanho de imóvel (pequenos, médios e grandes imóveis rurais). Também foi necessária uma atenção especial aos polígonos que representam os perímetros de assentamentos da reforma agrária, uma vez que eles, por definição, possuem sobreposições entre o perímetro do projeto de assentamento e cada um dos lotes que o compõem (Figura 2).



Figura 2 . Imóvel rural do perímetro do projeto do assentamento agrário (em destaque de contorno vermelho) e vários imóveis menores (ou lotes) em seu interior.

A título de esclarecimento, a seção I do capítulo V da Instrução Normativa N° 2/MMA, de 06 de Maio de 2014 descreve sobre os parâmetros dos quais os Assentamentos de Reforma Agrária precisam se enquadrar para efetivação da inscrição no Sistema de Cadastro Ambiental Rural (MMA, 2014). A normativa explica que o cadastramento é de responsabilidade do órgão fundiário e, tanto o perímetro do assentamento quanto os lotes individualizados, devem ser cadastrados no SICAR.

Procedeu-se em retirar as áreas de Projeto de Assentamento (PA) para que fosse evitada a superestimação da contagem de área e quantidade de imóveis rurais do CAR. Essa superestimação ocorre quando um polígono do Projeto de Assentamento – PA tem em seu interior vários polígonos menores, representando os lotes que o compõem, conforme apresentado na Figura 2. Pode ocorrer na base de dados polígonos iguais e sobrepostos. Eles também foram levados em consideração e retirados pela mesma razão, uma vez que essas áreas sobrepostas eram duplicatas de polígonos já existentes.

Por essa ocorrência de projetos de assentamento e de duplicadas acontecerem

mais em polígonos muito grandes, o processo foi realizado para imóveis rurais classificados em grandes e médios e somente aqueles polígonos de imóveis que foram declarados como projetos de assentamento (código AST) no *shapefile* do CAR. O procedimento deu-se pelo cruzamento entre as camadas de assentamento (código da modalidade: PA) do INCRA, Unidades de conservação de categoria extrativista e desenvolvimento sustentável (MMA, 2020) e polígonos grandes do próprio *shapefile* do CAR.

A segunda etapa refere-se à classificação geral entre imóveis predominantemente públicos e predominantemente privados. Esta classificação preliminar e aproximada foi feita para se ter uma ideia de como está a distribuição dos imóveis rurais dentro do CAR em função de entes majoritariamente públicos (como as unidades de conservação do tipo RESEX e RDS) e de entes majoritariamente privados (fora das áreas de RESEX e RDS). Essa aproximação levou em consideração o que se é declarado no CAR e sua intersecção dos imóveis rurais com áreas públicas. Este procedimento foi realizado para pequeno, médio e grande imóvel rural.

A terceira etapa deste processo foi a realização da seleção dos polígonos de desmatamento e dos imóveis rurais que contém os polígonos de desmatamento. Destacam-se duas situações: 1) neste primeiro momento, para os dados de desmatamento não será aplicada a ferramenta de geoprocessamento *clip (recortar)* para que seja evitada a duplicata de informações; 2) assumiu-se os imóveis rurais presentes no CAR de 2020 como já existindo no período entre 2008 e 2019. A ferramenta aplicada foi o *select by location* do *software ArcGIS* (ESRI, 2021).

As 3 etapas descritas acima como operações feitas isoladamente com o Sistema de Informação Geográfica (SIG) local, foram utilizadas como base conceitual para a definição da estrutura, preparação, sanitização e carregamento dos dados no *Data Warehouse*.

Para expressar computacionalmente essas situações de categorias de imóveis e capacidade de realizar análises espaciais, seguiu-se o estudo do desenho mais adequado para a estruturação do *Data Warehouse* que melhor representasse o relacionamento entre o fenômeno de desmatamento e os imóveis rurais do CAR.

3.3.2. Métodos para estruturação do *Data Warehouse* para Repartição Territorial de Desmatamentos no Bioma Amazônia e o Cadastro Ambiental Rural

O objetivo do *Data WareHouse* (DW) é centralizar os dados do projeto provenientes de diversas fontes (*shapefiles* de polígonos de desmatamento, limites dos imóveis rurais do CAR, limites de áreas públicas em unidades de conservação entre outros) e classificá-los de acordo com suas dimensões temporais e territoriais e estabelecer inter-relações sobre essas informações.

O DW facilita as consultas para extração e análise de dados pela forma como organiza a informação em suas estruturas internas. Possibilita realizar cruzamento das informações para responder os questionamentos existentes de maneira rápida e eficiente, pois o banco de dados possibilita não só a conexão com a plataforma *web*, mas com diferentes ferramentas de análises de dados, facilitando a geração de relatórios, gráficos, tabelas e mapas.

4. Resultado

Como resultados, apresentam-se uma demonstração parcial das 3 etapas apresentadas no item anterior e na sequência a estrutura do Data Warehouse. Uma vez com os polígonos de desmatamento que intersectam os imóveis rurais do CAR, foi possível verificar quanto efetivamente esses polígonos de desmatamento estão inseridos nos imóveis rurais. Para isso, a ferramenta *clip* do *ArcGIS* (ESRI, 2021) foi utilizada, de forma a não comprometer a quantidade de polígonos, mas captar a quantidade de área de desmatamento interna ao imóvel rural. Na Figura 3 pode-se ver um polígono de desmatamento dividido entre 3 imóveis rurais (a) e uma parte do polígono de desmatamento não coincidente com imóveis do CAR (b). A quantidade de polígonos foi preservada (1 desmatamento pela ferramenta *select by location* do *software ArcGIS*) e a área efetiva que ocorreu dentro de imóveis rurais (c) ficou mais próxima da realidade, ao descontar a parte em verde (fora de imóveis rurais).

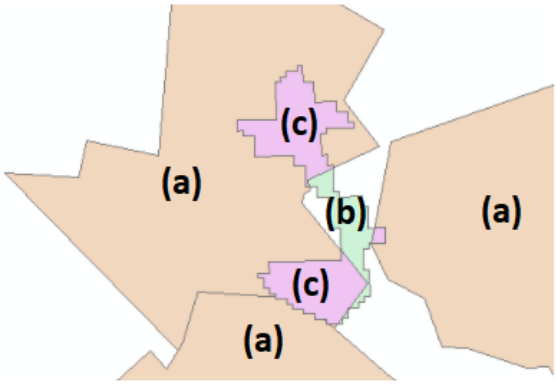


Figura 3. Interseção de um polígono de desmatamento (c)(b) com imóveis rurais (a) após aplicação da ferramenta *clip*.

Para atender as definições dadas no relacionamento entre o fenômeno de desmatamentos e os territórios dos imóveis rurais do CAR desenhou-se o DW utilizando-se do Método Kimball (ASTERA, 2020) com um DW desnormalizado, em esquema estrela (com tabelas fatos e dimensões), definindo-se que o ”fato” ou “assunto” (termos da estrutura do DW) seria o polígono de desmatamento, utilizando os dados de “tamanho da área desmatada em hectares” e o “polígono” como atributos. A partir desse “fato”, definiram-se as dimensões: territoriais e temporais. Essa estrutura permite o incremento de dimensões conforme as necessidades do projeto. Outra vantagem do conceito de “fato”, utilizado pelo *Data Warehouse*, é a possibilidade de identificar quais os territórios fazem intersecção em um único polígono de desmatamento.

Foram utilizados os conceitos de ETL – *Extract, Transform, Load* (Extrair, Transformar e Carregar) para criação dos scripts responsáveis pela extração e carregamento em uma base de dados relacional dos dados referente ao bioma Amazônia e aos imóveis rurais do CAR, além das áreas das Unidades Federativas e dos Municípios, todos presentes em arquivos do tipo *shapefile* e dos *scripts* responsáveis pela transformação e carregamento no *Data Warehouse*.

Tais scripts foram desenvolvidos com base nas definições de classificação e

organização dos dados espaciais do CAR com dois objetivos: 1) processar e armazenar os dados em um banco de dados relacional geoespacial para recuperação dos dados, execução de análises e apresentação em uma plataforma *GeoWeb*; 2) automatização da tarefa para posteriores atualizações dos dados de desmatamento e dos dados do CAR nos anos seguintes.

Por exemplo, uma área de desmatamento em uma região fronteiriça de divisões político-administrativas pode afetar um ou mais Estados da Federação, um ou mais municípios e um imóvel rural cadastrado no Cadastro Ambiental Rural (CAR) ao mesmo tempo (Figura 5). A estrutura desenhada no DW permitiu identificar quais são as áreas que fazem essas intersecções individualmente para cada polígono de desmatamento (ou "fato" do DW). Dessa forma, aumenta a capacidade de qualificação desses desmatamentos, tanto de forma individual quanto no conjunto.

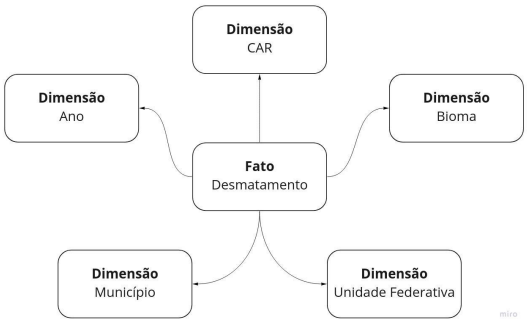


Figura 5. Desenho do *Data Warehouse* em esquema estrela, com a tabela fato ao centro e as dimensões ao redor.

5. Conclusão

A partir dessa proposta, o *Data Warehouse* permitirá a possibilidade de qualificar os polígonos de desmatamento de acordo com suas intersecções, dando visibilidade dos diferentes territórios que um mesmo polígono pode estar vinculado, além dos imóveis rurais do CAR. Também possibilitará agilidade nas consultas dos dados, na geração de estatísticas, relatórios e apresentações em uma plataforma *GeoWeb*.

Ao se trabalhar com os dados de imóveis rurais do CAR é necessário observar e detalhar como estes dados espaciais representam a informação de seus limites geográficos. A classificação de tamanho de módulo fiscal atribuída aos imóveis rurais, assim como a identificação dos polígonos duplicados, necessitou tanto de operações de geoprocessamento quanto uma navegação visual sobre os polígonos. Em algumas vezes, a execução de algumas funções de geoprocessamento presentes no *ArcGIS* (*clip, select by location*) precisou de ser acompanhada de uma identificação visual e pontual, pois alguns cruzamentos e interseções podem não ocorrer de forma como esperada, necessitando de um olhar mais apurado em situações de um polígono apenas tocar o outro.

Conclui-se portanto que os métodos utilizados se mostraram complementares, pois foi possível transpor o trabalho executado em uma estrutura, com ferramentas específicas de geoprocessamento e arquivos *shapefile,* para uma estrutura de *Data*

Warehouse mantendo a relação do fenômeno de desmatamento e das dimensões territoriais dos imóveis rurais do CAR e com a possibilidade de extensão de outras dimensões na mesma estrutura. E com as definições de scripts que automatizam as tarefas e carregam o DW de forma incremental para posterior consumo desses dados nas análises. Todo o material produzido, *scripts* de construção do banco, de sanitização, preparação e carregamento dos dados estão passando por revisões e refinamento para serem disponibilizados ao público ao término do projeto ao qual estão inseridos.

6. Referencias

ORACLE. **O que é um Data Warehouse:** *Data Warehouse Definido*. 2020. Disponível em: https://www.oracle.com/br/database/what-is-a-data-warehouse/. Acesso em: 22 out. 2021.

KIMBALL. ***Kimball Dimensional Modeling Techniques:*** *Basic Fact Table Techniques.* 2013. Disponível em https://www.kimballgroup.com/wp-content/uploads/2013/08/2013.09-Kimball-Dimensional-Modeling-Techniques11.pdf. Acesso em 22 out. 2021.

ASTERA. ***Data Ware House Concepts****: Kimball vs. Inmon Approach*. 2020. Disponível em: https://www.astera.com/type/blog/data-warehouse-concepts/. Acesso em: 20 jun. 2021.

SPOTLESSDATA. ***Exploring data warehouses and data quality:*** *Extract, transform, load.* 2017. Disponível em: https://web.archive.org/web/20180726071809/https://spotlessdata.com/blog/exploring-data-warehouses-and-data-quality. Acesso em 22 out. 2021.

BRASIL. **Lei nº 12.651, de 25 de maio de 2012.** Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12651.htm. Acesso em: 22 jun. 2021.

EMBRAPA. Sistema de Inteligência Territorial Estratégica do Bioma Amazônia. 2019. Disponível em: https://www.cnpm.embrapa.br/projetos/siteamazonia/

ESRI. ***ArcGIS* 10.7**. 2021. Disponível em: http://www.esri.com/software/arcgis/arcgis-for-desktop. Acesso em: 20 jun. 2021.

IBGE. **Biomas.** 2004. Disponível em: https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15842-biomas.html?edicao=16060&t=sobre. Acesso em: 12 jun. 2021.

IBGE. **Biomas e sistema costeiro-marinho do Brasil:** compatível com a escala 1:250000. 2019. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101676. Acesso em: 20 jun. 2021.

IBGE. **Estrutura Territorial.** 2020. Disponível em https://www.ibge.gov.br/geociencias/organizacao -do-territorio/estrutura-territorial.html. Acesso em 22/06/2021.

INCRA. **Acervo fundiário**. 2020a. Disponível em http://acervofundiario.incra.gov.br/acervo/acv.php. Acesso em: 12 jun. 2021.

INCRA. **Módulo Fiscal.** 2020b. Disponível em

https://antigo.incra.gov.br/pt/modulo-fiscal.html .html. Acesso em: 20 jun. 2021.

INPE. **TerraBrasilis - Incremento anual no desmatamento - *Shapefile* (2008/2018)**. 2020. Disponível em http://terrabrasilis.dpi.inpe.br/download/dataset/legal-amz-prodes/vector/yearly_deforestation_2008_2018.zip. Acesso em: 20 jun. 2021.

MMA. **Unidades de Conservação**. 2020. Disponível em: https://antigo.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs/dados-georreferenciados.html. Acesso em: 20 jun. 2021.

MMA. **Instrução Normativa N° 2/MMA, de 06 de Maio de 2014**. Disponível em: https://www.car.gov.br/leis/IN_CAR.pdf. Acesso em: 20 jun. 2021.

SICAR. **Sistema de Cadastro Ambiental Rural**. 2021. Disponível em https://www.car.gov.br/publico/imoveis/index. Acesso em: 20 jun. 2021.



Estudo comparativo entre Máquina de Vetores de Suporte e Rede Neural Convolucional na classificação do inseto *Diaphorina citri*

Mário Lúcio G. de Q. Pierre Júnior[1], José Leonardo dos S. Melo [2], Marcelo P. de Miranda[3], Michele Fúlvia Angelo[4]

[1]Instituto Federal de Educação, Ciência e Tecnologia Baiano
Senhor do Bonfim, Bahia, Brasil

[2]Instituto Federal de Educação, Ciência e Tecnologia da Bahia
Euclides da Cunha, Bahia, Brasil

[3]Fundo de Defesa da Citricultura
Araraquara, São Paulo, Brasil

[4]Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana
Feira de Santana, Bahia, Brasil

{mario.junior}@ifbaiano.edu.br, mfangelo@ecomp.uefs.br

Abstract. *Diaphorina citri is the main propagator of Huanglongbing citrus disease. To control the problem, the insects are captured by adhesive traps and manually counted to adjust the insecticide dosage. Thus, this work aimed to measure and analyze the performance of the automatic classification of Diaphorina citri from Machine Learning approaches. For the training of the models, an image dataset was created and the Support Vector Machine (SVM) and Convolutional Neural Network (CNN) approaches were used. CNN, with Inception v3 model, presented an average of 99.37% accuracy in the final classification of Diaphorina citri, a significant result in relation to the 97.99% presented by SVM.*

Resumo. *O inseto Diaphorina citri é o principal propagador da doença dos citros Huanglongbing. Para o controle do problema, os insetos são capturados por armadilhas adesivas e contados manualmente, para ajuste da dosagem de inseticida. Assim, este trabalho teve como objetivo medir e analisar o desempenho da classificação automática do Diaphorina citri a partir de abordagens de Aprendizado de Máquina. Para o treinamento dos modelos foi criado um banco de imagens e utilizadas as abordagens de Máquina de Vetores de Suporte*

(SVM) e Rede Neural Convolucional (CNN). A CNN, com modelo Inception v3, apresentou em média 99.37% de acurácia na classificação final do Diaphorina citri, resultado significativo em relação aos 97.99% apresentado pela SVM.

1. Introdução

O Brasil é o maior fornecedor mundial de frutas cítricas *in natura* e outros subprodutos [Neves et al. 2010]. No entanto, apesar de toda sua importância, os problemas fitossanitários causam grandes perdas econômicas aos produtores e tornam mais difícil a exportação do suco de laranja brasileiro [Nava et al. 2007].

Uma das preocupações mais sérias do setor é o Huanglongbing (HLB), também conhecido como *Greening*, considerada a pior doença de citros. Ela tem como característica a clorose, entupimento dos vasos responsáveis por levar água e nutrientes da raiz para a copa da planta, causando a redução e perda da qualidade do fruto, deixando-o com sabor amargo [Halbert e Manjunath 2004].

Sem métodos de cura, o manejo da doença é realizado com inspeções para identificação, eliminação das plantas sintomáticas e o controle químico, que é a maneira mais utilizada pelo citricultor para reduzir a população do *Diaphorina citri* [Yamamoto et al. 2009], principal inseto propagador da bactéria causadora da doença do HLB.

O psilídeo *Diaphorina citri* é um pequeno inseto sugador que mede de 2 a 3mm de comprimento, de cor cinza e manchas escuras nas asas [Gallo et al. 2002]. Devem ser realizadas pulverizações periódicas para o controle do inseto com o intuito de reduzir a probabilidade de ocorrência de novas infecções [Belasque Junior et al. 2009]. A tomada de decisão para a pulverização é realizada por meio do monitoramento. As armadilhas adesivas amarelas são utilizadas como forma de atração e captura de insetos e é um método de inspeção visual [Belasque Junior et al. 2009].

Um dos grandes problemas na utilização desse método de inspeção visual é que as armadilhas atraem outros insetos além do *Diaphorina citri* [Yamamoto et al. 2009]. Após dias de exposição no pomar, as armadilhas adesivas são retiradas e levadas para o laboratório. Através da identificação e contagem manual dos psilídeos encontrados, realizada por técnicos treinados, são feitos cálculos para se estimar a população da praga [Leonardo 2014].

O monitoramento é um importante componente do manejo integrado de pragas, pois indica o momento correto para a tomada de decisão, evitando aplicações desnecessárias de inseticidas e, consequentemente, o desequilíbrio biológico [Yamamoto et al. 2009]. Portanto, metodologias que automatizam esse processo de identificação e melhorem os resultados de acurácia são importantes para um processo eficiente de controle. Assim, o objetivo deste trabalho é medir e analisar o desempenho da classificação automática do inseto *Diaphorina citri* a partir de abordagens de Aprendizado de Máquina.

2. Materiais e Métodos

Os experimentos neste trabalho foram executados com as seguintes configurações de hardware e software: Processador Intel Corei7-7500U 2.70GHz; memória de 8GB; Disco

rígido de 1TB; Placa de vídeo NVIDIA R GeForce R MX150 de 4GB; Sistema operacional Ubuntu 16.04 LTS; Python 2.7.12; Keras 2.2.0; TensorFlow 1.9.

Para o desenvolvimento foram necessárias as seguintes etapas: Criação do *dataset* de imagens das armadilhas adesivas; Classificação automática do inseto Diaphorina citri; e Avaliação do classificador. A seguir serão descritas cada uma das etapas.

2.1. Criação do *dataset* de Imagens

O *dataset* de imagens foi criado com amostras digitalizadas das armadilhas adesivas, disponibilizadas pelo Fundo de Defesa da Citricultura (FUNDECITRUS). O recorte dos insetos, nas imagens digitalizadas, foi feito manualmente utilizando o software *ImageJ* e resultando em 551 amostras do *Diaphorina citri* e 551 amostras de outros insetos, com tamanhos variados e posições distintas no quadrante do recorte, conforme apresentado na Figura 1.

O uso da quantidades iguais de amostras é para impedir o desbalanceamento, ou desequilíbrio na quantidade de amostras por classe, com objetivo de evitar problemas de generalização na classificação. Segundo Japkowic [Japkowicz e Stephen 2002], o desequilíbrio de classes pode ter um efeito prejudicial no treinamento de classificadores, afetando a convergência no treinamento e a generalização do modelo no conjunto de teste.

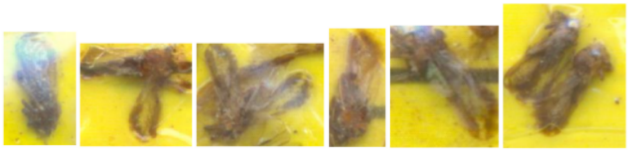


Figura 1. Recorte dos insetos

2.1.1. Incremento do *dataset*

Um dos métodos mais comuns para reduzir o super ajustamento durante o treinamento em aprendizado de máquina, aumentar a capacidade do classificador em generalizar e classificar padrões de testes que não foram utilizados no treinamento, é aumentar artificialmente o conjunto de imagens produzindo-se amostras transformadas a partir das imagens originais [Krizhevsky et al. 2012].

Para o aumento artificial das imagens foi utilizada a biblioteca *Augmentor*, resultando no aumento de 7.5 vezes do tamanho original do *dataset*, totalizando 8.354 amostras, divididas igualmente para *Diaphorina citri* e outros insetos nas dimensões de 224 x 224 e 299 x 299 pixels, dimessões de entrada para as abordagens SVM e CNN, respectivamente.

Neste trabalho, o *dataset* de imagens foi dividido em 70% das amostras destinadas para o conjunto de desenvolvimento e 30% das amostras para o teste final. O conjunto de desenvolvimento foi subdividido em outros três conjuntos, com 90% destas imagens divididas em partes iguais para o treinamento e validação e os 10% restantes destinadas para os testes de parâmetros do conjunto de desenvolvimento.

2.2. Classificação automática do Diaphorina Citri

Neste trabalho foi proposta a classificação do *Diaphorina citri* através de algoritmos de aprendizagem de máquina com as abordagens de Máquina de Vetores de Suporte e Rede Neural Convolucional. Os métodos serão descritos a seguir.

2.2.1. Support Vector Machine (SVM)

O SVM foi escolhido devido a sua ampla utilização em abordagens bens sucedidas de classificação de imagens [Boiman et al. 2008], [Chaplot et al. 2006], [Kotsia e Pitas 2006]. O SVM foi executado com sua configuração padrão e com *kernel* linear. Essas escolhas deveram-se a simplicidade de classificação desse núcleo e ao pequeno erro de generalização, o baixo custo computacional.

Nesta abordagem foram utilizados os algoritmos extratores de característica SIFT (*Scale Invariant Feature Transform*) [Lowe 2004] e o SURF (*Speeded Up Robust Features*) [Bay et al. 2008]. Para a implementação dos extratores foi criado um procedimento computacional que executou as tarefas de: ler os arquivos contendo os endereços das imagens, carregar cada imagem na memória, convertê-la para a escala de cinza, extrair suas características e salvar em arquivo o resultado dessas extrações, indexando-as de acordo com os rótulos: 1 para o *Diaphorina citri* e 0 para os outros insetos da base.

O método de anotação de características utilizado foi o Bolsa de Características (BoF - *Bag-of-Features*) [Tsai 2012] [Bosch et al. 2007]. A geração e representação dessas características pelo BoF segue a sequência: detecção de regiões ou pontos de interesse; descrição dos locais dessas regiões ou pontos de interesse; quantização dos descritores agrupando-os em palavras visuais para o livro de código (*Codebook*); contagem das palavras visuais específica do vocabulário visual para a construção do vetor de características BoF [Bosch et al. 2007].

Os livros de código foram criados usando o algoritmo de aprendizagem de máquina não-supervisionada *Mini Batch K-Means*. Esse algoritmo forma grupos (clusters) de coisas similares. Foi definida a quantidade de agrupamentos em 300 conjuntos, baseado no baseado no livro "*Mastering Machine Learning with Scikit-Learn* [Hackeling 2017]

Desta forma, foram criados três tipos distintos de livros de código. Um livro de código "Geral", que engloba tanto as características do *Diaphorina citri* quanto a dos outros insetos, agrupadas em 300 conjuntos. O segundo livro de código criado somente com as características do "Diaphorina citri", agrupadas em 300 conjuntos. E, no terceiro, foi criado um livro de código com 300 grupamentos de características das "Duas Classes", *Diaphorina citri* e outros insetos, totalizando 600 agrupamentos.

2.2.2. Rede Neural Convolucional (CNN)

Neste trabalho, utilizou-se a rede *Inception* v3, que é baseada na arquitetura *GoogLeNet* que venceu a competição ILSVRC (*ImageNet Large-Scale Visual Recognition Challeng*) em 2014. Esse modelo apresenta a sequência de módulos Inception A, B e C resultando

em 42 camadas de rede. Embora apresente uma grande quantidade de camadas, o custo computacional é apenas cerca de 2,5 maior que o da GoogLeNet [Szegedy et al. 2016].

Os experimentos foram realizados utilizando parâmetros semelhantes aos da abordagem anterior. Foi definido o valor de 50 *epochs* de iterações do algoritmo e com 32 *batch size*, correspondente ao grupo de imagens selecionadas para o treinamento da rede. Definida a função *categorical cross entropy* para o cálculo do valor de erro e a função de ativação *softmax* para calcular as probabilidades de pertencimento de uma amostra para as duas classes definidas. Foi utilizado o banco de imagens redimensionadas para 299 x 299, que é a dimensão de entrada para a rede *Inception v3* e os algoritmo de otimização Adam [Kingma e Ba 2014] e RMSprop [Tieleman et al. 2012].

2.3. Avaliação dos Classificadores

Quando os modelos são desenvolvidos, seu desempenho é avaliado no conjunto de validação e é selecionado o modelo com o melhor desempenho para a aplicação das amostras de teste final. Neste trabalho, as medidas de avaliação do modelo utilizadas foram: acurácia, precisão, *recall* e F1-score.

Além destas medidas, foi aplicada a Validação Cruzada que é um método para avaliar a capacidade de generalização de um modelo. Foi utilizada a validação cruzada *k-fold*, com o k=2, que subdividiu o conjunto de treinamento em k subconjuntos iguais, contendo amostras distribuídas aleatoriamente. Em seguida, k iterações de etapas de treinamento e validação foram aplicadas usando diferentes subconjunto de dados para a validação e o outro subconjunto para o treinamento do modelo. Este processo foi repetido cinco vezes. Para cada iteração, uma nova subdivisão de validação e treinamento foi gerada.

Também foram aplicados testes estatísticos para analisar se existem diferenças nas técnicas aplicadas para cada algoritmo de classificação. Pois, embora o desempenho de diferentes classificadores possa ser mostrado como diferente em conjuntos de dados específicos, é necessário confirmar se as diferenças observadas são estatisticamente significativas e não meramente coincidentes. Com os valores de acurácia resultantes dos classificadores, foi aplicado o teste não paramétrico Friedman, que indica que as informações não apresentam um padrão de distribuição, usado para comparar múltiplos algoritmos em vários domínios. E o uso do teste de *Nemenyi* para comparar os resultados dos experimentos em pares, os classificadores [Japkowicz e Shah 2011].

3. Resultados e Discussões

Ao executar o algoritmo SVM o resultado da classificação do inseto foi de 50.08% de acurácia. Os valores de agrupamentos foram reduzidos, empiricamente, de 300 para 250, 100 e 50 clusters. Tendo os resultados: 51.56% de acurácia para 250 clusters; 80.40% para 100 clusters. A quantidade de 50 clusters apresentou o melhor resultado com 97.99% de acurácia para a classificação do *Diaphorina citri* em imagens digitalizadas.

O trabalho de Venugoban [Venugoban e Ramanan 2014] mostrou um estudo simplificado na variação do número de agrupamentos de bolsas de caracterísitcas, aplicado a um contexto específico, e serviu de base para a definição de testes de variação de agrupamentos realizados nesta pesquisa. Os resultados dos testes também mostraram que o valor de 50 agrupamentos resultou em um melhor desempenho na classificação.

Definido em 50 clusters, o algoritmo de classificação foi novamente executado e desta vez sendo aplicado o procedimento de validação cruzada *2-fold x 5*. Na Tabela 1 estão os valores resultantes das 5 execuções do SVM. Na sexta coluna estão os resultados de acurácia e desvio padrão, em função da variação da acurácia alcançada no procedimento de validação cruzada. O tempo médio de cada classificação foi de 6.20 minutos.

Tabela 1. Resultado da classificação utilizando SVM com validação cruzada

Exec	Acurácia	Recall	Precision	F1-Score	Validação Cruzada	Tempo
1	97.99%	98.80%	97.23%	98.01%	94.05% (+/- 0.16%)	00:06:25
2	97.21%	97.97%	96.41%	97.18%	95.67% (+/- 0.09%)	00:06:10
3	97.57%	98.36%	96.78%	97.57%	94.49% (+/- 0.19%)	00:06:00
4	98.18%	98.00%	98.39%	98.19%	95.02% (+/- 0.08%)	00:06:40
5	98.07%	98.19%	97.99%	98.09%	95.53% (+/- 0.11%)	00:06:25

A Tabela 2 apresenta os valores de acurácia, da abordagem CNN, do processo de validação cruzada *2-fold x 5* com os otimizadores *Adam* e *RMSProp*. Um ponto que chama atenção nos dados é o tempo de execução do algoritmo que apresentou em média 26 horas de execução, aumento significativo de execução em relação ao SVM. Apesar do aumento do custo computacional, o modelo *Inception* apresentou números importantes com 99.82%, na primeira execução com o RMSProp, em relação a taxa de acurácia tanto no processo de validação cruzada como também na generalização do modelo.

Tabela 2. Resultados do treinamento com *Inception*

	ADAM		RMSProp	
Execução	Validação Cruzada	Tempo	Validação Cruzada	Tempo
1	99.48% (+/- 0.12%)	26:06:00	99.82% (+/- 0.06%)	25:56:00
2	99.80% (+/- 0.08%)	26:00:00	98.67% (+/- 0.10%)	26:16:00
3	99.78% (+/- 0.10%)	25:57:00	97.47% (+/- 0.12%)	26:01:00
4	99.84% (+/- 0.05%)	26:03:00	98.35% (+/- 0.24%)	26:00:00
5	99.56% (+/- 0.15%)	26:15:00	98.13% (+/- 0.19%)	26:06:00

A Tabela 3 apresenta os resultados de generalização do modelo *Inception* com suas medidas de avaliação para o conjunto de amostras de teste de parâmetros, com a aplicação dos otimizadores *Adam* e *RMSprop*. Os resultados de acurácia apresentaram valores de acertos próximos para os dois otimizadores, porém com o maior valor de acurácia sendo apresentado com o otimizador *RMSprop* na primeira e na quinta execução do processo de validação cruzada com 99.82%, consequentemente apresentando poucos erros na classificação do *Diaphorina citri*. O resultado da matriz de confusão, com dados de testes de parâmetros, previu incorretamente somente uma amostra de diaphorina citri como sendo outros insetos e classificou corretamente as amostras dos outros insetos.

Para analisar se existem diferenças nas técnicas aplicadas a cada algoritmo de classificação, foi necessário avaliar os resultados através de testes estatísticos. É necessário confirmar se as diferenças observadas foram estatisticamente significativas ou não meramente coincidentes. Assim, o teste verificou a diferença dos resultados dos classificadores quanto a questão de significância dos resultados obtidos.

Tabela 3. Resultados da generalização do modelo *Inception* com amostras para ajustes de parâmetros com os otimizados *ADAM* e *RMSProp*.

ADAM				RMSProp			
Acurácia	**Recall**	**Precision**	**F1-Score**	**Acurácia**	**Recall**	**Precision**	**F1-Score**
99.48%	99.31%	99.65%	99.48%	99.82%	100%	99.65%	99.82%
99.66 %	99.66%	99.66%	99.66%	99.66%	100%	99.32%	99.66%
99.48%	99.31%	99.65%	99.48%	99.49%	99.66%	99.32%	99.49%
99.66%	99.32%	100%	99.66%	99.49%	98.98%	100%	99.49%
99.14%	98.98%	99.32%	99.15%	99.82%	100%	99.65%	99.82%

Na avaliação dos resultados, foi aplicado o teste não paramétrico de *Friedman*. O resultado do teste apresentou o *p-value*, que representa a probabilidade de significância, de 0.04057. Como o valor de significância de 0.04057 é menor que o valor de confiança estabelecido em 0.05, foi rejeitada a hipótese nula H_0 que afirmaria serem iguais os classificadores, não apresentando diferença significativa em seus resultados. Ao ser rejeitada a hipótese nula, fez-se necessária a aplicação de um teste *Post Hoc* para o teste de *Friedman*.

Para o *Post Hoc* foi utilizado o teste de *Nemenyi* que compara os algoritmos e determina onde está a diferença entre eles [Japkowicz e Shah 2011]. No teste de *Nemenyi*, a comparação entre o algoritmo *Inception* com otimizador Adam x *Inception* com RMSProp não apresentou diferença significativa, com resultado maior que 0.05. Porém, a comparação entre o algoritmo *Inception* com o Adam x SVM teve como resultado o valor 0.004605 e o algoritmo *Inception* com RMSProp x SVM com o valor de 0.001291, apresentou diferença significativa entre o modelo *Inception* com os otimizadores Adam e RMSprop e a metodologia SVM.

Demonstrando maior confiabilidade na classificação correta do inseto *Diaphorina citri*, foi selecionada para a execução final a arquitetura *Inception*. Para esta execução final foram utilizadas as amostras de imagens do conjunto de teste, correspondente a 30% das amostras do banco de imagem, que ainda não haviam sido utilizadas em nenhum procedimento anterior deste trabalho. Ao aplicar o conjunto com amostras de teste ao modelo *Inception* foi obtida uma média de acurácia de 99.51% na validação cruzada e 99.37% na validação com o resultado da generalização do modelo com o conjunto de teste final, conforme Tabela 4.

Os modelos de aprendizagem de máquina podem ser salvos durante e após o final do treinamento, evitando assim longos períodos de treinamento e podendo esse modelo ser compartilhado. Desta forma, a execução dos modelos no teste final teve a média de tempo de 33.03 minutos, conforme Tabela 4.

Analisando-se os resultados apresentados por [Shen et al. 2018] e [Liu et al. 2016], que também utilizaram aprendizado profundo para classificação e identificação de insetos, é possível constatar que o presente trabalho, apesar de não ter o mesmo inseto como objeto de pesquisa, apresentou a média final de acurácia de 99.51% na validação cruzada e 99.37% na validação do modelo, sendo estes resultados superiores aos obtidos nos trabalhos citados.

Tabela 4. Resultado final do treinamento com o algoritmo *Inception*.

Exec	Acurácia	Recall	Precision	F1-Score	Validação Cruzada	Tempo
1	99.12%	99.44%	98.80%	99.12%	99.28% (+/- 0.08%)	00:34:55
2	99.32%	99.60%	99.04%	99.32%	99.48% (+/- 0.11%)	00:30:40
3	99.56%	99.68%	99.44%	99.56%	99.64% (+/- 0.13%)	00:32:17
4	99.32%	99.60%	99.04%	99.32%	99.48% (+/- 0.09%)	00:33:01
5	99.44%	99.60%	99.28%	99.44%	99.68% (+/- 0.27%)	00:34:23

4. Conclusão

Este trabalho teve como objetivo medir e analisar o desempenho da classificação automática do inseto *Diaphorina citri* a partir de abordagens de Aprendizado de Máquina. Para isso foram utilizadas uma SVM e uma CNN. A CNN com a arquitetura de rede *Inception* apresentou os melhores valores de acurácia e menor quantidade de erros na classificação do *Diaphorina citri*. Por esse motivo, foi aplicado ao modelo o teste final de generalização que resultou em uma média de acurácia de 99.51% na validação cruzada e 99.37% na validação do modelo com o conjunto de teste final.

Como melhorias para esta pesquisa, poderá ser desenvolvido um algoritmo de processamento de imagens para a extração de amostras dos insetos contidos nas armadilhas adesivas, de maneira que automatize a criação do *dataset* de imagem. E também, aprimorar o algoritmo CNN, utilizando outros parâmetros de otimização, objetivando reduzir o custo computacional. Poderão também ser experimentadas outras arquiteturas de CNN.

Referências

Bay, H., Ess, A., Tuytelaars, T., e Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.

Belasque Junior, J., Bergamin Filho, A., Bassanezi, R. B., Barbosa, J. C., Fernandes, N. G., Yamamoto, P. T., Lopes, S. A., Machado, M. A., Leite Junior, R. P., Ayres, A. J., et al. (2009). Base científica para a erradicação de plantas sintomáticas e assintomáticas de huanglongbing (hlb, greening) visando o controle efetivo da doença. *Tropical Plant Pathology*, 34(3):137–145.

Boiman, O., Shechtman, E., e Irani, M. (2008). In defense of nearest-neighbor based image classification. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

Bosch, A., Munoz, X., e Marti, R. (2007). Which is the best way to organize/classify images by content? *Image and vision computing*, 25(6):778–791.

Chaplot, S., Patnaik, L. M., e Jagannathan, N. (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical signal processing and control*, 1(1):86–92.

Gallo, D., Nakano, O., Silveira Neto, S., Carvalho, R. P. L., e Baptista, G. C. d. (2002). *Entomologia agrícola*. Fundação de Estudos Agrários Luiz de Queiroz, 632.7 edition.

Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd.

Halbert, S. E. e Manjunath, K. L. (2004). Asian citrus psyllids (sternorrhyncha: Psyllidae) and greening disease of citrus: a literature review and assessment of risk in florida. *Florida entomologist*, 87(3):330–353.

Japkowicz, N. e Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*. Cambridge University Press.

Japkowicz, N. e Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449.

Kingma, D. P. e Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kotsia, I. e Pitas, I. (2006). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE transactions on image processing*, 16(1):172–187.

Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105.

Leonardo, A. (2014). Otimização da leitura de cartão adesivo amarelo para o monitoramento de adultos de diaphorina citri kuwayama (hemiptera: Liviidae). Master's thesis, Fundo de Defesa da Citricultura.

Liu, Z., Gao, J., Yang, G., Zhang, H., e He, Y. (2016). Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific reports*, 6(1):1–12.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Nava, D., Torres, M., Rodrigues, M., Bento, J., e Parra, J. (2007). Biology of diaphorina citri (hem., psyllidae) on different hosts and at different temperatures. *Journal of Applied Entomology*, 131(9-10):709–715.

Neves, M. F., Trombin, V., Milan, P., Lopes, F., Cressoni, F., e Kalaki, R. (2010). O retrato da citricultura brasileira. *Ribeirão Preto: CitrusBR*, page 137.

Shen, Y., Zhou, H., Li, J., Jian, F., e Jayas, D. S. (2018). Detection of stored-grain insects using deep learning. *Computers and Electronics in Agriculture*, 145:319–325.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., e Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.

Tieleman, T., Hinton, G., et al. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31.

Tsai, C.-F. (2012). Bag-of-words representation in image annotation: A review. *International Scholarly Research Notices*, 2012.

Venugoban, K. e Ramanan, A. (2014). Image classification of paddy field insect pests using gradient-based features. *International Journal of Machine Learning and Computing*, 4(1):1.

Yamamoto, P. T., Fellipe, M. R., Sanches, A. L., Coelho, J. H., Garbim, L. F., e Ximenes, N. L. (2009). Eficácia de inseticidas para o manejo de diaphorina citri kuwayama (hemiptera: Psyllidae) em citros. *BioAssay*, 4.



Experimental evaluation of Data Augmentation heuristics for plant identification systems based on Deep Learning

Luciano Araújo Dourado Filho[1], **Rodrigo Tripodi Calumby**[1]

[1]University of Feira de Santana – Feira de Santana – BA – Brazil

lucianoadfilho@ecomp.uefs.br, rtcalumby@uefs.br

Abstract. *Data augmentation (DA) allows increasing datasets for training machine learning models that demands large amounts of data. In real-world applications in which data may not be abundant enough and data acquisition is not easy, DA enables increasing diversity and introducing model generalization. In this work we evaluate several DA techniques and combining approaches to extend image datasets used to train plant species recognition models. We experimentally validated Deep Convolutional Neural Networks (DCNN) with several datasets obtained from common augmentation techniques and combinations. The results allowed the identification of the Translate + Crop augmentation policy as the most effective within the scope of evaluation.*

1. Introduction

For a long time, machine learning algorithms for computer vision tasks were not effective or efficient enough for real-world applications, especially in the plant domain. Although innovative, these technologies mostly involved hand-craft designed solutions for single-organ detection (e.g., flower or leaves), recognition or segmentation from images. Therewith, ineffective models that were highly expensive to design and with low generalization power were a common reality. Hence, the use of those methods in critical systems (i.e., agriculture, pharmacology) was out of question. In turn, with the exploitation of high-end parallel architectures, those traditional solutions were considerably outperformed by *Deep Learning* (DL) models, specially the ones based on Deep Convolutional Neural Networks (DCNN) [Mehdipour Ghazi et al. 2017]. These advances allowed large-scale experiments and even the construction of models that surpassed human performance in some plant recognition tasks [Goëau et al. 2018].

The emergence of effective models and efficient models allowed the exploration of real-world experiments and the proposal of solutions for increasing and automatizing field production through DCNN-based systems in plant recognition tasks [Khan et al. 2018,

Brahimi et al. 2018, Zhao et al. 2020]. Despite that, the predictive models commonly include millions of parameters, demanding massive amounts of data for proper optimization, which still make the learning process unfeasible in some contexts. Besides that, additional challenges emerge from datasets with imbalanced class distributions or from the high cost for large-scale data acquisition. Such circumstances motivate researchers to use Data Augmentation (DA) for training DCNNs. DA allows to artificially extend a dataset through label-preserving transformations for the generation of novel samples from real ones. Indeed, DA has been reported as promising in overcoming dataset imbalance, reducing overfitting and introducing invariance in DL models [Shorten e Khoshgoftaar 2019].

In this context, recent works on automatic plant recognition through DCNN have commonly relied on DA [Sulc et al. 2018, Haupt et al. 2018], which shows a crucial role for achieving state-of-the-art classification. However, DA techniques have been used in an ad-hoc way, given the lack of a consensus regarding the suitability of the alternatives with respect to each context of application (i.e., organ type, photo type, background, herbarium sheet, etc). Besides, arbitrary costful heuristics can lead to performance bottlenecks and effectiveness decrease. Therefore, this work proposes a performance analysis of DA techniques commonly used for the optimization of plant species recognition models based on DL. Additionally, we assessed three heuristics for the combination of augmentation techniques, which allowed demonstrating the most effective techniques as well as the most suitable to perform plant species recognition.

2. Related Work

Data Augmentation enables to increase a dataset quantity in an artificial way [Buda et al. 2018]. The motivation behind that is to improve the DCNNs generalization power, through a invariance hardcoding procedure [Mehdipour Ghazi et al. 2017] that happens behind the process of training with images that presents variations of angle, position, light, brightness, simultaneously with the original corresponding versions. Consequently, the models become more invariant to these adversities and consequently, increase their capacity of performing well for unseen data [Shorten e Khoshgoftaar 2019, Mehdipour Ghazi et al. 2017].

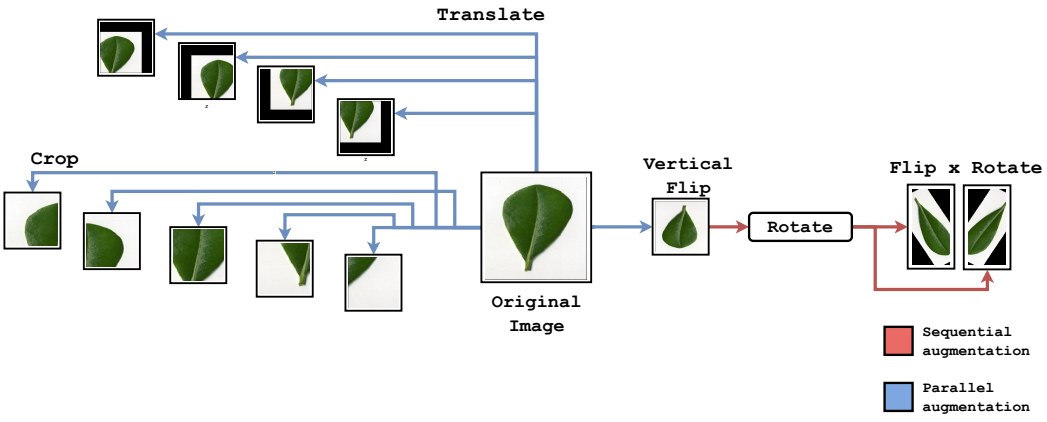


Figure 1. Illustration of geometric augmentation techniques.

Basically, augmentation techniques can be classified as geometric when they alter the geometric constitution of the images (Figure 1), or photometric when they only cause

changes over the color space [Taylor e Nitschke 2018]. Traditional geometric techniques includes: Crop, Flip (horizontally, vertically), Translate, Rotate, whereas photometric transformations mainly involves changes in brightness, light, color or saturation for example [Shorten e Khoshgoftaar 2019]. In terms of heuristics, these techniques (geometric or photometric) are usually applied individually or combined through sequential or parallel application heuristics to yield augmentation policies. In the *Sequential* modality, for example, samples are obtained from the chained application of multiple augmentation techniques over the same image. Figure 1 illustrates the sequential heuristic with a Flip x Rotate example. This process generates two rotated and vertically-flipped images, arising from the application of Flip followed by Rotate. In turn, the *Parallel* approach yields augmentation policies from the individual (parallel) application of two or more techniques and addition of the obtained images to a unique set. Besides that, some researchers have also proposed to use more complex heuristics, such as through Deep Learning based methods [Wang et al. 2017].

In the plant recognition context, the work in [Pawara et al. 2017] conducted an evaluation of augmentation techniques including Rotation, Blur, Contrast, Scaling, Illumination and Projection combined through a parallel approach. For that, three datasets were used: Folio (600 leaf images from 32 species), Swedish (1.000 images on a plain background of 15 Swedish tree species) and Agril Plant (3.000 images of 10 fruit species). Although promising results were achieved, it was not possible to draw generalizable techniques, possibly due to the low complexity of the datasets, with perfectly balanced classes, high background homogeneity (except for Agril Plant) and the small number of species/specimens.

In [Zhang et al. 2015] the authors proposed a CNN architecture for leaf classification and used Flavia dataset (1907 images of 32 species). The model was trained with traditional DA techniques (Translate, Scale, Rotate, Contrast and Sharpening) randomly selected according to a desired augmentation factor (5x, 10x or 20x). In a similar way, the authors observed a reasonable effectiveness improvement with DA, although, no expressive difference was observed for the techniques assessed and the low complexity of the dataset weakens the conclusion regarding the general effectiveness of the augmentation approaches. Similarly, in [Pandian et al. 2019] the authors assessed many traditional DA techniques, including Flip, Rotate, Crop, color transformation, PCA and noise injection, as well as DL-based techniques (WGAN, DCGAN, neural style transfer). An imbalanced plant dataset was used with around 54 thousand specimens from 38 classes including healthy and diseased leaves. The authors demonstrated that using all augmentation techniques (WGAN, DCGAN, neural style transfer and traditional techniques) at once lead to higher performance in relation to the use in isolation. Despite its complexity, higher costs and need for optimization, the DL-based augmentations applied in parallel allowed slightly superior results in contrast to the traditional techniques also combined in parallel.

As reported in [Mehdipour Ghazi et al. 2017], the combination by sequential and parallel application of vertical flip, rotation, and scaling augmentation techniques through a image patch extraction pipeline, allowed to achieve state-of-the-art performance in a multi-organ, large-scale plant classification task involving about 1,000 species. The authors observed a decrease of overfitting and the improvement of the benefits of fine-tuning. They also showed that an 80-fold augmentation outperformed a 10-fold by roughly 6% in

accuracy. This illustrates by how much the authors needed to increase the augmentation factor (no. of generated images) in order to obtain satisfactory results as well as how some augmentation heuristics can generate large datasets that increase training costs.

In general, large-scale studies that relied on DA demonstrated its important role towards developing real-world plant recognition systems. Nevertheless, most studies were not successful on determining general techniques or combining heuristics, mainly due to the low complexity of the data or the limited amount of combinations of DA techniques assessed. Hence, this work aims at assessing several DA techniques through different combination heuristics and a representative scenario of large-scale plant recognition in terms of class imbalance, visual image heterogeneity (multiple plant organs), number of species and specimens. Therefore, we believe to provide augmentation policies that are more suitable to the development of plant species recognition systems based on DCNN.

3. Experimental Setup

As discussed in [Pawara et al. 2017], the characteristics of the classification task and data used are decisive for robust DA studies involving DCNN. On the one hand, smaller datasets makes it difficult to conduct in-depth analysis of real-world challenges given the limited amount of samples and/or classes and its irregular distribution. On the other hand, datasets of larger scale/complexity makes comprehensive systematic analysis infeasible, specially considering the multiple possible combinations of DA techniques.

Considering the mentioned constraints, this work was conducted over the PlantCLEF2013 (PC2013) dataset [Goëau et al. 2013]. It presents 26,077 images of 250 plant species from the French flora, from which 5092 (20%) images are used for testing. Introducing additional complexity, the dataset includes images acquired with different methods and of multiple plant organs, resulting in two image types and 6 sub-types as illustrated in Figure 2. The two main types are: *Sheet as Background* (a homogeneous surface of uniform white background) representing 42% of the total images (11,031 samples) and *Natural Background*, with 15,046 samples representing 58% of total images, composed of natural photographs captured outdoors from different perspectives of different organs.

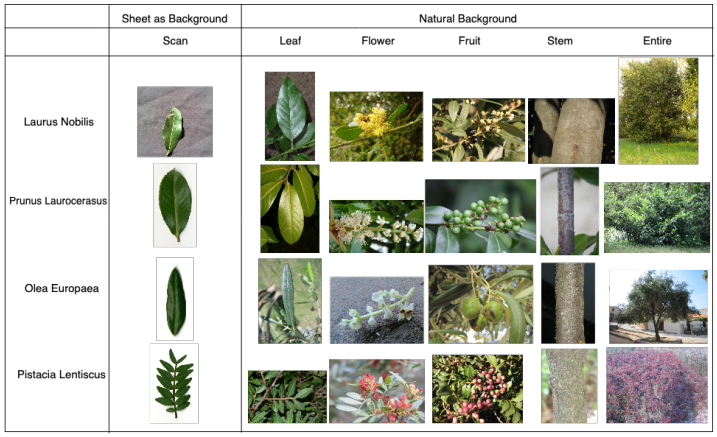


Figure 2. Samples of images with multiple organs of different species divided by the categories “Sheet as Background” and “Natural Background”.

For species recognition task we relied on transfer learning based on the ResNet [He et al. 2016] DCNN architecture[1]. In the preliminary experiments, ResNet50, ResNet101 and ResNet152 were fine-tuned with the individual application of each geometric augmentation technique: Translate, Crop, Flip and Rotate. The main goal of this phase was to assess the impact of the depth and the number of pre-adjusted (non-trainable) ImageNet weights over the results in order to select the most promising architecture for further analyzes. For that, we compared the performance of the baseline (training without augmentation) versus training also with the images generated by the DA technique. The preliminary experiments enabled selecting the better performing architecture to conduct the extended analysis. In this further analysis, we assessed 10 additional DA policies obtained through the sequential and parallel heuristics.

For an strict effectiveness assessment, the performance comparison relied on an independent test dataset. For test images, the same data augmentation policies were applied. Considering recognition is performed for 1+N images (1 original + N augmented versions), the class prediction is performed according to average Softmax scores. A summary of the DA policies assessed and the resulting amount images and augmentation factor are presented in Table 1. For these techniques, the following configurations were considered: Translate (4 different directions with offsets equivalent to 20% of the image width over the horizontal axis and 20% of the height over the vertical axis with black pixel padding); Rotate (30 degrees clockwise and counterclockwise); Crop (four corner patches and a central crop with 50% of the image size).

Table 1. Augmentation policies and resulting number of images. Bold values indicates the size in relation to the original dataset. *Augmentation factor in relation to the original dataset.

Dataset	Train+Validation	Test Set (*)
Original Dataset	20.985	5092
Flip	41.970	10.184 (**2x**)
Rotate	62.955	15.276 (**3x**)
Flip x Rotate	62.955	15.276 (**3x**)
Flip + Rotate	83.940	20.368 (**4x**)
Translate	104.925	25.460 (**5x**)
Crop	125.910	30.552 (**6x**)
Flip x Crop	125.910	30.552 (**6x**)

Dataset	Train+Validation	Test Set (*)
Flip + Translate	125.910	30.552 (**6x**)
Flip + Crop	146.895	35.644 (**7x**)
Translate + Rotate	146.895	35.644 (**7x**)
Crop + Rotate	167.880	40.736 (**8x**)
Translate x Rotate	188.865	45.828 (**9x**)
Translate + Crop	209.850	50.920 (**10x**)
Translate x Crop	440.685	106.932 (**20x**)

The evaluation was conducted with a stratified random sampling protocol to keep class proportion in training and validation sets. More specifically, we considered 80% of the training data for model construction and the 20% for validation, followed by the testing with a an independent held-out test set. To ensure comparability, instead of performing individualized optimizations, all models were trained with the same configurations, for a fixed number of epochs (75) and amount of trainable weights. The Categorical Crossentropy loss function and the Accuracy measure were considered. On the test phase, the Micro-F1 measure was also computed to account for class imbalance. The weight update strategy used was the SGD with the Adam optimizer [Kingma e Ba 2017] and a Batch size of 64. First and Second Moment exponential decays and Epsilon were set to default (0.9, 0.99 and 10^{-7}, respectively), and the learning rate was set to 2×10^{-6}.

[1]ResNet is a recognized effective network, having achieved the first place at the ILSVRC 2015 classification task with 3.57% error on the ImageNet test set [Russakovsky et al. 2015].

4. Results And Discussions

At the preliminary step, we tested the models trained over the individual application of Crop, Translate, Flip and Rotate techniques. The results are presented in Table 2. The models trained over the augmented datasets yielded expressively superior results in comparison to the baseline. It demonstrates how promising these techniques are, with emphasis to the Crop and Translate techniques, allowing Micro-F1 over 0.46 with the ResNet152 model. Considering the network depths (50, 101 and 152 layers), increasing the number of pre-trained weights did not demonstrated to significantly affect the results. Despite the ResNet152 presented overall higher performance, it was not expressive enough in relation to the 50 and 101 variations. Therefore, the ResNet50 was chosen for the further analysis, since it presents significantly lower computational burden in comparison to ResNet101 and ResNet152 [He et al. 2016].

Table 2. Preliminary Results (Micro-F1 on test set).

Network	Baseline	Crop	Translate	Flip	Rotate
ResNet50	0.3177	0.4636	0.4550	0.4094	0.4049
ResNet101	0.3157	0.4693	0.4705	0.4108	0.4065
ResNet152	0.3189	0.4738	0.4652	0.4049	0.4102

For a deeper analysis with the ResNet50, we assessed 10 extra combinations of augmentation policies considering the parallel and sequential heuristics. The results presented in Table 3 demonstrates that DA led to equivalent or superior performance in relation to the baseline. In terms of heuristics, the parallel application policies demonstrated superiority in relation to the sequential and individual approaches. Regarding the sequential application, for example, the overall worst performing policy (Flip + Rotate, Micro-F1=0.4255) enabled similar results to the best performing Sequential policy (Flip x Crop, Micro-F1=0.4271). Considering that the chained (sequential) application of multiple geometric augmentation policies increases the chances of violating the label-preservation principle of data augmentation, we believe that this effect degraded the performance of the models trained with data generated with that heuristic.

Table 3. ResNet50 results: Combination heuristics, Augmentation Policies and performance (Micro-F1 on test set).

Baseline	
Augmentation	**Micro-F1**
None (Original data)	0.3177
Individual Application	
Crop	**0.4636**
Translate	0.4550
Flip	0.4094
Rotate	0.4049
Sequential Application	
Flip x Crop	**0.4271**
Translate x Rotate	0.4010
Flip x Rotate	0.3800
Translate x Crop	0.3189

Parallel Application	
Augmentation	**Micro-F1**
Translate + Crop	**0.4919**
Flip + Crop	0.4672
Crop + Rotate	0.4573
Translate + Rotate	0.4522
Flip + Translate	0.4518
Flip + Rotate	0.4255

Considering the number of resulting images (Table 1), our findings demonstrate that the DA policies that yielded better results not necessarily were the ones that presented higher augmentation factor. The *Translate x Crop* policy, for example, increased the original dataset to over 440.000 images (20x factor) and yet presented only 0.37% gain in relation to the Baseline. In contrast to that, the single application of Flip (2x factor) enabled improving classification performance by roughly 28%. In general, the combination of techniques that presented reasonable individual performance through parallel heuristics allowed superior results in comparison to the individual and sequential.

5. Conclusions

Data augmentation was investigated in this work for leveraging DCNN models for plant species recognition from images. It included a broad assessment of data augmentation policies and combination heuristics for the mostly frequent techniques in the literature. Despite the difficulties behind the process of classification from imbalanced, heterogeneous multi-organ plant datasets, the most effective augmentation policies evaluated allowed expressive results with significant performance gains.

Our results can guide future development of DCNN-based plant recognition systems from imbalanced datasets. We believe that systems like these can play an important role in the development of novel production models, capable of improving precision and automatizing several tasks in fields such as agriculture, pharmacology, ecology, environmental engineering, etc. Finally, we believe novel experiments can be conducted to more deeply analyze the DA potential on overcoming class imbalance.

6. Acknowledgement

This work was partially supported by the National Council for Scientific and Technological Development (CNPq grant no. 158958/2020-9) and a Quadro® P6000 GPU donation by NVIDIA™ Corporation.

References

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., e Moussaoui, A. (2018). *Deep Learning for Plant Diseases: Detection and Saliency Map Visualisation*, pages 93–117. Springer International Publishing, Cham.

Buda, M., Maki, A., e Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249 – 259.

Goëau, H., Bonnet, P., e Joly, A. (2018). Overview of expertlifeclef 2018: how far automated identification systems are from the best experts?

Goëau, H., Joly, A., Bonnet, P., Bakic, V., Barthélémy, D., Boujemaa, N., e Molino, J.-F. (2013). The imageCLEF Plant Identification Task 2013. *Proceedings of the 2nd ACM International Workshop on Multimedia Analysis for Ecological Data*, (i):23–28.

Haupt, J., Kahl, S., Kowerko, D., e Eibl, M. (2018). Large-scale plant classification using deep convolutional neural networks. In *CLEF (Working Notes)*.

He, K., Zhang, X., Ren, S., e Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., e Saba, T. (2018). Ccdf: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep cnn features. *Computers and electronics in agriculture*, 155:220–236.

Kingma, D. P. e Ba, J. (2017). Adam: A method for stochastic optimization.

Mehdipour Ghazi, M., Yanikoglu, B., e Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*.

Pandian, J. A., Geetharamani, G., e Annette, B. (2019). Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, pages 199–204. IEEE.

Pawara, P., Okafor, E., Schomaker, L., e Wiering, M. (2017). Data augmentation for plant classification. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 615–626. Springer.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., e Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Shorten, C. e Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60.

Sulc, M., Picek, L., e Matas, J. (2018). Plant recognition by inception networks with test-time class prior estimation. In *CLEF (Working Notes)*.

Taylor, L. e Nitschke, G. (2018). Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1542–1547. IEEE.

Wang, J., Perez, L., et al. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, 11.

Zhang, C., Zhou, P., Li, C., e Liu, L. (2015). A convolutional neural network for leaves recognition using data augmentation. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 2143–2150. IEEE.

Zhao, B., Li, J., Baenziger, P. S., Belamkar, V., Ge, Y., Zhang, J., e Shi, Y. (2020). Automatic wheat lodging detection and mapping in aerial imagery to support high-throughput phenotyping and in-season crop management. *Agronomy*, 10(11).



Explorando técnicas de aprendizado de máquina para aprimoramento da previsão de geadas no sul e sudeste do Brasil

José Roberto Motta Garcia

Instituto Nacional de Pesquisas Espaciais (INPE)
Cachoeira Paulista – SP – Brasil

roberto.garcia@inpe.br

Abstract. *The CPTEC/INPE operational frost forecast is based on an index established from the weighted variation of meteorological variables from a weather forecasting model, which is compared to the operationally predicted values. Constant improvement of these forecasts is necessary to mitigate the economic and social damage caused by frosts. The objective of this work is to identify a methodology based on machine learning that surpasses the current prediction system. Therefore, this exploratory research applied different combinations between three algorithms, varying their parameterizations and the set of predictor variables from meteorological station data in southern and southeastern Brazil. Preliminary tests on 24-hour forecasts show that performance surpasses the current method, signaling that the search for the most suitable methodology should be continued.*

Resumo. *A previsão operacional de geadas do CPTEC/INPE é baseada em um índice estabelecido a partir da variação ponderada das variáveis meteorológicas de um modelo de previsão de tempo, que é comparado aos valores previstos operacionalmente. Um constante aprimoramento dessas previsões visando mitigar danos econômicos e sociais decorrentes de geadas se faz necessária. O propósito deste trabalho é identificar uma metodologia baseada em aprendizado de máquina que supere o desempenho do sistema de previsão atual. Para tanto, esta pesquisa exploratória aplicou diferentes combinações entre três algoritmos, variando suas parametrizações e o conjunto de variáveis preditoras de dados de estações meteorológicas do sul e sudeste do Brasil. Testes preliminares em previsões de 24h mostram que o desempenho supera o método atual, sinalizando que o processo de busca da metodologia mais adequada deve ser continuada.*

1. Introdução

Geadas causam impacto na agricultura e refletem na economia (Aguiar et al., 2004). Tecnicamente, uma geada é dada como ocorrida quando há formação de cristais de gelo nas superfícies de plantas e objetos expostos ao relento (Pereira et al., 2002). A definição de geada neste trabalho segue o critério estabelecido do Índice de Geadas do CPTEC de Rozante et al. (2020) (IG-CPTEC, de agora em diante) que considera a temperatura de 6°C como limiar para a ocorrência ou não de geadas. Como parte do protocolo de proteção de lavouras, prever ocorrências de geadas se apresenta como um fator de grande importância para minimizar esses impactos.

O IG-CPTEC é estabelecido a partir de pesos decorrentes de cálculos sobre as médias e desvios padrão das principais variáveis meteorológicas que favorecem ou se contrapõem à ocorrência de geadas, apenas para os casos em que elas ocorrem, conforme definido pela metodologia. Essas variáveis são extraídas do modelo regional Eta (Mesinger et al., 1988; Black, 1994; Chou et al., 2002) que são submetidas a um processo de calibração. Este processo sinalizou que a temperatura tem a maior contribuição para a ocorrência de geadas, seguida pela pressão e ventos. As demais variáveis foram ajustadas para que a soma de pesos fosse igual a 1.

A metodologia de predição aqui proposta usa técnicas de aprendizado de máquina (do Inglês Machine Learning, ML daqui em diante) para construir modelos estatísticos a partir de dados históricos visando prever geadas por classificação binária, assumindo 1 para a ocorrência ou 0 para a não ocorrência. Como este trabalho visa melhorar o IG-CPTEC foram mantidas todas as premissas nele contidas e dados por ele utilizados. Uma breve revisão bibliográfica mostra que vários algoritmos de aprendizado de máquina têm sido usados para esta tarefa, como em Naing e Htike (2015) e em Robinson e Mort (1997).

Como exploração foram usados os algoritmos Florestas Aleatórias (RF) de Liaw e Wiener (2002), Redes Neurais Artificiais (ANN) de Haykin (1999) e uma variação de Gradient Boosting Machine (GBM) de Friedman (2001) adaptados para classificação binária. Seus hiperparâmetros, além de variações do conjunto de preditores utilizados para treinamentos dos modelos também foram explorados. O resultado final mostra que a modelagem é bastante promissora pois supera o IG-CPTEC nas previsões de 24h em todos os experimentos.

2. Metodologia

2.1. Dados

Neste trabalho foram avaliadas apenas as previsões de 24h do modelo regional Eta. As variáveis utilizadas como preditoras são: latitude e longitude da estação meteorológica (LON, LAT); temperatura em 2m (TP2M), pressão ao nível médio do mar (PNMM), magnitude do vento em 10m (V10M), nebulosidade (NEBUL) e umidade relativa (UMID) previstas, além da topografia (altitude) da localização da estação meteorológica (TOPO). Além disso há foi atribuído 1 quando houve geada e 0 quando não houve geada na localidade e data válida da previsão, dado que foi reportado pelo Instituto Nacional de Meteorologia (INMET). Os dados de treinamento compreendem 5 anos de

previsão (2012 a 2016), dos meses de maio a setembro, totalizando 244.096 casos de treino. Para os dados de avaliação (teste) foram usadas as previsões do ano de 2017 para os mesmos meses, totalizando 47.112 casos. A distribuição geográfica das estações meteorológicas é mostrada na Figura 1, que mostra também as regiões utilizadas para avaliar o IG-CPTEC.

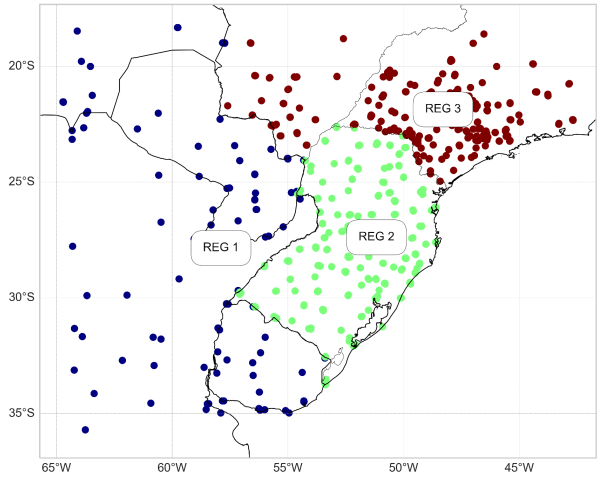


Figura 1. Domínio espacial

2.2. Métricas

As métricas aplicadas na avaliação são baseadas na matriz de confusão para classificação binária, que mostra os acertos e erros das estimativas (previsões) perante a real ocorrência ou não do evento (a geada). A partir dessa matriz de confusão podem ser calculados vários índices estatísticos (Stephenson, 2000 e Wilks, 2011) mas, para facilitar a visualização do desempenho, foi utilizado o diagrama de desempenho (em inglês, performance diagram) (Roebber, 2009), que sintetiza algumas desses índices num único diagrama, seguindo a metodologia praticada em IG-CPTEC.

2.3. Pesquisa exploratória: algoritmos de classificação binária e hiperparâmetros

Um objetivo secundário deste trabalho é descobrir a configuração ótima dos hiperparâmetros das funções que implementam os algoritmos. Foi feita uma exploração exaustiva *ad-hoc* pois podem influenciar positiva ou negativamente o modelo quando aplicadas a diferentes problemas (Yang L. e Shami, A. (2020). Toda a modelagem foi feita em Python (Python Org, 2021) e os algoritmos, explorados através do pacote Scikit-learn (Pedregosa et al., 2011), conforme descrito abaixo:

- **Random Forest (RF):** n_estimators (100, 200, 500, 1000, 1500, e 2000); max_leaf_nodes (None, 5, 25, 50, 100, 250, 500, 600, 700, 800, 1000, 1500, e 2000), e class_weight ("*balanced*" e "*balanced_subsample*").
- **XGBoost:** n_estimators (100, 200, 500, 1000, 1500, e 2000); max_depth (6, 8, e 10), e learning_rate (0.1, 0.05, 0.01, 0.005, e 0.001)

- **Redes Neurais Artificiais (ANN):** hidden_layer_sizes ([6,3], [6,6], [6,12], [12,6], e [12,6,3]); activation "relu" e "logistic"; solver ("sgd" e "lbfgs"), e learning_rate ("adaptive" e "invscaling")

2.4. Exploração dos preditores

Embora o IG-CPTEC tenha se baseado nas variáveis: TP2M, PNMM, V10M, UMID, e NEBUL para ser calculado, este trabalho também explorou variações deste conjunto de preditores, com o acréscimo de TOPO, visando verificar a influência de cada um deles sobre o resultado final dos modelos. Dessa maneira, houve a exploração do uso de todos estes preditores como entrada dos algoritmos até que restasse apenas um. O método de eliminação se baseou nos resultados obtidos em cada treinamento onde sempre o preditor menos influente era eliminado. Como mostrado na Figura 2.

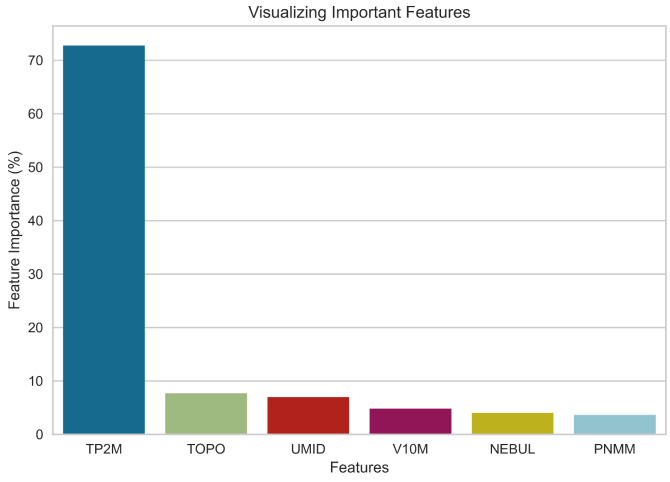


Figura 2. Exemplo da influência de cada preditor nos modelos estatísticos

2.5. Resultados da exploração de algoritmos e preditores

Para facilitar a visualização dos resultados, assume-se que os preditores TP2M, PNMM, V10M, NEBUL, UMID e TOPO são representados pelas letras a, b, c, d, e, e f. As Tabelas 1, 2 e 3 mostram os três melhores resultados da exploração dos algoritmos selecionados, dos seus hiperparâmetros e dos preditores. Há também o tempo gasto em segundos em cada treinamento (SEG).

Tabela 1. Rank dos três primeiros resultados da exploração via RF

PREDITORES	N_ESTIM	MAX_NODES	CLASS_WEIGHT	HR	SEG
a+b+c+d+e+f	1500	None	bal_subsample	0.963195	222
a+b+c+d+e+f	500	None	bal_subsample	0.963195	86

a+b+c+d+e+f	200	None	bal_subsample	0.963174	39

Tabela 2. Rank dos três primeiros resultados da exploração via XGBoost

PREDITORES	N_ESTIM	MAX_DEPTH	LEARN_RATE	HR	SEG
a+b+c+d+e+f	1500	8	0.010	0.967419	737
a+b+c+d+e+f	2000	8	0.010	0.967398	881
a+b+c+d+e+f	2000	10	0.005	0.967185	1426

Conforme pode ser visto nas Tabelas 1 e 2, os melhores resultados, ordenados de forma decrescente pelo Hit Rate (HR), foram conseguidos usando todos os preditores. Dessa maneira os testes utilizando redes neurais foram feitos sem variação nos preditores, uma vez que havia fortes indícios de que a utilização de todos os preditores era a melhor opção. A Tabela 3 mostra as configurações dos hiperparâmetros utilizados nos três melhores resultados utilizando ANN.

Tabela 3. Rank dos três primeiros resultados da exploração via ANN

HIDDEN_LAYERS	ACTIV	SOLVER	LEARN_RATE	HR	SEG
(6, 12)	relu	lbfgs	adaptive	0.962664	667
(6, 12)	relu	lbfgs	invscaling	0.962664	507
(6, 12)	logistic	lbfgs	invscaling	0.962643	521

A melhor dentre todas as variações exploradas foi conseguida via XGBoosting com a seguinte configuração de hiperparâmetros: N_ESTIMATORS=1500, MAX_DEPTH=8 e LEARNING_RATE=0.01, que obteve o melhor Hit Rate (HR) de 0.967419 levando 737s para ser treinado. Portanto esta configuração foi considerada para realizar as avaliações finais. Além disso, como já informado, os melhores resultados foram conseguidos usando todos os preditores. Embora os resultados sejam muito similares, XGBoost foi o algoritmo que obteve o melhor desempenho, medido pelo melhor Hit Rate (HR) e foi a opção considerada para as avaliações finais.

2.6. Experimentos

Visando descobrir a melhor maneira de agrupar os dados de entrada dos algoritmos, foram realizados quatro experimentos. Os resultados são apresentados por região, como feito em IG-CPTEC, e também forma generalizada do modelo para todo o domínio. Os seguintes experimentos foram conduzidos:

- EXP 1 - ONE-FITS-ALL: Um modelo XGBoost genérico, aplicado a todos os dados de teste. Neste experimento, todos os dados de treinamento foram submetidos para construir o modelo, sem qualquer distinção ou classificação, resultando em apenas um modelo. Visa verificar se existe um comportamento

generalizado da atmosfera quando há ou não há geada e se este comportamento foi capturado pelo modelo. A Figura 3a mostra a distribuição espacial considerada.

- EXP 2 - BY-REGION: Um modelo XGBoost para cada uma das três regiões definidas em IG-CPTEC, aplicado de forma distinta a todos os dados de teste, conforme a região que o dado se encontra, verificado pelo par LON+LAT. Neste experimento, os dados de treinamento e teste foram separados pela região que se encontram e, em seguida, submetidos para construir os modelos, resultando em três modelos estatísticos. Visa verificar se existe um comportamento específico da atmosfera para cada região definida em IG-CPTEC e se este comportamento foi capturado pelo modelo. A Figura 1 mostra as regiões consideradas.

Visando encontrar similaridades "comportamentais" entre os dados de treinamento, foram utilizadas duas abordagens de agrupamento dos preditores: por semelhança climática (CLUST-CLIM) e por semelhança entre localidades (CLUST-LOC). Foi utilizado o algoritmo KMeans do pacote Scikit-learn e a função z-score do pacote Scipy para normalizar os dados. A determinação da quantidade ideal de *clusteres* foi feita pelo método *elbow* (cotovelo), apresentado por Thorndike (1953). O processo resultou numa divisão dos dados em oito grupos, tanto para CLUST-LOC como para CLUST-CLIM, que foram associados aos dados de entrada. Conforme a Figura 3b, a linha azul refere-se a curva dos resultados do grau de distorção (escala azul à esquerda) obtidos pelo modelo para cada valor de K; o ponto de inflexão está anotado com uma linha tracejada (e também informado no canto superior direito do gráfico); a linha verde refere-se a escala de tempo gasto para resolver o modelo (escala verde à direita).

- EXP 3 - CLUST-CLIM: Um modelo XGBoost para cada um dos oito clusteres de comportamento climático similar, considerando todas as variáveis, exceto LON e LAT. Cada caso de treinamento foi associado a um dos oito grupos estimados para CLUST-CLIM e, em seguida, treinados, resultando em oito modelos estatísticos. Para testar a modelagem, cada caso de teste foi comparado aos centróides dos oito clusteres criados e foi usado o modelo estatístico referente ao centróide mais semelhante ao dado, medido por distância euclidiana. Visa verificar se a abordagem consegue distinguir comportamentos climáticos considerando cada medição da atmosfera. Não há representação espacial desse agrupamento.
- EXP 4 - CLUST-LOC: Um modelo XGBoost para cada uma das oito localidades de comportamento climático similar, considerando a média das variáveis para cada par único de LON+LAT, o que resultou em 593 diferentes pares LON+LAT. Cada caso de treinamento foi associado a um dos oito grupos estimados para CLUST-LOC, e em seguida treinados, resultando em oito modelos estatísticos. Para testar a modelagem, cada caso de teste foi comparado aos pares únicos de LON+LAT do *dataset* de treinamento e foi usado o modelo estatístico referente ao par de LON+LAT mais próximo do caso de teste, medido por distância euclidiana. Visa verificar se a abordagem consegue distinguir zonas climáticas de acordo com o

comportamento médio das estações meteorológicas. A Figura 3c mostra a clusterização das zonas climáticas de CLUST-LOC.

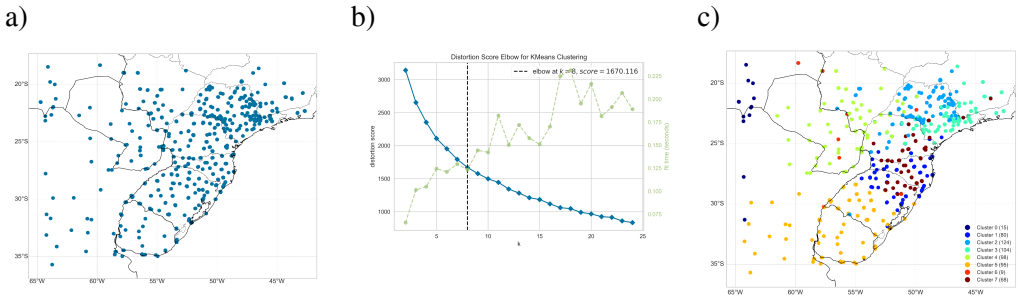


Figura 3. Domínio espacial do experimento ONE-FITS-ALL (a); exemplo de saída do método "Elbow" (b), e clusterização resultante da similaridade encontrada entre os pares LON+LAT únicos em CLUST-LOC (c).

3. Resultados

O desempenho comparativo entre IG-CPTEC e os experimentos deste trabalho pode ser analisado comparando as Figuras 4 e 5, referentes aos diagramas de desempenho das metodologias. Esse diagrama facilita a análise dos resultados, pois resume várias métricas obtidas a partir de estimativas dicotômicas (sim ou não) como: POD (Probability of Detection), FBias (Frequency Bias), CSI (Critical Success Index), FAR (False Alarm Ratio) e SR (Success Ratio). As métricas POD, CSI e SR são medidas de sucesso no desempenho então quanto maior seu valor melhor. POD pode ser visualizado no eixo vertical esquerdo, CSI é representado pelas curvas que convergem para o canto superior direito e sua escala é posicionada no eixo vertical direito. SR é representado pelo posicionamento do marcador dentro do gráfico. FBias é representado pelas linhas radiais de origem em (0,0) e indicam a frequência que as estimativas são sub ou superestimadas. Como FAR é uma medida de erro, o seu valor é subtraído de um para ser posicionado no eixo horizontal inferior, fazendo com que quanto mais à direita esteja melhor seja. Em resumo, quanto mais próximo o marcador da estimativa estiver posicionada do canto superior direito e sobre a diagonal principal melhor ela é.

A Figura 4 mostra os três Diagramas de Desempenho mostrados em IG-CPTEC, para a Região 1 (4.b), Região 2 (4.d) e Região 3 (4.f). Em cada um deles o círculo aberto menor representa as previsões de 24h, nas quais este trabalho foi baseado. Os outros círculos abertos maiores se referem a períodos de previsão até 120h, não considerados neste trabalho. Já os círculos fechados em cinza referem-se ao desempenho do modelo Eta, caso fosse considerado apenas a sua temperatura para prever a ocorrência ou não de geadas. A interpretação é que o IG-CPTEC conseguiu melhorar esta previsão através de seu índice.

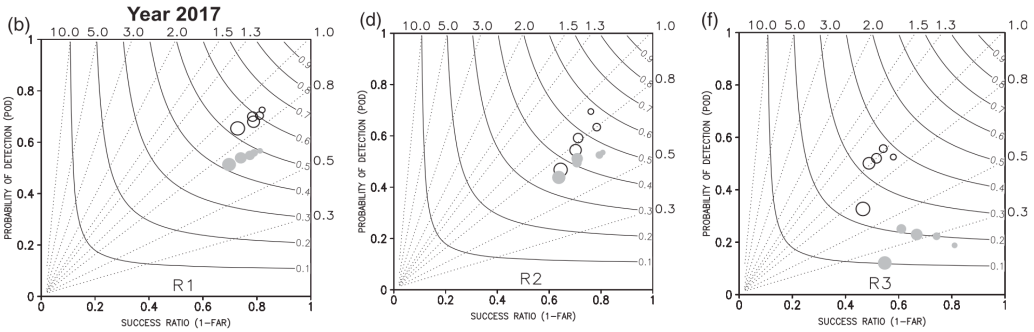
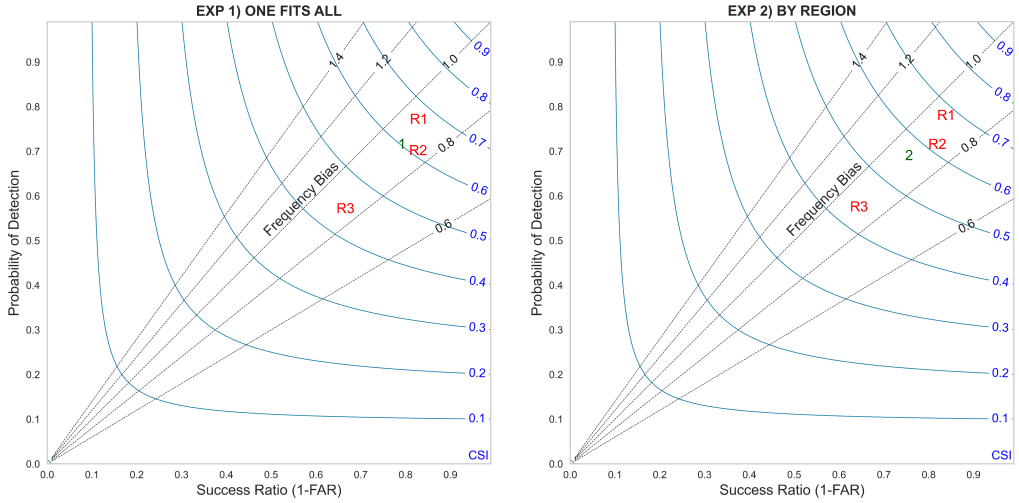


Figura 4. Resultado do desempenho do IG-CPTEC comparado ao modelo Eta, adaptado de Rozante et al. (2020, Figure 4)

A Figura 5 mostra um Diagramas de Desempenho para cada experimento deste trabalho: ONE-FIT-ALL (EXP 1), BY-REGION (EXP 2), CLUST-CLIM (EXP 3), e CLUST-LOC (EXP 4). Para todos os gráficos: O número em verde escuro representa o desempenho global referente ao experimento (1 a 4); em vermelho é mostrado o desempenho das regiões do IG-CPTEC (R1, R2 e R3); e em verde claro o desempenho dos clusteres formados no experimento (quando pertinente).



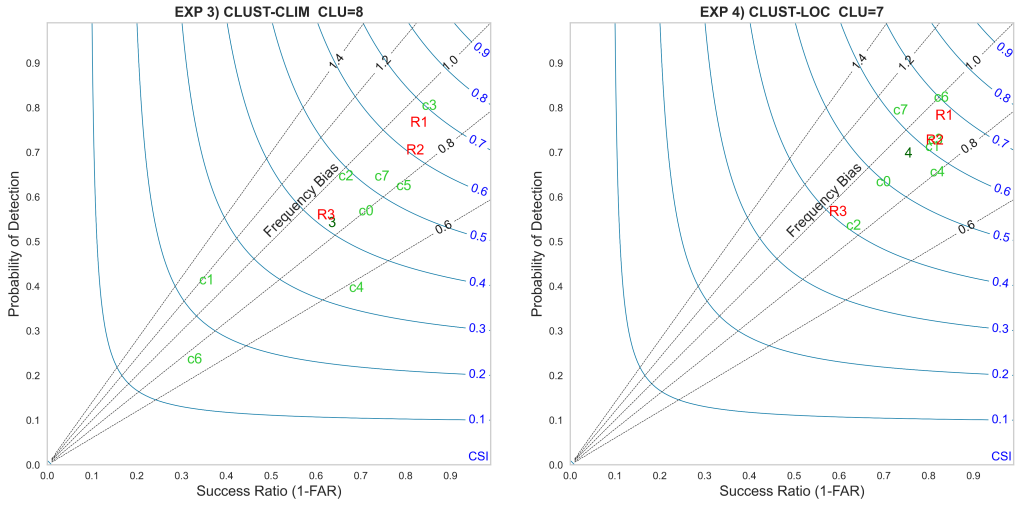


Figura 5. Resultado do desempenho dos experimentos 1 a 4

4. Conclusões

Usar aprendizado de máquina para a tarefa de previsão de geadas se mostra muito promissor. Os diagramas de desempenho mostraram que os quatro experimentos superam o IG-CPTEC para as previsões de 24h. Em relação ao desempenho global (números 1 a 4 em verde escuro), verifica-se que o ONE-FITS-ALL teve melhor desempenho, em seguida aparecem o BY-REGION e o CLUST-LOC praticamente idênticos. Por região, BY-REGION e CLUST-LOC desempenharam melhor que os demais nas regiões 1 e 2, enquanto que para a região 3 os experimentos BY-REGION e ONE-FITS-ALL se destacaram.

A região 3 apresenta-se sempre com desempenho inferior às demais, indicando que merece maiores estudos. O cálculo de importância dos preditores confirma que a temperatura é o aspecto que mais influencia na previsão de geadas, mas todos contribuem para a modelagem e não devem ser ignorados, conforme teste exploratório. Embora menos acentuado para a Região 3, os experimentos continuam tendendo a prever menos ocorrência de geadas como é o IG-CPTEC, que talvez seja explicado pelo fato de que desbalanceamento entre as ocorrências da classe 0 (sem geada) e da classe 1 (com geada) possa afetar a modelagem.

Os resultados motivam estudos futuros e mostram que não se pode descartar nenhuma das abordagens. Tarefas futuras incluem: extensão da metodologia para outros períodos de previsão; considerar outras variáveis preditoras, testar em outros *datasets* mais atuais e de outros modelos, estimar geada em multiclasses, usar modelos com capacidade de memória, fazer análise outras análises de relevância das variáveis, etc.

5. References

Aguiar, D., Mendonça, M. (2004) Climatologia das geadas em Santa Catarina. In: Simpósio Brasileiro de Desastres Naturais. Anais. Florianópolis.

Black, T.L. (1994) The new NMC mesoscale Eta model: description and forecast examples. Weather and Forecasting, 9, 265–278.

Chou, S.C., Tanajura, C.A.S., Xue, Y. and Nobre, C.A. (2002) Valida- tion of the coupled Eta/SsiB model over South America. Journal of Geophysical Research: Atmospheres, 107(D20), 8088. https://doi. org/10.1029/2000JD000270.

Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29, 1189-1232. https://doi.org/10.1214/aos/1013203451

Haykin, S. (1999) Neural Networks: A Comprehensive Foundation, 842 pp., Prentice-Hall, Old Tappan, N. J.

Liaw, A., Wiener, M., (2002) Classification and Regression by random forest. R. News 2, 18–22.

Mesinger, F., Janji?c, Z.I., Ni ckovi?c, S., Gavrilov, D. and Deaven, D.G. (1988) The step-mountain coordinate: model description and per- formance for cases of alpine lee cyclogenesis and for a case of an Appalachian redevelopment. Monthly Weather Review, 116, 1493–1518.

Naing, W.Y.N., Htike, Z.Z., (2015) Forecasting of monthly temperature variations using random forests. ARPN J. Eng. Appl. Sci. 10, 10109–10112.

Pedregosa et al. (2011) Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830.

Pereira, A. R., Angelocci, L. R., Sentelhas, P. C. (2002) Agrometeorologia: fundamentos e aplicações práticas. Guaíba: Agropecuária. 478p.

Robinson, C., Mort, N., (1997) A neural network system for the protection of citrus crops from frost damage. Comput. Electron. Agric. 16, 177–187. https://doi.org/10.1016/ S0168-1699(96)00037-3.

Roebber, P.J. (2009) Visualizing multiple measures of forecast quality. Weather and Forecasting, 24, 601–608.

Rozante J.R., Gutierrez E.R, da Silva Dias P.L., de Almeida Fernandes A., Alvim D.S., Silva V.M. (2020) Development of an index for frost prediction: Technique and validation. Meteorol Appl. 2020;27:e1807. https://doi.org/10.1002/ met.1807

Stephenson, D.B. (2000) Use of the “odds ratio” for diagnosing fore- cast skill. Weather and Forecasting, 15, 221–232.

Wilks, D.S. (2011) Statistical methods in the atmospheric sciences. Oxford and Waltham, MA: Academic Press and Elsevier Science.

Yang L., Shami, A. (2020) On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing, Volume 415, Pages 295-316, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.07.061.



H-Pasture: uma aplicação móvel para estimativa da disponibilidade de forragem e ajuste de carga em pastagens sob lotação contínua ou rotativa

Wendell Gasparoni[1], Ana Paula Lüdtke Ferreira[1], Teresa Cristina Moraes Genro[2], Márcia Cristina Teixeira Silveira[2]

[1]Universidade Federal do Pampa
Bagé – RS – Brasil

[2]EMBRAPA Pecuária Sul
Bagé – RS – Brasil

wendellgasparoni@gmail.com, anaferreira@unipampa.edu.br,
{cristina.genro,marcia.teixeira}@embrapa.br

Abstract. *This article introduces the mobile application H-Pasture, which uses linear regression models to provide tools for helping farmers and rural technicians perform grazing management activities using the pasture height metric. H-Pasture uses Google's Flutter/Dart technologies, makes forage weight estimations within the registered areas, builds reports, and produces recommendations for stocking rate under rotative or variable rate continuous pasture.*

Resumo. *Este trabalho apresenta o aplicativo móvel H-Pasture, que usa modelos de regressão linear para fornecer ferramentas que auxiliem produtores e técnicos rurais no manejo de pastagem por meio da métrica de altura do pasto. O aplicativo foi desenvolvido com a tecnologia Flutter/Dart e faz estimativas de massa de forragem nas áreas cadastradas, gera relatórios e produz recomendações para auxiliar no ajuste de carga em pastagens sob pastejo rotativo ou contínuo com taxa variável.*

1. Introdução

Modelos de negócio baseados em pecuária extensiva dependem de uma boa estimativa de disponibilidade de alimento no pasto, para que a movimentação de animais entre potreiros possa ser realizada com precisão, garantindo bons resultados no processo de produção animal. Contudo, as pastagens são um ecossistema dinâmico e o manejo do pasto ainda é apontado por produtores e técnicos que utilizam esse ecossistema para a produção animal como desafiador, pois orientações e definições relativas à lotação dos piquetes e

à definição do momento de entrada e saída dos animais nas áreas são pouco precisas [Genro e da Silveira 2018]. Frequentemente são observadas situações de superpastejo ou subpastejo, ambas acarretando em baixo desempenho animal, seja por produção vegetal aquém do potencial e degradação da pastagem, no primeiro caso, ou desperdício de forragem, no segundo [Hodgson 1990].

As pesquisas na área de manejo de pastagens têm trabalhado, principalmente, com dois conceitos de utilização do pasto: o primeiro diz respeito à oferta de forragem ou intensidade de pastejo, propondo o cálculo da oferta da quantidade de pasto existente na área em função de peso vivo animal (quilos de matéria seca ofertados para cada 100 quilos de peso vivo, ou % PV), por um determinado período de tempo [Maraschin 2004]; o segundo é a interceptação luminosa, assumindo que o momento ideal para pastejo é quando o pasto está interceptando, em média, 95% da radiação que chega da atmosfera [Carnevalli et al. 2006, Barbosa et al. 2007]. Ambos os conceitos exigem treinamento das pessoas que vão conduzir as avaliações e equipamentos específicos, os quais dificilmente são encontrados em propriedades rurais.

Com o objetivo de construir um sistema de apoio ao manejo das pastagens, buscou-se um dos componentes da estrutura do pasto que apresentasse uma boa relação com a massa de forragem disponível para pastejo e que, ao mesmo tempo, fosse de mensuração simples. Diversos estudos com produção animal em pastagens cultivadas ou nativas têm apontado que a altura do pasto apresenta uma relação linear positiva com massa de forragem [Abramides et al. 1982, Trindade et al. 2007, Carvalho et al. 2010, Rosa et al. 2017, Genro e da Silveira 2018, Souza Filho et al. 2019]. Essa relação entre a altura e os componentes da estrutura do pasto, principalmente com a massa de forragem, viabiliza a utilização da altura como critério prático para definir o momento ideal de pastejo, bem como permite identificar a necessidade ou não de realização de ajustes de carga animal, visando estabelecer condições ótimas de utilização do pasto mediante os principais processos envolvidos no crescimento e utilização das plantas forrageiras sob pastejo.

