

**PROGRAMA DE PÓS-GRADUAÇÃO
EM CIÊNCIA DA COMPUTAÇÃO**



Abordagens Guiadas por Dados para Estimativa de Produtividade de
Cana-de-açúcar: Métodos e Conjunto de Dados

Luiz Antonio Falaguasta Barbosa

INSTITUTO DE GEOCIÊNCIAS E CIÊNCIAS EXATAS
RIO CLARO

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

Instituto de Geociências e Ciências Exatas

Câmpus de Rio Claro

Luiz Antonio Falaguasta Barbosa

**Abordagens guiadas por dados para estimativa de
produtividade de cana-de-açúcar: métodos e conjunto de
dados**

Tese de Doutorado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Orientador: Prof. Dr. Daniel Carlos Guimarães Pedronette

Co-orientador: Prof. Dr. Ivan Rizzo Guilherme

Rio Claro - SP

2026

B238a

Barbosa, Luiz Antonio Falaguasta

Abordagens guiadas por dados para estimativa de produtividade de cana-de-açúcar :
métodos e conjunto de dados / Luiz Antonio Falaguasta Barbosa. -- Rio Claro, 2026
183 f. : il., tabs., fotos, mapas

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Instituto de Geociências
e Ciências Exatas, Rio Claro

Orientador: Daniel Carlos Guimarães Pedronette

Coorientador: Ivan Rizzo Guilherme

1. Estimativa de Produtividade. 2. Aprendizado de Máquina. 3. Aprendizado
Profundo. 4. Aprendizado Supervisionado. 5. Sensoriamento Remoto. I. Título.

Impacto potencial desta pesquisa:

Esta pesquisa contribui em duas frentes — metodológica e aplicada. No âmbito metodológico, a geração de *meta-features* por *stacking ensemble* demonstrou ser possível melhorar o desempenho preditivo sem coletar dados adicionais de campo, sendo transferível para outros problemas com pequenas amostras e alta dimensionalidade. O uso do kNN como metarregressor, explorando a estrutura de vizinhança no espaço de *meta-features*, abre perspectivas para pesquisas sobre representações aprendidas e algoritmos baseados em instâncias. No âmbito aplicado, as ferramentas desenvolvidas podem subsidiar decisões no setor sucroalcooleiro brasileiro, responsável por cerca de 40% da produção mundial de cana-de-açúcar, beneficiando a negociação de contratos futuros, o planejamento da colheita, a reforma de canaviais e estimativas oficiais de safra. A disponibilização pública do conjunto de dados multimodal com 2.157 observações de campos comerciais paulistas supera uma barreira reconhecida na literatura. Ao integrar aprendizado de máquina, sensoriamento remoto e conhecimento agrônômico, esta tese consolida uma agenda interdisciplinar com impacto potencial na segurança alimentar e energética do país.

Potential impact of this research:

This research contributes on two fronts — methodological and applied. Methodologically, the meta-feature generation approach based on stacking ensemble showed that predictive performance can be improved without additional field data collection, and is transferable to other regression problems with small sample sizes and high dimensionality. The use of kNN as a meta-regressor, exploiting the neighborhood structure in the meta-feature space, opens perspectives for research on learned representations and instance-based algorithms. From an applied standpoint, the tools developed can support decision-making in Brazil's sugar and ethanol sector, which accounts for approximately 40% of global sugarcane production, benefiting futures contract negotiation, harvest planning, ratoon field renewal, and official crop estimates. The public release of the multimodal dataset comprising 2,157 observations from commercial fields in São Paulo state overcomes a recognized barrier in the literature. By integrating machine learning, remote sensing, and agronomic knowledge, this thesis consolidates an interdisciplinary research agenda with potential impact on the country's food and energy security.

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Luiz Antonio Falaguasta Barbosa

**Abordagens guiadas por dados para estimativa de
produtividade de cana-de-açúcar: métodos e conjunto de
dados**

Tese de Doutorado apresentada ao Instituto
de Geociências e Ciências Exatas do Câmpus
de Rio Claro da Universidade Estadual Pau-
lista “Júlio de Mesquita Filho”, como parte
dos requisitos para a obtenção do título de
Doutor em Ciência da Computação.

Comissão Examinadora

Prof. Dr. Daniel Carlos Guimarães Pedronette
IGCE/UNESP - Rio Claro (SP)

Prof. Dr. Pedro Luiz Pizzigatti Corrêa
POLI/USP - São Paulo (SP)

Prof. Dr. José Paulo Molin
ESALQ/USP - Piracicaba (SP)

Prof. Dr. Antonio Maria Garcia Tommaselli
FCT/UNESP - Presidente Prudente (SP)

Prof. Dr. Wallace Correa de Oliveira Casaca
IBILCE/UNESP - São José do Rio Preto (SP)

Conceito: Aprovado.

Rio Claro (SP), 30 de junho de 2026.

Agradecimentos

Agradeço aos meus saudosos pais, à minha esposa e ao meu filho, pelo amor, pelo carinho e pelo incentivo.