Dados de pesquisa com forrageiras utilizadas nos sistemas de produção do Estado do Rio Grande do Sul têm reforçado a possibilidade do uso da altura como um método seguro para auxiliar o produtor no manejo do pasto da propriedade. Segundo Martins et al. (2015), a relação entre altura do pasto e a massa de forragem é uma importante ferramenta de manejo pois, ao controlar a altura, maneja-se indiretamente diferentes componentes do sistema. Adicionalmente, dentro da lógica de produção sustentável, tem sido possível observar que o manejo adequado do pasto contribui com serviços ambientais como a mitigação das emissões de gases de efeito estufa (GEE), por meio da redução das intensidades de emissão de metano por quilo de peso vivo produzido por área. Isso porque quando se usa a meta de altura adequada para manejo do pasto, os desempenhos individuais dos animais em pastejo são mais elevados [Savian et al. 2018] e mais carbono é sequestrado no solo [Franzluebbers e Stuedemann 2009].

Souza Filho et al. (2019) avaliaram o efeito de diferentes intensidades de pastejo por novilhos sobre o desempenho animal, o consumo de forragem e as emissões de CH_4 no período de lotação de um sistema integrado soja-gado de corte no sul do Brasil. Os tratamentos foram diferentes, definidos pelas alturas-alvo de pastejo (10, 20, 30 e 40 cm) em pastagem consorciada de aveia preta (*Avena strigosa Schreb.*) e azevém (*Lo-*

lium multiflorum Lam.), sob lotação contínua. O manejo do pastejo afetou a utilização da forragem e, consequentemente, o desempenho animal e a emissão de metano. Individualmente, a emissão de metano e o desempenho dos animais apresentaram valores ótimos quando a altura do pasto foi manejada dentro de uma faixa de 23 a 30 cm. Os dados agrupados mostraram um efeito linear positivo da intensidade do pastejo no ganho de peso vivo do animal por hectare e os custos ambientais associados ao uso da terra. O ganho de peso vivo aumentou em 90 g/ha/dia e as emissões de metano aumentaram em 500 g CO_2 eq/ha/dia para cada centímetro de redução da altura do pasto alvo. Esses resultados mostram o impacto ambiental, em termos de emissões de GEE, que pode ser gerado sem otimização do manejo do pasto.

Genro e Silveira (2018) indicam a altura como um guia de manejo recomendado para auxiliar produtores e técnicos nas tomadas de decisão relativas ao uso das pastagens. Como a avaliação da altura é rápida, não destrutiva e de fácil entendimento, juntamente com o fato da altura ser um elo de ligação entre plantas, animais e o meio, incrementos significativos em produtividade podem ser alcançados a baixo custo. O objetivo deste trabalho é usar esses resultados para desenvolver um aplicativo que permita aos produtores e técnicos valerem-se das informações teóricas disponíveis para melhorar sua tomada de decisão quanto ao ajuste de carga em pastagens. O modelo implementado usa os valores para o RS, visto que podem haver diferenças regionais e da vegetação envolvida, mas pode ser adaptado para valores de outras localidades.

O restante do texto está organizado como se segue: a Seção 2 discute trabalhos com o mesmo objetivo encontrados na literatura, a Seção 3 apresenta as tecnologias utilizadas no desenvolvimento e as telas e funcionalidades mais relevantes do software e a Seção 4 conclui o texto, discutindo resultados e desenvolvimentos futuros.

2. Trabalhos relacionados

O desenvolvimento deste trabalho foi feito para dispositivos Android, visto que essa plataforma possui mais de 80% em participação de mercado de sistemas operacionais móveis no Brasil [StatCounter 2021]. Sendo assim, a busca por trabalhos correlatos foi orientada para a análise de aplicativos disponíveis na Google Play [Google 2021] e para aplicativos e outras ferramentas de software citadas em fontes especializadas. Com objetivos similares ao deste trabalho foram encontradas as seguintes aplicações: *PastureMap* [PastureMap 2021], *Pasture.io* [Pasture.io 2021], *Grazing Wedge* [University of Missouri 2021], *Grazing Calculator* [South Dakota State University 2021] e *Grazing Calculator App* [Robert Hathaway 2021].

Visto que grande parte das soluções encontradas são pagas, não foi possível a realização de testes com todas as funcionalidades, sendo que em alguns casos, como nas ferramentas Pasture.io e Grazing Calculator App, a análise foi feita baseado-se na documentação da ferramenta no site oficial. A Tabela 1 resume as ferramentas encontradas, sendo a primeira coluna o nome da ferramenta, a segunda a descrição da ferramenta e a terceira o custo para utilização.

Entre as ferramentas analisadas, PastureMap e o Pasture.io são as que disponibilizam o maior número de funcionalidades para os usuários, sendo que o Pasture.io coleta imagens de satélites para a estimativa de forragem, tornando este processo mais prático, porém, resultando em um custo mais alto para utilização. Por fim, considerando

Tabela 1. Trabalhos Correlatos

Ferramenta	Descrição	Custo
PastureMap	Ferramenta com recursos oferecidos para a gestão do manejo de pastagem, possibilitando o cadastro de áreas, animais e extração de relatórios para a tomada de decisão durante o manejo; o algoritmo utiliza os dados do Departamento de Agricultura dos Estados Unidos, sem adequação a outras localidades.	paga
Pasture.io	Ferramenta com recursos oferecidos, com funcionalidades semelhantes ao PastureMap; a estimativa de forragem é calculada a partir de imagens extraídas de satélites, sem detalhes sobre a técnica utilizada.	paga
Grazing Wedge	Desenvolvida e disponibilizada pela Universidade do Missouri, possibilita o cadastro de propriedades rurais e suas áreas de pastejo, calculando taxas de crescimento da pastagem e a produção cumulativa de forragem no sistema de pastejo.	gratuita
Grazing Calculator	Calculadora disponibilizada pela Universidade do Estado de Dakota do Sul por meio de um website, que a partir de dados solicitados de entradas, como produção média de forragem, tipo dos animas, qualidade da forragem, etc. calcula o total de animais permitidos na área ou então a quantidade de dias para manter os animais na área.	gratuita
Grazing Calculator App	Aplicativo disponibilizado para Android ou iOS que produz diferentes cálculos sobre o processo de manejo, como quantidade de dias para manter os animais na área, kg de matéria seca por vaca por dia, entre outros cálculos.	paga

as soluções encontradas e comparando-as com os objetivos deste trabalho, recursos como a utilização de modelos matemáticos para tipos de forragens voltadas para a região da Campanha gaúcha, unidos à simplicidade de utilização e baixo custo, ainda tornam o H-Pasture uma solução interessante para auxiliar produtores e técnicos rurais na tomada de decisão durante o manejo.

3. O aplicativo H-Pasture

O documento de requisitos da aplicação, inicialmente elaborado, engloba todos os requisitos funcionais e não funcionais que a aplicação deve conter, refletindo o método no qual se baseia. Os requisitos do aplicativo envolvem a sua utilização em acordo com as orientações sobre o manejo baseado em altura, que envolvem registros de medições de altura (dependendo da heterogeneidade da estrutura da vegetação), cálculo de massa de forragem disponível e recomendações específicas para evitar sub ou sobrepastejo. Para minimizar a quantidade de dados inseridos, o registro prévio de propriedades, suas áreas internas e tipo de vegetação em cada área é facultado ao usuário.

O levantamento das tecnologias adequadas para o desenvolvimento do aplicativo móvel, a fim de permitir uma maior flexibilidade de utilização para o usuário, destacou quatro abordagens: React Native [Facebook 2015], Ionic [Drifty 2013], Xamarin [Microsoft 2011] e Flutter [Google 2017]. A análise dos requisitos em comparação às

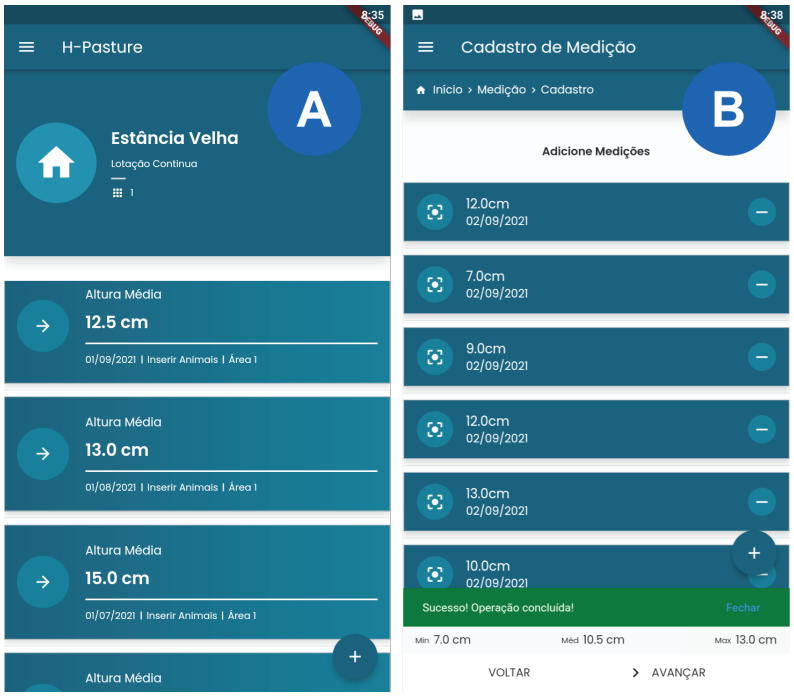


Figura 1. Tela inicial (A) e cadastro de medições (B)

tecnologias levantadas permitiu concluir que qualquer uma delas seria adequada. A tecnologia Flutter foi escolhida por pertencer a uma empresa de desenvolvimento sólida (Google) e pela sua curva de adoção pela comunidade estar aumentando consistentemente. A linguagem Dart, usada para desenvolvimento, contém construções que facilitam a portabilidade para diferentes sistemas operacionais e plataformas. Assim, ainda que o desenvolvimento tenha sido feito e testando no sistema operacionado Android, a migração para o sistema IOS pode ser feita sem alterações no código-fonte. Para possibilitar a utilização *offline*, foi utilizada uma base de dados local no dispositivo do usuário, com o banco de dados SQLite [D. Richard Hipp 2000].

O aplicativo contém informações sobre os tipos de pastagens mais comuns na região da Campanha gaúcha. A Tabela 2 apresenta os tipos de pastagem, a quantidade de massa seca estimada para cada centímetro de altura medido, recomendação de entrada e saída de animais, em função da altura do pasto em sistemas de pastejo rotativo e a indicação da revisão da taxa de lotação em áreas com pastejo contínuo.

Tabela 2. Indicações do aplicativo (cf. [Genro e da Silveira 2018])

Pastagem	**Massa (por cm)**	**Rotativo**		**Contínuo**
		Entrada	Saída	Ajuste
Azevém	127 kg/ha	20 cm	8 a 12 cm	15 a 20 cm
Festuca	165 kg/ha	20 cm	6 a 12 cm	–
Capim-sudão	63 kg/ha	60 cm	15 a 20 cm	30 a 40 cm
Pasto nativo	231,5 kg/ha	12 cm	6 cm	8 a 12 cm

Atualmente, as funcionalidades do aplicativo são: cadastro de uma ou mais pro-

priedades rurais com diferentes áreas para pastejo, visualização do histórico de todas as medições já cadastradas, registros de medição de altura, visualização de dados e geração de relatório para auxiliar na tomada de decisão durante o manejo de pastagem.

A tela inicial, ilustrada na Figura 1A, apresenta ao usuário a propriedade rural atualmente selecionada junto a todo histórico de medições efetuadas em cada uma das áreas cadastradas da propriedade. O histórico das medições é apresentado ao usuário por meio de cartões contendo informações sintetizadas da medição: data da medição, nome da área, altura média do pasto e a ação sugerida (inserir ou retirar animais). A qualquer momento o usuário pode iniciar o processo de cadastro de novas medições por meio do botão de adição no canto inferior direito da Figura 1A. O cadastro da medição é feito selecionando a área desejada da propriedade rural e inserindo todas as medições coletadas em campo para a geração da média aritmética (Figura 1B). A média é aritmética porque o número de medições deve ser proporcional à área, levando em conta a heterogeneidade. A tela de cadastro de medição ainda fornece, para fins de referência, os valores parciais da medição até o momento: valor mínimo, valor máximo e a média aritmética das medidas inseridas até o momento.

A Figura 2 mostra a fase final do cadastro de medição da altura do pasto para uma área da propriedade rural, após a inserção das diferentes medições da altura. O aplicativo gera um relatório com ilustração (Figura 2A) das condições do pasto. Como observado, são geradas duas linhas em laranja para demarcar a faixa ótima de uso para os melhores ganhos em termos de produção animal e vegetal. A linha em vermelho demarca a altura média do pasto calculada a partir das medições informadas, e está nessa cor para indicar uma situação fora da faixa ideal de uso do pasto; caso contrário, a linha da média tem a cor azul. Nessa situação pode ocorrer perda de valor nutritivo aproveitado (acima do valor da altura de referência), ou ainda, comprometer a persistência do pasto e ser limitante ao consumo dos animais em pastejo (abaixo do valor da altura de referência).

Após a visualização do relatório apresentado na Figura 2A, o usuário pode gerar outro relatório com maior detalhamento. Ao clicar na seta da parte inferior da tela, o aplicativo apresenta o conteúdo da Figura 2B, com informações sobre limite de referência inferior, limite de referência superior, altura média do pasto e estimativa da massa de forragem disponível. Como pode ser observado, tanto o relatório simplificado da Figura 2A, quanto o detalhado da Figura 2B apresentam ao usuário, na parte inferior da tela, a ação sugerida. A lógica desta funcionalidade é implementada por meio do cruzamento de diferentes informações cadastradas. Como ilustrado na Figura 3A, durante o cadastramento da propriedade rural é informado o tipo de sistema de pastejo utilizado, podendo ser continuo ou rotativo. A Figura 3B ilustra o cadastramento de uma área para a propriedade rural. Nesse cadastro é informado o tipo da forragem da área, podendo ser forragem nativa, festuca, capim-sudão, ou qualquer outra previamente cadastrada. Estas informações, junto à média da altura do pasto coletada durante o registro de medição, são cruzadas e analisadas conforme a Tabela 2. Após a análise, se for observado que a altura está acima do limite máximo, é recomendado um ajuste de carga com a inserção de animais na área. O aplicativo também informa qual o peso vivo por hectare que pode ser inserido, em função do excedente calculado de massa de forragem. Caso a altura esteja abaixo do limite mínimo é sugerido um ajuste através da remoção de animais da área. Os dados informados nas figuras são dados gerados aleatoriamente, para fins de apresentação do

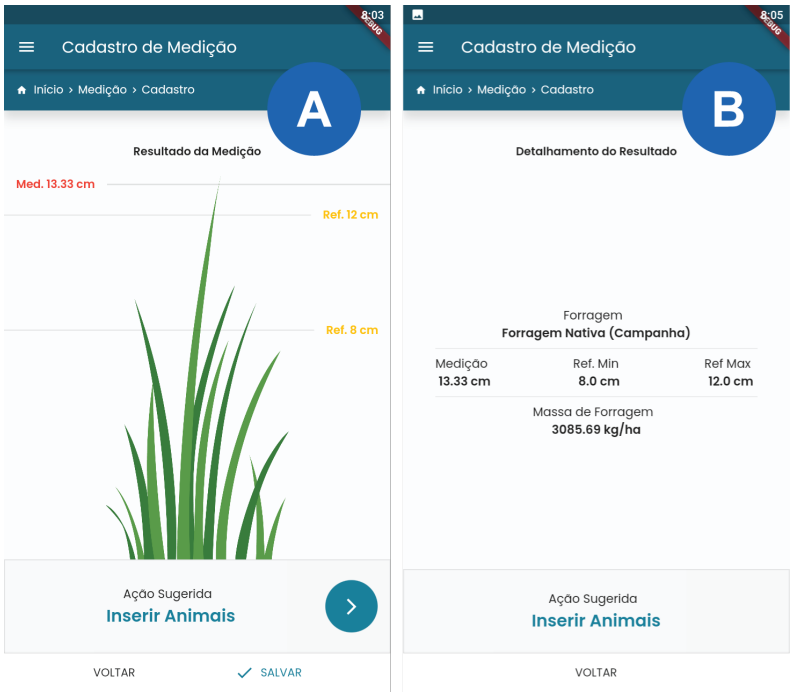


Figura 2. Telas de relatórios

aplicativo neste trabalho.

4. Conclusão

O manejo das pastagens e ajuste de lotação com base na altura do pasto é uma técnica simples que não requer equipamento ou pessoal especializado, ao contrário de outras técnicas de predição de massa de forragem. Ainda que bastante simples, a literatura aponta que é uma técnica que produz boas estimativas.

O H-Pasture é uma aplicação para tornar o processo de manejo de pastagem mais simples para produtores e técnicos rurais. Baseado nos modelos matemáticos e nas orientações publicadas em [Genro e da Silveira 2018] a aplicação fornece, ao produtor ou técnico, ferramentas didáticas para o manejo. A partir do cadastro de medições, a aplicação gera o cálculo de massa de forragem, ajudando o usuário na tomada de decisão durante o manejo de pastagem. O uso da altura do pasto como critério de manejo possibilita um ajuste mais preciso para evitar situações de sobre ou subpastejo, com a vantagem do baixo custo. O aplicativo H-Pasture facilita ainda mais o trabalho, visto que apresenta os cálculos e a recomendação de manejo de forma automática, funcionando sem necessidade de conexão à internet.

O H-Pasture somente possui dados para quatro tipos de pastagens, com dados voltados à região da Campanha gaúcha. Essa restrição não é permanente e o código pode ser facilmente estendido para adicionar mais dados de outras regiões do país, com culturas também pertinentes a essas regiões. Dados de GPS do próprio celular podem ser usados para orientar as recomendações, a partir da inserção de uma base de dados sobre culturas que possa ser acessada via internet.

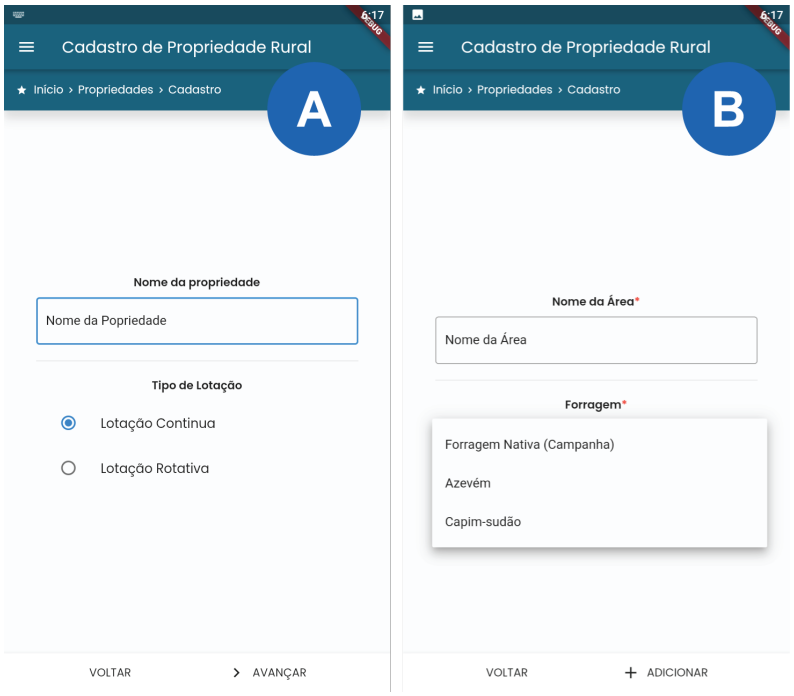


Figura 3. Telas de cadastros

A orientação sobre o processo de medição e sua frequência está sendo colocada no código do aplicativo. Também serão adicionados alertas para informar ao usuário o momento de coletar novas medidas de altura do pasto. A informação será dependente das estações do ano e do tipo de sistema de pastejo usado. Atualmente, o aplicativo está na fase de testes de aceitação, com o *feedback* de usuários especializados. Os testes com os usuários finais serão realizados nos próximos meses, quando o aplicativo será colocado à disposição da comunidade.

Referências

Abramides, P., Alcântara, P., Stafuzza, J., Foschine, A., Dower, J. (1982). Estimativa da quantidade de forragem em pastagens de capins prostrados tropicais, através da medida da altura média da vegetação. 20(1):17–41.

Barbosa, R. A., Nascimento Jr, D., Euclides, V. P. B., Silva, S. C., Zimmer, A. H., Torres Jr, R. A. A. (2007). Capim-tanzânia submetido a combinações entre intensidade e freqüência de pastejo. *Pesquisa Agropecuária Brasileira*, 42(3):329–340.

Carnevalli, R. A., Silva, S. C., Bueno, A. A. O., Uebele, M. C., Bueno, F., Hodgson, J., Silva, G. N., Moraes, J. P. G. (2006). Herbage production and grazing losses in panicum maximum cv. mombaça under four grazing managements. *Tropical Grasslands*, 40:165–176.

Carvalho, P. C. F., Rocha, L. M., Baggio, C., Macari, S., Kunrath, T. R., Moraes, A. (2010). Característica produtiva e estrutural de pastos mistos de aveia e azevém manejados em quatro alturas sob lotação continua. *Revista Brasileira de Zootecnia*, 39(9):1857–1865.

D. Richard Hipp (2000). SQLite. Disponível em: https://www.sqlite.org//. Acesso em: setembro 2021.

Drifty (2013). Ionic. Disponível em: https://ionicframework.com/. Acesso em: setembro 2021.

Facebook (2015). React native. Disponível em: https://reactnative.dev/. Acesso em: setembro 2021.

Franzluebbers, A. J. Stuedemann, J. A. (2009). Soil-profile organic carbon and total nitrogen during 12 years of pasture management in the southern piedmont usa. *Agric. Ecosyst. Environ.*, 129:28–36.

Genro, T. C. M. da Silveira, M. C. T. (2018). Uso da altura para ajuste de carga em pastagens. Comunicado Técnico 101, Embrapa Pecuária Sul.

Google (2017). Flutter. Disponível em: https://flutter.dev/. Acesso em: setembro 2021.

Google (2021). Google play. Disponível em: https://play.google.com/store. Acesso em: 11 out. 2021.

Hodgson, J. (1990). *Grazing management: science into practice*. John Willey, Longman Scientific and Technical, New York.

Maraschin, G. E. (2004). *Estratégias para valorizar sistemas pastoris sob a ótica de políticas de segurança alimentar, bem estar animal e social*, pages 67–83. Number 20. GRUPO CAMPOS, Salto.

Martins, A. P., aise Robinson Kunrath, Anghinoni, I., de Faccio Carvalho, P. C. (2015). *Integração Soja-Bovinos de Corte no Sul do Brasil*. Grupo de Pesquisa em Sistema Integrado de Produção Agropecuária, UFRGS, Porto Alegre.

Microsoft (2011). Xamarin. Disponível em: https://xamarin.com/. Acesso em: setembro 2021.

Pasture.io (2021). Pasture.io. Disponível em: https://pasture.io/. Acesso em: 11 out. 2021.

PastureMap (2021). Pasturemap. Disponível em: https://pasturemap.com/. Acesso em: 11 out. 2021.

Robert Hathaway, L. H. (2021). Grazing calculator. Disponível em: https://www.grazingcalculator.com/. Acesso em: 11 out. 2021.

Rosa, F. Q., Bremm, C., Machado, D. R. (2017). *Efeito da oferta de forragem na estrutura do pasto*, pages 23–25. UFRGS, Porto Alegre.

Savian, J. V., Schons, R. M. T., Marchi, D. E., de Freitas, T. S., Silva Neto, G. F., Mezzalira, J. C., de Faccio Carvalho, P. C. (2018). Rotatinuous stocking: a grazing management innovation that has high potential to mitigate methane emissions by sheep. *J. Clean. Prod.*, 186:602–608.

South Dakota State University (2021). Grazing calculator. Disponível em: https://extension.sdstate.edu/grazing-calculator. Acesso em: 11 out. 2021.

Souza Filho, W., Nunes, P. A. d. A., Barro, R. S., Kunrath, T. R., Almeida, G. M., Genro, T. C. M., Bayer, C., de Faccio Carvalho, P. C. (2019). Mitigation of enteric

methane emissions through pasture management in integrated crop-livestock systems: trade-offs between animal performance and environmental impacts. *J. Clean. Prod.*, 213:968–975.

StatCounter (2021). Mobile operating system market share brazil. Disponível em: https://gs.statcounter.com/os-market-share/mobile/brazil/#monthly-202010-202110. Acesso em: 11 out. 2021.

Trindade, J. K. et al. (2007). Composição morfológica da forragem consumida por bovinos de corte durante o rebaixamento do capim-marandu submetido a estratégias de pastejo rotativo. *Pesquisa Agropecuária Brasileira*, 42(6):883–890.

University of Missouri (2021). Grazing wedge. Disponível em: https://grazingwedge.missouri.edu/. Acesso em: 11 out. 2021.



Implantação da Rede Temática GO-FAIR Agro Brasil: Primeiros Passos

Debora P. Drucker[1], Sérgio Manuel Serra da Cruz[2,3], Antonio M. Saraiva[4], Juliana M. Fortaleza[5], Patrícia R. B. Bertin[5], Victor P. M. Simão[6], Milena A. Telles[7], Alessandra R. da Silva[8], Paulo S. S. Santos[7], Carla G. N. Macario[1]

[1]Embrapa Agricultura Digital - Campinas - SP - Brasil

[2]Departamento de Computação, Universidade Federal Rural do Rio de Janeiro (UFRRJ) - Seropédica - RJ - Brasil

[3]Programa de Pós-Graduação em Informática, Universidade Federal do Rio de Janeiro (UFRJ) - Cidade Universitária, RJ - Brasil

[4]Departamento de Engenharia de Computação e Sistemas Digitais, Escola Politécnica, Universidade de São Paulo (USP) - São Paulo, SP - Brasil

[5]Empresa Brasileira de Pesquisa Agropecuária - Secretaria de Desenvolvimento Institucional - Brasília, DF - Brasil

[6]Embrapa Meio Ambiente - Jaguariúna - SP - Brasil

[7]Empresa Brasileira de Pesquisa Agropecuária - Secretaria de Pesquisa e Desenvolvimento - Brasília - DF - Brasil

[8]Empresa Brasileira de Pesquisa Agropecuária - Secretaria Geral - Brasília - DF - Brasil

{debora.drucker, juliana.fortaleza, patricia.bertin, victor.simao, milena.telles, alessandra.silva, paulo.santos, carla.macario}@embrapa.br, serra@ufrrj.br, saraiva@usp.br

Abstract. *This paper aims to present the current efforts to create the GO-FAIR Brazil Agro Thematic Implementation Network, highlighting its relationship with the GO-FAIR International and GO-FAIR Brazil initiatives. The main pillars are also highlighted, focusing on cultural change, training and infrastructure, as well as the proposed objectives of the Brazilian network, aimed at promoting the sharing and reuse of research data in agricultural sciences, supported by the FAIR principles. The results show the trajectory traveled so far and points to the elaboration of a Manifesto agreed upon and built through the participation of the communities of Agrarian Sciences and related areas.*

Resumo. *O presente trabalho tem o objetivo de apresentar os esforços que estão em curso para a estruturação da Rede de Implementação Temática GO-FAIR Brasil Agro, sua relação com as iniciativas GO-FAIR Internacional e GO-FAIR Brasil. A importância dos pilares de mudança cultural, treinamento e infraestrutura também é destacada, bem como a proposta de objetivos da rede brasileira voltada para promover o compartilhamento e o reúso de dados de pesquisa em ciências agrárias, apoiado nos princípios FAIR. Como resultados, o trabalho mostra a trajetória percorrida até o momento e aponta para a elaboração de um Manifesto pactuado e construído através de participação das comunidades de Ciências Agrárias e afins.*

1. A Importância dos Princípios FAIR para as Ciências Agrárias

Dados são produtos valiosos da atividade de pesquisa [Heidorn, 2008] e boas práticas para sua gestão devem ser adotadas de modo a assegurar a confiabilidade e a reprodutibilidade do método científico e dos experimentos propriamente ditos. Nesse contexto, os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) – acrônimo em inglês para dados "localizáveis, acessíveis, interoperáveis e reutilizáveis" [Wilkinson et al., 2016], vêm sendo amplamente adotados como norteadores da gestão de dados de pesquisa e viabilizadores de seu reúso.

Com o aumento exponencial da quantidade de dados passíveis de análise, a transição para uma ciência intensiva em dados vem acontecendo em todas as áreas do conhecimento [Hey et al., 2009]. A adoção dos princípios FAIR torna-se ainda mais relevante em áreas do conhecimento multi e interdisciplinares como as Ciências Agrárias[1] e a Agricultura Digital - aqui entendida como a inserção de tecnologias digitais em todas as fases da cadeia de valor tendo em vista a promoção de vantagens competitivas e benefícios socioambientais [Massruhá et al., 2020] - que integram conhecimentos sobre os meios físico, biológico, social e econômico. Assim, dados e objetos digitais utilizados para representar parâmetros de interesse podem variar em termos de formato, sintaxe e semântica, bem como em métodos de aquisição, tratamento e análise e em práticas de curadoria e de preservação. Por exemplo, são relevantes para a pesquisa e a tomada de decisão na agropecuária dados de sensores coletados em tempo real para monitorar variáveis de interesse agroclimático, dados coletados em sensores em máquinas agrícolas, dados de espécies em Áreas de Preservação Permanente (APPs), dados de experimentos de campo ou de entrevistas com produtores rurais.

No caso dos dados de sensores, aspectos relacionados com a qualidade e manutenção dos dispositivos em campo e com a transmissão, o armazenamento e o reúso adequados dos dados, bem como com os fluxos de análise e ajustes de modelos são preocupações determinantes para assegurar a qualidade dos experimentos e seus

[1] A grande área do conhecimento Ciências Agrárias engloba Agronomia, Recursos Florestais e Engenharia Florestal, Engenharia Agrícola, Zootecnia, Medicina Veterinária, Recursos Pesqueiros e Engenharia de Pesca e Ciência e Tecnologia de Alimentos, conforme o diretório de grupos de pesquisa do CNPq - http://lattes.cnpq.br/web/dgp/ciencias-agrarias.

resultados. Já no caso dos levantamentos de espécies, há um alto custo para sua obtenção em campo e para o tratamento de amostras que serão depositadas em coleções biológicas, cujos detalhes devem ser adequadamente registrados e os requisitos da Lei de Patrimônio Genético e Conhecimento Tradicional Associado (Lei nº 13.123, de 20 de maio de 2015) obedecidos. Apenas posteriormente, com auxílio de especialistas, é possível associar um nome taxonômico aos espécimes coletados e, considerando que a taxonomia é também uma ciência viva, os nomes das espécies são atualizados ao longo do tempo e é preciso associar sinonímias para adequar a nomenclatura. Quando consideramos, ainda, dados de entrevistas com agricultores, muitas vezes os dados são oriundos de questionários semi-estruturados, trazendo a complexidade de tratar dados quanti e qualitativos e, do ponto de vista ético e de conformidade, é preciso atentar aos princípios da Lei Geral de Proteção de Dados Pessoais (LGPD - Lei nº 13.709, de 14 de agosto de 2018).

Considerando a diversidade de aplicações dos princípios FAIR para as Ciências Agrárias, sua relevância do contexto brasileiro para a produção de alimentos, agronegócios e sua relação com a saúde humana e do ambiente, o presente trabalho tem o objetivo de apresentar a Rede de Implementação Temática GO-FAIR Brasil Agro: sua relação com as iniciativas GO-FAIR Internacional e GO-FAIR Brasil e sua proposta de estruturação.

2. A Iniciativa GO-FAIR

A rede GO-FAIR é uma iniciativa internacional iniciada em 2018 na Europa e atualmente presente em diversos países. Ela é dirigida pelas partes interessadas e autogovernada, e visa implementar os princípios FAIR, tornando os dados localizáveis, acessíveis, interoperáveis e reutilizáveis. A rede GO-FAIR conta com uma estrutura de governança, escritórios regionais e redes de implementação temáticas, sendo estruturado em três pilares: GO-CHANGE, que trata de mudança cultural em vários níveis de organização institucional na transição para a ciência aberta; GO-TRAIN, relacionado com a capacitação de recursos humanos para implementar os princípios FAIR nas diferentes etapas do ciclo de vida dos dados e GO-BUILD, no qual padrões e serviços de dados FAIR são desenvolvidos e implementados[2]. Atualmente, a rede mantém estreita colaboração com o Committee on Data for Science and Technology do International Science Council (CODATA/ICSU)[3] e outras instituições internacionais relevantes como a Research Data Alliance (RDA)[4] e o World Data Systems (WDS)[5], que têm compromisso de trabalhar juntas para otimizar o ecossistema global de dados e identificar as oportunidades e necessidades que irão acionar as infraestruturas federadas para atender à nova realidade da ciência intensiva em dados.

[2] Mais detalhes podem ser encontrados em https://www.go-fair.org/go-fair-initiative.
[3] Informações sobre o Comitê podem ser encontradas em https://codata.org/.
[4] O portal da https://rd-alliance.org/acomoda detalhes sobre as iniciativas em andamento e detalhes dos grupos de trabalho e grupos de interesse.
[5] Mais informações sobre a WDS estão disponíveis em https://www.worlddatasystem.org/.

Há escritórios regionais da rede GO-FAIR em diversos países, como Alemanha, Áustria, Brasil, China, Dinamarca, França, Estados Unidos e Holanda, os quais são articulados pelo escritório de coordenação "GO-FAIR International Support and Coordination Office (GFISCO)", que promove a troca de experiências para avanço global da implementação dos princípios FAIR. Além disso, há dezenas de Redes de Implementação Temáticas no âmbito internacional (INs, acrônimo para Implementation Networks), que se comprometem a implementar os princípios FAIR para dados e serviços dentro dos três pilares: GO-BUILD (Tecnologia), GO-CHANGE (Cultura) e GO-TRAIN (Treinamento). Algumas INs estão relacionadas com domínios do conhecimento específicos, como Saúde e Biodiversidade, enquanto outras são aplicáveis a todos os domínios, como a IN para Centros de Competência em Custódia de Dados (tradução livre para "Data Stewardship Competence Centers" - DSCC). Há, ainda, iniciativas voltadas para desafios da atualidade, como a IN "Virus Outbreak Data Network (VODAN)", formada no início da pandemia de COVID-19 e que envolve diversos países, tendo contribuído com as necessidades globais de mobilização de dados para enfrentar os desafios que se apresentaram de forma abrupta para a humanidade e reforçando a importância da consonância com os princípios FAIR.

Diversas instituições brasileiras estão representadas nas INs existentes e as atividades do escritório regional GO-FAIR Brasil são articuladas com as iniciativas globais, promovendo o intercâmbio de conhecimentos e a troca de experiências. A seguir, descreveremos o escritório brasileiro em mais detalhes.

3. O Escritório Regional GO-FAIR Brasil

O Escritório do GO-FAIR Brasil foi criado em 2018, durante o evento de 20 anos da Rede Scientific Electronic Library Online (SciELO). Na ocasião, foi redigido e assinado o Manifesto da Rede GO-FAIR Brasil, coordenado pelo Instituto Brasileiro em Informação Científica e Tecnológica (IBICT) [Sales et al 2020]. Sua missão é:

> *"Promover, apoiar e facilitar a adoção dos princípios FAIR na gestão de dados no Brasil, contribuindo para ampliar a visibilidade dos saberes científicos, aprimorar a transparência e a reprodutibilidade da pesquisa, favorecer o reuso de dados e acelerar o desenvolvimento científico em todas as áreas, através da gestão de dados por domínios."*[6]

[6] O portal https://www.go-fair-brasil.org/ contém as informações gerais da rede.

O funcionamento da iniciativa GO-FAIR está descrito em sua Carta de Princípios[7], na qual constam também os objetivos, princípios e estrutura de governança da rede. Analogamente ao GO-FAIR internacional, a rede brasileira também se organiza em três pilares e em Redes de Implementação Temáticas, as quais contribuem para a disseminação e a adoção dos princípios FAIR no Brasil em diferentes domínios do conhecimento.

A criação de uma nova Rede de Implementação Temática (RI) se dá pela representação de instituições-chave no referido domínio, que engaja participantes a escrever e divulgar um Manifesto em evento significativo da área. Atualmente, a rede brasileira conta as Redes de Implementação Temáticas, em diferentes níveis de maturidade, nas áreas da Saúde e a sub-rede Enfermagem; Humanidades; Ciências Nucleares; Ensino, Ciência, Tecnologia e Inovação; Biodiversidade e Agro, a qual será descrita na próxima seção. A Rede brasileira promove encontros para a troca de experiências entre RIs e incentiva trabalhos conjuntos no âmbito de seus pilares, ou seja, promoção de treinamento em conteúdos em comum para diferentes domínios, ações de disseminação e promoção de mudança cultural e atividades voltadas à implantação de soluções tecnológicas para a adoção de padrões de dados e objetos digitais, em consonância com os esforços globais de implementação dos princípios FAIR.

O escritório brasileiro atua também em sincronia com os demais escritórios regionais distribuídos pelo mundo e participa ativamente de ações globais nessa temática, como a participação no FairFestival[8], que ocorreu em junho de 2021, de forma totalmente virtual, e contou com apresentações de vários membros da rede brasileira.

4. A Rede de Implementação Temática GO-FAIR Brasil Agro

A Empresa Brasileira de Pesquisa Agropecuária - Embrapa foi convidada pela Coordenação da Rede GO-FAIR Brasil para liderar a implementação da Rede Temática em pesquisa agropecuária, a qual foi denominada “GO-FAIR Agro”. O Centro de pesquisas em Informática da empresa, Embrapa Agricultura Digital, por sua vocação em atuação na área de Computação, foi designado para coordenar as atividades da Embrapa na Rede, agregando representantes de outros centros da Embrapa. Além da coordenação da Embrapa, representantes de outras instituições relevantes às Ciências Agrárias foram convidados a compor o colegiado de gestão da RI GO-FAIR Brasil Agro: Universidade de São Paulo (USP) e Universidade Federal Rural do Rio de Janeiro (UFRRJ). Essa é uma composição inicial e espera-se que o colegiado de gestão seja ampliado ao longo do tempo, agregando a representatividade do universo produtor e consumidor de dados das ciências agrárias com a participação de outras organizações.

[7] A Carta de Princípios da Rede Go-FAIR Brasil está disponível em https://www.go-fair-brasil.org/documentos .

[8] As gravações do evento e documentos associados estão disponíveis em https://www.go-fair.org/events/fair-festival-2021/.

O esboço do Manifesto da RI GO-FAIR Brasil Agro será apresentado no XIII Congresso Brasileiro de Agroinformática (SBIAgro), em 2021. A proposta inicial, a ser discutida em evento representativo da comunidade (conforme orientação da rede GO-FAIR), é que a missão da Rede GO-FAIR Brasil Agro seja "promover o compartilhamento e o reúso de dados de pesquisa (tanto por humanos quanto por máquinas) em Ciências Agrárias, apoiado nos princípios FAIR". Os objetivos específicos propostos são:

1. Promover pesquisas e inovações sobre:

a. repositórios de dados de pesquisa nacionais e internacionais que poderão armazenar e compartilhar, de forma confiável, os dados de pesquisa em ciências agrárias;

b. padrões de metadados e proveniência de dados específicos;

c. padrões de interoperabilidade tecnológica e semântica dos dados;

d. modelos de plano de gestão de dados.

2. Desenvolver metodologias voltadas para as práticas dos produtos e serviços FAIR, que atendam às necessidades disciplinares e operacionais da área das Ciências Agrárias.

3. Promover difusão do conhecimento através de encontros, cursos, *workshops* e seminários visando impulsionar e disseminar os princípios FAIR entre os membros da Rede GO-FAIR Brasil Agro e sociedade em geral.

4. Trabalhar de forma articulada e colaborativa com a Rede GO-FAIR Brasil e Redes de Implementação Temáticas Nacionais e com a Rede de Implementação *internacional Food Systems*.

O texto do Manifesto será trabalhado conjuntamente com a comunidade durante o XIII SBIAgro e a primeira versão do documento será produzida. Outros atores relevantes, tanto produtores quanto consumidores de dados, ainda não representados, serão convidados a integrar à rede e as etapas de implementação da RI GO-FAIR Brasil Agro serão construídas conjuntamente com a comunidade, em consonância com a Carta de Princípios da Rede GO-FAIR Brasil.

Ainda que os princípios FAIR estejam bem estabelecidos em termos conceituais, quando colocado em prática, geram dúvidas e incertezas. A rápida adoção internacional dos princípios – por instituições de pesquisa, universidades, editores, gestores de dados e de infraestrutura de dados e agências de fomento – levou a uma ambiguidade e a uma ampla interpretação do estado "FAIRness" dos dados e repositórios [Wilkinson et al., 2018], uma vez que os princípios são tecnologicamente agnósticos e não definem estritamente como alcançar esse estado, mas descrevem um conjunto de características, atributos e comportamentos que aproximam um objeto digital desse objetivo. Como resultado, várias abordagens, metodologias, *workflows* e métricas para a "FAIRificação" de dados e objetos digitais foram desenvolvidas [Chue Hong et al., 2021], [Cox et al. 2021], [Jacobsen et al., 2020], [RDA FAIR Data Maturity Model Working Group, 2020], porém, não necessariamente funcionais para todas as áreas do conhecimento, devido à complexidade e heterogeneidade dos dados e às especificidades das áreas de conhecimento.

Algumas iniciativas internacionais no contexto da agricultura digital já vem recomendando práticas mais efetivas de gestão de dados com a adoção dos Princípios FAIR, por exemplo na Plataform for Big Data in Agriculture (CGIAR[9]), no Global Open Data for Agriculture & Nutrition (GODAN[10]), no Centre for Agriculture and Bioscience International (CABI[11]), entre outros. No Brasil, esses esforços ainda são incipientes. Destacamos que operacionalizar os Princípios FAIR em sistemas e repositórios em um país continental e com tamanha abrangência e importância das cadeias do agronegócio é uma tarefa árdua em função do seu alto grau de subjetividade e complexidade, especialmente nas Ciências Agrárias, pois, para produzir mais safras com menos recursos e, ao mesmo tempo, mitigar os efeitos negativos sobre a sociedade e o meio ambiente, é preciso abordagens interdisciplinares intensivas nas análises de dados [Drucker et al., 2021]. A troca de dados entre os diferentes domínios – planta, animal, meio ambiente, ciências sociais, dentre outros – é um grande desafio, ainda maior no contexto da agricultura tropical.

Esforços para a disseminação e adoção dos princípios FAIR têm sido desenvolvidos em todo o mundo, por meio de redes de escritórios regionais, estudos e novas definições de padrões de dados e metadados; disseminação do uso de proveniência de dados, vocabulários controlados, ontologias e adoção de identificadores persistentes que proporcionam significado preciso aos dados e aos demais objetos a eles vinculados [Henning et al., 2019]. Espera-se que, com a implementação da Rede GO-FAIR Brasil Agro, pesquisadores e profissionais que permeiam toda a cadeia do agronegócio e que atuam na interface entre as Ciências Agrárias e da Ciência da Computação contribuam para alavancar a adoção dos princípios FAIR neste domínio científico, bem como para expandir a cultura da promoção de melhores práticas de gestão de dados e de outros objetos digitais. Além disso, o engajamento da comunidade proporciona novas ideias e colaborações entre indivíduos e instituições, podendo expandir para uma efetiva cooperaçãocom a IN Food Systems e outras redes com as quais há interesses comuns, como, por exemplo, a IN BiodiFAIRse e a RI Nacional Biodiversidade. Adicionalmente, avanços no sentido da adoção dos princípios FAIR em ações que independem de domínio devem ser promovidos pela atuação com as demais RIs da rede brasileira, bem como com as iniciativas dos três pilares para promover a mudança cultural, a capacitação de pessoas e o desenvolvimento tecnológico.

[9] Mais informações sobre a CGIAR estão disponíveis em https://bigdata.cgiar.org/gardian-tool/fair-data-principles/

[10] Mais informações sobre o GODAN estão disponíveis em https://www.godan.info/pages/data-ecosystem-and-farm-data-train

[11] Mais informações sobre o CABI estão disponíveis em https://www.cabi.org/

5. References

Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2021). FAIR Principles for Research Software (FAIR4RS Principles). Research Data Alliance. DOI: 10.15497/RDA00065.

Cox, S.J.D.; Gonzalez-Beltran A.N; Magagna B.; Marinescu M.-C. (2021). Ten simple rules for making a vocabulary FAIR. PLoS Comput Biol17(6): e1009041. https://doi.org/10.1371/journal.pcbi.1009041.

Drucker, D. P.; Fortaleza, J. M.; Bertin, P. R. B.; Vacari, I.; Macario, C. G. do N. (2021). Rumo à rede de implantação GO FAIR 'Agro' Brasil: a experiência de uma organização de PD&I na implantação dos princípios FAIR In: Sales, L. F.; Veiga, V. S. O.; Henning, P.; Sayão, L. F. (Ed.). Princípios FAIR aplicados à gestão de dados de pesquisa. Rio de Janeiro, RJ: IBICT. cap. 3. p. 53-68.

Jacobsen, A. et al. (2020). A generic workflow for the data FAIRification process. Data Intelligence 2, 56–65. https://doi: 10.1162/dint_a_00028.

Henning, P. et al. 2019. Desmistificando os princípios fair: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. Pesquisa Brasileira em Ciência da Informação e Biblioteconomia, Paraíba, v. 14, n. 3, p. 175-192. 10.22478/ufpb.1981-0695.2019v14n3.46969.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin, (ed.). 2009. The fourth paradigm: data-intensive scientific discovery. Redmond: Microsoft Research. Disponível em: https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf. Acesso em 1 ago 2021.

Massruhá, S. M. F. S.; Leite, M. A. De A.; Luchiari Junior, A.; Evangelista, S. R. M. (2020). A transformação digital no campo rumo à agricultura sustentável e inteligente. In: Massruhá, S. M. F. S.; Leite, M. A. De A.; Oliveira, S. R. De M.; Meira, C. A. A.; Luchiari Junior, A.; Bolfe, E. L. (Ed.). Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas. Brasília, DF: Embrapa. cap. 1, p. 20-45.

RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. DOI: 10.15497/RDA00050

Sales, L.; Henning, P.; Veiga, V.; Costa, M.M.; Sayão, L.F.; Santos, L.O.B.S.; Pires, L.F. (2020) GO FAIR Brazil: A Challenge for Brazilian Data Science. Data Intelligence 2 (1-2): 238–245. https://doi.org/10.1162/dint_a_00046.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Wilkinson, M.D. et al. (2018). Comment: A design framework and exemplar metrics for FAIRness. Scientific Data 5, 1–4. doi: 10.1038/sdata.2018.118.



IrrigaFeijão – Software para Manejo da Irrigação do Feijoeiro

Pedro M. Silveira[1], Marcelo G. Narciso[1], Alexandre B. Heinemann[1], Silvando C. Silva[1], Fábio F. Noleto[1], Luiz F. Stone[1],

[1]Embrapa Arroz e Feijão – Fazenda Capivara – 74.501-970 – Goiânia – GO – Brasil

{pedro.marques,marcelo.narciso, alexandre.heinemann, silvando.silva, fabio.noleto, luis.stone}@embrapa.br

Abstract. *To avoid loss of productivity in the bean plant, due to inadequate water supply for the plant, planned irrigation is important, in order to irrigate only what is necessary for the bean plant to develop and have the highest possible productivity. In this context, in the case of central pivot irrigation, there is a software for the bean crop, IrrigaFeijão, which allows planning irrigation throughout the plant cycle and with the minimum daily amount of water needed for growth healthy bean plant.*

Resumo. Para evitar a perda de produtividade do feijoeiro, por fornecimento inadequado de água para a planta, é importante uma irrigação planejada, de forma a irrigar somente o necessário para o feijoeiro se desenvolver e ter a maior produtividade de grãos. Neste contexto, para o caso da irrigação ser por pivô-central, existe um software, para a cultura de feijão, o IrrigaFeijão, que permite quantificar a água necessária para o desenvolvimento da planta.

1. Introdução

O feijão é uma cultura de ciclo curto, por isso pode ser muito afetada pela deficiência hídrica. A primeira safra de 2020, por exemplo, sofreu com a estiagem e teve impactos na sua produtividade [IRRIGA 2021]**.**

A irrigação é uma tecnologia que permite planejar melhor a lavoura e não contar somente com a "sorte" da chuva na hora certa. Assim, é importante um planejamento da irrigação.

O manejo da irrigação no feijoeiro consiste em aplicar água na cultura no momento adequado e em quantidades apropriadas. Isso é importante, visto que a produtividade de grãos da cultura é afetada pela condição hídrica do solo. Deficiência ou excesso de água reduz a produtividade em diferentes proporções.

Para facilitar o manejo da irrigação em uma lavoura de feijão, foi construído um software web para a irrigação do feijoeiro, por meio do pivô central, para o Estado de Goiás e alguns minicípios de MG, SP e MT, em sua versão inicial. Este software, cujo nome é IrrigaFeijão, mostra um calendário no qual tem a quantidade de água a ser irrigada em cada dia, de forma a suprir as necessidades de água da planta. O calendário tem início no momento do plantio, e a duração do calendário é conforme o ciclo vegetativo da planta, que pode ser 80, 90 ou 100 dias, e sob plantio direto.

O sistema foi implementado para irrigação por pivo-central, visto que quase a totalidade de irrigação do feijão é feita por esse sistema, como pode ser visto em [AGEITEC 2021]. A Figura 1, a seguir, ilustra um pivô central irrigando uma lavoura de feijão.



Figura 1. Irrigação de lavoura de feijão por pivô central.

Este artigo descreve o software IrrigaFeijão como ferramenta para o manejo da irrigação via pivô-central, com acesso pela web pelo link www.cnpaf.embrapa.br/irrigaFeijao.

2. Material e Métodos

O IrrigaFeijão foi construído para o Estado de Goiás e alguns municípios de SP, MG e MT. Para isto, foram necessários dados de estádio e coeficientes de cultura (Kc) do feijoeiro, para 80. 90 e 100 dias, e de evapotranspiração de referência (ETo) para cada dia durante o ciclo da cultura e para cada município considerado. Os dados relativos a ETo foram provenientes do estudo realizado por Xavier et al. (2016), e estes dados foram de 1980 a 2013. Estes dados foram retirados da base de dados do sistema Infoclima, que pode ser visto em [Infoclima 2021], e anexados à bade de dados do sistema IrrigaFeijão. Com estes 33 anos de dados, foi feita uma média para cada dia e para cada município, para que o sistema IrrigaFeijão tivesse todos os dados de ETo. O coeficiente de cultura (Kc) foi obtido pela relação entre a evapotranspiração da cultura (ETc) e a evapotranspiração de referência (ETo). Os dados de Kc foram obtidos experimentalmente na Embrapa Arroz e Feijão e podem ser vistos em [KC 2021]. Através desse coeficiente,

Kc, e conhecendo-se a ETo de um feijoeiro, pode-se estimar a lâmina de irrigação diária a ser aplicada. Para um dado dia i, o valor da lâmina a ser aplicada é igual a

$$lâmina_i = kc_i * ETo_i$$

O motivo do sistema ter 33 anos de dados, para cada dia do ano e por município, é para se ter uma média, para todos os dias do ano, de ETo, o qual é calculado pela fórmula de Penman-Monteith, a partir da temperatura e umidade do ar, velocidade do vento e radiação solar [PM 2006]. Com isso, é possível prever o quanto irrigar em qualquer dia do ano, a partir do dia atual (hoje). Por exemplo, se o plantio for começar em um dado dia e mês, é possível ver o quanto irrigar a partir daquela data inicial até ao final do ciclo da planta. Então, é possível saber quanto vai ser o gasto médio de água em todo o ciclo de cultivo feijoeiro antes de iniciar o plantio.

O sistema tem com entrada de dados o local da lavoura (município), o dia inicial do plantio, o ciclo vegetativo da planta (80, 90 ou 100 dias), a eficiência de distribuição de água do pivô (80% ou 85%) e sua lâmina de água aplicada a 100% do percentímetro. Ao ler estes dados, o sistema monta uma tabela de irrigação, a partir do dia inicial de plantio, até ao final do ciclo escolhido, calcula o quanto de água foi consumido pela planta e estabelece o quanto deve ser reposto, em cada dia do ciclo, e insere estes dados na planilha de saída do sistema, a qual pode ser salva ou impressa.

O sistema web foi feito com as linguagens javascript e PHP, além de HTML 5 e CSS. A base de dados do sistema foi feita usando o SGBD Mysql. O Aplicativo Dr Feijão, que está no google Play (Android) e App Store (IOs) , contém a versão mobile do sistema IrrigaFeijão.

3. Resultados

O sistema IrrigaFeijão possui um menu, e este direciona o usuário a escolher as páginas que explicam como o sistema funciona, além de considerações sobre o manejo da irrigação do feijoeiro. É recomendado que o usuário leia estas páginas antes de entrar com os dados no sistema. Na página relativa à entrada de dados, existe um formulário, com informações a serem preenchidas. Este formulário é a entrada de dados do sistema, o qual é ilustrado na Figura 2.

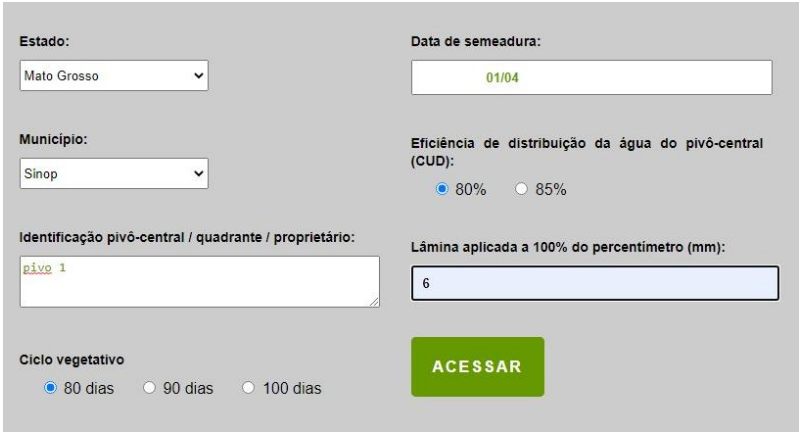


Figura 2. Formulário de entrada de dados do sistema IrrigaFeijão

Este formulário contém campos a serem preenchidos. O município deve ser escolhido para então o sistema obter os valores de ETo corretos para o cálculo da lâmina de irrigação. O ciclo vegetativo é usado para se escolher os valores de Estádio e Kc (coeficiente da cultura) correspondentes a 80, 90 ou 100 dias. A data de semeadura é importante para escolher ETo a partir desta data até ao final do ciclo. É a partir desta data que os cálculos começam e são mostrados na planilha, que está na Figura 3. A eficiência de distribuição de água (80% ou 85%) é usada para o cálculo da lâmina de água a ser aplicada da seguinte forma, para i-ésimo dia:

$lâmina_i = kc_i * Eto_i$ / eficiência, onde eficiência é igual a 0.80 ou 0.85.

O valor da lâmina aplicada a 100% do percentímetro (mm) é usado para calcular o valor do percentímetro do pivô central, em % para cada aplicação da lâmina de água, conforme ilustrado na Figura 3.

Após inserir e submeter os dados, o sistema irá mostrar a planilha de irrigação, tal como ilustra a Figura 3.

Município: **Sinop - MT**

Identificação: **pivo 1**

Eficiência de distribuição de água do pivô central (CUD): **80 %**

Lâmina aplicada a 100% do percentímetro (mm) : **6**

Ciclo vegetativo : **80 dias**

Dia	Dias após a semeadura	Estádio	Lâmina diária (mm)	Lâmina de irrigação (mm)	Percentímetro do pivô-central (%)
1/abril	0	--	--	7.0	86
2/abril	1	Vo	2.8		
3/abril	2	Vo	2.7	5.5	100
4/abril	3	Vo	2.9		
5/abril	4	Vo	2.7	5.6	100
6/abril	5	V1	2.8		
7/abril	6	V1	2.7	5.5	100
8/abril	7	V1	2.7		
9/abril	8	V1	2.9	5.6	100
10/abril	9	V1	2.8		
11/abril	10	V1	2.7	5.5	100
12/abril	11	V2	2.7		
13/abril	12	V2	2.6		
14/abril	13	V2	2.7	8	76
15/abril	14	V2	2.7		

Figura 3. Planilha de saída de dados do sistema IrrigaFeijão

Conforme ilustrado na Figura 3, foram considerados Turnos de Rega de 2 dias na fase inicial (as raízes ainda estão pequenas) e de 3 dias no restante do ciclo da cultura. Na definição do Turno de Rega de 3 dias, levou-se em consideração a capacidade de retenção de água dos solos do Cerrado, como também fatores inerentes ao tamanho da área de irrigação.

Ocorrendo chuva entre a última irrigação e a próxima (duas irrigações consecutivas), a quantidade da precipitação pluvial, em mm, medida por pluviômetro instalado próximo à área, deve ser subtraída da lâmina de água da próxima irrigação. Se essa precipitação pluvial for maior que a lâmina da próxima irrigação, considerar essa irrigação já atendida e não computar créditos dos milímetros (mm) de água excedentes em irrigações futuras.

Esta planilha foi elaborada para semeadura variando de 01 de abril a 20 de junho, época apropriada para semeadura de feijão de terceira safra (feijão irrigado). A Figura 4 ilustra o final da planilha, que vai até o final do ciclo da cultura.



3/junho	63	R8	3.3		
4/junho	64	R8	3.3	10	61
5/junho	65	R8	3.2		
6/junho	66	R8	3.2		
7/junho	67	R8	3.3	9.7	62
8/junho	68	R8	3.2		
9/junho	69	R8	3.2		
10/junho	70	R8	3.3	9.7	62
11/junho	71	R8	3.2		
12/junho	72	R8	3.2		
13/junho	73	R8	3.3	9.7	62
14/junho	74	R8	3.2		
15/junho	75	R8	3.2		
16/junho	76	R9	2.6	9	67
17/junho	77	R9	2.6		
18/junho	78	R9	2.6		
19/junho	79	R9	2.6	7.8	77
20/junho	80	R9	2.5		
			Total	279.4	

Figura 4. Planilha de saída de dados do sistema IrrigaFeijão, Valores finais da planilha.

4 Comentários finais

Neste contexto de planejamento e uso da irrigação, por pivô central, está o IrrigaFeijão. É um software fácil de usar, e que reúne as variáveis necessárias para auxiliar, de maneira simples e prática, o usuário a realizar o planejamento/manejo da irrigação na cultura do feijão, por pivô central, em Goiás e alguns municípios produtores de feijão de MT, MG e SP. Vale a pena ressaltar que, para a cultura do feijão, o sistema foi implementado para irrigação por pivo-central visto que quase a totalidade de irrigação do feijão é feita por esse sistema.

O planejamento/manejo apresenta-se na forma de um calendário de irrigação, com informações de quando e quanto irrigar, após o usuário informar o município, a data prevista de semeadura, a eficiência do pivô central e a lâmina mínima aplicada. Portanto, o IrrigaFeijão é um software de apoio a decisão, que aconselha o usuário quando e quanto irrigar.

Assim, os usuários que não dispõem de outro método de manejo da irrigação do feijoeiro na lavoura têm no no IrrigaFeijão uma planilha de planejamento prático da irrigação da cultura, a partir da data de semeadura, de modo fácil e rápido e com precisão aceitável.

O IrrigaFeijão pode ser adpatado para outros culturas que usam pivô central para irrigação. Para isto, basta saber o valor de Kc para a cultura específica que se desejar planejar/manejar a irrigação.

A quase a totalidade dos irrigantes não usam nenhum metodo de controle. O irrigafeijão vem atender essa demanda.

Referências Bibliográficas

AGEITEC. Sítio disponível em https://www.agencia.cnptia.embrapa.br/gestor/feijao/arvore/CONTAG01_86_131120021510 4.html. Sítio visitado em 07/09/2021.

Infoclima, Sítio disponível em www.cnpaf.embrapa.br/infoclima. Sítio visitado em 07/09/2021.

IRRIGA. Sítio disponível em https://blog.aegro.com.br/irrigacao-de-feijao/. Sítio visitado em 07/09/2021.

KC. Sítio disponível em https://www.agencia.cnptia.embrapa.br/gestor/feijao/arvore/CONTAG01_86_131120021510 4.html, Sítio visitado em 07/09/2021.

MP. Sítio disponível em https://ainfo.cnptia.embrapa.br/digital/bitstream/CNPUV/8815/1/cir065.pdf Sítio visitado em 07/09/2021.

Xavier, A. C.; King, C. W.; Scanlon, B. R. Daily gridded meteorological variables in Brazil: (1980.2013). International Journal of Climatology, v. 36, n. 6, p. 2644- 2659, mar. 2016.



Mineração de dados em rede social para avaliação de tendências de consumo do queijo artesanal no Brasil

Thallys da Silva Nogueira[1], Kennya Beatriz Siqueira[2], Priscila Vanessa Zabala Capriles Goliatt[1]

[1]Programa de Pós-Graduação em Modelagem Computacional – Universidade Federal de Juiz de Fora (UFJF) – 36036-900 – Juiz de Fora – MG – Brazil

[2]Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) - Gado de Leite – 36038-330 – Juiz de Fora – MG – Brazil

{thallysnogueira,capriles}@ice.ufjf.br, kennya.siqueira@embrapa.br

Abstract. *The present work refers to the results obtained by the development of a computational system called "Consumer Observatory". Based on the concepts of business intelligence, it is capable of collecting, storing, processing and extracting information from data of the Twitter social network. The objective is to evaluate the consumption trends by identifying the characteristics and consumption habits of artisanal cheese in Brazil. Using artificial intelligence techniques, data mining and natural language processing is possible to identify information about the scenario of consumption trends for artisanal cheeses in Brazil, unknown until now.*

Resumo. *O presente trabalho refere-se aos resultados obtidos pelo desenvolvimento de um sistema computacional denominado "Observatório do Consumidor". Baseando-se nos conceitos de business intelligence é capaz de coletar, armazenar, processar e extrair informações em dados da rede social Twitter. O objetivo é avaliar tendências de consumo por meio da identificação de características e hábitos de consumo do queijo artesanal no Brasil. Fazendo-se uso de técnicas de inteligência artificial, mineração de dados e processamento de linguagem natural, é possível identificar informações sobre o cenário das tendências de consumo dos queijos artesanais no Brasil, até então desconhecidas.*

1. Introdução

Desde sempre, entender bem o mercado em que se atua, assim como o consumidor, é uma tarefa muito importante para o desenvolvimento de estratégias em ambiente empresarial. O *Business Intelligence* (BI) tem sido bastante falado nos dias atuais e refere-se a um conjunto de conceitos e metodologias que visam auxiliar e tornar mais eficiente o processo de tomada de decisão de uma empresa, transformando dados históricos e correntes em informação, e informação em decisões mais assertivas [Turban *et al.* 2010]. Com o crescente uso das redes sociais, empresas estão cada vez mais atentas ao que é exposto nestes ambientes virtuais. Segundo Wasserman e Faust (1994), as redes sociais são definidas como um conjunto composto de dois elementos que são: (i) os atores (pessoas, instituições ou grupos) e (ii) as suas conexões (interações ou laços sociais).

Com tantas funcionalidades disponíveis, as redes sociais conectam pessoas de todo o mundo e o número de pessoas que possui um cadastro nessas redes tem aumentado ano após ano [DataReportal 2021a]. Entre abril e junho de 2019, a quantidade de pessoas que possuía cadastro em alguma rede social chegou a 3,5 bilhões de pessoas. Em janeiro de 2020, atingiu-se um total de 3,8 bilhões de usuários e, em janeiro de 2021, chegou-se à marca de 4,2 bilhões de usuários [DataReportal 2021a]. Estes números podem não representar de fato contas individuais de usuários, entretanto, sugerem que mais da metade da população atual do planeta utiliza alguma rede social [Worldometers 2021].

No Brasil, a realidade é semelhante. Em janeiro de 2021, o País possuía uma população de 213,3 milhões de pessoas [DataReportal 2021b]. Cerca de 160 milhões de internautas ficavam em média 10h 08min conectados na internet através de qualquer dispositivo. Neste intervalo de tempo, gastavam, em média, 3h 42min com redes sociais, 4h 02min em frente à TV (TV aberta, *streaming* e *on demand*), 1h 52min ouvindo músicas por serviços de *streaming* e 1h 17min usando consoles de videogame. Tem-se, então, que cerca de 150 milhões de brasileiros possuem algum cadastro em rede social, totalizando 70,3% da população [DataReportal 2021a].

Com tantas redes sociais disponíveis, um grande volume de dados é gerado e, por esse motivo, o termo *Big Data* está muito presente nestes ambientes, fazendo referência ao grande volume, variedade (podendo estes dados serem mensagens de texto, comentários, vídeos, áudios, fotos) e velocidade com que são gerados. Analisar dados provenientes destes ambientes demanda novas maneiras de processar os dados e gerar informação útil, melhorando a percepção e a tomada de decisão das empresas [Gartner 2021]. Áreas como mineração de dados, inteligência artificial e ciência de dados, têm ganhado destaque por abordarem técnicas e metodologias capazes de identificar tais informações neste vasto universo de dados.

Neste âmbito, alternativas para entender o perfil do consumidor e tendências de consumo têm sido estudadas como forma de suprir as deficiências da pesquisa de mercado tradicional. Para Barabba e Zaltaman (1991), tal pesquisa é definida como o "*processo de ouvir a voz do mercado e transmitir à administração informações a respeito*", tendo como principal objetivo direcionar a tomada de decisão. Essa pesquisa de mercado tradicional é feita há décadas pelas empresas e por esse motivo já é validada e consegue atingir as expectativas das empresas. Porém, a pesquisa de mercado

tradicional pode ser ineficiente por se apresentar, em alguns casos, como questionários que podem ser extensos e com longo tempo de aplicação, o que leva a não adesão de participantes; pela necessidade de profissionais qualificados para a aplicação dos testes; pela limitação geográfica em pesquisas de extensão (inter-)continental; e principalmente pela dificuldade em se obter uma amostra bem representativa e distribuída [Veríssimo *et al*. 2018].

Pensando nisso, a Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA Gado de Leite, inovou em parceria com a Universidade Federal de Juiz de Fora (UFJF) e o Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas - Campus Juiz de Fora (IF-Sudeste/MG), desenvolvendo o projeto "Observatório do Consumidor", uma plataforma que busca desenvolver alternativas às pesquisas de mercado tradicionais. Baseando-se na coleta e análise de dados de redes sociais e fazendo uso de técnicas de inteligência artificial, mineração de dados, *web*-semântica e processamento de linguagem natural, o sistema identifica o perfil dos consumidores, assim como suas tendências de consumo. A ideia é aplicar a transformação digital nas empresas, usando o grande volume de dados e informações disponíveis nas redes sociais para entender o mercado consumidor brasileiro, de modo a fornecer informações dinâmicas e em tempo real sobre as mudanças de comportamento e tendências de consumo no País.

O primeiro caso de estudo da plataforma do Observatório do Consumidor foi relacionado à mudança no comportamento de consumo de derivados lácteos no Brasil antes e durante a pandemia por COVID-19 [Siqueira *et al.* 2020a] [Siqueira *et al*. 2020b]. Nesses trabalhos, foi empregada a rede social Twitter [Twitter 2021], que é uma das redes sociais mais usadas no mundo com cerca de 328 milhões de contas ativas [Mohamed 2018]. Essa rede oferece aos usuários um espaço para conversação, publicação de conteúdos chamados de *tweets*, fotografias e vídeos. No contexto do Brasil, de acordo com DataReportal (2021b), cerca de 51,6% dos brasileiros entre 16 e 64 anos que usam a internet possuem uma conta no Twitter, tornando esta a sexta rede mais usada no Brasil, ficando atrás do Youtube, Facebook, Whatsapp, Instagram e FB Messenger.

Com intuito de mostrar o potencial que o Observatório do Consumidor tem para a extração de informações presentes em dados da rede social do Twitter, neste trabalho, avaliou-se em particular o queijo artesanal, um produto com forte apelo cultural no Brasil e que é responsável pela renda de muitas famílias em todo o País. Do lado do consumo, os queijos artesanais também se destacam por estarem alinhados com as principais tendências de consumo de alimentos atuais: regionalismo, volta às origens, produto natural, entre outras. Outro ponto interessante é que os queijos artesanais são muito versáteis dentro da cultura brasileira, podendo ser consumidos em diversas ocasiões, além de atender a diversos públicos, desde consumidores locais de baixa renda até famosos *chefs* de comida *gourmet*. No entanto, pouco se sabe sobre os hábitos e preferências de consumo de queijos artesanais no Brasil. Há uma carência de informações e estatísticas oficiais sobre esse mercado, que podem ser abordadas por meio do uso da *Business Intelligence,* facilitando a gestão de empresas de todos os portes desse setor.

2. Material e Métodos

A plataforma do Observatório do Consumidor possui três etapas principais: (i) coleta e armazenamento dos dados, (ii) processamento dos dados e (iii) pós-processamento dos dados e extração de informação. Inicialmente, foi necessário desenvolver uma lista contendo ao todo 179 palavras-chave (que neste caso, foram os nomes dos queijos artesanais de interesse), que foi construída com a supervisão e orientação de especialistas da Embrapa Gado de Leite e representantes da Empresa de Assistência Técnica e Extensão Territorial (EMATER).

Com a lista de palavras-chave validada, iniciou-se a etapa de coleta dos dados na rede social submetendo cada palavra-chave como parâmetro do algoritmo de coleta, armazenando cada *tweet* no banco de dados relacional desenvolvido. Com os dados devidamente armazenados, iniciou-se a etapa de processamento que padroniza os dados coletados removendo caracteres indesejados com o intuito de tornar análises futuras mais eficientes. Após esta etapa, com todos os *tweets* devidamente padronizados, iniciou-se a etapa de pós-processamento dos dados e extração de informação que realiza a análise de sentimentos dos *tweets* coletados, utilizando um aprendizado de máquina supervisionado para a classificação da polaridade dos sentimentos dos *tweets* em negativos, neutros e positivos e identificação de padrões relacionados à características do queijo artesanal e seus hábitos de consumo.

O modelo de análise de sentimentos foi desenvolvido na linguagem de programação Python [Van Rossum e Drake 2009] com a biblioteca *scikit-learn* [Pedregosa *et al.* 2011]. Para realização da tradução da linguagem humana para a linguagem de máquina, foi utilizada a abordagem *bag of words*[1] com o esquema TF-IDF [Sammut e Webb 2011], que é dada pela Frequência dos Termos - Inverso da Frequência do Documento[2]. Neste trabalho, foi proposto um classificador utilizando o *Voting Classifier* [Pedregosa *et al.* 2011] com voto majoritário a fim de combinar diversos classificadores (Regressão Logística, Naive Bayes, OneVSRest e OneVSOne) em um único classificador visando melhorar o desempenho e a precisão na classificação da polaridade dos sentimentos de cada *tweet.*

Para avaliação e verificação do desempenho do modelo, foi utilizada a técnica de validação cruzada utilizando o *K-fold* que consta em dividir o conjunto de dados de treinamento em K partes, sendo uma parte para treino e outra para teste seguindo um comportamento incremental. Utilizando a função *StratifiedKFold*, dividiu-se os dados em 10-*folds* usando o parâmetro *shuffle = true* a fim de garantir que os grupos fossem embaralhados mantendo-se sempre um percentual equivalente de dados de cada classe em cada uma das partições (tanto na de treinamento como na de teste). Para validação, cinco métricas foram implementadas que são a R^2, MSE (*Mean Squared Error*), RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*) e a matriz de confusão.

Para a construção do *dataset* de treinamento, foi utilizado cerca de 10,16% dos 82.959 *tweets* coletados com intuito de criar uma base própria e exclusiva do queijo artesanal com dados do próprio Twitter dado que não foi encontrado nenhum *dataset* específico na literatura para tal finalidade. Para a pré-classificação do conjunto de

[1] Vocabulário contendo todas as palavras contidas nos *tweets* de forma única.
[2] Técnica de determinação do peso que cada palavra possui em um documento.

treinamento, utilizou-se o *software* GOTIT [AI 2021], que usa redes neurais e análise semântica para a extração de informação em textos, fazendo uso de processamento de linguagem natural para a determinação dos sentimentos e polaridade presentes em uma frase. Ao todo, o *dataset* construído possui 8.432 *tweets* com 33,7% de classificações negativas, 13,5% para a classe neutra e 52,8% para a classe positiva.

Para analisar quais foram as características e hábitos de consumo mais mencionados nos *tweets* construiu-se mais duas listas de palavras-chave para a realização da mineração desses "padrões" nos dados. Com grande número de publicações coletadas, analisar de forma individual cada uma das postagens torna-se uma tarefa impraticável e por este motivo justifica-se o uso de técnicas de mineração de dados. Uma das listas de palavras-chave criadas têm a função de identificar as principais características do queijo artesanal (cor, modo de produção, sabor, textura dentre outros) e a outra, os principais hábitos de consumo (acompanhamentos, receitas culinárias, bebidas). Com o intuito de se obter informações positivas sobre o consumo dos queijos artesanais, foram selecionados todos os *tweets* classificados como positivos para esta análise de conteúdo.

3. Análise e Discussão dos Resultados

Ao todo foram coletados 82.959 *tweets* entre 30 de abril de 2020 e 18 de fevereiro de 2021, totalizando quarenta e duas semanas. Das 179 palavras-chave desenvolvidas para a coleta de dados, apenas 80 retornaram *tweets* ao longo das coletas realizadas, o que pode ser explicado pela existência de grande variedade de queijos artesanais que, muitas vezes, não são muito conhecidos pela população brasileira que utiliza o Twitter. Dentre as categorias de queijos analisados as 5 mais mencionadas pelos usuários foram **Coalho** (43.339 *tweets*), **Outros** (19.835 *tweets*), **Artesanal Paulista** (7.484 *tweets*), **Minas Artesanal** (5.687 *tweets*) e **Maturado** (2.925 *tweets*).

Após realizar o treinamento do modelo de análise de sentimentos, a Tabela 1 mostra os resultados obtidos por cada classificador.

Tabela 1 - Resultados dos classificadores. Em negrito está destacado o melhor resultado. Fonte: autores.

Modelo	Acurácia Média (%)	Intervalo de Confiança (%)
Regressão Logística	71,12	[68,22 - 74,02]
Naive Bayes	65,78	[62,49 - 68,08]
OneVSRest	70,94	[68,19 - 73,70]
OneVSOne	70,69	[68,16 - 73,23]
Voting Classifier	**71,20**	**[68,58 - 73,83]**

É possível verificar que após 10 rodadas realizadas pelo *k-fold*, o modelo proposto utilizando o *ensemble Voting Classifier* foi o que obteve o melhor resultado de acurácia média com 71,20% e um intervalo de confiança de [68,58 - 73,83]% sendo este modelo o utilizado para a realização da classificação da polaridade dos sentimentos dos *tweets* na ferramenta. Este resultado, em um primeiro momento, é bem expressivo pelo fato da base de treinamento do modelo ainda estar em desenvolvimento, possuir um

número reduzido de dados, além de possuir um desbalanceamento nas classes. Mesmo com tais limitadores, estes resultados estão bem próximos aos da literatura, como pode ser visto em Ankit (2018) em que o modelo de *ensemble* construído para classificações de sentimentos de *tweets* obteve uma acurácia média de 75,6%, sendo o tamanho da base de treinamento e o balanceamento de classes um dos principais fatores por este resultado.

Ao avaliar as métricas R^2, MSE, RMSE e MAE para o modelo *Ensemble Voting Classifier* que obteve o melhor resultado de acurácia média, foi possível observar que de um modo geral, tais métricas também obtiveram resultados bastante expressivos mesmo possuindo uma base de treinamento desbalanceada. Os resultados obtidos podem ser visualizados na Tabela 2.

Tabela 2 - Valores das métricas de validação onde R^2é o coeficiente de determinação, MSE é o erro médio quadrático, o RMSE é a raiz do erro médio quadrático e o MAE é o erro absoluto mediano. Fonte: autores.

Métrica	Valores
R^2	0.87
MSE	0.16
RMSE	0.33
MAE	0.09

Com o modelo validado, a Figura 1 exibe o quantitativo total de classificações negativas, neutras e positivas encontradas pelo modelo.

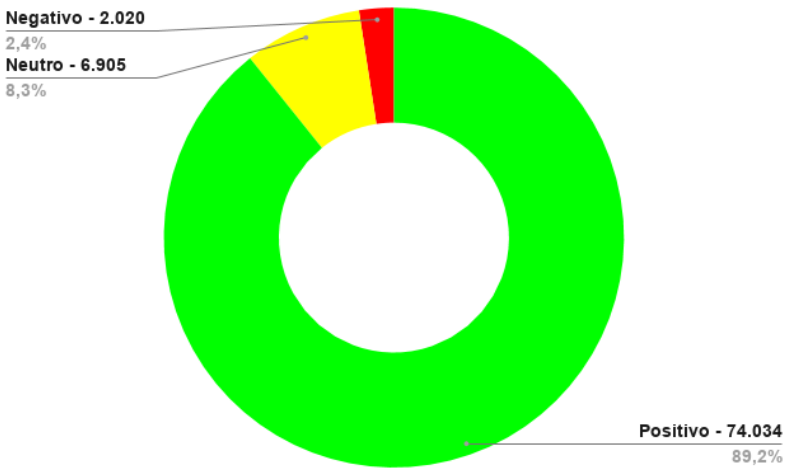


Figura 1 – Análise geral da polaridade dos sentimentos dos *tweets* sobre queijos artesanais ao longo das semanas de estudo (30/04/2020 a 18/02/2021). Fonte: autores.

De um modo geral, o sentimento positivo foi o que mais se destacou nos *tweets*, com 89,2% do total de classificações. Na maioria das vezes, a menção aos queijos artesanais nas publicações ocorreu pelo fato deste produto estar presente em diversos pratos da culinária brasileira, que são mencionados como sendo bastante saborosos.

Os *tweets* que não apresentaram adjetivos que os qualificassem tanto positivamente quanto negativamente foram classificados como neutros, e representaram cerca de 8,3%. Um exemplo sobre esta classificação seria a de um usuário somente afirmar que comeu um determinado queijo, não mencionando se gostou ou não do mesmo.

As classificações negativas apareceram em menor número representando um total de 2,4% do total. As classificações negativas no contexto dos queijos artesanais, ocorreram muito relacionadas a pessoas intolerantes à lactose. Também foi possível notar nas classificações negativas, afirmações sobre os incômodos proporcionados pelo mofo presente nos queijos e odor característico de alguns tipos de queijo.

Como mencionado anteriormente, para analisar as características e hábitos de consumo, selecionou-se todos os *tweets* com classificações positivas totalizando em 74.034 *tweets.* Ao realizar o processamento em cada *tweet*, foram identificadas cerca de 103.484 menções das características e hábitos de consumo analisadas neste estudo. Das 122 características do queijo artesanal, 98 foram encontradas e dos 104 hábitos de consumo, 103 foram retornados ao fim do processamento. Para fazer uma análise mais assertiva das tendências ao longo das semanas foi necessário analisar de forma separada cada uma das categorias devido à diferença no número de menções de atributos entre uma categoria e outra, o que dificultava muito as análises dos resultados.

Em resumo, das características do queijo artesanal, os atributos com maior número de citação foram: (i) produção **artesanal** e **tradicional**, (ii) a composição com **cebola** (sendo que esta pode não estar associada à composição do queijo) e **orégano** (presente em "dadinhos" de queijo coalho com orégano) e (iii) a categoria maturado com o termo **curado** se referindo ao processo de maturação dos queijos.

Já para os hábitos de consumo os de maior destaque foram: (i) acompanhamentos com **pão**, **carne** e **bolo**, (ii) receitas de **pão** (de/com) **queijo**, **tapioca**, ***pizza*** e **carne** (de) **sol** e (iii) bebidas, acompanhando **café** e **leite**. Com base nisso, é possível concluir que o consumo dos queijos artesanais está associado a diferentes refeições, podendo estar presente no café da manhã visto que **pão**, **café** e **leite** apareceram como acompanhamentos muito citados. Podem estar associados também a lanches, almoço e jantar com menções aos atributos **arroz**, **carne sol**, ***pizza*** (sendo um produto muito consumido no Brasil identificando a presença dos queijos artesanais como um dos ingredientes) e **tapioca** podendo esta ser consumida em qualquer refeição, inclusive como sobremesa. Isso indica que o consumo do queijo artesanal no Brasil é bastante diversificado e tem sido consumido amplamente nas mais diversas formas, refeições e acompanhamentos.

Devido a questões como características e hábitos de consumo regionais, definir uma tendência de consumo unificada para os queijos artesanais do Brasil mostrou-se muito complexo. A fim de facilitar essa identificação de tendências regionais, o Observatório do Consumidor atualmente conta com um sistema que fornece por meio de visualização gráfica a extração de informações de forma rápida e intuitiva, tornando assim a tomada de decisão local mais efetiva.

4. Considerações Finais e Indicações Futuras

Usando os "queijos artesanais" como estudo de caso, foi possível concluir que através da implementação da ferramenta do "Observatório do Consumidor" diversas

análises podem ser realizadas possibilitando identificar informações estratégicas e importantes sobre o cenário das tendências de consumo no Brasil. Através da combinação de técnicas de *Business Intelligence* e Inteligência Computacional em um só lugar, esta solução inova a maneira de se fazer pesquisas de mercado com análises das opiniões expressas pelos usuários/consumidores nas redes sociais, sem a influência e a necessidade de aplicação de questionários extensos, rompendo barreiras geográficas com maior diversidade e alcance de pessoas, mostrando quais são as principais características e hábitos/tendências de consumo. Como trabalhos futuros, pretende-se ampliar o universo de produtos e redes sociais a serem analisados, em que espera-se englobar outras categorias de derivados lácteos para a coleta, processamento e análise de dados.

References

AI, GOTIT. GOTIT - Sentiment. (2021), https://gotit.ai/en-us/Home/Sentiment.

Ankit Saleena, Nabizath. (2018). An ensemble classification system for twitter sentiment analysis. Procedia Computer Science, v. 132, p. 937–946. ISSN 1877-0509. International Conference on Computational Intelligence and Data Science, https://www.sciencedirect.com/science/article/pii/S187705091830841X.

Barabba, T. and Zaltaman, P. (1991). "Hearing the voice of the market". Harvard Business School Press.

DataReportal. (2021). "Digital 2021 Global Digital Overview", https://datareportal.com/reports/digital-2021-global-digital-overview.

DataReportal. (2021). "Digital 2021 Brazil", https://datareportal.com/reports/digital-2021-brazil.

Gartner. (2021). "Definition of Big Data", https://www.gartner.com/en/information-technology/glossary/big-data.

Mohamed, Sinkadar. (2018). "100 Social Media Statistics You Must Know [2018] + Infographic", https://statusbrew.com/insights/social-media-statistics-2018.

Pedregosa *et al.*, (2011). Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830.

Sammut C. and Webb G.I. (2011). TF–IDF. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_832

Siqueira, Kennya B. and Nogueira, Thallys S. and Campos, Emerson W. and Soares, Nedson D. and Moraes, Emerson A. P. and Villela, Regina M. M. B. and David, José Maria N. and Goliatt, Priscila V. Z. C. (2020). "Análise exploratória da imagem dos lácteos em tempos de coronavírus. Indústria de Laticínios", n. 143, p. 64–66, ISSN 1678-7250.

Siqueira, Kennya B. and Nogueira, Thallys S. and Campos, Emerson W. and Soares, Nedson D. and Moraes, Emerson A. P. and Villela, Regina M. M. B. and David, José Maria N. and Goliatt, Priscila V. Z. C. (2020). "O impacto da pandemia no consumo de lácteos no Brasil". Indústria de Laticínios, n. 147, p. 36–38, ISSN 1678-7250.

Turban, E. and Sharda, R. and Delen, D. (2010) “Decision Support and Business Intelligence Systems”. p.720

Twitter, Inc. (2021). “Twitter”, https://twitter.com.

Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Veríssimo, B and Lepre, L. and Tincani, D. (2018). “Diferenças entre pesquisa de marketing e pesquisa de neuromarketing”.

Wasserman, S. and Faust, k. (1994). “Social network analysis: methods and applications”. Cambridge University Press.

Worldometers. (2021). “World Population Clock: 7,8 Billion People”, https://www.worldometers.info.



MIP² - um sistema de informação especializado em manejo integrado de pragas

Rodolfo Chagas Marinho Nascimento[1], Monica Simão Giponi[2], Wellington Moreira de Oliveira[1], Vânia Maria Xavier[2]

[1] Departamento Acadêmico de Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais - Campus Rio Pomba, Rio Pomba, Minas Gerais, Brasil

[2] Departamento de Acadêmico de Agricultura e Ambiente, Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais - Campus Rio Pomba, Rio Pomba, Minas Gerais, Brasil

nsc.rodolfo@gmail.com, monicagiponi@gmail.com, wellington.moreira@ifsudestemg.edu.br, vania.xavier@ifsudestemg.edu.br

Resumo. *O Manejo Integrado de Pragas (MIP) é uma ferramenta que contribui com a tomada de decisão quanto ao momento ideal para a adoção de métodos de controle nos cultivos agrícolas. No entanto, a falta de conhecimento técnico e a quantidade de etapas a serem realizadas, torna a adoção do MIP custosa e passível de erros no recolhimento dos dados. Diante disso, o presente trabalho propõe um sistema que acompanha o agricultor na utilização do MIP em sistemas orgânicos de produção. Para o desenvolvimento deste sistema, identificaram-se os principais pontos para auxiliar nas etapas de implementação do MIP e a melhor forma de transferir esses conhecimentos aos usuários. Como resultado, o sistema apresenta diversos métodos de controle alternativos, informações para nortear o desenvolvimento das culturas e a correta identificação das pragas, além de gráficos que auxiliam no manejo dos cultivos.*

Abstract. *Integrated Pest Management (IPM) is a tool that contributes to decision making regarding the ideal time to adopt control methods in agricultural crops. However, the lack of technical knowledge and the number of steps to be carried out make the adoption of the IPM costly and minded to errors in data collection. Therefore, the present work proposes a system that accompanies the farmer in the use of IPM in organic production systems. For the development of this system, the main points to assist in the stages of implementation of the IPM and the best way to transfer this knowledge to*

users were identified. As a result, the system features several alternative control methods, information to guide the development of crops and the correct identification of pests, as well as graphics that help in crop management.

1. Introdução

A computação tem proporcionado avanços em diversos setores da sociedade e uma das grandes áreas beneficiadas é o setor agrícola. A integração computacional na agricultura ampliou-se desde a Revolução Verde, período entre 1960 e 1970 em que houve a difusão de tecnologias agrícolas que permitiram um aumento considerável na produção [Octaviano 2010]. Desde então, muitos agricultores vêm se adaptando a esta nova realidade para ampliar e melhorar seu sistema de produção.

Nos sistemas agrícolas, um dos grandes desafios enfrentados pelos produtores é o controle de pragas, que pela falta de manejo adequado pode causar grandes prejuízos. Atualmente, um modelo amplamente utilizado no controle de pragas em sistemas convencionais é o emprego de métodos químicos de controle, assim que observada a presença da praga nos cultivos. Ações desse tipo podem causar um desequilíbrio no agro ecossistema, além de não promover o controle adequado das pragas e elevar o custo de produção [Picanço, Gonring e Oliveira 2010]. A utilização de métodos de controle alternativos é pouco difundida e ainda apresenta entraves quanto a sua adoção. Dentre estes, destacam-se a escassez de conhecimentos sobre os métodos alternativos e a pouca transferência e divulgação dos mesmos aos agricultores [Campanhola e Bettiol 2003].

Uma metodologia utilizada para auxiliar os produtores no manejo das pragas é o Manejo Integrado de Pragas (MIP), um conjunto de técnicas que visa manter as pragas abaixo do nível de dano econômico [Gallo *et al*. 2002]. No entanto, sua correta adoção requer conhecimentos que na maioria das vezes o produtor não possui, relacionados ao diagnóstico da praga e a tomada de decisão, podendo dificultar e desestimular o agricultor a utilizar esta técnica. Diante dessas limitações, a utilização de um sistema de informação seria muito bem-vindo, devido aos benefícios gerados no armazenamento, organização e consulta de dados, facilitando a compreensão de todo processo e aplicação do MIP.

A crescente expansão da tecnologia e o advento dos aplicativos móveis tornou possível oferecer ferramentas de forma acessível. Sendo assim, atualmente, estão disponíveis aplicativos que contribuem desde a gerência agrícola até o manejo das culturas, como o aplicativo Aegro[1] e o Doutor Milho[2]. No entanto, mesmo que, de acordo com Rodrigues (2004), a utilização de softwares para área de MIP seja muito promissora, a presença e utilização de sistemas que introduzem as técnicas e acompanham na realização do MIP ainda é ínfima. Prevalecendo softwares de gerência das propriedades e de informações técnicas, como os citados acima.

Diante da demanda de tecnologias acessíveis para o controle, acompanhamento e difusão de informações relacionados ao MIP, este trabalho apresenta um sistema que

[1] https://play.google.com/store/apps/details?id=com.aegro.aegroapp&hl=pt_BR&gl=US

[2] https://play.google.com/store/apps/details?id=br.embrapa.app.doutormilho&hl=pt_BR&gl=US

auxilia na sua adoção e na utilização de métodos de controle considerados alternativos. Desenvolvido nas plataformas Web e mobile, o sistema visa guiar os produtores rurais na gerência de suas propriedades e na realização do manejo de pragas em sistemas de produção orgânica.

O presente trabalho foi estruturado da forma que segue. A Seção 2 descreve o contexto histórico do MIP no Brasil. A Seção 3 apresenta a metodologia e as etapas de desenvolvimento do sistema MIP². Na Seção 4 é realizado um comparativo com aplicativos existentes atualmente. A Seção 5 lista trabalhos que se relacionam com o presente trabalho. Por fim, na Seção 6, são feitas as devidas ponderações, conclusões e discussões sobre trabalhos futuros.

2. Manejo Integrado de Pragas

Para podermos entender a necessidade de se adotar o MIP precisamos realizar uma análise do contexto histórico acerca da situação em que se encontra o manejo de pragas no Brasil. Atualmente, o controle de pragas em propriedades rurais é realizado, em sua maioria, através de aplicações de produtos químicos, visando extinguir a população de insetos no campo. Normalmente, não se utilizam cálculos ou constatações da população das pragas em campo, esse tipo de produto é utilizado assim que o inseto-praga é identificado em campo. Desta forma, gerando aplicações de agrotóxicos em excesso e sem uma forma expressiva de controle. Tal prática causa inúmeras consequências como impactos ambientais e riscos à saúde humana, tanto para quem consome quanto para quem faz a aplicação destes produtos. Apesar do controle químico aparentar ser uma prática moderna, essa assertiva não é verdadeira. Segundo Gazzoni (2012), logo após a Segunda Guerra Mundial houve uma grande expansão quanto à utilização de inseticidas organossintéticos. Devido a este fato, rapidamente surgiram os primeiros problemas relacionados ao seu uso excessivo. Nas décadas vindouras, os entomologistas desenvolveram técnicas com o objetivo de diminuir esse rápido avanço dos agrotóxicos. Dentre essas técnicas podem-se destacar o “controle supervisionado de insetos” e o “controle integrado”, que visavam monitorar o inseto-praga, utilizando-se de pequenas doses de inseticidas para controlá-los concomitantemente com o uso do controle biológico [Gazzoni, 2012]. Além dessas premissas, era intrínseco realizar um levantamento para saber se a população das pragas havia atingido um nível em que o prejuízo causado fosse maior que custo das medidas de controle necessárias.

Objetivando diminuir a população de pragas e consequentemente os danos por elas causados, em conjunto com estudos e estratégias para atingir esse propósito, originou-se o que hoje chamamos Manejo Integrado de Pragas. Ao incorporar o conjunto de conceitos do MIP no sistema de produção seria possível garantir efeitos benéficos, social e ambientalmente. Porém, para integrar seus ideais aos cultivos é imprescindível o desenvolvimento de conjuntos tecnológicos que reforcem e auxiliem na aplicabilidade dessa metodologia, também se atendo à realidade dos cultivos em um futuro próximo. Assim, permitindo sua adoção em larga escala e expansão das técnicas do MIP. Além disso, podemos afirmar que apenas a tecnologia não seria suficiente para tornar o MIP totalmente difundido entre os produtores, também se fazem necessárias políticas públicas de encorajamento à aplicação do MIP e conscientização dos produtores quanto ao quadro de sustentabilidade na agricultura.

3. Manejo Integrado de Pragas apoiado por Sistema de Informação

Inicialmente, com objetivo de realizar uma análise geral da aplicabilidade do MIP, foi feito o acompanhamento de um experimento no setor de olericultura do IFSEMG – campus Rio Pomba. A partir deste acompanhamento, foi possível destacar quais eram os pontos mais importantes para a adoção do MIP por um usuário com pouca ou nenhuma experiência. Além de informações imprescindíveis para realizar um manejo de pragas corretamente, é disponibilizado um acompanhamento aos produtores durante as fases do manejo. Além disso, o sistema auxilia no levantamento populacional das pragas, identificando o momento ideal de se realizar o manejo. Desta forma, ao implantar esta ferramenta na propriedade, o produtor garante que o método de controle só será utilizado quando necessário, e não de forma aleatória como é comum no método convencional.

3.1. Engenharia de Requisitos

Para realizar o desenvolvimento deste projeto foram seguidas algumas etapas, sendo a primeira delas o levantamento de requisitos em conjunto com envolvidos na área de Agroecologia. A partir dos requisitos foram definidas as funções do sistema e a disposição das telas. Através da criação do diagrama de casos de uso[3] foi possível traçar o perfil dos usuários, tendo como base o acompanhamento em campo e as atividades exercidas pelos produtores. Considerando que o aplicativo deve auxiliar o produtor no monitoramento e na gestão de seu cultivo foram criadas duas classes de usuários: produtor e funcionário. Assim, é possível que o produtor adicione funcionários à sua propriedade e permite os funcionários a realizarem o recolhimento de dados nas culturas em que estão cadastrados.

Por meio do acompanhamento da execução manual do MIP, foram definidas quais as informações o sistema deveria organizar e gerenciar. Referente às pragas-chave e seus inimigos naturais é importante apresentar os dados que facilitem a identificação em campo e exemplifiquem qual tipo de dano é causado pelo inseto. Da mesma forma, as informações referentes às culturas acometidas foram disponibilizadas para auxiliar o manejo do produtor e, consequentemente, diminuir a ocorrência das pragas.

Em consonância com o conteúdo apresentado, foram definidas as melhores formas de apresentar a realização das amostragens de praga. Visando ajudar usuários pouco experientes no manejo de pragas, adotou-se o plano de amostragem convencional, que se baseia na coleta de um número fixo de amostras por área [Gallo *et al*. 2002]. Desta forma, foram realizadas pesquisas bibliográficas de fichas amostrais das pragas e das culturas escolhidas, onde já está estabelecido o número fixo de amostras e a porcentagem de infestação necessária para iniciar a adoção dos métodos de controle. Sendo assim, com base nos dados coletados durante a realização do plano de amostragem, é realizado um cálculo que compara o nível populacional identificado com os níveis já estabelecidos pela literatura. Como resultado, o sistema informa se é necessário adotar um método de controle ou não para o talhão amostrado.

Caso seja necessário a aplicação de um método de controle, o sistema disponibiliza uma listagem com as opções recomendadas para a praga em questão. Assim que o usuário realiza estas etapas, os dados são salvos e podem ser acessados por ele posteriormente para análise da flutuação da população de pragas ou métodos de controle aplicados. Estes dados são disponibilizados em forma de relatórios e gráficos,

[3] https://github.com/nscRodolfo/MIP-IniciacaoCientifica/tree/master/Documentação

que podem ser armazenados no dispositivo móvel, assim o usuário poderá ter um melhor controle de seu cultivo. Os relatórios, dados do usuário e demais informações também podem ser acessadas pela plataforma Web.

3.2. Modelo lógico

Após o levantamento inicial de informações e técnicas necessárias, iniciou-se a segunda etapa do desenvolvimento com a criação de um protótipo com o objetivo de medir a aceitação dos usuários quanto ao sistema. Este protótipo foi disponibilizado para testes, realizados por estudantes de Agroecologia e envolvidos na área, o que facilitou a identificação de falhas no layout e usabilidade. Após a correção dos problemas de navegação e acesso às informações, iniciou-se a terceira etapa com a implementação do aplicativo móvel e incorporação dos resultados das pesquisas bibliográficas ao banco de dados. Para isso, fez-se necessária a criação do modelo lógico do banco de dados (Figura 1) e a implementação do mesmo, assim tornando possível salvar os dados na nuvem através de uma plataforma de *host*.

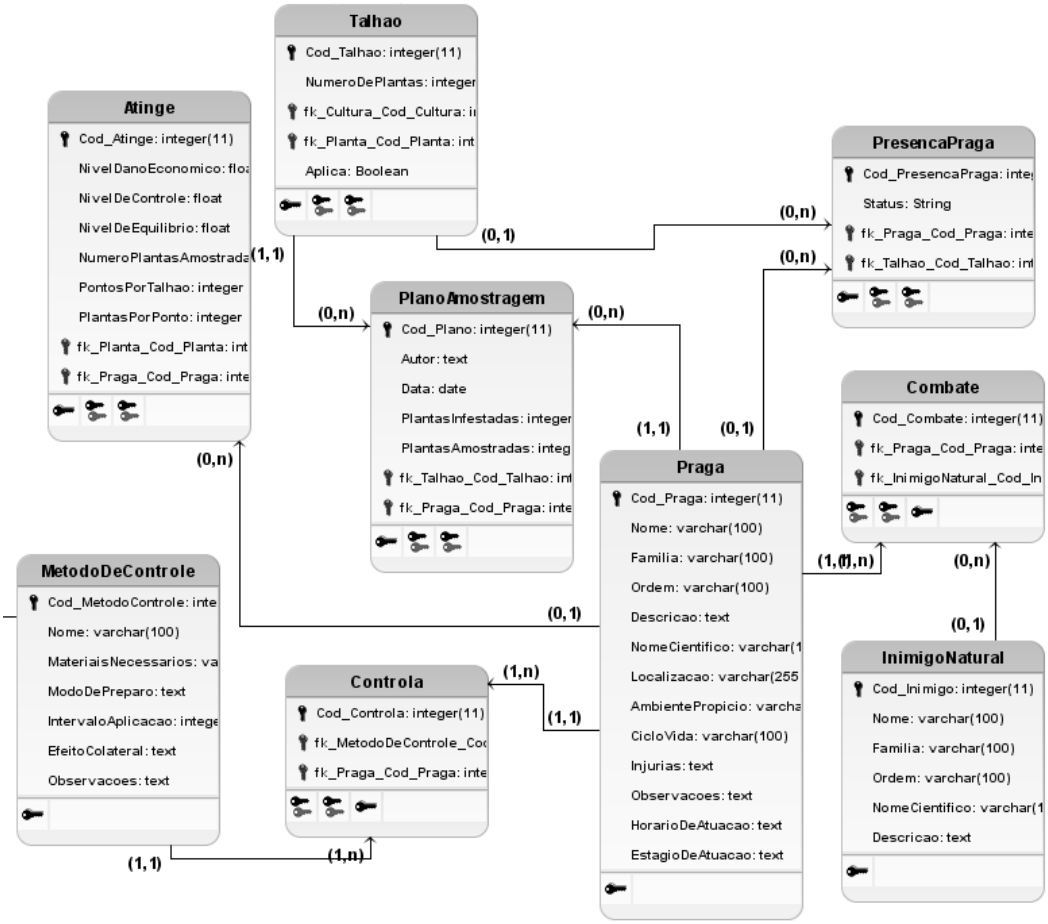


Figura 1. Parte do modelo lógico de banco de dados

A implementação foi realizada através do ambiente de desenvolvimento disponibilizado pela Google, *Android Studio*, que utiliza da linguagem *JAVA* para elaboração dos códigos e *XML* para desenvolvimento das interfaces. Assim que o desenvolvimento do aplicativo móvel foi finalizado, deu-se início à criação do módulo

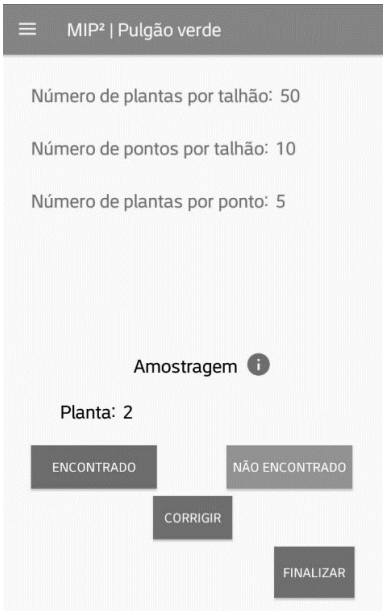
de suporte Web, feito na linguagem *Hypertext Markup Language* (HTML) com assistência de *JavaScript* e *Hypertext Preprocessor* (PHP) para comunicação com o Banco de Dados. Neste, definimos que seria criada uma forma mais simples de observação dos dados, deixando a parte de realização do monitoramento apenas para o aplicativo mobile. Desta forma, o módulo Web tem a função de visualização de informações e gerenciamento da propriedade. Assim que os dois módulos estavam finalizados foi realizada a interação entre ambas plataformas.

Mesmo com o desenvolvimento concluído ainda foi necessário aprimorar o aplicativo para suprir requisitos que se revelaram importantes durante a etapa de implementação do sistema. Dentre estes, destacou-se a necessidade de realizar o manejo de pragas de forma *offline*, já que a maioria das propriedades se encontram em zonas rurais, com pouco ou nenhum acesso à *internet*. Além disso, algumas áreas de cultivo são afastadas dos locais onde existe conexão *wi-fi* ou mesmo não possui rede móvel para conexão com a *internet*, isso acontece já que o acesso à *internet* é limitado e diversas vezes com alto custo em alguns locais da zona rural. Mesmo com este recurso é necessário que o agricultor consiga acesso à *internet* periodicamente para que o sistema possa realizar a sincronização dos dados. Caso o usuário não consiga se conectar à *internet* os dados estarão salvos apenas no dispositivo móvel.

4. Resultados e discussão

O sistema foi desenvolvido com o objetivo de auxiliar nas atividades de manejo de diversas culturas com importância comercial. Através dos dados disponibilizados, este *software* pretende capacitar o agricultor no entendimento e aplicação do MIP, mesmo que não tenha tido contato com essa prática. Para isso, guia o usuário no monitoramento das pragas (Figura 2(a)), orientando no caminhamento e na coleta das amostras. Além disso, contribui com a gerência da propriedade e permite observar os resultados da tomada de decisão, apresentados através de relatórios (Figura 2(b)). O MIP² [4] foi publicado na *Google Play Store*, após adaptações na interface e códigos, tornando-se assim disponível para os usuários. Em conjunto com o aplicativo também foi disponibilizado um módulo Web que pode ser acessado pelo endereço "*mip.software"*.

[4] https://play.google.com/store/apps/details?id=my.aplication.manejointeligentedepragas&hl=pt_BR&gl=US



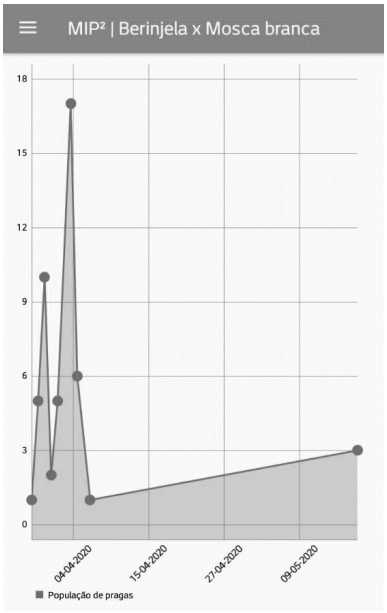


Figura 2(a) **Figura 2(b)**

Figura 2. Telas do aplicativo. Figura 2(a) – Plano de amostragem, Figura 2(b) – Relatório de população de pragas.

A presença de sistemas especializados no manejo de pragas ainda é pequena quando se trata de dispositivos móveis. Os sistemas relacionados a este objetivo e que estão disponíveis na *Google Play Store* foram utilizados para um comparativo e avaliação do presente trabalho. De modo a comparar as principais funcionalidades disponíveis em cada aplicativo em relação ao MIP², a Tabela 1 apresenta uma lista demarcando com um "X" a presença das mesmas em cada sistema.

Tabela 1. Variáveis a serem consideradas para aplicativos de manejo de pragas

	MIP²	ADAMA alvo	IZAgro	Pragueiro upcampo	Guia InNat
Tutorial	X	X		X	X
Plataforma Web e Mobile	X			X	X
Multicultura	X	X	X	X	
Informações sobre as culturas	X				
Apresenta as pragas chave	X	X	X	X	
Apresenta os inimigos naturais	X				X
Identificação de pragas	X	X			
Métodos de controle	X	X	X	X	
Métodos de controle alternativos	X				

Forma de preparo e intervalo de aplicação dos métodos de controle	X				
Relatórios	X			X	
Auxilia na tomada de decisão	X				

Conforme os dados apresentados na tabela acima, podemos confirmar que o MIP² é o *software* mais completo dentre os apresentados, tornando-se uma das mais completas soluções para aplicação dos conceitos do Manejo Integrado de Pragas. Um dos maiores diferenciais do sistema é a utilização apenas de métodos de controle alternativos, sendo vários deles recomendados e aprovados pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA[5]). Desta forma, valorizando e incentivando o crescimento da agricultura orgânica brasileira. Os *softwares* Agricontrole e AgroPro Monitor, embora atuem na área de manejo de pragas, não foram avaliados neste comparativo por não apresentar uma versão disponível na *Google Play Store* ou por dificuldade no contato com os desenvolvedores.

5. Trabalhos relacionados

Atualmente encontramos diversos sistemas voltados para a área agrícola, isto ocorre, principalmente, segundo Galvão (2010), por conta da utilização de microeletrônica, biotecnologia e informática na produção agrícola. Apesar disso, na sua maioria, são sistemas que ajudam na gerência agrícola ou na divulgação de informações.

Como exemplo podemos citar o aplicativo Agricontrole [Kuhn 2018], que apresenta mapeamento da área, controle do tamanho dos cultivos, controle de rebanho e de estoque. Todas essas funcionalidades e informações são de suma importância para produtores rurais, porém não o auxiliam no desenvolvimento de técnicas agrícolas com o fim de melhorar o manejo das culturas.

Aplicativos que acompanham o produtor na aplicação das técnicas do MIP ainda estão em falta no mercado. Podemos encontrar alguns como, por exemplo o AgroPro Monitor[6] que auxilia, também, no levantamento de população de pragas. Porém, os sistemas que encontramos ainda são voltados para o manejo em produções convencionais, além de não apresentar e demonstrar as etapas para adotar o MIP.

Visando suprir essa necessidade o MIP² foi criado, trazendo informações técnicas durante as etapas do processo de adoção do MIP. Além disso, com o intuito de fomentar a agricultura orgânica, apresenta apenas métodos de controle alternativos considerados orgânicos, trazendo os ingredientes, modo de preparo e intervalo de aplicação.

Este sistema acompanha o produtor de forma simples, explicando os processos e conduzindo-o durante a implantação e execução do MIP, com a possibilidade de verificar a eficiência dos métodos utilizados, além de monitorar a flutuação populacional por meio dos relatórios gerados.

[5] http://agrofit.agricultura.gov.br/agrofit_cons/principal_agrofit_cons
[6] https://play.google.com/store/apps/details?id=com.agromonitor&hl=pt_BR&gl=US

6. Conclusões

O desenvolvimento de ferramentas voltadas para a área agrícola é capaz de beneficiar os agricultores e contribuir com a difusão de técnicas que melhoram o setor produtivo. O sistema desenvolvido apresenta uma técnica específica que visa auxiliar e melhorar o manejo de pragas nas culturas cadastradas. Traz também, diversas informações sobre as pragas, culturas e métodos de controle que acrescentam ao produtor e contribuem com o manejo das culturas em um sistema de produção orgânica. Esta ferramenta também apoia a tomada de decisão quanto ao manejo adotado em cada talhão. Desta forma, por meio dos relatórios, o agricultor poderá verificar se o manejo para determinado talhão está contribuindo ou não com a baixa ocorrência de pragas em dado período, e assim realizar as possíveis adequações necessárias. Os relatórios apresentados, também podem colaborar com a rastreabilidade dos produtos quanto aos métodos de controle utilizados e auxiliar produtores que buscam uma certificação orgânica.

Como trabalho futuro pretende-se avaliar esta ferramenta com o público alvo para verificar sua utilidade e usabilidade, de modo a melhorar e adequar o *software* às demandas dos agricultores.

Agradecimentos

Esta pesquisa foi apoiada pelo IF Sudeste MG - Campus Rio Pomba e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico.

7. Referências

Campanhola, C. Bettiol, W. (2003) Métodos alternativos de controle fitossanitário. Jaguariúna: Embrapa Meio Ambiente.

Kuhn, F. (2018) “Agricontrole: aplicativo android para gerenciamento agrícola”. Trabalho de Conclusão de Curso. Universidade Tecnológica Federal do Paraná.

Gallo, D. *et al*. (2002) Entomologia agrícola. Piracicaba: FEALQ.

Galvão, W. C. (2010) “Sistema WEB para gerenciamento de informações agrícolas”. Dissertação de mestrado. Universidade Estadual de São Paulo.

Gazzoni, D. L. (2012) “Perspectivas do manejo de pragas”, In: Soja: manejo integrado de insetos e outros Artrópodes-praga, Editado por Clara Beatriz Hoffmann-Campo, Beatriz Spalding Corrêa-Ferreira e Flavio Moscardi. Embrapa, Brasília.

Picanço, M. C. (2010) Manejo integrado de pragas. Viçosa: UFV.

Rodrigues, W. C. (2004). Utilização da Informática na Entomologia. Rev. Info Insetos - p. 1-10.

Octaviano, C. (2010). Muito além da tecnologia: os impactos da Revolução Verde. ComCiência, p. 0-0.



Modelagem de uma RSSF para manejo de pragas em pomares usando o simulador Eboracum

Felipe Frantz Zanini[1], Weslen Schiavon de Souza[1], Lisane Brisolara de Brisolara[1], Paulo R. Ferreira Jr.[1]

[1]Programa de Pós-Graduação em Computação
Cyber-Physical Systems Engeneering Laboratory
Centro de Desenvolvimento Tencológico
Universidade Federal de Pelotas
Pelotas, Brasil

{felipe.zanini, wsdsouza, lisane, paulo.ferreira}@inf.ufpel.edu.br

Abstract. *Eboracum is an extensible modeling simulation framework for evaluating wireless sensor networks (WSN). This provides high-level primitives for the modeling of platform and application aspects of WSNs, and thus allowing to evaluate efficiency of application-specific WSN solutions. In this work, we employ Eboracum for modelling of solutions for integrated pest management based on smart traps. A small extension of simulator was proposed and implemented in order to model the smart trap as sensor nodes. Based on the proposed modeling and simulation preliminary results, the efficiency of the solution is discussed, and optimizations are planned.*

Resumo. *O Eboracum é um framework extensível para modelagem e simulação de redes de sensores sem fio (RSSF). Este fornece primitivas de alto nível de abstração para modelagem da plataforma de hardware bem como aspectos da aplicação, permitindo assim avaliar a eficiência de soluções de RSSF dedicadas a uma dada aplicação. Neste trabalho, empregamos o Eboracum para a modelagem de uma solução de RSSF para o manejo integrado de pragas baseado em armadilhas eletrônicas. Uma pequena extensão do simulador foi proposta para a modelagem das armadilhas como nodos sensores. Com base na modelagem proposta e em resultados preliminares de simulação, a eficiência desta solução é discutida e prospeta-se otimizações.*

1. Introdução

As Redes de Sensores sem Fio (RSSF) têm sido empregadas no sensoriamento e monitoramento de ambientes para as mais diversas aplicações, dentre estas destaca-se a Agricultura

de Precisão (AP), onde redes de sensores tem sido amplamente empregadas para gerenciamento de irrigação, fertilização e controle de pragas [Manikandan e Jayapriya 2018]. Estas redes são normalmente auto-organizáveis e formadas por nós sensores, sendo estes geralmente conectadas entre si através de módulos de rádio frequência (RF), com baixo custo energético [Alsharif et al. 2019].

O aprimoramento da captura de dados e a evolução das tecnologias de informação e comunicação, alavancaram consideráveis melhorias na eficiência, efetividade e produtividade de vários setores agrícolas [Zhang et al. 2018]. A Agricultura de precisão é uma das áreas de aplicação de RSSF, onde sensores são geralmente empregados para monitorar a qualidade do solo e da plantação, coletando dados como umidade, pressão e temperatura. Ainda, RSSF têm sido amplamente empregadas nessa área para o gerenciamento de irrigação, fertilização e para o controle de pragas [Manikandan e Jayapriya 2018].

Com o intuito de identificar a proliferação de pragas em lavouras, armadilhas são comumente espalhadas pelo cultivo, de modo que um responsável fique encarregado de verificá-las periodicamente [Murtaza et al. 2019]. A coleta de dados das armadilhas é feita por humanos, o que não garante a precisão dos dados e a periodicidade requeridas, prejudicando o manejo. A frequência da coleta é muito importante para o controle de infestações em uma dada plantação e RSSF são empregadas para realizar este monitoramento [Shelton e Badenes-Pérez 2006]. Estas redes são usualmente compostas por nodos mantidos por bateria e, portanto, o consumo energético deve ser minimizado de forma a prolongar o tempo de vida da rede.

Para mensurar a eficiência de uma solução de RSSF dedicada para uma aplicação, experimentos precisam ser realizados que permitam avaliar o tempo de vida da rede, bem como a cobertura de sensoriamento e comunicação. Experimentos no campo apesar de oferecerem maior precisão nas análises demandam muitos esforços e custo de aquisição prévia de equipamentos. Por outro lado, simuladores permitem avaliar diferentes soluções com menor custo e esforço, postergando a aquisição de equipamentos e os testes diretamente em campo. No entanto, na avaliação através de simulação, é necessário emular os fenômenos do ambiente real, considerando sua distribuição espacial na área de interesse e a frequência de ocorrência destes eventos usando os modelos estocásticos apropriados para a aplicação alvo [Haghighi e Cliff 2013].

A ferramenta Eboracum [2019] permite a modelagem e avaliação de RSSF, fornecendo primitivas de alto nível para modelar tanto aspectos da camada de hardware quanto de aplicação. Em [Marques et al. 2016], a modelagem de alguns aspectos da aplicação de manejo de pragas com armadilhas eletrônicas foram discutidos. No entanto, neste estudo preliminar, extensões do simulador não foram de fato implementadas e modelos estocásticos para representação da infestação foram generalizados. Estes modelos precisam ser mais precisos uma vez que emulam a carga de trabalho a qual a rede será submetida no campo e afetam diretamente a avaliação da eficiência das soluções de rede.

Este artigo contribui com a modelagem e simulação de soluções de RSSF dedicadas a uma dada aplicação empregando os modelos de alto nível de abstração suportados pelo Eboracum. Por meio de um estudo de caso empregando a aplicação de manejo de pragas, será demonstrado como empregar esta infraestrutura para modelar e simular soluções de RSSF. Neste estudo de caso, uma solução dedicada de RSSF reativa será avaliada.

O artigo está organizado da seguinte forma: a Seção 3 revisa o framework Eboracum, enfatizando as primitivas empregadas ou estendidas para realização dos estudos de caso. Na Seção 4 é apresentado o modelo de armadilha inteligente que será considerado para modelagem dos estudos de caso. O modelo construído para o estudo de caso será detalhado na Seção 5 e os resultados preliminares são apresentados na Seção 6. Por fim, conclusão e trabalhos futuros serão discutidos na Seção 7.

2. Trabalhos Relacionados

O emprego de armadilhas eletrônicas para o manejo de pragas vem sendo estudado em vários grupos de pesquisa e por equipes de pesquisa e desenvolvimento de empresas [Parsons et al. 2019] [Ding e Taylor 2016] [Remboski et al. 2018] [Martins et al. 2019]. Em [Parsons et al. 2019] uma revisão é apresentada de várias soluções para controle de pragas em geral envolvendo armadilhas inteligentes. No trabalho de [Remboski et al. 2018] foi discutido o emprego de armadilhas eletrônicas inteligentes no manejo de pragas em pomares, empregando redes neurais para a identificação dos insetos. A identificação de moscas da fruta também foi objeto de estudo de [Martins et al. 2019], onde diferentes técnicas de aprendizado de máquina foram avaliadas para aplicação nas armadilhas eletrônicas. A técnica proposta por [Martins et al. 2019] está sendo prototipada em uma parceria com o Grupo de Pesquisa de Ciberfísicos da UFPel e a empresa Partmon. Este protótipo serviu de base para a caracterização apresentada neste trabalho.

As armadilhas inteligentes podem ser integradas à módulos comunicadores, como discutido em [Parsons et al. 2019], podendo assim ser conectadas de forma a compor uma rede de sensores sem fio. A ideia de modelar uma RSSF dedicada para manejo de pragas no simulador Eboracum já foi proposta em [Marques et al. 2016]. No entanto, o referido trabalho possui limitações pois vários pontos foram apenas prospectados porém sem de fato terem sido experimentados, além de várias caracterizações ainda não estarem definidas pela ausência de protótipos ou trabalhos que suportassem esta caracterização. O presente trabalho contribui não só com o modelo de simulação e as respectivas extensões do simulador mas também com caracterizações que propiciarão novos experimentos desta aplicação no simulador Eboracum, incluindo avaliações de otimizações.

3. Eboracum

Eboracum é um framework para projeto e avaliação de RSSFs, que visa modelar e simular RSSFs em alto nível de abstração. Esta infraestrutura de simulação permite avaliar o consumo energético dos nodos da rede ao executar uma aplicação, e consequentemente estimar o tempo de vida da rede.

O Eboracum dispõe de classes que permitem representar os componentes que compõem a rede. O nodo sensor da rede é representado por instâncias da Classe *SimpleWSNNode*, já a estação base ou Sink é representada pela classe *NetworkMainGateway*. Estes representam as principais primitivas para a representação dos componentes que compõem a plataforma de hardware. Ambos possuem capacidade de comunicação, e um parâmetro define o raio de alcance do módulo transmissor. Mas, o nodo sensor também tem capacidade de captura de fenômenos e esta é delimitada por um raio de sensoriamento. Além disso, o nodo sensor tem capacidade de processamento, assim o nodo processa os dados capturados, gerando a informação que será transmitida pela rede.

Este simulador permite dois modos de simulação uma sincronizada com relógio, e outra seguindo um modelo de simulação de eventos discretos. O simulador suporta a avaliação do consumo, permitindo analisar o tempo de vida da rede considerando uma dada carga de trabalho. A carga de trabalho de trabalho será modelada através de eventos gerados pelo simulador. A descarga da bateria deverá ser caracterizada no modelo de simulação considerando três modos de operação para os nodos sensores: idle, processamento e transmissão.

Nesta simulação baseada em eventos discretos, as atualizações só ocorrem quando algum evento acontece, o que permite acelerar a simulação. Desta forma a carga de trabalho da aplicação é representada por eventos que acionam os atores do modelo. O Eboracum oferece várias classes para representação de diferentes fenômenos da natureza. A classe *PeriodicEvent* representa o acionamento de nodos sensores para capturas periódicas, com períodos pré-definidos. Dentre as primitivas providas, destaca-se ainda classes que modelam eventos usando distribuições de probabilidade, tal como a classe *StochasticPeriodicJumperEvent*. Esta classe permite representar um conjunto grande de eventos não simultâneos que são distribuídos espacialmente e temporalmente seguindo modelos estocásticos. Muitos fenômenos da natureza, inclusive infestações de pragas, requerem modelos estocásticos para uma simulação mais precisa. Então em nosso estudo de caso, esta classe será empregada no modelo proposto a fim de emular a incidência de insetos à área de interesse que representa o cultivo.

4. Armadilhas Inteligentes

Existe uma considerável variedade de insetos que afetam negativamente a produção de frutas [Sharma et al. 2017]. No sul do Brasil, pesquisadores apontaram a mosca-das-frutas Ceratitis Capitata como uma das pragas mais críticas presentes em árvores frutíferas [Mello e Caimi 2008]. Para mitigar as infestações por este inseto, os produtores usam pesticidas. O monitoramento da população de moscas-das-frutas e a tomada de decisões sobre sua mitigação baseiam-se em armadilhas de feromônios.

As armadilhas inteligentes consideradas neste trabalho são baseadas no trabalho de [Martins et al. 2019] e utilizam feromônios para atrair insetos, que ao entrarem nas armadilhas ficam presos em seu interior.

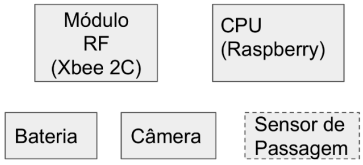


Figura 1. Diagrama de blocos da armadilha inteligente

A Figura 1 ilustra os principais módulos que são utilizados no protótipo da armadilha inteligente, que são: CPU, Módulo Raio-frequência (RF), Câmera e uma bateria para alimentar todo o sistema. A câmera captura a imagem do piso adesivo da armadilha. A abordagem proposta usa uma técnica de processamento digital de imagens para detectar regiões de interesse (ROI - *Region of Interest*) que provavelmente representem moscas e uma rede neural artificial para analisar e identificar a classe do inseto capturado. Essas

técnicas são executadas em uma Raspberry Pi III. O módulo de comunicação empregado neste trabalho é o XBee S2C [Digi International Inc. 2020]. Um sensor infra-vermelho de passagem é usado para ativar o sistema quando uma mosca entra na armadilha, fornecendo assim um comportamento reativo para nodos sensores. Estas informações sobre o hardware empregado nas armadilhas são importantes para a construção dos modelos, sobretudo para a caracterização do consumo dos nodos , que será discutida na Seção 5.

5. Modelagem Proposta

Neste trabalho, considera-se o emprego das armadilhas propostas em [Martins et al. 2019] para compor uma RSSF dedicada ao manejo de pragas. Estas armadilhas eletrônicas são baseadas em feromônio e possuem recursos para captura de imagens, processamento e comunicação, como detalhado na Seção 4. Cada armadilha é um nodo da rede, que deve processar a imagem capturada e comunicar o total de moscas da fruta detectadas fazendo chegar esta informação através de múltiplos saltos (*multihop*) até a estação base. Estas mensagens serão de 36 Bytes empregando o protocolo Zigbee [Zigbee Alliance], considerando 24 Bytes de cabeçalho e 12 Bytes de Payload.

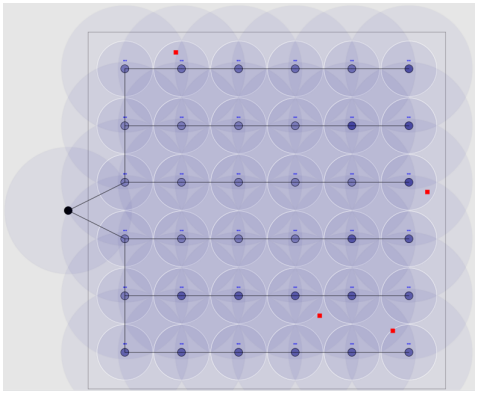


Figura 2. RSSF dedicada ao manejo de pragas

Para avaliar nosso modelo, foi construído um cenário com 36 nodos sensores implantados através de uma área quadrada de 810.000 m2 (900 m x 900 m) e uma estação base localizada na lateral esquerda, conforme ilustrado na Figura 2. Todos os nodos possuem um módulo RF S2C. O alcance do radio foi considerado equivalente a 160 metros conforme experimentos em campo relatados em [Freitas et al. 2019].

A estação base é representada por uma instância da classe *NetworkMainGateway*. Este nodo só tem capacidade de comunicação. Já os nodos sensores (os quais representam as armadilhas), além da capacidade de comunicação, também são caracterizados por um raio de sensoriamento de 70 metros. Este raio foi definido considerando dados obtidos com especialistas da Partmon que sugerem uma armadilha a cada um hectare quando o feromônio Bio Trimedlure é utilizado.

O nodos sensores também foram caracterizados pela quanto ao custo energético referente ao processamento da imagem capturada pelas armadilhas e técnica de classificação. Este custo depende do número de insetos ou regiões de interesse (ROI) desta imagem. O número de regiões de interesse dependem do número de insetos capturados. Assim, a classe *SimpleWSNNodeWithEventCounter* foi criada no simulador de

forma a representar as armadilhas reativas e contabilizar o número de insetos capturados. Esta classe estende a classe *SimpleWSNNode*, permitindo que a descarga da bateria dos nodos possa considerar a contagem de eventos capturados para a estimativa do consumo energético referente ao processamento

Assim, quando o nodo captura um evento, este nodo deve incrementar seu número de regiões de interesse. Este valor será multiplicado pelo custo energético referente ao processamento de um único ROI para determinar a descarga da bateria do nodo quando este estiver processando. Assim como empregado no protótipo de armadilha desenvolvido, os nodos sensores da solução modelada são mantidos por uma bateria de lítio, a qual fornece uma carga de 25200000 mAs. Este valor portanto é usado como bateria inicial dos nodos sensores em nossos experimentos.

Nosso modelo considera as armadilhas reativas, assim os nodos reagem quando algum evento é detectado em seu raio de sensoriamento. O evento acionará a captura da imagem e o tratamento dos dados, e consequentemente o envio da nova informação para a estação base. Na modelagem proposta para a aplicação alvo, os eventos representam os insetos que entram nas armadilhas. Para que a armadilha seja acionada desta forma, um sensor infra-vermelho foi incluído no hardware do nodo sensor. Com a passagem do inseto, este sensor aciona a captura da imagem e na sequência aciona a Raspberry para o processamento da imagem capturada. Concluído o processamento, uma mensagem será enviada via rede até alcançar a estação base.

Os eventos aparecem no espaço seguindo uma distribuição normal invertida, visando emular as moscas chegando prioritariamente no pomar pelas das bordas da área de monitoramento. Para representar aspectos temporais da incidência de insetos na rede, o modelo aqui proposto considera que a maior probabilidade de chegada de insetos ocorre ao entardecer. A frequência dos eventos é dada por uma distribuição de Poisson com uma média de intervalo de 68399 segundos (ou 19 h), desvio padrão de 199 s (2 h) e intervalo de 0 a 86399 segundos (24h). Para representar estes eventos foi criada a classe *StochasticDailyJumperEvent*, que similar a *StochasticPeriodicJumperEvent*, permite criar vários eventos com esta característica, espaçados em um período de 24h. Como o pomar vai estar constantemente monitorado e pesticidas serão usados sempre que necessário, nunca o nível de insetos será muito alto. Assim, forma considerados quatro insetos por dia nesta primeira modelagem, representados por quatro instâncias de *StochasticDailyJumperEvent*.

Tabela 1. Caracterização do modelo proposto - Custo energético

Parâmetro	Custo (mAs)
CPUCost	720 x NROI
ComCost	0,038016
IdleCost	38

A Tabela 1 apresenta os custos energéticos dos três modos de operação dos nodos sensores correspondente ao modelo de solução proposto. O custo de processamento neste caso terá um custo dependente do número de ROIs (NROI) na imagem. Cada nodo terá este controle e na descarga da bateria este número será considerado. Assim, o custo *CPUCost* é 720 multiplicado pelo NROI. Para processar a imagem de 1 ROI, a Raspberry

trabalha a 100% de sua capacidade, consumindo 800mA por 0,9 segundos, o que equivale a uma descarga de 720 mAs. O custo de *idle* considera o consumo do módulo RF quando não está transmitindo, mas está com recepção habilitada (28 mA - *idle*) acrescido do consumo do sensor infra-vermelho que consome 10 mA. O custo de comunicação refere-se ao consumo do módulo RF quando transmitindo mensagem. Este custo é mensurado considerando o tamanho do pacote (36 Bytes), taxa de transmissão do módulo (250 Kbps) e a corrente elétrica de transmissão (33 mA), ambos informados no *datasheet* do equipamento.

6. Experimentos e Resultados Preliminares

O modelo proposto foi empregado em experimentos de simulação visando avaliar o tempo de vida da rede considerando a carga de trabalho representada neste modelo. Os experimentos foram repetidos trinta vezes e os resultados médios foram considerados nas análises. O teste estatístico *t-student* foi empregado para avaliar a significância dos resultados obtidos e as médias das trinta execuções foram comparadas.

O modelo proposto com sua caracterização de custos foi empregado em um primeiro conjunto de experimentos. Em média a solução permanece ativa por sete dias. Na carga modelada acontecem 4 eventos por dia, mas alguns destes eventos podem ocorrer fora da área de cobertura dos nodos e portanto não são detectados pela rede.

Embora dentre os custos considerados no modelo proposto, o maior deles seja associado ao processamento, como as armadilhas são reativas, estas só consomem este valor quando um evento acontece e aciona-se a captura e classificação da imagem. Desta forma, o custo que se torna mais impactante para a eficiência das soluções é o custo referente ao modo *idle*. Isso porque em um maior número de ciclos de simulação a rede opera neste modo. Qualquer otimização que seja aplicada para reduzir o consumo do processamento ou da transmissão terá pouco impacto, se o custo de *idle* permanecer nesta proporção.

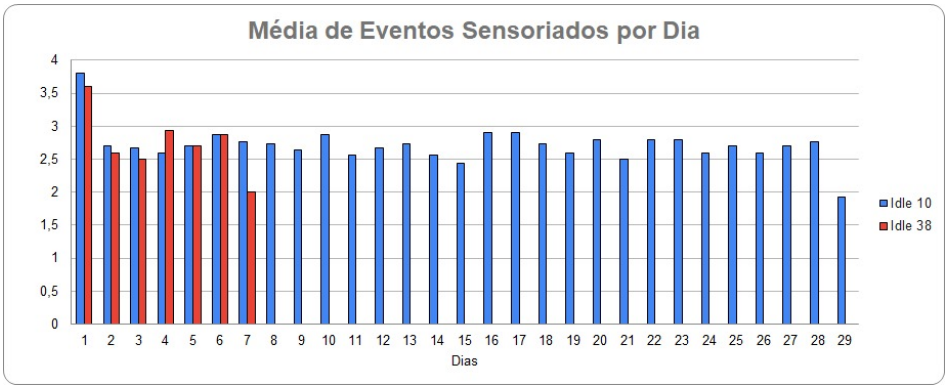


Figura 3. Número de eventos sensoreados por dia - comparativo das soluções

Uma maneira de reduzir este consumo seria colocar o módulo RF a operar em modo *sleep*, assim o nodo consumiria uma corrente de 10,001 mA durante os ciclos inativos. Porém, quando operando neste modo, o nodo não recebe mensagens, diferentemente do modo *idle* suportado pelo Eboracum atualmente. Então, para permitir a otimização de

consumo através deste quarto modo, o simulador precisa ser estendido de forma a permitir simular o chaveamento de nodos do modo *sleep* para modo *idle* quando mensagens tiverem de ser capturadas e retransmitidas.

Para demonstrar esta possível otimização e prospectar sobre o seu ganho, foi realizado um segundo conjunto de experimentos, considerando este valor reduzido de consumo no lugar do custo *idle* no modelo de simulação. Os resultados apresentados na Figura 3 representam a média de eventos sensoreados pela rede por dia nas duas soluções, idle de 38 e idle de 10, e permitem observar um aumento significativo no tempo de vida da rede com a redução do custo de *idle*. No entanto, é preciso destacar que neste segundo conjunto de experimentos foi considerada uma simplificação dos custos energéticos do módulo RF, a qual despreza o custo das recepções. Para avaliar com maior precisão o ganho de uma solução de RSSF reativa que mantém os nodos em modo *sleep* sempre que possível, migrando para *idle*, processamento ou comunicação somente quando necessário, o simulador precisará ser estendido e o modelo de simulação deverá prever quando o chaveamento de *sleep* para *idle* ocorrerá.

7. Conclusão e trabalhos futuros

O simulador Eboracum provê uma infra-estrutura extensível para a modelagem e simulação de RSSF. Neste trabalho é proposta a modelagem de RSSF dedicadas ao manejo de pragas no simulador Eboracum. Neste modelo os nodos são armadilhas eletrônicas que atraem e contabilizam moscas da fruta, auxiliando assim no controle de pragas em pomares.

O artigo contribui para a discussão da modelagem de aspectos da aplicação em um simulador de alto nível de abstração. Ainda, o presente trabalho contribui com extensões para o simulador Eboracum, além de caracterizações realizadas para consumo energético que serão de suma importância a trabalhos que visem a otimização soluções dedicadas para esta aplicação alvo para estender ao máximo as baterias dos nodos.

A distribuição de probabilidade atualmente empregada para determinar a ocorrência dos eventos no tempo, embora relacione a presença do inseto ao horário do dia pode ser aprimorada através de discussão com especialistas. Como trabalhos futuros, pretende-se discutir a definição de um modelo mais preciso.

Os transmissores disponíveis no mercado, inclusive o S2C, possuem um recurso de modo *sleep*, onde o consumo reduzido. Nossos experimentos demonstraram que o tempo de vida da rede pode ser aumentado significativamente ao manter os rádios neste modo. Porém, este modo não é suportado pelo simulador atualmente, que considera apenas o modo *idle* onde o radio está apto a receber mensagens e portanto tem um consumo mais alto. A inclusão do suporte ao modo *sleep* no Eboracum permitirá uma avaliação mais precisa de soluções otimizadas para esta aplicação alvo, bem como representará uma melhoria significativa no simulador como uma ferramenta para exploração do espaço de soluções dedicadas de RSSF.

8. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

(2019). Eboracum. Disponível em: `https://sourceforge.net/projects/eboracum/`. Acesso em: 25 de julho 2021.

Alsharif, M. H., Kim, S., e Kuruoğlu, N. (2019). Energy harvesting techniques for wireless sensor networks/radio-frequency identification: a review. *Symmetry*, 11(7):865.

Digi International Inc. (2020). Xbee/xbee-pro s2c zigbee: Rf module. Disponível em: `https://www.digi.com/resources/documentation/digidocs/pdfs/90002002.pdf`. Acesso em: 13 de julho 2021.

Ding, W. e Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Computers and Electronics in Agriculture*, 123:17–28.

Freitas, L., Martins, V., Ferreira Jr, P., e Brisolara, L. (2019). Aplicação de redes de sensores sem fio para o manejo de pragas em pomares: Avaliação de módulos de comunicação. In *CONGRESSO DE INICIAÇÃO CIENTIFICA DA UFPEL*.

Haghighi, M. e Cliff, D. (2013). Multi-agent support for multiple concurrent applications and dynamic data-gathering in wireless sensor networks. In *7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS*.

Manikandan, S. e Jayapriya, P. (2018). Precision agriculture using wireless sensor network system: Opportunities and challenges. *International Journal Of Engineering And Computer Science*.

Marques, M. R. S., Brisolara, L., Jr., P. R. F., e Indrusiak, L. S. (2016). Eboracum: An extensible framework for high-level modeling and evaluation of reactive and adaptable wsns. In *IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8.

Martins, V. A. M., Freitas, L. C., de Aguiar, M. S., de Brisolara, L. B., e Ferreira, P. R. (2019). Deep learning applied to the identification of fruit fly in intelligent traps. In *IX Brazilian Symposium on Computing Systems Engineering (SBESC)*, pages 1–8.

Mello, B. A. d. e Caimi, L. L. (2008). Simulação na validação de sistemas computacionais para a agricultura de precisão. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 12:666 – 675.

Murtaza, G., Ramzan, M., Ghani, M. U., Munawar, N., Majeed, M., Perveen, A., e Umar, K. (2019). Effectiveness of different traps for monitoring sucking and chewing insect pests of crops. *Egyptian Academic Journal of Biological Sciences. A, Entomology*, 12(6):15–21.

Parsons, L., Ross, R., e Robert, K. (2019). A survey on wireless sensor network technologies in pest management applications. *SN Applied Sciences*, 2(1).

Remboski, T., Souza, W., Aguiar, M., e Ferreira Jr, P. (2018). Identification of fruit fly in intelligent traps using techniques of digital image processing and machine learning. In *33rd Annual ACM Symposium on Applied Computing*, pages 260–267.

Sharma, S., Kooner, R., e Arora, R. (2017). Insect pests and crop losses. In *Breeding insect resistant crops for sustainable agriculture*, pages 45–66. Springer.

Shelton, A. e Badenes-Pérez, F. (2006). Concepts and applications of trap cropping in pest management. *Annual review of entomology*, 51:285–308.

Zhang, Y., Chen, D., Wang, S., e Tian, L. (2018). A promising trend for field information collection: An air-ground multi-sensor monitoring system. *Information Processing in Agriculture*, 5.

Zigbee Alliance. Zigbee specification. Disponível em: http://www.zigbee.org/. Acesso em: 25 de setembro 2021.



Modelos de aprendizagem de máquina para predição da presença de desoxinivalenol em grãos de trigo por meio de análises multiespectrais

Eduarda Zanini[1], Deividi Felipe Zaions[1], Alexandre Lazaretti Zanatta[1]; Carlos Amaral Hölbig[1]; Willingthon Pavan[2]

[1]Programa de Pós Graduação em Computação Aplicada (PPGCA) – Universidade de Passo Fundo (UPF) - 99052-900 - Passo Fundo - RS – Brasil

[2] International Fertilizer Development Center - 46 David Lilienthal Dr, - Muscle Shoals, AL 35661 - USA

{158223, 185579, zanatta, holbig,}@upf.br, wpavan@ifdc.org

Abstract. *This work aimed at an introductory study applied in agriculture, concerning data modeling to predict the condition of wheat grains through multispectral analysis of the AS7265x sensor. Ten commercial samples and 1 healthy sample of wheat grains were analyzed. Among commercial samples, 8 of them are contaminated with deoxynivalenol (DON). Fifty readings were collected from each typifying, totaling 550 data inputs. The paper does a performance analysis of supervised machine learning algorithms: k- nearest neighbors, support vector machine and random forest. As a result, the random forest algorithm had the highest performance among the 3. From the results, we conclude that this solution is a first step to help a wheat farmer in the decision-making process.*

Keywords: *Data Modeling. Machine Learning. K- Nearest Neighbors. Support Vector Machine. Random Forest.*

Resumo. *Este trabalho apresenta um estudo sobre modelagem de dados para predizer o estado dos grãos de trigo através da análise multiespectral do sensor AS7265x. Foram analisadas 10 amostras comerciais e 1 amostra sadia de grãos de trigo. Dentre amostras comerciais, 8 estão contaminadas com desoxinivalenol (DON). Foram coletadas 50 leituras de cada amostra, totalizando 550 entradas de dados. O trabalho faz uma análise do desempenho dos algoritmos supervisionados de machine learning: k- nearest neighbors, support vector machine e random forest. Como resultado, o algoritmo random forest teve o melhor desempenho entre os 3. Com base nos resultados conclui-*

se que, esta solução é um primeiro passo para ajudar um produtor de trigo no processo de tomada de decisão.

Palavras chave: *Modelagem de Dados. Machine Learning. K- Nearest Neighbors. Support Vector Machine. Random Forest.*

1. Introdução

A tecnologia da informação está presente em muitos setores, como por exemplo, na agricultura para avaliação da qualidade da colheita, previsões de rendimento, detecção de doenças, e demais aspectos importantes que auxiliam na tomada de decisão dos produtores e que tornam esse setor cada vez mais competitivo e eficiente.

O trigo é um dos cereais mais produzidos no mundo, justamente por conta da sua grande adaptação com o solo e o clima. No Brasil, o trigo tornou-se uma opção de grande valor econômico, principalmente para os produtores que utilizam mão de obra familiar, pelo seu bom preço de mercado e à significativa demanda pelo produto por ser utilizado em diversos tipos de preparações. (CONAB, 2017)

Nesse sentido, percebeu-se a necessidade de analisar o estado dos grãos do trigo por meio da análise multiespectral de amostras coletadas, aplicando os conceitos de modelagem de dados e *machine learning*. O local de maior relevância para os cultivares de inverno, entre eles o trigo, é a Região Sul do Brasil, justamente por seu clima temperado, que de certa forma, favorece o desenvolvimento desses cereais, pois fornece adaptabilidade a eles em relação aos seus centros de origem. (CONAB, 2020)

A doença giberela (*Gibberella zeae*), é um tipo de fungo que com frequência é encontrado no cultivar do trigo sendo responsável pela redução do rendimento e a qualidade dos grãos e derivados. Também, pode provocar danos à saúde humana e animal, resultando na produção de micotoxinas. As principais micotoxinas são a desoxinivalenol (DON), a zearalenona (ZEA) e o nivalenol (NIV). (Embrapa, 2016)

A espectroscopia é uma ferramenta muito explorada na área da agricultura ao longo de 50 anos (Embrapa, 2018). Como a resposta espectral das plantas se altera conforme seus índices nutricionais, existem outras ferramentas e técnicas para manejo dentro da computação que também auxiliam na informação e segurança para a tomada de decisão como o uso de inteligência artificial.

Diante do exposto, tem-se o seguinte problema de pesquisa: *"Como auxiliar os agricultores no processo de identificação do estado do cultivar de trigo através de análises multiespectrais de amostras e predizer a partir de algoritmos de machine learning se um determinado grão esta sadio ou contaminado com a micotoxina (DON)".* Assim, o objetivo deste trabalho é introduzir uma abordagem sobre modelagem de dados com algoritmos de *machine learning* para predição do estado dos grãos de trigo por meio da coleta de amostras por um sensor multiespectral, bem como, avaliar o desempenho dos algoritmos propostos para a solução do problema.

Para alcançar o objetivo pretendido, o trabalho divide-se em cinco capítulos. No primeiro capítulo, que é a introdução, é apresentado o tema da pesquisa, sua justificativa de estudo, objetivo geral e os específicos e o questionamento norteador do estudo. Na sequência é apresentada a fundamentação teórica da pesquisa que aborda sobre, espectroscopia na identificação de características de amostras, como é realizado o processo das técnicas de *machine learning* e como funcionam os algoritmos: *k- nearest*

neighbors, *support vector machine*, e *random forest*. No terceiro capítulo apresenta-se a aplicação, descrevendo os métodos utilizados e a discussão sobre os resultados e por fim, no quarto capítulo apresentam-se as considerações finais e o quinto as referências.

2. Revisão de literatura

2.1. Espectroscopia

A espectroscopia é um estudo que remete a compreensão da geração da radiação eletromagnética (uma forma de energia propagada na forma de ondas eletromagnéticas) e da sua interação com a matéria. Se divide em muitas áreas que se dedicam a estudar faixas relativamente estreitas do espectro eletromagnético, de acordo com suas energias e, com os fenômenos que elas podem produzir ao interagir com a matéria. Dentre as diversas formas de interação da radiação com a matéria, a absorção da radiação pelos constituintes da amostra é de grande interesse para a espectroscopia analítica, pois este fenômeno gera os espectros de absorção que contém as informações analíticas qualitativas e quantitativas a respeito de uma amostra. (Embrapa, 2018)

As ondas eletromagnéticas conseguem ser classificadas com base nos seus diversos comprimentos de onda/frequências, que são denominadas como espectro eletromagnético. (Hollas, 2004)

Uma imagem digital é a representação da figura de um objeto pela combinação da intensidade dos raios de luz provenientes da mesma. Uma imagem espectral é aquela que reproduz, a partir da análise de um objeto, o comprimento de onda do objeto em questão. (Habibi, 2014). Os sensores ópticos de câmeras comuns possuem uma faixa de captação do comprimento de onda eletromagnético dentro espectro visível, a qual varia entre 400 nm a 750 nm. (Zhou, 2019). Os sensores multiespectrais conseguem capturar frequências além da faixa de luz visível, ou seja, capturam dados de imagens com uma faixa de captação do comprimento de onda específicas em todo o espectro eletromagnético, e podem ser separados por filtros ópticos ou detectados por meio do uso de equipamentos que são sensíveis a comprimentos de onda específicos, é possível realizar, por exemplo, um levantamento do número de plantas em determinada área, verificar a saúde das plantas e detectar pragas na plantação. (Giannoni et al., 2018)

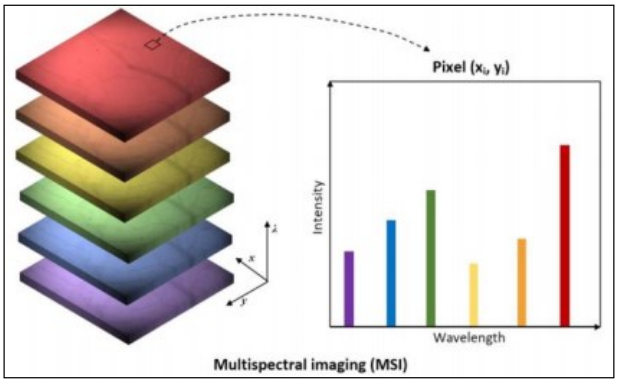


Figura 1. Representação do comprimento de onda de uma imagem multiespectral.

Fonte: Giannoni et al., 2018.

2.2. Machine Learning

O *machine learnig* (ML) tem com o objetivo criar novas possibilidades para resolver, quantificar e compreender processos intensivos de dados em ambientes operacionais de determinadas áreas como a agricultura. O ML envolve um processo de aprendizagem, a partir de dados de treinamento para realizar uma tarefa. Para calcular o desempenho dos modelos e algoritmos de ML, vários modelos estatísticos e matemáticos são usados. Após o final do processo de aprendizagem, o modelo treinado pode ser usado para classificar, prever ou agrupar novos exemplos. As técnicas de ML podem ser divididas em diferentes categorias amplas dependendo do tipo de aprendizado classificam-se em: supervisionado, não supervisionado ou por reforço. (Liakos et al., 2018)

2.2.1. Aprendizado supervisionado

As redes de aprendizado supervisionado aprendem a partir de dados de treinamento pré-classificados, e assim a partir desse classificador é utilizando como exemplo, onde contêm a informação da saída esperada e por fim conseguem classificar os dados de entrada mais precisamente (Coppin, 2015).

2.2.2. Aprendizado não supervisionado

O aprendizado não supervisionado aprende a classificar um conjunto de dados de entrada, sem informação alguma sobre quais são as classificações e sem receber nenhum dado de treinamento, o próprio aprendizado irá identificar padrões e classificá-los (Coppin, 2015).

2.2.3. Aprendizado por reforço

O aprendizado por reforço, posiciona-se entre o aprendizado supervisionado e o aprendizado não supervisionado. Esses métodos costumam ser úteis quando apenas rótulos incompletos estão disponíveis, ou seja, os dados de entrada podem estar apenas parcialmente disponíveis com alguns dados de saídas ausentes (VanderPlas, 2016).

2.2.4. Algoritmo *k-nearest neighbors*

O algoritmo *K-Nearest Neighbor* (KNN), do português "K-Vizinho Mais Próximo", é um método de aprendizado com base em instâncias, que consiste em, armazenar os dados de treinamento e os usar para definir uma classificação para cada dado novo de entrada. (Coppin, 2015) O KNN é um algoritmo supervisionado do tipo classificador não paramétrico, possuí três elementos principais: um conjunto de dados para exemplo, uma métrica de distância e o valor k número de vizinhos. (Oliveira, 2017). Cada instância pode ser formada por um vetor de *n* dimensões, onde *n* é o número de atributos usados para descrever cada instância e as classificações para valores numéricos discretos. Os dados de treinamento são armazenados, e quando uma nova instância é encontrada ela será comparada aos dados de treinamento para encontrar os seus vizinhos mais próximos. E isso é realizado pela computação chamado de distância Euclidiana entre instâncias em um espaço de *n* dimensões (Coppin, 2015).

2.2.5. Algoritmo *Support Vector Machine*

O algoritmo *Support Vector Machine* (SVM), do português "Máquina de Vetor de Suporte" têm como finalidade a especificação de limites de decisão que produzam um ótimo desempenho entre classes por meio da minimização dos erros. O funcionamento do SVM serve para problemas de reconhecimento de padrão, e aplica uma teoria estatística de aprendizagem, encontra uma linha de separação, chamada de hiperplano

entre dados de duas classes. Assim buscando maximizar a distância entre os pontos mais próximos em relação a cada uma das classes. Esse algoritmo possui quatro funções: linear, quadrática, polinomial e função de base radial. (Vapnik, 2009). Neste estudo será utilizada a função linear, justamente com base na simulação da distribuição das amostras coletadas.

2.2.6. Algoritmo *Random Forest*

No algoritmo *random forest* são construídas várias árvores de decisão para classificar um novo dado. A partir disso, será construída a primeira árvore de decisão, sendo definido o primeiro nó da árvore, que será a primeira condição analisada e criará os dois primeiros ramos, para isso, é necessário aplicar a função de entropia ou o índice *Gini*, justamente para escolher a melhor variável para compor o nó raiz. O algoritmo definirá aleatoriamente duas ou mais variáveis, e então realizará os cálculos com base nas amostras selecionadas para definir qual dessas variáveis será utilizada no primeiro nó.

Para escolha da variável do próximo nó, novamente serão escolhidas duas (ou mais) variáveis, excluindo as já selecionadas, e o processo de escolha se repetirá. Desta forma, a árvore será construída até o último nó. Quanto mais árvores criadas, melhores serão os resultados do modelo, até determinado ponto, onde uma nova árvore não conseguirá uma melhora significativa no do modelo. Cada árvore tem o seu resultado. O resultado que mais vezes for gerado será o optado. (Didática Tech, 2020)

3. Metodologia

Para a parte física do dispositivo, foi escolhido o sensor *AS7265x Smart Spectral Sensor* do fabricante AMS para realizar a análise das amostras de trigo, pois possui tecnologia multiespectral que, forma um conjunto de chips do sensor multiespectral de 18 canais *AS7265x*. Sendo visível (VIS) e infravermelho próximo (NIR) de 410nm a 940nm cada com 20nm FWHM. (AMS, 2018). A partir da decisão exposta sobre o uso do sensor, foi realizada a impressão de modelos em impressoras 3D, para realizar a coleta de dados espectrais sobre as amostras de trigo.

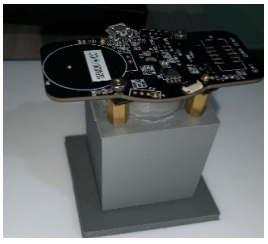


Figura 2. Modelo 3D para coleta das amostras multiespectrais.

Na coleta de dados foram analisadas pelo sensor dez (10) amostras comerciais e uma amostra sadia de grão de trigo. Dentre as amostras comerciais, oito (8) estão contaminadas com níveis diferentes de (DON). Foram coletadas 50 leituras pelo sensor espectral sobre todas as amostras, totalizando 550 entradas de dados. Destes, aplicou-se uma razão 70/30 para treinamento e validação. Na Tabela 1 podem ser observados os índices de (DON) identificados nas amostras.

Tabela 1. Índices de (DON) identificados nas amostras.

	21_195	21_196	21_197	21_198	21_199	21_200	21_201	21_202	21_210	21_227	Sadio

DON (ug/kg)	1788	483,6	2113,8	1508,1	2009,1	1943,4	0	0	799,2	307,5	0

Para a aplicação dos algoritmos utilizou-se das bibliotecas *scikit-learn*, *pandas* e *numpy* na linguagem de programação Python. A partir da modelagem dos dados, para a aplicação dos algoritmos de ML é necessário normalizar estes dados para que os modelos não fiquem com maiores ordens de grandeza, ou seja, normalizar dados tem como objetivo colocar as variáveis dentro do intervalo de 0 e 1. A função que se aplica está operação é dada pela Equação 1:

$$Z = x\text{-}min(x) / (max(x)\text{-}min(x)) \quad (1)$$

A aplicação da função de normalização apresentou os dados dos espectros L 940nm e A 410nm com valores zerados, sendo que, estes dados foram removidos dos testes, pois estavam influenciando no treinamento da rede, o que pode ser caracterizado por *outliers*. Após, foram aplicados os testes de classificação para os três algoritmos: KNN, SVM e *random forest.* Na Tabela 2 pode ser observado um comparativo do desempenho da classificação por decisão binária dos grãos de trigo. Todas as amostras cujos níveis de (DON) foram verificados tem uma classificação de contaminados (1) e todas as demais se classificam como sadio (0).

Tabela 2. Teste de decisão binária (sadio ou contaminado).

	KNN	SVM	Random Forest
Accuracy	0.84	0.89	0.92
Recall (sadio)	0.59	0.70	0.85
Recall (contaminado)	0.94	0.97	0.94
Precision (sadio)	0.79	0.89	0.85
Precision (contaminado)	0.85	0.89	0.94

Dos testes realizados verificou-se que o algoritmo *random forest* teve um maior desempenho de acurácia, ou seja, a proporção de previsões que o modelo classificou corretamente com ele. Esse algoritmo também conseguiu uma maior precisão de contaminados, a proporção de identificações positivas estava realmente correta para contaminados. No entanto, o SVM teve uma maior precisão para dados sadios. A maior sensibilidade (*Recall*) de sadios também foi o *random forest*. Já a sensibilidade para dados contaminados tanto o algoritmo KNN quando *random forest* resultaram no mesmo valor, e o SVM teve o melhor resultado nesta métrica. Além desses dados a partir da matriz confusão de cada um o *random forest* teve menor índice de dados falsos negativos, ou seja, o resultado que o modelo previu como incorreto é na verdade positivo. Por fim o melhor modelo para esse tipo de classificação dentre os três foi o *random forest*.

Para os testes de classificação a partir dos níveis de DON como classificador, os seguintes resultados estão disponíveis na Tabela 3.

Tabela 3. Teste de classificação.

	KNN	SVM	Random Forest
Accuracy	0.70	0.68	0.82
Recall (sadio)	0.72	0.63	0.87
Precision (sadio)	0.85	0.89	0.82

Embora o KNN tenha sido efetivo em identificar corretamente as amostras sadias, o *random forest* também teve o melhor desempenho para classificar as amostras de acordo com seu nível de contaminação.

4. Conclusões e trabalhos futuros

Este trabalho mostrou como modelar dados coletados por meio de análises multiespectrais dos grãos de trigo e aplicou os algoritmos de *machine learning*: KNN, SVM *e random forest.* A partir dos dados previu se os níveis desses grãos estão sadios ou contaminados. Conclui-se, que o *random forest* foi o algoritmo que teve a melhor *performance* em comparação com os demais.

Por fim, nota-se que as taxas de acerto e erros relativos obtidos desses algoritmos em geral ainda são razoáveis e que a pesquisa carece de evolução, por exemplo, utilizou-se um "número muito pequeno" de grãos de trigo dissipados ou não com uma micotoxina desoxinivalenol. Por outro lado, essa abordagem atingiu o objetivo do trabalho, pois é um primeiro passo para uma solução que ajude os agricultores no processo de tomada de decisão.

Como trabalho futuro, buscar outras classificações, não se limitando a apenas em duas classes (não contaminado e contaminado). Também, buscar-se-á analisar mais amostras deste sensor e validar os mesmos algoritmos propostos para qualificar a performance dos resultados obtidos, bem como, usar o método *random forest* para outros tipos de classificadores e outros tipos de cultivares.

5. Referências

AMS. (2018). AS7265x Smart 18-Channel VIS to NIR Spectral_ ID 3-Sensor Chipset with Electronic Shutter. 63 p. Disponível: https://ams.com/documents/20143/36005/AS7265x_DS000612_1-00.pdf/08051c8a-a7f6-6231-7993-2d3fe0bf38b8. Acesso: junho/2021.

Companhia Nacional de Abastecimento. (2017). A cultura do trigo. CONAB. 218 p. Disponível: https://www.conab.gov.br/uploads/arquivos/17_04_25_11_40_00_a_cultura_do_trigo_versao_digital_final.pdf. Acesso: junho 2021.

Companhia Nacional de Abastecimento. (2020). Acompanhamento da safra brasileira de grãos, v. 7 - Safra 2019/20 - Décimo segundo levantamento. CONAB. 45 p. Disponível: https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analises-do-mercado/historico-mensal-de-trigo. Acesso: junho, 2021.

Coppin, B. (2015). Inteligência Artificial / Ben Coppin; tradução e revista técnica Jorge Duarte Pires Valério. [Reimpr.]. Rio de Janeiro: LTC. Tradução de: Artificial intelligence illuminated, 1st ed.

Didática Tech. (2020). Inteligência Artificial & Data Science. O que é e como funciona o algoritmo RandomForest. Disponível: https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-randomforest/. Acesso: junho/2021.

Empresa Brasileira de Pesquisa Agropecuária. (2016). Trigo: o produtor pergunta, a Embrapa responde / Claudia De Mori et al., editores técnicos. Brasília, DF: Embrapa. 309 p. Disponível: https://www.embrapa.br/busca-de-publicacoes/-

/publicacao/1040211/trigo-o-produtor-pergunta-a-embrapa-responde. Acesso: junho, 2021.

Empresa Brasileira de Pesquisa Agropecuária. (2018). Espectroscopia no Infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos / Casiane Salete Tibola et al., editores técnicos. Brasília, DF: Embrapa. 200 p.

Giannoni, L. et al. (2018). Hyperspectral imaging solutions for brain tissue metabolic and hemodynamic monitoring: past, current and future developments. Computer Science, Medicine. Journal of Optics, v. 20, n. 4. p. 25. Disponível: https://iopscience.iop.org/article/10.1088/2040-8986/aab3a6/meta. Acesso: julho/2021.

Habibi, M. (2014) Image sensors. In: Measurement, Instrumentation, and Sensors Handbook. 2. ed. [S.l.]: CRC Press. cap. 4. p. 1921.

Hollas, J. M. (2004). Modern Spectroscopy. Hoboken: John Wiley & Sons, Ltd. 482 p.

Liakos, K.G. et al. (2018). Machine Learning in Agriculture: A Review. *Sensors*. *18*, 2674. Disponível: https://doi.org/10.3390/s18082674. Acesso: junho/2021.

Oliveira, A.R. and Roesler. (2017). Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes. ELSA-Brasil: accuracy study. Sao Paulo Medical Journal, v.135, n. 3, p. 234-46.

VanderPlas, J. (2016). Python Data Science Handbook. Disponível: https://jakevdp.github.io/PythonDataScienceHandbook/. Acesso: junho/2020.

Vapnik, V. (2009). The Nature of Statistical Learning Theory. 2. nd. New York: Springer. 745 p.

Zhou, X. et al. (2019). A novel combined spectral index for estimating the ratio of carotenoid to chlorophyll content to monitor crop physiological and phenological status. International Journal of Applied Earth Observation and Geoinformation, v. 76, p. 128–142. Disponível: https://www.sciencedirect.com/science/article/abs/pii/S030324341830566X. Acesso: junho/2020.



Monitoramento de plantas em casas de vegetação para assimilação de dados

Monique Pires Gravina de Oliveira[1], Luiz Henrique Antunes Rodrigues[1]

[1]Faculdade de Engenharia Agrícola – Universidade Estadual de Campinas (Unicamp)

moniquepgoliveira@gmail.com, lique@unicamp.br

Abstract. *Data in agriculture is being produced by new sources with new roles and great potential. They may be captured on the field using the internet of things or in plant factories and they may be used on digital twins, which aim at connecting observations with a model of the system. This connection may happen using data assimilation and, particularly in controlled environments, plants may be more continuously monitored, allowing for more data to be available. In this work, we performed data assimilation in a tomato model using data from cameras and load cells, observing these sources provide good estimates of aboveground biomass and that assimilation improves estimates obtained by the non-calibrated Reduced-State Tomgro model.*

Resumo. *Mais dados têm sido gerados em aplicações agrícolas por novas fontes e com grande potencial de uso. Obtidos no campo ou em fazendas verticais, eles podem ser usados, por exemplo, em gêmeos digitais, que visam conectar observações a um modelo do sistema. Essa conexão pode ocorrer por assimilação de dados e em ambientes protegidos, em que as plantas podem ser monitoradas mais intensamente, mais dados estariam disponíveis. Neste trabalho, realizamos assimilação em um modelo de tomateiro usando dados coletados por câmeras e por células de carga, observando que essas fontes fornecem boas estimativas de biomassa da parte aérea e que a técnica melhora as estimativas obtidas pelo modelo Tomgro Reduzido sem calibração.*

Introdução

O uso de dados em modelagem tem sido tradicionalmente associado ao começo do ciclo de modelagem, no seu desenvolvimento, ou ao fim, para calibração e avaliação dos modelos. Existem, no entanto, novas expectativas para seu uso que são baseadas em sua abundância e qualidade. [Keating and Thorburn 2018] mencionaram uma nova forma de se relacionar com dados que expandiria esse papel por meio de, por exemplo,

modelagem inversa e fusão de modelo e dados, incluindo novos papéis para sensoriamento remoto e proximal em sua inicialização e calibração.

A assimilação de dados é um método amplamente utilizado em hidrologia e meteorologia e consiste em combinar observações aos estados estimados pelos modelos, levando em consideração a incerteza que existe nas estimativas dos modelos e nas observações [Pellenq e Boulet 2004]. No caso da modelagem de culturas agrícolas, a assimilação de dados tem sido usada com imagens de sensoriamento remoto para estimar variáveis de estado relacionadas ao dossel ou a propriedades do solo, visando obter melhores estimativas de produtividade, índice de área foliar e umidade do solo [Dorigo et al. 2007; Jin et al. 2018].

Revisões recentes discutem as diferentes abordagens para realizar a assimilação e suas limitações [Huang et al. 2019; Jin et al. 2018]. Elas apresentam os aspectos teóricos das técnicas bem como os diversos algoritmos que têm sido utilizados, incluindo diferentes contextos e objetivos. Dentre as limitações elencadas pelos autores, frequentemente são apontadas questões relacionadas ao uso em grandes áreas, visto que assimilação de dados tem tido sua aplicação frequentemente limitada às resoluções temporais e espaciais de satélites. Trabalhos recentes exploraram outras fontes de dados, como veículos aéreos não-tripulados (VANTs) e imagens digitais. [Yu et al. 2020] se valeram dessas outras fontes para melhorar as estimativas de produtividade por meio da assimilação da altura da cana-de-açúcar. [Linker and Ioslovich 2017] usaram imagens digitais do dossel para estimar a cobertura do dossel a ser assimilada no modelo Aquacrop, bem como dados de análises destrutivas como observações de biomassa. Dados de análises destrutivas também foram usados por [Ruíz-García et al. 2014] com alfaces e o modelo Nicolet.

Dado que é possível monitorar mais intensamente o crescimento da vegetação em cultivo protegido, também é possível, com maiores resoluções temporais e espaciais, realizar a assimilação com observações coletadas automaticamente em modelos de crescimento de hortaliças e outras culturas normalmente cultivadas em estufas agrícolas e outros ambientes protegidos. Medições automáticas relacionadas à planta em ambientes protegidos tem sido aplicada para obtenção de diferentes informações. O monitoramento do crescimento da planta por meio de células de carga foi registrado na Holanda por [De Koning and Bakker 1992] e sugerido por trabalhos como [Helmer et al. 2005; Lee and Son 2019] e em [Chen et al. 2016], para fazendas verticais. Outras aplicações possíveis estão relacionadas à determinação da demanda de água pela planta [De Graaf et al. 2004] e em "gêmeos digitais" [Verdouw et al. 2021]. A assimilação dessas observações poderia aprimorar estratégias de modelagem de [Marcelis et al. 2000], com informações de feedback por sensores, ou complementar as estratégias que têm explorado principalmente métodos de aprendizado de máquina para lidar com dados em tempo real [Gong et al. 2021; Hemming et al. 2019, 2020]. E como a variabilidade da produção em estufas pode ser detectada até mesmo por modelos [Bojacá et al. 2009], esta abordagem pode melhorar a avaliação da produção pelos produtores.

Como a assimilação de dados com dados de diferentes fontes pode se mostrar útil neste ambiente, valendo-se do monitoramento intenso e de modelos de crescimento vegetal, e como o tomate é uma cultura mais estudada e relevante para a economia, é possível que técnicas de assimilação de dados sejam capazes de melhorar as estimativas

obtidas por um modelo de crescimento de tomateiros mesmo sem calibração. O objetivo deste trabalho foi então investigar, com os dados obtidos em um ambiente de produção, se uma técnica de assimilação seria capaz de aprimorar as estimativas de biomassa da parte aérea realizadas pelo modelo Tomgro Reduzido (Reduced State Tomgro) [Jones et al. 1999].

Material e Métodos

Modelo de crescimento

O modelo Tomgro Reduzido (RT) é um modelo que resume as principais características do modelo TOMGRO e foi desenvolvido para utilização em sistemas de controle de cultivos protegidos. O modelo possui apenas cinco variáveis de estado — número de nós, índice de área foliar, biomassa seca da parte aérea, biomassa seca de frutos e biomassa seca de frutos maduros — e foi desenvolvido visando o controle do ambiente. É um modelo simples que não inclui crescimento radicular ou irrigação, e essa simplicidade pode ajudar nesta primeira abordagem.

Com base em dados horários de temperatura e radiação solar, o modelo quantifica o crescimento e o desenvolvimento do tomateiro quando a água e os nutrientes não são limitantes. Seu desenvolvimento é indicado por meio do número de nós e o seu crescimento, por meio dos demais estados. O cálculo da biomassa seca da parte aérea se dá pela diferença entre a fração de carboidratos destinada à parte área da planta e a fração removida por senescência das folhas ou desfolha. A fração de carboidratos, por sua vez, é calculada como a diferença entre a fotossíntese e a respiração, multiplicada por um fator de que caracteriza a eficiência de conversão dos fotoassimilados em carboidratos, e por um coeficiente que indica qual a proporção da massa que não é destinada às raízes. Mais informações sobre variáveis e parâmetros podem ser encontradas em [Jones et al. 1999].

Coleta de dados

Foram realizados três ciclos de crescimento de minitomates, conduzidos em estufas agrícolas de pesquisa na Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas, localizada no município de Campinas, SP. O Ciclo 1 teve início em julho de 2019 e término em outubro de 2019, tendo sido utilizada a cultivar Fercan/Milla, o Ciclo 2 foi iniciado em novembro de 2020 e concluído em fevereiro de 2021, com a cultivar Feltrin/Carolina e o Ciclo 3 foi realizado entre março de 2021 e junho de 2021, com a cultivar Seminis/DRC-564.

Os dados ambientais medidos referem-se à temperatura do ar e à radiação fotossinteticamente ativa (RFA). A temperatura foi determinada por transdutores SHT75 protegidos por cápsulas porosas que, por sua vez, eram protegidas por tubos de policloreto de vinila revestidos com folha de alumínio. Os tubos incluíam ventiladores a jusante. Os sensores foram instalados em plataforma de hardware para redes de sensores sem fio (Radiuino BE900), com backup diário. Para RFA, foram utilizados sensores Licor LI-190SA com um datalogger Licor LI-1400.

As plantas foram caracterizadas por análises destrutivas e por observações não destrutivas. Os dados não-destrutivos referem-se às fotos tiradas lateralmente, por módulos fixos de câmera Raspberry Pi v2, conectados a um computador Raspberry Pi

Zero. Também se referem ao monitoramento contínuo — a cada minuto — de peso utilizando transdutores de força HBM S2M com força nominal de 10 N (precisão de 0,02%) e armazenados em um datalogger PMX WGX002, cartão de medida PX455. Nos Ciclos 02 e 03, o tempo total de irrigação foi definido como aquele que não permitiria déficit ao final do dia, para minimizar as flutuações de massa de um dia para o seguinte, causadas pela variação nas irrigações.

Os dados destrutivos referem-se à caracterização do peso seco da planta. Em intervalos de uma a três semanas, três plantas eram retiradas da casa de vegetação. Folhas, haste e frutos maduros e verdes foram separados para pesagem. O material vegetal foi pesado antes e depois da secagem por quatro dias ou até que o peso constante fosse atingido.

Antes da remoção, essas plantas foram fotografadas lateralmente com uma câmera de smartphone. Todas as fotos incluíram referências de tamanho. A anotação dos órgãos das plantas nas imagens foi feita manualmente, por meio do software GIMP[1]. Foram marcadas apenas as áreas em que havia confiança de que o órgão correspondia à planta correta, o que fazia com que se houvesse incerteza ou oclusão a área não fosse marcada. Os órgãos foram coloridos de forma diferente, sendo então detectados por um script em Python. A área correspondente foi calculada usando a referência.

Os três ciclos apresentaram desenvolvimentos distintos, que são explorados neste trabalho: o primeiro pode ser caracterizado por baixa irrigação, submetendo as plantas a déficit hídrico, o segundo, por excesso de fertilização com nitrogênio e ataque do ácaro do bronzeamento ao final do ciclo, e o terceiro ciclo foi conduzido de forma mais próxima ao potencial.

Os dados das análises destrutivas foram usados para calibrar o modelo de Tomgro Reduzido. A calibração seguiu a abordagem de minimizar o erro quadrado relativo. Dado que havia três ciclos diferentes, cada um com condições diferentes, as simulações denominadas calibradas usaram dados do respectivo ciclo. Simulações não calibradas usaram dados da calibração original do modelo.

Modelos de observação

Dados obtidos por formas não-destrutivas para as plantas submetidas às análises destrutivas, bem como os resultados destas análises foram usados para criar modelos de observação. Nossa estratégia de modelagem focou na obtenção de relações empíricas simples e o método dos mínimos quadrados generalizado foi usado para lidar com a heterocedasticidade e a correlação entre os resíduos. Para evitar vazamento de dados, apesar das diferentes condições de crescimento, os dados do ciclo não foram usados para obter a relação usada naquele ciclo. Todos os valores obtidos por planta foram multiplicados pela densidade de plantas para compatibilizar as unidades com os valores estimados pelo modelo para o estado.

Usamos a área e a altura extraídas de imagens laterais para inferir a biomassa acima do solo e o peso total, conforme determinado pelo sistema de pesagem para

[1] GNU Image Manipulation Program

inferir a biomassa acima do solo. Nesse caso, realizamos uma conversão que excluiria as raízes, dependendo do estágio de desenvolvimento, e que converteria de biomassa fresca em seca.

Assimilação de dados

Foi usada a técnica Unscented Kalman Filter (UKF), que requer especificações de incerteza que foram atribuídas como a seguir. A incerteza nos estados iniciais foi determinada como a variância das amostras analisadas no dia do transplantio. A incerteza no modelo para o UKF foi determinada como a raiz do erro quadrático médio calculado pela diferença entre as observações no ciclo e o valor simulado do modelo não calibrado. Embora estejamos cientes de que isso corresponde a vazamento de dados, acreditamos que essa foi a melhor maneira de fornecer uma estimativa para este parâmetro. A variância nas medições foi determinada como o erro padrão residual para o modelo de observação no ciclo. Acreditamos que esta seja uma forma de incorporar erros de medição e erros na representação do estado por meio dessa abordagem. Neste trabalho, a variável de estado estimada foi a biomassa seca da parte aérea.

Uma vez que nosso objetivo era determinar o quanto as estimativas do modelo não calibrado poderiam ser melhoradas assimilando observações de diferentes fontes, avaliamos nossa abordagem comparando as séries não calibradas, calibradas e filtradas.

Resultados e discussão

Modelos de observação

Na Figura 1, mostramos as correlações entre os pares de variáveis que deram base aos modelos de observação que obtivemos. É possível notar também as diferenças entre os ciclos. Notam-se a menor amplitude da biomassa total no Ciclo 1, decorrente dos períodos de estresse hídrico, e os maiores valores de biomassa aérea no Ciclo 3, provavelmente ligados ao menor estresse das condições de cultivo. Nos dois primeiros ciclos, ainda que a biomassa total produzida pudesse ser equivalente à do Ciclo 3, o estresse poderia levar uma parte maior dos assimilados produzidos a serem destinados à raiz.

A figura mostra que as três fontes utilizadas fornecem abordagens razoáveis para estimativa de biomassa. As menores correlações se referem ao uso da altura para estimativa da biomassa no Ciclo 1, em que apesar da altura máxima ter sido atingida, houve pouca produção de folhas e frutos, diminuindo a massa aérea, e ao uso da área de folha nas imagens para caracterização da biomassa no Ciclo 2, em que houve produção excessiva de folhas, inclusive em detrimento da produção de frutos. Pode-se perceber também que, exceto para a altura, as observações estão concentradas nos valores mais baixos, e se por um lado, isso distorceu a distribuição das observações e pode ter destacado observações não informativas, ao mesmo tempo, se a maior parte do ciclo é representada por esses valores baixos, é razoável usá-los para obter os modelos. Uma anotação automática de imagens permitiria monitorar mais plantas e tornar os resultados mais robustos e baseados em mais observações.

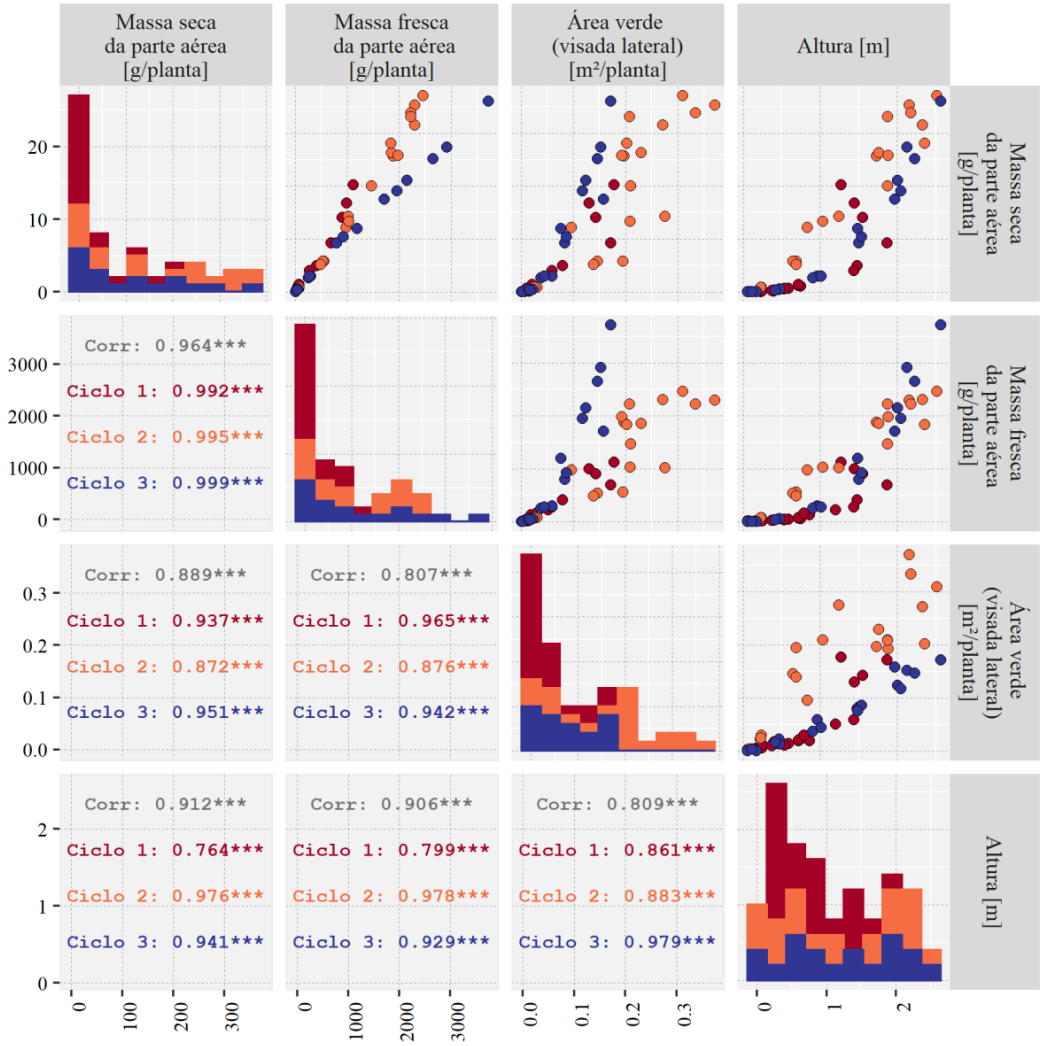


Figura 1. Os gráficos da parte superior mostram os gráficos de dispersão pareados comparando a biomassa da parte aérea às medidas diretas e indiretas obtidas em cada ciclo. Os gráficos da parte inferior apresentam a correlação geral e as específicas para cada ciclo e a diagonal mostra a distribuição dos dados, colorida de acordo com o ciclo de cultivo.

Assimilação de dados

Nossos experimentos levaram a resultados diferentes com relação ao aprimoramento das estimativas (Figura 2). Frequentemente, o UKF levou a melhorias nas estimativas ao longo do ciclo em comparação ao modelo não calibrado, incluindo resultados tão bons quanto os obtidos pela calibração. Entretanto, esse desempenho variou dependendo da fonte de dados utilizada.

O uso da altura permitiu o aumento das baixas estimativas realizadas pelo RT nos três ciclos. No Ciclo 1, os valores obtidos foram superiores aos das demais plantas no ambiente, enquanto nos Ciclos 2 e 3, embora os valores tenham sido compatíveis aos das demais no fim do ciclo, foram obtidos valores negativos para a massa na fase inicial do crescimento. Isso ocorreu porque a relação entre a biomassa e a altura foi sido tratada

como linear quando, na verdade, as observações do início do crescimento não têm esse comportamento. Já o uso das áreas verdes identificadas nas imagens também permitiu aumento das estimativas do modelo para o Ciclo 2, tornando-as bem próximas aos valores obtidos pelo modelo calibrado, mas não afetou grandemente o Ciclo 3. Neste caso, o modelo de observação obtido para o ciclo relacionou as áreas observadas a valores muito baixos de biomassa. Deste modo, não houve muitas diferenças entre as estimativas e as observações, o que não promoveu correções dos valores estimados. Mais ciclos de cultivo permitiriam a obtenção de modelos mais robustos, que melhor representassem as relações entre as observações e os estados. Por fim, a massa do sistema de pesagem ajustou quase perfeitamente o Ciclo 2 e melhorou as estimativas do Ciclo 3, ainda que subestimando o valor final. Para o Ciclo 1, as estimativas foram muito comprometidas em razão do erro na medição causado por flutuação da água durante o crescimento.

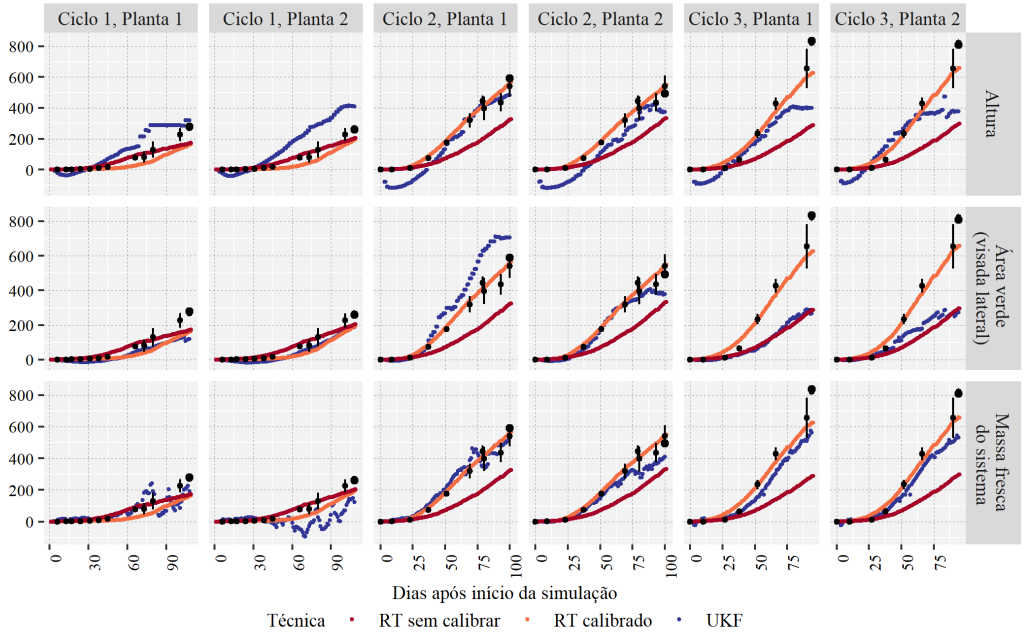


Figura 2. Curvas de crescimento de cada planta monitorada incluindo assimilação utilizando o UKF, simulação calibrada e simulação não calibrada. Os pontos se referem aos valores determinados nas análises destrutivas e a barra representa o desvio-padrão dessas amostras. O valor final da planta monitorada está representado por um ponto maior.

Um experimento preliminar de assimilação da biomassa seca da parte aérea no RT foi realizado por [Torres-Monsivais et al. 2017]. Embora os autores apresentem poucos detalhes das observações utilizadas, eles concluem que a simulação perturbada com ruídos de até 10% implica erros de estimativa aceitáveis após a aplicação do UKF. Todavia, os resultados dos autores se referem a uma avaliação em que os erros do modelo são muito baixos por premissa, o que não é necessariamente verdade para modelos não-calibrados. No nosso caso, os erros tanto das observações quanto do modelo atingem magnitudes muito maiores do que os valores simulados, particularmente por serem valores fixos para todo o ciclo. Além disso, a conclusão deles compara erros de magnitudes distintas e, sem uma avaliação dos erros relativos, é possível que ruídos maiores pudessem ser aceitáveis. Sendo assim, as conclusões são

limitadas e dificultam comparações. A despeito disso, estudos sobre valores adequados para os parâmetros de incerteza são necessários, uma vez que a decisão de como atribuir esses valores de incerteza é um problema em aberto e diversos autores a tomam de formas distintas: alguns autores trabalham com erros variáveis [Lu et al. 2021], outros usam valores fixos também relacionados aos erros do modelo de conversão [Huang et al. 2016], outros estimam com base na variância de observações coletadas em campo [Zhao et al. 2013] enquanto outros usam estratégias que independem dessa incerteza [Chen et al. 2018; Vazifedoust et al. 2009].

Para o Ciclo 03, em que o crescimento se aproximou do potencial, é possível comparar nossos resultados aos obtidos por [Jones et al. 1999]. Os autores também avaliaram o desempenho do modelo sem realizar calibração em duas localidades e concluíram que, embora fosse possível a obtenção de bons resultados para a biomassa, isso não seria possível para a massa de frutos e frutos maduros. Nossos resultados apontaram que a calibração teria sido necessária e que os parâmetros originais implicariam baixa produção de biomassa, possivelmente por incompatibilidade na área foliar, mas a realização de assimilação, permitiu que valores mais próximos aos reais fossem obtidos.

É comum que a assimilação de dados seja considerada uma forma de adequar o crescimento potencial simulado por modelos ao crescimento real. Por exemplo, [Hu et al. 2019] realizaram a assimilação de dados no modelo SWAP-Wofost com o objetivo de contabilizar a interferência no IAF da cana-de-açúcar por remoção de folhas e tempestades naturais. No nosso caso, temos três cenários diferentes: estresse hídrico, estresse nutricional e fertilização e irrigação adequadas. Observamos que o método foi capaz de adaptar a simulação, que espera crescimento potencial, em valores mais realistas ao mesmo tempo em que, para o caso de alta flutuação da água no Ciclo 01, a estimativa foi gravemente deteriorada.

Conclusão

Nosso trabalho mostra resultados promissores para a assimilação de observações de crescimento por duas perspectivas distintas: ocorre a melhora das estimativas quando não há calibração do modelo e as estimativas obtidas são mais adequadas às condições reais de produção, ao invés da limitação existente na simulação à condição de produção potencial. Também mostramos como observações de diversas fontes, incluindo fotos capturadas por câmeras de baixo custo, foram capazes de caracterizar o crescimento das plantas. Embora nem sempre as relações obtidas tenham sido boas, elas foram suficientemente boas para que informações imperfeitas de observações contínuas pudessem ser usadas juntamente a um modelo imperfeito. Sendo assim, há espaço ainda para que fatores, como o desenvolvimento de modelos de observação com mais dados, sejam mais explorados, de modo que eles possam ser considerados mais generalizáveis e que as análises destrutivas necessárias para a calibração possam ser de fato dispensadas.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e pelo processo nº 2018/12050-6, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências bibliográficas

Bojacá, C. R., Gil, R. and Cooman, A. (mar 2009). Use of geostatistical and crop growth modelling to assess the variability of greenhouse tomato yield caused by spatial temperature variations. Computers and Electronics in Agriculture, v. 65, n. 2, p. 219–227.

Chen, W.-T., Yeh, Y.-H. F., Liu, T.-Y. and Lin, T.-T. (31 mar 2016). An Automated and Continuous Plant Weight Measurement System for Plant Factory. Frontiers in Plant Science, v. 7, n. MAR2016, p. 392.

Chen, Y., Zhang, Z. and Tao, F. (1 nov 2018). Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. European Journal of Agronomy, v. 101, p. 163–173.

De Graaf, R., De Gelder, A. and Blok, C. (2004). Advanced weighing equipment for water, crop growth and climate control management. Acta Horticulturae, v. 664, p. 163–167.