Aos meus orientadores, pelas ideias, pelos ensinamentos e pela credibilidade.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“Estude muito o que mais lhe interessa da maneira mais indisciplinada, irreverente e original possível.”
(Richard Feynman)

Resumo

Estimar, com meses de antecedência, a quantidade a ser produzida no cultivo da cana-de-açúcar é de suma importância para a tomada de decisão no setor sucroalcooleiro do Brasil. Esta tese investiga abordagens guiadas por dados para a predição da produtividade da cana-de-açúcar, propondo e avaliando diferentes estratégias baseadas em Aprendizado de Máquina (AM) para a integração de dados multimodais provenientes de sensoriamento remoto, de sensores LiDAR e de variáveis agronômicas. As principais contribuições desta tese são: (i) um estudo comparativo sistemático de regressores (*Support Vector Regression*, *Random Forest*, *Gradient Boosting*, AdaBoost e Regressão Linear) aplicados à predição de produtividade, demonstrando que o regressor AdaBoost apresenta os melhores resultados no conjunto de dados analisado; (ii) a proposição de um modelo de geração de *meta-features* (MFs) baseado em *stacking ensemble*, no qual 25 vetores de MFs são gerados a partir de cinco configurações de *features* originais e cinco regressores base, permitindo a avaliação comparativa de diferentes representações; (iii) a extensão da abordagem de MFs com a incorporação do regressor *k-Nearest Neighbors* (kNN), tanto como gerador de MFs quanto como metarregressor, alcançando valores de $\bar{R}^2$ próximos a 0,99; (iv) a avaliação de características visuais extraídas por redes profundas pré-treinadas (ResNet-101 e DINOv2) combinados com técnicas de redução de dimensionalidade (PCA e UMAP); e (v) a construção e disponibilização de um novo conjunto de dados multimodal de cana-de-açúcar proveniente de campos comerciais no estado de São Paulo, compreendendo 2.157 observações com dados agronômicos, meteorológicos e imagens de satélite. Os resultados demonstram que a geração de MFs a partir de diferentes regressores captura informações preditivas complementares às *features* originais e que a combinação dessas MFs com o metarregressor AdaBoost produz ganhos significativos de desempenho, com $\bar{R}^2$ médio superior a 0,90 nas configurações agregadas e alcançando $\bar{R}^2$ de até 0,9946 com o metarregressor kNN. As contribuições desta tese têm implicações práticas para o setor sucroalcooleiro, oferecendo ferramentas para estimativas mais precisas de produtividade e implicações metodológicas para a área de Ciência da Computação, particularmente no que se refere à fusão de dados multimodais em cenários com pequeno número de observações e alta dimensionalidade.

Palavras-chave: Estimativa de Produtividade, Aprendizado de Máquina, Aprendizado Profundo, Métodos de Aprendizado Supervisionado, Índices Vegetativos, Dados Agronômicos, Dados Meteorológicos, Dados Visuais, Sensoriamento Remoto, RPA.

Abstract

Estimating, months in advance, the amount of sugarcane to be produced is of paramount importance for decision-making in Brazil's sugar and alcohol sector. This thesis investigates data-driven approaches to sugarcane yield prediction, proposing and evaluating various machine learning strategies for integrating multimodal data from remote sensing, LiDAR, and agronomic variables. The main contributions of this thesis are: (i) a systematic comparative study of regressors (Support Vector Regression, Random Forest, Gradient Boosting, AdaBoost, and Linear Regression) applied to yield prediction, demonstrating that AdaBoost achieves the best results on the analyzed dataset; (ii) the proposition of a meta-feature generation model based on stacking ensemble, in which 25 MF vectors are generated from five original feature configurations and five base regressors, enabling comparative evaluation of different representations; (iii) the extension of the MF approach by incorporating the k-Nearest Neighbors (kNN) regressor both as a MF generator and as a meta-regressor, achieving $\bar{R}^2$ values close to 0.99; (iv) the evaluation of visual features extracted by pre-trained deep networks (ResNet-101 and DINOv2) combined with dimensionality reduction techniques (PCA and UMAP); and (v) the construction and public availability of a new multimodal sugarcane dataset from commercial fields in São Paulo state, comprising 2,157 observations with agronomic, meteorological, and satellite image data. The results demonstrate that MFs generated from different regressors capture predictive information complementary to the original features, and that combining MFs with the AdaBoost meta-regressor yields significant performance gains, with average $\bar{R}^2$ exceeding 0.90 in aggregated configurations and reaching $\bar{R}^2$ up to 0.9946 with the kNN metaregressor. The contributions of this thesis have practical implications for the sugar-alcohol sector, providing tools for more accurate yield estimates, and methodological implications for Computer Science, particularly regarding multimodal data fusion in scenarios with small sample sizes and high dimensionality.

Keywords: Crop Yield Prediction, Machine Learning, Deep Learning, Supervised Learning Methods, Vegetative Indices, Agronomic Data, Meteorological Data, Visual Data, Remote Sensing, RPA.