De Koning, A. N. M. and Bakker, J. C. (mar 1992). In situ plant weight measurement of tomato with an electronic force gauge. Acta Horticulturae, n. 304, p. 183–186.

Dorigo, W. A., Zurita-Milla, R., De Wit, A. J. W., et al. (may 2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation, v. 9, n. 2, p. 165–193.

Gong, L., Yu, M., Jiang, S., Cutsuridis, V. and Pearson, S. (1 jul 2021). Deep Learning Based Prediction on Greenhouse Crop Yield Combined TCN and RNN. Sensors, v. 21, n. 13, p. 4537.

Helmer, T., Ehret, D. L. and Bittman, S. (2005). CropAssist, an automated system for direct measurement of greenhouse tomato growth and water use. Computers and Electronics in Agriculture, v. 48, n. 3, p. 198–215.

Hemming, S., De Zwart, F. De, Elings, A., Petropoulou, A. and Righini, I. (11 nov 2020). Cherry Tomato Production in Intelligent Greenhouses—Sensors and AI for Control of Climate, Irrigation, Crop Yield, and Quality. Sensors, v. 20, n. 22, p. 6430.

Hemming, S., De Zwart, F., Elings, A., Righini, I. and Petropoulou, A. (16 apr 2019). Remote Control of Greenhouse Vegetable Production with Artificial Intelligence—Greenhouse Climate, Irrigation, and Crop Production. Sensors, v. 19, n. 8, p. 1807.

Hu, S., Shi, L., Huang, K., et al. (15 feb 2019). Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. Field Crops Research, v. 232, p. 49–61.

Huang, J., Gómez-Dans, J. L., Huang, H., et al. (15 oct 2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. Agricultural and Forest Meteorology, v. 276–277, p. 107609.

Huang, J., Sedano, F., Huang, Y., et al. (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. Agricultural and Forest Meteorology, v. 216, p. 188–202.

Jin, X., Kumar, L., Li, Z., et al. (jan 2018). A review of data assimilation of remote sensing and crop models. European Journal of Agronomy, v. 92, n. November 2017, p. 141–152.

Jones, J. W., Kenig, A. and Vallejos, C. E. (1999). Reduced state-variable tomato growth model. Transactions of the ASAE, v. 42, n. 1, p. 255–265.

Keating, B. A. and Thorburn, P. J. (22 oct 2018). Modelling crops and cropping systems—Evolving purpose, practice and prospects. European Journal of Agronomy, v. 100, n. April, p. 163–176.

Lee, J. W. and Son, J. E. (18 oct 2019). Nondestructive and Continuous Fresh Weight Measurements of Bell Peppers Grown in Soilless Culture Systems. Agronomy, v. 9, n. 10, p. 652.

Linker, R. and Ioslovich, I. (2017). Assimilation of canopy cover and biomass measurements in the crop model AquaCrop. Biosystems Engineering, v. 162, p. 57–66.

Lu, Y., Chibarabada, T. P., Ziliani, M. G., et al. (2021). Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. Agricultural Water Management, v. 252, n. April, p. 106884.

Marcelis, L. F. M., Van den Boogaard, R. and Meinen, E. (2000). Control of Crop Growth and Nutrient Supply by the Combined Use of Crop Models and Plant Sensors. IFAC Proceedings Volumes, v. 33, n. 19, p. 161–166.

Pellenq, J. and Boulet, G. (may 2004). A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. Agronomie, v. 24, n. 4, p. 197–204.

Ruíz-García, A., López-Cruz, I. L., Ramírez-Arias, A. and Rico-Garcia, E. (may 2014). Modeling uncertainty of greenhouse crop lettuce growth model using Kalman Filtering. Acta Horticulturae, v. 1037, n. 1037, p. 361–368.

Torres-Monsivais, J. C., López-Cruz, I. L., Ruíz-García, A., Ramírez-Arias, J. A. and Peña-Moreno, R. D. (2017). Data assimilation to improve states estimation of a dynamic greenhouse tomatoes crop growth model. Acta Horticulturae, n. 1170, p. 433–440.

Vazifedoust, M., Van Dam, J. C., Bastiaanssen, W. G. M. and Feddes, R. A. (2009). Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. International Journal of Remote Sensing, v. 30, n. 10, p. 2523–2545.

Verdouw, C., Tekinerdogan, B., Beulens, A. and Wolfert, S. (1 apr 2021). Digital twins in smart farming. Agricultural Systems, v. 189, n. January, p. 103046.

Yu, D., Zha, Y., Shi, L., et al. (2020). Improvement of sugarcane yield estimation by assimilating UAV-derived plant height observations. European Journal of Agronomy, v. 121, n. February, p. 126159.

Zhao, Y., Chen, S. and Shen, S. (dec 2013). Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation. Ecological Modelling, v. 270, p. 30–42.



Multi-objective methods for crop insurance premiums: framework proposal and a case study in sugarcane

Roberto F. Silva[1], Marcos R. Benso[2], Gabriela C. Gesualdo[2], Eduardo M. Mendiondo[2], Antônio M. Saraiva[3], Patrícia A. A. Marques[4], Alexandre C. B. Delbem[1]

[1]Institute for Mathematics and Computer Science - University of São Paulo
São Carlos, Brazil

[2]São Carlos School of Engineering - University of São Paulo
São Carlos, Brazil

[3]Polytechnic School - University of São Paulo
São Paulo, Brazil

[4]Luiz de Queiroz Colege of Agriculture - University of São Paulo
Piracicaba, Brazil

roberto.fray.silva@gmail.com, {marcosbenso,gabriela.gesualdo}@usp.br

emm@sc.usp.br,{saraiva,paamarques,acbd}@usp.br

Abstract. *The increase in extreme climate events due to climate change has resulted in crop losses, quality losses, environmental and social impacts in agricultural areas. Insurance against extreme events is a vital tool adaptation to deal with the impacts of those hazards. However, few works consider the optimization of different dimensions related to this tool. This work proposes a framework to use multi-objective optimization models to better design and evaluate crop insurance premiums and conducts a case study for sugarcane premiums at São Paulo state in 2010. The framework can be adopted for different crops, objectives, and models. The case study showed that around 20% of the policies evaluated were efficient solutions from the farmer's point of view.*

1. Introduction

Extreme climate events affect every region globally and are becoming more frequent and intense due to continuous human-induced climate change [IPCC 2021]. Particularly in South America, projections indicate an increase in the temperature of the hottest days at about two times the rate of global warming [IPCC 2021]. Consequently, events such as heatwaves, heavy precipitation, and agricultural and ecological droughts tend to become

increasingly common. Therefore, to reduce the future impacts of these extreme events or natural hazards, it is crucial to understand the spatial and temporal variability of the vulnerability of people and economic assets [Formetta and Feyen 2019].

There is additional complexity in extreme events vulnerability and uncertainty regarding developing economies due to their greater economic dependence on climate-sensitive primary activities, such as agriculture. According to [Kim et al. 2019] from 1983 to 2009, droughts caused 166 billion USD of losses in maize, soy, rice, and wheat production worldwide, affecting three-fourths of the global harvested area.

These losses can be minimized and rewarded through adaptation tools such as extreme event insurance. In this context, insurance prevents financial damage from turning into long-term economic damage [Seifert-Dähnn 2018]. Therefore, models that work on identifying and quantifying risk, estimating costs, and optimizing the premium are needed to enable risk transfer [Merz et al. 2010] and [Winsemius et al. 2013].

This research aims at proposing a theoretical framework for using multi-objective optimization (MOO) tools to evaluate crop insurance premiums considering multiple dimensions, such as the premium value, total insurable value (TIV), and coverage, among other essential aspects from the multi-dimensional model proposed [Navarro et al. 2021]. This framework encompasses the main stages of the data lifecycle and is model-agnostic. Then, a case study will be conducted for sugarcane crop insurance in 2010 at São Paulo state at the city level from the point of view of the sugarcane grower. The empirical attainment function (EAF) multi-objective model will be used for evaluating the efficient and non-efficient solutions for a bi-objective problem.

The two main research questions in this work are: (i) how can multi-objective optimization models be used for evaluating crop insurance premiums, considering multiple dimensions and objectives?; and (ii) what percentage of the crop insurance policies for sugarcane at São Paulo state cities could be considered efficient for the year of 2010, considering the maximization of the TIV per area and the minimization of the premium value per area? The main contributions of this work are: (i) proposing a theoretical framework for using MOO to evaluate crop insurance premiums that is model-agnostic; and (ii) evaluating the proposed framework on a case study with real data.

2. MOO problems on the agricultural domain

Crop insurance is an *ex-ante* risk management tool that can be used to mitigate the impacts of extreme climate events. In this context, the farmer pays a premium to an insurance company to be protected when crop yields are lower than a stipulated level [Bowers et al. 1997]. Nonetheless, the optimization of insurance contracts requires the development of methods that take into account multiple variables, as demonstrated by [Mohor and Mendiondo 2017]. Other works, such as [Fadhliani et al. 2019], incorporated participatory data to improve farmer's satisfaction and participation in insurance programs, improving coverage areas and farmer's adhesion in vulnerable rural areas.

Several important works in the literature evaluated the use of MOO on agricultural domains. Two examples are the ones by [Klein et al. 2013] and [Dunnett et al. 2018]. [Klein et al. 2013] proposed a system containing a MOO model and a self-organizing map to adapt agricultural land management considering the potential impacts of climate

change. Their main contributions were : (i) considering objectives related to farm productivity and environmental aspects; (ii) evaluating different climate scenarios and land uses; and (iii) evaluating the use of a generic crop model to estimate productivity.

[Dunnett et al. 2018] evaluated environmental and economic aspects related to agricultural production, but from a climate-smart strategy perspective. Those authors proposed and implemented a framework named Climate Smart Agricultural Prioritization (CSAP). Its significant contributions were: (i) considering different climate adaptation strategies for the different scenarios; (ii) considering socio-economic criteria; and (iii) considering both biophysical and agricultural emissions models on evaluating land use. However, that are also more specific MOO models for individual farming systems, such as the ones by [Groot et al. 2012] and [Jiang et al. 2020].

Lastly, this work evaluates the use of an important MOO model, the EAF. It was proposed by [Da Fonseca et al. 2001]. According to [López-Ibáñez et al. 2010], the EAF can be defined as the distribution of the different results of a stochastic local search algorithm applied to a multi-objective problem, considering maximization or minimization objectives. Different algorithms can be used for the local search. After several runs of the chosen algorithm with different random seeds, the model can identify several curves representing the relations between the dominated and non-dominated solutions. Additionally, an important use of the EAF model is to compare the behavior of different MOO models in a graphical format. For an in-depth description of the EAF model, we refer the reader to the works by [Da Fonseca et al. 2001] and [López-Ibáñez et al. 2010].

3. Methodology

The methodology used in this work was comprised of the following steps:

1. Framework proposal for the use of MOO for crop insurance premiums definition and evaluation, considering all the steps of the data lifecycle (as described by [Ball 2012]) and multiple dimensions (as proposed by [Navarro et al. 2021]);
2. Data gathering from sugarcane crop insurance premiums and coverage on the state of São Paulo for the year of 2010 from the official Government database, "Sistema de Subvenção Econômica ao Prêmio do Seguro Rural" ([SISSER 2021]);
3. Data processing, considering three main steps: (i) objective definition, generating the two objectives: TIV per area and premium per area; (ii) identifying and eliminating cities that presented missing data; and (iii) detecting outliers using the Boxplot method and eliminating them from the dataset;
4. Framework implementation on a bi-objective problem for sugarcane insurance in all cities at the São Paulo state at the year of 2010, using the EAF model through the EAF tools library [1]. The following objectives were considered: (i) maximizing the TIV per area, which is defined as the maximum value that the insurance company will pay for the farmer if the policy is used; and (ii) minimizing the premium value per area, which is the price that the farmer will pay for the insurance policy;
5. Evaluation of the optimal curve generated by the model and the different percentiles, considering: (i) the percentage of non-dominated points; (ii) the number of points that presented the worst strategies (according to the MOO model); and (iii) an overall analysis of the "efficiency" of the crop insurance premiums for sugarcane at São Paulo state in the year of 2010.

[1] https://mlopez-ibanez.github.io/eaf/

4. Proposed framework and its components

As described by [Navarro et al. 2021, Hajkowicz and Collins 2007], and [de Brito and Evers 2016], it is essential to consider MOO and multi-criteria decision-making to address the potential effects of extreme climate events in agricultural areas. Therefore, we propose a theoretical framework that considers multiple dimensions and objectives to improve crop insurance design and evaluation. It also considers the main steps of the data lifecycle [Ball 2012] and the standard workflow for machine learning models [LeCun et al. 2015]. It is composed of the following ten components:

1. **Data collection**, considering historical geographic data from official databases. The main objectives to be evaluated are related to the dimensions described by [Navarro et al. 2021] and to important criteria that are used on crop insurance premium design: coverage area, coverage production volume, TIV per area, the premium per area, government subsidies per area, probability of occurrence and potential impact of specific natural hazards, and willingness to pay for the insurance. Nevertheless, additional data sources can be considered;
2. **Objective definition**, generating the objectives that may be relevant for the specific use case. It is essential to observe the importance of the spatial and temporal dimensions for defining the objective functions;
3. **Data processing**, considering detection and handling outliers and missing data;
4. **Exploratory data analysis**, considering statistical analyses of the different criteria selected at the second component, to better understand their distributions and identify local patterns;
5. **Definition of the objective functions**, considering the multiple dimensions, and maximization and minimization objectives;
6. **MOO model selection**, considering their different aspects and characteristics, as explored in the works by [de Brito and Evers 2016, Hajkowicz and Collins 2007];
7. **Data standardization**, normalization, or transformation, if needed, to improve the model's overall results;
8. **MOO model implementation**, considering hyperparameters analysis and the objectives defined in step 3. It is also possible to conduct experiments with different combinations of the criteria and to compare different MOO models;
9. **Final model evaluation and analysis**, considering the following analyses: (i) identification of the optimal strategy; (ii) spatial analyses, considering dominant strategies for each region; and (iii) an in-depth analysis of the worst strategies identified. The main objective of this step is to extract the most important insights from the model's results. Additional analyses can be conducted if needed.
10. **Decision-making** based on the results of step 9. This may involve several actions, such as: redesigning the crop insurance policy, redesigning the crop insurance premium, among others.

5. Case study - Sugarcane crop insurance in 2010

This section contains the results of implementing the framework proposed in Section 4 on a case study of sugarcane insurance premiums at São Paulo state. First, we will present the results of the exploratory data analysis. Then, we will present and explore the results of the EAF model implementation.

Figure 1 illustrates the kernel density estimations (KDE) of the distributions for the premium per area (on the left, item ”a”) and the TIV per area (on the right, item ”b”). It is essential to observe that: (i) both distributions are considerably different, indicating that the two objectives present a degree of conflict, justifying the multi-objective analysis; (ii) the normal distribution cannot be considered as a proxy for any of those distributions; (iii) the premium per area has a minimum value of R$13.39, a maximum of R$61.59, and a peak of around R$20.00; and (iv) the TIV per area has a minimum value of R$1180.62, a maximum of R$2530.18, and a peak of around R$2000.00. Among the main reasons that may explain those observations are: (i) the characteristics of the different regions, in terms of technology, soil, climate, and productivity;

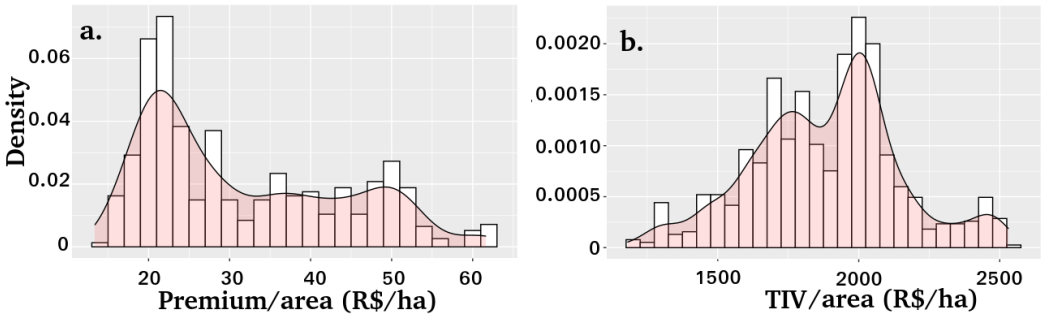


Figure 1. KDEs of the premium per area (on the left, item ”a”) and the TIV per area (on the right, item ”b”) objectives.

Figure 2 illustrates the results of the EAF model implementation, considering the bi-objective analysis described in Section 3. The EAF model is a stochastic optimization algorithm. Several simulations are run considering the data distributions and the two objective functions. The different colors in Figure 2 illustrate the areas that are related to the percentiles of the results obtained by those simulations. For example, the dark green area (”Best”) represents the area that can be considered optimal for the MOO. The points that are on the upper border of the dark green area represent the efficient solutions found by the simulations conducted with the EAF model for the two objectives analyzed. The remainder of the points in that area (which are not along the upper border) represent solutions that can be considered good in practical terms, but that are not efficient (due to the fact that they are dominated by the solutions in the upper border).

The percentiles on the chart (represented by the different colors) represent the results of the run of the EAF model. Therefore, there may be combinations of the objectives that are not feasible in the real world. The efficient solutions (closer to the upper border of the green area on Figure 2) are candidates for Pareto-optimal solutions, based on the simulations that were conducted using the EAF model. These are more efficient from the farmer’s point of view, as they will cost less and provide higher loss coverage.

It is vital to observe in Figure 2 that 162 points can be considered efficient or non-dominated. This indicates that around 20% of the total points in the dataset are non-dominated. Therefore, it is possible to conclude that around 20% of the total insurance policies in the dataset evaluated provide the best possible solution for the farmers. Additionally, around 30% of the policies in the dataset are encompassed in the dark green area (”Best”), meaning that they are also advantageous for the farmers.

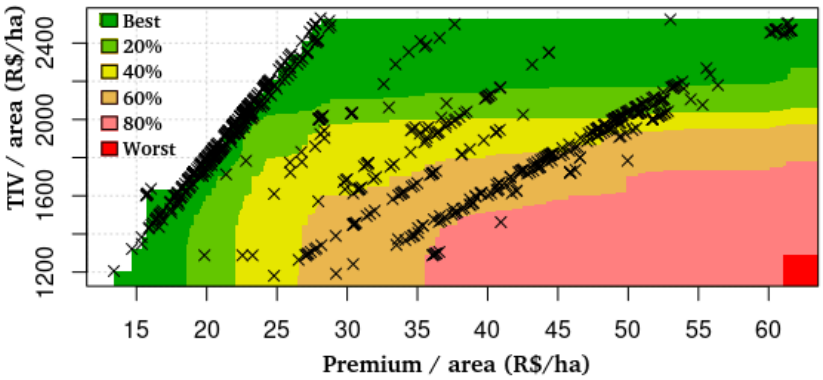


Figure 2. Final results of the EAF model implementation.

Lastly, no insurance policies are part of the dark red area ("Worst"), and very few insurance policies (7 data points in the evaluated dataset) are part of the 80% percentile (in light red). That means that very few policies are significantly inefficient for the farmers. Therefore, we can infer that most of the insurance premium prices for sugarcane in 2010 were aligned with the farmers' best interests. However, several insurance policies could be improved by reducing the premium per area or increasing the TIV per area (areas represented in medium and light green, orange, and light red).

6. Discussions

The results of the study case have shown that the use of multi-objective models could lead to interesting insights into the evaluation of crop insurance policies and premium rates. Therefore, the answer to the first research question is that, due to the several dimensions that may conflict with each other on designing and evaluating crop insurance policies, it is suitable to use an approach that considers partially or entirely conflicting objectives and decision criteria. The MOO models are specifically defined to deal with situations with conflicting objectives and decision criteria that must be maximized or minimized. As for the second research question, it was observed that 20% of the policies evaluated in the case study could be considered efficient from the farmer's point of view. It was also observed that an additional 30% of the policies were advantageous for the farmers.

Different supply chain stakeholders could use the results obtained from the case study for improving decision-making. Banks and insurance companies could focus their efforts on designing insurance policies with premium rates that increased the insurance coverage (in terms of area or production volume). Farmers could use the model's results to evaluate better and negotiate their insurance policies. Additionally, they could compare their policies with other policies in the same region. Lastly, it is essential to emphasize that an in-depth study considering different years, additional objectives (especially objectives related to the potential impacts of extreme climate events), and MOO models could improve the quality of the analysis. This could be used as an important tool for improving the decision-making processes related to public policies.

Several extensions of the case study conducted can be implemented, such as: considering additional areas, crops, objective functions, criteria, and MOO models. However, it is vital to observe that all extensions can be implemented following the framework pro-

posed in Section 4. Finally, the main limitations of this research were: (i) the lack of data available for implementing all the framework dimensions; (ii) the wide variety of options of MOO models; and (iii) the lack of applications of MOO for crop insurance premiums, which limited the comparisons that could be done with previous works in the literature. However, we believe that those limitations were correctly addressed in our work.

7. Conclusions and future works

A critical impact of climate change is the increase in frequency and impact of extreme climate events, such as droughts, floods, heatwaves, among others. These will increasingly impact agricultural areas and products, leading to crop losses, quality loss, and social and environmental impacts. One essential tool to reduce the impact of those natural hazards is extreme event insurance or crop insurance. Additionally, the use of MOO models can improve the quality of this tool by incorporating several dimensions into its design. Among those dimensions, it is important to consider: the probability of occurrence of the event, its potential impacts, the farmer's willingness to pay, among others.

This work proposes a theoretical framework for using MOO models to evaluate crop insurance premiums, considering multiple dimensions and objective functions. This is essential to address better the impacts of climate change on agroindustrial supply chains. The framework contains ten components and encompasses the whole data lifecycle. It also considers the main stages of the machine learning workflow, allowing the use of different machine learning models on the different components. A case study was conducted to evaluate sugarcane premiums at São Paulo state in 2010, concluding that around 20% of the policies were candidates for Pareto-optimal solutions, and that around 30% can be considered efficient policies from the farmer's point of view.

We recommend as future works: (i) implementing and analyzing other crops, objectives, and MOO models; (ii) and conducting a temporal analysis of the efficiency of the insurance premiums for a specific crop. To study the robustness of the EAF approach with TIV-Premium output, we recommend future works on: (i) attainment goal optimization with special attention to the slackness nature of some decision-making variables; (ii) setting the EAF simulations with TIV-Premium constraints. and (iii) scaling effects (upscaling/dowscaling) of spatial and temporal dimensions which strongly define both the objective functions and outputs.

Acknowledgements: This work was carried out at the Center for Artificial Intelligence (C4AI-USP), with support by: the São Paulo Research Foundation (FAPESP grant #2019/07665-4), IBM Corporation, FAPESP 2014/50848-9 INCT-Climate Change Phase II, and the National Council for Scientific and Technological Development (CNPq). We also acknowledge the support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, finance code 001).

References

Ball, A. (2012). *Review of data management lifecycle models*. University of Bath.

Bowers, N., Gerber, H., Hickman, J., Jones, D., and Nesbitt, C. (1997). *Actuarial Mathematics*. Society of Actuaries.

Da Fonseca, V. G., Fonseca, C. M., and Hall, A. O. (2001). Inferential performance assessment of stochastic optimisers and the attainment function. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 213–225. Springer.

de Brito, M. M. and Evers, M. (2016). Multi-criteria decision-making for flood risk management: a survey of the current state of the art. *Natural Hazards and Earth System Sciences*, 16(4):1019–1033.

Dunnett, A., Shirsath, P., Aggarwal, P., Thornton, P., Joshi, P. K., Pal, B. D., Khatri-Chhetri, A., and Ghosh, J. (2018). Multi-objective land use allocation modelling for prioritizing climate-smart agricultural interventions. *Ecological modelling*, 381:23–35.

Fadhliani, Z., Luckstead, J., and Wailes, E. J. (2019). The impacts of multiperil crop insurance on indonesian rice farmers and production. *Agricultural Economics*, 50(1):15–26.

Formetta, G. and Feyen, L. (2019). Empirical evidence of declining global vulnerability to climate-related hazards. *Global Environmental Change*, 57:101920.

Groot, J. C., Oomen, G. J., and Rossing, W. A. (2012). Multi-objective optimization and design of farming systems. *Agricultural Systems*, 110:63–77.

Hajkowicz, S. and Collins, K. (2007). A review of multiple criteria analysis for water resource planning and management. *Water resources management*, 21(9):1553–1566.

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, Cambridge University Press.

Jiang, S., Zhang, H., Cong, W., Liang, Z., Ren, Q., Wang, C., Zhang, F., and Jiao, X. (2020). Multi-objective optimization of smallholder apple production: Lessons from the bohai bay region. *Sustainability*, 12(16):6496.

Kim, W., Iizumi, T., and Nishimori, M. (2019). Global patterns of crop production losses associated with droughts from 1983 to 2009. *Journal of Applied Meteorology and Climatology*, 58(6):1233 – 1244.

Klein, T., Holzkämper, A., Calanca, P., Seppelt, R., and Fuhrer, J. (2013). Adapting agricultural land management to climate change: a regional multi-objective optimization approach. *Landscape ecology*, 28(10):2029–2047.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *N ature*, 521(7553):436–444.

López-Ibáñez, M., Paquete, L., and Stützle, T. (2010). Exploratory analysis of stochastic local search algorithms in biobjective optimization. In Bartz-Beielstein, T., Chiarandini, M., Paquete, L., and Preuss, M., editors, *Experimental Methods for the Analysis of Optimization Algorithms*, pages 209–222. Springer, Berlin, Germany.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A. (2010). Review article "Assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, 10(8):1697–1724.

Mohor, G. S. and Mendiondo, E. M. (2017). Economic indicators of hydrologic drought insurance under water demand and climate change scenarios in a brazilian context. *Ecological Economics*, 140:66–78.

Navarro, F. A. R., Gesualdo, G. C., Ferreira, R. G., Rápalo, L. M. C., Benso, M. R., de Macedo, M. B., and Mendiondo, E. M. (2021). A novel multistage risk management applied to water-related disaster using diversity of measures: A theoretical approach. *Ecohydrology & Hydrobiology*.

Seifert-Dähnn, I. (2018). Insurance engagement in flood risk reduction – examples from household and business insurance in developed countries. *Natural Hazards and Earth System Sciences*, 18(9):2409–2429.

SISSER (2021). Apólices de seguros agrícolas - 2006 a 2016. *Available at: https://dados.gov.br/dataset/sisser3. Accessed on: 01.09.2021*.

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., and Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, 17(5):1871–1892.



Ontologias para modelos de negócio na pecuária de corte: uma revisão da literatura

Matheus Collares Machado[1], **Ana Paula Lüdtke Ferreira**[1]

[1]Universidade Federal do Pampa - UNIPAMPA

{matheusmachado.aluno,anaferreira}@unipampa.edu.br

Abstract. *Ontologies are formal representations of concepts and their relationships, with semantics expressed by terms in Description Logics. Business models form the framework for testing the feasibility and financial results of enterprises. The description of a business model as an ontology allows the inference of knowledge about the business and the generation and combination of models to build software applications. This paper presents a literature review on business model and beef cattle ontologies as a basis for identifying research opportunities in the area.*

Resumo. *Ontologias são representações formais de conceitos e de seus relacionamentos, com semântica expressa por Lógicas de Descrições. Modelos de negócio formam o arcabouço para teste de viabilidade e resultados econômicos de empreendimentos. A descrição de um modelo de negócio como uma ontologia permite tanto a inferência de conhecimento sobre o negócio quanto gerar e combinar modelos na forma de aplicações de software. Este trabalho apresenta uma revisão da literatura sobre ontologias de modelos de negócio e sobre pecuária de corte, como base para a identificação de oportunidades de pesquisa na área.*

1. Introdução

A evolução dos sistemas produtivos agropecuários acompanhou as chamadas *revoluções industriais*: de maquinário movido a tração animal ou humana, passou-se ao trabalho mecanizado, informatização dos processos e, atualmente, uma miríade de soluções de cunho tecnológico [Massruhá e Leite 2017]. As implementações feitas fazem uso de diferentes tipos de interfaces, formas de organização da informação, linguagens de programação, sistemas gerenciadores de bancos de dados e todo um conjunto de artefatos tecnológicos independentes e sem comunicação entre si, causando uma sobrecarga cognitiva que pode comprometer os processos de controle e tomada de decisão.

A evolução da Agricultura Digital exige que a tecnologia se torne transparente para os usuários, integrada em dispositivos físicos com interfaces que façam sentido para quem delas fará uso. Devido aos impactos econômicos, ambientais e éticos relacionados à atividade agropecuária, é preciso garantir que o funcionamento dos artefatos de hardware e software é seguro e está de acordo com as suas especificações. Para tanto, a informação e os procedimentos devem estar mapeados e organizados, de forma a permitir que novas tecnologias e sistemas possam ser integrados aos que já existem com um mínimo de esforço humano e com garantias de correção e qualidade. Modelos formais de especificação e integração de sistemas podem servir de base para que este objetivo seja atingido.

Um modelo é dito formal quando possui sintaxe e semântica definidas por meio de algum domínio matemático. Uma das vantagens de linguagens de especificação formal é a capacidade de tradução das especificações para outras linguagens. A vantagem da tradução orientada por semântica – e não somente por sintaxe – é a possibilidade de gerar e integrar aplicações de maneira correta. Ontologias [Staab e Studer 2009] podem ser um caminho nessa direção. Uma ontologia é "uma especificação explícita de uma conceptualização" [Gruber 1993]. [Borst 1997] complementa a definição adicionando o aspecto formal das ontologias e sua natureza contratual, no sentido da conceptualização ser bem definida e compartilhada. Ou seja, ontologias expressam um modelo conceitual, formal, explícito e acordado de algum aspecto do conhecimento humano.

Ontologias podem ser desenvolvidas em diferentes níveis de abstração. Inicialmente usadas para construção de taxonomias e tesauros (dicionários de sinônimos), sua natureza formal evoluiu para o desenvolvimento de modelos de inferência, expressos em lógica de descrições [Baader et al. 2007]. A lógica de descrições é um fragmento decidível da lógica de primeira ordem, com vantagens computacionais imediatas, como o uso de algoritmos para verificação de propriedades. As bases de conhecimento construídas com lógica de descrições são divididas em duas partes: (i) a representação da terminologia dos conceitos que se quer descrever, ou o vocabulário do domínio de interesse e (ii) as propriedades e relações existentes entre os elementos anteriores. Assim, não somente a sintaxe de uma representação pode ser definida, mas também sua semântica. A descrição semântica, usando o motor de inferência da lógica de descrições, permite que conclusões corretas sejam produzidas sobre os modelos. Ou seja, pode-se unificar conceitos e suas relações e inferir novos conhecimentos sobre os fatos inicialmente estabelecidos.

O advento da web semântica [Bizer et al. 2009] trouxe a especialização das linguagens para descrição de ontologias [Horrocks et al. 2003], com um consórcio para o desenvolvimento e padronização das tecnologias relacionadas à web semântica (https://www.w3.org/). Essas tecnologias incluem padrões para a construção de páginas web, criação de repositórios de dados interligados, construção de vocabulários e criação de regras para manipulação desses dados. As linguagens usadas para representação de conhecimento na web semântica interligam-se com as linguagens usadas para descrição de ontologias [Gasevic et al. 2006], originadas em XML (*eXtensible Markup Language*), que usa símbolos de marcação para a construção de gramáticas de vários tipos, XML namespaces e identificadores de recursos web. As linguagens padrão correntes são RDF Schema (*Resource Description Framework*) e a família de linguagens OWL (*Web Onto-*

logy Language), com diferentes poderes de expressão e computabilidade.

Ontologias têm sido usadas em diferentes domínios de aplicações, da Educação à Medicina. A motivação para o seu uso no desenvolvimento de sistemas agropecuários engloba: (i) a diversidade de termos regionais para expressar os mesmos conceitos requer uma adequação de linguagem dos sistemas para refletir a linguagem dos usuários locais e (ii) a possibilidade de descrever produtos relacionados aos conceitos dos sistemas produtivos agropecuários, facilitando processos de integração de dados e sistemas. O uso de ontologias permite que esses processos possam ser conduzidos (semi)automaticamente.

O objetivo deste trabalho é investigar o uso de ontologias como ferramenta de modelagem conceitual e buscar seus usos em modelos de negócio da pecuária de corte por meio de uma revisão sistemática da literatura. O foco em modelo de negócios é derivado de uma recente pesquisa, conduzida com produtores rurais por um consórcio entre EMBRAPA, SEBRAE e INPE, que conclui que diversos produtores gostariam de ter condições de melhor planejar e gerir seus negócios, avaliar seus custos de produção e de quantificar o seu lucro [Bolfe et al. 2020]. Essa dificuldade pode ser contornada pela definição explícita de um modelo de negócio no qual todos os aspectos do sistema produtivo, incluindo custos e receitas possam ser formalmente descritos e computados, a partir de sistemas gerados com base nas descrições.

O restante deste trabalho está assim organizado: a Seção 2 descreve o protocolo de revisão aplicado; a Seção 3 apresenta e discute os resultados obtidos; a Seção 4 conclui com a análise dos resultados e direções para futuras pesquisas.

2. Material e métodos

A condução do processo de revisão foi adaptada de [Kitchenham 2004] e [Dermeval e Bittencourt 2020]. As questões de pesquisa da revisão deste trabalho são:

- Q1 Existem ontologias já construídas para pecuária de corte?
- Q2 Existem ontologias já construídas para sistemas de produção agropecuária?
- Q3 Existem ontologias já construídas para modelos de negócio?
- Q4 Para que fins são utilizadas as ontologias construídas, referentes às questões de pesquisa anteriores?

A definição das *strings* de busca foi feita com base nas questões de pesquisa e em artigos relacionados, previamente selecionados. As duas *strings* de busca construídas, S1 e S2, são as seguintes:

- S1 (ontology OR ontologies OR ontologia OR ontologias OR ontological) AND (agriculture OR agricultural OR agronomic sciences OR precision agriculture OR agricultura OR pecuária OR pecuária de precisão OR agropecuária)
- S2 (ontology OR ontologies OR ontologia OR ontologias OR ontological) AND (business OR business model OR business models OR modelo de negócio OR modelos de negócio)

As *strings* foram usadas nos repositórios digitais ACM Digital Library (https://dl.acm.org/), IEEE Xplorer (https://ieeexplore.ieee.org/), Wiley Online Library (https://onlinelibrary.wiley.com/) e Science Direct (https://www.sciencedirect.com/). As buscas foram feitas somente nos campos de título e resumo, de trabalhos publicados entre

Tabela 1. Resultados da revisão sistemática da literatura

Fonte	Encontrados		Excluídos		Incluídos	
	S1	**S2**	**S1**	**S2**	**S1**	**S2**
ACM Digital Library	1	8	1	5	0	3
IEEE Xplorer	39	21	33	20	6	1
Wiley Online Library	1	2	1	2	0	0
Science Direct	63	164	61	162	2	2

os anos de 2015 e 2021. Os critérios de inclusão para o trabalhos obtidos exigiram que o artigo versasse sobre ontologias projetadas para pecuária ou para modelos de negócio e que estivessem escritos em Português ou Inglês. Foram excluídos os artigos que se tratavam de versões anteriores do mesmo trabalho, duplicatas ou que não fossem capazes de responder a pelo menos uma das questões de pesquisa. As referências dos artigos incluídos foram verificadas e, se em conformidade com os critérios de inclusão, foram incluídas na revisão. Essa ação garante que trabalhos mais antigos possam ser incluídos com menor esforço. Os artigos usados para a construção das *strings* de busca também foram incluídos. Os resultados da busca e análise dos artigos produzidos na revisão sistemática estão apresentados na Tabela 1.

3. Resultados

3.1. Ontologias para modelos de negócio

[Liang et al. 2019] introduz a ontologia de domínio GDBPO para a construção de modelos dinâmicos de processos de negócios orientados a objetivos. A ontologia proposta fornece quatro perspectivas, na forma de subontologias: processos, objetivos, regras de negócio e tomadas de decisão. Ela é capaz de auxiliar na decomposição de metas de alto nível em refinadas metas de nível operacional que são alinhadas com regras de negócio e atividades disponíveis para dinamicamente construir o processo de negócio. As subontologias especificam conceitos sobre os processos da empresa, variáveis para diferentes níveis de objetivos e metas da organização, restrições e requisitos a seguir durante a execução dos processos e especificações sobre como uma decisão é tomada e quais os fatores que a afetam, incluindo análise de riscos, metas e operações alternativas para cumpri-las. A ontologia construída não foi aplicada em nenhum domínio específico. Também não houve nenhum processo de inferência realizado, ainda que o trabalho aponte nessa direção.

[Zeleti e Ojo 2017] apresenta a ontologia de domínio OGDBMO (*open government data business model ontology*), objetivando criar um arcabouço para novas ferramentas de gestão de dados abertos. Os autores identificaram seis conceitos fundamentais usados para criar o modelo *6-Values Business Model Framework*, usado para guiar o desenvolvimento dos requisitos funcionais para a ontologia proposta. A concepção de requisito funcional, no trabalho, foram identificadas com questões de competência [Ren et al. 2014] como método de elicitação de requisitos. Estas questões consistem em perguntas que a ontologia deve ser capaz de responder quando completa e caso a ontologia seja capaz de responder a questão, ela é considerada atendida.

[Pease et al. 2020] desenvolve uma ontologia de domínio projetada para um sis-

tema automatizado de apoio à tomada de decisão para produtos *end-of-life*: produtos que estão no final de sua vida útil ou serão descontinuados. Decisões sobre continuidade de produtos são tomadas por pessoas que podem não ter total ciência das informações necessárias para aquela tomada de decisão, não podendo ser efetivadas em tempo hábil. O autor propõe um sistema capaz apoiar as decisões com foco na diminuição do risco da decisão tomada não resultar em lucro. O sistema desenvolvido é construído sobre uma ontologia cujos conceitos e relações explicitam lucratividade, previsão de preços e custos de produção.

[Yingbo e Xia 2019] propôs um método de modelagem de negócios e raciocínio baseada em uma ontologia de descrição de processos, com o intuito de resolver o problema do grande volume e alta complexidade dos dados no processamento de negócios bancários. Os autores decompõem e refinam a função principal do processo de negócio, completando a definição com os relacionamentos, resultando em um modelo da ontologia de processos voltados ao negócio bancário. Os elementos das operações bancárias são incorporados ao modelo e um conjunto customizado de regras de inferência é utilizado para verificar a efetividade e completude do modelo criado. Por fim o autor explica que a falta de dados experimentais abre caminho para melhora no trabalho e que o objetivo futuro é estender a ontologia para outros domínios de negócio.

[Zou et al. 2019] criou uma ontologia para descrever conceitos de *product-service-systems*, representados sob três pontos de vista: criador de estratégias de negócio, fabricantes e consumidores. Utilizando a ontologia desenvolvida o autor foi capaz de verificar a consistência entre modelos de negócio e modelos de manufatura de maneira semiautomática, provê raciocínio e evita potenciais inconsistências, além de auxiliar na seleção de sócios, mercados e tecnologias. Por fim o autor explica que, apesar do teste das funcionalidades, a aplicabilidade do modelo exige verificação em escalas maiores.

[Afify et al. 2017] elaborou uma ontologia de domínio de negócios para serviços em nuvem, englobando conhecimentos sobre modelos de computação em nuvem (entrega, *deployment* e preço), componentes, provedores e detalhes do serviço. A abordagem é orientada aos tipos de serviços da computação em nuvem: infraestrutura como um serviço (IaaS), plataforma como um serviço (PaaS) e software como um serviço (SaaS). A ontologia foi desenvolvida na linguagem OWL 2.0 por sua expressividade e semântica que permite modelagem explicita de seus conceitos e descrições do domínio além de possibilitar raciocínio automatizado.

3.2. Ontologias para a agropecuária

[Boshkoska et al. 2019] apresenta um sistema de apoio à decisão para avaliação de limites do compartilhamento de conhecimento na área de *agri-food value chains*. O desenvolvimento do sistema se dá em duas fases: identificação dos limites do conhecimento utilizando aprendizado de máquina e ontologias e transformação da ontologia obtida em um sistema de apoio à decisão para a avaliação dos limites desse conhecimento. A efetividade do sistema desenvolvido foi avaliada aplicando-o a três cadeias de valor distintas, sendo que nas três aplicações foi possível encontrar pontos que necessitavam de melhoria.

[Walisadeera et al. 2015] apresenta uma ontologia com conceitos relacionados a condições ambientais, colheitas e uso de fertilizantes, com o objetivo de inferir conhecimentos a partir das informações registradas na ontologia. As informações representadas

foram usadas para auxiliar os agricultores na seleção dos vegetais que seriam plantados. O sistema foi testado com um grupo de agricultores do Sri Lanka e a ontologia desenvolvida foi disponibilizada na nuvem em conjunto com uma interface para requisição de informações. O motor de inferência usado foi um *plug-in* chamado *FaCT ++ reasoner* e suas requisições foram codificadas em SPARQL e em lógica de descrições.

[Chukkapalli et al. 2020a] desenvolve um ecossistema de fazendas cooperativo e conectado que gerencia sensores e sua comunicação entre diferentes entidades do ecossistema através de um *co-op hub* na nuvem, desenvolvido a partir de duas ontologias: a ontologia membro e a ontologia co-op. A ontologia membro descreve os conceitos de uma fazenda individual membro da rede e a ontologia co-op descreve interações e conhecimento de domínio especifico sobre o ecossistema cooperativo de fazendas. Ambas ontologias são utilizadas em conjunto para capturar os dados do ecossistema e gerenciar suas interações. Em seu outro trabalho [Chukkapalli et al. 2020b] foi construído um sistema de segurança para controle de acesso baseado em atributos para a ontologia proposta anteriormente. O sistema é capaz de controlar o acesso dos trabalhadores ao maquinário (e.g. tratores), controlar o acesso de trabalhadores aos dados dos sensores e manter as permissões de uso dos sistemas automatizados (e.g. irrigação) por meio de regras escritas em SWRL em conjunto com conceitos e atributos da ontologia.

[Bi et al. 2016] apresenta um método de integração de dados da agricultura baseado em ontologia, motivado pelas inconsistências nos diferentes métodos de representação do conhecimento na agricultura. A ontologia foi construída a partir da análise da estrutura comum dos dados de gerência da produção agrícola de instituições, com construção de um modelo lógico de estruturas e múltiplas relações entre dados da agricultura são estabelecidas. As múltiplas relações entre os dados são convertidos em um modelo orientado a objetos e aplicados em um sistema de análise econômica em Jilin, província rural da China.

[Pani e Mishra 2016] desenvolve uma ontologia para formalização de conceitos da agricultura. A ontologia é descrita na linguagem OWL DL para uso em um sistema web. A ontologia é construída com base em um diagrama de classes criado anteriormente por meio da análise das variadas fases necessárias em um sistema para agricultura e as informações que o agricultor necessita em cada fase. Por fim são definidos os axiomas da ontologia que com base nestes axiomas e na estrutura hierárquica da ontologia poderão ser feitas consultas e inferências.

[Abrahao e Hirakawa 2017] criaram um modelo para representar conhecimento técnico de operações agrícolas de campo. As operações agrícolas foram analisadas para determinar quem são os agentes, as funções, recursos de entrada, controle de fluxo, conceitos das tarefas, atributos e relacionamentos. Como resultado foi obtida uma representação formal do domínio em conjunto com um modelo conceitual construído usando diagrama de classe UML e diagrama de atividades. A consistência do modelo foi verificada em estudo de caso sobre as operações da colheita de cana de açúcar. O trabalho termina com as perspectivas para futuras pesquisas envolvendo metodologias para construir implementações com linguagens como OWL e a exploração de outras maneiras de representar o controle de fluxo das operações como BPMN.

[Da Silva et al. 2019] descrevem um modelo ontológico para classificação de

doenças que acometem rebanhos de bois e búfalos. A ontologia proposta foi dividida em seis classes principais: anatomia, Bioma Amazônia, doenças, raças, sistema de criação e tipo de produção. O modelo foi capaz de relacionar doenças presentes na pecuária bovina e bubalina, tipos de produção, sistemas de criação e os ecossistemas do Bioma Amazônia. A linguagem utilizada foi OWL em conjunto com RDF. O trabalho apresenta consultas demonstrativas a base de dados utilizando a linguagem SPARQL.

3.3. Discussão

Apesar do protocolo estabelecer a busca de trabalhos nas bases com vultoso acervo de artigos tanto sobre Computação como sobre Ciências Agrárias, acrescidos aos achados relevantes dos referenciais bibliográficos dos trabalho inicialmente selecionados, foram encontrados relativamente poucos trabalhos de formalização de modelos de negócio/sistemas produtivos agropecuários com ontologias. Os trabalhos encontrados, por outro lado, mostram possibilidades promissoras de desenvolvimento nessa direção.

Dentre as ontologias voltadas para a modelagem de negócios, foram observadas aplicações variando entre níveis estratégicos e operacionais. Ontologias empregadas em níveis estratégicos foram usadas para desenvolver sistemas de apoio a decisão (e.g. [Pease et al. 2020]) e ontologias empregadas a níveis operacionais se preocupam com a definição dos processos de negócio da empresa (e.g. [Zou et al. 2019, Afify et al. 2017, Yingbo e Xia 2019]). Seja com modelos descrevendo parcialmente ou de forma mais completa (como [Liang et al. 2019]) a ideia de modelo de negócio, todos os trabalhos apresentam relações entre diferentes níveis. Essa é uma das características de uma ontologia: a descrição dos conceitos e de seus relacionamentos permite particionar e hierarquizar ideias em diferentes níveis, o que é importante para o controle da complexidade. Por outro lado, relacionamentos explícitos mantém os elementos conectados com precisão sintática (em todos os casos) e semântica (em alguns).

A revisão sistemática encontrou apenas um artigo projetado especificamente para pecuária, com foco em saúde animal ([Da Silva et al. 2019]). Outros trabalhos incluem ontologias aplicadas à gestão de fazendas inteligentes que também podem ser usadas para gerir sensores empregados na pecuária. Entre os trabalhos analisados, um aspecto geográfico das ontologias projetadas para a agropecuária foi observado onde os conceitos e informações relevantes para o produtor variam entre localizações, biomas, espécies e tipos de produção (e.g. [Walisadeera et al. 2015],[Da Silva et al. 2019]). Nota-se que, com exceção de [Pani e Mishra 2016], os demais trabalhos são voltados a descrições de processos específicos no setor, ainda que também com ligações aos seus conceitos de maior abstração, com as mesmas vantagens já descritas o parágrafo anterior. Nenhum trabalho no domínio de sistemas de produção de pecuária de corte foi encontrado durante a revisão.

As linguagens e ferramentas usadas nos trabalhos estudados é apresentada na Tabela 2. Pode-se notar a predominância de OWL como linguagem para descrição das ontologias. Essa predominância faz sentido, dado que a linguagem é suportada pelo consórcio W3C, além de que seus três níveis de expressividade/computabilidade bem definidos permitem que o desenvolvedor possa adequar a ferramenta certa às suas necessidades. O software Protégé [Musen 2015] segue sendo uma escolha frequente, dada sua robustez, tempo de desenvolvimento e tamanho da comunidade de usuários.

A Tabela 3 apresenta as ferramentas de inferência empregados nos trabalhos em

Tabela 2. Uso de ferramentas nos trabalhos analisados

Ferramenta	**Linguagem de descrição da ontologia**	**Trabalhos**
Protégé	OWL	[Afify et al. 2017], [Yingbo e Xia 2019], [Zou et al. 2019], [Pani e Mishra 2016], [Walisadeera et al. 2015]
Não especificado	OWL	[Chukkapalli et al. 2020b], [Chukkapalli et al. 2020a], [Pease et al. 2020]
Protégé	RDF e OWL	[Zeleti e Ojo 2017], [Da Silva et al. 2019]
OntoGen	Não especificado	[Boshkoska et al. 2019]
Não especificado	UML	[Abrahao e Hirakawa 2017]
Não especificado	Não especificado	[Liang et al. 2019], [Bi et al. 2016]

que havia algum tipo de raciocínio ou verificação de consistência automatizado, contabilizando aproximadamente 40% dos trabalhos analisados. Todos os trabalhos listados na Tabela 3 realizaram verificação de consistência em suas ontologias com exceção de [Da Silva et al. 2019]. Note-se que em [Pease et al. 2020] duas ferramentas de inferência foram usadas para verificação de consistência e, portanto, aparece em duas linhas da tabela. As inferências realizadas nos trabalhos analisados incluem o processamento de operações bancárias ([Yingbo e Xia 2019]), a extração de informações sobre saúde animal ([Da Silva et al. 2019]) e sobre plantações ([Pani e Mishra 2016], [Walisadeera et al. 2015]) e, por fim, estimativas de risco e lucro em sistema de apoio a decisão ([Pease et al. 2020]).

Tabela 3. Ferramentas de inferência utilizadss

Motor de inferência	**Trabalhos**
Jena	[Yingbo e Xia 2019], [Da Silva et al. 2019]
FaCT ++	[Walisadeera et al. 2015]
HermiT	[Pease et al. 2020]
Pellet	[Pani e Mishra 2016], [Pease et al. 2020], [Zeleti e Ojo 2017]

4. Conclusão

Ontologias são formalizações de conceitos e seus relacionamentos, com possibilidade de especificação da semântica dessas descrições e inferência sobre o conhecimento existente, por meio de lógicas de descrições e seus sistemas de prova. Ontologias têm sido usadas em várias áreas do conhecimento. Neste trabalho, apresentou-se uma revisão da literatura sobre o uso de ontologias para especificação de modelos de negócio e para a área da agropecuária. A intenção do trabalho é avaliar possíveis caminhos de pesquisa, especificamente na descrição de modelos de sistemas produtivos da pecuária de corte, com foco no controle e gerenciamento da produção.

A literatura encontrada mostrou que existem iniciativas em ambas as áreas, mas que ainda não atingiram seu potencial. Todos os trabalhos apresentam conceitos e seus relacionamentos, mas poucos implementam uma semântica, análises de consistência e inferência ou geração automatizada de aplicações. Essas últimas definições são imprescindíveis para que se atinja todo o potencial de uso de ontologias em sistemas produtivos agropecuários.

Como trabalho futuro, vislumbra-se a construção de uma ontologia voltada para modelos de negócio na pecuária de corte. Ontologias permitem a construção de tesauros, então os regionalismos existentes podem ser incorporados naturalmente ao sistema. O foco do trabalho será no gerenciamento e controle do sistema produtivo, com previsões relacionadas ao ganho financeiro do processo e melhor controle do processo como um todo. Inicialmente, as definições serão construídas com apoio de produtores de gado de corte na região do Pampa no Rio Grande do Sul com a parceria da EMBRAPA. A ontologia deverá abarcar detalhes que permitam a geração (semi)automatizada de aplicações, a partir das necessidades do negócio.

Referências

Abrahao, E. Hirakawa, A. R. (2017). Task ontology modeling for technical knowledge representation in agriculture field operations domain. *Proceedings - 2017 2nd International Conference on Information Systems Engineering, ICISE 2017*, 2017-Janua(1):12–16.

Afify, Y. M., Badr, N. L., Moawad, I. F., Tolba, M. F. (2017). A comprehensive business domain ontology for cloud services. *2017 IEEE 8th International Conference on Intelligent Computing and Information Systems, ICICIS 2017*, 2018-Janua(Icicis):134–143.

Baader, F., Calvanese, D., McGuinness, D. L., McGuinness, D. L., Nardi, D., Patel-Schneider, P. F. (2007). *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, Cambridge, UK., 2 edition.

Bi, L., Di, X. Q., Zhang, Y. (2016). Agricultural data modeling method based on semantics. *Proceedings of 2016 IEEE International Conference on Big Data Analysis, ICBDA 2016*.

Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5:1–22.

Bolfe, E. L., Junior, A. L., de Castro Victoria, D., Grego, C. R. (2020). Agricultura digital no Brasil - tendências, desafios e oportunidades (resultado de pesquisa online). Technical report, Campinas.

Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, University of Twente, Enschede, The Netherlands.

Boshkoska, B. M., Liu, S., Zhao, G., Fernandez, A., Gamboa, S., del Pino, M., Zarate, P., Hernandez, J., Chen, H. (2019). A decision support system for evaluation of the knowledge sharing crossing boundaries in agri-food value chains. *Computers in Industry*, 110:64–80.

Chukkapalli, S. S. L., Mittal, S., Gupta, M., Abdelsalam, M., Joshi, A., Sandhu5, R., Joshi, K. (2020a). Ontologies and artificial intelligence systems for the cooperative smart farming ecosystem. *IEEE Access*, 8:164045–164064.

Chukkapalli, S. S. L., Piplai, A., Mittal, S., Gupta, M., Joshi, A. (2020b). A smart-farming ontology for attribute based access control. In *IEEE 6th Intl Conference on Big Data Security on Cloud*, pages 29–34.

Da Silva, B. A., Da Ponte, M. J. M., Filho, J. R. B., Lima, C. P. (2019). Ontological model for classification of diseases in bovines and buffaloes. *Iberian Conference on Information Systems and Technologies, CISTI*, 2019-June(June):19–22.

Dermeval, D. Bittencourt, J. A. P. M. C. I. I. (2020). *Mapeamento Sistemático e Revisão Sistemática da Literatura em Informática na Educação*, volume 2, chapter 3, pages 1–26. CEIE/SBC.

Gasevic, D., Djuric, D., Devedzic, V. (2006). *Model Driven Architecture and Ontology Development*. Springer, Berlin Heidelberg.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.

Horrocks, I., Patel-Schneider, P. F., Van Harmelen, F. (2003). From SHIQ and RDF to OWL: The making of a Web Ontology Language. *Web Semantics*, 1(1):7–26.

Kitchenham, B. (2004). Procedures for performing systematic reviews. Joint Technical Report TR/SE-0401, Keele University, Keele.

Liang, Y., Wen, Z., Liu, L., Li, G., Guo, B. (2019). Towards a goal-driven dynamic business process ontology. In *ACM International Conference Proceeding Series*, pages 295–299.

Massruhá, S. M. F. S. Leite, M. A. d. A. (2017). *Agro 4.0 – rumo à Agricultura Digital*. Centro Paula Souza, São Paulo, 2 edition.

Musen, M. A. (2015). *The Protégé project: A look back and a look forward*, volume 1. Association of Computing Machinery Specific Interest Group in Artificial Intelligence.

Pani, S. Mishra, J. (2016). Building semantics of E-agriculture in India: Semantics in e-agriculture. *Proceedings - 2015 International Conference on Man and Machine Interfacing, MAMI 2015*, pages 15–18.

Pease, S. G., Sharpe, R., van Lopik, K., Goodall, P., Young, B., Conway, P., West, A. (2020). An interoperable semantic service toolset with domain ontology for automated decision support in the end-of-life domain. *Future Generation Computer Systems*, 112:848–858.

Ren, Y., Parvizi, A., Mellish, C., Pan, J. Z., Van Deemter, K., Stevens, R. (2014). Towards competency question-driven ontology authoring. *Lecture Notes in Computer Science*, 8465:752–767.

Staab, S. Studer, R., editors (2009). *Handbook on Ontologies*. International Handbooks on Information Systems. Springer, 2 edition.

Walisadeera, A. I., Ginige, A., Wikramanayake, G. N. (2015). User centered ontology for Sri Lankan farmers. *Ecological Informatics*, 26(P2):140–150.

Yingbo, D. Xia, H. (2019). Business modeling and reasoning based on process ontology. In *ACM International Conference Proceeding Series*, pages 143–147.

Zeleti, F. A. Ojo, A. (2017). An ontology for open government data business model. In *ACM International Conference Proceeding Series*, volume Part F1280, pages 195–203.

Zou, M., Basirati, M. R., Bauer, H., Kattner, N., Reinhart, G., Lindemann, U., Böhm, M., Krcmar, H., Vogel-Heuser, B. (2019). Facilitating consistency of business model and technical models in product-service-systems development: An ontology approach. *IFAC-PapersOnLine*, 52(13):1229–1235.



Perfil tecnológico e de automação das unidades armazenadoras de grãos do Estado do Paraná

Fernando Carvalho[1] **and Luiz Antonio Rodrigues**[1]

[1]Programa de Pós-Graduação em Ciência da Computação – PPGComp
Universidade Estadual do Oeste do Paraná – UNIOESTE
Cascavel – PR – Brasil

{fernando.carvalho3, luiz.rodrigues}@unioeste.br

Abstract. *Paraná is one of the leading states in grain production in Brazil and has a static storage capacity of approximately 30 million tons. The storage conditions and its monitoring have a direct impact on the quality of the grains, which avoids losses and adds value to the product. The objective of this work was to investigate the level of adoption of information technologies and automation of storage units in Paraná, through an online data collection instrument applied to those responsible for units registered in SICARM. From the main results, it was observed that the level of automation is still low, especially in the Internet of Things (IoT) scenario, which indicates new market opportunities.*

Resumo. *O Paraná é um dos estados líderes na produção de grãos no Brasil e possui uma capacidade estática de armazenamento de aproximadamente 30 milhões de toneladas. As condições de armazenamento e o seu monitoramento têm impacto direto na qualidade dos grãos, o que evita perdas e agrega valor ao produto. O objetivo deste trabalho foi investigar o nível de adoção de tecnologias da informação e automação das unidades armazenadoras do Paraná, por meio de um instrumento de coleta de dados online aplicado aos responsáveis pelas unidades cadastradas no SICARM. A partir dos resultados, observa-se que o nível de automação ainda é baixo, especialmente no cenário de Internet das Coisas (IoT), o que indica novas oportunidades no mercado.*

1. Introdução

A etapa de pós-colheita é constituída pelas fases do processo produtivo inerentes à conservação e o armazenamento dos alimentos, conforme Grolleaud (2002). A armazenagem está ligada ao conceito de abastecimento, que consiste em um conjunto de medidas necessárias para prover com regularidade, os alimentos que atendem às necessidades da

população onde, no Brasil, destacam-se os estoques de determinados grãos cereais, como soja e milho, além do arroz, feijão, trigo, café, entre outros. Segundo Puzzi (2000), quaisquer mudanças na disponibilidade, qualidade ou preço destes produtos ou seus derivados, afetam a rotina socioeconômica de uma região ou país.

Para o armazenamento em perfeitas condições para posterior redistribuição e industrialização, os grãos são acondicionados nas Unidades Armazenadoras, também denominadas UAs, comumente chamados de silos e armazéns. O decreto 3.855/2001 institui como Unidades Armazenadoras as edificações, instalações e equipamentos destinados à guarda e conservação dos produtos agropecuários, seus derivados, subprodutos e resíduos de valor econômico [BRASIL 2001].

Em geral, as edificações mais utilizadas para o armazenamento de grãos são os silos metálicos, os armazéns graneleiros e os armazéns convencionais [Puzzi 2000, Paturca 2014]. Os silos são compartimentos estanques ou herméticos que, em virtude da compartimentação disponível, permitem o controle das características físico-biológicas dos grãos, que são armazenados separadamente conforme sua espécie e padrões agrícolas. Já os armazéns graneleiros, também denominados silos horizontais, são estruturas de concreto ou alvenaria, com cobertura metálica, e que separam o local em depósitos onde a massa de grãos é armazenada. Por ser uma estrutura para grandes volumes, há uma maior dificuldade de controle homogêneo e um maior risco de deterioração. Finalmente, os armazéns convencionais são estruturas de piso plano, comumente construídos em alvenaria, estruturas metálicas ou mistas, onde os produtos são estocados em sacaria, e devidamente empilhados e dispostos em blocos. Embora tenham baixo custo de instalação, a movimentação dos produtos é prejudicada.

Ressalta-se que após a colheita de gêneros alimentícios, não se pode adicionar qualidade aos produtos, sendo possível apenas a sua manutenção ou a minimização da deterioração natural de suas propriedades orgânicas. Destaca-se, pois, a importância do aprimoramento das tecnologias de suporte ao armazenamento de grãos, devido a necessidade da manutenção da qualidade dos mesmos durante todo o tempo em que são acondicionados, antes de serem transformados em diversos subprodutos industrializados.

A CONAB (Companhia Nacional de Abastecimento), é o órgão que controla e regula a atividade de armazenagem de grãos no Brasil, responsável também por manter o SICARM (Sistema de Cadastro Nacional de Unidades Armazenadoras) [SICARM 2021]. Este cadastro *online* contém dados, como o endereço e a localização geográfica dos armazéns, a razão social do proprietário e demais informações para contato, o tipo da estrutura edificada, e a sua capacidade estática de estocagem. No cadastro atual, atualizado em 23/08/2021, constam 2.469 UAs no Paraná, que somam aproximadamente 30 milhões de toneladas. Destas, 1.286 são Bateria de Silos (52,09%), 528 Graneleiros (21,39%), 457 Armazéns Convencionais (18,51%), 169 Silos (6,84%) e 29 outras estruturas (1,17%).

Segundo relatório da Embrapa (Empresa Brasileira de Pesquisa Agropecuária) [EMBRAPA 2017], os sistemas baseados em Internet das Coisas (IoT) têm sido dedicados a monitorar ambientes das propriedades rurais, desde a análise de condições do solo até o monitoramento dos equipamentos e implementos utilizados em todas as fases da produção agrícola. No entanto, são poucas as iniciativas tecnológicas que apoiem o avanço da qualidade do armazenamento de grãos no Brasil.

Assim, o objetivo deste trabalho foi investigar o nível de adoção de tecnologias da informação, incluindo a utilização de IoT, no ambiente de Unidades Armazenadoras no Estado do Paraná, bem como o nível de conhecimento técnico relacionado a automação destas estruturas por parte dos gestores e operadores de silos e armazéns.

O restante deste trabalho está organizado nas seguintes seções. A Seção 2 apresenta a metodologia de aplicação do instrumento de coleta de dados. A Seção 3 discute os resultados obtidos com base nas respostas de questionários completos. Por fim, a Seção 4 apresenta a conclusão e trabalhos futuros.

2. Metodologia

Para realizar a pesquisa, foi tomado por base o cadastro público do SICARM e todos os dados de disponíveis das pessoas jurídicas proprietárias de silos e armazéns com base no extrato de 2020, filtradas aquelas localizadas no Paraná. Para tanto, um questionário *online* foi elaborado e enviado para os 1.170 gestores ou responsáveis por armazéns no Paraná cadastrados no SICARM, considerando os e-mails únicos.

O formulário para captação das respostas foi desenvolvido no Google Forms[1], contendo a caracterização da pesquisa, o TCLE (Termo de Consentimento Livre e Esclarecido), os dados de contato dos pesquisadores e as seções de perguntas. O envio e a gestão da sensibilização dos respondentes foi realizado com o Sendinblue[2]. Com o apoio desta ferramenta é possível verificar, além da quantidade de respostas completas e erros de entrega, quantos contatos realizaram a leitura do e-mail, abriram o *link* de resposta e optaram por não responder a pesquisa. Na primeira etapa foram enviados 1.323 convites. Destes, 1.051 (79,4%) tiveram confirmação de entrega e o restante foi marcado com problemas na caixa de correio ou no cadastro. Dos entregues, 208 (15,7%) foram abertos, mas apenas 46 (3,48%) acessaram o questionário. Após cada rodada de sensibilização, as informações atualizadas foram utilizadas para o envio de lembretes mensais.

Uma vez que o avanço na captação de respostas via e-mail seguiu mais lento do que o esperado, por diversos motivos, desde a utilização de contas de e-mail que são pouco verificadas ou geridas por pessoas que não fazem parte do grupo de pesquisa, até mesmo pelo bloqueio do recebimento do contato por filtros *anti-spam*, foi iniciada uma fase de contatos por telefone. Os contatos telefônicos dos respondentes também foram obtidos no SICARM e, somente nos casos em que não havia tal informação, o número foi localizado em sites de busca pelo nome da empresa responsável pelo armazém. Nos meses de junho a outubro de 2020 foram contatadas 106 empresas para a sensibilização de resposta ao questionário, e convidadas a responder de imediato por telefone e registradas as respostas manualmente, ou solicitado o acesso posterior ao *link* do formulário. A partir deste processo foram obtidas 15 respostas.

No total, somadas as iniciativas de contato por e-mail e telefone, foram obtidas 45 respostas completas, devidamente registradas no formulário criado para a captação dos dados, seja por iniciativa dos próprios respondentes, seja por lançamento manual pelo responsável da pesquisa. Isto equivale a 4,28% dos e-mails válidos.

[1]Google Forms: `https://forms.google.com`
[2]Sendinblue: `https://pt.sendinblue.com`

3. Resultados e discussão

Os resultados da pesquisa foram obtidos por meio de um questionário constituído por 34 perguntas objetivas divididas nas seções apresentadas a seguir.

3.1. Seção I - Perfil do respondente e ambiente de armazenamento da empresa

Na primeira seção do instrumento são abordadas as questões de formação e atuação do respondente, bem como sobre seu conhecimento relativo ao ambiente de armazenagem da empresa. Inclui ainda o perfil das UAs sob responsabilidade da empresa.

Em relação ao nível de formação escolar, a maioria possui ensino superior (53,3%) ou especialização (40%). Ensino médio e nível técnico em agropecuária totalizam 4,4%. Mestrado ou doutorado, 2,2%. O nível hierárquico dos respondentes é majoritariamente de Supervisores ou Coordenadores (46,7%), seguido de Gerente ou Diretor (40%) e Técnico/Analista (13,3%). Quanto ao tempo de atuação na empresa, a maior parte atua há mais de 10 anos (48,9%). A atuação de 1 a 5 anos representa 28,9% e outros 20% atuam entre 6 e 10 anos. Apenas um respondente trabalha há menos de um ano na função, que representam 2,2% do total.

A capacidade de armazenamento das unidades sob gerência dos respondentes está registrada na Tabela 1. A maioria das empresas (cerca de 64%) possuem capacidade de até 50 mil toneladas. O restante, cerca de 35%, possuem capacidade de armazenamento superior a 50 mil, sendo em 12 delas, superior a 100 mil.

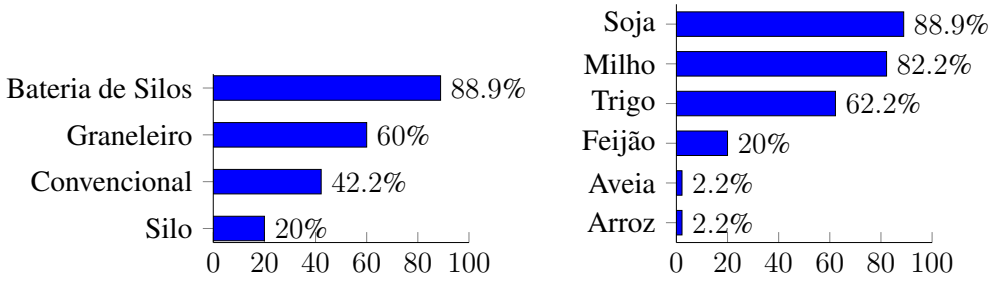
Tabela 1. Capacidade aproximada de armazenamento de grãos da empresa.

Capacidade de Armazenamento	Resultado	
De 5 mil até 10 mil t.	11	(24,4%)
De 10 mil até 50 mil t.	18	(40,0%)
De 50 mil até 100 mil t.	4	(8,9%)
Acima de 100 mil t.	12	(26,7%)
Total	45	(100.0%)

O gráfico da Figura 1(a) representa os tipos de armazéns agrupados individualmente conforme as respostas. Verifica-se que o tipo Bateria de Silos é o mais utilizado nas empresas pesquisadas, correspondendo a 88,9%. O segundo tipo é o Graneleiro, com 60%, e o tipo Silo individual foi indicado por 20%. Já a categoria Convencional, onde os grãos não são estocados a granel, totaliza 42,2% das respostas. Por se tratar de uma questão de múltipla escolha, a soma ultrapassa os 100%.

Verifica-se pela Figura 1(b) que as empresas consultadas armazenam majoritariamente soja e milho, tendo a soja 88,9% de indicações, e o milho com 82,2%. Observa-se ainda uma considerável parcela de armazenamento de trigo (62,2%), assim como algumas ocorrências de armazenadores de feijão, aveia e arroz.

Na Tabela 2, que contém os dados da idade média das edificações de armazenamento, é possível observar que 80% das mesmas possuem mais de 10 anos, sendo que neste arranjo, 23 informaram idade entre 10 e 20 anos, e 13 informaram que as UAs possuem mais de 20 anos. Desta forma, 20% dos respondentes indicaram que as instalações possuem menos de 10 anos, sendo que apenas 2 destas foram construídas há menos de 5 anos.



(a) Tipos de armazéns (b) Grãos armazenados

Figura 1. Resultados agrupados para a) tipos de armazéns instalados e b) tipos de grãos armazenados nas empresas consultadas.

Tabela 2. Idade estimada das unidades armazenadoras de grãos.

Idade Estimada das UAs	Resultado
Menos de 5 anos	2 (4,4%)
Entre 5 e 10 anos	7 (15,6%)
Entre 10 e 20 anos	23 (51,1%)
Mais de 20 anos	13 (28,9%)
Total	45 (100,0%)

Para descobrir também o entendimento dos operadores em relação aos conceitos de "Automação de Silos", 88,9% consideraram-se parcialmente ou já familiarizados com o assunto, enquanto 86,7% indicaram conhecer o tema "Internet das Coisas", ao menos em partes, como pode ser visto na (Tabela 3).

Tabela 3. Familiaridade dos respondentes com os conceitos de "Automação de Silos" e "Internet das Coisas".

Questão	Não	Sim	Parcialmente
Você está familiarizado com o conceito de Automação em Silos e Armazéns de grãos?	5 (11,1%)	28 (62,2%)	12 (26,7%)
Você está familiarizado com o conceito de IoT ou "Internet das Coisas"?	6 (13,3%)	18 (40,0%)	21 (46,7%)

3.2. Seção II - Uso de tecnologias de automação

Na segunda seção de perguntas, o questionário aborda sobre o uso de tecnologias de informação nos ambientes de armazenagem das empresas. As questões foram organizadas em uma única tabela, na qual as linhas separam diferentes tecnologias como sensores digitais, sistemas MES/SCADA ou CAD/CAM/CAE, simulações virtuais, análise de *Big Data* e utilização de Computação em Nuvem, e as colunas indicam o nível de adoção ou intenção de utilização para cada uma delas.

De acordo com os dados contidos na Tabela 4, 84,4% das empresas pesquisadas utilizam sensores sem automação. A indicação de não utilização de sensores (15,6%), pode ser justificada por equívocos no preenchimento, pelo desconhecimento do respondente quanto a existência de sensores na UA, ou ainda por se tratar de armazém do tipo Convencional que não dispõe deste recurso. Já a utilização de sensores baseados em

terminais é indicada por 55,6% dos pesquisados. Com relação as demais tecnologias, destaca-se a baixa adesão às tecnologias mais recentes, como a simulação e análise através de modelos virtuais indicada por 6,7%, o emprego de processamento de grandes bases de dados (*Big Data*) e de dispositivos IoT, apontados por 8,9% dos pesquisados. Ressalta-se, porém, a intenção dos pesquisados em aderir a estas novas tecnologias nos próximos anos, tendo indicado a possibilidade de adotar, em até 5 anos ou mais, a computação em nuvem (20%), *Big Data* (26,7%) e IoT (33,3%).

Tabela 4. Utilização de tecnologias de automação, monitoramento e informação nas unidades armazenadoras de grãos.

Tecnologia	Utiliza	Não Utiliza	Pretende adotar em Até 5 anos	Pretende adotar em 5 anos ou mais
Sensores sem automação	38 (84,4%)	7 (15,6%)	0	0
Automação digital s/ sensores baseada em terminais	25 (55,6%)	17 (37,8%)	3 (6,7%)	0
Automação digital c/ sensores p/ controle de processos ou ident. de produtos	12 (26,7%)	21 (46,7%)	8 (17,8%)	4 (8,9%)
Monitoramento e controle remoto de produção com sist. MES/SCADA	11 (24,4%)	38 (84,4%)	3 (6,7%)	3 (6,7%)
Sistemas Integrados de Eng. (CAD/CAM/CAE/PDM)	12 (26,7%)	27 (60,0%)	4 (8,9%)	2 (4,4%)
Simulação/análise de modelos virtuais	3 (6,7%)	34 (75,6%)	6 (13,3%)	2 (4,4%)
Coleta, proc., análise de gdes. qtds. de dados (*Big Data*)	4 (8,9%)	29 (64,4%)	7 (15,6%)	5 (11,1%)
Utilização de serviços em nuvem relacionados à operação (*Cloud Computing*)	12 (26,7%)	24 (53,3%)	5 (11,1%)	4 (8,9%)
Incorporação de serviços digitais aos equipamentos da operação (IoT)	4 (8,9%)	26 (57,8%)	9 (20,0%)	6 (13,3%)

3.3. Seção III - Desafios na adoção de novas tecnologias

O objetivo desta seção era determinar quais são os itens possivelmente impeditivos, na visão dos entrevistados, relacionados aos custos de implantação e manutenção voltados a adoção de novas tecnologias no processo de armazenagem de grãos em suas unidades.

Verifica-se na Tabela 5 que 60% das empresas consideraram irrelevante ou pouco impeditiva a falta da padronização de protocolos de comunicação industrial que possam ser utilizados na operação de armazenamento de grãos. Todavia, 82,2% indicaram ser moderado ou altamente impeditivo o custo para adoção destas tecnologias e sistemas.

Em relação às despesas de manutenção, 71,1% dos respondentes informaram que este fator é impeditivo em um nível moderado ou alto, assim como outros 71,1% apontaram também a falta de profissionais capacitados na operação dos sistemas. A respeito dos riscos ou a incerteza sobre o Retorno do Investimento (ROI) em novas tecnologias, houve um equilíbrio nas respostas, onde 53,3% consideram como desafio pouco impeditivo ou irrelevante, enquanto 47,7% julgam ser moderadamente ou altamente impeditivo.

Tabela 5. Nível dos desafios enfrentados pelas empresas para a adoção de novas tecnologias ou sistemas nas unidades de armazenamento.

Desafios para adoção de tecnologias	Irrelevante	Pouco Impeditivo	Moderadamente Impeditivo	Altamente Impeditivo
Falta de padronização nos protocolos de comunicação industrial	9 (20,0%)	18 (40,0%)	14 (31,1%)	4 (8,9%)
Custo para adoção das tecnologias/sistemas	2 (4,4%)	6 (13,3%)	18 (40,0%)	19 (42,2%)
Despesas de manutenção das tecnologias/sistemas	1 (2,2%)	12 (26,7%)	13 (28,9%)	19 (42,2%)
Riscos ou incerteza sobre o ROI nas tecnologias/sistemas	5 (11,1%)	19 (42,2%)	10 (22,2%)	11 (24,4%)
Falta de recursos financeiros	3 (6,7%)	17 (37,8%)	17 (37,8%)	8 (17,8%)
Falta de profissionais capacitados na operação das tecnolog/sist.	4 (8,9%)	9 (20,0%)	18 (40,0%)	14 (31,1%)
Falta de capacidade técnica em TI	7 (15,6%)	18 (40,0%)	12 (26,7%)	8 (17,8%)
Risco para operação	5 (11,1%)	18 (40,0%)	17 (37,8%)	5 (11,1%)
Risco para a segurança dos dados/informação	6 (13,3%)	23 (51,1%)	6 (13,3%)	10 (22,2%)

A falta de recursos financeiros para a adoção de novas tecnologias é considerada um desafio alto ou moderado para 55,6%. Já no aspecto técnico, 71,1% apontaram a falta de capacidade técnica na operação como impeditivo, ao passo que 44,4% julgam a falta de capacidade técnica na equipe de TI. Com relação aos riscos, 51,1% consideram irrelevante ou pouco impeditivo, o risco para a operação de armazenagem, enquanto 64,4% subestimam o risco à segurança da informação.

3.4. Seção IV - Resultados esperados quando da adoção de novas tecnologias

Na última seção, as perguntas são relacionadas aos resultados pretendidos com a adoção de novas tecnologias em UAs, com foco no aumento das receitas pela manutenção da qualidade dos produtos armazenados, a manutenção do mercado atual e a entrada em novos mercados, a melhoria na operação ou processo de armazenagem, e a inovação.

Tabela 6. Resultados pretendidos no caso de adoção de novas tecnologias ou sistemas nas Unidades de Armazenamento.

Resultados pretendidos	Nenhum Resultado	Baixos Resultados	Alguns Resultados	Altos Resultados
Aumento das receitas pela qualidade do produto	0	2 (4,4%)	19 (42,2%)	24 (53,3%)
Entrada em novos mercados	6 (13,3%)	10 (22,2%)	14 (31,1%)	15 (33,3%)
Manutenção do mercado atual	1 (2,2%)	10 (22,2%)	19 (42,2%)	15 (33,3%)
Melhoria na operação ou processo	0	1 (2,2%)	14 (31,1%)	30 (66,7%)
Melhoria por inovação	1 (2,2%)	5 (11,1%)	14 (31,1%)	25 (55,6%)

A melhoria na operação ou no processo, foi o resultado de mais alta expectativa indicado pela maioria dos respondentes (98,8%), seguido pelo aumento nas receitas pela qualidade do produto, com 95,6%. Na sequência foram apontados ainda a melhoria por inovação (86,7%) e a manutenção dos mercados atuais (75,5%). Já a entrada em novos mercados foi o resultado de menor expectativa (64,4%).

3.5. Discussões

Verifica-se que a grande maioria dos profissionais responsáveis pelas UAs do Paraná possuem ensino superior ou pós-graduação em algum nível, *lato sensu* ou *stricto sensu*, e que em geral, atuam há pelo menos 5 anos neste setor.

Quanto ao perfil das UAs, observa-se que os tipos Bateria de Silos e Graneleiros predominam nas empresas, assim como destaca-se a idade avançada das estruturas, onde a maioria pesquisada encontra-se nas faixas acima de 10 anos ou mais. Denota-se também que os grãos armazenados por estas estruturas são, em sua maioria, soja e milho.

Nota-se também, que há familiaridade dos respondentes com os conceitos de automação de silos, tendo aproximadamente 88% indicado conhecê-los parcial ou totalmente. Além disso, cerca de 86% afirmaram entender a definição de Internet das Coisas.

Ainda com relação às aplicações de tecnologias na operação de silos e armazéns, se observa a predominância no emprego de sensores sem automação ou até mesmo de sensores digitais automatizados. No entanto, o uso de tecnologias mais recentes na automação e monitoramento das estruturas e dos produtos estocados, como a Computação em Nuvem e a Internet das Coisas, são pouco frequentes na rotina dos entrevistados. Este cenário é semelhante ao discutido por [Puzzi 2000] acerca do monitoramento em UAs que utilizam sensores termopares e pela observação de dezenas de armazéns que ainda utilizam este sistema de controle.

Finalmente, os principais desafios indicados para a utilização de novas tecnologias na operação de UAs, foram os custos das tecnologias e sistemas para este propósito, tanto para a implantação quanto para a manutenção. Em contrapartida, esperam o aumento das receitas pela maior qualidade do produto, e a melhoria no processo operacional, quando da aplicação destas tecnologias.

4. Conclusão

O objetivo desta pesquisa foi verificar qual é o nível de automação das UAs do Estado do Paraná, incluindo o perfil dos gestores e profissionais que as operam, as suas expectativas em relação a adoção de novas tecnologias que permitam o aprimoramento da atividade, bem como quais os desafios mais relevantes para adoção destas.

Para tanto, um questionário online foi inicialmente enviado aos e-mails dos gestores cadastrados no banco de dados do SICARM. Posteriormente, considerando o grande número de falhas de entregas e o baixo retorno de respostas, foram também realizadas algumas ligações telefônicas.

Embora considerável parcela dos respondentes afirme conhecer os conceitos de automação de silos (62,2%) e Internet das Coisas (40%), verifica-se pelos resultados que há uma baixa utilização pelos sistemas destas categorias nas UAs, especialmente para a coleta de informações ambientais, tão importantes para a manutenção da qualidade dos grãos. Embora o custo seja considerado impeditivo, há interesse na adoção de novas tecnologias em até 5 anos ou mais, mas ainda pouco significativa.

Referências

BRASIL (2001). Decreto no 3.855, de 3 de julho de 2001. Acesso em: 22 jul. 2020.

EMBRAPA (2017). Internet das coisas (iot): inovação para o agronegócio. *EMBRAPA - Secretaria de Inteligência e Macroestratégia: Diálogos Estratégicos*.

Grolleaud, M. (2002). Post-harvest losses: discovering the full story. Overview of the phenomenon of losses during the Post-harvest. Acesso em: 10 nov. 2019.

Paturca, E. Y. (2014). *Caracterização das estruturas de armazenagem de grãos: um estudo de caso no Mato Grosso*. PhD thesis, Universidade de São Paulo.

Puzzi, D. (2000). *Abastecimento e armazenagem de grãos*. Instituto Campineiro de Ensino Agricola.

SICARM (2021). Sistema de Cadastro Nacional de Unidades Armazenadoras. Acesso em: 27 abr. 2021.



Processamento de Linguagem Natural para consultas de invasores na cultura da soja

Carolinne Roque e Faria[1], Cinthyan Renata Sachs C. de Barbosa[1]

[1]Programa de Pós Graduação em Ciência da Computação –
Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – 86.057-970 – Londrina – PR – Brazil

carolinne.rf@outlook.com, cinthyan@uel.br

Resumo. *As tecnologias para auxiliar na produtividade do campo estão em constantes desenvolvimentos com a expectativa de estarem em sintonia com as demandas da área. Assim, a proposta deste trabalho é apresentar ferramentas computacionais em Processamento de Linguagem Natural para auxiliar os produtores rurais no controle das principais pragas e doenças na cultura da soja para não prejudicar a produtividade e, a partir da extração de suas características que serão armazenados em um repositório, seja possível identificar os vilões que atacam a lavoura ajudando-os nas tomadas de decisões de acordo com as consultas feitas a essa Base de Dados.*

Abstract. *The technologies to assist in productivity in the field are in constant development with the expectation of being in harmony with the demands of the area. Thus, the proposal of this paper is to present computational tools in Natural Language Processing to assist farmers in the control of the main pests and diseases in soybean crops in order to avoid affecting productivity and, from the extraction of their characteristics that will be stored in a repository, it is possible to identify the villains that attack the crop, helping them in making decisions according to the queries made to this database.*

1. Introdução

No Brasil, a modernização do campo caminha em grandes passos com o intuito de aumentar a produção e melhorar as condições econômicas por meio de investimentos em inovações tecnológicas, pesquisas científicas e créditos rurais. A agricultura é uma das principais bases da economia de diversos países, principalmente no Brasil [Hiraoka e Jacopini 2018] que nesse setor tem seu potencial reconhecido globalmente. O Centro de Estudos Avançados em Economia Aplicada [CEPEA 2019] da ESALQ da USP, em

parceria com a Confederação da Agricultura e Pecuária do Brasil e com a Fundação de Estudos Agrários Luiz de Queiroz, calcularam o Produto Interno Bruto (PIB) do Agronegócio Brasileiro em 2019 e o desempenho representou significativamente 21,4% do PIB brasileiro total, o que revela uma ampliação da sua participação na economia.

A cultura da soja representa um dos principais propulsores do agronegócio brasileiro transformando a vida de milhões de cidadãos com a geração de emprego e renda, melhorias de infraestrutura e aumento do IDH *(Índice de Desenvolvimento Humano)* nas regiões produtoras [EMBRAPA 2019]. A CONAB *(Companhia Nacional de Abastecimento)* divulgou no mês de junho o 9º levantamento de grãos para a safra 2018/2019 no mercado nacional, que relata o aumento de produção de soja como vimos.

Descrição/Safra	2016/17	2017/18	2018/19 (*)
Estoque Inicial	5.405,4	7.779,8	1.390,7
Produção	114.075,3	119.281,7	114.843,3
Importação	253,7	187,0	150,0
Suprimento	119.734,4	127.248,5	116.384,0
Consumo total	43.800,0	42.600,0	44.000,0
Exportação	68.154,6	83.257,8	68.000,0
Estoque Final	7.779,8	1.390,7	3.184,0

Figura 1. Soja em grãos. Fonte: [CONAB 2019]

Devido às ciências e à produção de conhecimento, os sistemas avançados são fundamentados em práticas agrícolas. Ter uma ferramenta que auxilie o produtor sobre as principais pragas e doenças na cultura, bem como garantir a sanidade das plantas significa melhor produção e, desse modo, essa pode lidar melhor com as necessidades da lavoura. O gerenciamento de dados agrícolas depende de informações adquiridas de diversas tecnologias e sistemas que sejam capazes de auxiliar na tomada de decisão nas técnicas agrícolas. Assim, é fundamental mapear os dados (principais características de identificação de pragas e doenças na cultura da soja) e padronizá-los.

Indurkhya e Damerau (2010) salientam que o Processamento de Linguagem Natural (PLN) tem como objetivo extrair representações e significados mais completos de textos livres escritos em linguagem natural. PLN responde de forma inteligente por meio de algoritmos em uma linguagem não só escrita, mas também falada.

Assim, este presente trabalho faz uso de uma grande quantidade de informações armazenadas e disponibilizadas em rede aberta e aplica técnicas de PLN em um dicionário para consultas de domínio agrícola para auxiliar o agricultor a melhorar a sua produção e evitar prejuízos. O artigo está dividido da seguinte maneira: a seção 2 descreve os materiais e métodos, na seção 3 são expostos os resultados e discussões e na seção 4 as conclusões.

2. Materiais e Métodos

O PLN desempenha um papel imprescindível na comunicação, pois suas técnicas refletem na compreensão da Linguagem Natural (LN), passando por Análise Léxico-Morfológica *(tokenização)*, Análises Sintática *(parsing)*, Semântica e Pragmática.

Para a construção deste trabalho são utilizadas algumas ferramentas auxiliares necessárias para o desenvolvimento do software em PLN para agricultura que serão descritas ao decorrer do artigo. Um chatbot será apresentado para conversar com estudantes de agronomia, agronegócio, engenheiros, produtores rurais ou pessoas que

gostariam de adquirir conhecimentos para evitar o ataque das principais pragas, insetos, nematoides, doenças e fungos na cultura da soja. A interface desse é desenvolvida em LN, o que facilita a comunicação entre o sistema e o usuário. Para permitir esse tipo de comunicação, os chatbots podem utilizar recursos de PLN [Khanna et al. 2015]. Alguns chatbots foram encontrados para uso na agricultura como, por exemplo, os trabalhos de Sawant et al. (2019) e Mostaço et al. (2018), porém nenhum deles são sobre a soja.

Para a elaboração do chatbot em PLN para consulta em Banco de Dados foi feito um sistema de interpretação e geração em linguagem natural. Alguns recursos são essenciais durante a fase de interpretação, bem como na de geração, e foram baseados em Silva et al. (2007) para este trabalho envolvendo: **Léxico:** conjunto de palavras que quando acessado por analisadores léxico, sintático e semântico reconhece os *tokens* e fornece além dos traços gramaticais, os traços semânticos de suas entradas para possibilitar a verificação semântica. **Gramática:** conjunto de regras gramaticais que definem as frases válidas em uma sentença em linguagem natural. **Modelo de Domínio:** fornece informações sobre o domínio específico da aplicação. **Modelo de Usuário:** permite reconhecer características do significado textual a partir do contexto do discurso

Após isso foi possível estabelecer os inúmeros modos de escrever, fazendo com que o sistema manuseie automaticamente a geração da tarefa e que o processo dessa transmita informações continuamente em uma base de dados em PLN.

Para este trabalho contou-se com as mais diversas fontes, tais como manuais e livros [Santos 1995] [Moreira e Aragão 2009] [Sosa-Gomez 2010] [Ávila 2017], teses/ dissertações, *sites* como da *Empresa Brasileira de Pesquisas Agropecuária* [EMBRAPA 2019], artigos e profissionais da área para selecionar as palavras para a criação do dicionário, com a finalidade de fornecer informações, permitindo que a mesma pergunta seja feita de várias maneiras, relacionada às características que se manifestam durante a produção de soja.

As palavras armazenadas em um dicionário totalizam 108 ameaças que são associadas às informações gramaticais. O registro de cada palavra foi feito de acordo com algumas categorias morfológicas, as quais são avaliadas isoladamente e separadas em classes gramaticais pelo processo de *Part of Speech* (POS) *Tagging*, isto é, para cada palavra de um texto é feita a identificação da classe gramatical a que ela pertence baseada em sua definição e no contexto, visto que uma palavra pode pertencer a mais de uma classe ou possuir mais de um significado [Moraes 2019].

Este trabalho gera uma estrutura de árvore que represente a estrutura sintática da sentença analisada e foi desenvolvido por meio da linguagem de programação Python e o analisador sintático escolhido foi o spaCy (2020) que utiliza-se de uma biblioteca de PLN, deixando-o com um desempenho superior a outras bibliotecas da área. Assim, foi possível fazer o processo de POS *Tagging* (classificar todas as sentenças do texto por categorias gramaticais e a relação de dependência entre palavras) com tal biblioteca.

A ferramenta spaCy (2020) classifica cada palavra do texto conforme às suas classes gramaticais como artigo, conjunção, numeral, pronome, símbolos, substantivo, verbo etc., com o intuito de encontrar as características específicas sobre um determinado assunto. Ela é uma biblioteca para análise sintática e permite fazer o NER (*Named-entity recognition* - Reconhecimento de Entidades Nomeadas) que tem a função

de identificar e classificar palavras ou frases em um texto de acordo com classes definidas para o modelo [Nadeau 2007].

Neste trabalho optou-se por um Banco de Dados Não Relacional, também conhecido como Banco de Dados (BD) NoSQL (*Not Only Structured Query Language*). Esses são uma família de gerenciadores de dados que não seguem o modelo de dados relacional. O NoSQL escolhido é o tipo Grafo. Como é mostrada na Figura 2 tem-se a estruturação dos dados armazenados em nós dos Ácaros/Fitófagos e associação às informações por meio das arestas do Ácaro-Rajado, Ácaro-Branco e Ácaro-Verde.

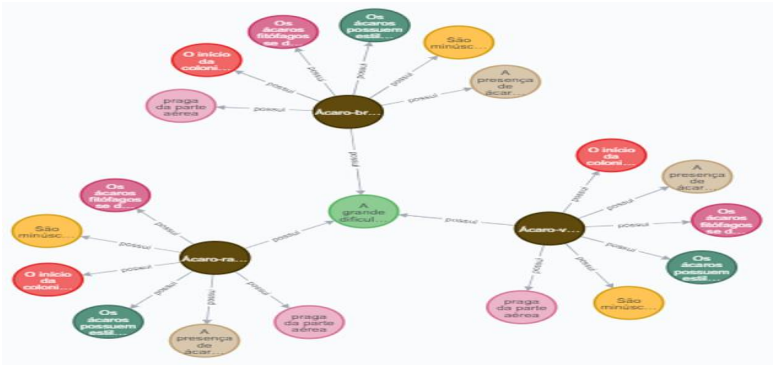


Figura 2. Dados dos Ácaros Fitófagos no Banco de Dados NoSQL Neo4j

O objetivo é criar uma interface computacional em LN para garantir um bom desempenho das tarefas na área da agricultura, como uma ferramenta de conversação entre o sistema e o usuário que possibilita um fácil acesso às informações em um repositório da base de dados e visa aumentar esse conforme ocorre o diálogo e, então, os dados crescem exponencialmente. Assim, o BD armazena não só os dados, mas as suas relações de maneira eficiente.

Mitchell (1997) define Aprendizado de Máquina (AM) como a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência. Técnicas de AM são essenciais para calcular o desempenho dos dados para o futuro e foram aplicadas a fim de detectar pragas automaticamente a partir de textos. Com o grande volume de dados propõe-se treinar os modelos de classificação para avaliar o desempenho de doenças da soja.

A fase de pré-processamento de dados pode ser considerada a mais importante para a aplicação das tarefas de AM, pois são feitas, segundo Bird, Klein and Loper (2009): - remoção de *stopwords* que são palavras como artigos ou verbos de ligação que aparecem nos textos várias vezes, mas praticamente não influenciam a classificação; - lematização que é a redução de palavras a seus radicais, removendo flexões de tempo verbal, gênero, número; - tokenização que é o processo de criação de um vetor de termos de um documento, onde cada termo ocupa um índice do vetor. Nesse contexto aplicam-se técnicas para extração e classificação de textos na identificação de pragas e doenças a partir das características da praga na planta por meio de AM, a fim de analisar em menor tempo e avaliar o grau de severidade do dano na lavoura.

Os resultados obtidos a partir da predição para representar o comportamento real da identificação é predizer o desempenho do modelo para o futuro e a principal fonte é calculada pela matriz de confusão [Olson and Delen 2008]. Ao realizar a aplicação do

modelo faz-se a medição da precisão para calcular a taxa de acerto (razão entre os valores previstos e valores reais) e erro. As predições podem ser Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN). A partir dela apresentam-se os valores previstos e reais por meio de forma tabular e baseada nessa contagem, a qual várias métricas podem ser calculadas.

Aqui foram aplicadas as principais métricas em problemas de classificação [Abonizio et al. 2019]: **(a)** acurácia que informa a taxa de acerto; **(b)** a métrica de precisão é uma equação que é medida a proporção dos valores previstos negativos com os verdadeiros negativos, ou seja, essa apresenta quão corretas estão as predições; **(c)** a revocação que é conhecida como sensibilidade, a qual calcula a proporção dos valores previstos positivos com os verdadeiros positivos e retorna a fração dos documentos que foram relevantes; **(d)** métrica F1-*score* que calcula a média da precisão e revocação.

Para validar o modelo de AM é utilizada a técnica de Árvore de Decisão, cuja fase de treinamento é feita por meio de um sistema de indução de árvores baseada na divisão recursiva [Steinberg and Colla 1995]. Após a construção da árvore pode-se classificar novos exemplos a partir dessa [Silva e Vieira 2007]. A classificação é feita percorrendo a árvore até chegar à folha que determina a classe na qual o exemplo pertence ou sua probabilidade de pertencer àquela classe.

No presente trabalho foi aplicado o aprendizado supervisionado para fornecer ao modelo os dados que já possuem significados e serão categorizados, ou seja, o algoritmo irá conferir a precisão do modelo criado utilizando o conjunto de dados para predizer a eficácia. O método de análise de dados descrito automatiza o processo de criação de modelos. Utilizando algoritmos que iterativamente aprende com os dados, o AM permite que os computadores encontrem padrões escondidos nos dados sem terem sido programados para essa finalidade.

A principal dificuldade no desenvolvimento do sistema para identificação de pragas e doenças por meio de PLN é dialogar com o usuário e reconhecer as suas intenções a partir de uma frase e respondê-lo automaticamente. Assim, este sistema inteligente de pré-atendimento aos profissionais do campo agronômico tem o intuito de ser um canal alternativo de comunicação para facilitar o acesso às informações e auxiliar no ensino para identificar o patógeno. Para o pré-processamento da LN foram aplicados algoritmos de AM à base de dados para a construção de classificadores que pudessem predizer a causa dos sintomas das plantas da soja. A partir dos resultados é possível tomar decisões, o que permite otimizar as atividades guiadas à agricultura.

Para treinar as métricas, foram escolhidos os algoritmos de classificação Florestas aleatórias (*Random Forest* – RF) [Breiman 2001], Máquinas de Vetores Suporte (*Support Vector Machine* - SVM) [Scholkopf 2002], K-vizinhos mais próximos (*K-Nearest Neighbors* - KNN) [Weinberger, Blitzer and Saul 2006].

Os resultados do treino de classificação (Tabela 1) são compostos pela média das métricas de precisão, revocação e f1-*score* para cada praga que danifica a planta. As métricas apresentadas indicam que o modelo obteve uma performance alta, conseguindo uma média de F1-*score* de 94% e 97% de acurácia. Por meio do resultado experimental foi possível analisar o desempenho do classificador RF na identificação de 101 pragas que prejudicam a lavoura da soja e isso se mostrou robusto atingindo acurácia de 99%.

Tabela 1. Comparação entre os algoritmos

Algoritmos	Precision	Recall	F1-score	Accuracy
KNN	0.87	0.92	0.88	0.96
Random Forest	0.97	0.93	0.95	0.96
SVM	0.94	0.94	0.94	0.97
Algoritmos	**Precision**	**Recall**	**F1-score**	**Accuracy**
SVM - Kernel	1.00	0.98	0.99	0.99

Foi desenvolvida a Ferramenta CAROLINA (acrônimo para **C**onversação **A**gronômica **RO**botizada em **LI**nguagem **NA**tural) para identificação de pragas e doenças na cultura da soja e adotado o modelo em cascata para o desenvolvimento dessa que utiliza das fases de análises de requisitos, projeto, implementação, testes (validação), integração e manutenção de software. O sistema foi desenvolvido na linguagem de programação Python, com a utilização da biblioteca spaCy, para a criação de um chatbot, que elabora, por exemplo, as seguintes questões proporcionando ricas consultas, como essas: • *Os cupins são comuns em qual região?* • *Quais os prejuízos causados pela Lagarta-rosca?* • *O Ácaro-branco é o mesmo que o Ácaro-do-Bronzeamento?* • *O Piolho-de-cobra se alimenta do quê?* • *Tem uma larva atacando as folhas, flores e vagens. Que larva é?* • *O Cupim-subterrâneo ataca que região da planta da soja?* • *Minha cultura está demorando para se desenvolver. Por quê?* • *Minha planta está enfraquecendo. Por quê?* • *A minha planta teve queda das folhas durante o período de produção. Por quê?* • *As folhas estão amareladas. Por quê?* Esse sistema recebe o texto por meio de um diálogo (Figura 3) com o utilizador, analisa em um primeiro momento as palavras das sentenças isoladamente e passando pela compreensão da frase como um todo até concluir uma resposta da frase requisitada.

3. Resultados e Discussões

As etapas para o desenvolvimento do Sistema da Soja são compostas por: • definições das funcionalidades do sistema para permitir que o usuário controle as funções desse; • levantamentos das informações e técnicas necessárias para a construção das partes mais importantes do sistema para determinar a melhor maneira de realizar uma tarefa; • definições das ferramentas necessárias para a construção do sistema para o desenvolvimento de aprendizado de máquina; • construção do sistema seguindo uma metodologia de desenvolvimento de software para norteá-lo; • análises dos resultados obtidas com os testes para chegar às conclusões e verificar se o objetivo foi atingindo.

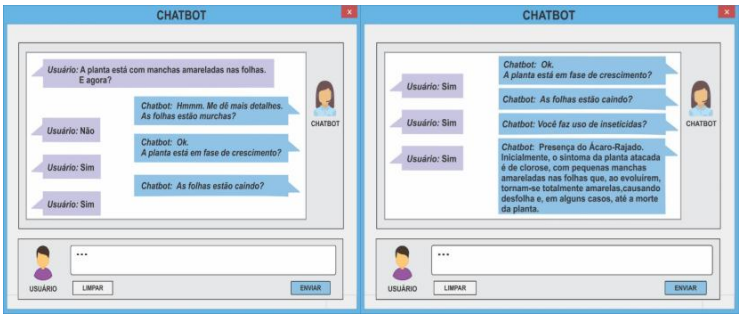


Figura 3. Interface de Consulta à Base de Dados das Pragas da Soja

A elaboração deste trabalho permitiu contribuir com a área agrícola, principalmente no que diz respeito à viabilidade de um assistente virtual, utilizando técnicas de PLN na identificação de pragas e doenças nas plantas da soja.

A partir das classificações das características das pragas e doenças da soja, um modelo foi construído por meio da biblioteca spaCy (2020), utilizando técnicas de PLN para aplicar regras gramaticais às sentenças, reconhecendo suas estruturas e extraindo seus significados, onde o pré-processamento envolveu as seguintes fases:
• **Tokenização:** subdivide a base de dados em tokens para uma pergunta como "*As folhas atacadas ficam com grandes áreas recortadas ou são completamente consumidas?*"
• **Lematização e Stemização**: a *lematização* é o processo de redução de palavras à sua base (raiz), enquanto a *stemização* permite checar se uma palavra é raiz de outra. São etapas importantes para ánalise de grandes volumes de textos.
• **POS *tagging*:** classifica corretamente todas as palavras de cada sentença do texto por categorias gramaticais.
• **Análise Léxico-Morfológica:** responsável por manipular o léxico que é composto por palavras que armazenam seus significados. Identificar as palavras individualmente é essencial porque ajuda a entender as frases de entrada e constrói com mais exatidão as saídas (respostas).
• **Analise Sintática:** responsável por organizar o conjunto das palavras e aplicar regras gramaticais à sentença para reconhecer sua estrutura; Posteriormente, essa ferramenta rotula todos os *tokens* a partir da análise para prever qual *tag* provavelmente se aplica nesse contexto. Para isso, conta-se com as seguintes tarefas: **Text** (texto puro), **Lemma** (reduz as palavras em seu formato base/raiz), **POS** (*tags* simples que determinam as categorias gramaticais de um *token*), **Dep** (Dependência sintática é a relação entre os *tokens* presentes dentro de uma sentença para entender o seu significado), **Shape** (classificação da palavra em maiúscula ou minúscula), **Alpha** (especificação das palavras em alfanuméricas ou não), **Stop** (indica se as palavras são consideradas *stopwords*), como é detalhada na Figura 4 para a frase: "*As folhas atacadas ficam com grandes áreas recortadas ou são completamente consumidas?"* O spaCy (2020) possibilita descrever a relação sintática das palavras que se conectam na formação da árvore. Isso permite percorrê-la toda, retornar uma sequência ordenada de *tokens* e verificar os atributos e domínios das palavras. Nessa fase conta-se com o Head Text (relação entre as palavras nos *tokens*), Head Pos (rotula as palavras em categorias) e Children (dependentes sintáticos do *token*).



```
for token in doc:
    print(token.text, token.lemma_, token.pos_, token.tag_, token.dep_,
          token.shape_, token.is_alpha, token.is_stop)
```

```
As As DET <artd>|ART|F|P|@>N det Xx True True
folhas folhar NOUN <np-def>|N|F|P|@SUBJ> nsubj xxxx True False
atacadas atacar VERB <mv>|V|PCP|F|P|@ICL-N< acl xxxx True False
ficam ficar VERB <mv>|V|PR|3P|IND|@FS-STA ROOT xxxx True False
com com ADP PRP|@<ADVL case xxx True True
grandes grande ADJ ADJ|F|P|@>N amod xxxx True True
áreas área NOUN <np-idf>|N|F|P|@P< obl xxxx True False
recortadas recortar VERB <mv>|V|PCP|F|P|@ICL-N< acl xxxx True False
ou ou CCONJ <co-fcl>|<co-fmc>|<co-vfin>|KC|@CO cc xx True True
são ser VERB <cjt>|<mv>|V|PR|3P|IND|@FS-STA aux:pass xxx True True
completamente completamente ADV ADV|@ADVL> advmod xxxx True False
consumidas consumir VERB <pass>|<mv>|V|PCP|F|P|@ICL-AUX< conj xxxx True False
. . PUNCT PU|@PU punct . False False
```

Figura 4. Parsing

• **Reconhecimento de Entidades Nomeadas:** spaCy também permite fazer o NER e tem a função de identificar e classificar palavras ou frases em um texto, de acordo com classes definidas para o modelo [Nadeau 2007]. As principais atividades feitas pelo NER são identificar os *tokens* em um texto não estruturado e classificá-los em tipos de entidades definidas de acordo com a peculiaridade do domínio [Speck e Ngomo 2014].

Ao utilizar a função displaCy por meio da ferramenta spaCy é possível destacar visualmente as entidades e os tipos para a frase: *"No Brasil há registros da ocorrência de Percevejo-castanho em várias regiões, embora os danos dessa praga tenham sido mais frequentes nos estados de Mato Grosso, Goiás e Mato Grosso do Sul"*, como é apresentado na Figura 5.

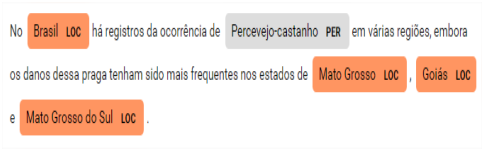


Figura 5.NER com displaCy

• ***Parsing* de dependências:** depois de classificar corretamente todas as palavras de cada sentença do texto por categorias gramaticais é feita a relação de dependência entre palavras.

4. Conclusões

A criação da Ferramenta CAROLINA teve o intuito de otimizar a análise dos dados das pragas e doenças na cultura da soja em um repositório e auxiliar o usuário na tomada de decisão por meio de uma agente conversacional, o que pode facilitar o trabalho dos profissionais da área que precisam se envolver com um amplo volume de informações na tomada de decisões. Pretende-se abranger futuramente outras culturas agrícolas na base de dados, visto que as mesmas pragas prejudicam várias plantações.

Um modelo foi criado utilizando o formato JSON *Schema* para extrair os textos do banco de dados Neo4j. Foi aplicada a função sintática em um grande volume de textos, responsável por organizar as estruturas gramaticais. Assim, foi possível a construção de um modelo por meio da ferramenta spaCy para aplicar regras gramaticais às sentenças e reconhecer as estruturas e extrair seus significados. Além disso, foi avaliado um conjunto de dados reais na identificação das doenças causadas pelas pragas na cultura da soja, a partir das extrações das características do BD. Foram escolhidos três algoritmos de classificação (RF, SVM, KNN) e dentre as validações aplicando as métricas de AM resultou em 99% de acurácia, o que é considerada uma alta taxa de acertos.

Destaca-se que para esse modelo seja realmente efetivo para auxiliar na identificação de pragas e doenças na sojicultura é necessário: analisar outras bases de dados, pois pode causar perda de eficiência treinar somente com uma base de dados estática e, ainda, para comprovar a eficácia do modelo com mais textos; melhorias quanto ao método de seleção de palavras de um documento para termos uma maior assertividade; e avaliar o desempenho de mais classificadores considerando o processamento e tempo de execução.

Ressalta-se ainda que o trabalho está diretamente relacionado aos Objetivos de Desenvolvimento Sustentável (ODS), pois a partir da construção de um sistema inteligente de pré-consulta para a identificação de pragas e doenças é possível promover a agricultura sustentável, o que evita o uso descontrolado de agroquímicos, traz oportunidade de aprendizagem sobre as pragas e doenças na planta analisada, melhora a produtividade e supera os desafios no meio agrícola, como mudanças climáticas, ausência de mão-de-obra, quantidade de terras agricultáveis disponíveis e custos.

Referências

Abonizio, H. Q., Barbosa, C. R. S. C. e Artoni, A. A. (2019) Detecção Automática dos Heterônimos de Fernando Pessoa por Aprendizado de Máquina. *XII Simpósio em Tecnologia da Informação e Linguagem Humana*, Salvador, SBC. p. 144-153

Ávila, C. J. (2017) "Pragas da soja e seu controle". https://pragas.cpao.embrapa.br/, julho.

Bird, S, Klein, E. and Loper, E (2009) *Natural Language Processing with Python.* 1st edition. Cambridge: O'Reilly Media Inc.

Breiman, L. (2001) Random forests. *Machine Learning*, Califórnia, Springer, v.45, n.1, Out. p. 5-32.

CEPEA (2019). "PIB-AGRO/CEPEA: PIB do agronegócio encerra 2019 com alta de 3,81%" https://www.cepea.esalq.usp.br/br/releases/pib-agro-cepea-pib-do-agronegocio-encerra-2019-com-alta-de-3-81.aspx, Dezembro.

EMBRAPA (2019). "Pragas da Soja". https://pragas.cpao.embrapa.br/ , Julho.

Hiraoka, E. e Jacopini, V. (2018) O papel da tecnologia na evolução da agricultura. *Sociedade Nacional de Agricultura*. http://www.sna.agr.br/o-papel-da-tecnologia-na-evolucao-da-agricultura, março.

Indurkhya, N. and Damerau, F. J. (2010) In: *Handbook of Natural Language Processing.* 2nd edition. Chapman & Hall/CRC.

Khanna, A. et al. (2015) A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence. In *International Journal of u-and e-Service, Science and Technology*, v.8, pages 277–284.

Mitchell, T. M. (1997) *Machine learning.* 37ª ed. Burr Ridge, IL: McGraw Hill.

Moraes, M. P. (2019) *Mineração de Dados Aplicada à Identificação de Notícias Falsas.* Departamento de Ciência da Computação da Universidade Federal do Rio de Janeiro, Rio de Janeiro. Trabalho de Conclusão de Curso.

Moreira, H. J. C. e Aragão, F. D. (2009) *Manual de Pragas da Soja.* Campinas: FMC Agricultural Products.

Mostaço, G. M., Campos, L. B., Souza, I. R. C. and Cugnasca, C. E. (2018). AgronoBot: a smart answering Chatbot applied to agricultural sensor networks. *14th International Conference on Precision Agriculture*, p.1-13.

Nadeau, D. (2007) *Semi-supervised named entity recognition: learning to recognize 100 entity types with little supervision.* Institute for Computer Science de Ottawa-Carleton. Ottawa, Canada. Master Thesis.

Neo4J (2018) "Neo4j Basics". https://neo4j.com/ product/#basics, Julho.

Olson, D. L. and Delen, D. (2008) *Advanced data mining techniques.* Berlim: Springer Science & Business Media.

Santos, O. S. (1995) *A Cultura da Soja, 1. Rio Grande do Sul-Santa Catarina-Paraná.* 2ª edição. São Paulo: Globo.

Sawant, D., Jaiswal, A., Singh, J. and Shaw, P. (2019) AgriBot – An Intelligent interactive interface to assist farmers in agricultural activies. In *IEEE Bombay Section Signature Conference*, pages 1-6.

Silva, B. C. D., Montilha, G., Rino, L. H. M., Specia, L., Nunes, M. G. V., Oliveira Junior, O. N., Martins, R. T. e Pardo, T. A. S. (2007) *Introdução ao Processamento das Línguas Naturais e suas Aplicações.* Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional da Universidade de São Paulo. São Carlos.

Silva, C. F. e Vieira, R. (2007) Categorização de Textos da Língua Portuguesa com Árvores de Decisão, SVM e Informações Linguísticas. *V Workshop em Tecnologia da Informação e da Linguagem Humana*, Rio de Janeiro, SBC. p.1650-1658

Sosa-Gomez, D. R. et al. (2010) *Manual de identificação de insetos e outros invertebrados da cultura da soja*. Londrina, PR.: Embrapa Soja. Documentos 269.

Spacy (2020) https://spacy.io/ julho.

Speck, R. and Ngomo, A. C. N. (2014) Ensemble learning for named entity recognition. In *International semantic web conference*, v.8796, n.1, Out., pages 519–534. Switzerland, Springer.

Steinberg, D. and Colla, P. (1995) *CART: Tree-Strutured NonParametric Data Analysis.* San Diego, CA: Salford Systems.

Weinberger, K. Q., Blitzer, J. and Saul, L. K. (2006) Distance metric learning for large margin nearest neighbor classification. In *Advances in Neural Information Processing Systems,* pages 1473-1480. MIT Press.



Proposta de um modelo de processo para aplicação de métodos de interpolação

Gustavo E. Kida[1], Raul T. Santos[1]

[1]Instituto de Computação – Universidade Federal do Mato Grosso (UFMT)
Cuiabá – MT – Brasil

gustavokida@live.com, raul@ic.ufmt.br

Abstract. *In the interpolation of spatial data there are several methods and each one of them have some advantages in specific scenarios. Furthermore, some methods needs treatments to perform a reliable interpolation, such as removing outliers. Hence, this work objective is to present a process model of the application of interpolation methods, serving as a good practice guide. The results demonstrate the efficacy in the maps quality, obtaining the best results when the process model is followed.*

Resumo. *Na interpolação de dados espaciais existem vários métodos e cada um deles possui algumas vantagens em cenários específicos. Além disso, alguns métodos precisam de tratamentos para realizar uma interpolação confiável, como a remoção de outliers e anisotropia. Dessa forma, o objetivo deste trabalho é propor um modelo de processo para aplicação de métodos de interpolação, servindo como um guia de boas práticas. Os resultados demonstram a efetividade dos tratamentos na qualidade dos mapas, obtendo o melhor resultado quando seguido o modelo de processo.*

1. Introdução

A Agricultura de Precisão busca atingir a melhor produtividade de forma sustentável, utilizando a tecnologia como apoio. Entre suas várias etapas é realizada a interpolação de amostras coletadas de uma área, a fim de reduzir o custo da coleta e produzir com isso um mapa de zonas de manejo [Bottega et al. 2017]. Devido a grande quantidade de métodos de interpolação é necessário realizar uma análise dos dados para decidir o método e tratar os dados de acordo com os requisitos do método selecionado [Webster and Oliver 2007; Li and Heap 2008; Adhikary et al. 2017; Li et al. 2015; Pebesma et al. 2011]. O objetivo desse artigo é propor um modelo de processo para aplicação adequada de métodos de interpolação, visando contribuir para obtenção de melhores mapas de

zonas de manejo, e consequentemente melhorando a produtividade do campo, reduzindo o uso de insumos e evitando o desperdício.

2. Metodologia

O modelo proposto foi baseado em uma revisão da literatura de livros e artigos com palavras chaves como “geostatistics”, “spatial interpolation”, “spatial interpolation methods”, “kriging” e “anisotropic inverse distance weighted”, buscando aqueles que detalhavam o processamento dos dados para melhorar os resultados de interpolações espaciais [Webster and Oliver 2007; Li and Heap 2008; Oliver and Webster 2014; Pebesma et al. 2011; Merwade et al. 2006]. Para a especificação do modelo de processo foi utilizado o *Business Process Model and Notation* (BPMN), oferecendo uma notação simples e de fácil entendimento [OMG 2011].

2.1. Construção do modelo

O modelo inicia-se pela remoção de outliers. Essa atividade é primordial para aumentar a precisão, pois os outliers podem gerar efeitos negativos substanciais nas análises estatísticas [Osborne and Overbay 2004]. Após essa etapa o modelo verifica se existem mais de 50 pontos de dados, considerado o mínimo para criar um variograma [Webster and Oliver 2007].

Quando houver menos de 50 pontos será feito a verificação da presença de anisotropia, característica que ocorre quando a distribuição espacial de um elemento varia mais intensamente para uma direção e menos em outra [Camargo et al. 2001]. Esse problema pode ser tratado realizando uma transformação das coordenadas, tornando-os isotrópicos [Boisvert 2010]. Após isso é utilizado o IDQ, método que interpola os dados atribuindo pesos para cada amostra com base na distância em que estas se encontram do ponto a ser interpolado [Alvarenga et al. 2010], e então é revertida a transformação da anisotropia caso tenha ocorrido e finalizado o processo.

Já para quando houver mais de 50 pontos, verifica-se a distribuição normal utilizando o teste de assimetria, considerando assimétricos os valores maiores que 1,0 [Webster and Oliver 2007]. Caso o conjunto de dados seja assimétrico deve ser feita uma transformação para distribuição normal [Oliver and Webster 2014]. A próxima etapa é construir um variograma e verificar se há tendência nele, ou seja, o variograma não deve crescer indefinitivamente. Caso exista tendência serão calculados os resíduos e verificado a presença de anisotropia nos resíduos, se houver anisotropia é realizada a remoção e em seguida a interpolação com o método Krigagem Universal (KU), que tem como característica conseguir tratar a tendência [Webster and Oliver 2007]. Após realizada a interpolação nos dados transformados é necessário reverter a transformação de assimetria [Webster and Oliver 2007].

Não existindo tendência, a anisotropia é analisada da mesma forma que nos casos anteriores, se não for encontrada então será gerado um variograma para analisar a presença do efeito pepita puro, que se caracteriza por um variograma em que o efeito pepita tem o mesmo valor do patamar [Webster and Oliver 2007; Oliver and Webster 2014]. Se não for detectado esse efeito será realizada a KO, a qual leva em conta a correlação espacial dos dados ao utilizar um variograma para determinar os pesos dos pontos vizinhos ao ponto que será estimado [Requia et al. 2019], então é revertida a transformação de assimetria e finalizado o processo.

Caso seja encontrada a anisotropia é preciso verificar se existem ao menos 250 pontos de dados, pois é a quantidade de amostras recomendadas para realizar o tratamento do variograma [Webster and Oliver 2007]. Se não houver 250 pontos, será desfeita a transformação de assimetria se foi realizada e o processo seguirá o caminho de testar a anisotropia e utilizar o método IDQ. Se existirem pontos suficientes será feita a remoção da anisotropia e então realizado a KO, finalizando o processo após reverter a transformação de assimetria se foi feita. Por fim, para ambos os casos em que o efeito pepita puro for detectado, será desfeita a transformação de assimetria e então o processo seguirá o caminho de testar a anisotropia e interpolar utilizando o IDQ.

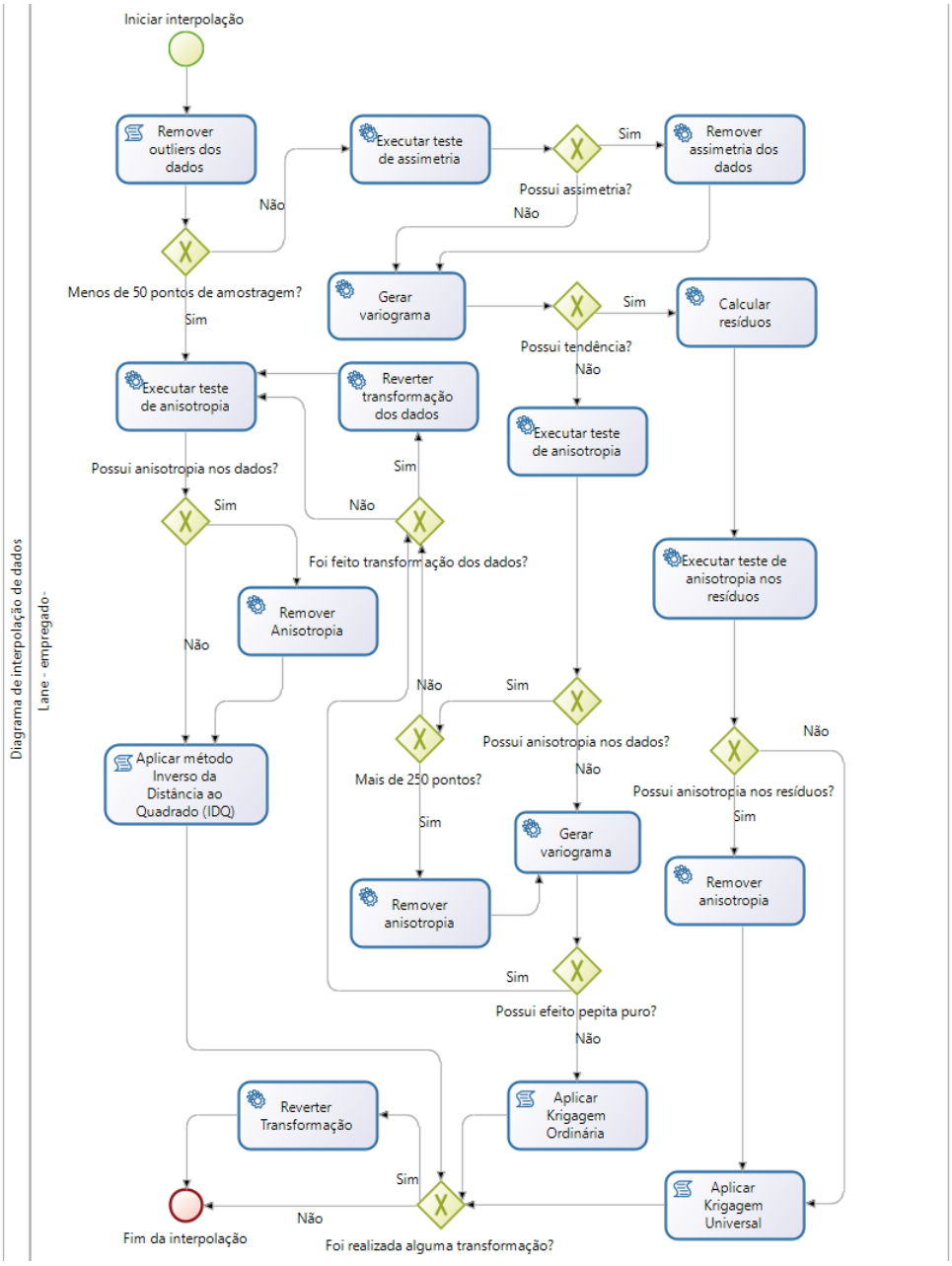


Figura 1. Modelo de processo para aplicação de métodos de interpolação

O modelo está representado pela Figura 1, seguindo todas as etapas que foram explicadas.

2.2. Validação

O modelo foi validado utilizando validação cruzada pelo método *leave-one-out*, aplicando o *Root Mean Square Error* (RMSE) para verificar a acurácia, sendo que quanto mais próximo de 0, melhor a precisão. Os testes foram feitos seguindo o modelo proposto na Figura 1 para interpolar o conjunto de dados disponíveis para o projeto de pesquisa e comparados com os resultados dos testes que não seguem o modelo, utilizando o software R [Team 2013].

A validação utilizou um conjunto de dados de solo de um talhão de 133,96 ha, coletado próximo ao município de Tibagi no estado do Paraná, possuindo 62 pontos com distanciamento de 165 metros aproximadamente. Os atributos Acidez potencial (H+Al) e Fósforo (P) foram selecionados para a validação.

3. Estudo de caso

No atributo H+Al a remoção de outliers removeu 4 pontos, utilizando o método boxplot por ser mais estável [Bento and Santos 2018]. Já a assimetria foi de 1,904404 para o conjunto sem tratamentos e 0,8529702 para o conjunto com tratamentos, não necessitando transformar a assimetria. A anisotropia foi testada e não foi encontrada nos dados tratados.

Dessa forma, o caminho correto do modelo de processo é realizado pelo tratamento de outliers, verificação da assimetria, tendência no variograma, anisotropia e efeito pepita puro. Por fim os dados são interpolados com a KO.

Após a interpolação foi realizada a validação cruzada e foram obtidos valores de RMSE do atributo H+Al conforme a Tabela 1. A precisão melhorou 38,8% em relação ao IDQ sem tratamentos e 32,4% quando comparado com a KO sem tratamentos.

Tabela 1. Resultados RMSE da interpolação do atributo H+Al

Sem tratamento (KO)	11,10533
Sem tratamento (IDQ)	12,27034
Seguindo o modelo	7,516667 (KO)

A Figura 2 mostra os variogramas finais utilizados para a krigagem do atributo H+Al, e todos apresentaram um variograma aceitável, sem tendência ou efeito pepita [Oliver and Webster 2014].

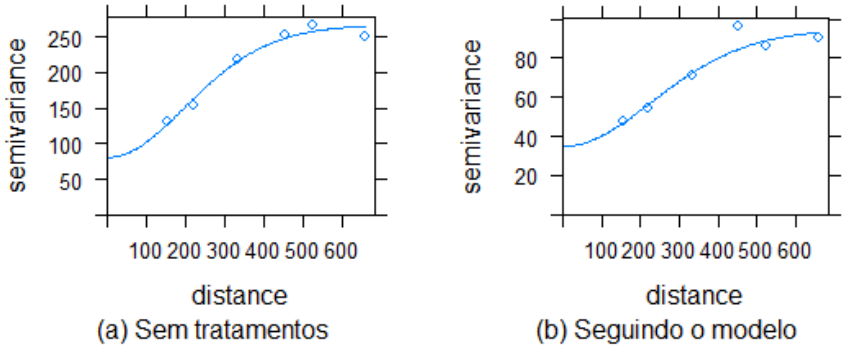


Figura 2. Variogramas do atributo H+Al

A Figura 3 apresenta os mapas interpolados do atributo H+Al, sendo que a Figura 3c foi considerada a melhor pela validação.

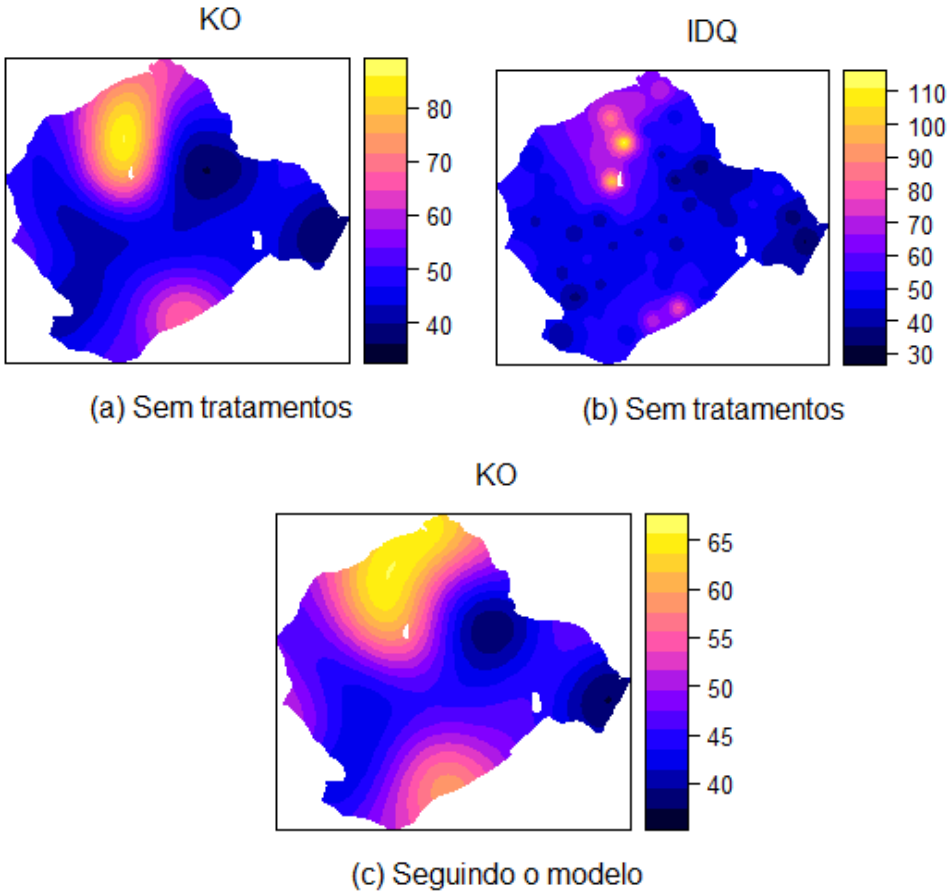


Figura 3. Mapas da interpolação do atributo H+Al

Para o atributo P, a remoção de outliers resultou em apenas 1 ponto removido, sobrando 61 pontos. A assimetria do atributo P foi de 0,764862 para o conjunto sem outliers, sendo assim, não houve a necessidade de uma transformação Box-Cox. O

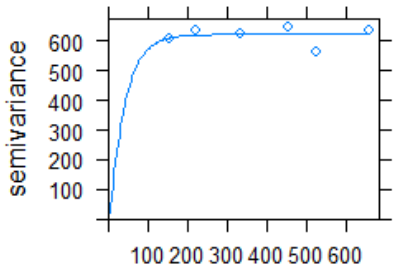
sofware encontrou anisotropia nos dados sem outliers e retornou os valores para o ângulo de -38,785155 e 1,721492 para razão. Assim, o modelo realizou a remoção de outliers, não encontrou assimetria e nem tendência, mas detectou a anisotropia. Como não há mais de 250 pontos a KO não foi aplicada, por isso foi feito a verificação de anisotropia no caminho para o IDQ, removeu a anisotropia detectada, realizou a interpolação com o método IDQ e reverteu as transformações.

Feito os testes, foram observados e comparados os valores RMSE obtidos pela validação cruzada apresentados na Tabela 2, demonstrando um aumento na acurácia seguindo o modelo de 12,1% em relação ao método KO e 9,5% se comparado ao IDQ, ambos os métodos sem tratamentos.

Tabela 2. Resultados RMSE da interpolação do atributo P

Sem tratamento (KO)	25,9916
Sem tratamento (IDQ)	25,23343
Seguindo o modelo	22,8612 (IDQ)

A Figura 4 mostra o variograma gerado para ser utilizado na KO do atributo P e apresentou um patamar estável.



(a) Sem tratamentos

Figura 4. Variograma do atributo P

A Figura 5 mostra os mapas gerados pela interpolação de cada teste. Dentre os 3 testes, o melhor resultado é apresentado na Figura 5c que seguiu o modelo, pois obteve o menor valor de RMSE.

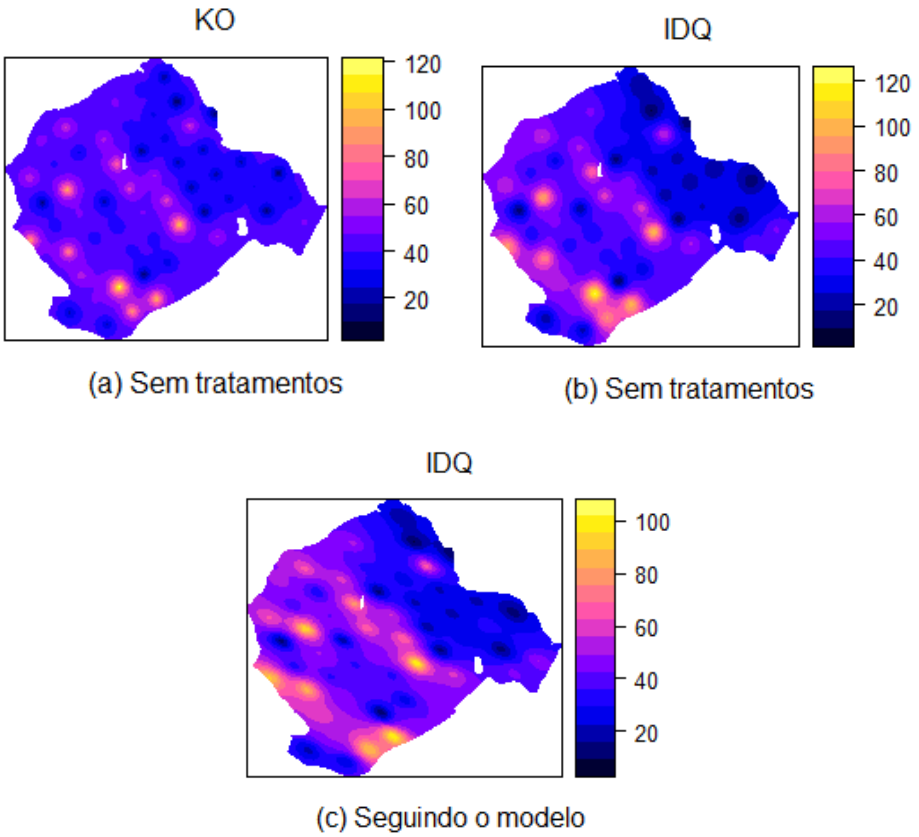


Figura 5. Mapas da interpolação do atributo P

4. Conclusões e trabalhos futuros

Os resultados obtidos seguindo o modelo foram melhores para ambos os atributos H+Al e P, demonstrando a aplicabilidade na área de ciências da terra, como na gestão ambiental que geralmente coletam pontos de uma área, mas necessitam de um mapa de superfície contínuo [Li and Heap 2008], na criação de mapas de zonas de manejo para reduzir a variância dos dados [Schenatto et al. 2016] e na redução de custos de amostragem de acordo com a estrutura espacial [Li and Heap 2008]. No atributo H+Al a melhora foi significativa utilizando apenas a remoção de outliers, já no atributo P houve um pequeno aumento na precisão com os tratamentos de anisotropia e outliers.

Os resultados também evidenciam que a remoção de outliers sempre melhora os resultados da interpolação, independente do atributo ou do método utilizado, além disso, no atributo H+Al a remoção de outliers removeu a assimetria como consequência, podendo assim reduzir a complexidade de todo o processo de interpolação e consequentemente contribuir para a redução de erros humanos.

Dessa forma, os resultados permitem concluir que o modelo consegue reduzir erros e escolher o melhor método de interpolação e seus tratamentos necessários, portanto é recomendado seguir o modelo de processo, servindo como um guia de boas práticas para a interpolação de mapas, permitindo reduzir os custos de amostragem de solo e interpolar mapas com a melhor representação possível da área.

Como trabalho futuro pretende-se validar as etapas que não foram utilizadas, buscando dados com tendência para testar o caminho do processo que utiliza o método KU. Além disso, será buscado melhorar o modelo como a inclusão de outros métodos de interpolação.

5. Agradecimentos

Os autores agradecem à FAPEMAT pelo apoio financeiro ao projeto e pela concessão de bolsa de iniciação científica através da cooperação técnica com a UFMT.

Referências

Alvarenga, L. H. V., Pinto, A. L. R., Silva, S. T., Altoé, T. F., Morais, V. A., and MELLO, J. (2010). Comparação de procedimentos de amostragem na precisão de Inventário Florestal em fragmento de Floresta Estacional Semidecidual. In *XIX Congresso de Pós-graduação da UFLA*.

Adhikary, S. K., Muttil, N., and Yilmaz, A. G. (2017). Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological processes*, *31*(12), 2143-2161.

Bento, G. M., and Santos, R. T. (2018) "Avaliação de Métodos de remoção de outliers e seus impactos na precisão dos métodos de interpolação", Simpósio Mato-grossense de Mecanização Agrícola e Agricultura de Precisão.

Boisvert, J. (2010). Geostatistics with locally varying anisotropy.

Bottega, E. L., de Queiroz, D. M., de Assis de Carvalho Pinto, F., de Souza, C. M. A., and Valente, D. S. M. (2017). Precision agriculture applied to soybean: Part I-Delineation of management zones. *Australian Journal of Crop Science*, *11*(5), 573-579.

Camargo, E. C. G., Felgueiras, C. A., and Monteiro, A. M. V. (2001). A importância da modelagem da anisotropia na distribuição espacial de variáveis ambientais utilizando procedimentos geoestatísticos. *Simpósio Brasileiro de Sensoriamento Remoto*, *10*, 395-402.

Li, J., Duan, P., Sheng, Y., & Lv, H. (2015, December). Spatial interpolation approach based on IDW with anisotropic spatial structures. In *International Conference on Intelligent Earth Observing and Applications 2015* (Vol. 9808, p. 98081O). International Society for Optics and Photonics.

Li, J., and Heap, A. D. (2008). A review of spatial interpolation methods for environmental scientists.

Merwade, V. M., Maidment, D. R., & Goff, J. A. (2006). Anisotropic considerations while interpolating river channel bathymetry. *Journal of Hydrology*, *331*(3-4), 731-741.

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, *113*, 56-69.

OMG (2011). Business Process Model and Notation (BPMN), Version 2.0 Object Management Group (Technical report, Object Management Group).

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, *9*(1), 6.

Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G. B., Hristopulos, D., Pilz, J., ... & Skøien, J. O. (2011). INTAMAP: the design and implementation of an interoperable automated interpolation web service. *Computers & Geosciences*, *37*(3), 343-352.

Requia, W. J., Coull, B. A., and Koutrakis, P. (2019). Evaluation of predictive capabilities of ordinary geostatistical interpolation, hybrid interpolation, and machine learning methods for estimating PM2. 5 constituents over space. *Environmental research*, *175*, 421-433.

Schenatto, K., Souza, E. G., Bazzi, C. L., Bier, V. A., Betzek, N. M., & Gavioli, A. (2016). Data interpolation in the definition of management zones. *Acta Scientiarum. Technology*, *38*(1), 31-40.

Team, R. C. (2013). R: A language and environment for statistical computing.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.



Semantic Sensor Network Ontology como modelo de referência em soluções de monitoramento agrícola

Julio C. Nardi[1], Vitor N. Alves[1], Carlos A. R. Junior[1], Fabiano B. Ruy[2], Robson Prucoli Posse[3]

[1] Instituto Federal do Espírito Santo (Ifes) – Campus Colatina
Av. Arino Gomes Leal, nº 1700, Santa Margarida, Colatina-ES, Brasil

[2] Instituto Federal do Espírito Santo (Ifes) – Campus Serra
Rodovia ES-010, Km 6,5 – Manguinhos, Serra-ES, Brasil

[3] Instituto Federal do Espírito Santo (Ifes) – Campus Itapina
Rodovia BR-259, Km 70, Colatina-ES, Brasil

julionardi@ifes.edu.br, vitornicchioalves@gmail.com,
junior_recla@hotmail.com, {fabianoruy, robson.posse}@ifes.edu.br

Abstract. *Solutions of Internet-of-Things in Agriculture deal with devices and software applications that do not share the same information model. This limits the semantic interoperability between them. This work presents and discusses the use of SSN ontology (W3C) in two initiatives related to monitoring of environmental variables. Thus, it aims at supporting data integration of devices, and development of software applications more interoperable, in so far they use a conceptual reference model.*

Resumo. *Soluções de Internet das Coisas na Agricultura lidam com dispositivos e aplicações de software que não compartilham um mesmo modelo de informação. Isso limita a interoperabilidade semântica entre eles. Este trabalho apresenta e discute o uso da ontologia SSN (W3C) em duas iniciativas relacionadas ao monitoramento de variáveis ambientais. Como resultado, busca-se promover integração de dados de dispositivos e desenvolvimento de aplicações de software mais interoperáveis, na medida em que se utiliza um modelo conceitual de referência.*

1. Introdução

As áreas de Agricultura de Precisão (AP) e Internet das Coisas (*Internet of Things* - IoT) têm se aproximado cada vez mais na medida em que o monitoramento automático de eventos e variáveis ambientais tem levado ao aumento da produção agrícola e da redução do consumo de recursos. Por meio de dispositivos interconectados e associados a aplicações de softwares é possível capturar, manipular e compartilhar dados reais de condições de solo, clima e outros para otimizar a produção [Shafi et al., 2019].

Esse contexto se caracteriza pelo uso de diversos dispositivos e aplicações "heterogêneos", desenvolvidos por diferentes fabricantes e sem compartilharem um

mesmo modelo de informação. Como consequência, um dos grandes desafios, reconhecidos tanto pela Indústria [Ericsson Blog, 2021] quanto pela Academia [Barnaghi et al., 2012] [IERC, 2015] [Honti and Abonyi, 2019], é a promoção de "interoperabilidade semântica" entre dispositivos, entre aplicações e entre dispositivos e aplicações em soluções de IoT. Para além da integração em nível técnico (de hardware e de plataforma), por meio da promoção da interoperabilidade semântica busca-se alinhar o significado dos dados trocados entre dispositivos e aplicações para que haja o correto entendimento desses dados, a devida comunicação e o uso efetivo.

Modelos conceituais de referência e padrões têm sido desenvolvidos e adotados [Honti and Abonyi, 2019], tais como: IoT-A Reference Model [Bauer et al., 2013], The Internet of Things Reference Model [CISCO, 2014] e IoT-Lite Ontology [W3C, 2005]. Tais modelos/padrões buscam estabelecer um referencial conceitual e uma interlíngua a ser adotada na integração de dispositivos e aplicações de software. A ontologia SSN (*Semantic Sensor Network*) [W3C, 2017] é um padrão internacional amplamente utilizado cujo objetivo é descrever o domínio de sensoriamento abordando conceitos como sensores, suas observações, os procedimentos envolvidos, as propriedades observadas, os atuadores e suas ações etc. Ontologias têm sido reconhecidas como um importante artefato para enriquecer os dados sensoriados e promover interoperabilidade semântica [Honti and Abonyi, 2019]. Assim, SSN tem sido aplicada em diferentes cenários aqueles relacionados à Agricultura de Precisão [W3C, 2017] [Li et al., 2018].

Entretanto, o uso de ontologias para promoção de interoperabilidade semântica, seja na construções de soluções de software ou na integração de dados requer sistematicidade. Abordagens que tratam da derivação/transformação de ontologias em esquemas de banco de dados [Guidoni et al., 2020] e da integração de dados de aplicações e/ou dispositivos têm sido desenvolvidas [Honti and Abonyi, 2019], mas ainda se tem que avançar no entendimento e na adoção de abordagens dessa natureza.

Este trabalho, portanto, apresenta e discute o uso de SSN em duas iniciativas relacionadas ao monitoramento de variáveis ambientais: (i) desenvolvimento de aplicação de software para monitoramento *online* e (ii) realização de integração de dados sensoriados advindos de arquivos (.XLS). São discutidos aspectos relacionados: (i) a derivação de modelo de classes a partir da ontologia SSN e (ii) a alinhamento semântico entre estruturas de dados geradas por dispositivos heterogêneos e um modelo de referência (SSN) visando a integração de dados. A partir dos resultados deste trabalho, espera-se contribuir para o desenvolvimento de outras iniciativas correlatas, em especial, aquelas baseadas em ontologias, uma vez que pesquisas atuais evidenciam o uso de ontologias como um caminho promissor no sentido de obter soluções de IoT mais adaptáveis e interoperáveis [Honti and Abonyi, 2019].

2. Referencial Teórico

Interoperabilidade semântica no âmbito de soluções de IoT tem sido foco de vários trabalhos [Honti and Abonyi, 2019] [Drury et al., 2019]. Tais soluções, em geral, fazem uso de inúmeros sensores, formando uma rede de dispositivos que apresentam capacidades complementares e especificações distintas, requerendo abordagens de integração no nível semântico para alinhar a troca de dados entre eles [Li et al., 2018]

Nesse contexto, a interoperabilidade semântica se coloca como um dos grandes desafios em soluções IoT, na medida em que o enriquecimento semântico de dados sensoriados brutos (*raw sensory data*) é crucial para suportar integração de informação, que levará à produção de conhecimento e à inovação [Honti and Abonyi, 2019].

Também no domínio de Agricultura de Precisão, abordagens e tecnologias semânticas de várias naturezas tem sido desenvolvidas [Honti and Abonyi, 2019] [Drury et al., 2019], dentre as quais estão aquelas relacionadas à integração e compartilhamento de dados sensoriados. Nesse contexto, a (re)utilização de ontologias e

modelos de referência como “recursos semânticos” se destaca [Drury et al., 2019]. Mais especificamente, ontologias se apresentam como modelos conceituais que buscam, dentre outros, minimizar problemas de comunicação (p.ex., ambiguidade) entre dispositivos, sistemas, pessoas e organizações [Guizzardi, 2007].

A ontologia SSN (*Semantic Sensor Network*), como um padrão internacional, é amplamente utilizada. Como tal, busca formalizar, explicitamente, conceitos e relações de um determinado domínio (sensoriamento), que são compartilhados (“consensuais”) dentro de uma comunidade [Gruber, 1993].

A Figura 1 apresenta um fragmento de SSN cujo foco é na descrição de conceitos relacionados à observação de uma determinada propriedade. O fragmento descreve, em linhas gerais, que sensores (*Sensor*) fazem observações (*Observation*), que estão relacionadas a propriedades observáveis (*ObservableProperty*) e utilizam determinados procedimentos (*Procedure*). Além da representação gráfica (ilustrada pela Figura 1), SSN conta com uma formalização em OWL (*Ontology Web Language*), permitindo também sua utilização/processamento por máquinas.

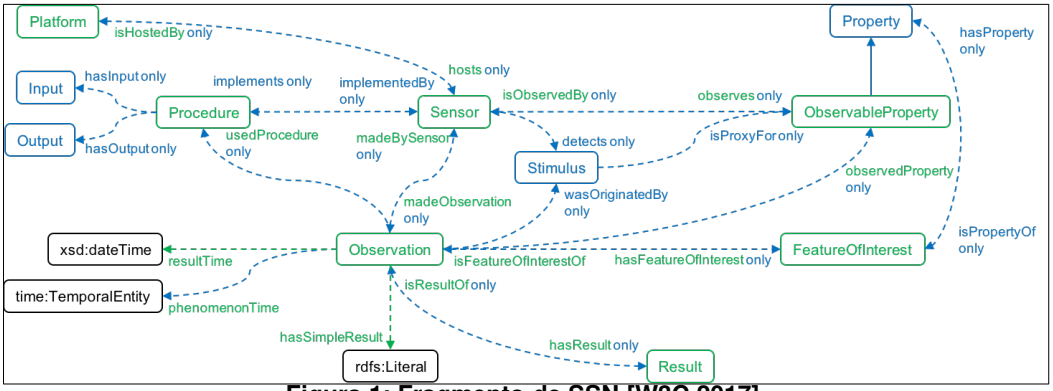


Figura 1: Fragmento de SSN [W3C 2017].

Ontologias, como no caso de SSN, têm sido utilizadas para suportar o desenvolvimento e à integração de sistemas de softwares e soluções IoT, que necessitam lidar com um determinado domínio [Abinaya et al., 2015] [Nardi et al., 2013] [Honti and Abonyi, 2019]. Dessa forma, o conhecimento compartilhado e explicitado em uma ontologia é reutilizado na medida em que é incorporado nos sistemas de software, p.ex., por meio de transformação de modelos [Guidoni et al., 2020] dando origem a modelos de classes, modelos relacionais, regras de negócio etc. Como resultado, espera-se que tais soluções tenham um maior grau de reuso e interoperabilidade, por fazerem uso de dados/informações alinhados a modelos conceituais “consensuais”.

3. Da Ontologia SSN para um Modelo de Informação

Com vistas à utilização do conhecimento formalizado em SSN como referência na construção de soluções de software, o primeiro passo foi a derivação de um modelo de classes usando a UML (*Unified Modeling Language*), conforme ilustra a Figura 2.

O processo de derivação é iterativo, incorporando, gradualmente, os conceitos, relações e restrições da ontologia no modelo de classes. Pode-se entender esse processo como uma espécie de "transformação de modelos" [Guidoni et al., 2020], na medida em que o modelo ontológico de SSN é transformado em um modelo de classes UML. Ao se realizar a transformação, construtos da linguagem de representação do modelo de origem são mapeados em construtos da linguagem de representação do modelo de destino. Assim, conceitos, relações e restrições (da ontologia) são representados no modelo resultante (de classes) acomodando-se de acordo com a expressividade da linguagem utilizada. Ademais, como derivou-se um modelo de classes visando a

construção de soluções de software, alguns aspectos de *design* de software foram inseridos (p.ex., tipos e navegabilidades). A Figura 3 apresenta um fragmento do modelo de classes gerado.

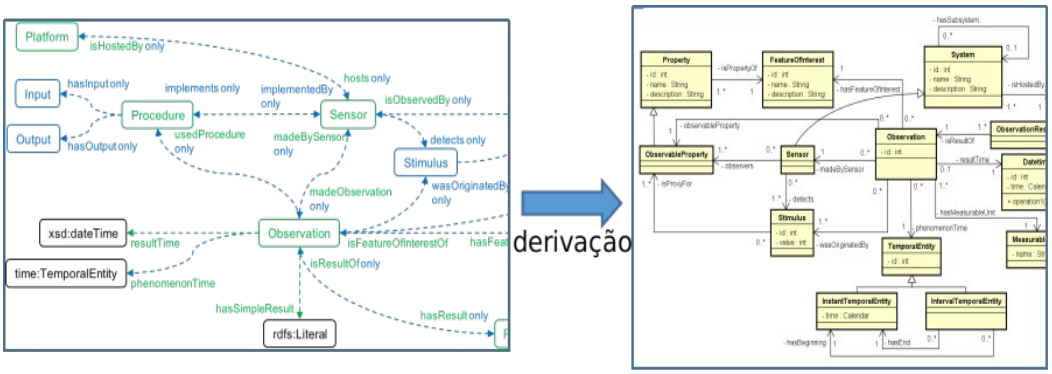


Figura 2: Ilustração da derivação de SSN em um modelo de classes.

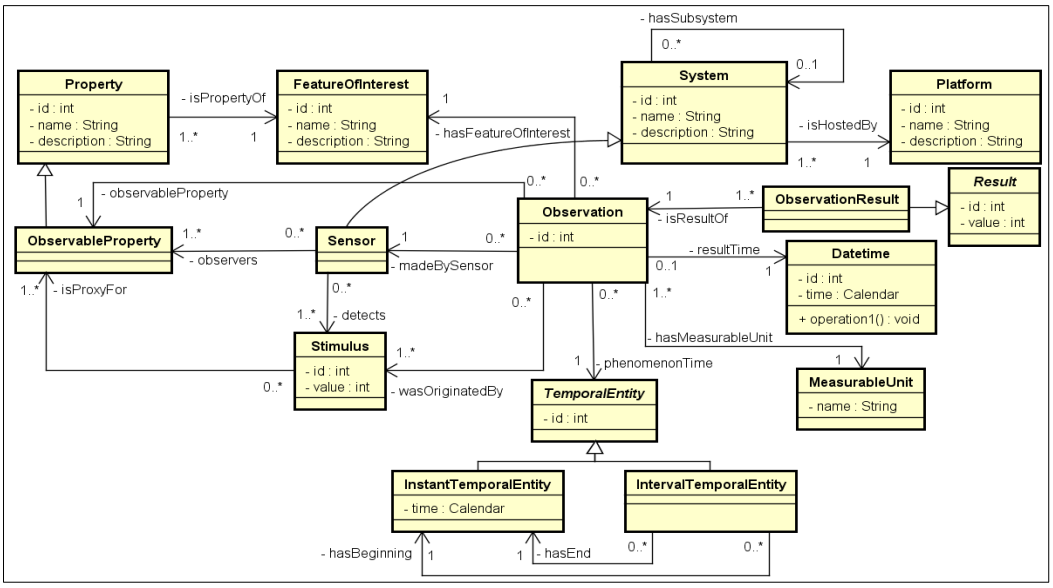


Figura 3: Fragmento do modelo de classes gerado a partir de SSN.

De acordo com o modelo, uma *Platform* (p.ex., um viveiro de mudas) hospeda/contem vários *System*. *System* é um conceito especializado em *Sensor*. Um *Sensor* (um sensor de umidade) observa uma ou várias *ObservableProperty* (umidade). Uma *Observation* (leitura da umidade do solo em um dado momento t1) é feita por um *Sensor*, sendo originada por um *Stimulus* (acionamento de um temporizador), que é detectado pelo *Sensor*. Uma *Observation* pode acontecer instantaneamente ou ao longo de um intervalo temporal, gera *ObservationResults* (20, valor sensoriado de umidade). *ObservableProperty* é uma especialização do conceito *Property*, que, por sua vez, é uma propriedade de uma *FeatureOfInterest* (solo do viveiro de mudas).

3.1. Iniciativa 1: SSN Aplicada no Desenvolvimento de Aplicação de Software

Nesta iniciativa foi construída uma solução de software para monitoramento de irrigação usando sensores capturam dados de variáveis de ambiente como temperatura do ar e umidade do solo e os transmitem para um *middleware* que, por meio de serviços de software, os envia a uma aplicação web que realiza o armazenamento. De posse

desses dados, o usuário pode realizar o monitoramento remoto das variáveis ambientais usando aplicações específicas em diferentes plataformas (p.ex., computador e celular).

Conforme apresenta a Figura 4, o sistema foi estruturado em 4 (quatro) camadas: a *Camada de Sensoriamento* (onde os sensores se encontram localizados no ambiente monitorado), a *Camada de Conexão* (possui um *middleware* rodando em um Raspberry Pi, que estabelece a conexão entre os sensores e a aplicação de software), a *Camada de Gerenciamento* (onde se encontra a aplicação de software (JavaEE) juntamente com o banco de dados (MySQL)) e a *Camada de Aplicação do Usuário* (onde o usuário realiza o monitoramento por meio de aplicações de software – *desktop* e *mobile*). Todo o *setup* da solução como, por exemplo, cadastros dos sensores, plataforma, medidas e propriedades observadas são realizados por um usuário administrador. Em tempo real, os dados sensoriados são lidos e enviados, via serviços web, à aplicação, a qual os analisa, identifica a origem, o que representam e, assim, os armazena na base de dados.

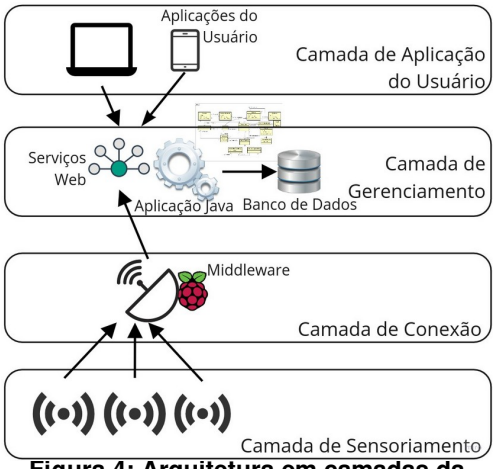


Figura 4: Arquitetura em camadas da solução proposta.

Para implementação das classes de domínio, dos serviços *web* (REST) e da estrutura da base de dados, utilizou-se o modelo de classes derivado de SSN (vide Figura 3), fornecendo, dessa forma, a referência conceitual provida pela ontologia. Assim, o banco de dados armazena tanto os dados sensoriados quanto as configurações de *setup* da infraestrutura de sensoriamento como, por exemplo, marca/modelo dos dispositivos utilizados. A Figura 5 ilustra a implementação da classe *Sensor* na aplicação de software baseada em SSN.

```
1  package core;
2
3  import java.util.ArrayList;
4  import java.util.List;
5  import java.util.Set;
6  import javax.persistence.Entity;
7  import javax.persistence.FetchType;
8  import javax.persistence.Inheritance;
9  import javax.persistence.InheritanceType;
10 import javax.persistence.ManyToMany;
11
12 @Entity
13 @Inheritance(strategy = InheritanceType.TABLE_PER_CLASS)
   public class Sensor extends System {
15
16     @ManyToMany(fetch = FetchType.EAGER)
17     private List<ObservableProperty> observableProperty;
18
19     @ManyToMany(fetch = FetchType.EAGER)
20     private Set<Stimulus> stimulus;
```

Figura 5: Exemplo de implementação baseada em SSN. Classe Sensor.

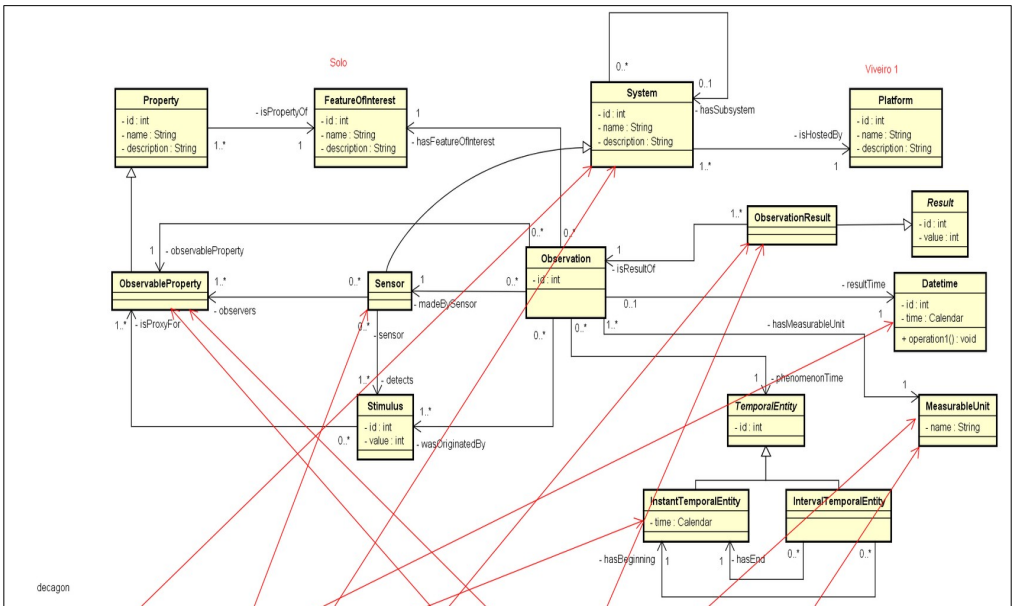
A classe *Sensor*, que herda de *System,* tem relação com duas outras classes

implementadas: *ObservableProperty* e *Stimulus*. Assim, pode-se observar que a implementação em código fonte (Java) mapeia os mesmos conceitos e relações do modelo de classe, que, por sua vez, foram derivados de SSN.

3.2. Iniciativa 2: SSN Aplicada na Integração de Dados de Diferentes Dispositivos

Nesta iniciativa, usou-se o modelo de classes gerado a partir de SSN para criar uma aplicação que realizasse a leitura de dados fornecidos em arquivos .XLS gerados a partir de dois dispositivos utilizados no monitoramento de condições climáticas/ambientais - (i) uma estação climática HOBO® U30 Station; e (ii) um dispositivo Decagon® EM43375 - e inserisse esses dados em um banco de dados único.

O processo de construção da solução consistiu, inicialmente, no entendimento das estruturas e dados dos arquivos gerados por cada dispositivo. Em seguida, passou-se para a construção do alinhamento semântico entre os dados e estrutura de cada arquivo e o modelo de classes gerado à luz de SSN. Isso se faz necessário na medida em que os dados dos arquivos são importados para uma base de dados única, cuja estrutura é baseada no modelo de classes de SSN. Deve haver uma correspondência entre cada dado contido nos arquivos e as estruturas de informação (classes e tabelas) correspondentes da solução. O modelo de classes baseado em SSN foi utilizado como modelo de referência para o alinhamento semântico. As figuras 6 e 7 ilustram os mapeamentos entre os dados e estruturas de cada arquivo e o modelo de classes.

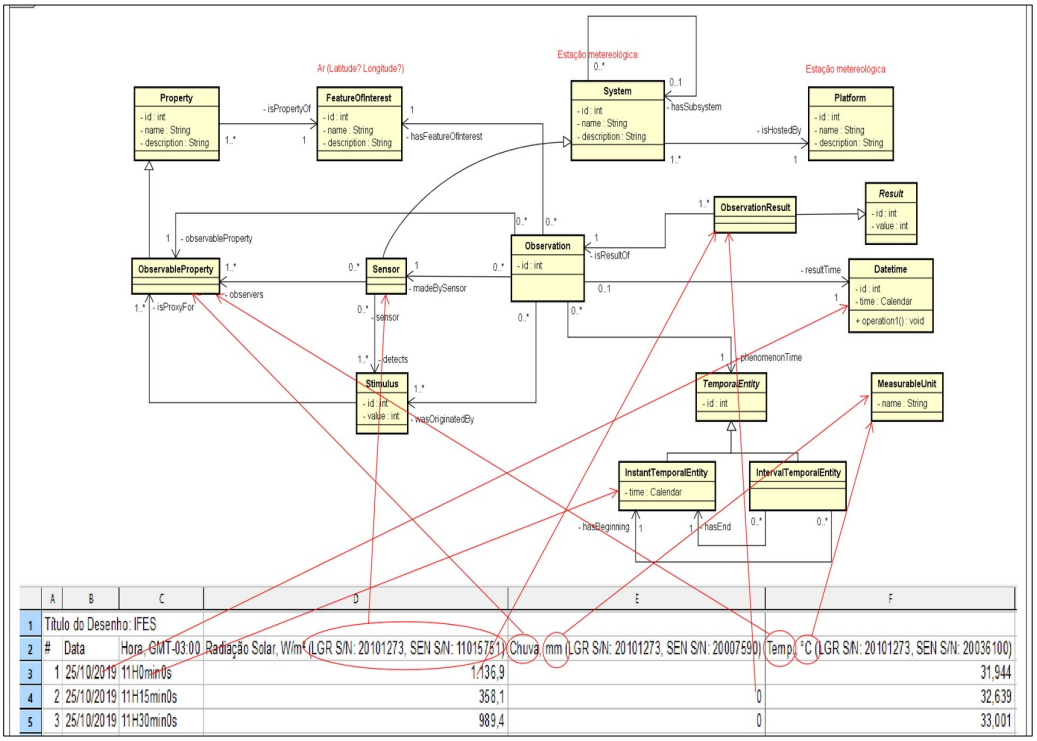


	A	B	C	D	E	F
1	EM43375	Port 1	Port 1	Port 2	Port 2	Port 2
2	2015 records	MPS-6 Water Potential/Temp	MPS-6 Water Potential/Temp	5TE Moisture/Temp/EC	5TE Moisture/Temp/EC	5TE Moisture/T
3	Measurement Time	kPa Potential	°C Temp	m³/m³ VWC	°C Temp	mS/cm EC Bull
4	04/06/2018 14:50	-13,6	27,2	0,305	27,4	
5	04/06/2018 14:55	-13,6	27,1	0,305	27,3	

Figura 6: Mapeamentos conceituais entre a planilha de dados do Decagon® e o modelo de classes.

Após a definição dos mapeamentos, a aplicação de software foi implementada. As classes de domínio (em Java) e a estrutura da base de dados relacional (MySQL) refletiram o modelo de classes baseado em SSN. Em seguida, foi implementado o módulo de importação, que tem a função ler cada arquivo de dados e, com base nos mapeamentos, realizar a persistência. Assim, ao final da importação, todos os dados

foram inseridos na base de dados, considerando, agora, o mesmo modelo de informação fornecido pela ontologia. Isso, tende a facilitar consultas e análises dos dados, então, integrados. A Figura 8 ilustra a arquitetura de alto nível da solução.



	A	B	C	D	E	F
1	Título do Desenho: IFES					
2	#	Data	Hora, GMT-03:00	Radiação Solar, W/m² (LGR S/N: 20101273, SEN S/N: 11015781)	Chuva (mm (LGR S/N: 20101273, SEN S/N: 20007590)	Temp. °C (LGR S/N: 20101273, SEN S/N: 20036100)
3	1	25/10/2019	11H0min0s	1.136,9		31,944
4	2	25/10/2019	11H15min0s	358,1	0	32,639
5	3	25/10/2019	11H30min0s	989,4	0	33,001

Figura 7: Mapeamentos conceituais entre a planilha de dados da estação e o modelo de classes.

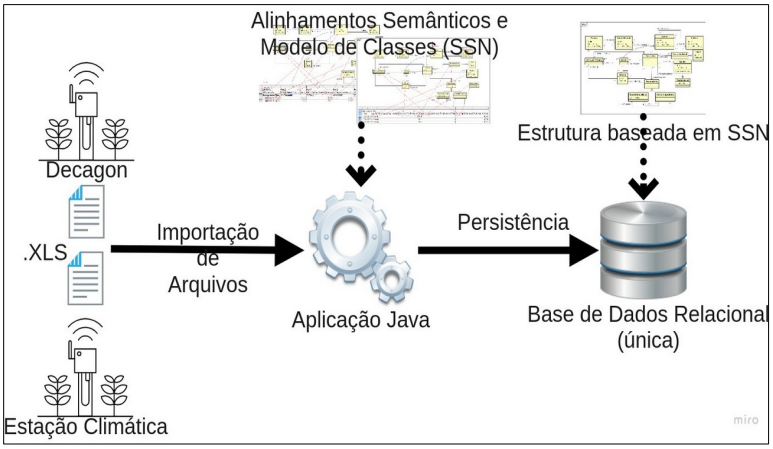


Figura 8: Arquitetura de alto nível da solução.

4. Discussão

Considerando o processo de derivação do modelo de classes a partir de SSN e o desenvolvimento das 2 (duas) iniciativas, alguns aspectos merecem destaque.

Aspecto 1. A ontologia SSN apresenta, basicamente, 3 (três) perspectivas: *Observation*,

Actuation e *Sampling*. As iniciativas apresentadas neste trabalho focaram na perspectiva *Observation*, devido ao seus requisitos. De todo modo, uma vez que SSN estabelece uma estrutura padrão entre suas perspectivas, acredita-se que avançar na ampliação do modelo de classes derivado rumo às duas outras perpectivas (*Actuation* e *Sampling*) seja análogo ao processo apresentado neste trabalho.

Aspecto 2. Ao derivar o modelo de classes, foram incluídos atributos mínimos (p.ex., IDs, nome e descrição), para dar suporte ao aspecto informacional do modelo. Ademais, foram definidas navegabilidades, para atender a requisitos não funcionais como, por exemplo, economia de memória. De todo modo, tais alterações não descaracterizam SSN. No caso de atributos, entende-se que, havendo necessidade, pode-se especializar o modelo incluindo novas classes que venham a acrescentar eventuais atributos/relações. Com a especialização, é possível manter a conceitualização fundamental baseada em SSN e, mesmo assim, adequar o modelo a cenários específicos.

Aspecto 3. Em relação ao conceito *TemporalEntity*, existente em SSN, esse foi especializado no modelo de classes gerado em dois outros conceitos: *InstantTemporalEntity* e *IntervalTemporalEntity*. Tal ajuste foi julgado necessário para melhor descrever os possíveis tipos de momentos em que um fenômeno pode ocorrer, ou seja: instantaneamente ou durante um intervalo de tempo. De todo modo, a superclasse *TemporalEntity*, mantém relação direta com conceito original de SSN.

Aspecto 4. A tarefa de mapeamento entre cada arquivo de dados gerado por um modelo/marca de dispositivo e o modelo de classes utilizado precisa ser realizada unicamente. Isso é feito pelo modelador/desenvolvedor com base na especificação técnica do dispositivo e com o apoio de um especialista de domínio (p.ex., um agrônomo que conhece os dados sensoriados). Essa tarefa, quando realizada para uma primeira solução, consome tempo. Depois, entretanto, o mapeamento poderá ser reutilizado em outras iniciativas que vierem a fazer uso do mesmo modelo/marca de dispositivo e mesmo modelo de classes. Daí a importância de se utilizar um modelo de referência. Pode-se pensar em uma abordagem semelhante ao uso *drivers* (dispositivo – *driver* – plataforma). Cada dispositivo terá o seu *driver* de importação/mapeamento dos dados sensoriados para uma base de dados padrão/referência, otimizando seu uso.

5. Conclusões

Este trabalho discute o uso de SSN em duas iniciativas relacionadas ao monitoramento de variáveis ambientais na Agricultura: (i) desenvolvimento de aplicação de software para monitoramento *online* e (ii) integração de dados sensoriados advindos de arquivos (.XLS) em uma base de dados única. A despeito das particularidades de cada iniciativa, há uma abordagem comum cujas etapas merecem destaque: (i) recepção do dado sensoriado; (ii) caracterização do dado à luz de um modelo conceitual de referência (SSN); e (iii) uso e armazenamento do dado em uma base de dados compartilhada.

Dentre as contribuições do trabalho destacam-se: (i) derivação da ontologia SSN em um modelo de classes aplicado nas iniciativas apresentadas e que pode ser reutilizado em outras iniciativas correlatas; (ii) estratégia para mapeamento de dados de diferentes fontes de sensoriamento para um modelo comum (de referência); e (iii) discussão e criação de bases de dados alinhadas a uma ontologia padrão W3C (SSN), o permite que consultas e análises de dados sejam feitas à luz de conceitos mais amplamente aceitos pela Academia e Indústria.

Como trabalhos futuros espera-se: (i) integrar as soluções das duas iniciativas apresentadas em uma plataforma única de monitoramento, (ii) aplicar tal plataforma no âmbito de culturas agrícolas e na produção de mudas, (iii) ampliar o modelo de classes gerado a fim de comtemplar também as perspectivas *Actuation* e *Sampling* de SSN e (iv) avançar nas pesquisas e aplicações de SSN, buscando propor melhorias na especificação padrão da ontologia.

Referências

Abinaya, Kumar, V., & Swathika. (2015). "Ontology Based Public Healthcare System in Internet of Things (IoT)". *Procedia Computer Science*, *50*, 99–102. https://doi.org/http://dx.doi.org/10.1016/j.procs.2015.04.067

Barnaghi, P., Wang, W. E. I., Henson, C., & Taylor, K. (2012). "Semantics for the Internet of Things : early progress and back to the future". *International Journal on Semantic Web and Information Systems*. https://doi.org/10.4018/jswis.2012010101.

Bauer, M., Bui, N., De Loof, J., Magerkurth, C., Nettstra¨ter, A., Stefa, J., & Walewski, J. W. (2013). "IoT Reference Model". In A. Bassi, M. Bauer, M. Fiedler, T. Kramp, R. van Kranenburg, S. Lange, & S. Meissner (Eds.), *Enabling Thing to Talk: Designing IoT solutions with the IoT Architectural Reference Model* (p. 352). Springer.

CISCO. (2014). *The Internet of Thing Reference Modelo (2014)*.

Drury, B., Fernandes, R., Moura, M.-F., & Lopes, A. de A. (2019). "A survey of semantic web technology for agriculture". *Information Processing in Agriculture*, *6*, 487– 501.

Ericsson Blog. (2021). *What is semantic interoperability in IoT and why is it important?* https://www.ericsson.com/en/blog/2020/7/semantic-interoperability-in-iot

Gruber, T. R. (1993). "A Translation Approach to Portable Ontologies". *Knowledge Acquisition*, *5*(2), 199–220.

Guidoni, G. L., Almeida, J. P. A., & Guizzardi, G. (2020). "Transformation of Ontology-Based Conceptual Models into Relational Schemas". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 315–330. https://doi.org/10.1007/978-3-030-62522-1_23

Guizzardi, G. (2007). "On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models". In O. Vasilecas, J. Edler, & A. Caplinskas (Eds.), *Frontiers in Artificial Intelligence and Applications, Databases and Information Systems IV* (Vol. 15, pp. 18–39). IOS Press.

Honti, G. M., & Abonyi, J. (2019). "A Review of Semantic Sensor Technologies in Internet of Things Architectures". *Hindawi. Complexity.*, *2019*, 1–22. https://doi.org/https://doi.org/10.1155/2019/6473160

IERC. (2015). *Internet of Things - IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps*.

Li, S., Chen, L., & Chen, S. (2018). "An SNN Ontology Based Environment Monitoring Method for Intelligent Irrigation System". *Journal of Shanghai Jiaotong University (Science)*, *23*(4), 577–583. https://doi.org/10.1007/s12204-018-1959-0

Nardi, J. C., Falbo, R. de A., & Almeida, J. P. A. (2013). "Panorama of the Semantic EAI Initiatives and the Adoption of Ontologies by these Initiatives". *International IFIP Working Conference on Enterprise Interoperability (IWEI 2013)*, 198–211.

Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S. A., Zaidi, S. A. R., & Iqbal, N. (2019). "Precision Agriculture Techniques and Practices: From Considerations to Applications". *Sensors (Basel)*, *19*(17), 3796. https://doi.org/10.3390/s19173796

W3C. (2005). *IoT-Lite Ontology*. W3C Member Submission 26. https://www.w3.org/Submission/iot-lite/

W3C. (2017). *Semantic Sensor Network Ontology (SSN)*. https://www.w3.org/TR/vocab-ssn/



Sementesuite: desenvolvimento de uma aplicação móvel para determinação do teor de água em sementes

Érik Alves Almeida[1], Humberto B. da Cunha Júnior[2], Cosmo Rufino de Lima[1], Fábio José Marques[3], Pablo Radamés C. de França[4],

[1]Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), *Campus* Afogados da Ingazeira - PE – Brasil

[2]Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), *Campus* Garanhuns- PE – Brasil

[3]Instituto Federal de Educação, Ciência e Tecnologia de Alagoas (IFAL), *Campus* Piranhas - AL – Brasil

[4]Instituto Federal de Educação, Ciência e Tecnologia do Pará (IFPA), *Campus* Castanhal - PA – Brasil

eaa@discente.ifpe.edu.br, humberto.junior@garanhuns.ifpe.edu.br, cosmo.rufino@afogados.ifpe.edu.br, fabio.marques@ifal.edu.br, pablo.radames@ifpa.edu.br

Abstract. *The moisture content present in seeds plays an important role mainly in formation and maturation, and it is necessary to be calculated at all stages of the seed technology process. Considering its importance, this work aimed to present an application for mobile devices entitled SementeSuite, aiming to determine the moisture content of seeds in a practical and agile way. In addition to the main features related to the calculation of the moisture content, one can analyze the data stored on the device, as check the process through charts view. Preliminary results have shown that the application has the potential to optimize laboratory seed analysis.*

Resumo. *O teor de água presente nas sementes assume papel importante principalmente na formação e na maturação, sendo necessário sua determinação em todas as etapas do processo de tecnologia de sementes. Face a sua importância, este trabalho teve como objetivo apresentar uma aplicação para dispositivos móveis intitulada SementeSuite, conferindo agilidade e praticidade na determinação do teor de água de sementes. Além do cálculo do teor de água, os dados são armazenados no dispositivo do próprio usuário que pode acompanhar, por meio de gráficos, como tem se dado o*

comportamento dos experimentos realizados. Os resultados preliminares mostram que a aplicação tem potencial para otimizar as análises feitas em laboratório.

1. Introdução

A manutenção da qualidade das sementes é um fator de extrema importância para obtenção de alta produtividade. Assim, o teor de água presente nas mesmas é de vital importância para sua conservação e manejo, uma vez que sua atividade fisiológica pode ter seu processo acelerado ou minimizado em função do teor de água adequado para cada espécie [Lima Junior et al 2011].

Na área de produção e tecnologia de sementes, antes de instalar um determinado experimento ou realizar análises de um determinado lote de sementes, uma das avaliações realizadas no laboratório é a determinação do teor de água presente nas sementes. Esse procedimento que, na sua forma padrão, costuma se dar por meio de uma estufa a 105 ± 3 ºC [Brasil 2009], é seguido da computação dos dados obtidos para análise a posteriori, e confere pouca praticidade em virtude da natureza semi automatizada.

Desse modo, um dos problemas enfrentados é a necessidade de informatizar o processo como um todo, otimizando o tempo com o qual os resultados são obtidos. Nessa perspectiva, as aplicações móveis podem apresentar relevante contribuição, já que por meio de um aplicativo que facilite a forma de coleta das informações, os cálculos podem ser realizados para que finalmente se obtenha o percentual do teor de água nas sementes.

O procedimento de análise tradicional considera incluir os dados observados após um período de vinte e quatro horas, em planilhas eletrônicas, onde o cálculo do teor de água é então realizado. Contudo, essa etapa é precedida pela computação manual para, em seguida, ser feito o uso das fórmulas em computador, o que pode levar a falhas, sobretudo se for considerado um grande número de tratamentos, repetições e análises.

Vale salientar que com o desenvolvimento de um aplicativo simples, versátil e eficiente para uso nos laboratórios que realizam análise de sementes, torna-se o trabalho do usuário mais rápido e fácil, principalmente para iniciantes com trabalhos voltados à área. Assim, a proposta deste trabalho é apresentar uma aplicação para dispositivos móveis intitulada SementeSuite e que atua com o objetivo de conferir agilidade e praticidade na determinação do teor de água em sementes.

2. Metodologia

O aplicativo foi desenvolvido no Laboratório de Informática pertencente ao Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), *Campus* Afogados da Ingazeira - PE.

O desenvolvimento da aplicação considerou uma atividade interdisciplinar, partindo do levantamento dos requisitos com profissionais da área de ciências agrárias, até o desenvolvimento propriamente dito. O intuito foi identificar quais os principais recursos, bem como a melhor forma de apresentá-los aos usuários. Nesse sentido, partiu-se da construção de protótipos que foram passando pelo refinamento necessário até a

versão funcional do aplicativo. A implementação foi feita utilizando o framework Flutter, visando a possibilidade da ampla utilização do aplicativo pelas principais plataformas móveis existentes, sendo o SQLite utilizado como mecanismo para armazenar, no próprio dispositivo do usuário, as análises feitas. A representação do aplicativo por meio do seu respectivo diagrama de caso de uso é ilustrada na Figura 1.

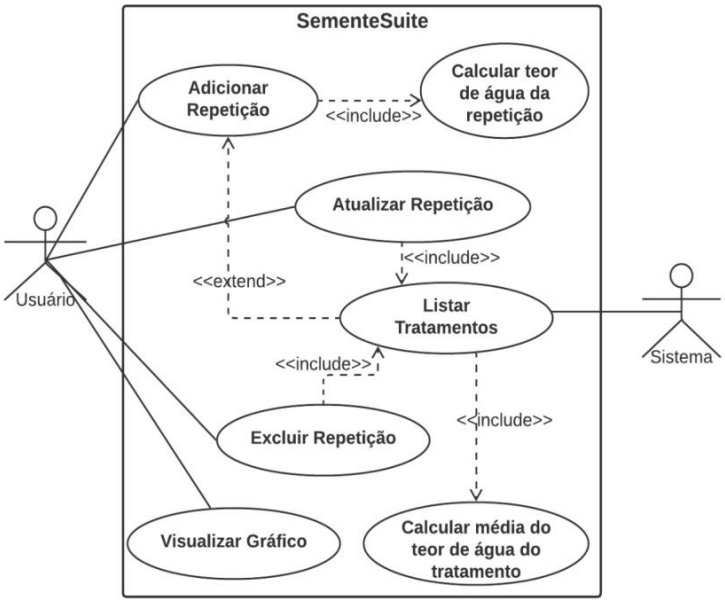


Figura 1. Diagrama de caso de uso do aplicativo SementeSuite.

O procedimento de análise consistiu em utilizar no SementeSuite os dados da determinação do teor de água de sementes dos trabalhos realizados por Lima et al. (2014) e França et al. (2018) e comparar os resultados gerados no aplicativo com os obtidos nesses experimentos. Em ambos os trabalhos, os autores realizaram o teor de água das sementes, utilizando o método padrão da estufa a 105 ± 3 ºC por 24 h, sendo os dados obtidos, validados estatisticamente através do teste de Scott-Knott a 5% de probabilidade. No trabalho de Lima et al. (2014), as sementes analisadas foram de *Poincianella pyramidalis* (Tul.) L. P. Queiroz e no de França et al. (2018) foram analisadas sementes de *Gossypium hirsutum* L.

3. Resultados e discussões

O SementeSuite utiliza um ambiente informatizado móvel para auxiliar em experimentos acadêmicos, bem como em análise da determinação do teor de água de sementes realizadas em laboratório, pois a implementação dos dados de forma rápida e precisa torna-se essencial quando tratamos de informações importantes. Testes conduzidos em laboratório tornam-se mais rápidos graças aos avanços tecnológicos atuais, no qual, para o contexto de análise de sementes, faz-se necessária a criação de uma ferramenta capaz de otimizar as diversas análises, dentre elas o teor de água.

Cada tratamento a ser analisado no aplicativo é identificado normalmente por um número inteiro (que pode, ou não, ser precedido pelo caractere “T”) ao qual estão relacionados um determinado conjunto de repetições. Estas, por sua vez, incluem algumas características, a saber: o número da lata, correspondente à identificação do

recipiente onde as sementes ficam armazenadas; o peso do recipiente de armazenamento juntamente com a tampa (denotada por M1); o peso da lata com a respectiva tampa acrescido do peso da semente fresca (denotado por M2) e o peso da lata com a respectiva tampa acrescido do peso da semente seca (denotado por M3).

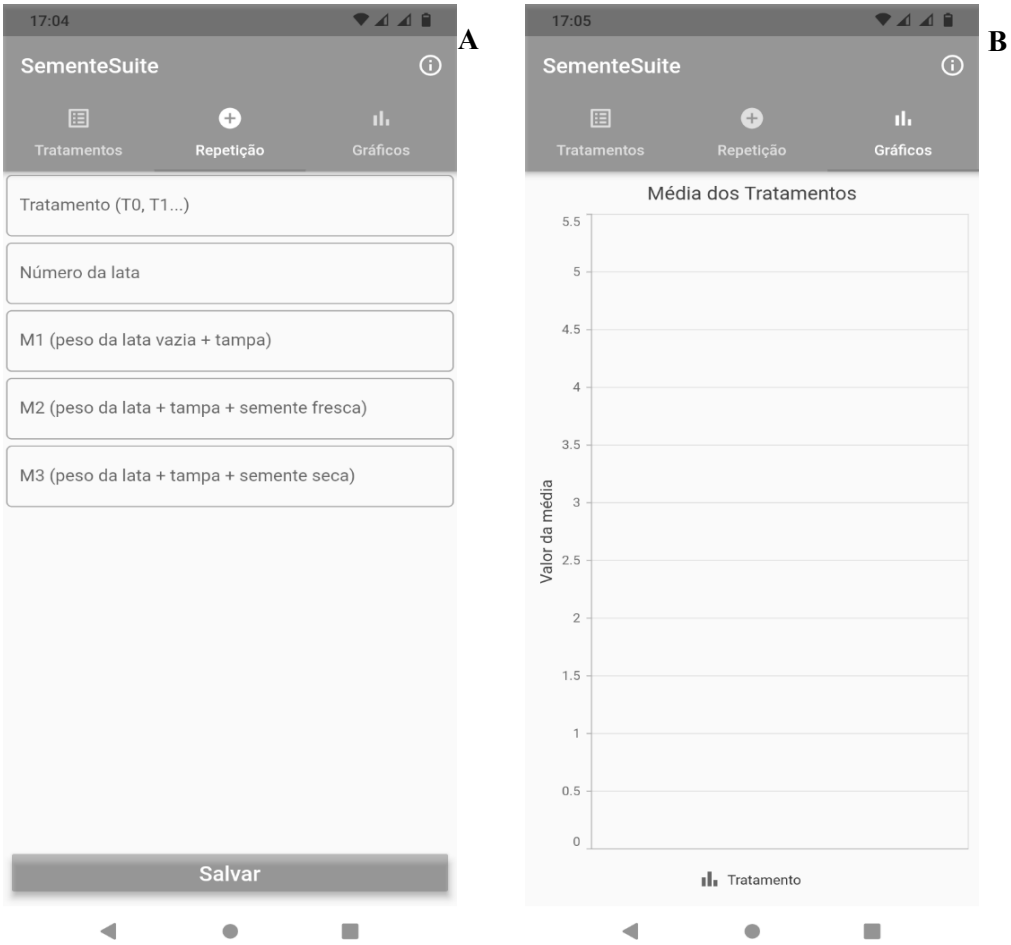


Figura 2. Telas do aplicativo SementeSuite: aba +Repetição (A) e aba Gráficos (B).

Os dados são inseridos a partir da opção +Repetição, representada na Figura 2A. Ao terminar de preencher, o botão Salvar é pressionado e os dados passam a compor a repetição do respectivo tratamento. Ao visualizar a Figura 2B, é apresentado a tela de Gráficos, onde serão mostrados todos os tratamentos com suas respectivas médias das repetições em um gráfico de colunas, sendo possível identificar de forma imediata o melhor tratamento.

Os testes para validação do aplicativo foram realizados com base nos dados obtidos em análises feitas em laboratório e publicados em dois trabalhos. A Tabela 1 sumariza os resultados das análises de teor de água feitas com o aplicativo SementeSuite a partir dos dados do trabalho desenvolvido por Lima et al (2014).

Tabela 1. Resultados do teor de água de sementes de *Poincianella pyramidalis* desenvolvido por Lima et al. (2014) e testados no aplicativo SementeSuite.

Tratamento	Nº da lata	M1	M2	M3	**TA (%)**	**Média (%)**
14	14	31,27	33,09	32,87	**12,09**	**11,58**
	46	27,99	29,87	29,66	**11,06**	
15	66	26,4	28,06	27,85	**12,17**	**11,95**
	75	28,42	30,22	30	**11,72**	
20	38	26,03	27,72	27,514	**12,19**	**12,34**
	69	28,64	30,28	30,07	**12,50**	
21	81	31,46	33,08	32,86	**13,02**	**12,91**
	25	28,97	30,61	30,4	**12,80**	
22	38	31,51	33,5	33,259	**12,46**	**12,47**
	8	28,42	30,52	30,26	**12,47**	
24	20	28,16	29,89	29,66	**13,28**	**13,30**
	11	26,93	28,47	28,26	**13,32**	
28	34	29,86	31,52	31,309	**12,71**	**12,83**
	38	27,93	29,56	29,349	**12,94**	

Como pode ser observado, os valores indicados para a média do teor de água para cada tratamento (ou matriz) foram os mesmos valores encontrados no trabalho desenvolvido por Lima et al (2014), considerando cada tratamento com duas repetições (Tabela 1). Esse aspecto permite inferir que a contribuição do aplicativo SementeSuite é promissora, pois os valores obtidos não diferem das análises feitas pelos autores no trabalho supracitado.

Na Figura 3A são apresentadas as médias dos tratamentos com suas respectivas repetições na tela do aplicativo. A determinação do teor de água das sementes é considerado, quando associado às outras características, como um dos principais índices que evidencia o processo de maturação e, muitas vezes, é sugerido como ponto de referência para indicar a condição fisiológica das sementes [Silva 2002].

É importante salientar que o sucesso na colheita das sementes após a maturidade fisiológica depende da porcentagem de água presentes nas mesmas. Assim, um aplicativo como o SementeSuite pode auxiliar de forma prática e segura na época mais apropriada para realização da colheita, com redução de perdas no campo.

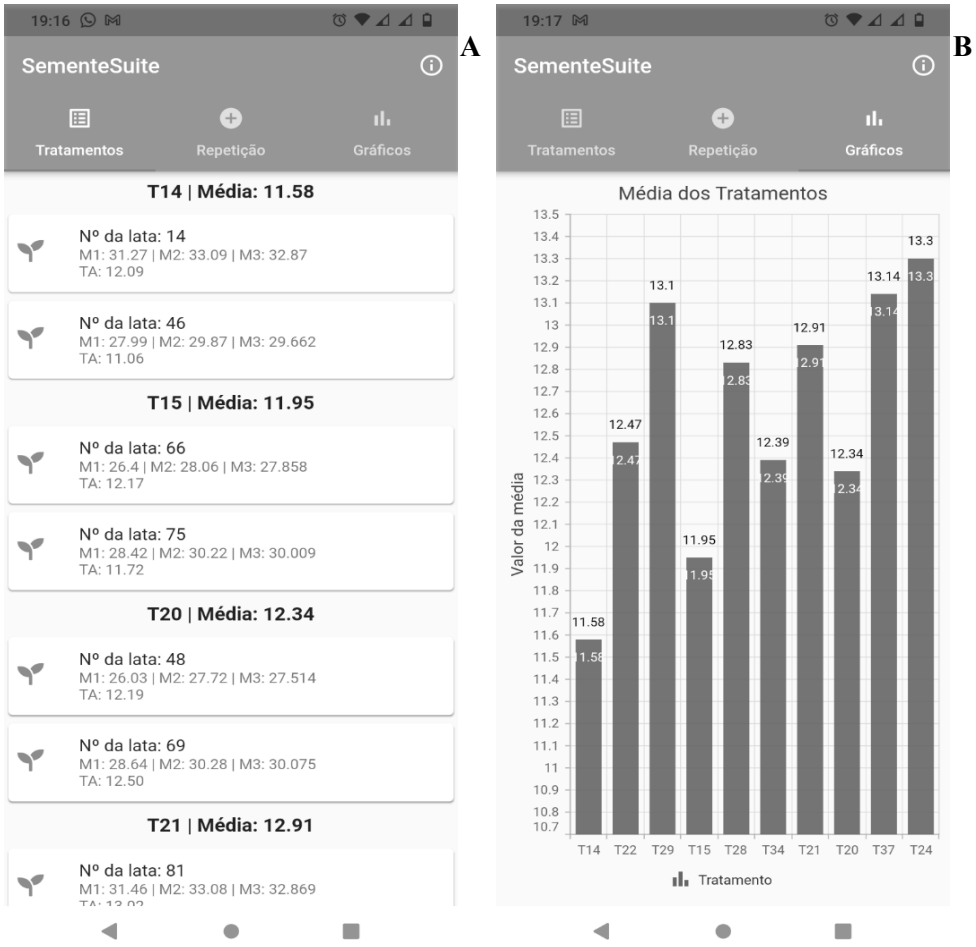


Figura 3. Telas do aplicativo SementeSuite com dados de experimentos científicos: ilustração de alguns tratamentos e repetições para fins de visualização (A) e gráfico de colunas com as médias de todos os tratamentos inseridos (B). Dados provenientes de trabalho realizado por Lima et al. (2014).

Em estudos de maturação das sementes, o teor de água assume papel importante, permanecendo elevado nas fases em que ocorre a transferência de matéria seca da planta, decrescendo rapidamente a partir do momento em que atingem o máximo conteúdo de matéria seca e desligam-se da planta mãe [Marcos Filho 2005].

Um diferencial do SementeSuite é a apresentação do gráfico de forma rápida e com precisão, como pode ser visualizado na Figura 3B. Isso permite que o usuário consiga identificar imediatamente qual o melhor tratamento do experimento conduzido, dispensando o uso de cadernetas e planilhas para computação dos dados para uma análise apenas a posteriori.

A Tabela 2 trata das análises feitas visando a comparação com os resultados do teor de água encontrados no trabalho desenvolvido por França et al (2018), utilizando os dados de sementes de algodão, sendo o experimento validado pelos autores no trabalho mencionado através do teste Scott-Knott a 5% de probabilidade.

Tabela 2. Resultados do teor de água de sementes de *Gossypium hirsutum* L. desenvolvido por França et al. (2018) e testados no aplicativo SementeSuite.

Tratamento	Nº da lata	M1	M2	M3	**TA(%)**	**Média (%)**
T1	1	36,67	45,9	45,11	**8,56**	**8,53**
	2	36,15	45,56	44,77	**8,40**	
	3	34,54	43,52	42,73	**8,80**	
	4	35,88	45,82	44,99	**8,35**	
T2	5	34,39	47,09	44,83	**17,80**	**17,75**
	6	36,53	46,98	45,07	**18,28**	
	7	35,42	45,42	43,71	**17,10**	
	8	34,49	44,52	42,73	**17,85**	
T3	9	36,2	47,25	46,25	**9,05**	**8,80**
	10	35,94	45,36	44,52	**8,92**	
	11	34,74	45,07	44,17	**8,71**	
	12	35,91	45,2	44,41	**8,50**	
T4	13	33,57	43,83	42,93	**8,77**	**8,78**
	14	35,02	44,65	43,8	**8,83**	
	15	35,0	44,28	43,49	**8,51**	
	16	35,34	44,67	43,83	**9,00**	
T5	17	36,37	45,52	44,75	**8,42**	**8,81**
	18	34,68	43,65	42,84	**9,03**	
	19	36,4	45,65	44,82	**8,97**	
	20	31,56	45,08	44,24	**8,82**	
T6	21	35,22	44,71	43,89	**8,64**	**9,01**
	22	34,19	44,26	43,37	**8,84**	
	23	36,34	45,68	44,8	**9,42**	
	24	36,78	45,64	44,83	**9,14**	
T7	25	35,55	45,77	44,84	**9,10**	**8,68**

26	34,57	42,75	42,16	**8,44**
27	35,72	45,03	44,23	**8,59**
28	37,31	47,66	46,77	**8,60**

Os dados encontrados na Tabela 2 correspondem aos mesmos valores encontrados por França et al (2018) e cuja representação na interface do aplicativo é indicada na Figura 4A.

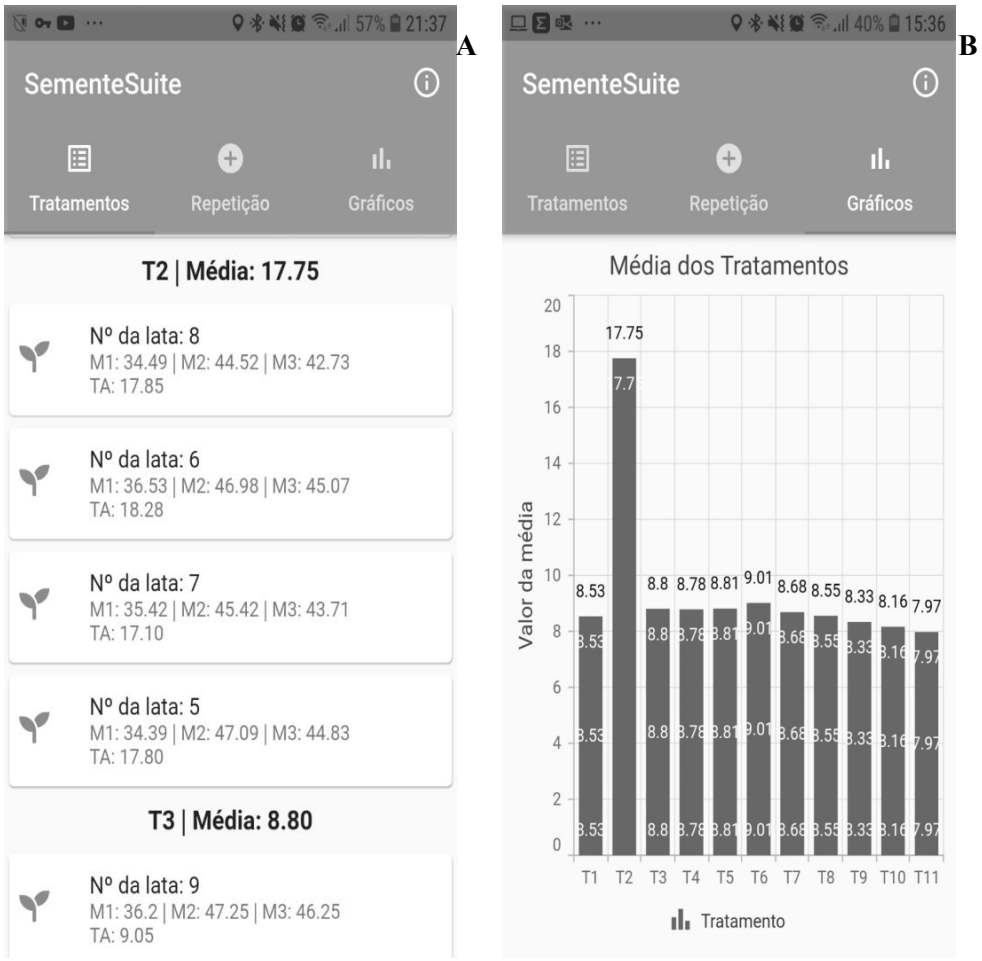


Figura 4. Telas do aplicativo SementeSuite com dados de experimentos científicos: ilustração de alguns tratamentos e repetições para fins de visualização (A) e gráfico com todos os tratamentos inseridos (B). Dados provenientes de trabalho realizado por França et al. (2018).

A Figura 4B corresponde à indicação do gráfico com os onze tratamentos que foram inseridos no aplicativo, e assim é possível constatar que independentemente da quantidade de dados, mesmo existindo valores discrepantes, como o valor médio do tratamento T2, o aplicativo não apresentou erros nos valores gerados, conforme pode ser observado na interface da Figura 4A, e na Figura 4B, correspondente ao gráfico.

Diante do que foi exposto, podemos inferir que o aplicativo SementeSuite poderá contribuir com uma rápida resposta na decisão de realização da colheita de frutos e sementes, em virtude do teor de água presente naquele momento. Assim, vale salientar que a qualidade das sementes torna-se muito prejudicada quando a colheita é realizada antes de atingir a maturidade fisiológica ou após esse estágio [Barbosa et al 2015].

4. Conclusões

O SementeSuite é um aplicativo que visa otimizar o processo de análise de sementes em laboratório. A proposta é que partindo de uma interface objetiva, o cálculo do teor de água possa ser realizado, conferindo mais celeridade nas analises.

Os resultados obtidos indicam a capacidade do aplicativo em promover uma análise otimizada em atividades de pesquisa de tecnologia de sementes, sendo previsto, como trabalhos futuros, a análise de mais de uma espécie de semente sendo realizada pelo aplicativo em um mesmo experimento, além da inserção de outros tipos de análises, como o teste de germinação, por exemplo. Essa melhoria permitirá que o aplicativo possa ter sua finalidade escalada para além do uso em laboratório.

5. Agradecimentos

Os autores agradecem pelo apoio do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE).

6. Referências

Barbosa, J. M.; Rodrigues, M. A.; Barbério, M.; Araújo, A. C. F. B. Maturação de sementes de espécies florestais tropicais In: Piña-Rodrigues; Figliolia, M. B.; Silva, A. *Sementes florestais tropicais*: da ecologia à produção. Londrina, ABRATES, 2015. p.180-189.

França, P. R. C.; Almeida, F. A. C.; Silva, O. R. R. F.; Jerônimo, J. F. Physiological quality of *Gossypium hirsutum* L. seeds after the thermal delinting process. *Comunicata Scientiae*. v. 9, n.3, p. 465-474, 2018.

Brasil, Ministério da Agricultura e Reforma Agrária. *Regras para Análise de sementes*. Brasília: SNDP/DNDV/CLAV, 2009. 395p.

Lima, C. R.; Bruno, R. L. A.; Silva, K. R. G.; Pacheco, M. V.; Alves, E. U.; Qualidade fisiológica de sementes de diferentes árvores matrizes de *Poincianella pyramidalis* (Tul.) L.P. Queiroz. *Revista Ciência Agronômica*, v. 45, n. 2, p. 370-378, 2014.

Lima Junior, M. J. V.; Gentil, D. F. O.; Figliolia, M. B.; Ferraz, I. D. K.; Calvi, G. P.; Rodrigues, F. C. M. P.; Silva, V. S.; Souza, M. M. Manual de procedimentos para análise de sementes florestais. *Associação Brasileira de Tecnologia de Sementes*, Londrina-PR, 2011. 83p.

Marcos Filho, J. *Fisiologia de sementes de plantas cultivadas***.** Piracicaba: FEALQ, 2005. 495p.

Silva, L. M. M. *Maturação fisiológica de sementes de Cnidosculus phyllacanthus Pax & K. Hoffm*. 2002. 61f. Tese (Doutorado em Agronomia) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, 2002.



Simulação de Sistemas de Produção de Novilho Precoce: uma Aplicação Prática

Roberto A. F. Abreu[1], Thiago Basso[2], Camilo Carromeu[3], Rodrigo da Costa Gomes[3], Fernando R. T. Dias[3], Débora Maria Barroso Paiva[1]

[1]Faculdade de Computação – Universidade Federal de Mato Grosso do Sul (UFMS) - Campo Grande, MS – Brasil

[2]Universidade Federal da Grande Dourados (UFGD) – Dourados, MS – Brasil

[3]Embrapa Gado de Corte - Campo Grande, MS – Brasil

abreu.roberto@ufms.br, thiagobasso@ufgd.edu.br, camilo.carromeu@embrapa.br, rcostagomes@gmail.com, fernando.dias@embrapa.br,debora.paiva@ufms.br

Abstract. *The development of software to support the sciences as a new platform for scientific research and experimentation has generated contributions to several areas of knowledge and is considered a trend for the coming years. In this article, the +Precoce platform is presented, focused on the field of Livestock, which offers models of beef cattle production systems and allows rural producers to make simulations considering their particular data. The results obtained show that the platform can contribute to decision making in rural areas and generate several benefits, such as improving the quality of brazilian meat.*

Resumo. *O desenvolvimento de software para apoiar as ciências como uma nova plataforma de pesquisa e experimentação científica tem gerado contribuições para diversas áreas de conhecimento e é considerada uma tendência para os próximos anos. Neste artigo apresenta-se a plataforma +Precoce, voltada para o domínio da Pecuária, que oferece modelos de sistemas de produção de gado de corte e permite que produtores rurais façam simulações considerando seus dados particulares. Os resultados obtidos mostram que a plataforma pode contribuir para a tomada de decisão no meio rural e gerar diversos benefícios, tais como a melhoraria da qualidade da carne brasileira.*

1. Introdução

O setor de agronegócio possui papel importante para o desenvolvimento da economia e sociedade brasileira. O notável crescimento do agropecuário brasileiro tem se sustentado no aumento da produtividade, muito mais do que na ocupação de novas terras, pela aplicação de tecnologias e conhecimento acumulado tanto pelos produtores quanto por instituições que integram as cadeias produtivas do agronegócio. A inovação é a base para o desenvolvimento sustentável e atributo potencializador para criação de novas oportunidades e enfrentamento de desafios impostos por mercados globalizados (Santos et al., 2011).

O crescente uso de diferentes tecnologias envolvendo, por exemplo, biotecnologia e tecnologia da informação tem elevado o potencial de criação de produtos e processos disruptivos e de alto impacto para o setor de agropecuária. Considerando ainda a intensificação do mercado de agricultura e pecuária digital, tanto os avanços da biologia sintética e das novas tecnologias aplicadas a sistemas genéticos complexos, quanto a expansão da bioinformática na análise e no compartilhamento de dados científicos são caminhos que tendem a se intensificar (Embrapa Visão 2030). Considerando estas tendências e o volume de dados que pode ser gerado, surge a necessidade de prover apoio computacional à experimentação científica (e-Ciência ou e-Science) no sentido de auxiliar pesquisadores a explorar coleções de dados. Estas novas demandas certamente impõem grandes desafios que precisam ser enfrentados, considerando a modelagem, o gerenciamento e a apresentação de dados.

A plataforma +Precoce (Abreu, 2020, Basso, 2018, Basso et al., 2018) é um projeto desenvolvido em parceria entre a Embrapa Gado de Corte, Embrapa Pantanal e a Faculdade de Computação da Universidade Federal de Mato Grosso do Sul (UFMS). O projeto recebe financiamento da Embrapa e está inserido no contexto do "Precoce MS", programa responsável pelo incentivo à produção de bovinos precoces no estado de Mato Grosso do Sul, vinculado à Secretaria de Meio Ambiente, Desenvolvimento Econômico, Produção e Agricultura Familiar do governo do estado. A Plataforma +Precoce provê acesso aos técnicos rurais e produtores de gado de corte aos sistemas melhorados de produção pesquisados pela Embrapa. Utilizando um simulador o produtor pode realizar mudanças em parâmetros visando reproduzir as características de sua propriedade rural e simular estes sistemas de produção.

De forma geral, a plataforma +Precoce considera sistemas de produção de gado de corte e simplifica sua conversão para modelos possíveis de serem simulados computacionalmente. O objetivo deste artigo é apresentar a evolução da plataforma +Precoce, em especial enfatizando seu modelo de simulação voltado para o domínio de pecuária e de apresentação de dados.

Este artigo está apresentado da seguinte forma: a Seção 2 apresenta os principais conceitos e trabalhos relacionados, a Seção 3 apresenta os principais módulos da plataforma +Precoce e a Seção 4 apresenta as considerações finais.

2. *Background* e Trabalhos Relacionados

A pecuária de precisão é a área que estuda o uso das tecnologias de informação e comunicação na produção de bovinos. Nesta área, a tecnologia da informação é utilizada para garantir o uso de boas práticas no processo de produção pecuário almejando, assim, impulsionar a produtividade. A utilização de tecnologias pode ser feita em vários níveis e em diversas áreas da indústria bovina, lançando novos desafios para as disciplinas de Ciência da Computação no que tange o desenvolvimento de novas ferramentas e dispositivos (software e hardware) que aumentem a competitividade de estabelecimentos pertencentes à indústria pecuária (Cáceres et al., 2011).

A concorrência global impõe forte ritmo e dinamicidade na produção de gado de corte. Neste sentido, a Embrapa Gado de Corte conduz estudos há décadas visando à construção de modelos de sistemas melhorados de produção que sirvam de referência ao produtor rural. Estes modelos indicam técnicas empregadas na produção, coeficientes técnicos estimados e indicadores de desempenho econômico.

No entanto, esses modelos de sistemas geralmente estão disponibilizados no formato de publicações técnico-científicas e os indicadores são elaborados utilizando planilhas eletrônicas disponíveis apenas para pesquisadores. Essa característica dificulta o alcance e utilização pelos produtores rurais, além de dificultar o compartilhamento de atualizações e criação de novos modelos. Neste contexto, ferramentas de software têm papel cada vez mais importante uma vez que, dado o volume de informações e quantidade de variáveis que existem atualmente, é impossível tomar decisões rápidas e seguras sem o uso de ferramentas robustas.

Existem alguns simuladores disponíveis no setor agropecuário, sendo mencionados três deles a seguir:

- APSIM: simulador de sistemas agrícolas, contém módulos que permitem simulações envolvendo solo, plantas, animais e clima (APSIM Initiative, 2021);
- PaSim: provê simulações dos impactos das alterações climáticas nas pastagens e fornece *feedback* sobre emissão de gases do efeito estufa na atmosfera. É composto por submodelos para plantas, animais, microclima, biologia e física do solo e gerenciamento (PASIM, 2021);
- Embrapa Invernada: fornece apoio para o planejamento da produção de bovinos de corte. Permite realizar simulações do crescimento de pastagens, pastejo e crescimento de animais, com possibilidade de formular dietas otimizadas e comparar e analisar cenários de produção (Invernada, 2021).

A plataforma +Precoce apresenta como diferencial o fato de ser uma ferramenta Web para a organização e disponibilização de informações a respeito de cria, recria e engorda para a produção de bovinos destinados a programas e parcerias que bonificam carcaças por qualidade. Chamada de plataforma +Precoce em alusão ao “novilho precoce”, um termo relacionado à qualidade, permite a geração de indicadores econômicos e ambientais para auxiliar o produtor na tomada de decisão e caracteriza a adequação esperada para os diversos programas e parcerias de carne de qualidade dos animais produzidos em cada sistema disponibilizado.

3. Plataforma +Precoce

A plataforma é composta por quatro módulos: o Módulo Desenhista, o Módulo Simulador, o Módulo Administrador e o Módulo Portal. De forma geral, o Módulo Desenhista permite aos técnicos da Embrapa cadastrar e manter os modelos computacionais dos sistemas de produção; o Módulo Simulador é responsável pela execução das simulações dos modelos computacionais dos sistemas de produção; o Módulo Administrador gerencia as funcionalidades da plataforma relativas a cadastro de usuários perfis de acesso, logs, etc. e o Módulo Portal é a interface com o usuário final (pesquisadores, produtores e técnicos rurais).

A plataforma +Precoce permite que usuários façam a simulação e o acesso às informações de sistemas melhorados de produção de gado de corte, construídos a partir do conhecimento acumulado de experimentos da Embrapa Gado de Corte e parceiros. A aplicação deve ser flexível o suficiente para que seus usuários possam simular tais sistemas com dados o mais próximo possível de suas propriedades, de modo a colaborar com a produção de bovinos e contribuir para melhorar a qualidade da carne brasileira.

3.1. Módulo Desenhista

A plataforma +Precoce suporta diversos modelos de sistemas de produção de gado de corte e simplifica sua conversão para modelos passíveis de serem simulados computacionalmente. A gerência para estes modelos é propiciada pelo Módulo Desenhista, em que o ator denominado "Desenhista", ficará responsável por modelar e gerenciar os sistemas de produção considerando o modelo computacional. Este papel é exercido por pesquisadores instruídos para esta função.

Um modelo de sistema de produção é uma caixa preta em que recursos entram por suas fronteiras, são processados e geram como saídas coprodutos, resíduos ou emissões e um produto. Possuem estações e fluxos de entrada e saída. Os fluxos carregam recursos, organizados em hierarquias de tipos. Estações utilizam fórmulas de duração e fluxos utilizam fórmulas de coeficientes que envolvem parâmetros. Os valores desses parâmetros são informados pelo usuário na hora de solicitar uma simulação. A execução da simulação calcula as quantidades e datas dos fluxos e, a partir destes, os valores dos indicadores de desempenho.

Um usuário desenhista pode cadastrar, editar e excluir modelos, parâmetros e suas categorias e hierarquias, fórmulas de coeficiente, duração, valor inicial, valor residual ou vida útil, parâmetros associados ao sistema, definir os valores mínimo e máximo aceitáveis e o valor padrão. O Módulo Desenhista é responsável por gerenciar o modelo computacional permitindo que o usuário desenhista consiga inserir, ler, editar e excluir dados das entidades do modelo computacional. Este módulo foi construído utilizando o Titan *Framework* (Titan Framework, 2021). Na Figura 1 é possível observar uma tela que exibe os modelos cadastrados. Ao final de cada linha da tabela são exibidas as ações de visualizar, editar, excluir e copiar modelo.

Na Figura 2 é indicado um exemplo de tela de cadastro de dados. Essa tela apresenta os campos necessários para o cadastro de um fluxo. O canto superior direito apresenta o menu de ações disponíveis. É possível salvar os dados preenchidos (ação que irá processar o *form*, validar dados e persistir em base, no caso de sucesso na validação ou retornar erro).

3.2. Módulo Simulador e Módulo Administrador

O Módulo Simulador é responsável pela execução das simulações requisitadas e, de forma geral, este módulo faz o carregamento da base de dados do modelo computacional escolhido, a execução das rotinas de cálculo de simulação e a entrega dos resultados na forma de indicadores de desempenho.

O Módulo Simulador é responsável pela execução das simulações dos modelos computacionais dos sistemas de produção, processando as requisições de simulação. Para isso, ele carrega na memória o modelo do sistema de produção e os valores dos parâmetros escolhidos para simulação e dispara a execução das rotinas de cálculo de simulação, entregando os resultados na forma de *Indicadores*. Devido à limitação de espaço, o Módulo Simulador não será detalhado neste artigo. Os resultados obtidos após uma simulação são apresentados no Módulo Portal, que será apresentado na subseção 3.3. De forma semelhante, o Módulo Administrador, responsável por gerenciar as funcionalidades da plataforma relativas a cadastro de usuários, perfis de acesso e outros cadastros, não será apresentado em detalhes neste artigo.

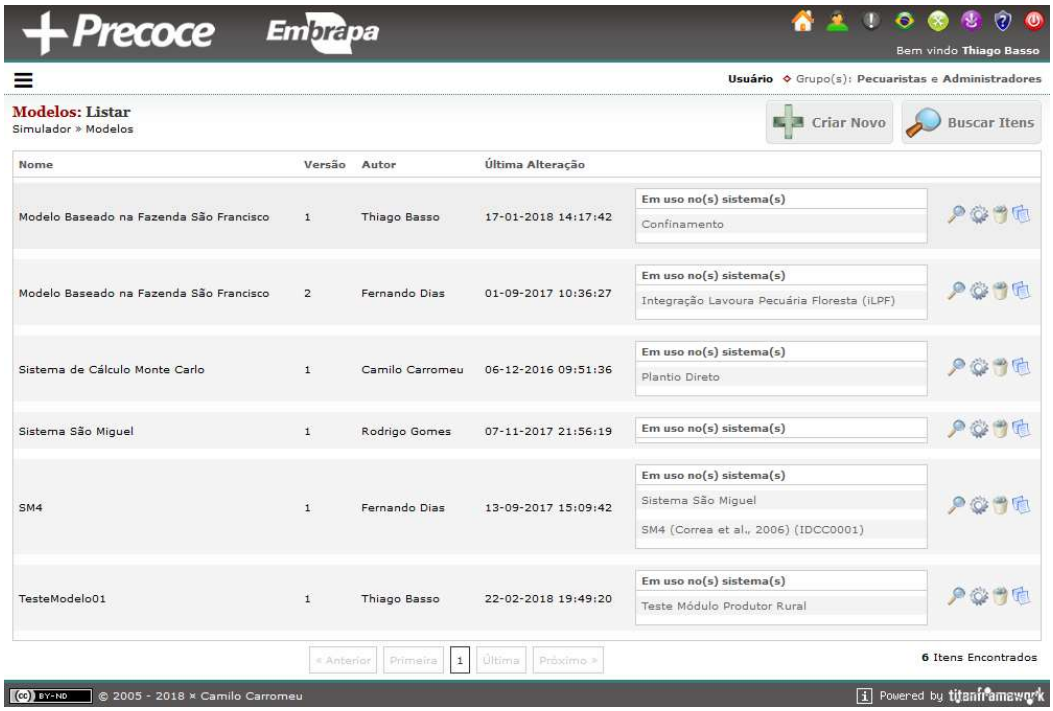


Figura 1 – Módulo Desenhista - Tela de Listagem de Modelos

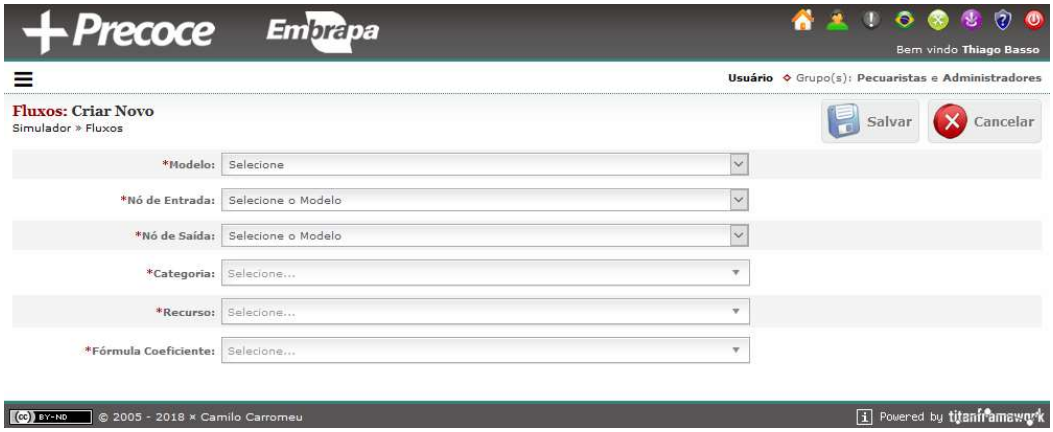


Figura 2 – Módulo Desenhista - Tela de Cadastro ou Edição de Fluxo

3.3 Módulo Portal

O Módulo Portal é a interface com o usuário final (pesquisadores, produtores e técnicos rurais). Neste módulo acontece a escolha do modelo para simulação, consulta à descrição e protocolo dos modelos de produção, consulta a simulações já realizadas, exclusão, edição ou cópia de simulações realizadas, ajuste de parâmetros das simulações e exibição dos resultados de até três simulações para comparação.

Foi utilizada a teoria de Gráfico de Gantt para representar a simulação do ciclo de reprodução de gado de corte. A representação de tarefas em barras horizontais na linha do tempo, mostra dados relevantes para tomada de decisão de forma intuitiva. Por exemplo: quando uma tarefa começa ou termina, qual tarefa se inicia antes ou depois, quais tarefas podem ser executadas em paralelo, etc. Além disso, por meio de cores é possível representar grupos de tarefas que têm algo em comum como a necessidade de um recurso ou de determinada ação.

É possível encontrar vários exemplos práticos dos benefícios do Gráfico de Gantt na literatura e que motivaram sua utilização neste trabalho. Por exemplo, Borges et al. (2013) utilizaram o Gráfico de Gantt para obter um melhor planejamento e controle da produção de uma empresa e obtiveram uma redução de 56% no *lead time*, dobraram a produtividade, diminuíram os gastos com horas extras e aumentaram a satisfação do cliente.

De forma geral, o simulador da Plataforma +Precoce retorna um JSON com o resultado da simulação e o Módulo Portal executa como principais tarefas: 1) Ler o JSON; 2) Abstrair as informações necessárias para a construção de um Gráfico de Gantt que mostre as principais atividades no manejo (Cria, Recria, Monta, Engorda, etc.), em uma linha de tempo em que fique claro início e fim de cada atividade, dando uma visão cronológica de toda a cadeia servindo como um mapa para a tomada de decisões e para provisão antecipada de recursos e 3) Desenhar o gráfico na tela.

A partir do JSON retornado os dados são parseados em formato SVG (*Scalabre Vectorial Graphics*, criado pelo W3C - *World Wide Web Consortium*) para permitir a manipulação de imagens vetorizadas no navegador web. A biblioteca 3d.js (*Data-Drive Documents*) é utilizada para visualização dos dados.

A Figura 3 apresenta um exemplo de uso do software desenvolvido. O usuário escolhe de uma até quatro das simulações geradas pelo simulador, se deseja comprimir as datas (traz a data das atividades para o mais cedo possível respeitando dependências) e se deseja ver as simulações escolhidas sequencialmente ou intercaladas (esta última não será apresentada neste artigo).

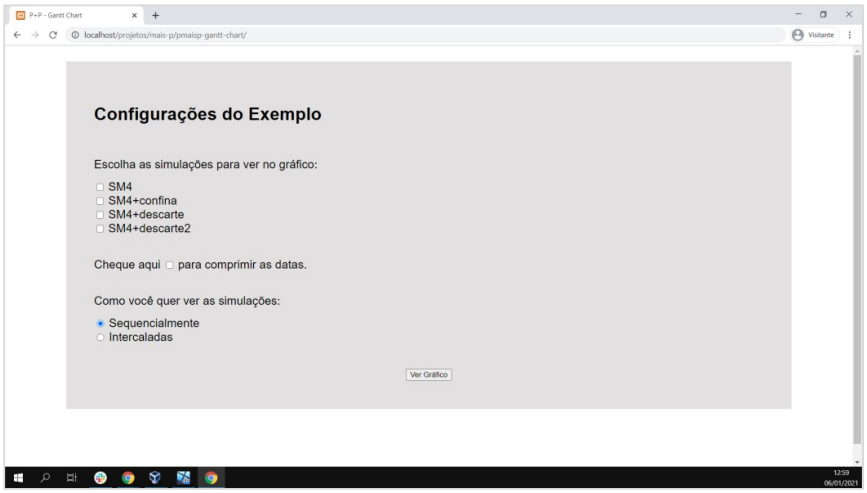


Figura 3 – Seleção de parâmetros para apresentação da simulação

Neste exemplo, o SM4 foi baseado no "Sistema Melhorado 4", descrito em https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/326450/1/COT102.pdf (Corrêa et al., 2006). Os demais sistemas foram definidos apenas para testes da aplicação, e não representam sistemas reais ou recomendados pela Embrapa:

- SM4+confina: maior tempo de confinamento e menor tempo de engorda a pasto (tarefa imediatamente anterior);
- SM4+descarta: descarta (vende) todas as vacas que não emprenharam na última monta, logo após a desmama, eliminando as atividades de "reserva" destas vacas para a monta do ano seguinte;
- SM4+descarta2: semelhante ao SM4+descarta, mas vendendo também as vacas em fim de ciclo reprodutivo, que foram selecionadas para não irem para a próxima monta, com as crias, logo após o puerpério.

A Figura 4 apresenta o cronograma da simulação. Neste cronograma é apresentada a cadeia mais longa com todas as suas atividades, seguida de cadeias menores que podem ser executadas em paralelo. Ao posicionar o ponteiro do mouse sobre uma atividade, aparece uma dica (*tip*) informando: nome da estação, data de início e fim da atividade. Além disso, as estações são apresentadas na mesma cor facilitando a identificação dos momentos em que ocorrerá cada estação no ciclo de produção.

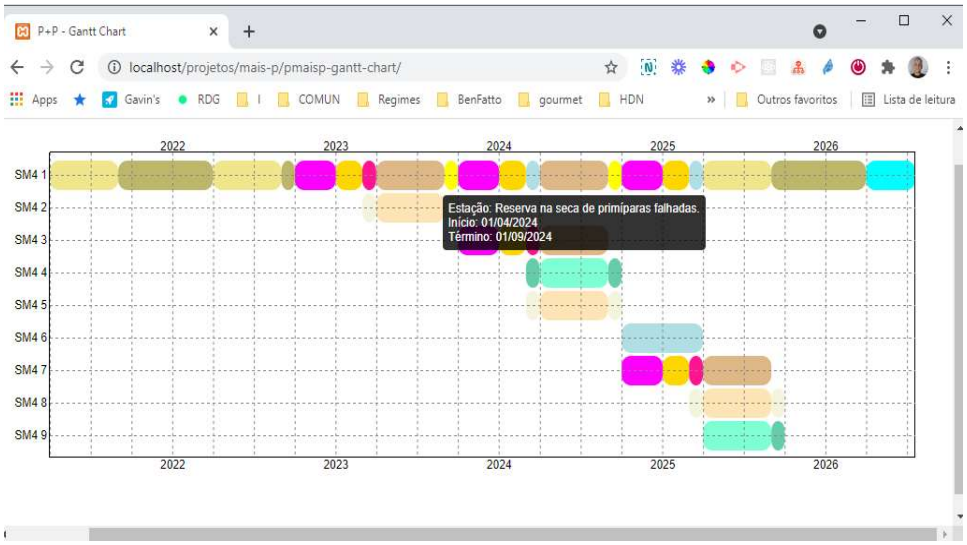


Figura 4 – Cronograma de simulação

4. Conclusões

A plataforma +Precoce apresenta, como principal diferencial, o fato de ser uma ferramenta computacional que promove a criação de modelos computacionais dos sistemas de produção de forma fácil, ou seja, cria um local centralizado em que o produtor rural ou pesquisadores podem criar seus próprios modelos matemáticos, cadastrar dados necessários para a simulação (de acordo com a realidade de cada usuário final) e obter informações que são apresentadas de forma simples, valorizando a usabilidade.

A principal vantagem observada é que, anteriormente ao desenvolvimento da plataforma +Precoce, os dados importantes para os produtores rurais eram disponibilizados em planilhas eletrônicas e exigiam um cuidadoso estudo e interpretação. Espera-se, a partir de agora, contribuir no sentido de prover apoio computacional à experimentação científica no domínio de pecuária. Atividades futuras estão relacionadas aos testes extensivos usando a plataforma, sobretudo, realizando estudos de caso com dados de diferentes propriedades rurais.

Referências

Abreu, R. A. F. Interface Gráfica para a Plataforma +Precoce: Uma Visão em Gantt Chart. 2020. Dissertação (Mestrado Profissional em Computação Aplicada) - Universidade Federal de Mato Grosso do Sul.

APSIM Initiative. *The Agricultural Production Systems sIMulator*. 2021. Disponível em: <https://www.apsim.info/>. Acesso em: 22.03.2021.

Basso, T.; Plataforma +Precoce: Simulador de Sistemas de Produção de Novilho Precoce. 2018. Dissertação (Mestrado Profissional em Computação Aplicada) - Universidade Federal de Mato Grosso do Sul.

Basso, T.; Gomes, R. C.; Dias, F. R. T. ; Carromeu, C. ; Cagnin, M. I. ; Paiva, D. M. B. Desenvolvimento da Plataforma +Precoce para Simulação de Sistemas de Produção de

Novilho Precoce. In: IX Congresso Brasileiro de Software - Trilha da Indústria, 2018, São Carlos, SP.

Borges, J. P. V. et al. Planejamento e controle da produção de uma indústria de cataventos apoiado pelo gráfico de gantt: um estudo de caso. XXXIII Encontro Nacional da Engenharia de Produção. Salvador, 2013.

Cáceres, E. N. et al. Computational precision livestock-position paper. In: II Workshop of the Brazilian Institute for Web Science Research. [Rio de Janeiro, RJ: no. 02-03], 2011.

Corrêa, E. S. et al. Sistemas de produção melhorados para gado de corte em Mato Grosso do Sul. Campo Grande: Embrapa Gado de Corte, 2006. Disponível em https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/326450/1/COT102.pdf (acesso em 06/01/2021).

Embrapa Visão 2030: O futuro da Agricultura Brasileira – Brasília, DF: Embrapa, 2018.

Invernada. *Embrapa Invernada*. 2021. Disponível em: <https://www.invernada.cnptia.embrapa.br/>. Acesso em: 22.03.2021

PASIM. *The Pasture Simulation Model (PaSim)*. 2021. Disponível em: <https://www1.clermont.inra.fr/urep/modeles/pasim.htm>. Acesso em: 22.03.2021.

Santos, A. R. et al. Sw agro: Estudo do mercado brasileiro de software para o agronegócio. Estudo do mercado brasileiro de software para o agronegócio. Campinas: Embrapa Informática Agropecuária, 2011.

Titan Framewotk *Titan Famework*. 2021. Disponível em: <http://www.titanframework.com/>. Acesso em: 22.03.2021.



Sistemas de suporte à decisão nas fazendas inteligentes: um mapeamento sistemático

Jonas S. Gomes[1], José Maria N. David[1], Regina Braga[1], Wagner Arbex[2], Bryan Barbosa[1], Wneiton Luiz Gomes[2], Leonardo M. Gravina Fonseca[2]

[1] Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG – Brazil

[2] Empresa Brasileira de Pesquisa Agropecuária – Embrapa, Juiz de Fora, MG – Brazil

jonas.gomes@estudante.ufjf.br, {jose.david, regina.braga}@ufjf.edu.br, bryan.barbosa@ice.ufjf.br, {wagner.arbex, wneiton.gomes, leonardo.gravina}@embrapa.br

Abstract. *This work presents a systematic mapping related to Decision Support Systems (DSS) in precision agriculture and livestock, in smart farming context. Through this systematic mapping, we aimed to investigate solutions that are used to support data integration from IoT sensors. The results point to the use of a knowledge base and machine learning to enrich data for decision-making. Moreover, few works exist that use semantics to enrich decision-making in the context of precision livestock.*

Resumo. *Este trabalho apresenta um mapeamento sistemático relacionado a Sistemas de Suporte à Decisão (SSD) na agricultura e pecuária de precisão, no contexto da agricultura inteligente. Por meio desse mapeamento sistemático, buscamos investigar soluções que são usadas para dar suporte à integração de dados de sensores IoT. Os resultados apontam para a utilização de base de conhecimento e aprendizado de máquina para enriquecimento de dados para tomada de decisão. Além disso, existem poucos trabalhos que utilizam a semântica como forma de enriquecer a tomada de decisão no contexto da pecuária de precisão.*

1. Introdução

Nos últimos anos, a Internet das Coisas (IoT) passou a conectar diversos dispositivos e sensores, o que gerou oportunidades em vários setores, dentre eles o da agricultura e pecuária [Zhai et al 2020]. Nesse sentido, a Internet das Coisas em Saúde Animal usa biossensores e software para monitorar e manter os registros da saúde do animal. Isso gera um grande volume de dados que pode ser interpretado por sistemas inteligentes e,

posteriormente, apoiar produtores e gestores na tomada de decisão relacionada à saúde e à produção do animal [Karthick, Sridhar e Pankajavalli 2020].

Estudos secundários recentes foram conduzidos para responder questões pertinentes a Sistemas de Suporte a Decisão no domínio da pecuária de precisão. O trabalho Villa-Henriksen et al. (2020), por exemplo, apresenta uma revisão dos padrões de interoperabilidade na pecuária de precisão que utilizam dados abertos. Esse estudo secundário discute os desafios de integrar os diversos dados gerados nas fazendas. Em [Bahlo et al. 2019] os autores revisam os problemas de rede, dispositivos físicos, aplicações e destacam o papel dos middlewares na integração de dados, uma vez que isso é um requisito fundamental para o Big Data e Inteligência Computacional. No trabalho de Zhai et al. (2020) é apresentado um survey, em que os autores ressaltam a importância da integração de dados no domínio da agricultura. Afirmam que, para a tomada de decisão, são necessários mais elementos como dados de estação meteorológica, devido à influência de temperatura e umidade na produção do animal. Nos estudos previamente apresentados, os autores abrangem a importância da integração de dados para tomada de decisão, mas não trazem uma perspectiva geral sobre como utilizar os dados coletados nas fazendas, e quais os modelos e arquiteturas de software são utilizados nos sistemas de suporte à decisão.

Dessa forma, este artigo apresenta um mapeamento sistemático cujo objetivo é responder a quatro questões de pesquisa definidas a seguir: (i) **RQ1: Como integrar os dados dos Sistemas de Suporte à Decisão provenientes dos dispositivos de IoT?** Através dessa pergunta, buscamos investigar o estado da arte sobre a integração de dados gerados nas fazendas pelos dispositivos IoT para dar suporte às decisões; (ii) **RQ2: De que forma a semântica dos dados é tratada para integrar os Sistemas de Suporte à Decisão?** Com essa pergunta buscamos identificar, nos resultados de RQ1, quais as técnicas semânticas de integração estão sendo utilizadas para enriquecer as decisões; (iii) **RQ3: Quais modelos inteligentes são usados para apoiar os Sistemas de Suporte à Decisão na agricultura e pecuária de precisão?** Dada a crescente utilização da inteligência computacional, apontada nos trabalhos secundários, essa questão visa explorar os modelos, técnicas e algoritmos inteligentes que estão sendo utilizados na smart farming; (iv) **RQ4: Como a computação ubíqua está sendo aplicada na agricultura e pecuária de precisão?** Considerando que a Internet das Coisas vem sendo amplamente utilizada na Agricultura 4.0, como mostra os estudos secundários, o intuito dessa questão é analisar como as soluções utilizadas contribuem para o avanço da computação ubíqua nas fazendas, seja através do monitoramento através de sensores, no uso do celular, de "drones" ou de dispositivos inteligentes.

Através de um mapeamento sistemático, nosso foco é elencar as técnicas e abordagens que são usadas para dar suporte à integração de dados nas fazendas para que os Sistemas de Suporte à Decisão (SSD) possam apoiar as decisões no contexto da agricultura e pecuária de precisão. Este artigo está organizado da seguinte forma: a Seção 2 discute a metodologia da pesquisa. Na Seção 3 apresentamos o processo de coleta de dados. Em seguida, discutimos os resultados obtidos e, na Seção 5, as apresentamos as considerações finais.

2. Metodologia de Pesquisa

A metodologia utilizada neste trabalho segue as orientações propostas por Kitchenham (2004). Durante a condução do mapeamento, foi utilizado a ferramenta de suporte denominada Parsifal[1]. O detalhamento completo do mapeamento, como questões de qualidade e critérios de inclusão e exclusão, está disponível online[2].

2.1. Etapas de Pesquisa

De acordo com as recomendações de Kitchenham (2004), a pesquisa foi dividida em três partes, planejamento, condução e relatório. Durante o planejamento foram identificadas as necessidades para revisão, as questões de pesquisa que deveriam ser respondidas e o protocolo a ser seguido. Feito isso, iniciou-se o processo de condução, através do qual identificamos e selecionamos os estudos, e executamos as técnicas de *backward* e *forward snowballing*, conforme a técnica de busca híbrida [Mourão et al. 2017]. A técnica de busca híbrida consiste em executar a *string* de busca um uma base que indexa outras bases de dados, como a Scopus. Em seguida, executar a técnica de *backward* e *forward snowballing* nos artigos retornados pela *string*. Após isso, foi realizada uma extração e síntese dos dados obtidos para uma análise de qualidade, e apresentamos os resultados.

2.2. Termos de Pesquisa

Os termos de pesquisa foram definidos considerando a População, Intervenção, Comparação, Resultado e Contexto (PICOC) [Kitchenham 2007], a fim de identificar palavras-chave, sinônimos e construir a *string* de busca com base na combinação dos termos-chave. Comparação não é relevante neste trabalho, uma vez que este mapeamento é um estudo exploratório. Para selecionar os artigos a serem analisados a fim de responder às questões de pesquisa, alguns critérios foram definidos. Os critérios de inclusão visaram adicionar aqueles trabalhos que apresentam arquiteturas, integração de dados e modelos inteligentes para apoiar as decisões na agricultura e pecuária de precisão, e estudos em inglês. Os critérios de exclusão adotados foram: (i) capítulos de livro, (ii) estudos publicados apenas como resumo, (iii) estudos cuja versão é mais antiga em relação a outro estudo já considerado, (iv) artigos que não estivessem em inglês, (v) trabalhos que não apresenta um resumo e (v) as palavras-chave não estão presentes no resumo. Mais detalhes do processo de seleção podem ser vistos em: (*https://files.fm/f/65d7wxg3p*).

2.3. String de Busca

A busca foi feita na base de dados eletrônica da Scopus, uma vez que ela indexa diversas outras bases, possibilitando a execução de apenas uma *string* de busca [Mourão et al. 2017]. Um conjunto de estudos primários (artigos de controle) [Villa-Henriksen et al. 2020] e [Helfer et al. 2019] foi definido, como sugerido em [Budgen et al. 2008], para a precisão da *string* de busca nas bases de dados selecionadas e se a pesquisa recuperou estudos relevantes. As palavras-chave desses artigos foram analisadas para encontrar novos termos relevantes a serem incluídos como parte da *string* de busca. Especialistas

[1] https://parsif.al/

[2] https://files.fm/f/65d7wxg3p

na área de Sistemas de Suporte à Decisão também participaram da revisão dos termos da *string*. A *string* de busca final foi descrita da seguinte maneira:

("internet of things" OR iot) AND (dss OR "decision support system") AND (agriculture OR livestock).

3. Coleta de dados

O primeiro passo nesta fase foi executar a *string* de busca, considerando as fontes selecionadas. Posteriormente, foi realizado o processo de seleção dos estudos, ilustrado na Figura 1, que compreendeu 4 etapas principais.

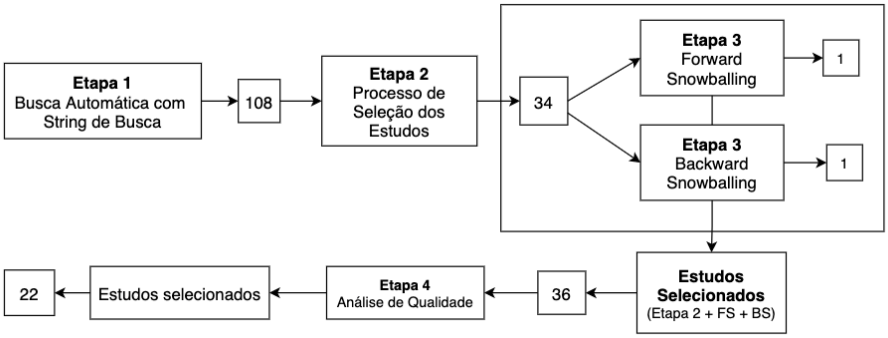


Figura 1: Processo de coleta de dados [Mourão et al. 2017]

Na Etapa 1 os resultados da busca nas bases de dados foram mesclados no Parsifal, e os trabalhos duplicados foram removidos, totalizando 108 artigos. A Etapa 2 envolveu a análise do título e do resumo considerando os critérios de inclusão/exclusão. Aqueles artigos que não abordavam Sistemas de Suporte à Decisão e IoT foram excluídos. Foram lidas as seções de Introdução, Referencial Teórico e Conclusão dos artigos. Como resultado, 74 artigos (68%) foram excluídos e 34 (32%) incluídos.

De acordo com as diretrizes propostas em [Kitchenham 2004], os pesquisadores devem desenvolver listas de verificação de qualidade para avaliar os estudos individuais. Durante a Etapa 3, aplicou-se o *backward e forward snowballing* [Webster e Watson 2002] aos 34 artigos selecionados resultando em mais 1 artigo pelo *backward snowballing* e 1 pelo *forward snowballing*.

Na Etapa 4, os trabalhos foram lidos na íntegra e sua qualidade foi avaliada por meio da lista de perguntas de avaliação de qualidade, que foi personalizada para o contexto deste mapeamento. Nesta etapa também executamos a extração de dados, verificando qual abordagem as soluções utilizaram para apoiar Sistemas de Suporte à Decisão e sua aplicação na agricultura ou pecuária de precisão. Esta seleção foi realizada por especialistas nas áreas de Sistemas de Suporte à Decisão. Ao final da etapa 4, dos 36 artigos, 22 (64%) foram incluídos. Esta redução nos estudos pode ser justificada pelos seguintes aspectos: (i) ausência de propostas de soluções para apoiar a tomada de decisão; (ii) muitos estudos discutiram a relevância da integração no contexto de IoT e tomada de decisão, sem introduzir uma solução para sua realização; e (iii) a pontuação na lista de verificação de avaliação de qualidade. Uma vez mapeados os trabalhos, foi necessária uma análise mais profunda a fim de identificar, avaliar e interpretar os estudos selecionados para responder às questões de pesquisa do mapeamento sistemático.

4. Resultados

De acordo com os resultados encontrados no mapeamento é possível observar que a utilização de web semântica e bases de conhecimento semântico estão sendo amplamente utilizados. A aplicação dessas soluções varia desde a irrigação de plantações e fertilização do solo, ao cuidado e manejo dos animais para que possam aumentar sua produção. Cerca de 11 artigos (50%) selecionados para o mapeamento abordam alguma técnica de web semântica, seja utilização de ontologias para inferir novos conhecimentos e marcações semânticas para análise automática de algoritmos inteligentes. Os demais artigos variam sua abordagem em relação à utilização de inteligência computacional. Redes neurais, *Support Vector Machine* (SVM), aprendizado de máquina e aprendizado profundo também são algumas das técnicas utilizadas. A Tabela 1 sintetiza os achados do mapeamento.

Tabela 1: Resumo das descobertas do relatório de mapeamento.

Artigos	Aplicação
S1	Consciência de Contexto e Middleware
S2	Aprendizado de máquina
S3, S18	Ontologias e Web Semântica
S4, S5	Inteligência computacional com árvore de decisão
S6, S15, S17, S22	Web Semântica
S7, S8, S12, S14, S20	Base de conhecimento semântica
S9	Inteligência computacional com redes de Petri
S10, S11, S19	Análise Estatística
S13	Aprendizado profundo
S5, S16	Inteligência computacional com SVM
S21	Inteligência computacional com Redes neurais

RQ1: Como integrar os dados dos Sistemas de Suporte à Decisão provenientes dos dispositivos de IoT? As pesquisas conduzidas em [S6, S15, S17, S19, S22] demonstram a utilização de uma rede de sensores sem fio (WSN) para integrar todos os sensores IoT do campo, a fim de fornecer o maior conjunto de dados para os sistemas de suporte à decisão. Em [S15] é proposto um protótipo de aplicativo móvel integrado com a rede de sensores sem fio, capaz de emitir alertas via SMS para que os produtores possam tomar decisões rapidamente sobre a situação do solo. Kamath et al. [S17] descrevem a implementação de uma rede de sensores sem fio para monitorar culturas de arroz com captura de imagem. Os autores fizeram a integração do sistema com o RabbitMQ [S17], um serviço de mensageria capaz de persistir os dados dos sensores. Em [S19] são propostos um modelo conceitual e projeto de sistema para apoio à decisão na smart farming. O trabalho utiliza um middleware para dar suporte à comunicação e à interoperabilidade de dispositivos inteligentes, a detecção e caracterização de eventos em tempo real e a coleta de eventos para análise de big data. Khanna e Kaur [S22] utilizam

uma rede de sensores sem fio integrados com servidores em nuvem. Dessa forma, os produtores podem monitorar diferentes informações das plantações, como temperatura e umidade do solo, e intervir nos locais físicos para equilibrar o ambiente. A pesquisa proposta em [S3] descreve o Agri-IoT, um framework baseado em IoT capaz de integrar via Global Sensor Network (GSN) vários fluxos de dados, fornecendo um processamento semântico completo.

RQ2: De que forma a semântica dos dados é tratada para integrar os Sistemas de Suporte à Decisão? Os trabalhos [S1, S3, S18] propõem o uso de uma camada que trata da semântica para interpretar os dados de dispositivos IoT, e [S1] ainda agrega dados de contexto como temperatura, pressão, velocidade do vento, umidade, raios UV, entre outros dados que podem influenciar o comportamento dos animais, das colheitas e plantações. Uma vez que esses dados são levados em consideração, é possível que os sistemas indiquem decisões mais apropriadas no manejo das fazendas. Kamilaris et al. [S3] exploram como a integração semântica de informações de várias fontes, tais como: sensores, mídia social, fazendas conectadas, alertas governamentais, regulamentações, entre outras, pode aumentar a produtividade em smart farming. Já em [S5] os autores propõem o uso de web semântica para enriquecer os dados gerados pelos dispositivos IoT e, em seguida, aplicar algoritmos de aprendizado de máquina para otimizar a fertilização do solo. Com isso, a evolução da agropecuária tradicional para agricultura e pecuária de precisão, sistemas conectados e baseados em conhecimento semântico, têm transformado o ambiente físico e tecnológico das propriedades rurais. Os proprietários, produtores e veterinários ganham um novo papel nesse contexto: enriquecer os sistemas através de uma base de conhecimento compartilhado, de tal forma que as decisões sejam cada vez mais eficientes.

RQ3: Quais modelos inteligentes são usados para apoiar os Sistemas de Suporte à Decisão na agricultura e pecuária de precisão? Os artigos [S2, S4, S5, S9, S13, S16, S21] apresentaram soluções de inteligência computacional variando entre as técnicas de SVM, árvore de decisão, redes de Petri e redes neurais. Através da coleta dos dados gerados pelos dispositivos IoT no campo, é possível utilizar técnicas de aprendizado de máquina para predição de eventos como colheita, irrigação e fertilização do solo [S5, 16, S21]. Todos esses trabalhos apresentaram a computação na nuvem como componente importante da solução, uma vez que a comunicação entre sensores, sistemas e repositórios se dá através da web. Os trabalhos [S7, S8, S11, S12, S14, S20] utilizam bases de conhecimento com regras específicas do domínio da agricultura. Essas regras servem de bases tanto para os sistemas indicarem decisões, quanto para que os algoritmos inteligentes tenham os parâmetros corretos para realizar uma predição confiável. Nesses estudos é possível notar variações de abordagem, tais como a automatização de decisões através da predição de eventos indicados pelos algoritmos de aprendizado de máquina [S2, S9, S13, S5, S16, S21], até conhecimentos que podem ser inferidos utilizando web semântica [S6, S15, S17, S22]. Adicionalmente, camadas que tratam da semântica são utilizadas para refinar a análise de aprendizado de máquina para apoiar Sistemas de Suporte à Decisão, tornando cada vez mais ricas e confiáveis as decisões.

RQ4: Como a computação ubíqua está sendo aplicada na pecuária de precisão? Após a análise dos estudos, foi possível perceber como a computação ubíqua está presente nas fazendas. Em [S10] os autores propõem duas aplicações móveis integradas, uma para o produtor e outra para o vendedor de produtos agrícolas. Na pesquisa, o produtor pode monitorar e enviar os dados de sua plantação para o vendedor,

que por sua vez, pode indicar o melhor fertilizante, pesticida e quantidade de água necessária. Alguns trabalhos, como [S3], demonstram a utilização de "drones" para coleta de imagens de ambientes agrícolas. Isto pode ser estendido para localizar animais no pasto, monitorar suas atividades de movimentação e detectar animais doentes. Dessa maneira, os "drones" podem agilizar o processo de decisão fornecendo informações em tempo real. A criação de uma rede de sensores sem fio também contribui para que um sistema possa gerenciar mais de uma fazenda geograficamente distribuída, podendo indicar diferentes decisões simultaneamente de acordo com os dados coletados pelos sensores.

4.1. Ameaças à Validade

Este mapeamento teve como objetivo explorar questões relacionadas aos Sistemas de Suporte à Decisão no contexto da agricultura e pecuária de precisão. Entretanto, algumas ameaças à validade podem ser ressaltadas. Em relação à *string* de busca, mesmo utilizando um processo de revisão dos termos é possível que algum termo não tenha sido considerado para o contexto de agricultura de precisão. Para mitigar essa ameaça, utilizamos a técnica de *snowballing*.

A utilização da base Scopus também representa uma ameaça à validade. Outras bases não indexadas pela base utilizada podem conter pesquisas relevantes para este mapeamento. Para mitigar essa ameaça caberia explorar outras bases específicas, sobretudo voltadas para as pesquisas no agronegócio. Entretanto, consideramos que as bases utilizadas nesta pesquisa foram capazes de oferecer uma ampla visão geral do estado da arte das pesquisas relacionadas à Sistemas de Suporte à Decisão voltadas para o Agronegócio.

5. Conclusão e Trabalhos Futuros

Este artigo apresentou um mapeamento sistemático para identificar, classificar e analisar soluções computacionais em Sistemas de Suporte à Decisão na agricultura e pecuária de precisão. Inicialmente, 108 artigos foram identificados pela fonte selecionada, os quais foram filtrados, e isso resultou em 34 artigos. Após essa seleção, foi executada a técnica de *snowballing* agregando mais 2 novos artigos, totalizando 36 trabalhos. Após a análise de qualidade, o resultado totalizou 22 artigos.

Como contribuição, este mapeamento oferece oportunidades de pesquisas a partir da busca pelo estado da arte das pesquisas em SSD. Ao avançar no conhecimento sobre o tema, apresentamos oportunidades e técnicas de integração utilizadas nos trabalhos, demonstrando a importância de se utilizar dados de sensores IoT para apoiar as decisões. Como resultado, decisões complexas voltadas para o agronegócio poderão ser enriquecidas com informações existentes em outros sistemas utilizados pelas propriedades rurais. Para trabalhos futuros, outras questões de pesquisa podem ser exploradas sobretudo no que diz respeito à utilização da integração entre sistemas e dados gerados em dispositivos da IoT, para o agronegócio. Adicionalmente, é importante desenvolver pesquisas que tratam da proveniência da decisão, levando em consideração o contexto dos dados de IoT, para enriquecer e melhorar os SSD na agropecuária.

Agradecimentos

Gostaríamos de agradecer aos pesquisadores que participaram da avaliação e expressar nossos agradecimentos ao Centro Nacional de Pesquisa da Pecuária de Leite (Embrapa Gado de Leite) da Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Este trabalho foi parcialmente financiado pela UFJF / Brasil, CNPq / Brasil (bolsa 311595 / 2019-7) e FAPEMIG / Brasil (bolsa APQ-02685-17), (bolsa APQ-02194-18). Este estudo também foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

Bahlo, C., Dahlhaus, P., Thompson, H., Trotter, M. (2019) "The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review". Computers and electronics in agriculture. Vol. 156, pp. 459-466, doi: 10.1016/j.compag.2018.12.007.

Budgen, D., Turner, M., Brereton P., Kitchenham, B. (2008) "Using mapping studies in software engineering," in Proceedings of PPIG.

Helfer, G. A., Barbosa, J. L., Costa, A. B. D., Martini, B. G., Santos, R. D. (2019) "A model for productivity and soil fertility prediction oriented to ubiquitous agriculture". In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, pp. 489–492, doi: 10.1145/3323503.3360637.

Karthick, G. S., Sridhar, M., Pankajavalli, P. B. (2020) "Internet of things in animal healthcare (IoTAH): review of recent advancements in architecture, sensing technologies and real-time monitoring". SN Computer Science, 1(5).

Kitchenham B. (2004) "Procedures for performing systematic review". Vol. 33, Keele University, Keele, UK.

Kitchenham, B., Charters S. (2007) "Guidelines for performing Systematic Literature Reviews in Software Engineering". Keele University and University of Durham, EBSE Technical Report Version 2.3.

Mourão E., Kalinowski M., Murta L., Mendes E., Wohlin C. (2017) "Investigating the Use of a Hybrid Search Strategy for Systematic Reviews," 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 193-198, doi: 10.1109/ESEM.2017.30.

Villa-Henriksen, A., Edwards, G. T., Pesonen, L. A., Green, O., Sørensen, C. A. G. (2020) "Internet of Things in arable farming: Implementation, applications, challenges and potential". Biosystems Engineering, Volume 191, pp. 60-84, doi: 10.1016/j.biosystemseng.2019.12.013.

Webster, J., Watson, R. T. (2002). "Analyzing the past to prepare for the future: Writing a literature review". MIS quarterly, Vol. 26, No.2, pp. xiii-xxiii, doi: http://www.jstor.org/stable/413231.

Zhai Z., Martínez F. J., Beltran V., Martínez L. N. (2020) "Decision support systems for agriculture 4.0: Survey and challenges". Computers and Electronics in Agriculture, Volume 170, March 2020, 105256, doi: 10.1016/j.compag.2020.105256.

ESTUDOS SELECIONADOS

[S1] Symeonaki E., Arvanitis K., Piromalis D. (2020) "A Context-Aware Middleware Cloud Approach for Integrating Precision Farming Facilities into the IoT toward Agriculture 4.0". Applied Sciences.10(3), 813; https://doi.org/10.3390/app10030813.

[S2] Kakamoukas, G. et al., (2019) "A Multi-collective, IoT-enabled, Adaptive Smart Farming Architecture," 2019 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1-6, doi: 10.1109/IST48021.2019.9010236.

[S3] Kamilaris, A., Gao, F., Prenafeta-Boldu, F. X., Ali, M. I. (2016) "Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications," 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), pp. 442-447, doi: 10.1109/WF-IoT.2016.7845467.

[S4] Rezk, N. G., Hemdan, E. E., Attia, A., El-Sayed, A., Elrashidy, M.A. (2021) "An efficient IoT based smart farming system using machine learning algorithms". Multimedia Tools and Applications, vol. 80, pp. 773-797.

[S5] Sowmya B. J., Krishna Chaitanya S., Seema S., K.G. Srinivasa, (2020) "Data Analytic Techniques for Developing Decision Support System on Agrometeorological Parameters for Farmers," International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), IGI Global, vol. 14(2), pp. 92-107. doi: 10.4018/IJCINI.2020040106.

[S6] Visconti P., de Fazio R., Velázquez R., Del-Valle-Soto C., Giannoccaro N. I. (2020) "Development of Sensors-Based Agri-Food Traceability System Remotely Managed by a Software Platform for Optimized Farm Management". Sensors. 20(13), 3632. doi: 10.3390/s20133632.

[S7] Mikhaylenko, I. M., Timoshin, V. N. (2019). "Expert strategic management systems in precision farming". In Journal of Physics: Conference Series. IOP Publishing. doi: 10.1088/1742-6596/1419/1/012030.

[S8] Rekha, P., Rangan, V. P., Ramesh, M. V., Nibi, K. V. (2017) "High yield groundnut agronomy: An IoT based precision farming framework," 2017 IEEE Global Humanitarian Technology Conference (GHTC). pp. 1-5, doi: 10.1109/GHTC.2017.8239287.

[S9] Borisenko, I., Gangur, M.. Kobylko, A., Romanov, D., Smirnova, O., (2019) "Information support system for management decisions in the agricultural sector", 2019 International Scientific Conference on Agricultural Machinery Industry, Volume 403, Issue 1. doi: 10.1088/1755-1315/403/1/012083.

[S10] Dabre, K. R., Lopes, H. R., D'monte, S. S. (2018) "Intelligent Decision Support System for Smart Agriculture," 2018 International Conference on Smart City and Emerging Technology. pp. 1-6, doi: 10.1109/ICSCET.2018.8537275.

[S11] Yusianto, R. R., Marimin, Suprihatin, Hardjomidjojo, H. (2020) "Intelligent Spatial Decision Support System Concept in the Potato Agro-Industry Supply Chain," 2020 International Conference on Computer Science and Its Application in Agriculture (ICOSICA). pp. 1-7, doi: 10.1109/ICOSICA49951.2020.9243233.

[S12] Loret, N., Affinito, A., Bonanomi, G. (2020). "Introducing Evja-Rugged Intelligent Support System for precision farming". ACTA IMEKO, 9(2), pp. 83 – 88.

[S13] Kale, A. P., Sonavane, S. P. (2019) “IoT based Smart Farming: Feature subset selection for optimized high-dimensional data using improved GA based approach for ELM”. Computers and Electronics in Agriculture, Volume 161, pp. 225-232. doi: 10.1016/j.compag.2018.04.027

[S14] Zhang X., Zhang J., Li L., Zhang Y., Yang G. (2017) “Monitoring Citrus Soil Moisture and Nutrients Using an IoT Based System”. Sensors (Basel). 2017;17(3):447. doi:10.3390/s17030447.

[S15] Karim, F., Karim, F., Ali frihida, (2017) “Monitoring system using web of things in precision agriculture”. Procedia Computer Science, Volume 110, pp. 402-409. doi: 10.1016/j.procs.2017.06.083.

[S16] Tripathy, P. K., Tripathy, A. K., Agarwal, A., Mohanty, S. P. (2021) "MyGreen: An IoT-Enabled Smart Greenhouse for Sustainable Agriculture," IEEE Consumer Electronics Magazine. vol. 10, no. 4, pp. 57-62, 1 July 2021, doi: 10.1109/MCE.2021.3055930.

[S17] Kamath, R., Balachandra, M., Prabhu, S. (2019) "Raspberry Pi as Visual Sensor Nodes in Precision Agriculture: A Study," in IEEE Access, vol. 7, pp. 45110-45122, doi: 10.1109/ACCESS.2019.2908846.

[S18] Rotondi, D., Straniero, L., Saltarella, M., Balducci, F., Impedovo, D., Pirlo, G. (2019) "Semantics for Wastewater Reuse in Agriculture*," 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 598-603, doi: 10.1109/SMC.2019.8913949.

[S19] Suakanto, S., Engel, V. J. L., Hutagalung, M., Angela, D. (2016) "Sensor networks data acquisition and task management for decision support of smart farming," 2016 International Conference on Information Technology Systems and Innovation (ICITSI). pp. 1-5, doi: 10.1109/ICITSI.2016.7858233.

[S20] Sakthi, U., Rose, J. D. (2020) "Smart Agricultural Knowledge Discovery System using IoT Technology and Fog Computing," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). pp. 48-53, doi: 10.1109/ICSSIT48917.2020.9214102.

[S21] Balan T., Dumitru C., Dudnik G., Alessi E., Lesecq S., Correvon M., Passaniti F., Licciardello A. (2020) “Smart Multi-Sensor Platform for Analytics and Social Decision Support in Agriculture”. Sensors. 20(15), 4127. doi: 10.3390/s20154127.

[S22] Khanna, A., Kaur, S. (2020) “Wireless Sensor and Actuator Network (s) and its significant impact on Agricultural domain”.

2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 384-389, doi: 10.1109/PDGC50313.2020.9315822.



SQUADRO-i: Uma plataforma para o monitoramento da sustentabilidade de sistemas produtivos agropecuários

Kimberlly da Rosa Moreles[1], Vinícius do Nascimento Lampert[2], Érico Marcelo Hoff do Amaral[1]

[1]Programa de Pós-graduação em Computação Aplicada
Universidade Federal do Pampa (UNIPAMPA)
Bagé – RS – Brasil

[2]Embrapa Pecuária Sul (CPPSUL)
Bagé, RS – Brasil

kim.moreles@hotmail.com, vinicius.lampert@embrapa.br

erico.amaral@unipampa.edu.br

Abstract. *The demand for sustainability assessments of agricultural systems has been growing in recent years. Based on a specific methodology for dairy production, a generic platform called SQUADRO-i has been developed, aimed at diagnosing the sustainability of different types of systems. This study presents the SQUADRO-i pilot project, a software that allows the construction of sustainability indicators and assessments. In the pilot project, it is already possible to build quantitative indicators with their respective parameters and calculate their sustainability index. The project is still in progress, but the possibility of the platform adapting to other assessment methodologies has already been raised.*

Resumo. *A demanda por avaliações de sustentabilidade de sistemas agropecuários tem sido crescente nos últimos anos. A partir de uma metodologia específica para a produção leiteira, têm-se desenvolvido uma plataforma genérica denominada SQUADRO-i, visando o diagnóstico de sustentabilidade de diferentes tipos de sistemas. Este estudo apresenta o projeto-piloto da SQUADRO-i, um software que permite a construção de indicadores e de avaliações de sustentabilidade. No projeto-piloto já é possível construir indicadores quantitativos com seus respectivos parâmetros e realizar o cálculo de índice de sustentabilidade dos mesmos. O projeto ainda se encontra em andamento, porém já levantou-se a possibilidade da plataforma se adaptar a outras metodologias de avaliação.*

1. Introdução

As preocupações com a sustentabilidade datam da década de 1970, porém foi o Relatório de Brundtland que trouxe a conceituação de desenvolvimento sustentável [Imperatives 1987]. Neste conceito, o desenvolvimento busca suprir as necessidades da geração atual sem prejudicar os recursos necessários às gerações futuras [Imperatives 1987, Bell e Morse 2012, das Nações Unidas ONU 2016], visando ir ao encontro das necessidades básicas humanas e possibilitando condições dignas de vida. Neste contexto, a agricultura possui um papel de destaque por demandar de grandes áreas para sua produção [Bell e Morse 2012] e por possuir um papel importante na economia brasileira, representando em torno de 20% do PIB nacional [de Oliveira e Carraro 2019].

Nos últimos anos, a demanda por avaliações de sustentabilidade de sistemas agropecuários tem sido crescente e para esta tarefa popularizou-se o uso de indicadores e índices de sustentabilidade. Os indicadores de desenvolvimento sustentável são instrumentos que medem as modificações de um sistema [Deponti et al. 2002] e precisam ser construídos de forma a contribuir para uma sustentabilidade autorregulada [Sitarz 1993]. A partir deles é possível determinar a necessidade de mudanças na condução de uma determinada estratégia ou até mesmo comparar o desempenho deste indicador em diferentes localidades e contextos. Para esta comparação, empregam-se índices que normalizam esses indicadores a uma mesma unidade de medida e que quando construídos com dados relevantes são capazes de inclusive servir de referência para o direcionamento de investimentos [Negra et al. 2020]. As avaliações construídas precisam levar em consideração não somente as características inerentes do sistema produtivo, como também os objetivos dos avaliadores. Neste sentido, apesar de existirem amplas possibilidades de metodologias de avaliação, ainda há uma escassez de ferramentas que se adaptem para o diagnóstico do desenvolvimento sustentável de sistemas produtivos e que gerem informações relevantes de forma eficaz e não onerosa.

O objetivo desta pesquisa é propor uma plataforma genérica que por meio de uma ferramenta computacional auxilie na sistematização de indicadores de sustentabilidade, para que desta forma os diagnósticos dos diferentes sistemas possam ser feitos conforme suas características e necessidades. A plataforma em construção, nomeada SQUADRO-i, se propõe a ser um ambiente onde os usuários poderão construir avaliações de sustentabilidade personalizadas, onde também será possível gerar um formulário de coleta de dados a campo que estará disponível para acesso em smartphones. A partir desta coleta, a SQUADRO-i disponibilizará o resultado não somente do índice geral da avaliação, como também os índices dos indicadores e dimensões avaliadas, podendo estas informações serem disponibilizadas a demais interessados autorizados pelo gerenciador da avaliação.

Em uma perspectiva técnica este trabalho visa contribuir para que o ambiente avaliativo seja adaptável, sendo possível delimitar as variáveis que melhor definirem o sistema estudado, sem ser necessário construir uma nova ferramenta específica para cada diagnóstico, e também, proporcionar um ambiente de coleta de dados dinâmico via aplicação móvel. Já em um panorama científico, este estudo visa criar algoritmos específicos para o cálculo de índices de indicadores e também para o cálculo de demais índices necessários à avaliação.

Neste artigo são dispostos os resultados obtidos até o momento com a plataforma SQUADRO-i, e as próximas etapas a serem concluídas e perspectivas futuras de estudo

com a mesma. Para tal, na seção 2 é descrita a metodologia deste estudo, na seção 3 é apresentado o referencial teórico, na seção 4 é expressa a proposta da plataforma e os resultados já alcançados, e por fim, na seção 5 são elencadas as conclusões obtidas até o momento.

2. Metodologia

Quanto a sua natureza a pesquisa é classificada como pesquisa aplicada, pois pretende resolver um problema específico do desenvolvimento sustentável quanto a utilização de um sistema computacional para a sua avaliação. Empregou-se o método dedutivo, pois é pretendido avaliar a capacidade da ferramenta construída gerenciar o processo de avaliação de indicadores de forma genérica. Quanto aos seus objetivos a pesquisa é classificada como descritiva por pretender analisar o comportamento da ferramenta diante indicadores que avaliam diferentes sistemas produtivos agropecuários. Do ponto de vista da forma de abordagem do problema ela é classificada como quantitativa devido à plataforma manipular valores numéricos e fórmulas. Quanto aos procedimentos técnicos a pesquisa é classificada como bibliográfica, documental e experimental.

As etapas empregadas na condução da pesquisa até o momento e que levam à construção do projeto-piloto seguem a linha do tempo apresentada na Figura 1. Determinou-se que inicialmente, seguindo a engenharia de requisitos, seriam realizadas entrevistas junto a um profissional da Embrapa para a delimitação do problema de pesquisa e para elencar os requisitos necessários à solução. Durante estas, ficou convencionado o uso da metodologia mencionada no estudo de [Brum et al. 2017] como base para a implementação da solução, a mesma teve como objetivo o diagnóstico do desenvolvimento sustentável de propriedades familiares com atividade leiteira. Esta possui uma planilha Excel que foi utilizada como auxílio não só para o desenvolvimento, como também para a aplicação de testes na SQUADRO-i. O profissional entrevistado foi o responsável pelo desenvolvimento da metodologia modelada na planilha eletrônica. A fórmula de interpolação linear utilizada nesta planilha para o cálculo do índice dos indicadores também serviu como base para o algoritmo de mesma função na plataforma SQUADRO-i.

Após a etapa das entrevistas, determinou-se que iriam ser realizadas pesquisas abrangendo tanto a temática de indicadores quanto a de diagnósticos de sustentabilidade. Para a etapa seguinte ficou especificada a escrita de um documento de requisitos, contendo a modelagem do sistema, sua arquitetura e seus requisitos, e a partir deste documento, a construção de protótipos de tela necessários à validação dos requisitos. Por fim, levantou-se a necessidade de implementação de um projeto-piloto da plataforma para que se pudesse verificar a possibilidade de implementação das funcionalidades iniciais e cálculos essenciais em tempo hábil. Foi dado seguimento à implementação da aplicação e nela codificado o algoritmo de cálculo de índice dos indicadores, este calculando o índice com base nos parâmetros definidos para o dito indicador, sendo que qualquer valor intermediário aos parâmetros especificados é calculado para um valor aproximado, ou seja, uma interpolação linear. Para que as funcionalidades do projeto-piloto da plataforma fossem validadas, os resultados do mesmo foram comparados com os obtidos na ferramenta Excel. Para tal, procedeu-se a construção da cópia de um indicador, constante na ferramenta Excel, na plataforma SQUADRO-i, contendo a mesma estrutura e parâmetros de cálculo de índice, e posteriormente injetado no banco de dados o valor que este indicador



Figura 1. Etapas da pesquisa

apresentou na coleta de dados a campo para uma unidade avaliativa aleatória.

3. Referencial Teórico

A sustentabilidade engloba a capacidade de um sistema utilizar recursos de forma que este não seja perdido ou venha a escassear. Já o desenvolvimento sustentável faz menção ao uso de recursos no presente de forma que este uso não afete as necessidades das gerações futuras [Imperatives 1987]. Quando se fala neste último é necessário levar em consideração as dimensões ambiental, econômica e social [Bell e Morse 2012], logo é imprescindível que estes aspectos estejam presentes na construção de uma avaliação de sustentabilidade.

Ao se estudar um determinado sistema agrícola ou outro sistema qualquer quanto a sua sustentabilidade é necessário levar em consideração não somente estes aspectos, como também as características inerentes deste sistema. Estes aspectos, ou dimensões, por vezes são complementados por outras como, por exemplo, a dimensão produtiva e a dimensão institucional, a depender dos objetivos da avaliação em seguimento.

Outra característica denotada como importante para a construção de uma avaliação de sustentabilidade é a temporalidade. Segundo [López-Ridaura et al. 2002], este tipo de avaliação só pode se dar de forma comparativa e a comparação temporal é uma das alternativas. Para [Zhen e Routray 2003] é importante avaliar não somente práticas específicas empregadas no contexto da fazenda com o uso de indicadores adequados, como também a janela de tempo em que a avaliação irá ocorrer. Assim, este tipo de avaliação necessita não somente de um objetivo claro de avaliação e de quais informações são necessárias, como também a especificação do período de tempo em que a mesma irá ocorrer.

Em relação aos indicadores envolvidos na avaliação vale ressaltar que existe a demanda de que estes reflitam o grau de sustentabilidade dos sistemas produtivos [López-Ridaura et al. 2002], porém grandes quantidades dos mesmos não denotam qualidade na avaliação. A seleção de indicadores está diretamente relacionada ao sistema a

ser avaliado e a região que a avaliação irá abranger tornando cada avaliação única, com poucas possibilidades de ser reaproveitada sem adaptações.

Há amplas possibilidades de metodologias de diagnóstico de sustentabilidade que são aplicáveis a sistemas produtivos agropecuários [Bohanec e Rajkovic 1990, Astier et al. 1999]. Destas metodologias, algumas contam com ferramentas computacionais específicas para uso, já outras necessitam da implementação de um sistema ou elaboração de uma planilha eletrônica.

Observa-se na literatura [Rodrigues et al. 2016, Lampert et al. 2018] que planilhas eletrônicas são utilizadas com certa frequência para diagnósticos de sustentabilidade, isto devido à flexibilidade proporcionada pelas mesmas em relação à adaptabilidade. Porém, também é observado que para este tipo de ferramenta há a necessidade de realização das coletas em formulários, e em uma etapa posterior, manualmente realizar a inserção dos dados. Também foram encontrados trabalhos que empregam sistemas computacionais [Chen et al. 2014, Lampert et al. 2015, Rezaei et al. 2018, Whitehead et al. 2020], alguns destes com soluções específicas e outros com a proposta de flexibilizar o processo avaliativo. Nestes últimos as principais limitações encontradas foram: uma seleção fechada de opções de personalização da avaliação e possibilidade de tratar apenas características qualitativas.

4. Proposta e Resultados

Na maioria das vezes, a construção de uma ferramenta de diagnóstico de sustentabilidade na agricultura envolve diversos interessados, gerando uma certa complexidade no processo de criação da mesma [Whitehead et al. 2020]. Envolver o usuário na construção, torna o sistema mais compreensível, adaptável e com mais chances de ser melhorado, sendo necessário que a usabilidade do mesmo seja garantida [Goumopoulos e Mavrommati 2020]. Porém, a complexidade de utilização é um fator-chave quando se fala na adesão desta ferramenta [Goumopoulos e Mavrommati 2020]. Ferramentas de avaliação podem abranger propósitos diversos e possuírem abordagens distintas [Ness et al. 2007], mas são as pessoas envolvidas na construção da ferramenta de sustentabilidade e as funcionalidades desenvolvidas para a mesma que irão definir o seu nível de sucesso [Goumopoulos e Mavrommati 2020].

Visando suprir a necessidade de construção de uma ferramenta específica para cada diagnóstico a ser realizado, neste estudo é apresentada a solução denominada SQUADRO-i. Esta irá possuir dois módulos: um módulo web contendo a parte gerencial de construção de uma avaliação de sustentabilidade e a possibilidade de visualização de resultados da avaliação em relatórios e gráficos, e um módulo móvel dedicado à coleta de dados a campo, substituindo assim a necessidade de formulários de papel.

A plataforma SQUADRO-i tem por base a metodologia mencionada por [Brum et al. 2017], onde uma avaliação é composta por dimensões e esta última por indicadores. Tanto estes, quanto as dimensões são ponderados por pesos, para que desta forma seja possível determinar indicadores e dimensões mais ou menos expressivos que os demais. Cada indicador também irá estar atrelado a um conjunto de parâmetros, estes serão o conjunto de regras que serão aplicadas para o cálculo de interpolação linear de seu índice. A estrutura completa de uma avaliação pode ser visualizada na Figura 2.

O presente trabalho teve início com a realização de entrevistas junto a um pro-

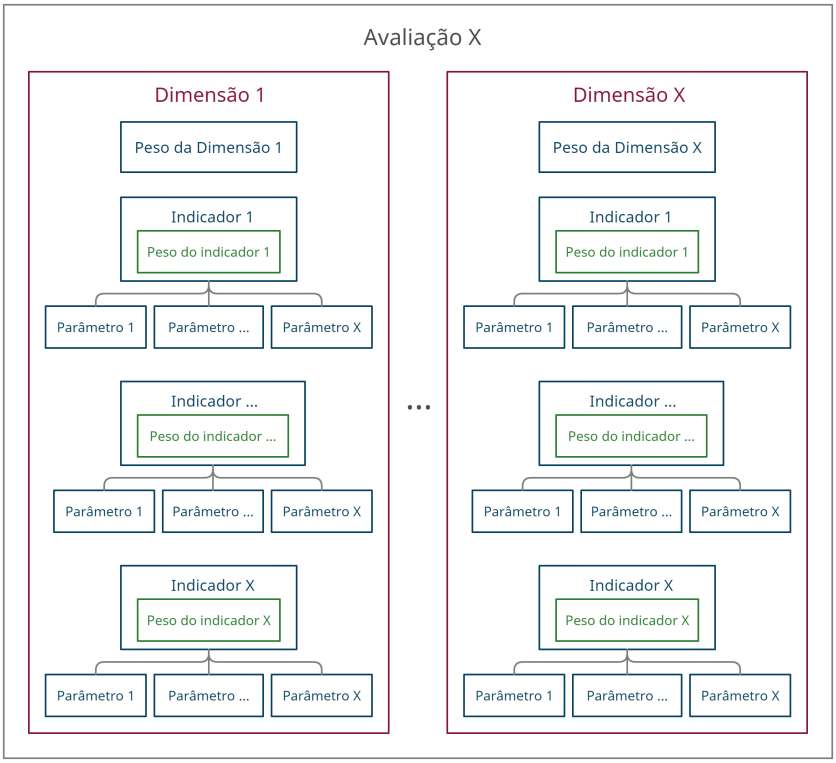


Figura 2. Estrutura da avaliação

fissional da Embrapa para que os requisitos necessários à plataforma fossem elencados. Logo após, para que a SQUADRO-i atenda adequadamente às necessidades levantadas, foram realizadas pesquisas bibliográficas e documentais sobre a temática de indicadores e de diagnóstico de sustentabilidade. Com estas informações, foi possível a construção de um documento de requisitos contendo a modelagem e arquitetura do sistema.

Em seguida, foi construído um protótipo da aplicação web para que os requisitos pudessem ser validados e a partir deste, iniciada a implementação do projeto-piloto que abrange as funcionalidades de inserção de indicadores quantitativos simples, componentes e parâmetros na SQUADRO-i, e também o algoritmo de cálculo do índice dos indicadores quantitativos simples. A partir do estudo e análise das tecnologias disponíveis no mercado, definiu-se como plataforma para a construção da solução web a linguagem PHP aliada às funcionalidades do JavaScript e o PostgreSQL para o gerenciamento do banco de dados.

Nesta etapa da pesquisa, a fim de validar o projeto-piloto, foram utilizados os dados e resultados da planilha Excel da metodologia mencionada por [Brum et al. 2017]. Para tal, comparou-se o resultado do cálculo do índice do indicador na SQUADRO-i com o resultado para o mesmo cálculo na planilha original do Excel.

Para apresentação desta fase da ferramenta elencou-se o indicador “POTENCIAL PRODUTIVO” e a unidade avaliativa “Nome 57”. Na Figura 3 é apresentado este indicador na planilha Excel. Como podemos ver, ele é avaliado por uma única variável, esta

chamada "Área Total", que por sua vez possui valor 10 para a unidade avaliativa "Nome 57". Também é possível observar na Figura 3 quais os parâmetros de notas utilizados para o cálculo do índice deste indicador, que para o valor 10 recebeu índice 0,30.

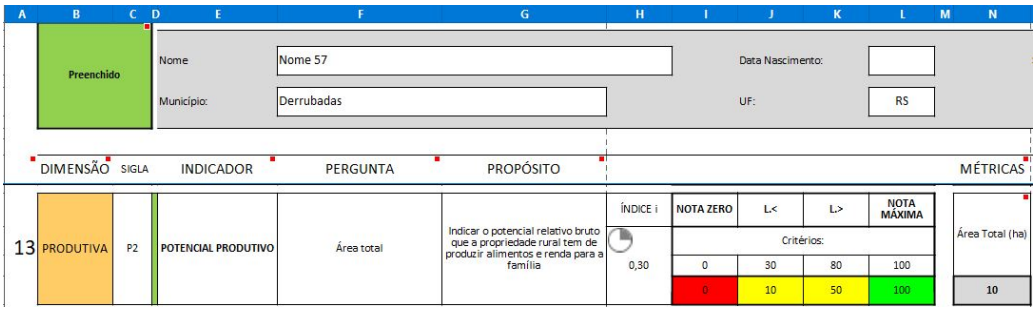


Figura 3. O indicador "POTENCIAL PRODUTIVO" para a unidade avaliativa "Nome 57" na ferramenta Excel da metodologia SQUADRO

Na plataforma SQUADRO-i procedeu-se à inserção do indicador "POTENCIAL PRODUTIVO", a atribuição dos mesmos parâmetros de cálculo de índice, presentes na planilha Excel para este, e a injeção dos mesmos valores recebidos pela unidade avaliativa "Nome 57" na coleta de dados a campo. Na Figura 4 é apresentado o indicador "POTENCIAL PRODUTIVO" com valor 10 e seu índice já calculado pelo clique no botão "Calcular índice". Nesta etapa o indicador obteve índice 0,3, o mesmo resultado obtido na planilha Excel.



Figura 4. O indicador "POTENCIAL PRODUTIVO" para a unidade avaliativa "Nome 57" na plataforma SQUADRO-i

Visto que o trabalho ainda se encontra em andamento e que o tipo de indicador testado não engloba todas as possibilidades necessárias à construção de uma avaliação de sustentabilidade, nas próximas etapas serão construídos e testados indicadores qualitativos e indicadores resultantes de fórmulas.

5. Conclusões

A construção de um sistema de avaliação de sustentabilidade é uma tarefa complexa que precisa envolver diferentes interesses e no contexto agropecuário envolver também as

necessidades do produtor rural. As ferramentas disponíveis, em geral, são construídas especificamente para o diagnóstico a ser realizado, demandando tempo também para a implementação da mesma. A SQUADRO-i é uma ferramenta que tem a proposta ser customizável para diferentes tipos de avaliação. Para tal, ela irá contar com um módulo web onde será possível construir e consultar resultados de uma avaliação de sustentabilidade e um módulo móvel que possibilitará a coleta de dados a campo sem a necessidade de formulários de papel.

Apesar da complexidade envolvida no processo de diagnóstico de sustentabilidade e de ser um projeto ainda em andamento, a plataforma já foi capaz de realizar o cálculo do índice dos indicadores, por meio de uma interpolação linear, para indicadores do tipo quantitativo simples. Os passos futuros envolvem a implementação das demais funcionalidades da ferramenta e a adequação da mesma para suportar outros tipos de indicadores.

A plataforma SQUADRO-i é adaptável a diversos tipos de avaliação, pois a mesma foi criada na intenção de ser genérica e customizável para diferentes cenários avaliativos. Inicialmente ela está sendo testada no monitoramento da sustentabilidade de sistemas produtivos agropecuários, porém nada impede que a mesma seja aplicada em outras áreas. É pretendido verificar também a possibilidade de esta plataforma ser capaz de se adequar a outras metodologias de avaliação como, por exemplo, a metodologia MESMIS [Astier et al. 1999, López-Ridaura et al. 2002].

Por fim, os autores agradecem o suporte financeiro fornecido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pelo programa de auxílio da pós-graduação (PAPG) da Universidade Federal do Pampa (UNIPAMPA).

Referências

Astier, M., Masera, O., e Lopez-Ridaura, S. (1999). *Sustentabilidad y manejo de recursos naturales: El Marco de evaluación MESMIS*.

Bell, S. e Morse, S. (2012). *Sustainability indicators: measuring the immeasurable?* Routledge.

Bohanec, M. e Rajkovic, V. (1990). DEX: An expert system shell for decision support. *Sistemica*, 1:145–157.

Brum, L. d. L., Lampert, V. d. N., Camargo, S. d. S., e Eickhoff, F. (2017). Aplicação de técnicas de mineração em dados de propriedades leiteiras do município de Derrubadas-RS. In *Embrapa Pecuária Sul-Artigo em anais de congresso (ALICE)*. In: SIMPÓSIO DA CIÊNCIA DO AGRONEGÓCIO, 5., 2017, Porto Alegre.

Chen, D., Thiede, S., Schudeleit, T., e Herrmann, C. (2014). A holistic and rapid sustainability assessment tool for manufacturing SMEs. *CIRP Annals*, 63(1):437 – 440.

das Nações Unidas ONU, O. (2016). A agenda 2030 para o desenvolvimento sustentável.

de Oliveira, E. C. e Carraro, N. C. (2019). Análise do comportamento e participação do agronegócio na composição do produto interno bruto (PIB) brasileiro: Um estudo da série temporal de 1996 a 2017. *Brazilian Journal of Development*, 5(11):24042–24064.

Deponti, C. M., Eckert, C., e Azambuja, J. d. (2002). Estratégia para construção de indicadores para avaliação da sustentabilidade e monitoramento de sistemas. *Agroecologia e desenvolvimento rural sustentável*, 3(4):44–52.

Goumopoulos, C. e Mavrommati, I. (2020). A framework for pervasive computing applications based on smart objects and end user development. *Journal of Systems and Software*, 162:110496.

Imperatives, S. (1987). Report of the world commission on environment and development: Our common future. 10.

Lampert, V. d. N., da Silva, G., e Weiller, O. (2018). Nova planilha de sistematização da produção (Nova PSP): ferramenta de apoio no diagnóstico e intervenção em unidades de produção familiar com atividade leiteira. *Embrapa Pecuária Sul-Documentos (INFOTECA-E)*.

Lampert, V. d. N., do Amaral, E., Lopes, P. P., Fracari, J. R., Dantas, T. F., e Oliveira, M. (2015). Uma ferramenta para gestão de indicadores na produção de bovinos de corte: simplificando a organização de processos. In *Embrapa Pecuária Sul-Artigo em anais de congresso (ALICE)*. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 10., 2015, Ponta Grossa.

López-Ridaura, S., Masera, O., e Astier, M. (2002). Evaluating the sustainability of complex socio-environmental systems: the MESMIS framework. *Ecological Indicators*, 2(1):135 – 148. Hyatt S.I.

Negra, C., Remans, R., Attwood, S., Jones, S., Werneck, F., e Smith, A. (2020). Sustainable agri-food investments require multi-sector co-development of decision tools. *Ecological Indicators*, 110:105851.

Ness, B., Urbel-Piirsalu, E., Anderberg, S., e Olsson, L. (2007). Categorising tools for sustainability assessment. *Ecological Economics*, 60(3):498–508.

Rezaei, M. E., Barmaki, M., e Veisi, H. (2018). Sustainability assessment of potato fields using the DEXi decision support system in Hamadan Province, Iran. *Journal of Integrative Agriculture*, 17(11):2583 – 2595.

Rodrigues, G. S., Barros, I. d., e Martins, C. (2016). Análise de sustentabilidade em diferentes sistemas de produção de coco no norte e nordeste. In *Embrapa Tabuleiros Costeiros-Artigo em anais de congresso (ALICE)*. In: SEMINÁRIO DE INTENSIFICAÇÃO ECOLÓGICA DA FRUTICULTURA TROPICAL, 4.

Sitarz, D. (1993). Agenda 21: The earth summit strategy to save our planet.

Whitehead, J., MacLeod, C. J., e Campbell, H. (2020). Improving the adoption of agricultural sustainability tools: A comparative analysis. *Ecological Indicators*, 111:106034.

Zhen, L. e Routray, J. K. (2003). Operational indicators for measuring agricultural sustainability in developing countries. *Environmental Management*, 32(1):34–46.



Towards an IoT-Based Architecture for Monitoring and Automated Decision-Making in an Aviary Environment

Vinícius C. Lopes[1], **Roberto Felício de Oliveira**[2], **Valdemar Vicente Graciano Neto**[1]

GoIn-Sight: **G**oiás **In**formation **S**ystems and Software Engineering Research **T**eam
Instituto de Informática (INF) – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – CEP 74001-970 – Goiânia – GO – Brazil

[2]Universidade Estadual de Goiás (UEG) - UnU-Posse
CEP 73900-000 – Posse – GO – Brazil

vinilopes03@discente.ufg.br, roberto.oliveira@ueg.br, valdemarneto@ufg.br

Abstract. *Internet of Things (IoT) is a prominent technology in which everyday objects can be equipped with identifying, sensing, networking, and processing capabilities that allow them to communicate with each another and with other devices and services over the Internet to achieve some goal. In agribusiness, the use of technologies, such as IoT, is called precision livestock farming, which includes the use of different technologies in production and care of livestock animals. Brazil is one of the largest poultry producers in the world, being the first in exportation. Despite that, poultry production faces difficulties due to sensitivity that birds have to numerous environmental factors in aviaries, such as lightning, sounds, harmful gases, air humidity, food quality and clean water. When these variables are not well controlled, problems in meat quality and poultry production are likely to occur. As a result, farmers may face severe financial losses. Thus, providing a healthy environment is essential, and to achieve this, accurate monitoring and fast decision making are required in order to solve the problem as soon as possible. In this paper a detailed process of requirements elicitation and architectural design for IoT-Based aviary monitoring systems is proposed along with an informal literature review from the area. As preliminary results, we delivered a requirements document with functional and nonfunctional requirements closer to the real needs of farmers and an architectural proposal that can be used as a reference for further studies.*

1. Introduction

Brazil is one of the largest poultry producers in the world, being the first in exportation, according to the Brazilian Association of Animal Protein (ABPA) [Pessoa et al. 2013]. Nevertheless, domestic poultry production often faces difficulties due to the sensitivity that birds have to numerous factors of ambience in the aviary. The temperature is one of the factors that most impacts on losses of aviculture production [da Silva Oliveira et al. 2019], which can be due, for instance, to the imbalance of the birds' thermal comfort zone. The imbalance of the environment is not only caused by the temperature. There are other variables that influence the welfare of poultry, such as lightening, harmful gases, air humidity, food quality and clean water [Mumbelli et al. 2020].

Providing a healthy environment requires a monitoring and control practices, such as decision-making [Mumbelli et al. 2020]. A human must decide when and how to solve aviary's problems, such as the need to warm up or cool down the environment, also controlling the process of renewing aviary's air and the supply of water and feed, without having any waste [Choukidar and Dawande 2017]. However, a human by himself cannot execute these tasks with precision needed, since a person is subject to fatigue and physiological needs, such as sleeping, preventing 24/7 monitoring, which can expose the birds to temperature variations and other factors that can be fatal [Camargo et al. 2019]. Therefore, the need for a system that could automate and assist in this process is evident.

Developing a system that successfully addresses the major needs of the agribusiness is far from trivial. Despite huge difficulty, it is something that needs to be achieved, since the loss of production would have a huge impact on producers and in the poultry's market, mainly on the financial issue. For example, in Brazil, chicken consumption in 2020 was around 45 kilos of chicken meat/*per capita*, which is an increase of 5% compared to 2019 (42.84 kilos). For that reason, poultry production must keep up with the increase in consumption and failures must not happen during the monitoring activity, otherwise, it would negatively impact the country's economy. So, systems and technologies developed for this area must achieve a high level of quality and reliability, since it can be classified as critical systems. Thus, a well-structured and defined architecture is needed.

Thus, the main contribution of our study so far is providing preliminary results of an exploratory study on software architecture for monitoring an environment of an aviary and to support and automate the decision-making. Our **general goal** is to propose an architectural design that solve the problems faced in the daily basis of real farmers, specifically: (i) the need for a faster and more accurate decision-making in poultry management, (ii) the need to monitor the environment 24 hours a day, (iii) the stress level of the poultry inside the aviary. All of these problems can be related to critical systems due to potential losses, and those losses may strongly affect producers financially.

The remainder of this paper is structured as follows: Section 2 presents the background, Section 3 the methodology adopted by this study, Section 4 the results obtained so far with this study and Section 5 the final considerations of this paper.

2. Background

In agribusiness sector, which includes rural producers, farmers and ranchers, some terms are relevant, such as **Precision livestock farming**. This term refers to the application

of technology in the production and care of livestock animals. The aim of this technology is to manage animals by continuous real time health monitoring, welfare, production/reproduction, and environmental impact [Berckmans 2006]. The monitoring activity can be done by sensors, cameras connected to the internet and the aid software application it can assist the farmer's decision-making. Another synonymous term with precision livestock farming also used in this study is livestock 4.0. However, for conceiving any engineering system, requirements elicitation is an essential activity.

In software engineering, *software requirements* refer to the needs and constraints placed on a software product that contribute to the solution of some real-world problem [Bourque and Fairley 2014]. Requirements engineering involves elicitation, analysis, specification, and validation of software requirements as well as the management of requirements during the whole life cycle of the software product [Bourque and Fairley 2014]. The software requirements are commonly classified in two groups. First, **functional requirements** are related to the functions that should be performed by the system [Bourque and Fairley 2014]. Secondly, the **nonfunctional requirements** are related to constraints or quality requirements [Bourque and Fairley 2014].

From the set of requirements elicited, we can design the software architecture. **Software architecture** corresponds to the structure or structures of the system, involving software components, the externally visible properties of those components, and the relationships among them [Bass et al. 2003]. In the process of architectural design, there are some decisions over the solutions to be applied, such as the architectural styles and the views that will be described and represented by architecture. **Architectural Styles** refer to a specialization of element and relation types, together with a set of constraints on how they can be used [Horowitz et al. 2007]. In software architecture, a **view** is a representation of the system from the perspective of a particular viewpoint [ISO 2011].

Related Work. First, two search strings were defined seeking to reduce bias in this activity. The search started on May 24, 2021 with the strings application on Google Scholar digital base. As a result, five studies [Mumbelli et al. 2020], [Lashari et al. 2018], [Raj and Jayanthi 2018], [Debauche et al. 2020] and [Manshor et al. 2019] were retrieved. Another similar paper found was [Feijó et al. 2021], which describes a decision support system for agribusiness. This support takes place through the indication of technical documents about environmental legislation with more sustainable environmental practices. The results gave us a better idea about the existing gaps in the area. It was possible to notice that part of the studies so far developed are not concerned about software engineering solutions, and thus, we did not find specifications about requirements elicitation process and just a few reports about architectural solutions.

A recent systematic mapping showed that there are some studies on monitoring aviaries. The study [Mumbelli et al. 2020] reports the elaboration of a low-cost prototype of a Poultry House Monitoring System. [Raj and Jayanthi 2018] provides some interesting results regarding the monitoring activity of poultry diseases. The study [Debauche et al. 2020] proposes an IoT-Based aviary monitoring system; however, the focus is on an artificial intelligence algorithm to predict environment parameters, not monitoring and operationalizing actions on the environment, as we propose.

In face of all the studies presented, our study conducted differently from all

these related studies. One of our differentials is that our architectural design supports decision-making automation, which our requirements elicitation process showed to be a relevant function for farmers' monitoring work. That automation process seeks to reduce the need for a human being to control the temperature and ventilation of the aviary, in addition to speeding up the reaction time to correct an error in the environment, which can decrease the stress of the poultry and increase production. Another differential of our proposal are the adoption of a architectural design process [Hofmeister et al. 2007] and model [Kruchten 1995], which allowed us to design and define five different views of the system, a relevant document to critical systems since it must have a high level of quality to avoid failures and future losses.

Thus, we can also state that these studies are not concerned with elaborating a specific architecture for aviary monitoring that can serve as a basis for future works. The software solutions are not the main concern of them, as they are focused on how the system works as a hardware, or algorithms. Of course, some studies described some software architecture, such as [Raj and Jayanthi 2018] and [Lashari et al. 2018], but with few details, just describing system data flows. The contribution of our proposal is an architecture that can serve as a reference for a monitoring system with automated decision-making support, since the results obtained show an architecture that is concerned with the quality that this type of system must have. Besides that, this study presents requirements closer to the real needs of farmers than the previously mentioned studies, since this paper presents an interview process that supported the requirements elicitation with real farmers.

3. Methodology

In order to achieve the general goal of this study, research questions (RQ) were defined in Table 1.

Table 1. Research questions.

ID	Description
RQ01	What monitoring difficulties do poultry producers face?
RQ02	Is it possible to define an architecture to support the construction of monitoring systems that solve one or more typical problems of poultry production?
RQ03	How does the proposed architecture meet the needs of poultry producers?

A methodology was defined to guide this study, which made possible to answer the research questions previously defined. The methodology's phases for software engineering's studies proposed by Neto, Spínola and Travassos [Dias-Neto et al. 2010] was used as a guideline. It can be seen in Figure 1.

Literature review: An informal literature review was performed seeking to provide an overview of the context under study and investigate gaps in the area. Part of the results are shown in *Related work*. Even though the literature review was not systematic, the reading process was based on the studies [Lopes et al. 2020], [Batista et al. 2021] and [Santos et al. 2021]. The second step refers to **Requirements elicitation**. This process was performed aiming to understand the real needs for the project. First, we have defined a requirements elicitation process based on an interviews. Thus, the interviews were

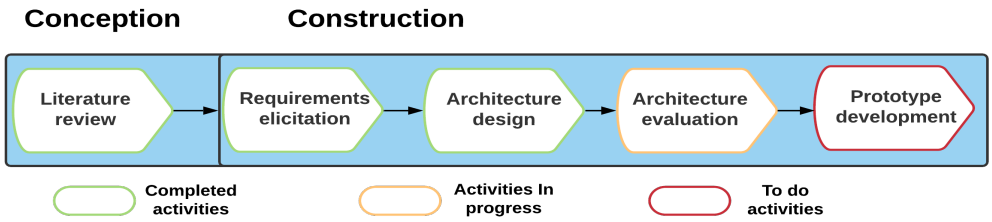


Figure 1. Scientific methodology proposed by Neto, Spínola and Travassos [Dias-Neto et al. 2010].

performed with real farmers, which enabled us to understand the rural worker needs and elaborate a requirements document. The requirements elicitation helped to answer **RQ01**.

Architecture design: An architecture was planned and specified. At the end of this activity an architectural document was elaborated. This document can be used to assist in the development and study of future work, since it is a new solution for the problems faced in this context. **Architecture evaluation:** An activity that is still in progress. The process consists of (i) a research in existing architectures in the context of critical systems, precision agriculture and precision livestock and (ii) the comparison of the architectures found with the one developed in this study. **Prototype development:** It refers to the instantiation of the conceived architecture in a proof-of-concept artifact. The software will be used for monitoring, gathering data and generate results on the benefits of IoT-Based applications in this area of study.

4. Preliminary Results

This section answers the research questions presented in Section 3. The **RQ01** is answered by the Requirements elicitation process that is presented in this section. The **RQ02** and **RQ03** are answered by the architectural design also presented in this section.

4.1. Requirements elicitation process

The process was applied in an interview model. Three rural workers were interviewed. The answers were recorded and documented. The requirements document provided by this process is composed by three tables. The functional requirements table contains twenty-two requirements, the nonfunctional requirements table contains six items and a third table that refers to the business rules of the project is composed of eleven items. We chose to present and exemplify two tables that better represent the requirements document. The Table 2 represents the functional requirements and Table 3 exemplify the nonfunctional requirements. These two requirements of each table were chosen to represent the automation of the system and also the prevention of failures and interferences, which is an important part of our solution.

4.2. Architectural design

Our second activity is the architectural design. This activity is necessary because the solutions found so far in the literature [Mumbelli et al. 2020], [Debauche et al. 2020],

Table 2. Table of functional requirements

Functional requirements	Description	Priority
RF01 - Automatically turn on the lamp	The system should turn on the lamp(s) automatically based on temperature parameters.	High
RF02 - Send failure notification	The system must send some notification or warning to system users if it fails to receive information from its sensors.	High

Table 3. Table of nonfunctional requirements

Nonfunctional requirements	Description	Priority
RNF01 - Availability	The system must monitor the aviary 24 hours a day.	High
RNF02 - Reliability	The system must not be interfered with by any other device, sensor or actuator that is not present in the system.	High

[Raj and Jayanthi 2018] do not support some essential features for this type of monitoring, such as the automation of decision making and failure prevention, which can avoid possible production losses, Figure 2 exemplifies these differences. Therefore, it is necessary to propose something new, thus, an architectural design process was conducted in conformance with the one prescribed by Hofmeister [Hofmeister et al. 2007], which defines activities to guide the architectural design process. The diagrams were produced in conformance to those defined in 4+1 view model proposed by Kruchten [Kruchten 1995]. The 4+1 view model [Kruchten 1995] defines five different views that were defined in our architecture document, the scenario view, the logical view, the developer view, the process view and the physical view.

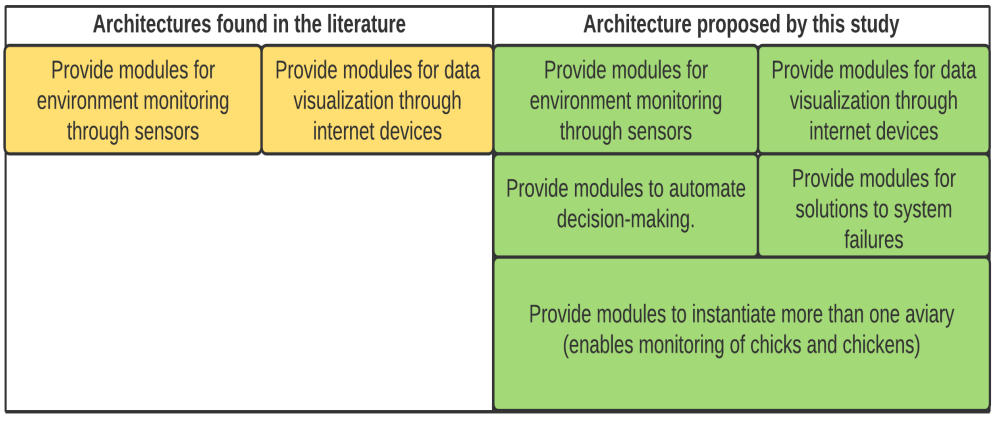


Figure 2. Deployment diagram

First, Figure 3 is a representation of the overall architecture view. The monitoring device is responsible for monitoring and gathering data of luminosity, temperature, air

humidity and device battery. This device is composed by a JARM ESP32 board, a SHT20 temperature and humidity sensor, a LoRa module, a SD card to store data from sensors, a DC plug that makes the device run directly on energy and a battery holder so that the device can still work in case of power failure. The gateway is also composed by a JARM ESP32, a LoRa module, a DC plug, a battery holder and a WIFI antenna to communicate with the ThingSpeak cloud and the platform through HTTP requests. There was a need for a LoRa/WiFi gateway to solve the problem of low WiFi connections in farms, thus enabling the sensor to communicate with the gateway at greater distances, where there is a better WiFi connection. This architectural design was made for each module to work individually. For example, the monitoring process is done only at the monitoring device part. All of the automation algorithms are located in a firmware on the ESP32 board, and all date gathered are stored in a SD card in the same device. Thus, if any failures happen on the gateway or on the platform, it would not affect the monitoring process.

Second, Figure 4 is a component diagram. The diagram describes the software components, their interfaces and their dependencies. In the diagram, there are thirteen components and three subsystems. The component called **SensorNode**, which is composed of sensors and actuators, communicates with the subsystem called **DataControl**. The **DataControl** subsystem is responsible for maintaining and processing data coming from the sensors, in addition to allowing the export of this type of data. **DataControl** communicates with the **Accounting** subsystem, which is responsible for maintain data regarding system user's accounts, in which each aviary is managed by an account. **Accounting** communicates with the **WebDataView** subsystem, which performing actions in the system, such as viewing sensor data and performing the user authentication process provided by the **DataView** and **Authentication** components respectively. Finally, the **ClientBrowser** component represents the user's current application.

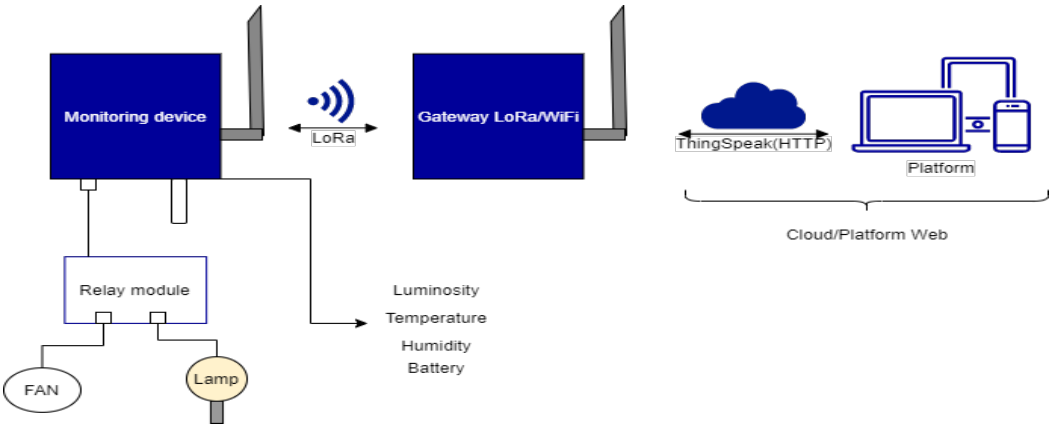


Figure 3. Overall Architecture

5. Final Remarks

This paper presented preliminary results of a study conducted on precision livestock farming envisioning the use of IoT-Based for automated decision in aviary monitoring systems. The main contributions so far are (i) results of an informal literature review, (ii) a proposal of a requirements elicitation process that can be successfully used with farmers and the

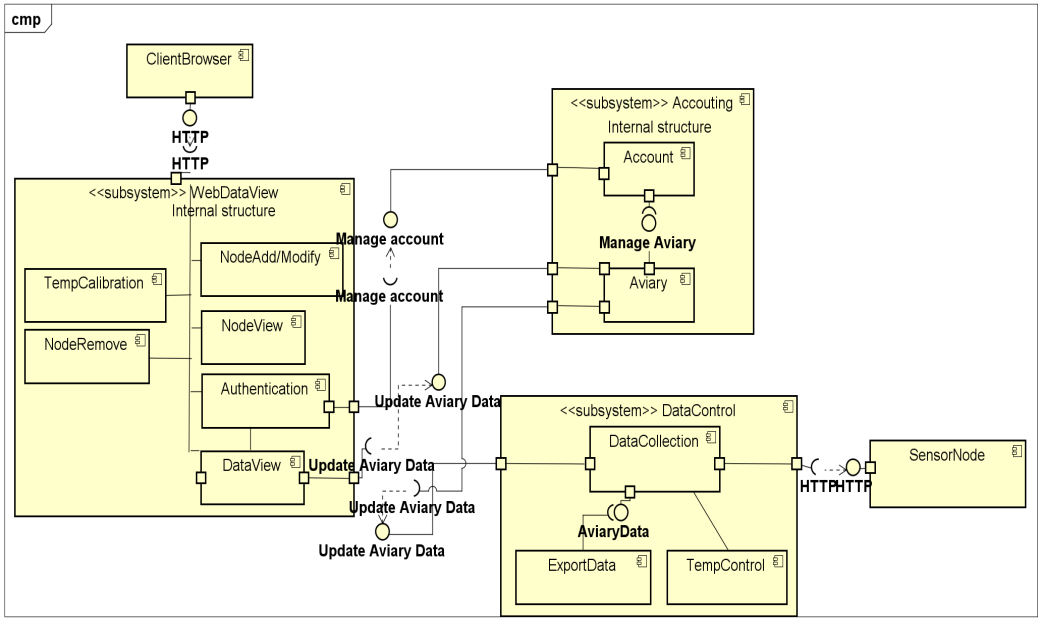


Figure 4. Component diagram

requirements themselves and (iii) an architectural design adapted to this context to satisfy the requirements elicited and to meet the automate and support decision-making main requirement. Based on our findings, we externalized them into a requirement document, which is composed of functional and non-functional requirements and an architecture document, composed by five diagrams. We are now finalizing the process of refining our architectural design. At the end, a prototype will be developed, deployed and evaluated.

References

(2011). ISO/IEC/IEEE Systems and software engineering. *ISO/IEC/IEEE 42010:2011(E) (Revision of ISO/IEC 42010:2007 and IEEE Std 1471-2000)*, pages 1–46.

Bass, L., Clements, P., and Kazman, R. (2003). *Software Architecture In Practice*.

Batista, P. E. P., Bulcão Neto, R. d. F., Paes, C. E. d. B., Lima, E. A. d., Rohling, A. J., and Graciano Neto, V. V. (2021). Reference architecture for systems of systems: a systematic mapping. *iSys - Brazilian Journal of Information Systems*, 14(1):119–153.

Berckmans, D. (2006). Automatic on-line monitoring of animals by precision livestock farming. *Livestock production and society*, 287:27–30.

Bourque, P. and Fairley, R. E., editors (2014). *SWEBOK: Guide to the Software Engineering Body of Knowledge*. IEEE Computer Society, Los Alamitos, CA, version 3.0 edition.

Camargo, T. F., Silva, R. L., Higa, M., Coutinho, M. R., de Oliveira, J. C., and Conceição, W. A. d. S. (2019). Monitoramento do conforto térmico em aviários mediante sistemas de aquisição de dados em tempo real. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 23(9):694–701.

Choukidar, G. A. and Dawande, N. (2017). Smart poultry farm automation and monitoring system. In *2017 ICCUBEA*, pages 1–5. IEEE.

da Silva Oliveira, G., dos Santos, V. M., Rodrigues, J. C., and Nascimento, S. T. (2019). Protótipo para o estudo do comportamento e da zona de conforto térmico de frangos de corte. *REVISTA EIXO*, 8(1).

Debauche, O., Mahmoudi, S., Mahmoudi, S. A., Manneback, P., Bindelle, J., and Lebeau, F. (2020). Edge computing and artificial intelligence for real-time poultry monitoring. *Procedia Computer Science*, 175:534–541.

Dias-Neto, A. C., Spinola, R., and Travassos, G. H. (2010). Developing software technologies through experimentation: experiences from the battlefield. In *XIII CIbSE*.

Feijó, T., David, J. M., Braga, R., Otenio, M. H., Paula, V. R., Santos, G. M., Campos, F., and Stroele, V. (2021). @grogestambiental: A web-based decision support system for agribusiness. WebMedia '21, page 1–8, New York, NY, USA. Association for Computing Machinery.

Hofmeister, C., Kruchten, P., Nord, R. L., Obbink, H., Ran, A., and America, P. (2007). A general model of software architecture design derived from five industrial approaches. *Journal of Systems and Software*, 80(1):106–126.

Horowitz, E., Sahni, S., and Rajasekaran, S. (2007). *Computer Algorithms*. Silicon Press, USA, 2nd edition.

Kruchten, P. B. (1995). The 4+ 1 view model of architecture. *IEEE software*, 12(6):42–50.

Lashari, M. H., Memon, A. A., Shah, S. A. A., Nenwani, K., and Shafqat, F. (2018). Iot based poultry environment monitoring system. In *2018 IEEE IOTAIS*, pages 1–5.

Lopes, V. C., Norberto, M., R. S., D. W., Kassab, M., da Silva Soares, A., Oliveira, R., and Neto, V. V. G. (2020). A systematic mapping study on software testing for systems-of-systems. SAST 20, page 88–97, New York, NY, USA. Association for Computing Machinery.

Manshor, N., Rahiman, A. R. A., and Yazed, M. K. (2019). Iot based poultry house monitoring. In *2019 2nd ICCET*, pages 72–75. IEEE.

Mumbelli, A., Brito, R. C., Pegorini, V., and Priester, L. F. (2020). Low cost iot-based system for monitoring and remote controlling aviaries. In *2020 3rd ICICT*, pages 531–535. IEEE.

Pessoa, G. T., de Sousa, G. V., Ferraz, M. S., Feitosa, M. L. T., and de Miranda Sampaio, A. (2013). Estratégias inovadoras no manejo de frangos de corte em avicultura industrial: fases pré-inicial, inicial, engorda e final. *Pubvet*, 7:1002–1136.

Raj, A. A. G. and Jayanthi, J. G. (2018). Iot-based real-time poultry monitoring and health status identification. In *2018 11th ISMA*, pages 1–7.

Santos, D., Basso, F., Luizelli, M., and Cabrera, S. (2021). Emprego de simulações computacionais em problemas envolvendo agricultura: Um estudo de mapeamento sistemático. In *Anais do III Workshop em Modelagem e Simulação de Sistemas Intensivos em Software*, pages 20–29, Porto Alegre, RS, Brasil. SBC.



Um estudo comparativo entre as arquiteturas de redes neurais profundas *AlexNet* e YOLO aplicadas ao problema de visão computacional em Ambientes Agrícolas

Danielly O. A. de Oliveira[1], **Josué Pereira de Castro**[1]

[1]Ciência da Computação - Universidade Estadual do Oeste do Paraná (UNIOESTE)
Caixa Postal 701 – 85.819-110 – Cascavel – PR – Brazil

danielly_omori@hotmail.com, josue.castro@unioeste.br

Abstract. *The verification and control of the uniformity of irrigation water distribution is one of the most important tasks to ensure the best use of water resources. This type of activity requires the robot to be able to move in rural environments, recognizing obstacles in its path. In this article, a comparative study was developed between two convolutional neural network architectures, AlexNet and YOLO, with the objective of being embedded in a Raspberry PI processor mounted on a small robot to be used to assist in the automated adjustment of adjustable emitters for irrigation located. The networks were trained using databases containing images related to the agricultural environment and then, data regarding the performance of both architectures were collected and a comparison was made between them, using several criteria, as the execution time, the consumption of main and secondary memory, and classification accuracy. The architecture that showed the best performance was YOLO, which was embedded in the robot, undergoing the necessary adaptations so that it was able to use the images transmitted in real time by the built-in camera existing in the robot.*

Resumo. *A verificação e o controle da uniformidade de distribuição de água para irrigação é uma das tarefas de maior importância para garantir o melhor uso dos recursos hídricos. Este tipo de atividade exige que o robô seja capaz de deslocar-se em ambientes rurais, reconhecendo obstáculos em seu caminho. Neste artigo foi desenvolvido um estudo comparativo entre duas arquiteturas de redes neurais convolucionais, AlexNet e YOLO, com o objetivo de serem embarcadas em um processador Raspberry PI montado em um robô de pequeno porte a ser usado para auxiliar no ajuste automatizado de emissores ajustáveis para irrigação localizada. As redes foram treinadas, utilizando bases de dados contendo imagens relacionadas ao meio agrícola e a seguir, foram coletados dados*

referentes ao desempenho de ambas arquiteturas e realizada uma comparação entre elas, usando vários critérios, sendo os principais o tempo de execução, o consumo de memória principal e secundária, e a precisão da classificação. A arquitetura que apresentou melhor desempenho foi a YOLO, a qual foi embarcada no robô, passando pelas devidas adaptações para que a mesma fosse capaz de utilizar as imagens transmitidas em tempo real pela câmera embutida existente no robô.

1. Introdução

Desde o início da década de 1990 é possível observar um grande desenvolvimento na área de Tecnologia da Informação (TI) e como sua utilização em processos de produção poderiam aumentar a produtividade industrial [Santos et al. 2018]. Uma das áreas que mais tem recebido incentivo para entrar no mundo tecnológico é a área agrícola, que por ser uma das bases da economia brasileira, torna essencial que sejam utilizadas tecnologias para aumentar o rendimento das lavouras [Massruhá e Leite 2017, Jardim 2019].

Alguns projetos já vem sendo desenvolvidos com o intuito de estimular este mercado a investir no desenvolvimento de soluções voltadas para o agronegócio. O programa Agro 4.0 [ABDI 2019], criado pelos ministérios da Agricultura, Economia e da Ciência, Tecnologia e Inovações, juntamente com a Agência Brasileira de Desenvolvimento Industrial (ABDI) tem o objetivo de promover o aumento de eficiência e produtividade agrícolas, empregando métodos computacionais para processar grandes volumes de dados e construir sistemas de tomada de decisão.

A partir do ano 2000 tem sido dada muita importância ao uso consciente de recursos hídricos [Paz et al. 2000], com destaque para os crescentes períodos de seca, que afetam a produção agrícola. Esta situação torna imprescindível otimizar o uso de recursos hídricos, de forma a reduzir o consumo de água na produção agrícola. Segundo a UNESCO [UNESCO 2020], a agricultura é responsável por 70% da exploração global de água doce, com possibilidade de chegar a 90% em países em rápido crescimento, sendo a irrigação é responsável por 40% da produção mundial de alimentos. Disto conclui-se que há necessidade premente de desenvolver tecnologias que otimizem e reduzam o consumo de água para irrigação.

Visão computacional e robótica podem ser ferramentas úteis para diversas finalidades dentro da área de irrigação. Há vários trabalhos que utilizam visão computacional em ambiente agrícolas, por exemplo: ASTRAND [Åstrand e Baerveldt 2002] descreve um robô móvel agrícola capaz de realizar o controle de ervas daninhas em uma plantação de beterrabas; JODAS [Jodas et al. 2013]. apresenta um sistema capaz de fazer um robô móvel deslocar-se por plantações de forma autônoma, com o foco na realização de tarefas como detecção de pragas e controle da aplicação de agrotóxicos. Contudo, aplicações deste tipo voltadas para irrigação localizada ainda são poucas, por exemplo: [Lam et al. 2006] apresenta um sistema de controle de feedback de sensoriamento remoto baseado no solo para monitorar o avanço da água em sulcos e controlar automaticamente a descarga de água na entrada do sulco. [Berenstein et al. 2018] apresenta uma nova abordagem para ajustar emissores usando um manipulador robótico móvel autônomo.

Esse contexto, considerando ainda o fato da região oeste do Paraná ter vocação agrícola, motivaram o desenvolvimento deste trabalho. O principal objetivo é desenvol-

ver um sistema de visão em um robô embarcado de pequeno porte, com uma câmera embarcada, que seja capaz de detectar objetos em ambientes agrícola, localizando-os e identificando-os corretamente, para auxiliar na tarefa de ajuste automático de gotejadores. Para determinar quais seriam estes objetos, foi feita uma análise visual de uma área rural, identificando quais objetos encontravam-se dispostos neste local. Procurou-se fazer uma classificação genérica dos objetos, evitando-se detalhamentos desnecessários que não contribuíssem para a aplicação desejada, de modo que optou-se por uma classe genérica "planta"do que classes específicas como "soja"ou "feijão", por exemplo, já que este tipo de classificação não influencia na tarefa de ajuste dos gotejadores.

2. Materiais e Métodos

Esta seção tem como objetivo apresentar brevemente os recursos e a metodologia utilizados no desenvolvimento deste trabalho.

2.1. Raspberry PI

Este projeto utiliza uma placa Raspberry Pi modelo 3B integrada a um robô Wi-Fi com chassi TH *Robot Tank* produzido por *Xiao R Geek Techonology* e possui uma câmera USB (*Universal Serial Bus* - Porta Serial Universal) em um suporte com dois graus de liberdade, e um braço manipulador com 4 graus de liberdade[1]. O modelo do robô pode ser visto na Figura 1.

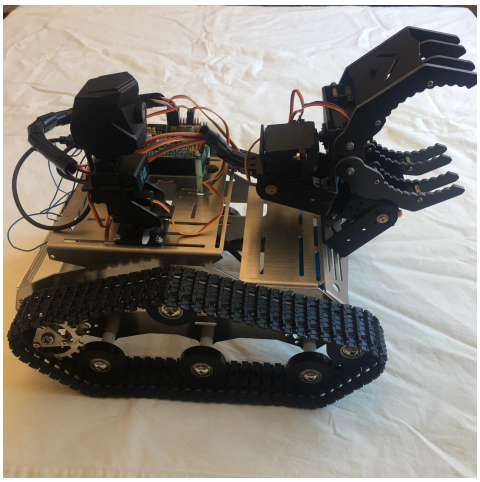


Figura 1. Modelo de robô utilizado

2.2. Linguagem Python

Python é uma linguagem de programação de alto nível, interpretada, orientada a objetos, de tipagem dinâmica e forte, de código livre, publicada no ano de 1991, por Guido van Rossum [itMídia 2019]. Neste trabalho foi utiliza esta linguagem em com suas bibliotecas OpenCV[2] e *Tensor Flow*[3].

[1]quantidade de juntas do manipulador que restringem a sua liberdade de movimento
[2]https://opencv.org
[3]https://www.tensorflow.org/learn?hl=pt-br

2.3. Estruturas de Redes Neurais Convolucionais Utilizadas

2.3.1. *AlexNet*

O trabalho desenvolvido por Krizhevsky [Krizhevsky et al. 2017] tinha como objetivo classificar a base de dados ImageNet [Yang et al. 2020]. A rede desenvolvida possuía oito camadas, com cinco delas sendo camadas convolucionais e três camadas totalmente conectadas. A saída da última camada totalmente conectada era utilizada como entrada para uma camada de ativação, que fazia uso da função *softmax* para realizar a classificação entre as 1000 classes existentes. Todo esse processo foi realizado utilizando duas GPUs (*Graphics Processing Unit* - Unidade de Processamento Gráfico), o que facilitou o processamento da rede.

2.3.2. YOLO (*You Only Look Once*)

YOLO (*You Only Look Once*) é um sistema de reconhecimento em tempo real, desenvolvido com o objetivo de detectar vários objetos em um único frame, além de gerar caixas delimitadoras (*Bounding Box Predictions*) que indicam em que região da imagem os objetos estão localizados e a qual classe eles pertencem [Orac 2020].

Seu funcionamento se difere das de diversas arquiteturas de redes neurais convolucionais pois, ele não somente realiza o treinamento da rede, como também a divisão da imagem em áreas de interesse, permitindo que o algoritmo também seja capaz de delimitar a área em que os objetos serão encontrados [Techzizou 2020].

2.4. Ambientes Utilizados

O treinamento dos protótipos criados para esse projeto foram executados no ambiente Google Colaboratory[4]. As configurações utilizadas no ambiente para a execução dos modelos foram as padrões gratuitas do Colab, consistindo em:

- Processador: Intel(R) Xeon(R), 2.30GHz, ;
- RAM: 12GB;
- Espaço Disponíve: 25GB;
- GPU: Nvidia K80, 12GB.

A máquina utilizada para a realização dos testes foi a placa Raspberry Pi, modelo 3B, embutida no robô. Suas configurações consistiam em:

- Sistema Operacional: Raspberry PI OS 32-bits. *Released*: 2021-03-04;
- Processador: Quad Core 1.2GHz Broadcom BCM2837 64bit CPU;
- RAM: 1GB.

2.5. Bases de Dados

Com objetivo de abranger objetos existentes no ambiente rural, criou-se uma base de dados contendo seis classes: animais, flores, gotejador, veículos, pessoas e plantas. Elas foram escolhidas após um estudo sobre quais objetos eram mais encontrados em ambiente agrícola local. Para construí-la, foram reunidas imagens de bases de dados disponíveis no

[4]https://colab.research.google.com

site Kaggle[5] e em alguns casos buscou-se imagens em outras fontes. Maiores informações referentes à composição das bases podem ser vistas à seguir:

- **Animais**: Foram utilizadas duas bases de imagens para o desenvolvimento dessa classe. A base *256-Objects*, utilizando as classes *dog*, *horse* e *duck* e a base *Animal-10*, utilizando as classes *cat*, *chiken*, *cow*, *dog*, *horse* e *sheep*;
- **Flores**: Para a classe flor, foi utilizada a Base *Hackathon Blossom (Flower Classification)*;
- **Gotejador**: Para a construção dessa classe, foram utilizadas imagens disponíveis no Google, utilizando a palavra chave "gotejador".
- **Veículos**: Foi utilizada a base *Vehicle Data Set*, selecionando apenas as classes *Car*, *Truck*, *Van*, *Taxi*, *Ambulance* e *Bus*. Além disso, se utilizou a classe *bulldozer* da base *256-Objects* para que fosse possível abranger veículos de maior porte;
- **Pessoas**: A base *1 Million Fake Faces* contendo imagens de rostos de pessoas falsas. Também foram utilizadas imagens disponíveis no Google (utilizando as palavras chaves "fotos paparazzi famosos") para poder completar a base de dados, pois não foi possível encontrar base de dados que representassem o corpo inteiro de pessoas;
- **Plantas**: Inicialmente utilizou-se as bases *V2 Plant Seedlings Dataset* e *Agriculture crop images* para fazerem parte da classe planta. Notou-se, porém, que utilizar essas bases apenas não seriam suficientes, pois elas não eram capazes de representar tão bem a vegetação brasileira. Por este motivo, foi realizada uma visita ao Núcleo Experimental de Engenharia Agrícola (NEEA) da UNIOESTE, para realizar a coleta de imagens de outros tipos de plantações como soja e milho.

Foram utilizadas 170 instâncias para as classes planta, veículo, animal, flor e pessoa e 150 para a classe gotejador, que por ser mais característica, não necessitava de tantas imagens. A base final, composta por um total de mil imagens, foi dividida em 70% para treinamento, 15% para validação e 15% para testes.

2.6. Experimentos Realizados

Apresentaremos aqui os protótipos desenvolvidos e os comparativos realizados com as redes escolhidas. Para cada uma das redes foi implementado um protótipo em linguagem Python e realizado o treinamento da rede neural, que ao final de cada iteração, passava por uma etapa de validação para verificar se os valores sendo obtidos durante a etapa anterior estavam se aproximando do valor esperado ou não. Ao fim do treinamento, foram colhidos os resultados referentes ao desempenho da rede e a partir de uma análise desses valores, foi escolhido o protótipo mais adequado a ser embarcado no robô.

2.6.1. Protótipo I

O primeiro protótipo criado utilizou a arquitetura *AlexNet*, utilizando o modelo desenvolvido por ALAKE [Alake 2020], redimensionando as imagens para o tamanho igual à 200×200 pixels. Para o treinamento, foram utilizadas 1.000 épocas, com 7 iterações cada, totalizando 7.000 iterações. Esse número de épocas foi escolhido após testes preliminares que verificaram que caso fossem utilizados valores maiores, a rede entraria em processo

[5]https://www.kaggle.com

de *overfitting*. Além disso, optou-se por usar uma etapa de validação após cada iteração. Os parâmetros utilizados foram taxa de aprendizado de $0,00261$[6] e *Batch Size* [7] igual à 100. Além disso, também utilizou-se o parâmetro *Sparse Categorical Crossentropy* para realizar o cálculo de perda do modelo e otimizador Adam. O código completo pode ser encontrado no Google Colaboratory[8].

2.6.2. Protótipo II

Ao analisar o primeiro modelo gerado, notou-se que ele não seria adequado para ser embarcado no robô, principamente por demandar muita memória. Portanto, para o segundo protótipo, buscou-se uma alternativa voltada para a utilização em ambiente embarcado e *mobile*, como é o caso do robô. A solução encontrada para contornar esse problema a foi o YOLO-V4 *Tiny*, versão reduzida do sistema YOLO, capaz de criar um modelo de rede neural mais simples, com número de parâmetros reduzidos, visando utilização de redes neurais em ambientes que não possuem tanto poder computacional disponível.

Conforme solicitado nas documentações do YOLO-V4 *Tiny*, foram utilizadas configurações específicas para a realização do treinamento, tais como imagens redimensionadas para o tamanho 416×416, taxa de aprendizado igual à $0,00261$, e número máximo de *batches* igual ao número de classes multiplicados por 2.000, o que totalizou em 12.000 lotes. A organização das camadas seguiu o modelo desenvolvido por [Jacob e Samrat 2020] e está disponível no Google Colaboratory[9].

3. Resultados e Discussões

3.1. Análise de Desempenho

Durante a realização dos experimentos, foram analisados os seguintes para verificação do desempenho da rede:

- Consumo máximo de memória que o treinamento da rede exigiu;
- Tempo gasto para a realização do treinamento;
- Precisão obtida durante a etapa de testes.

Além disso, ao fim do treinamento, foram exportados os pesos gerados pela rede para um arquivo que pudesse ser utilizado em um ambiente externo ao que foi utilizado nesta etapa. Os resultados podem ser vistos na Tabela 1.

3.2. Protótipo I

O protótipo I obteve uma precisão média de 72% durante a etapa de testes, atingindo um uso máximo de memória RAM de 2,56 GB e levando um total de 36 minutos e 22 segundos para finalizar seu treinamento.

[6]Parâmetro que controla o tamanho do salto na descida do gradiente e controla a velocidade e precisão do treinamento [Academy 2019]

[7]Refere-se ao número de exemplos de treinamento utilizados em cada iteração [Academy 2019]

[8]https://colab.research.google.com/drive/1vqYAJEXrIi3dtMOgSA4mqeCQwO6lpyBU?usp=sharing

[9]https://colab.research.google.com/drive/1JIHClGomsXoLw86aWzK5T05u6uoh0cxJ?usp=sharing

Apesar de ter bons resultados referentes ao desempenho, a *AlexNet* se mostrou uma rede "pesada", gerando um total de 171.557.638 parâmetros quando compilada utilizando o *TensorFlow*. Isto influenciou diretamente no tamanho do arquivo contendo os pesos finais da rede, que ao ser exportado, totalizou 670 MB.

3.3. Protótipo II

O protótipo II apresentou uma precisão média de 77,40% durante a etapa de testes, utilizando no máximo 1,28 GB de memória RAM do ambiente de treinamento, porém, levando um total de 3 horas, 56 minutos e 41 segundos para a realização dessa etapa.

Apesar da demora na etapa de treinamento, o modelo YOLO apresentou uma rede menor que a do protótipo anterior, gerando um total de 5.891.874 parâmetros em seu modelo quando convertido para o formato *Tensorflow*. O arquivo contendo os pesos finais da rede totalizou o tamanho de 23 MB.

3.4. Comparações

Como pode ser observado na Tabela 1, o protótipo I apresentou uma única vantagem em relação ao concorrente, referente ao tempo de treinamento.

Tabela 1. Comparação entre os protótipos

COMPARAÇÃO PROTÓTIPOS		
Parâmetros de Análise	**Protótipo I**	**Protótipo II**
Número de Iterações	7.000	12.000
Tempo de Execução	36m 22s	3h 56m 41s
Gasto Máximo de Memória	2,6 GB	1,8 GB
Precisão	72,00%	77,40%
Número de Parâmetros Gerados	171.557.638	5.891.874
Tamanho do Arquivo de Pesos Gerados	670MB	23MB
Possui algum marcador para a área que os objetos estão?	Não	Sim

Já o protótipo II, além de apresentar melhor precisão, obteve destaque em dois pontos. O primeiro no que diz respeito ao tamanho da rede gerada e o segundo referente ao gasto de memória RAM máximo durante o treinamento. Como visto anteriormente, a placa Raspberry Pi possui limitações referente ao poder de processamento, tendo disponível apenas 1 GB de memória RAM. Por conta disso, o modelo a ser embarcado nele deveria ser algo mais simples, que não demandasse tanto poder computacional.

Outro ponto positivo do protótipo II em relação ao seu componente é que o sistema YOLO já possui integrado em seu algoritmo de treinamento o processo de criação de caixas delimitadoras, que indicam a área em que o objeto classificado se encontra na imagem de entrada. Para que a *AlexNet* seja capaz de realizar essa tarefa, seria necessário utilizar algoritmos próprios para o tratamento e segmentação de imagem e para o desenho dessas caixas delimitadoras. Esse foi o caso do trabalho realizado por [Wang et al. 2019], que utilizou a *AlexNet* para realizar o treinamento da rede, mas também apresentou em seu projeto algoritmos de tratamento de imagens para a realização da segmentação da imagem em objeto de interesse e fundo. Técnicas como Curva ROC (Característica de

Operação do Receptor - *Receiver Operating Characteristics*), *F-Measure*, Coeficiente *Dice* [Prabha e Kumar 2016] poderiam ser utilizadas nesta etapa, porém ocasionariam em mais gasto da memória do robô.

Por esses motivos escolheu-se o protótipo II para ser utilizado no processo de desenvolvimento da visão computacional do robô. Além de apresentar um melhor uso da memória RAM que o outro protótipo e maior acurácia, o protótipo também possui integrado ao seu processo de treinamento toda a parte de segmentação da imagem em área de interesse.

As Figuras 2 e 3 apresentam alguns resultados obtidos após rodar o sistema YOLO em imagens de diferentes classes.

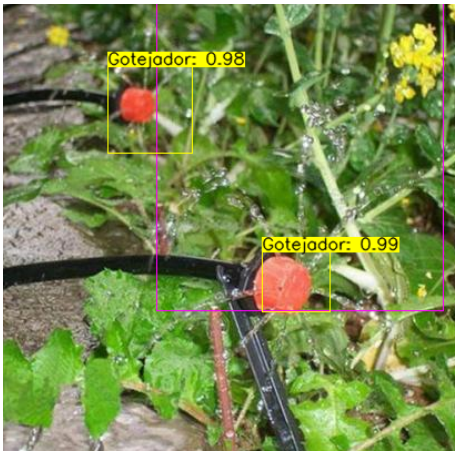


Figura 2. Exemplo Gotejador



Figura 3. Exemplo Planta

4. Considerações Finais

A arquitetura YOLO mostrou melhores resultados quando comparada à arquitetura AlexNet em relação a desempenho, número de parâmetros e arquivos gerados, gasto máximo de memória, entre outros. Além disso, ele já possui um sistema integrado que permite marcar a área que os objetos identificados estão. Por este motivo, foi escolhido o protótipo II para ser embarcado no robô. O sistema de visão desenvolvido foi então embarcado

e se mostrou capaz de detectar e identificar objetos presentes em ambiente agrícola, tais como plantas, veículos, pessoas, com uma taxa de acerto de 77.40%.

Referências

ABDI (2019). Agro 4.0.

Academy, D. S. (2019). Deep learning book.

Alake, R. (2020). Implementing alexnet cnn architecture using tensorflow 2.0+ and keras.

Åstrand, B. e Baerveldt, A.-J. (2002). An agricultural mobile robot with vision-based perception for mechanical weed control. *Autonomous robots*, 13(1):21–35.

Berenstein, R., Fox, R., McKinley, S., Carpin, S., e Goldberg, K. (2018). Robustly adjusting indoor drip irrigation emitters with the toyota hsr robot. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia. IEEE, IEEE.

itMídia (2019). Python: 10 motivos para aprender a linguagem em 2019.

Jacob, S. e Samrat, S. (2020). Train yolov4-tiny on custom data - lightning fast object detection.

Jardim, A. (2019). Inovação para o agro 4.0. *AgroANALYSIS*, 38(4):48.

Jodas, D. S., Marranghello, N., Pereira, A. S., e Guido, R. C. (2013). Desenvolvimento de um sistema para navegaçao de robôs móveis por caminhos em plantaçoes. *Interciência & Sociedade*, 2(1).

Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.

Lam, Y., Slaughter, D., Wallender, W., e Upadhyaya, S. (2006). Computer vision system for automatic control of precision furrow irrigation system. In *2006 ASAE Annual Meeting*, page 1. American Society of Agricultural and Biological Engineers.

Massruhá, S. M. F. S. e Leite, M. d. A. (2017). Agro 4.0-rumo à agricultura digital. In *Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE)*. In: MAGNONI JÚNIOR, L.; STEVENS, D.; SILVA, WTL da; VALE, JMF do; PURINI, SR ....

Orac, R. (2020). What's new in yolov4?

Paz, V. P. d. S., Teodoro, R. E. F., e Mendonça, F. C. (2000). Recursos hídricos, agricultura irrigada e meio ambiente. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 4:465–473.

Prabha, D. S. e Kumar, J. S. (2016). Performance evaluation of image segmentation using objective methods. *Indian J. Sci. Technol*, 9(8):1–8.

Santos, B. P., Alberto, A., Lima, T. D. F. M., e Charrua-Santos, F. M. B. (2018). Industry 4.0: challenges and opportunities. *Revista Produção E Desenvolvimento*.

Techzizou (2020). Yolov4 vs yolov4-tiny.

UNESCO (2020). Relatório mundial das nações unidas sobre desenvolvimento dos recursos hídricos 2020: O manejo dos recursos hídricos em condições de incerteza e risco.

Wang, R., Xu, J., e Han, T. X. (2019). Object instance detection with pruned alexnet and extended training data. *Signal Processing: Image Communication*, 70:145–156.

Yang, K., Qinami, K., Fei-Fei, L., Deng, J., e Russakovsky, O. (2020). Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In *Conference on Fairness, Accountability, and Transparency*.



Uma plataforma de dados geoespaciais para caracterização e combate à desertificação da Caatinga

Jaudete Daltio, André Rodrigo Farias,
Lucíola Alves Magalhães, Daniela Maciel Pinto

[1]Empresa Brasileira de Pesquisa Agropecuária (Embrapa)
Soldado Passarinho 303, Fazenda Chapadão – Campinas – SP – Brasil

{jaudete.daltio, andre.farias,
luciola.magalhaes, daniela.maciel}@embrapa.br

Abstract. *Desertification is the land-degradation process that occurs in arid, semi-arid, and dry sub-humid zones as a result of the joint action of natural source and anthropic factors. In Brazil, most of the susceptible areas are located in the Caatinga biome. This paper presents a data platform for characterizing and combating desertification that combines two main aspects: a collection of technologies to prevent, stem, or reverse desertification and a set of features that provides an overview of the regional agriculture trends. We expect to support technological disclosure strategies in suitable contexts and territories, and thus contribute to promoting sustainable agricultural development.*

Resumo. *Desertificação é o processo de degradação das terras que ocorre nas zonas áridas, semiáridas e sub-úmidas secas, resultado da ação conjunta de fatores de origem natural e antrópica. No Brasil, grande parte das áreas suscetíveis localiza-se no bioma Caatinga. Este artigo apresenta uma plataforma de dados para caracterização e combate à desertificação que integra dois aspectos principais: um acervo de tecnologias para prevenir e mitigar a desertificação e um conjunto de variáveis que provê uma visão geral das tendências da agropecuária regional. Espera-se subsidiar estratégias de disseminação tecnológica em contextos e territoriais adequados, contribuindo com a promoção do desenvolvimento agropecuário sustentável.*

1. Introdução e Motivação

De acordo com a Convenção das Nações Unidas de Combate à Desertificação, desertificação é o processo de degradação das terras que ocorre nas zonas áridas, semiáridas e sub-úmidas secas, resultado da ação conjunta de múltiplos fatores que vão

desde características naturais (principalmente solo, clima e relevo) a intervenções humanas no ambiente [UNCCD 1994]. Dada sua relevância, o combate à desertificação é um dos temas no Objetivo de Desenvolvimento Sustentável (ODS 15) [1]. O item 15.3 estabelece a meta de, até 2030, combater a desertificação, restaurar a terra e o solo degradado e lutar para alcançar um mundo neutro em termos de degradação do solo.

O conceito de Neutralidade da Degradação da Terra (*Land Degradation Neutrality* - LDN), destacado na definição da meta 15.3, refere-se a um estágio em que os recursos da terra encontram-se em um nível suficiente para a manutenção de todas as funções ecossistêmicas e da segurança alimentar das populações, em diferentes escalas temporais e espaciais [UNCCD 2019]. A neutralidade, segundo essa abordagem, se manifesta pela ausência de perda líquida no balanço de terras em processo de degradação e não degradadas. Assim, qualquer aumento de degradação deve ser necessariamente acompanhado por um crescimento proporcional de redução ou mitigação nessa ou em outras áreas, de forma que as perdas sejam compensadas pelos ganhos.

Associada a essa definição, desenvolveu-se uma estrutura científica para planejamento, implementação e monitoramento da LDN em diferentes contextos e sua instituição como uma meta global, na qual os países foram convidados a se comprometer voluntariamente com a LDN em nível nacional [Cowie et al. 2018]. Esse compromisso foi firmado, de forma institucional, a partir da adesão à Estratégia de Neutralidade da Degradação da Terra, mecanismo global de avaliação e implementação da LDN. Atualmente, esse instrumento reúne 128 países, conjunto do qual o Brasil faz parte.

Sob esta perspectiva, diversas iniciativas internacionais têm sido realizadas para atender aos objetivos estabelecidos pela LDN, principalmente no fomento a criação e difusão de tecnologias e práticas associadas à Gestão Sustentável da Terra (*Sustainable Land Management* - SLM). A SLM, por sua vez, define-se como o uso dos recursos da terra, incluindo solos, água, animais e plantas, visando a produção de bens para atender às necessidades humanas em constante mudança, onde há a garantia de manutenção simultânea de sua produtividade de longo prazo, a preservação de seus recursos e a manutenção de suas funções ambientais [Critchley et al. 2021].

Em âmbito nacional, iniciativas relacionadas ao tema são coordenadas pelo Ministério do Meio Ambiente (MMA), por meio do Programa de Ação Nacional de Combate à Desertificação e Mitigação dos Efeitos da Seca [PAN-Brasil 2004] e da Política Nacional de Combate à Desertificação (Lei nº 13.153 de 30 de julho de 2015). Nesse sentido, propõe-se uma série de estratégias voltadas ao atendimento da neutralidade da terra que envolvem: oferta e disponibilização de tecnologias; proposição de sistemas integrados de produção sustentável e adequados às características locais; ampliação das possibilidades de agregação de valor e de processamento de produtos agropecuários de âmbito regional; capacitação e iniciativas de extensão rural.

Uma cooperação institucional entre a Embrapa e o MMA foi estabelecida em 2019 com o objetivo de auxiliar a atualização da estratégia ministerial de combate à desertificação no bioma Caatinga e a implementação da LDN. A adoção da Caatinga como recorte territorial para essa iniciativa justifica-se por suas características únicas em termos fitofisionomias e biodiversidade, a existência de complexos arranjos sociais e econômicos

[1] https://brasil.un.org/pt-br/sdgs/15

em âmbito regional e por reunir parcela importante de áreas suscetíveis a processos de desertificação.

A proposta da cooperação entre a Embrapa e o MMA tem dois pilares: (1) a caracterização das dinâmicas territoriais no bioma, principalmente associadas à produção agropecuária e as áreas suscetíveis à desertificação e; (2) a sistematização de soluções tecnológicas aderentes ao tema e relevantes para as estratégias de neutralidade da terra. Estes dois pilares foram integrados na **Plataforma de Informações para Combate à Desertificação da Caatinga**, elaborada para materializar os resultados desta cooperação. O objetivo deste artigo é descrever esta plataforma.

Analisar as características territoriais tem se mostrado um importante instrumento para iniciativas de promoção de inovação, a adoção de tecnologias e disseminação de informações [da Silva et al. 2018]. Espera-se que os resultados deste trabalho contribuam para o diagnóstico qualificado das cadeias produtivas da Caatinga e possam oferecer aos produtores rurais, extensionistas, gestores públicos e privados um canal para divulgar e disseminar as soluções tecnológicas disponíveis para a sua região.

2. Bases de Dados da Plataforma

2.1. Tecnologias da Embrapa

Além de referência na produção agrícola mundial, o Brasil se destaca entre os que mais geram e publicam conhecimento em ciências agrárias. A Empresa Brasileira de Pesquisa Agropecuária (Embrapa) tem uma contribuição significativa na pesquisa agrária nacional. Ao longo de sua existência, a Embrapa registra mais de 270 mil publicações técnico-científicas e 1.860 soluções tecnológicas aptas a serem transferidas e aplicadas pelo setor produtivo, disponíveis para acesso público [2].

Apesar do grande volume de informações e tecnologias disponíveis, atingir diferentes públicos-alvo é um constante desafio. Há uma carência de dados organizados, sistematizadas e disponíveis em sistemas de informação e comunicação que cheguem efetivamente aos produtores rurais [Penteado-Filho 2010]. Além disso, a indexação deste conteúdo no âmbito corporativo e, consequentemente, seus mecanismos de busca, não contempla aspectos territoriais e nichos de atuação. Muitas tecnologias lançadas pela Embrapa associadas de forma direta ou indireta aos temas prevenção à desertificação, degradação da terra e recuperação de áreas por manejo florestal ou manejo integrado, se encontram dispersas em diversas categorias de soluções.

Para o desenvolvimento da plataforma descrita neste trabalho, realizou-se de uma seleção inicial de tecnologias pertinentes ao tema a partir da combinação de filtros de busca (por exemplo, desertificação, caatinga, degradação). Adotou-se, ainda, os procedimentos registrados por [Pinto et al. 2019]. A seleção inicial foi refinada a respeito de sua aderência ao conceito de desertificação (controle de processos erosivos, tecnologias de conservação da água e solo) e práticas sustentáveis. A partir desta seleção, pesquisadores de 27 diferentes centros de pesquisa da Embrapa (responsáveis por cada uma das tecnologias selecionadas) e atuantes na Caatinga participaram de um processo interativo de validação e qualificação destas tecnologias, no qual algumas foram excluídas e novas foram elencadas para compor a plataforma.

[2] www.embrapa.br

Neste processo, uma atenção especial foi dada na organização dos materiais de apoio que auxiliam o produtor a efetivamente conhecer e adotar a tecnologia em questão (informações e instruções técnicas registradas em diferentes tipos de mídias como cartilhas/publicações, vídeos e programas de áudio). Além deste, o atributo **onde encontrar** (unidade da Embrapa responsável ou parceiros comerciais) foi revisado e tabulado para identificar os campos importantes que podem facilitar o acesso a cada tecnologia (endereço, telefone de contato e web site).

2.2. Censo Agropecuário

Apesar das fragilidades impostas pelas restrições hídricas e da imprevisibilidade do regime climático, a produção agropecuária da Caatinga é extremamente diversa. Esta diversidade é expressa tanto em variedade e volume produzido como nas diferentes realidades onde as atividades agropecuárias são desenvolvidas. As especificidades regionais, resultado de combinações únicas entre características naturais e sociais, influenciam diretamente as formas de se produzir no campo. Compreender os perfis produtivos da Caatinga é essencial para o planejamento e a adequada execução de políticas públicas de combate à desertificação, dado que o processo de degradação da terra se relaciona diretamente com os resultados da produção agropecuária.

Nesta plataforma, a análise deste perfil produtivo foi baseada na sistematização e qualificação de dados do Censo Agropecuário [3]. Coordenado e executado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em períodos decenais, o censo agropecuário é uma das principais referências para levantamento de dados sociais, econômicos e ambientais sobre o mundo rural brasileiro, e sua pesquisa mais recente foi realizada entre os anos de 2016 e 2017.Trata-se do único levantamento em que a totalidade dos produtores rurais é inserida na amostra, sendo a pesquisa de maior amplitude em termos de conhecimento de características das áreas rurais e da produção agropecuária existente em todo o território nacional.

O Censo é centrado no conceito de *estabelecimento agropecuário*, adotado como unidade básica para obtenção, caracterização e agregação de dados para divulgação. Diferentemente do imóvel rural, o estabelecimento agropecuário é definido como toda unidade de produção/exploração dedicada, total ou parcialmente, a atividades agropecuárias, florestais ou aquícolas, independente de seu tamanho, de sua forma jurídica ou de sua localização, com sua produção destinada à comercialização ou à subsistência [IBGE 2019]. Este conceito admite unidades de produção sem área associada (caso de apicultores ou pescadores, por exemplo).

Foram utilizados dois conjuntos de variáveis do Censo Agropecuário, com finalidades distintas na caracterização dos perfis agropecuários da Caatinga. O primeiro deles apresenta um quadro territorial das principais atividades agropecuárias da região, em termos econômicos e de uso da terra, a partir de duas variáveis e suas categorias específicas: **valor da produção** e **utilização das terras**. Enquanto o valor de produção representa os recursos financeiros gerados a partir da produção e sua comercialização, a utilização das terras apresenta a distribuição do uso das terras em diferentes atividades no interior dos estabelecimentos, considerando inclusive áreas destinadas à preservação ambiental.

[3] https://sidra.ibge.gov.br/pesquisa/censo-agropecuario/censo-agropecuario-2017

O segundo conjunto apresenta indicadores mais específicos, a partir da seleção de produtos agropecuários. O objetivo é evidenciar as dinâmicas inerentes a cada cadeia produtiva e sua representatividade no bioma a partir de quatro variáveis que caracterizam o estabelecimento: enquadramento como agricultura familiar, tamanho de propriedade, condição legal e grupo de atividade econômica (principal). Estas características permitem o diagnóstico detalhado dos estabelecimentos agropecuários e identificação de perfis para políticas públicas, oportunidades de fomento e disseminação tecnológica.

3. Resultados

A plataforma está disponível para acesso público no portal da Embrapa [4] e possui dois pontos de entrada: **Tecnologias para prevenção e combate à desertificação** e **Perfil da agricultura e da pecuária regional**. Visando viabilizar o acesso à plataforma em múltiplos dispositivos, todos os elementos visuais foram customizados para apresentação adequada na perspectiva *mobile*, favorecendo e ampliando o alcance do público alvo que pode fazer uso da ferramenta. As próximas seções descrevem os aspectos tecnológicos que embasam a plataforma e os resultados em cada um dos pontos de entrada.

3.1. Aspectos Tecnológicos

Os dados que embasam a plataforma foram organizados em um banco de dados relacional espacial (PostgreSQL + PostGIS). O diagrama de dados foi elaborado visando a normalização e consistência dos atributos e tendo em vista generalizar as particularidades diante dos seletores previstos para navegação na plataforma.

Os painéis interativos foram construídos utilizando-se o software de visualização de dados Tableau [5]. Conceitualmente, o Tableau categoriza os atributos em dois grupos: dimensões e medidas. Dimensões são valores qualitativos usados para categorização. Medidas correspondem a valores numéricos e quantitativos, passíveis de agregação (soma ou média, por exemplo). Os dados espaciais foram materializados como mapas e os demais atributos foram acomodados em diferentes apresentações visuais, de acordo com a natureza e classes de valores. Buscou-se a representação que melhor comportasse a maior parte dos casos. Os painéis forma publicados na nuvem pública do Tableau, onde os dados estão disponíveis para *download* em diferentes formatos.

Para a estruturação do conteúdo, foram construídas páginas web (HTML/CSS e JavaScript) que incorporam os painéis publicados. Através da API JavaScript do Tableau, os painéis podem ser acessados e filtrados de forma integrada com outros conteúdos descritivos, possibilitando uma maior interação. As páginas resultantes foram publicadas no gerenciador de conteúdos Liferay (ambiente corporativo da Embrapa), materializando a plataforma.

3.2. Tecnologias para prevenção e combate à desertificação

A Figura 1 apresenta o painel publicado junto à página **Tecnologias para prevenção e combate à desertificação**. As tecnologias foram indexadas por 8 categorias (de forma não unívoca): *Conservação da Água e do Solo*, *Sistemas Integrados*, *Manejo Vegetal*, *Manejo Animal*, *Cultivares*, *Insumos e Implementos*, *Gestão e Serviços*, *Agroindústria*.

[4] www.embrapa.br/bioma-caatinga/desertificacao/

[5] www.tableau.com

Estas categorias são utilizadas na navegação da plataforma conforme mostrado no topo Figura 1. Com exceção das duas primeiras categorias, as demais possuem um segundo seletor de navegação referente ao produto agropecuário ao qual esta tecnologia se associa no processo produtivo.

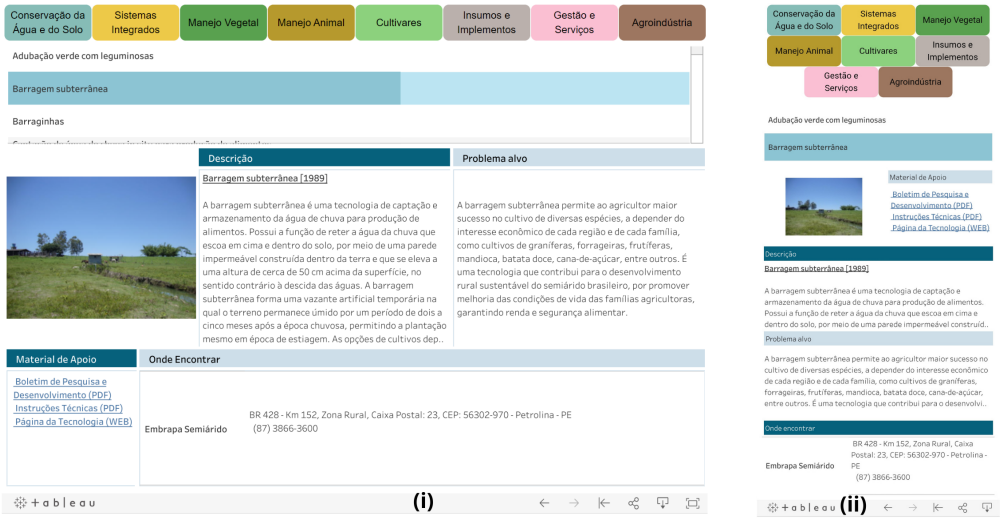


Figura 1. Painel de Tecnologias da Embrapa: (i) Desktop e (ii) *Mobile*

Como pode ser visto na Figura 1 **item (i)**, a listagem de tecnologias é mostrada em uma tabela com rolagem, logo abaixo dos botões para seleção de categoria. Para uma tecnologia selecionada (na figura, **Barragem subterrânea**), é exibida uma imagem representativa da tecnologia, seguida horizontalmente de sua descrição e problema alvo. Abaixo, são mostrados os materiais de apoio disponíveis (um boletim de pesquisa e desenvolvimento, um folder de instruções técnicas e a página da tecnologia na web) e onde é possível encontra-la (**Embrapa Semiárido**). O clique neste último campo encaminha ao site quando disponível (neste caso, o portal institucional da Embrapa Semiárido).

3.3. Perfil da agricultura e da pecuária regional

A Figura 2 apresenta o primeiro painel publicado junto à página **Perfil da agricultura e da pecuária regional**. Ele apresenta dois aspectos: (a) as principais atividades econômicas da Caatinga em termos do valor gerado pela produção agropecuária (gráfico em barras na cor verde, na parte superior); e (b) a distribuição de utilização das terras nos estabelecimentos agropecuários da região (gráfico em barras roxas, na parte inferior). Além da apresentação gráfica, estes dados podem ser explorados de acordo com sua dispersão espacial (mapa em gradiente de cores terrosas, à esquerda do painel). O seletor no topo do mapa permite escolher a variável a ser espacializada no mapa.

É possível interagir com os conteúdos por meio da seleção de componentes em qualquer um dos gráficos (uma barra) ou do mapa (uma microrregião, por exemplo). A partir de uma interação, o painel se recalcula para apresentar as informações da seleção realizada. Ao selecionar uma microrregião no mapa de utilização das terras, por exem-

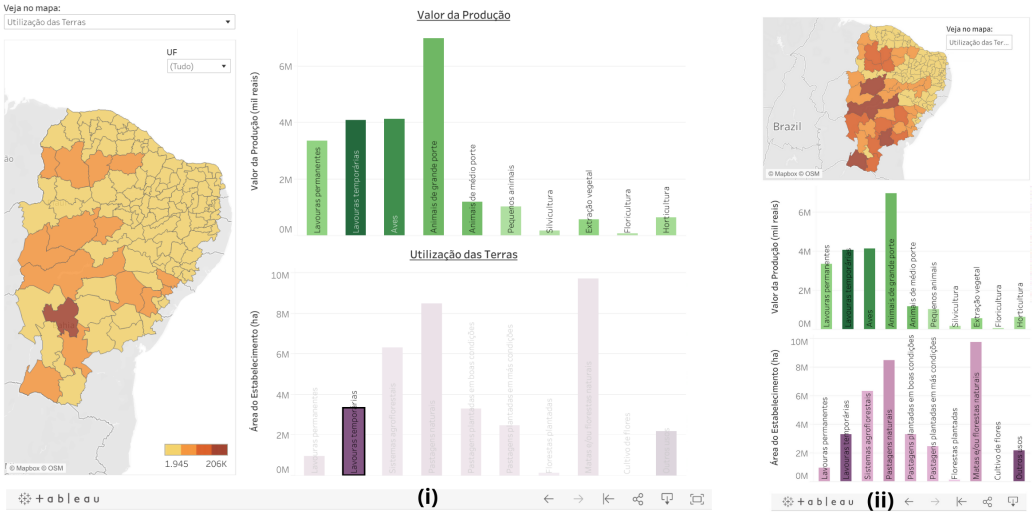


Figura 2. Dados Gerais do Censo Agropecuário na Caatinga: (i) Desktop e (ii) *Mobile*

plo, o gráfico correspondente irá apresentar distribuição das classes de utilização da terra apenas da microrregião selecionada.

O segundo painel, mostrado na Figura 3, apresenta uma visão dos estabelecimentos a partir dos produtos agropecuários. Os produtos foram organizados em 9 grupos, mostrados no topo do painel: *Agroindústria*, *Aquicultura*, *Extração Vegetal*, *Lavoura Permanente*, *Lavoura Temporária*, *Leite de Vaca*, *Pecuária*, *Silvicultura* e *Total* (representando todos os produtos). A partir da seleção de um grupo, é possível selecionar o produto e variável de interesse. De maneira geral, foram disponibilizadas duas variáveis: o número de estabelecimentos agropecuários e o volume produzido, que pode ser a quantidade em toneladas para lavouras, o número de caixas de colmeias para apicultura ou a quantidade de cabeças para pecuária.

O painel mostra, além da visão espacial dos dados (mapa à direita do painel, em gradiente de tons verdes e amarelos), quatro aspectos de caracterização: (a) enquadramento como agricultura familiar (em gráfico de pizza, à esquerda superior); (b) faixas de tamanho das propriedades (em gráfico de barras azuis, ao centro); (c) grupo de atividade econômica principal da propriedade (em mapa de árvore, à esquerda inferior); e (d) condição do produtor em relação à terra (em formato tabular, centralizado). Este painel tem um comportamento análogo ao anterior – a partir de seleção de uma microrregião no mapa (ou um conjunto de microrregiões), o painel se recalcula para apresentar as informações detalhadas de acordo com a seleção realizada.

A integração com o componente espacial nestes painéis permite obter tanto a visão geral do bioma como também o perfil detalhado de um estado, uma área de interesse ou uma microrregião. Isto permite, de forma concomitante, uma análise do perfil de variáveis com a distribuição territorial, evidenciando especificidades regionais produtivas, como padrões de uso e ocupação da terra ou presença da agricultura familiar na região.

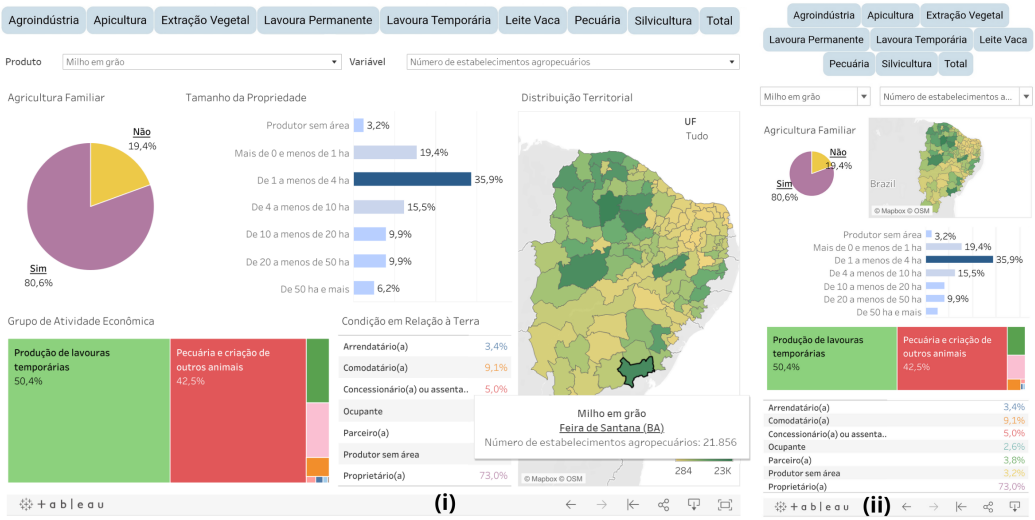


Figura 3. Perfil da agricultura e da pecuária da Caatinga por produto: (i) Desktop e (ii) *Mobile*

4. Contribuições e Trabalhos Futuros

Após um vasto trabalho de filtragem e organização dos dados da Caatinga, foi possível construir uma base de dados espacial que agrega diferentes aspectos do bioma e, a partir dela, elaborar um conjunto de painéis interativos para exploração e análise dos dados. Os conteúdos presentes na plataforma por meio dos painéis permitem visualizar tendências sobre as principais características territoriais da agropecuária da Caatinga. Assim, essa visão é complementar à disponibilização das tecnologias da Embrapa no tema da prevenção e combate à desertificação na medida em que a utilização desses ativos deve ser necessariamente ancorada em contextos territoriais adequados à sua adoção.

A integração das informações, no âmbito das tecnologias e das características produtivas das regiões, amplia as possibilidades de associação entre os potenciais oferecidos por uma solução tecnológica ou por um conjunto de soluções frente às demandas e as características existentes em cada local, em diferentes realidades sociais, econômicas e ambientais. Quanto maior o conhecimento acerca das diferentes realidades geográficas associadas à desertificação, maior é a capacidade de planejamento e execução de políticas públicas assertivas, de promover o desenvolvimento agropecuário sustentável e de propor ações diferenciadas. Espera-se que esta visão espacial possa fornecer subsídios à elaboração de ações público/privadas no que se refere a este tema.

A plataforma está atualmente em fase de validação pelo MMA e parceiros da extensão rural. Os próximos passos abrangem a verificação da adequabilidade das apresentações visuais propostas (tipos de gráficos e adequações de cores e rótulos) e a estruturação de um processo para atualização contínua dos dados das tecnologias da Embrapa, dada a dinâmica necessária para evitar que a plataforma se torne obsoleta.

Referências

Cowie, A. L., Orr, B. J., Castillo Sanchez, V. M., Chasek, P., Crossman, N. D., Erlewein, A., Louwagie, G., Maron, M., Metternicht, G. I., Minelli, S., Tengberg, A. E., Walter, S., e Welton, S. (2018). Land in balance: The scientific conceptual framework for land degradation neutrality. *Environmental Science & Policy*, 79:25–35.

Critchley, W., Harari, N., e Mekdaschi-Studer, R. (2021). Restoring Life to the Land: The Role of Sustainable Land Management in Ecosystem Restoration. UNCCD and WOCAT.

da Silva, J. P., Guarneri, H., Arenas, F. C., de Paula, E. V., e Camboim, S. P. (2018). Uso de um Dashboard Geoespacial como ferramenta de suporte para o diagnóstico socioeconômico e ambiental da Reserva Biológica Bom Jesus – Litoral do Paraná. In *Proceedings XIX GEOINFO*, pages 152–157, Campina Grande, PB.

IBGE (2019). *Censo Agropecuário 2017: Resultados Definitivos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro.

PAN-Brasil (2004). Programa de ação nacional de combate a desertificação em mitigação dos efeitos da seca. Relatório Técnico, Ministério do Meio Ambiente/Secretaria de Recursos Hidrícos, Brasília.

Penteado-Filho, R. C. (2010). Diagnóstico e propostas de soluções para a transferência de tecnologia numa empresa pública de pesquisa agropecuária: o caso Embrapa. *Prisma.com (Portugual)*, 11:86–102.

Pinto, D. M., Magalhães, L. A., Takemura, C. M., Farias, A. R., de Almeida, B. T., dos Santos Brandão, V. V., e da Conceição, A. R. (2019). Metodologia para espacialização das soluções tecnológicas da Embrapa. Circular Técnica 34, Embrapa Territorial, Campinas.

UNCCD (1994). United Nations Convention to Combat Desertification in Those Countries Experiencing Serious Drought and/or Desertification Particularly in Africa: Text with Annexes. United Nations Convention to Combat Desertification. UNEP, Nairobi.

UNCCD (2019). Land Degradation Neutrality for Biodiversity Conservation. United Nations Convention to Combat Desertification. Briefing Note. Bonn, German.



Uso de *workflows ETLH* para integrar *datasets* pedológicos: estudo para adequação aos princípios FAIR

Sabrina Santos Cruz de Oliveira[1], Emerson de Barros Duarte[1], Élton Carneiro Marinho[2], Sérgio Manuel Serra da Cruz[1,2]

[1]Programa de Pós-Graduação Interdisciplinar em Humanidades Digitais – Universidade Federal Rural do Rio de Janeiro
Caixa Postal 74.583 – 26.285-060 – Nova Iguaçu – RJ – Brasil

[2]Programa de Pós-Graduação em Informática – Universidade Federal do Rio de Janeiro
{sabrina, serra}@pet-si.ufrrj.br, emersonbd@ufrrj.br,
elton.marinho@ppgi.ufrj.br

Abstract. *Currently, datasets from pedological projects are isolated in silos under the most varied formats. The objective of this work is to present an approach and experiments based on ETLH and FAIRification workflows capable of loading, cleaning, transforming, identifying and harmonizing large masses of legacy data on the OpenSoils platform. Additionally, we discuss the main steps of the FAIRification process.*

Resumo. *Atualmente, os datasets oriundos de projetos pedológicos se encontram isolados em silos sob os mais variados formatos. O objetivo deste trabalho é apresentar uma abordagem e experimentos baseados em workflows ETLH e de FAIRificação capazes de carregar, limpar, transformar, identificar e harmonizar grandes massas de dados legados na plataforma OpenSoils. Adicionalmente, discutimos as principais etapas do processo de FAIRificação.*

1. Introdução

Estudos recentes da Embrapa, SEBRAE e INPE[1] indicam que a agricultura digital apoiada na Ciência de Dados e na acelerada adoção de tecnologias digitais têm o potencial de transformar importantes setores da cadeia do agronegócio brasileiro. O estudo indica que a internet no campo é o principal atrativo para os agricultores que buscam ampliar ou conquistar novos mercados, reduzir custos ou agregar valor à produção. Portanto, é possível conjecturar que, em breve, essa cadeia, demandará novas práticas

[1]https://www.embrapa.br/en/busca-de-noticias/-/noticia/54770717/pesquisa-mostra-o-retrato-da-agricultura-digital-brasileira

correlacionadas com a gestão de grandes volumes de dados para auxiliar nos processos de tomada de decisão.

Segundo Jonquet *et al*. (2018), os dados agronômicos são de difícil integração e interoperabilidade tanto do ponto de vista técnico quanto semântico. Análises mais refinadas dos *datasets* produzidos nesta cadeia podem auxiliar tanto em sistemas mais inteligentes quanto numa produção mais sustentável [Cruz *et al*., 2018]. No caso dos solos, as dificuldades também são grandes. Portanto, disponibilizar sistemas mais transparentes e sustentáveis que indiquem os melhores tipos de solos para usos adequados, sequestro de carbono, racionalização do uso de recursos naturais e redução da aplicação fertilizantes e produtos químicos são de grande relevância.

Os desafios e as oportunidades para desenvolver pesquisas e inovação na cadeia do agronegócio são variadas, mas ainda timidamente exploradas pela comunidade brasileira de computação. Neste trabalho focaremos em um dos principais recursos da cadeia: os dados de solos. Segundo Bessa e Cruz (2021), esse é um dos elementos centrais na ligação entre diversas áreas do conhecimento, tais como computação, segurança de solos e humanidades digitais.

Como objeto de estudo, investigaremos as necessidades do processo de FAIRificação de dados relacionados aos *datasets* de dados pedológicos. Destacamos que a Pedologia é uma ciência complexa que ainda não se apropriou da perspectiva da Ciência Aberta para a gestão de seus *datasets* apoiados pelos princípios FAIR [Wilkinson *et al*., 2016]. Do ponto de vista da computação, determinados aspectos da Pedologia podem ser considerados como uma ciência tão intensiva em dados quanto a Bioinformática pois envolve a correlação de múltiplas variáveis ambientais [Cruz *et al*., 2018].

Segundo Oliveira *et al*. (2021), os *datasets* de dados pedológicos caracterizam-se por serem grandes conjuntos de dados oriundos de grandes projetos de mapeamento de solos que ocorreram nos últimos 60 anos. Por exemplo, o RADAMBRASIL[2], projeto governamental criado em 1970, realizou um inventário de grande parte do País usando mapeamento em escala 1:1.000.000. Em geral, esses dados não são disponibilizados para a sociedade brasileira como dados abertos governamentais, muitas vezes permanecendo inacessíveis. Sob a ótica da Ciência de Dados, a maioria desses *datasets* apresentam duas categorias de problemas principais:

- ***Problemas Estruturais*** – Segundo Rosa & Anjos (2020), há pouco acesso e estímulos a políticas de dados primários abertos na área. Adicionalmente, os arquivos são dispersos em vários servidores, sendo de tipos e formatos diversos, muitas vezes incompletos, irregulares, contendo *outliers* e mesmo falhas de preenchimentos. Adicionalmente, os dados são pouco transparentes e estão desconectados em silos de dados.

- ***Problemas Semânticos*** – Via de regra, os solos do Brasil são classificados de acordo com taxonomias que são periodicamente atualizadas. Desde 2017 é utilizada a 5ª edição do Sistema Brasileiro de Classificação de Solos (SiBCS) [Santos, 2018]. As classes taxonômicas descrevem propriedades e atributos pedológicos cujas classificações podem variar no tempo. No entanto, devido a uma série de condicionantes, os solos

[2] https://www.embrapa.br/en/pronasolos

previamente classificados com regras anteriores não são reclassificados automaticamente à medida que novas taxonomias são liberadas. Essa dinâmica, involuntariamente, agrega inconsistências semânticas nos *datasets* se forem (re)analisados à luz da taxonomia corrente. Adicionalmente, se constata que atualmente inexistem *datasets* públicos e abertos totalmente alinhados aos princípios FAIR ou aderentes aos princípios da Ciência Aberta.

Essas categorias de problemas se evidenciam em duas bases públicas de dados de solos do Brasil: BDSolos[3] e no repositório FeBR[4]. Ambas são de grande valor, armazenam expressivos volumes de dados legados contendo estruturas e atributos distintos. No entanto, são de difícil integração e por conseguinte de difícil atualização, sua integração demanda esforços de processamento, isso muitas vezes é oneroso para pesquisadores e instituições, levando baixo reuso por parte da comunidade.

Como contribuição, este trabalho apresenta *workflows* que desempenham tarefas de Extração-Transformação-Carga-Harmonização (*ETLH*) e de FAIRificação que consideram a taxonomia SiBCS atual no contexto do tratamento e estabelecimento de relacionamentos entre dados legados de projetos pedológicos. Os experimentos foram capazes de tratar grandes volumes de dados pedológicos legados, transformando-os em dados harmonizados e anotados com proveniência retrospectiva e carregando-os diretamente na plataforma *OpenSoils* [Cruz *et al*., 2018, 2019]. Além disso, este trabalho aprofunda discussões sobre o trabalho de Oliveira *et al*. (2021), pontuando as adequações dos dados para que se alinhem aos princípios FAIR, onde discutimos os pontos que ainda necessitam de evolução e melhorias.

Este trabalho está organizado da seguinte forma: as seções 2, 3 e 4 abordarão uma visão geral sobre *datasets* pedológicos, princípios FAIR e proveniência de dados, respectivamente. A seção 5 apresenta os trabalhos relacionados, a seção 6 os materiais e métodos e na seção 7 discutimos os resultados obtidos até o momento. Na seção 8 apresentamos as conclusões e indicações de trabalhos futuros.

2. *Datasets* Pedológicos

Fisicamente o solo é composto por diversos materiais dispostos em camadas; sua análise é conduzida através de aberturas denominadas tradagens, trincheiras (ou perfis) que são realizadas diretamente no campo onde pedólogos, agrônomos ou geólogos coletam diversos tipos de dados ambientais e morfológicos georreferenciados (imagens, descrições das camadas, profundidades, composição física e química, transições, cores, erosões, texturas, entre outros atributos das camadas) [Santos, 2018].

Os dados de solos são obtidos através de observações diretas e em experimentos que algumas vezes podem ser considerados de difícil reprodutibilidade. Essas atividades são realizadas diretamente no campo ou em laboratórios "de campanha" e podem envolver muitos profissionais, instrumentos e sensores. Os dados são posteriormente complementados por diversos tipos de análises laboratoriais realizadas em ambientes especializados. Em geral, os *datasets* pedológicos são volumosos e gerados por equipes distintas com especialidades diversas e dispersas no tempo e, por vezes, geograficamente

[3] https://www.bdsolos.cnptia.embrapa.br/consulta_publica.html
[4] https://www.pedometria.org/febr/

distantes. Com isso, acabam produzindo, involuntariamente, *datasets* desconectados que contém várias falhas em séries de dados não harmonizadas.

Uma parte dos dados de solos já está dispersa na Web ou em repositórios públicos e privados formando silos de dados desconectados. Constata-se que a recuperação dos *datasets* não é um processo trivial e seu acesso não é transparente e nem mesmo automatizado, o que dificulta sua reutilização tanto por gestores quanto por pesquisadores. Porém, essa situação penaliza especialmente os agricultores que muitas vezes se veem excluídos do acesso a esse patrimônio digital.

Muitos *datasets* não são classificáveis como dados abertos. Por exemplo, existem apenas seis *datasets* abertos no portal de dados abertos do Governo Federal[5]. Mesmo no sítio do recentíssimo programa PRONASOLOS[6], os *datasets* não estão abertos e nem disponíveis ao público. Neste caso, atualmente, são ofertadas apenas interfaces de visualização de dados consolidados sob a forma de mapas.

Muitas vezes tais dados ainda estão dispersos em mapas de papel ou mesmo em arquivos fechados e planilhas armazenadas em servidores sob os mais variados formatos. Não raro, os dados são apresentados de modo agregado na forma de tabelas com estatísticas ou possuem diferentes estruturas semânticas, com poucos metadados descritivos, ausência quase absoluta de rastros de proveniência [Cruz *et al*., 2019, Bessa *et al*., 2021], por fim, os dados não são harmonizados. Isto é, vários *datasets* não usam as mesmas classificações taxonômicas, variáveis ou padrões de unidades de medidas para os mesmos atributos pedológicos ao longo do tempo, acentuando ainda mais as inconsistências dos dados.

Do ponto de vista da Ciência de Dados, esse conjunto de características prejudica e retarda o desenvolvimento de trabalhos de pesquisa e as entregas de soluções para a sociedade civil. Resumidamente, elas reduzem a vida útil dos dados. Portanto, estudar e difundir práticas mais eficiente ligadas à Ciência Aberta e uso de repositórios públicos confiáveis e focados em preservação digital e baseadas em princípios FAIR e "FAIRificação" de dados poderão trazer maior reprodutibilidade das pesquisas e visibilidade aos *datasets*.

3. Princípios FAIR

Os princípios FAIR foram originalmente publicados em 2016 com o objetivo de fornecer orientações para a aplicação em repositório de dados [Wilkinson *et al*., 2016]. FAIR é um acrônimo para *Findable* (Localizável), *Accessible* (Acessível), *Interoperable* (Interoperável) e *Reusable* (Reutilizável). São princípios orientadores de alto nível que podem ser aplicados em diversas áreas do conhecimento, incluindo dados da cadeia do agronegócio. No caso das ciências agrárias, os princípios FAIR visam promover a agregação e uniformização dos dados e dos sistemas, incluindo os dados pedológicos, sob protocolos que consideram aspectos éticos, legais, culturais e barreiras técnicas, reduzindo custos de gestão dos dados e assegurando sua qualidade e transparência.

Os dados pedológicos e seus metadados devem ser fáceis de serem localizados e individualizados por identificadores persistentes, para que sejam consumidos tanto por humanos quanto por máquinas (*Findable*). Após localizados, os dados precisam ser acessíveis, ou recuperáveis através do seu identificador, mesmos que os dados originais já

[5] https://dados.gov.br/dataset?q=solos&tags=Brasil
[6] http://pronasolos.agenciazetta.ufla.br/

não estejam disponíveis, possivelmente incluindo autenticação e autorização e protocolos comuns (*Accessible*). Os dados precisam se integrar a outros dados (*Interoperable*), precisam interoperar com aplicativos ou fluxos de trabalho distintos para suportar análises, armazenamento e processamento mais eficazes. Logo, o suporte de estruturas semânticas (ontologias, taxonomias, etc.) é essencial. Tudo isso para que os *datasets* possam ser mantidos e reutilizados ao longo do tempo. Os metadados e os dados devem ser bem descritos, para que possam ser replicados e/ou (re)combinados em diferentes configurações (*Reusable*)[7]. Cada um dos quatro princípios possui subdivisões, além das descrições e características sobre sua aplicação. Graças a isso, hoje somos capazes de realizar análises e traçar metas para tornar os dados, independentemente de sua origem e projetar processos de FAIRificados de dados.

FAIRificação é o processo de tornar os dados brutos aderentes aos princípios FAIR, tornando-os inteligíveis para humanos ou máquinas [Veiga *et al.*, 2019]. Esse processo ainda é pouco explorado na agricultura digital e no agronegócio com seus dados agrícolas e pedológicos. No Brasil, os primeiros passos começam a ser trilhados na área Agro através do GO-FAIR Brasil[8].

4. Proveniência de dados

O tema proveniência de dados ainda é pouco difundido na cadeia do agronegócio. O conceito, consagrado na área da computação, refere-se à origem ou à procedência de um determinado objeto [Buneman; Khanna; Tan, 2000 e Davidson & Freire, 2008]. Entretanto, os aspectos fundamentais da proveniência não se resumem apenas aos dados, mas também aos processos e aos agentes transformadores. Apesar de serem aplicados há tempos nas áreas de banco de dados, e-ciência, e humanidades digitais, os estudos de proveniência de dados na área da pedologia e mesmo em ciência de dados ainda são incipientes.

A proveniência dos dados aumenta a confiança do consumidor de dados [Allemang; Bobbin, 2016]. A família de documentos PROV[9] da W3C define um modelo e serializações para permitir o intercâmbio interoperável de informações de proveniência em ambientes heterogêneos como a web.

A proveniência pode ser classificada como prospectiva quando captura a especificação de uma tarefa computacional, seu fluxo de trabalho. Descreve as etapas que devem ser seguidas para gerar um determinado tipo de dado. É a captura de uma especificação abstrata de fluxo de trabalho como uma receita para derivação futura de dados [Lim *et al*., 2010]. A proveniência retrospectiva captura os dados, as etapas que foram executadas e informações sobre este ambiente [Da Cruz; Do Nascimento, 2016]. Neste trabalho adotamos a retrospectiva de baixa granularidade nos *workflows ETLH* e de FAIRificação dos dados.

5. Trabalhos relacionados

O Brasil possui poucos bancos de dados de solos, a maioria dos dados encontram-se em repositórios dispersos e isolados. O BDSolos é um banco de dados relacional

[7] https://www.go-fair.org/fair-principles/
[8] https://www.go-fair.org/go-fair-initiative/go-fair-offices/go-fair-brazil-office/
[9] Um conjunto de documentos que definem vários aspectos que são necessários para alcançar a visão de intercâmbio interoperável de informações de proveniência em ambientes heterogêneos como a web.

desenvolvido pela EMBRAPA que demanda alto conhecimento de solos para que seja plenamente consultado, além disso apresenta limitações na extração de dados. Por outro lado, o FeBR surgiu como um disco virtual público, mas evoluiu para um conjunto de planilhas de projetos pedológicos. Em linhas gerais, adotou alguns princípios da Ciência Aberta e, importou parte dos dados do BDSolos, disponibilizando-o sob a forma de planilhas que possibilitam que qualquer pessoa, mesmo com poucos conhecimentos sobre pedologia, consiga ter acesso aos projetos pedológicos.

A organização no FeBR consiste, basicamente, em três arquivos no formato .txt para cada projeto, com valores separados por tabulação. Um contém dados de identificação do projeto, outro os dados das observações realizadas no projeto e por fim, um com os dados pedológicos das descrições das camadas. Até o momento, não se verificou maior alinhamento aos princípios FAIR ou mesmo aporte do leque de tecnologias de Web Semântica. Em termos internacionais, organizações tais como a ISRIC[10] e a OGC[11] já advogam abertamente a necessidade da adoção dos princípios FAIR como um dos mecanismos de padronização de dados para a área de solos.

6. Materiais e métodos

6.1. Opensoils

OpenSoils (www.opensoils.org) é uma plataforma computacional gratuita, aberta, elástica, distribuída, multiusuário e multicamada. É orientada para armazenar dados curados e harmonizados (primários e secundários) de solos do Brasil e seus metadados de proveniência [Cruz *et al*., 2018, 2019].

Atualmente, o *OpenSoils* conta com uma versão Web e aplicativos móveis, que se comunicam através de APIs e troca síncrona de dados. O banco de dados do *OpenSoils* é do tipo relacional, seu esquema lógico é capaz de armazenar uma grande quantidade de dados pedológicos que seguem definições presentes na 5a. edição da taxonomia SiBCS.

O *schema* completo possui 46 tabelas. É possível correlacionar dados produzidos por processos observacionais de campo com os produzidos em laboratórios. Por exemplo, há integração de dados coletados em campo (tabelas *projeto, observação, relevo, descrição geral, horizontes, morfologia, entre outros*), além de dados obtidos em laboratório através de experimentos físicos (*curva de retenção de água*) e químicos (*ataque sulfúrico, pasta saturada*).

6.2. *Workflows ETLH* e de Fairificação de Dados

Para a carga de dados legados no banco de dados relacional do *OpenSoils*, foram desenvolvidos *workflows ETLH* a partir da ferramenta Pentaho Data Integration (PDI)[12]. Os *workflows ETLH*, originalmente descritos por Cruz *et al*. (2021), são compostos por fluxos de *tasks* e *jobs*. As *tasks* são as unidades "mínimas de processamento" dentro do processamento dos *jobs* de um *workflow*. As *tasks* variam desde simples tarefas de conexão e captura de dados, verificação de nulos, duplicatas, outliers, por exemplo validação, harmonização de unidades/dados, ajuste de sistemas de coordenada, até tarefas classificatórias descritas pela taxonomia.

[10] https://www.isric.org/international-soil-standards
[11] https://www.ogc.org/
[12] https://help.hitachivantara.com/Documentation/Pentaho/7.1/0D0/Pentaho_Data_Integration

As primeiras *tasks* dos *jobs* são responsáveis pela conexão e ingestão de dados no *schema* conceitual do banco de dados da plataforma, as demais são responsáveis pelas validações e harmonizações segundo a SiBCS (por exemplo: validações de classes e outros atributos tais como relevo, descrição geral, observação, erosão, horizonte, propriedade química do horizonte, propriedade física do horizonte, propriedade morfológica, etc).

Um dos maiores desafios enfrentados na construção dos *workflows ETLH*, foi a inexistência no PDI de uma funcionalidade que realizasse automaticamente o relacionamento entre os registros dos *datasets* e as tabelas da plataforma *OpenSoils* (característica essencial no processo de FAIRificação de dados). Nesse sentido, utilizamos a *task* Execute SQL, um script capaz de executar qualquer declaração SQL diretamente no banco de dados. Assim, após a carga dos dados, os *scripts* SQL são executados para realizar as atualizações das chaves estrangeiras e manter o relacionamento entre os dados de um mesmo projeto.

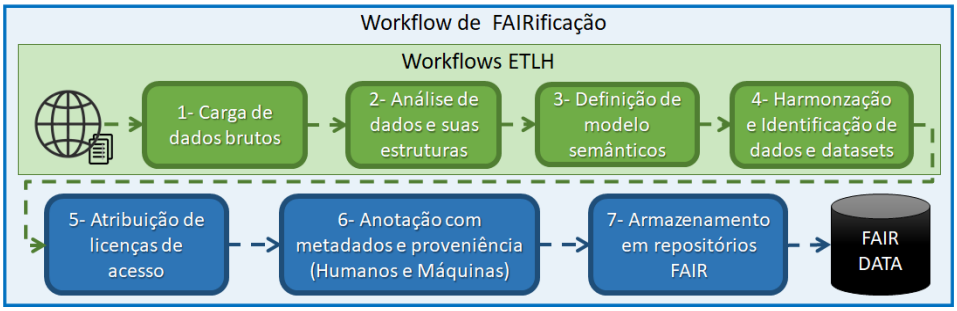


Figura 1. Representação conceitual das etapas do processo de *ETLH* e FAIRificação de dados.

A Figura 1, ilustra, o relacionamento entre os *workflows ETLH* e de FAIRificação de dados da plataforma *OpenSoils*. Em geral, a FAIRificação dos dados representa um processo de alto nível que pode ser materializado sob diversas etapas e tecnologias [Goble *et al*., 2020]. As principais etapas da FAIRificação são: 1) mapeamento e carga de *datasets (*dados brutos) disponíveis na Web; 2) análise dos dados para verificar conteúdo; conceitos representados, estruturas, as relações entre os elementos que constituem os dados; 3) definição/escolha de um modelo semântico para a representação do conjunto de dados (idealmente a partir de vocabulários ou ontologia bem fundamentadas), fornecendo uma estrutura para organizar/estruturar os dados sem ambiguidades; 4) permitir a identificação e harmonização dos dados, promovendo a interoperabilidade e a integração com outros tipos de dados e sistemas; 5) atribuir uma licença/autorização para acessar aos dados; 6) anotar os dados com metadados e proveniência permitindo que tanto seres humanos como máquinas possam localizá-lo; 7) promover o armazenamento (dados e metadados) a longo-prazo em repositórios FAIR e/ou publicar os dados FAIRificados adicionados de licença para que, os metadados possam ser indexados e localizados por mecanismos de pesquisa.

7. Resultados e Discussão

7.1 Experimentos

Os experimentos iniciais, foram mediados pelos *workflows ETLH* coletaram dados de perfis de solos dispersos em centenas de arquivos registrados no FeBR, além das tabelas do BDSolos. A execução dos experimentos modelados como um *job* e composto por *tasks*, processou 193 projetos pedológicos, totalizando mais 9 mil perfis de solos e mais de 800 mil registros carregados na plataforma *OpenSoils* em pouco mais de 2 horas de processamento. Cada um deles passou pelo processo de *ETLH*, renomeação de colunas, relacionamento e consistência de chaves e registro de proveniência.

Os dados produzidos pelos experimentos são abertos e oriundos de projetos pedológicos realizados durante décadas em todo o Brasil, podendo ser integralmente acessados tanto pela plataforma *OpenSoils Web* e sua API quanto pelo app *OpenSoils* Edu [Cruz, 2018], já disponível na *Google Play Store* e na *App Store*.

7.2 Discussão sobre a FAIRificação

Até o momento, o banco de dados do *OpenSoils* adequa-se a pelo menos uma ou mais subdivisões de cada um dos princípios FAIR. Nesta seção pontuamos as adequações já existentes e as que ainda precisam ser concluídas em novos desdobramentos da pesquisa.

7.2.1. Localizável

Embora seja um requisito funcional da plataforma, os dados harmonizados ainda não estão plenamente linkados na Web de Dados, como na DBPedia. Apesar de identificáveis, por enquanto, encontram-se em banco de dados relacional. Entretanto, os metadados de descrição são padronizados e um dicionário de dados possibilita a compreensão de cada uma das tabelas e colunas que compõem o banco atualmente. Essa característica poderá ser explorada nas etapas do *workflow* de FAIRificação de dados.

7.2.2. Acessível

Os dados podem ser acessados pela plataforma Web ou pela plataforma móvel, como citado na seção 7.1. Podem ser acessados por outros sistemas abertos através de API. Os metadados possibilitam a construção de diversas formas de busca dos dados e se mantêm acessíveis constantemente, mesmo que, em algum momento, algum dado não se encontre mais disponível. Assim como os dados, os metadados possuem acesso facilitado pelas plataformas.

7.2.3. Interoperável

A partir da harmonização dos dados realizada pelos *workflows ETLH,* conseguimos atingir uma maior qualidade dos dados. Os dados encontram-se prontos para serem compartilhados e utilizados na representação do conhecimento. Os dados fazem referência a outros dados existentes na plataforma, integrando relacionamentos persistentes entre si.

7.2.4. Reutilizável

Desenvolvemos, junto aos *workflows ETLH*, um processo de registro e coleta de proveniência retrospectiva bem definidos apoiada na especificação PROV. Após a carga de novos *datasets* ou alteração de dados já existentes, uma tabela é atualizada com um novo registro que descreve quem, o que e como foi feita aquela operação, além de guardar o novo e o antigo valor. Um dos pontos a ser incorporado ao projeto é a anonimização

dos dados de usuários/projetos de forma a permitir sua reutilização, assegurando a privacidade e sob as normas definidas pela Lei Geral da Proteção de Dados Pessoais[13] (LGPD). No caso de necessidade de exportação para repositórios FAIR deverá proceder-se a técnicas de pseudo minimização, tais como a encriptação e utilização de códigos identificadores de documentos.

8. Conclusão

Podemos inferir que nesta era de constantes avanços da agricultura digital teremos grandes demandas relacionadas a gestão de dados devido a necessidade de reutilizar dados para desenvolver novos produtos ou serviços inovadores para atender os diversos atores da cadeia do agronegócio brasileiro. Os solos do Brasil são um elemento central nessa cadeia, porém, verificam-se poucas pesquisas sobre a gestão eficiente de longo prazo dos dados pedológicos.

Os atuais sistemas e repositórios apresentam gargalos relacionados à alta dispersão e baixa integração de dados, resultando em pouca transparência e informações limitadas de proveniência de dados. Essas condições reduzem a acessibilidade, interoperabilidade e reuso de dados pedológicos.

O estudo dos princípios FAIR em agricultura digital e humanidades digitais são recentes e implementados de forma tímida devido à resistência dos investigadores. Este artigo apresenta um esforço para compreender esse contexto e oferecer novos elementos que possam colaborar com a mitigação das inconformidades relacionadas às limitações dos *datasets* pedológicos e produzir dados de solos FAIRificados e harmonizados. Desenvolvemos *workflows ETLH* acopláveis à plataforma *OpenSoils* que, até o momento, se mostraram capazes em apoiar a FAIRificação segundo a taxonomia da SIBICS, anotando dados com metadados de proveniência retrospectiva e identificadores. Além disso, buscamos analisar as lacunas existentes para adequação dos dados aos princípios FAIR na plataforma *OpenSoils*.

A forma como os projetos de FAIRificação são conduzidos depende do orçamento disponível e do tipo e tamanho da organização. Nossos experimentos processaram centenas de milhares de registros de perfis de todos os tipos de solos em todo o Brasil. Na prática, esses dados já podem ser exportados ou acessados com maior praticidade e agilidade a partir dos aplicativos móveis ou da API do *OpenSoils*. Como trabalhos futuros, além da adequação integral aos princípios FAIR, pretendemos linkar os dados na Web de Dados e incorporar uma estrutura de *FAIR Digital Objects*, possibilitando assim a integração com diferentes bases de dados já disponíveis.

Referências

Bessa, A. C. F.; Cruz, S. M. S. (2021). Investigando a adoção de Princípios FAIR e Proveniência de Dados na Agricultura Digital sob a perspectiva das Humanidades Digitais: um estudo de caso na plataforma OpenSoils. In: HDRIO2021, 2021, Rio de Janeiro. II Congresso Internacional em Humanidades Digitais.

[13] http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Cruz, S. M S., *et al*. (2019). Desenvolvendo Sistemas Agrícolas de Próxima Geração: Um Estudo em Ciência de Solos. In Anais do X Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais (pp. 135-144). SBC.

Cruz, S. M. S. *et al*. (2018) "Towards an e-infrastructure for Open Science in Soils Security". In: XII BRESCI 2018, 2018, Recife. Proceedings of the XII Brazilian E Science Workshop. Porto Alegre: SBC.

Da Cruz, S. M. S; Do Nascimento, J.A.P. (2016). SisGExp: Rethinking Long-Tail Agronomic Experiments. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). [S. l.: s. n.], vol. 9672, p. 214–217. https://doi.org/10.1007/978-3-319-40593-3_24.

Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., ... & Schober, D. (2020). FAIR computational workflows. Data Intelligence, 2(1-2), 108-121.

Oliveira, S. S. C., de Barros Duarte, E., Marinho, E. C., & da Cruz, S. M. S. (2021, September). Integração de Data Lakes Pedológicos através de Workflows ETLH. In Anais da VII Escola Regional de Sistemas de Informação do Rio de Janeiro (pp. 48-55). SBC.

Marinho, E. C. *et al*. (2020). "Proteção de Dados: Proposta de gerenciamento de dados de solos usando os princípios FAIR e a tecnologia blockchain". In: 10ª. Conferencia de Directores de Tecnología de Información y Comunicación en Instituciones de Educación Superior, TICAL2020 y 4° Encuentro Latinoamericano de e-Ciencia. Equador.

Rosa, A. S., Anjos, M. A. (2020). Uma plataforma para facilitar o acesso aos dados do Repositório Brasileiro Livre para Dados Abertos do Solo. SEI-SICITE.

Santos, H. G. et al. (2018). Sistema brasileiro de classificação de solos. 5. ed. rev. e ampl. Brasília, DF: Embrapa.

Simitsis, A. (2003, September). Modeling and managing ETL processes. In *VLDB PhD Workshop* (Vol. 76).

Veiga, V. S. de O., Henning, P., Dib, S., Penedo, E., Lima, J. D. C., Silva, L. O. B. da, & Pires, L. F. (2019). Plano de gestão de dados fair: uma proposta para a Fiocruz | Fair data management plan: a proposal for Fiocruz. Liinc Em Revista, 15(2). https://doi.org/10.18617/liinc.v15i2.5030

Wilkinson, M.; Dumontier, M.; Albersberg, I. *et al*. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Avaliação da eficiência do classificador Iso Cluster para o monitoramento de desmatamento na Amazônia Legal

Adriana M. Iora[1], Andressa Boni[1], Igor Vani[1], Sandro L. B. Fensterseifer[2]

[1] Graduando curso Engenharia Florestal, Universidade Federal Santa Maria - Campus Frederico Westphalen (UFSM).

[2] Professor Doutor em Engenharia Florestal – Departamento Engenharia Florestal - Universidade Federal Santa Maria - Campus Frederico Westphalen (UFSM).

adriiora26@gmail.com, andressa.boni@acad.ufsm.br, vani.igor@acad.ufsm.br, sandrolbf@gmail.com.

Abstract: *Tropical forests are essential for various ecosystem functions, but deforestation is responsible for its destruction.Therefore, the importance of monitoring programs that use remote sensing images to analyze,identify and quantify changes in forest cover. The objective of this work was to evaluate the efficiency of the Iso Cluster classifier to monitor deforestation in the Legal Amazon. For the conduction of this study, the images of four scenes, located in the state of Mato Grosso, were used,obtained by sentinel 2 satellite and then processed in arcgis 10.3.1 software. Despite the need for some corrections, the Iso Cluster classifier showed good results.*

Resumo: *As florestas tropicais são essenciais para diversas funções ecossistêmicas, porém o desmatamento é responsável pela sua destruição. Por isso, a importância de programas de monitoramento que utilizam imagens de sensoriamento remoto para analisar, identificar e quantificar mudanças na cobertura florestal. O objetivo deste trabalho foi avaliar a eficiência do classificador Iso Cluster para monitorar o desmatamento na Amazônia Legal. Para a condução deste estudo foram utilizadas as imagens de quatro cenas, localizadas no estado do Mato Grosso, obtidas pelo satélite Sentinel 2 e após realizou-se o processamento delas no software ArcGIS 10.3.1 Apesar da necessidade de algumas correções, o classificador Iso Cluster apresentou bons resultados.*

INTRODUÇÃO

A Amazônia Legal possui superfície aproximada de 5.015.067,75 km², correspondente a cerca de 58,9 % do território brasileiro [IBGE, 2020]. Segundo o Instituto Nacional de Pesquisas Espaciais [INPE] foram desmatados um total de 813.063,44 km^2 e 729.781,76 km^2 do Bioma Amazônico, até o ano de 2020. Conforme Florenzano [2008 *apud* AMORIN et al., 2021 p.89] a análise do meio físico por meio de técnicas de geoprocessamento e produtos de sensoriamento remoto permite monitorar e avaliar áreas de vegetação natural, com maior agilidade no processamento e atualização de dados e menor custo [VAEZA et al, 2010 *apud* AMORIN et al., 2021 p.89]. Os métodos classificadores de imagens podem ser divididos em dois grupos, os supervisionados, onde o analista define as classes de interesse e os não-supervisionados, onde somente alguns parâmetros, como o número mínimo e máximo de classes, são fornecidos pelo observador [QUEIROZ et al., 2017]. Segundo Faria (2015), o Iso Cluster baseia-se na proximidade do padrão do pixel de referência com as bandas de entrada para separá-los em classes. Os arquivos de entrada podem ser bandas da imagem, índices de vegetação ou qualquer outro arquivo raster que contenha informações relevantes para a classificação [FARIA, 2015]. De acordo com Rosa [2009 *apud* AMORIN et al., 2021 p.89] as classes são definidas por meio de algoritmo de classificação, este procedimento tem grande utilidade quando não há informações sobre a área imageada, ou pretende-se realizar classificação exploratória.

O objetivo do presente estudo foi avaliar estatisticamente, através de um coeficiente de eficiência, a metodologia de monitoramento do desmatamento de parte da Amazonia Legal, utilizando-se o classificador Iso Cluster e as imagens de Satélite do Sentinel 2, como produto de informação.

METODOLOGIA

A metodologia foi estabelecida em etapas distintas, porém interligadas. As imagens utilizadas para o presente projeto, foram geradas pelo sensor orbital Sentinel 2 e obtidas no site Copernicus Open Acess Hub, composta pelas seguintes cenas, 20LPP, 20KGQ, 20LPN e 20KRG, com imageamento em maio de 2021. Realizou-se a importação para o software ArcGIS versão 10.3.1, onde adicionou-se as bandas 02 (Blue), 03 (Green), 04 (Red); as imagens possuíam resolução espacial de 10 metros e o comprimento de onda variava de 490 a 665 nm. Inicialmente, elas estavam em tons de cinza, para facilitar a visualização e torná-las coloridas, realizou-se a composição de bandas. Neste procedimento atribuiu-se cores básicas, vermelho, verde e azul (Red, Green e Blue). E para uma melhor observação inverteu-se as ordens das bandas, ficando desta forma: Red 3, Green 2 e Blue 1.

Com a área de cada cena definida, o passo seguinte foi a classificação não-supervisionada através do classificador Iso Cluster. Para as cenas 20KRG e 20KQG foram necessárias 5 classes e para as cenas 20LPP e 20LPN foram 4 classes. Além disso, utilizou-se a rotina Raster to Polygon, através da tabela de atributos, e com GRIDCODE estabelecido, excluiu-se as classes que não eram de interesse (floresta e nuvens). Os pontos de desmatamento que o classificador não identificou, foram classificados manualmente. Realizou-se os últimos acabamentos, onde adicionou-se os

cinco elementos básicos de um mapa temático: Título, legenda, grade de coordenadas, escala e orientação. Os mapas temáticos foram exportados em formato TIFF.

Para determinar a eficiência da metodologia, gerou-se um quadro descritivo (Tabela1), realizou-se uma Análise de Correlação Direta da área total da amostra desmatada, que corresponde a 100 %, com a área não identificada de desmatamento, que consiste nas áreas que a metodologia utilizando-se do classificador não identificou desmatamento.

RESULTADOS E DISCUSSÃO

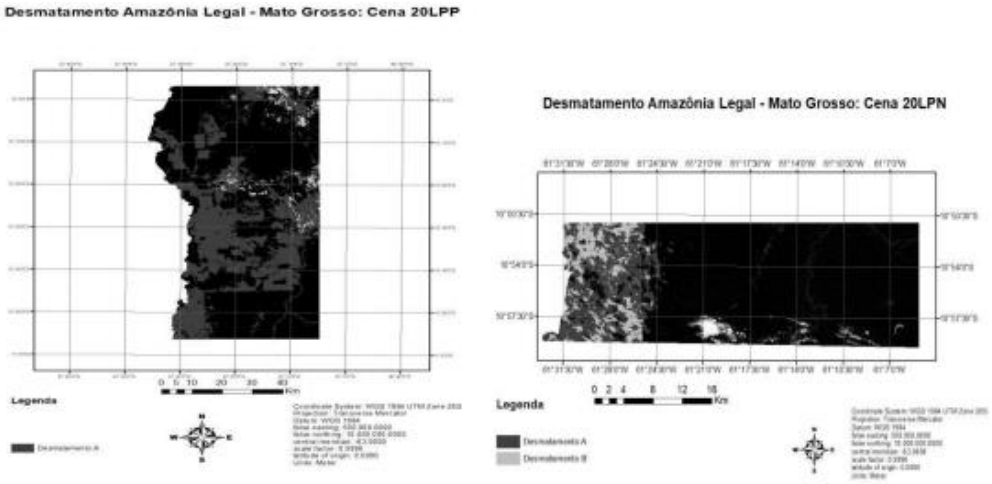


Figura 1 e 2: Mapa classificado da cena, Fonte: Autores (2021).

Tabela 1. Coeficiente de eficiência

Cena	Área total da cena (ha)	Área total desmatada da cena (ha)	Área amostral (ha)	Área desmatada na amostra (ha)	Área não identificada desmatada na amostra (ha)	Coef. eficiência (%)
20 KRG	29.318,19	16.502,27	29.318,19	16.502,27	305,63	98,18
20 KQG	439,83	364,61	439,83	364,61	8,21	97,79
20 LPN	74.218,07	13.205,16	74.218,07	13.205,16	125,51	99,06
20LPPA	525.944,99	102.005,73	100.000	15.457,61	2.599,98	85,60
20LPPB	525.944,99	102.005,73	100.000	37.762,53	2.123,76	94,67

Fonte: Autores (2021).

Em relação, a avaliação das áreas desmatamento das cenas, as imagens obtidas pelo Sentinel-2, mostraram-se eficientes. No estudo de, Mas et al. (2021) os dados obtidos pelo Sentinel-2 foram o que apresentaram melhor resolução temporal, aquisição mais sistemática, além de possuírem um maior número de observações válidas por mês, o que as capacita como ferramenta de análise para desmatamentos.

Quanto a metodologia utilizando-se o algoritmo do classificador não-supervisionado Iso Cluster, conforme demonstra a tabela 1, em todas as amostras o valor do coeficiente de eficiência foi acima de 85 %. A amostra A da cena 20 LPP obteve 85,60 %, um fator que contribuiu para esse índice menor, pode ter sido a presença de nuvens. Em seu estudo Faria (2015) considerou o algoritmo de classificação Iso Cluster como satisfatório, pois atingiu bons resultados. No trabalho desenvolvido por Nascimento et al (2016) também houve a necessidade de edição para a correção manual de classes temáticas que o Isso Cluster não distinguiu. Além disso, conforme Nascimento et al (2016) o Iso Cluster não apresentou uma boa acurácia.

CONSIDERAÇÕES FINAIS

Observa-se que o principal fator causador de erros no processo de classificação não supervisionada utilizando-se o algoritmo Iso Cluster foi a presença de nuvens. Nas imagens na qual eram esparsas ou inexistentes o classificador junto com a metodologia utilizada teve índices acima de 95 %, quase o total das áreas desmatadas foram identificadas. Tratando-se de uma metodologia que não apresenta custos de aquisição de imagens, comparados a outros métodos, que vem sendo empregados no monitoramento do desmatamento da Amazonia Legal, logo este torna-se extremamente operante e confiável.

REFERÊNCIAS BIBLIOGRÁFICAS

Amorim, J. V. A; Valladares, G. S.; Portela, M. G. T. Classificação não-supervisionada de imagens RAPIDEYE no mapeamento da cobertura das terras do Delta do Parnaíba, Piauí. **Geosaberes:** Revista de Estudos Geoeducacionais, Ceará, v. 12, n. 1, p. 88-104, fev./2021.

Faria, R. M. **Classificação temporal de imagens LANDSAT 8 para o monitoramento das mudanças do uso da terra.** 2015. 52 p. Monografia (Bacharel em Geografia) – Universidade Federal de Viçosa, Viçosa/MG, 2015..

Instituto Nacional De Pesquisas Especiais (INPE). **Monitoramento do desmatamento da Floresta Amazônica Brasileira por Satélite.** Disponível:http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes. Acesso em: 06 de agosto de 2021.

Mas, J. F.; et al. Análise da disponibilidade de imagens Landsat e Sentinel-2 para o Brasil. **Geografia Ensino & Pesquisa**: Santa Maria, v.24, e. 47, 2020.

Nascimento, I. S. do; et al. Avaliação da exatidão dos classificadores Maxver e Iso Cluster do Software ArcGIS for Desktop, com uso de imagem LANDSAT 8 do município de Cáceres/MT. **Revista Continentes:** (UFRRJ), ano 5, n. 8, p. 48-62, 2016.

Queiroz, T. B.; et al. Avaliação do desempenho da classificação do uso e cobertura da terra a partir de imagens LANDSAT 8 e RAPIDEYE na região central do Rio Grande do Sul. **Geociências**: São Paulo, UNESP, v. 36, n. 3, p. 569 - 578, 2017.

Classificação climática de thornthwaite (1948) para a região Centro-Oeste: uma abordagem inédita para a maior região produtora de grãos do Brasil

Rafael F. Lima[1], Lucas E. O. Aparecido[2], Guilherme B. Torsoni[3], Gabriel H. O. Souza[1]

[1]Departamento de Agronomia – Instituto Federal de Mato Grosso do Sul (IFMS) – Naviraí – MS – Brasil

[2]Departamento de Agronomia – Instituto Federal de Mato Grosso do Sul (IFMS) – Naviraí – MS – Brasil

[3]Departamento de Física – Instituto Federal de Mato Grosso do Sul (IFMS) – Naviraí – MS – Brasil

rafael.lima2@estudante.ifms.edu.br, lucas.aparecido@ifms.edu.br, guilherme.torsoni@ifms.edu.br, gabriel.souza4@estudante.ifms.edu.br

Abstract. *A simple way to capture climate variation over a location is through the use of climate classification systems. Thus, the objective of this work is to classify the climate of the Central-West region of Brazil using Thornthwaite's (1948) climate classification system. Climatic data from the NASA / POWER station were collected daily in the period 1990 – 2020, where potential evapotranspiration and the climatological water balance were calculated. The climatic classification was generated by the Thornthwaite system (1948). Twenty-four climate classes were obtained for the entire study region, with the most predominant classes being ArA'a' and B1rA'a' with 15.0% and 12.5% respectively.*

Resumo. *Uma forma simples de captar a variação do clima sobre um local é através do uso de sistemas de classificação climática. Dessa forma o objetivo desse trabalho é classificar o clima da região Centro-Oeste do Brasil utilizando o sistema de classificação climática de Thornthwaite (1948). Foram coletados dados climáticos da estação NASA / POWER de forma diária no período de 1990 – 2020, onde foram calculados a evapotranspiração potencial e o balanço hídrico climatológico. A classificação climática foi gerada pelo sistema de Thornthwaite (1948). Foram obtidas 24 classes*

climáticas para toda a região de estudo, sendo as classes mais predominantes a ArA'a' e B1rA'a' com 15,0% e 12,5% respectivamente.

Introdução

Buscando uma forma simplificada para analisar características climáticas de um determinado local, a classificação climática surge afim de sintetizar e agrupar elementos climáticos comuns em tipos climáticos, delimitando espacialmente regiões climáticas homogêneas [Tapiador et al. 2019]. Os sistemas de classificação climática são compostos por métodos estatísticos eficientes para capturar informações climáticas, gerando um arranjo eficiente de informações de uma forma simples e generalizada [Baker et al. 2010].

Na classificação climática elaborada por Thornthwaite, a planta é vista como um meio físico capaz de retirar umidade do solo e transferir para o sistema atmosférico através de processos fisiológicos, sendo esse sistema mais indicado para fins agrícolas [Elguindi et al. 2014]. A variabilidade do clima influencia diretamente na produção agrícola, afetando o desenvolvimento e a produtividade das lavouras.

No Brasil, a região Centro-Oeste se destaca na produção de diversas comoditys do agronegócio sendo elas: Soja, Milho, e algodão [Conab 2021]. Assim, buscou-se com o trabalho classificar o clima da região Centro-Oeste do Brasil utilizando o sistema de classificação climática de Thornthwaite (1948).

Material e métodos

Foram coletados dados climáticos de temperatura média do ar em graus celsius (°C) e precipitação em milímetros (mm) de forma diária para 467 municípios localizados na região Centro-Oeste do Brasil no período de 1990 – 2020, obtidos pela plataforma *National Aeronautics and Space Administration / Prediction of Worldwide Energy Resources* – NASA / POWER. Essa plataforma disponibiliza dados meteorológicos modelados por satélite com resolução espacial de 1° (latitude - longitude) [Stackhouse et al. 2015].

A evapotranspiração potencial (mm) foi calculada pelo método de Camargo (1971) e o balanço hídrico climatológico (BHC) foi gerado para todas as localidades estudadas seguindo a metodologia de Thornthwaite e Mather (1955), utilizando uma capacidade de água disponível no solo igual 100 mm. A classificação climática foi realizada pelo método proposto por Thornthwaite (1948), utilizando os atributos de deficiência hídrica (mm) e excedente hídrico (mm) obtidos do extrato do BHC associados a evapotranspiração potencial [Rolim e Aparecido 2016].

Com a utilização de um sistema de informações geográficas (SIG) foi realizado a espacialização dos atributos climáticos utilizando krigagem ordinária com um modelo esférico [Krige 1951], e através da sobreposição dos mapas foi obtido a delimitação espacial para as classes climáticas de Thornthwaite (1948).

Resultados e discussões

A região Centro-Oeste apresentou predomínio de 6 tipos climáticos (Figura 1) sendo eles A (Super úmido), B1 (úmido), B2 (úmido), B3 (úmido), B4 (úmido) e C2 (Sub-úmido). Foram encontradas 24 classes climáticas de Thornthwaite (1948), fator

determinado pela alta especificidade do sistema em captar variações mínimas nos atributos climáticos da região (Figura 1). Os tipos climáticos com maior destaque na região foram B2 (úmido) e B3 (úmido) com ocorrência de 6 classes climáticas em cada.

As classes ArA'a' e B1rA'a' (Figura 1) foram mais expressivas representando 15,0% e 12,5% da região de estudo respectivamente. A classe C2rA'a' (Figura 1) apresentou maior predomínio no bioma pantaneiro caracterizando a região como clima Sub-úmido com pouca deficiência hídrica, evapotranspiração acima 1140 mm anuais (megatérmico) sendo 48% concentrados na estação do verão. O bioma pantaneiro corresponde a uma extensa superfície alagada, considerada a maior planície alagada do mundo [Cardoso e Marcuzzo 2010]. As classes B1rA'a' e B2rA'a' (Figura 1) representam grande parte das regiões agrícolas do Mato Grosso do Sul.

Localidades ao norte do Estado do Mato Grosso apresentaram predomínio das classes ArA'a' e AwA'a' (Figura 1) caracterizando a região como clima superúmido com moderada deficiência hídrica no verão, evapotranspiração acima 1140 mm anuais (megatérmico) sendo 48% concentrados na estação do verão. Essas localidades apresentam altos índices de precipitação anual, pouca deficiência hídrica e temperaturas mais elevadas, principalmente no verão. As altas temperaturas no verão podem influenciar negativamente as lavouras de soja, provocando o abortamento de vagens causado pelo estresse térmico [Djanaguiraman et al. 2019]

No Distrito federal, a classe predominante foi a B2wB'3a' (Figura1). As regiões sul de Goiás e do Mato Grosso do Sul apresentaram desenvolvimento em comum da classe climática B2rB'4a' (Figura 1). Essas localidades apresentam temperatura do ar e índices pluviométricos semelhantes.

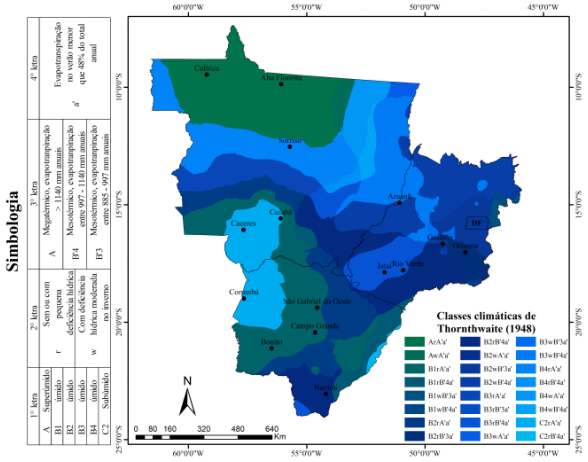


Figure 1. Classificação climática de Thornthwaite (1948) para a região Centro-Oeste do Brasil.

A combinação entre os elementos extraídos do extrato do BHC explorados na classificação climática de Thornthwaite (1948), permitiu explorar a grande variabilidade climática presente na região Centro-Oeste do Brasil.

Conclusões

A região Centro-Oeste apresentou 6 tipos climáticos principais e 24 classes climáticas. As classes climáticas de maior predomínio foram ArA'a' e B1rA'a' representando 15,0%

e 12,5% de toda região respectivamente. A classe C2rA'a' apresentou maior predomínio no bioma pantaneiro. Localidades ao norte do Estado do Mato Grosso apresentaram predomínio das classes ArA'a' e AwA'a'.

Referencias

Baker, B., Diaz, H., Hargrove, W., Hoffman, F. (2010). Use of the Köppen–Trewartha climate classification to evaluate climatic refugia in statistically derived ecoregions for the People's Republic of China. Climatic Change, p. 113-131.

Camargo A. (1971). São Paulo State Water Balance. Bol Inst Agronômico Camp, p. 1–24.

Cardoso, M. R. D. e Marcuzzo, F. F. N. (2010). Mapeamento de três decênios da precipitação pluviométrica total e sazonal do bioma Pantanal.

CONAB - Companhia Nacional de Abastecimento (2021). Acompanhamento safra brasileira de grãos, v. 8 - Safra 2020/2021, p. 1-108. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>.

Djanaguiraman, M., Schapaugh, W., Fritschi, F., Nguyen, H., Prasad, P. V. (2019). Reproductive success of soybean (Glycine max L. Merril) cultivars and exotic lines under high daytime temperature. Plant, cell & environment, p. 321-336.

Elguindi, N., Grundstein, A., Bernardes, S., Turuncoglu, U., Feddema, J. (2014). Assessment of CMIP5 global model simulations and climate change projections for the 21 st century using a modified Thornthwaite climate classification. Climatic change, p. 523-538.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy, p. 119-139.

Rolim, G. S., Aparecido, L. E. O. (2016). Camargo, Köppen and Thornthwaite climate classification systems in defining climatical regions of the state of São Paulo, Brazil. International Journal of Climatology, p. 636-643.

Stackhouse, P. W., Westberg, D., Hoell, J. M., Chandler, W. S., & Zhang, T. (2015). Prediction of Worldwide Energy Resource (POWER)-Agroclimatology methodology-(1.0 latitude by 1.0 longitude spatial resolution). Prediction of Worldwide Energy Resource (POWER)-Agroclimatology methodology-(1.0 latitude by 1.0 longitude spatial resolution).

Tapiador, F. J., Moreno, R., Navarro, A., Sánchez, J. L., García-Ortega, E. (2019). Climate classifications from regional and global climate models: Performances for present climate estimates and expected changes in the future at high spatial resolution. Atmospheric research, p. 107-121.

Thornthwaite C. e Mather J. (1955). The water balance publications in Climatology. DIT Lab Climatol Centerton NJ USA, p. 1-104.

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. Geographical review, p. 55-94.



Deep Features para sistemas de identificação de plantas

Luciano A. D. Filho[1], **Rodrigo T. Calumby**[1]

[1]Universidade Estadual de Feira de Santana – Feira de Santana – BA – Brazil

lucianoadfilho@ecomp.uefs.br, rtcalumby@uefs.br

Abstract. *The selection of a Deep Convolutional Neural Network (DCNN) architecture is a decisive task. In this paper we propose to evaluate the generalization power of Deep Features extracted from a broad set of DCNNs without fine-tuning for plant species recognition from images using a multi-organ dataset. The conventional Softmax classification approach and an alternative based on SVM and feature selection are compared. The experimental validation allowed to identify the most promising methods, with Softmax and SVM reaching 0.76 and 0.82 of Micro-F1, respectively.*

Resumo. *A seleção de uma arquitetura de Rede Neural Convolucional Profunda (RNCP) é uma tarefa decisiva. Neste trabalho avaliamos o poder de generalização de Deep Features extraídas à partir de um amplo conjunto de RNCPs sem fine-tuning para classificação de espécies de plantas em imagens, utilizando um conjunto de dados de imagens multi-órgão. São comparadas as abordagens de classificação convencial com Softmax e uma alternativa baseada em SVM e seleção de features. A validação experimental permitiu indentificar os métodos mais promissores, com o métodos baseados em Softmax e SVM alcançando 0.76 e 0.82 de Micro-F1, respectivamente.*

1. Introdução

A indústria agrícola ao redor do globo tem lidado cada vez mais com os desafios oriundos da necessidade de sustentabilidade à medida que a população mundial cresce. Com isso, a otimização dos resultados tem desempenhado um papel crucial, objetivando maximização da produção com menor uso de área possível. Nesse sentido, diversas tecnologias têm sido empregadas, não apenas no desenvolvimento de novos modelos de produção, mas buscando a melhoria da precisão em várias tarefas [Shi et al. 2015, Meng et al. 2015], bem como a automação de muitas delas, por exemplo, por meio de tecnologias de reconhecimento de padrões. Neste contexto, diversos trabalhos têm proposto a utilização de RNCPs para realização de tarefas relacionadas à identificação automática de doenças

em plantas [Khan et al. 2018, Brahimi et al. 2018], identificação de espécies invasivas [V. Sivakumar et al. 2020], , entre outros. De modo geral, esses estudos demonstram o desenvolvimento de ferramentas auxiliares nas tarefas de monitoramento e administração do campo, principalmente relacionadas à identificação automática de plantas. Isso possibilita também a implementação de sistemas que auxiliem na tomada de decisões, contribuindo para maior eficácia no controle de situações adversas que impactem a produção.

Neste contexto, a seleção de arquiteturas de RNCPs adequadas é uma tarefa decisiva [Maeda-Gutiérrez et al. 2020], ainda mais se tratando do domínio de plantas, com um vasto número de espécies e cenários de classificação com alta heterogeneidade visual nas imagens. Considerando a diversidade de aplicações de RNCPs relacionadas a plantas na agricultura, este trabalho tem por objetivo avaliar experimentalmente a eficácia de RNCPs para classificação sobre um conjunto de dados de propósito geral voltado para identificação multi-órgão de espécies de plantas. Para isso, realiza-se uma análise considerando a abordagem convencional de classificadores neurais com ativação *Softmax*, além da extração e seleção explícita de *Deep Features* e classificação com o algoritmo *Support Vector Machines* (SVM).

2. Trabalhos Correlatos

Diversos trabalhos propõem a utilização de RNCP para extração de *Deep Features* e identificação de plantas com classificadores tradicionais. Em [Khan et al. 2018], foi proposto o uso de *Deep Features* e classificadores tradicionais (*e.g., Decision Trees e SVM*) para segmentação e identificação de doenças em frutas. Similarmente, em [Khan et al. 2020], também é proposta a detecção e identificação de doenças em imagens de folhas de pepino. Foram utilizadas as arquiteturas VGG-19 e o VGG-M pré-treinadas para extração de *Deep Features*, além um método para seleção de características. Para identificar doenças em imagens de folhas de plantas, em [Li et al. 2020] avaliou-se o uso de classificadores (*Random Forests* e SVM) e *Deep Features*, com objetivo de compará-los com as RNCP. Além de reportar a superioridade dos métodos tradicionais, os autores demonstraram aspectos positivos da seleção de *features*, principalmente em relação à redução da complexidade computacional. Especificamente para a tarefa de identificação de espécies de plantas, em [Feitoza et al. 2019] os autores demonstraram resultados promissores com extração de *features* usando as redes InceptionV3, VGG-16 e VGG-19 e o classificador SVM. Contudo, não foram considerados procedimentos de otimização dos classificadores.

Diferentemente dos trabalhos descritos, este trabalho propõe avaliar um amplo conjunto de arquiteturas de RNCP, para a tarefa de reconhecimento de espécies de plantas sobre dados de maior complexidade em termos de heterogeneidade visual, número de amostras, classes e múltiplos órgãos. Além disso, estende-se significativamente o trabalho de [Feitoza et al. 2019] considerando o dobro das arquiteturas de RNCP e otimizações de hiperparâmetros, analisando-se a estabilidade dos modelos, além de compararmos os resultados obtidos com as arquiteturas treinadas à partir da abordagem de classificação convencional (classificador neural com ativação *Softmax*).

3. Metodologia

Esse trabalho utilizou o conjunto de dados do PlantCLEF 2013, com 26,077 imagens de 250 espécies de plantas, incluindo imagens de órgãos individuais *e.g.,* fruta, flor, ga-

a)

Protocolo	Parâmeteros	Configuração
10-Folds Estratificado	Batch size	16
		32
		64
	N. épocas	20
		40
		60
	Learning rate	1×10^{-4}
		5×10^{-4}
		1×10^{-3}
		5×10^{-3}

b)

Protocolo	Parâmetros	Configuração
10-Folds Estratificado	Cost	0.01
		0.1
		1
		10
	Gamma	0.0001
		0.001
		0.01
		"auto"

Figura 1. Configurações de otimização de hiperparametros *Grid Search.*

lhos, tronco, além da planta inteira. Considerando um cenário de classificação de mundo real, os experimentos foram realizados de maneira agnóstica em relação aos tipos de imagem. Na primeira etapa treinamos as redes VGG-16, VGG19, InceptionV3 (GoogLeNet), NasNET-A (NasNETLarge), ResNet50 e ResNet152_V2 com otimização de hiperparâmetros via *Grid Search* (Figura 1-a). As redes foram inicializadas com pesos pré-treinados sobre a base ImageNet, havendo otimização apenas no componente de classificação. Para isso, consideramos a função de custo *Categorical Crossentropy* e otimizador *Adam* com as configurações padrão propostas pelos autores. As redes pré-treinadas foram utilizadas para a extração de *Deep Features*, com posterior aplicação de PCA com 90% de variância junto com o classificador SVM com configurações de Kernel: *Linear, Sigmoid, Polynomial* e método de composição *One-vs-Rest*. Nesta etapa realizou-se também otimização de hiperparâmetros (Figura 1-b). Para avaliação da eficácia, as médias das medidas Micro-F1 foram calculadas para considerar o desbalanceamento de classes. Utilizou-se protocolo de validação cruzada com k *folds*, sendo, em cada rodada, 9 *folds* usados para treino e 1 para teste. Para análise de estabilidade e significância estatística, são apresentados intervalos de confiança de 95%.

4. Resultados e Discussão

A Figura 2 apresenta os resultados de Micro-F1 para as melhores configurações obtidas com a otimização de hiperparâmetros. Na Figura 2-c apresentamos um comparativo dos resultados entre Softmax e SVM. De modo geral, os resultados com SVM foram superiores aos com o classificador baseado em *Softmax*. Nos casos em que a média de Micro-F1 do Softmax foi maior, os resultados podem ser considerados equivalentes quando analisados os intervalos de confiança. Além disso, o tamanho dos intervalos de confiança para as redes treinadas com o *Softmax* foram muito maiores, indicando menor estabilidade dos métodos para diferentes amostras de treino e teste.

a)

Arquitetura	Batch size	N. epochs	Learning rate	Micro-F1
ResNet50	64	60	5×10^{-3}	$(\mathbf{0.7683} \pm 0.0223)$
NasNetLarge	32	60	1×10^{-3}	(0.7172 ± 0.0403)
ResNet152_V2	64	40	5×10^{-4}	(0.6698 ± 0.0467)
VGG19	64	60	5×10^{-3}	(0.6554 ± 0.0478)
VGG16	16	60	5×10^{-3}	(0.6438 ± 0.0059)
InceptionV3	64	60	1×10^{-3}	(0.6127 ± 0.0212)

b)

Arquitetura	Kernel	Cost	Gamma	Micro-F1
ResNet50	Polynomial	0.01	0.01	$(\mathbf{0.8180} \pm 0.0044)$
VGG16	Polynomial	0.01	0.01	(0.7025 ± 0.0047)
VGG19	Polynomial	0.01	0.01	(0.6951 ± 0.0045)
InceptionV3	Linear	1	*auto*	(0.6905 ± 0.0051)
NasNetLarge	Linear	1	*auto*	(0.6881 ± 0.0059)
ResNet152_V2	Polynomial	1	*auto*	(0.6438 ± 0.0059)

c)

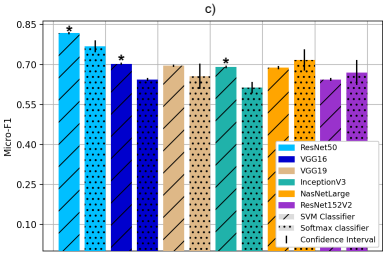


Figura 2. Resultados dos modelos otimizados: a) Softmax; b) SVM; c) Comparativo de Softmax e SVM com *deep features* e intervalos de confiança. (*) Significância em relação ao Softmax.

5. Conclusões

Os resultados possibilitaram identificar as arquiteturas de RNCP mais eficazes para extração de *features* de plantas, com destaque para a ResNet50, que com a utilização de PCA e do SVM alcançou Micro-F1=0.82. Além disso, os resultados evidenciaram que apesar do tipo de classificador ter desempenhado um papel significativo sobre a eficácia e precisão dos modelos, podemos observar que estatisticamente não houve diferença significativa na ordem de eficácia das arquiteturas para os diferentes classificadores. Isso sugere a consistência dos resultados em relação à eficácia das redes avaliadas. Por fim, estes resultados contribuem diretamente para o desenvolvimento de sistemas de identificação automática de plantas e tarefas relacionadas à aplicação de RNCP para classificação de órgãos de plantas e as inúmeras aplicações no contexto da agricultura e biodiversidade.

Referências

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., e Moussaoui, A. (2018). *Deep Learning for Plant Diseases: Detection and Saliency Map Visualisation*, pages 93–117. Springer International Publishing, Cham.

Feitoza, M. C., da Silva, W. B., e Calumby, R. T. (2019). Exploring deep features and transfer learning for plant species recognition. In *Proceedings of the XV Brazilian Symposium on Information Systems*, pages 1–8.

Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., e Saba, T. (2018). Ccdf: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep cnn features. *Computers and electronics in agriculture*, 155:220–236.

Khan, M. A., Akram, T., Sharif, M., Javed, K., Raza, M., e Saba, T. (2020). An automated system for cucumber leaf diseased spot detection and classification using improved saliency method and deep features selection. *Multimedia Tools and Applications*, pages 1–30.

Li, Y., Nie, J., e Chao, X. (2020). Do we really need deep cnn for plant diseases identification? *Computers and Electronics in Agriculture*, 178:105803.

Maeda-Gutiérrez, V., Galván Tejada, C., Zanella Calzada, L., Celaya Padilla, J., Galván Tejada, J., Gamboa-Rosales, H., Luna-Garcia, H., Magallanes-Quintanar, R., Carlos, G.-M., e Olvera-Olvera, C. (2020). Comparison of convolutional neural network architectures for classification of tomato plant diseases. *Applied Sciences*, 10:1245.

Meng, Q., Qiu, R., He, J., Zhang, M., Ma, X., e Liu, G. (2015). Development of agricultural implement system based on machine vision and fuzzy control. *Computers and Electronics in Agriculture*, 112:128–138. Precision Agriculture.

Shi, Y., Wang, N., Taylor, R., e Raun, W. (2015). Improvement of a ground-lidar-based corn plant population and spacing measurement system. *Computers and Electronics in Agriculture*, 112:92–101. Precision Agriculture.

V. Sivakumar, A. N., Li, J., Scott, S., Psota, E., Jhala, A. J., Luck, J. D., e Shi, Y. (2020). Comparison of object detection and patch-based classification deep learning models on mid- to late-season weed detection in uav imagery. *Remote Sensing*, 12(13).

Estimativa da Cercosporiose do café usando clima em Varginha, Minas Gerais

Camila D. Cabral[1], Rafael F. de Lima[2], Lucas Eduardo de O. Aparecido[1], Fábio dos S. Corsini[3], Geraldo G. de O. Júnior[4]

[1]Departamento Agronomia – Instituto Federal do Sul de Minas Gerais campus Muzambinho (IFSULDEMINAS) – Muzambinho – MG – Brasil
camilacabralagro@gmail.com, lucas.aparecido@muz.ifsuldeminas.edu.br

[2]Departamento Agronomia – Instituto Federal do Mato Grosso do Sul campus Naviraí (IFMS) – Naviraí – MS – Brasil
rafael.lima2@estudante.ifms.edu.br

[3]Departamento Computação – Instituto Federal do Sul de Minas Gerais campus Machado (IFSULDEMINAS) – Machado – MG – Brasil
fabio.corsini@muz.ifsuldeminas.edu.br

[4]Departamento Segurança do Trabalho – Instituto Federal do Sul de Minas Gerais campus Muzambinho (IFSULDEMINAS) – Muzambinho – MG – Brasil
geraldo.junior@muz.ifsuldeminas.edu.br

A cafeicultura é uma das principais atividades do agronegócio brasileiro, sendo cada vez mais acometida ao ataque de doenças. O objetivo deste trabalho foi estimar a incidência de Cercosporiose no município de Varginha em função de elementos climáticos. Foram coletados dados de incidência de Cercosporiose em cafeeiros de Boletins Fitossanitários da Procafé e informações climatológicas da plataforma NASA/POWER. Os dados foram submetidos a uma Análise de Correlação Múltipla. O modelo criado para estimar a incidência da doença foram submetidos a avaliações. As variáveis climáticas que demonstraram maior influência na previsão da Cercosporiose foram pressão de saturação de vapor e velocidade média do vento.

1. Introdução

O Brasil é o maior produtor e exportador de café e segundo maior consumidor da bebida no mundo (MAPA, 2018). Minas Gerais é o maior estado produtor do país, com destaque para o Sul de Minas com 19,15 milhões de sacas (CONAB, 2020).

A Cercosporiose do café é causada pelo fungo *Cercospora coffeicola* Berk & Cook. e atinge todos os estágios de desenvolvimento do cafeeiro (MESQUITA et al., 2016). As perdas na produção devido a esta doença podem chegar a 20%. O fungo ataca

folhas e frutos do cafeeiro. Nas folhas, ele penetra por aberturas naturais na face superior ou cutícula (SILVA; GIUNTI, 2014).

Em vista da importância da cafeicultura e dos danos potenciais que esta doença pode causar ao cafeeiro e a sua produção, objetiva-se com o presente trabalho estimar a incidência de Cercosporiose no município de Varginha, Sul de Minas Gerais, em função de variáveis climáticas.

2. Material e Métodos

Para o desenvolvimento deste trabalho, foram coletados dados de incidência de Cercosporiose em Boletins de Avisos Fitossanitários da Procafé para Varginha (Latitude 21° 34’ 00’’S; Longitude 45° 24’ 22’’W; Altitude: 940m), Minas Gerais, de Setembro de 2007 a Março de 2021. Os valores da doença correspondem as cultivares de cafeeiros *Mundo Novo* e *Catuaí*, coletados em lavouras com carga alta e baixa, e estão em porcentagem.

Os dados climáticos de armazenamento (ARM), déficit hídrico (DEF), duração do período de molhamento foliar (DPM), evapotranspiração potencial (ETP), evapotranspiração real (ETR), excedente hídrico (EXC), precipitação pluviométrica (P), radiação solar incidente (Qg), radiação solar no topo da atmosfera (Qo), saldo de radiação (Rn), temperatura do ponto de orvalho (TDEW), temperatura máxima (TMAX), temperatura média (TMED), temperatura mínima (TMIN), umidade relativa do ar (UR), velocidade máxima do vento (VMAX), velocidade média do vento (VMED), velocidade miníma do vento (VMIN), pressão atual de vapor (ea) e pressão de saturação de vapor (es) foram coletados para esta mesma localidade na plataforma National Aeronautics and Space Administration/Prediction of WorldWide Energy Resources – NASA/POWER.

Para estimar a incidência de Cercosporiose em função dos elementos climáticos, realizou-se *Regressão Linear Múltipla*. Os valores de Cercosporiose são as variáveis dependentes e os elementos climáticos as independentes. Devido ao grande número de variáveis climáticas, foram selecionadas somente os elementos climáticos que haviam maior correlação com a incidência da doença, através da análise de *Correlação de Pearson*. Todas as análises foram realizadas com auxílio do software Microsoft® Office Excel®.

A análise de correlação é um método estatístico que estuda o grau de relacionamento entre variáveis. O Coeficiente de correlação de Pearson mede a direção e grau com que duas variáveis, de tipo quantitativo, se associam linearmente, tendo seus valores variando entre -1 a +1. O sinal indica direção (perfeita correlação negativa ou inversa e positiva ou direta, respectivamente) e o valor a força entre as variáveis. Assim, quando a correlação é positiva, ambos os valores das variáveis aumentam (diretamente proporcionais), e negativa, á medida que uma variável cresce as outras decrescem (indiretamente proporcionais). Escores em proximidade de +1 ou -1 apresentam maior grau de dependência estatística linear entre as variáveis e próximos de zero menor força de relação (MARTINS, 2014).

Segundo Chein (2019), na análise estatística de regressão, o objetivo é sempre verificar a existência de uma relação funcional (dependência estatística) entre variáveis. Na Regressão Linear Múltipla tem-se uma variável dependente (Y) que será explicada por duas ou mais variáveis independentes. É uma extensão da análise linear simples.

Esta análise visa estabelecer uma equação que estime valores de Y para valores dados das variáveis independentes.

De acordo com Minitab (2019), depois de ajustar um modelo linear usando análise de regressão, precisa-se determinar se o modelo foi bem ajustado aos dados (conjunto de observações). Para a avaliação do modelo, entre incidência observada de Cercosporiose e os estimados, foi utilizado o r2 (coeficiente de determinação) para precisão.

3. Resultado e Discussões

Na análise de Correlação de Pearson, as variáveis que se mostraram mais significativas, foram ETP, ETR, P, Qo, TMAX, TMED, TMIN, VMAX, VMED e es. E por esse motivo foram as variáveis independentes do modelo ajustado.

Esses elementos climáticos foram submetidos à Análise de Regressão Linear Múltipla, associados à incidência de Cercosporiose em Varginha/MG. Da estatística de Regressão obtiveram-se os coeficientes (Tabela 2) deles, o R múltiplo, R-quadrado (r2), R-Quadrado ajustado, e F de significação da mesma (Tabela 1).

Tabela 1. Parâmetros de avaliação do modelo de Regressão Linear Múltipla.

Estatística de regressão	
R múltiplo	0,358862546
R-Quadrado	0,128782327
R-quadrado ajustado	0,101124623
F de significação	3,33225E-06

Tabela 2. Coeficientes calibrados para previsão da Cercosporiose do cafeeiro.

Variáveis Climáticas	Interseção	ETP	ETR	P	Qo	TMAX	TMED	TMIN	VMAX	VMED	es
Coeficientes	23,40	0,04	-0,05	0,00	0,16	-1,57	0,86	-0,88	3,44	-7,51	5,54

O F de significância da regressão (<0,05) indicou que ela foi significativa, uma vez que ele determina se o relacionamento entre o modelo e a variável resposta é estatisticamente significativo (MINITAB, 2019). Assim, mesmo com o valor de r2 mediano, o modelo de previsão de Cercosporiose em função do clima é considerado satisfatório.

Ao analisar o modelo que contém as variáveis preditoras e seus respectivos coeficientes, verifica-se que a variável mais significativa positivamente é o es (5,54) e negativamente o VMED (-7,51). Segundo Minitab (2019), os coeficientes de regressão descrevem o relacionamento entre cada variável preditora (explicativa ou independente) e a resposta (dependente). Assim, os seus valores representam a mudança média na resposta, dado o aumento de uma unidade no preditor, sendo multiplicados pelos valores da dependente de modo a obter os valores estimados da mesma.

A variação dos dados de Cercosporiose do café observados e estimados pode ser observada na Figura 1.

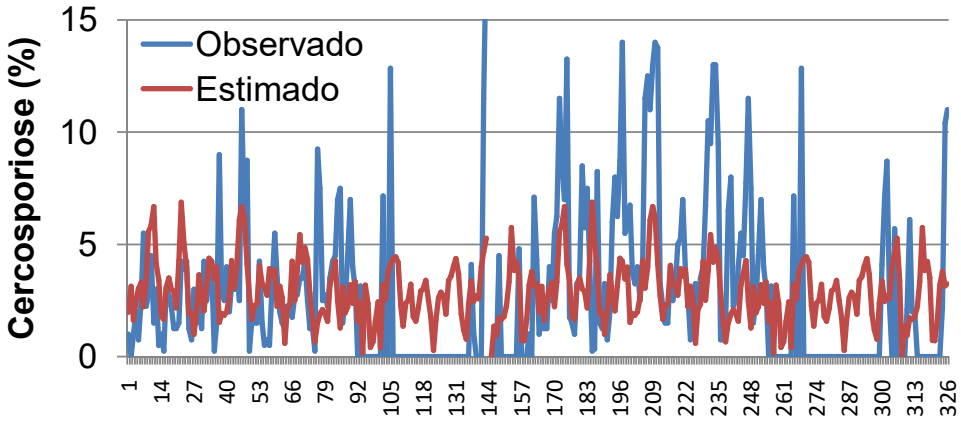


Figura 1. Variação dos dados de Cercosporiose do café observados e estimados na localidade de Varginha, Minas Gerais.

4. Conclusões

É possível estimar a incidência de Cercosporiose do cafeeiro, em Varginha/MG, usando de variáveis climáticas. As que demonstraram maior influencia na previsão da doença foram pressão de saturação de vapor (es) e velocidade média do vento (VMED). O modelo de previsão da doença em função do clima é considerado satisfatório devido ao baixo valor do F de significação e r2 mediano.

5. Agradecimentos

Agradecemos ao IFSULDEMINAS pela concessão de bolsa ao primeiro autor (Edital nº. 171/2020).

6. Referências Bibliográficas

CHEIN, F. **Introdução aos modelos de Regressão Linear**. Brasília: Enap, 2019.

CONAB. **Acompanhamento da Safra Brasileira de Café - Safra 2020**, Quarto Levantamento. Brasília, DF: Conab, 2020.

MARTINS, M. E. G. Coeficiente de correlação amostral. **Ciência Elementar**, Campo Grande, Lisboa, v. 2, n. 2, p. 1-2, 2014.

MAPA. **Café no Brasil**. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/politica-agricola/cafe/cafeicultura-brasileira>. Acesso em: 29 abr. 2021.

MESQUITA, C. M. de. et al.. **Manual do Café:** Distúrbios fisiológicos, pragas e doenças do cafeeiro. Belo Horizonte: EMATER-MG, 2016.

MINITAB. **Análise de regressão: Como interpretar o R-quadrado e avaliar a qualidade de ajuste?**. Disponível em: <https://blog.minitab.com/pt/analise-de-regressao-como-interpretar-o-r-quadrado-e-avaliar-a-qualidade-de-ajuste>. Acesso em: 05 jul. 2021.

SILVA, A. V.; GIUNTI, O. D. **Cafeicultura:** manejo de pragas, doenças e plantas invasoras do cafeeiro. Muzambinho, MG: Rede e-Tec Brasil, 2014.

Impacto das mudanças climáticas no período 2040-2080 sobre a aptidão do cultivo de *Coffea arabica* no Brasil

João Antonio Lorençone[1], Lucas E. de O. Aparecido[2], Pedro Antonio Lorençone[1], Guilherme Botega Torsoni[1]

[1]Instituto Federal de Mato Grosso do Sul (IFMS)
Naviraí – MS – Brazil

[2]Instituto Federal do Sul de Minas (IFSULDEMINAS)
Muzambinho – MG - Brazil

joao.lorencone@estudante.ifms.edu.br,
lucas.aparecido@muz.ifsuldeminas.edu.br,
pedro.lorencone@estudante.ifms.edu.br,
guilherme.torsoni.ifms.edu.br

Abstract. *The objective of this work was to evaluate the impact of climate change on areas suitable for Arabica coffee cultivation under different climate change scenarios. The work was carried out for all of Brazil, using data from 1960-2020. The BCC - CSM 1 climate model was used to obtain future climate data for all RCP 2.6, 4.5, 6.0, 8.5 scenarios. All future scenarios of climate change showed a reduction in the total areas suitable for coffee cultivation in Brazil, with an average reduction of 50%. Furthermore, areas restricted by excessive heat and water deficiency were the most common throughout the country in the future scenarios, with an average of 63% of the entire territory.*

Resumo. *O objetivo desse trabalho foi avaliar o impacto das mudanças climáticas sobre as áreas aptas para o cultivo do café arábica nos diferentes cenários de mudanças climáticas. O trabalho foi realizado para todo o Brasil, sendo utilizado dados no período de 1960-2020. Ainda, para obtenção dos dados climáticos futuros foi utilizado o modelo climático o BCC – CSM 1, de todos os cenários RCP 2.6, 4.5, 6.0, 8.5. Todos os cenários futuros de mudanças climáticas demonstraram redução no total de áreas aptas ao cultivo do cafeeiro no Brasil, com redução média de 50%. Ainda, as áreas com restrição por excesso térmico e deficiência hídrica foram as mais comuns por todo o país nos cenários futuros, com média de 63% de todo território.*

1. Introdução

O cultivo de café tem papel fundamental na geração de empregos e renda, estando muito presente na agricultura familiar (Muñoz-Rios et al. 2020). É uma das commodities mais

comercializadas no mundo, sendo o Brasil o maior produtor e exportador do grão (WTO 2020).

As mudanças climáticas se tornaram um grande problema no século 21. Estudos comprovam que a produção de alimento será impactada negativamente, ao mesmo tempo em que a população continua com seu crescimento. No último século, a temperatura global aumentou 1 °C (± 0,2), as previsões para 2100 são de aumentos entre 1,5 °C e 6 °C (IPCC, 2018). Essas condições podem representar um risco para a vegetação devido ao aumento da evaporação, períodos mais longos de seca e redução disponibilidade de água no solo para as plantas.

Dessa forma, o objetivo desse estudo é avaliar o impacto das mudanças climáticas sobre as áreas aptas para o cultivo do café arábica nos diferentes cenários de mudanças climáticas.

2. Material e Métodos

Foram coletados os elementos climáticos Temperatura do ar (°C) e Precipitação (mm), de 4.947 estações meteorológicas do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) do Instituto Nacional de Meteorologia do Brasil - INMET (Brasil, 1992), no período de 1960-2020. Os dados provenientes do IPCC são resultados de simulações de modelos globais de alguns centros de pesquisa que contribuem para a confecção do relatório do IPCC-AR5, forçadas pelas concentrações observadas de gases de efeito estufa durante o século XX e forçadas por uma concentração estimada para o século XXI (IPCC, 2013). O modelo climático global utilizado foi o BCC – CSM 1.1 desenvolvido no Centro Climático de Pequim (BCC), possui resolução de 125 x 125 km, com 26 níveis verticais, seus componentes são: atmosfera, superfície de terra, oceano, gelo marinho, ciclo do carbono terrestre, ciclos biogeoquímicos do oceano (Flato et al., 2014). Os dados climáticos dos cenários RCPs foram coletados através da plataforma WorldClim 2.

Para poder identificar as áreas aptas, restritas ou inaptas ao cultivo do café foram utilizadas as faixas de aptidão climáticas ideais de temperatura do ar (°C), temperatura do ar do mês de novembro (°C), temperatura do ar do mês mais frio do ano (°C) e déficit hídrico para o C.arabica, conforme os seguintes autores Camargo (1977), Matiello (1991) e Camargo and Pereira (1994). Um fluxograma explicativo das etapas de elaboração do trabalho pode ser acessado a partir do link:

https://photos.app.goo.gl/sgXYvLXeJP75vJ4g6

Os softwares utilizados no trabalho foram o Quantum GIS (QGis) 3.16.4 para interpolação dos dados e elaboração do mapeamento, Python 3.10.0 e o Excel Versão 2109 para elaboração das análises estatísticas.

3. Resultados e Discussão

O zoneamento agroclimático do Coffea arabica nos cenários de mudanças climáticas, apresentou um aumento significativo das áreas inaptas, com grande redução das áreas aptas ao cultivo do cafeeiro (Figura 1 E). Grande parte do estado de Minas Gerais, apresentou redução das áreas aptas em todos os cenários, da mesma maneira para o estado de São Paulo. Por outro lado parte da região sul do país, passaram de inaptas por insuficiência térmica para aptas ao cultivo (Figura 1). Essas alterações de aptidões

climáticas no Brasil podem resultar em redução da produção de café, tendo em vista que as regiões com maiores altitudes no país podem apresentar restrições climáticas e grande parte das possíveis novas regiões aptas ao cultivo não possuem uma altitude que estimule o melhor desenvolvimento do cafeeiro (Carvalho et al., 2019).

Grande parte do território brasileiro demostrou mudar de classe nos cenários possíveis de mudanças climáticas em relação ao atual, tanto nos cenários de 2041 – 2060 quanto nos de 2061 – 2080 (Figura 1). Regiões como o oeste de Mato Grosso e o estado do Mato Grosso do Sul passaram das classes com restrição por excesso térmico para restrição térmica e hídrica, demostrando que o possível aumento do déficit hídrico resultará em grandes impactos às lavouras de café, tendo em vista que o café é muito suscetível ao estresse hídrico. A utilização de irrigação complementar para os cafeeiros pode ser uma alternativa para contornar as possíveis cenários e mudanças climáticas, proporcionando um impacto menor na produção final das lavouras do país (Moreira et al. 2021).

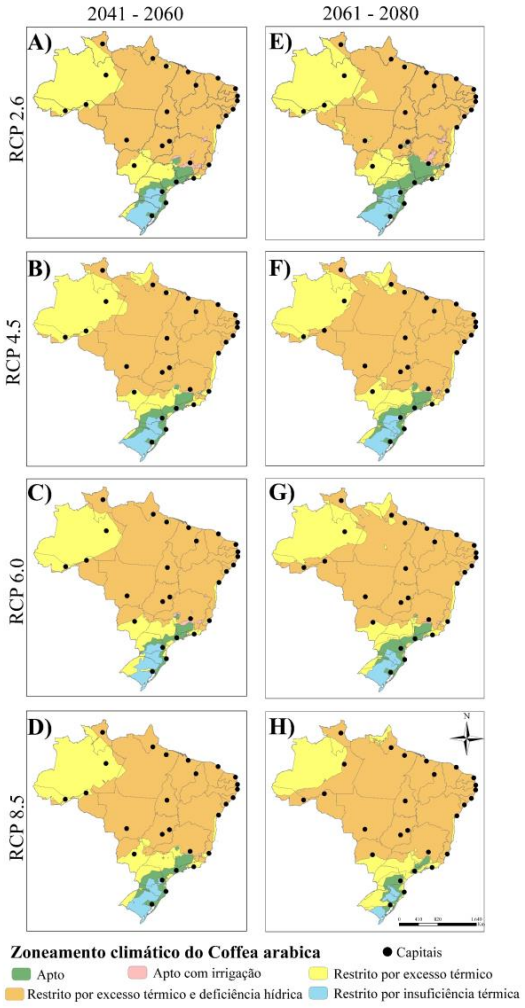


Figure 1. Zoneamento climático do Coffea arábica para o Brasil nos diferentes cenários de mudanças climática.

4. Conclusões

As mudanças climáticas futuras podem impactar negativamente a cultura do cafeeiro em todos os cenários RCPs estudados. Todos os cenários do modelo BCC – CSM 1.1 apresentam redução no total de áreas aptas ao cultivo do cafeeiro no Brasil, com redução média de 50%. Ainda, as áreas com restrição por excesso térmico e deficiência hídrica são as mais comuns por todo o país nos cenários futuros, com média de 62,95% de todo território.

References

Camargo, A. de. (1977). Zoneamento de aptidão climática para a cafeicultura de arábica e robusta no Brasil. *Fundação IBGE, Recursos, meio ambiente e poluição*, 68–76.

Camargo, A. de, & Pereira, A. (1994). Agrometeorology of the coffee crop. *Geneva: World Meteorological Organization*, 92.

Carvalho, C. F., Carvalho, S. M., & Souza, B. (2019). Coffee. In B. Souza, L. L. Vázquez, & R. C. Marucci (Orgs.), *Natural Enemies of Insect Pests in Neotropical Agroecosystems* (p. 277–291). Springer International Publishing. https://doi.org/10.1007/978-3-030-24733-1_23

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., & Eyring, V. (2014). Evaluation of climate models. In *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 741–866). Cambridge University Press.

IPCC. (2018). *Global Warming of 1.5°C, Summary for Policymakers. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE.

Matiello, J. B. (1991). *O café: Do cultivo ao consumo*. Editora Globo São Paulo.

Moreira, T. R., da Silva, S. F., da Silva, N. B., dos Santos, G. M. A. D. A., & dos Santos, A. R. (2021). Global Warming and the Effects of Climate Change on Coffee Production. In L. Louzada Pereira & T. Rizzo Moreira (Orgs.), *Quality Determinants In Coffee Production* (p. 65–100). Springer International Publishing. https://doi.org/10.1007/978-3-030-54437-9_2

Muñoz-Rios, L. A., Vargas-Villegas, J., & Suarez, A. (2020). Local perceptions about rural abandonment drivers in the Colombian coffee region: Insights from the city of Manizales. *Land Use Policy*, *91*, 104361. https://doi.org/10.1016/j.landusepol.2019.104361

WTO, W. T. O. (2020). *Statistics on merchandise trade*. WORLD TRADE ORGANIZATION. timeseries.wto.org/



PDAgro: Uma proposta de protocolo para *compliance* à LGPD

Samara B. Marques, Aline P. A. Lisboa,
Érico M. H. do Amaral [1], Vinícius do N. Lampert[2]

[1] Universidade Federal do Pampa (UNIPAMPA)
Bagé – RS – Brasil

[2]Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA – CPPSUL)
Bagé – RS – Brasil

{samaramarques.aluno,alinelisboa.aluno,ericoamaral}

@unipampa.edu.br, vinicius.lampert@embrapa.br

Abstract. *This paper addresses the compliance process of the General Data Protection Law in small and medium-sized agribusiness companies that processes personal data. Through studies on compliance with the law in agribusiness, there was a lack of works dedicated to this area. Based on this reality, this study proposes a protocol to help companies with compliance, facilitating this process.*

Resumo. *Este artigo aborda o processo de conformidade com a Lei Geral de Proteção de dados (LGPD) em empresas de pequeno e médio porte do agronegócio que manipulam dados pessoais. Através de estudos acerca da adequação com a lei no agronegócio, verificou-se falta de trabalhos dedicados a esta área. Com base nesta realidade o presente estudo propõe um protocolo para auxiliar empresas com a compliance, facilitando o processo de adequação.*

1. Introdução

A Lei Geral de Proteção de Dados (LGPD), vigente desde 14 de agosto de 2018, contempla direitos como titularidade de dados pessoais, privacidade de dados, protege direitos fundamentais e estabelece penalidades vigentes desde 1º de agosto deste ano [Brasil 2018].

A *compliance* com a LGPD deve ocorrer em todos os setores da economia, inclusive no agronegócio, onde a agropecuária obteve um crescimento de 5,7% no primeiro trimestre, sendo o setor com maior resultado [IBGE 2021]. A aplicabilidade da LGPD se destina à todas empresas e cidadãos brasileiros que tratam dados pessoais.

Além da LGPD, o *General Data Protection Regulation* (GDPR) é um regulamento europeu com regras de proteção das pessoas naturais quanto ao tratamento de dados pessoais e direito à proteção dos dados. Está implementada em maior grau de adequação, onde as empresas do agronegócio podem usar selo de transparência, código de conduta e ferramentas de diagnóstico de conformidade [EU 2016]. Portanto, é possível encontrar estudos referentes ao tema, mas nenhum trabalho sobre a LGPD no agronegócio.

Diante do exposto, este artigo trata sobre o processo de conformidade das empresas Agro de pequeno e médio porte por meio de uma proposta de protocolo que servirá como guia para entender o que a lei exige e como a adequação pode ser feita. Neste sentido, o presente artigo está organizado na forma que segue, na metodologia é apresentado o método científico, o referencial descreve a visão dos autores com o embasamento teórico e por fim, é apresentada uma proposta de protocolo para *compliance* à LGPD.

2. Metodologia

Para o desenvolvimento deste estudo adotou-se uma metodologia de pesquisa, conforme Prodanov e de Freitas (2013), baseada em autores que descrevem os procedimentos legais e de segurança referentes a privacidade e proteção de dados. Quanto à sua natureza esta pesquisa é classificada como pesquisa aplicada, pois "objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos". Quanto aos procedimentos técnicos é classificada como bibliográfica e experimental, objetivando propor uma solução com base na teoria.

3. Referencial

Nessa seção será apresentado o embasamento teórico para o presente trabalho. Inicialmente, será abordado o contexto de privacidade de dados e tecnologia, na sequência são trazidos pontos sobre a lei europeia GDPR, a lei brasileira LGPD e sobre a norma ISO 27701 relacionadas à proteção e privacidade de dados. Por fim, serão apresentados alguns trabalhos correlatos de Lima (2020), Gomes (2019), Rolim e Gibran (2021) e Silva et al. (2021), pois trazem discussões relevantes que contribuem com esta pesquisa.

O desenvolvimento de novas tecnologias da informação, durante o século XX, fizeram com que a privacidade se tornasse um problema da linha de frente em todo o mundo, de acordo com Solove (2008), que propõe uma taxonomia para compreensão da privacidade de forma contextual, dividida em 4 principais grupos de atividades: coleta de informações, processamento de informações, disseminação da informação e invasão.

Ainda em relação à privacidade, Brito e Machado (2017) reforçam a sua necessidade, independente do tempo, do meio inserido e da importância dada pelos indivíduos. Com a facilidade da transmissão da informação, a vulnerabilidade à diversos tipos de ataques é uma realidade quando trata-se de dispositivos conectados à rede. Nesse sentido, existe um conjunto de leis internacionais e nacionais que visam a privacidade de dados.

Na Europa, a GDPR aplica-se ao tratamento de dados pessoais envolvendo um controlador ou um processador, situado no território da União Europeia, independentemente de o tratamento em si ocorrer dentro ou fora da União [EU 2016]. No Brasil, o artigo 3° da LGPD informa sobre a questão territorial, em que o tratamento ou coleta dos dados seja realizado no território nacional, independentemente do país de sua sede ou

localização dos dados. O artigo 6º apresenta os 10 princípios que deverão ser observados no tratamento de dados pessoais [Brasil 2018]. Desta forma, uma norma de privacidade é abordada para implementar, manter e melhorar o processo de conformidade.

Diante dos impactos da LGPD nas empresas, a ISO/IEC 27701 passa a ter importância no processo de adequação, pois especifica os requisitos e fornece as diretrizes para implementação, manutenção e melhoria contínua de um Sistema de Gestão de Privacidade da Informação (SGPI) na forma de uma extensão das NBR ISO/IEC 27001 e NBR ISO/IEC 27002 para a gestão da privacidade. Destacam-se o Anexo D, o qual traz uma tabela com o mapeamento entre as subseções da ISO/IEC 27701 e os artigos da GDPR, e principalmente o Anexo N/A que relaciona a estrutura da ISO/IEC 27701 com os artigos da LGPD [ABNT 2019]. Neste contexto, os esforços são concentrados para evitar incidentes com dados pessoais manipulados pelas organizações.

Mesmo com a LGPD em vigor, ocorreram incidentes que deixaram expostos os dados pessoais de milhares de brasileiros [Lima 2020]. Os prejuízos causados à credibilidade das empresas, segundo Gomes (2019), podem superar as multas impostas e obrigação de adequação. A proteção dos dados demonstra uma gestão transparente como um diferencial no mercado [Rolim e Gibran 2021]. Para verificar a conformidade ferramentas de tecnologia são utilizadas para questionar a empresa sobre seus processos, se existe tratamento de dados e ao final apresentam o grau de adequação [Silva et al. 2021].

4. Proposta de Protocolo para Compliance à LGPD

Os estudos realizados nesta pesquisa sobre LGPD, assim como a identificação das boas práticas descritas na GDPR e orientadas pela ISO 27701, culminaram em uma proposta de protocolo, representada na Figura 1, destinada ao alinhamento de empresas de pequeno e médio porte da área Agro com a nova Lei Geral de Proteção de Dados.

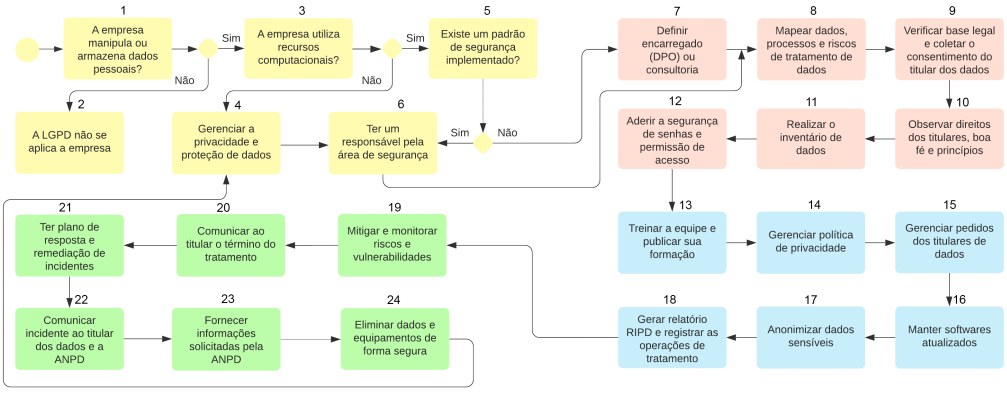


Figura 1. Protocolo - O fluxo inicia questionando o que existe na empresa, em amarelo, na sequência sugere a execução de práticas, em rosa, a seguir o acompanhamento das ações, na cor azul, por fim, ações de padronização, na cor verde, e o retorno do ciclo para constante aperfeiçoamento

O foco da presente proposta é possibilitar a identificação, de forma simples, dos principais aspectos da lei, auxiliando as empresas na tomada de decisões relacionadas a compliance com a LGPD. A execução deste protocolo vislumbra garantir agilidade necessária para implementar a LGPD dentro da organização ou até mesmo, acompanhar o

processo de implantação de terceiros. Este protocolo é um resultado inicial de um trabalho de iniciação científica, o qual pretende-se chegar a uma solução computacional que permita a empresa executar o sistema e identificar todos os passos que precisará seguir.

5. Conclusão

Como resultado desta pesquisa, foi possível fazer uma análise da Lei Geral de Proteção de Dados (LGPD), destacar a importância do agronegócio e a necessidade de adequação das pequenas e médias empresas deste setor, que manipulam dados pessoais, à lei. Com base em uma revisão da bibliografia, uma tabulação sobre os requisitos da lei e da norma ISO 27701, foi implementada uma proposta para *compliance* à LGPD para proporcionar a dimensão das ações que devem ser realizadas por uma empresa do agronegócio que necessite alinhar seus processos e atingir a conformidade.

Reconhecendo que este é um estudo em andamento, tem-se como trabalhos futuros o desenvolvido de um sistema que permita a execução desse modelo de forma automatizada, podendo ser feito o acompanhamento das ações e o alinhamento da empresa a LGPD.

Referências

ABNT (2019). ABNT NBR ISO/IEC 27701 – Técnicas de segurança — Extensão da ABNT NBR ISO/IEC 27001 e ABNT NBR ISO/IEC 27002 para gestão da privacidade da informação.

Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. Disponível em `http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`.

Brito, F. T. e Machado, J. C. (2017). Preservação de privacidade de dados: Fundamentos, técnicas e aplicações. *Jornadas de Atualização em Informática*.

EU (2016). General Data Protection Regulation (GDPR): Regulation (EU) 2016/679. Disponível em `https://eur-lex.europa.eu/eli/reg/2016/679/oj`.

Gomes, H. d. S. (2019). Lei Geral De Proteção de Dados (LGPD): uma análise dos impactos da lei na cultura e tratamento de dados no Brasil.

IBGE, I. B. D. G. E. E. (2021). Contas nacionais trimestrais: Indicadores de volume e valores correntes - jan.-mar. 2021. *Indicadores IBGE*.

Lima, V. H. (2020). LGPD Análise dos impactos da implementação em ambientes corporativos: Estudo de caso.

Prodanov, C. C. e de Freitas, E. C. (2013). *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. Editora Feevale.

Rolim, M. d. C. L. M. e Gibran, S. M. (2021). Lei Geral De Proteção De Dados (Lei Nº 13.709/2018) E Terceiro Setor: Principais Desafios E Alternativas Rumo à Adequação. *CAMPO JURÍDICO*, 9(1):723.

Silva, R. H. d. et al. (2021). Framework para identificar o nível de conformidade das empresas brasileiras do setor químico no processo de adequação à lei geral de proteção de dados pessoais.

Solove, D. J. (2008). Understanding privacy. *Harvard University Press, May*.



Redes Neurais na Simulação de Condições Climáticas Potenciais ao Cultivo de Bambu no Brasil

João Antonio Lorençone[1], Lucas E. de O. Aparecido[2], Pedro Antonio Lorençone[1], Guilherme Botega Torsoni[1]

[1]Instituto Federal de Mato Grosso do Sul (IFMS)
Naviraí – MS – Brazil

[2]Instituto Federal do Sul de Minas (IFSULDEMINAS)
Muzambinho – MG - Brazil

joao.lorencone@estudante.ifms.edu.br,
lucas.aparecido@muz.ifsuldeminas.edu.br,
pedro.lorencone@estudante.ifms.edu.br,
guilherme.torsoni.ifms.edu.br

Abstract. *This study aimed to perform the agricultural zoning of climatic risk for bamboo in Brazil by means of artificial neural networks. It was used climatic data of air temperature (TAIR, ºC) and rainfall (P). The Feed Forward Artificial Neural Network, Multilayer Perceptron (MLP) with backpropagation learning algorithm for multilayers was employed. The agroclimatic zoning allowed the classification of regions by climatic suitability and showed that 71% of the national territory was suitable for bamboo cultivation. The use of the neural network allowed an accurate and fast classification of climate suitability.*

Resumo. *Esse estudo teve como objetivo realizar o zoneamento agrícola de risco climático do bambu no Brasil por meio de redes neurais artificiais. Foram utilizados dados climáticos de temperatura do ar (TAIR, ºC) e precipitação pluviométrica (P). Foi empregado a Rede Neural Artificial Feed Forward, Multilayer Perceptron (MLP) com algoritmo de aprendizagem "backpropagation" para multicamadas. O zoneamento agroclimático permitiu a classificação das regiões por aptidão climática e demonstrou que 71% do território nacional foram aptos para o cultivo do bambu. O uso da rede neural permitiu uma classificação precisa e rápida das aptidões climáticas.*

1. Introdução

O bambu destaca-se no comércio internacional pela sua utilização em diversos fins, como na produção de papel, carvão, artesanato e como substituto a madeira na construção civil.

A biomassa do bambu tem grande potencial para a produção de bioenergia por apresentar alta taxa de crescimento, bem como alto teor de celulose, alto valor calorífico bruto e baixo teor de cinzas. Grande parte dessas caracteristicas do bambu são influênciadas pelo clima da região.

O Zoneamento agrícola de risco climático (ZARC) tem sido uma ferramenta fundamental para minimizar os riscos relacionados aos fenômenos climáticos adversos [de Oliveira Aparecido et al., 2018]. O ZARC pode fornecer informações para identificar as melhores datas de plantio, colheita, assim como identificar os períodos que as condições climáticas atendem às necessidades de um determinado do cultivo [Gelcer et al., 2018]. Neste contexto, objetivou-se realizar o ZARC para a cultura do bambu usando redes neurais, e assim minimizar os riscos de perdas dos produtores por adversidades climáticas incontroláveis.

2. Material e Métodos

Este trabalho foi realizado utilizando dados climáticos do Brasil, foram utilizados dados de temperatura do ar e precipitação pluviométrica de 4.947 estações meteorológicas do Brasil, entre 1950 a 2016. Esta série temporal climática foi coletada pelo Instituto Nacional de Meteorologia do Brasil (INMET).

As áreas aptas para o cultivo do bambu foram definidas a partir dos dados climáticos de Temperatura do ar (TAIR) média anual (ºC), precipitação total anual (P, mm ano-1) e precipitação total do inverno (Pwinter, mm ano-1). Ao se conhecer as necessidades térmico-hídricas do bambu, foram estabelecidas as classes de aptidão climática para cultura.

Foram consideradas as regiões aptas climaticamente para o cultivo quando a temperature média anual do ar (T) variou entre 18 e 35 ºC, a precipitação pluvial (P) apresentou valores entre 500 a 2800 mm ano^{-1} ou Precipitação durante o inverno (Pwinter) entre 90 e 180 mm ano-1. A ocorrência de T acima de 35 ºC foi considerada inapta, pois acima dessa temperatura do ar as plantas tendem a reduzir a fotossíntese líquida, pois a respiração é maior do que o processo fotossintético, inviabilizando a plantação da cultura. Ainda as áreas nas quais apresentaram T abaixo de 10°C foram consideradas inaptas.

Também foram estabelecidas as seguintes restrições: 1ª = T entre 10 e 18 ºC, o crescimento do cultivo é retardado devido ao baixo acúmulo de graus dias durante o ciclo; 2ª = P < 500 mm ano-1 ou Pwinter < que 90 mm ano-1 (desenvolvimento irregular devido ao déficit hídrico); 3ª = P > 2800 mm ano-1 ou Pwinter > de 180 mm ano-1 (excesso de umidade), alto índice de doenças nos cultivos.

Com base nesses critérios de classificações, foi criado o vetor Yj para calibração do modelo, sendo o mesmo representado pela equação 1.

$$\mathrm{Y}_j = [y_1 \dots y_{4947}](y \in \{APTO, INAPTO \dots, \dots RESTRIÇÕES\}) \quad (1)$$

Empregou-se a Rede Neural Artificial Feed Forward, Multilayer Perceptron (MLP), onde os valores de entradas foram as variáveis climáticas selecionadas (T e P) e a saída foram às classes do zoneamento agrícola (output). O aprendizado da MLP foi o supervisionado e o algoritmo de aprendizagem foi o backpropagation para multicamadas.

Uma MLP com N camadas, H neurônios escondidos e um neurônio de saída pode ser expresso pela equação 2:

$$\mathbf{ZONING} = \mathbf{S_k}\left(\sum_{\mathbf{h=1}}^{\mathbf{H}} \mathbf{O_h W_0 W_h}\right) \quad (2)$$

em que: ZONING = é a saída da RNA; Oh = valor de saída do h-ésimo neurônio escondido; W0 = são valores iniciais do algoritmo de treinamento da RNA; Wh = são os pesos sinápticos entre os neurônios escondidos da RNA.

Para realizar a calibração do algoritmo da rede neural foram utilizados 35% dos dados climáticos das localidades, sendo que nesses locais o zoneamento agrícola foi feito de maneira manual, utilizando o software Excel. Com a rede neural calibrada e acurada, foi estabelecida a classificação climática para o bambu para todas as 4.947 estações climáticas do Brasil..

3. Discussão e Resultados

O ZARC para a cultura do bambu demonstrou que 71% do Brasil é apto climaticamente ao cultivo (Figura 1). Os Estados brasileiros que foram classificados como aptos em sua totalidade foram o Mato Grosso do Sul, Goiás, Tocantins, Rio de Janeiro, Espírito Santo, Sergipe, Alagoas, Ceará, Piauí, Maranhão, Rondônia e Acre (Figura 1). Essa enorme faixa de regiões aptas ocorre devido o cultivo do bambu apresentar uma ampla aptidão climática. Estas regiões com condições climáticas ideais ao cultivo associadas aos fatores do manejo, apresentam elevadas produtividades e consequentemente, um maior retorno econômico aos produtores agrícolas. As regiões que apresentam restrições por baixas TAIR representaram apenas 11% do território brasileiro (Figura 2), sendo parte da região Sul do Brasil (região do Planalto e Norte de Santa Catarina, região Sul do Paraná e Sul do Rio Grande do Sul) e o Sul de Minas Gerais. Ressalta-se que nestas regiões o cultivo do bambu pode ser implantado, entretanto, apresentar um crescimento retardado, em função do baixo acúmulo de determinada quantidade de energia, denominada graus-dia acumulados, que o cultivo vai apresentar [Chen et al. 2018].

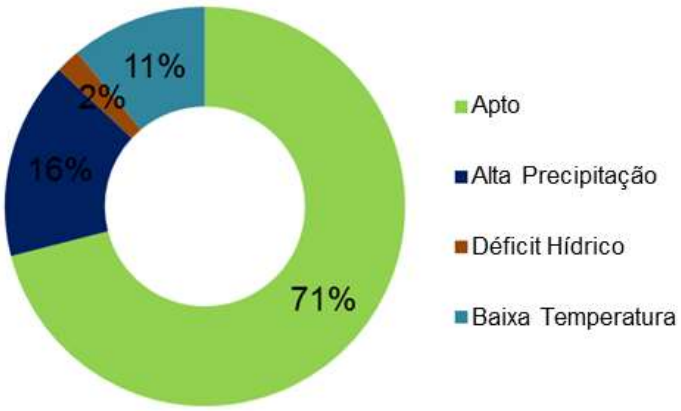


Figure 1. Porcentagem das áreas para cada classe de zoneamento agrícola de risco climático do bambu no Brasil.

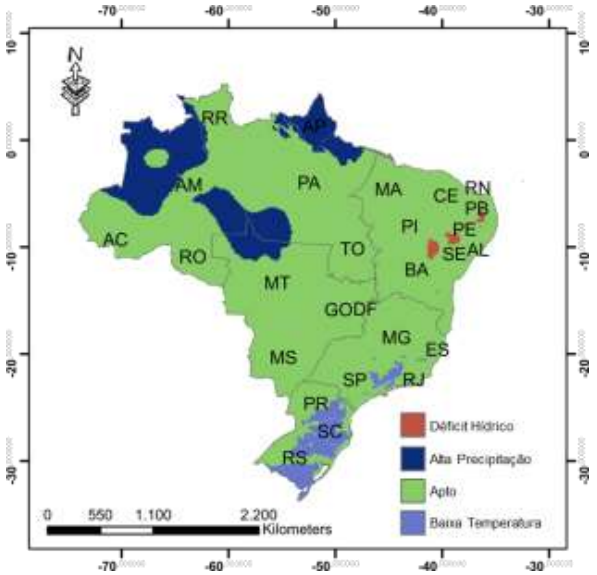


Figura 2. Zoneamento agrícola de risco climático da cultura do bambu para o Brasil.

Conclusões

A rede neural é uma ferramenta eficiente na realização de zoneamento agrícola de risco climático, permitindo uma classificação rápida e acurada das aptidões climáticas.

O zoneamento agrícola de risco climático demonstra que 71% do território brasileiro são aptos climaticamente ao cultivo do bambu.

As regiões com restrição foram parte do Sul do Brasil, por apresentar baixas TAIR e parte da região no Norte por apresentar elevadas P que proporcionam elevados índices de doenças

References

Chen, S., Jiang, H., Cai, Z., Zhou, X., & Peng, C. (2018). The response of the net primary production of Moso bamboo forest to the On and Off-year management: A case study in Anji County, Zhejiang, China. *Forest Ecology and Management*, *409*, 1–7. https://doi.org/10.1016/j.foreco.2017.11.008

de Oliveira Aparecido, L. E., Moreto, V. B., de Souza Rolim, G., da Silva Cabral de Moraes, J. R., Valeriano, T. T. B., & de Souza, P. S. (2018). Climatic potential for summer and winter wine production: Climatic potential for summer and winter wine production. *Journal of the Science of Food and Agriculture*, *98*(4), 1280–1290. https://doi.org/10.1002/jsfa.8575

Gelcer, E., Fraisse, C. W., Zotarelli, L., Stevens, F. R., Perondi, D., Barreto, D. D., … Southworth, J. (2018). Influence of El Niño-Southern oscillation (ENSO) on agroclimatic zoning for tomato in Mozambique. *Agricultural and Forest Meteorology*, *248*, 316–328. https://doi.org/10.1016/j.agrformet.2017.10.002