Lista de abreviaturas e siglas

AM	<i>Aprendizado de Máquina</i>
ANN	<i>Artificial Neural Networks</i>
AP	<i>Agricultura de Precisão</i>
APSIM	<i>Agricultural Production Systems sIMulator</i>
CAR	<i>Cadastro Ambiental Rural</i>
CNN	<i>Convolutional Neural Networks</i>
CONAB	<i>Companhia Nacional de Abastecimento</i>
DAC	<i>Dias Após Colheita</i>
DDA	<i>Dia do ano</i>
DNN	<i>Deep Neural Networks</i>
EMBRAPA	<i>Empresa Brasileira de Pesquisa Agropecuária</i>
EN	<i>Elastic Net</i>
EVI	<i>Enhanced Vegetation Index</i>
FAO	<i>Food and Agriculture Organization</i>
GIS	<i>Geographic Information System</i>
GNSS	<i>Global Navigation Satellite System</i>
IAC	<i>Instituto Agronômico de Campinas</i>
IAPAR	<i>Instituto Agronômico do Paraná</i>
IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>
IBSNAT	<i>International Benchmark Sites Network for Agrotechnology Transfer</i>
ICASA	<i>International Consortium for Agricultural Systems Applications</i>
ICIS	<i>International Crop Information System</i>
INMET	<i>Instituto Nacional de Meteorologia</i>
INPE	<i>Instituto Nacional de Pesquisas Espaciais</i>

IRGA	<i>Infrared Gas Analyzer</i>
kNN	<i>k-Nearest Neighbor</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
LiDAR	<i>Light Detection And Ranging</i>
LR	<i>Linear Regression</i>
LST	<i>Land Surface Temperature</i>
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MARS	<i>Monitoring Agricultural ResourceS</i>
MCYFS	<i>MARS Crop Yield Forecasting System</i>
MF	<i>Meta-feature</i>
MODIS	<i>Moderate Resolution Imaging Spectroradiometer</i>
MSE	<i>Mean Squared Error</i>
NASA	<i>National Aeronautics and Space Administration</i>
NASA POWER	<i>NASA Prediction of Worldwide Energy Resource</i>
NDMI	<i>Normalized Difference Moisture Index</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NDWI	<i>Normalized Difference Water Index</i>
NIR	<i>Near InfraRed</i>
NLP	<i>Natural Language Processing</i>
NRMSE	<i>Normalized Root Mean Squared Error</i>
OLS	<i>Ordinary Least Squares</i>
OOB	<i>out-of-bag</i>
PCA	<i>Principal Component Analysis</i>
R^2	<i>Coefficient of Determination</i>
ReLU	<i>Rectified Linear Unit</i>

ResNet	<i>Residual neural network</i>
RF	<i>Random Forest</i>
RFR	<i>Random Forest Regression</i>
RGB	<i>Red, Green, Blue</i>
RMSE	<i>Root Mean Squared Error</i>
RNN	<i>Recurrent Neural Network</i>
RPA	<i>Remotely Piloted Aircraft</i>
SAR	<i>Synthetic-aperture radar</i>
SLAM	<i>Simultaneous Localization and Mapping</i>
TCH	<i>Toneladas de Cana por Hectare</i>
VGGNet	<i>Very Deep Convolutional Network from Visual Geometry Group</i>
SAVI	<i>Soil-Adjusted Vegetation Index</i>
SSTNN	<i>Spatial-Spectral-Temporal Neural Network</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
SWIR	<i>Short-wave infrared</i>
USDA	<i>United States Department of Agriculture</i>
VARI	<i>Visible Atmospherically Resistant Index</i>
ViT	<i>Vision Transformer</i>

Lista de ilustrações

Figura 1 – Evolução da área plantada de cana-de-açúcar no Brasil	20
Figura 2 – Evolução da participação de São Paulo na área plantada de cana-de-açúcar	21
Figura 3 – Visão geral estruturada das contribuições da tese	29
Figura 4 – Visão geral da estrutura conceitual da tese	31
Figura 5 – Satélites públicos	37
Figura 6 – Diferentes tipos de câmeras	38
Figura 7 – Estação meteorológica automática	43
Figura 8 – Rede de estações meteorológicas do INMET	44
Figura 9 – Dados meteorológicos disponíveis no NASA POWER	45
Figura 10 – Tipos de aprendizado de máquina	49
Figura 11 – Hiperplano de separação de classes no SVM	51
Figura 12 – SVM de margens suaves	52
Figura 13 – <i>Kernel trick</i>	52
Figura 14 – Parte de uma rede neural convolucional	55
Figura 15 – A função não linear ReLU	55
Figura 16 – Imagem original subdividida	56
Figura 17 – Visão geral do modelo	57
Figura 18 – Imagens de entrada e suas respectivas saídas do mecanismo de atenção	58
Figura 19 – Estrutura das características para a estimativa da produtividade	63
Figura 20 – Resumo do método para estimar a produtividade da cana-de-açúcar	68
Figura 21 – $\bar{R}^2$ de cada modelo em cada pesquisa para predição de biomassa	69
Figura 22 – Estrutura conceitual, na qual os dados de produção não são uma entrada	72
Figura 23 – Visão geral do SSTNN proposto	73
Figura 24 – Visão geral do <i>workflow</i> MCYFS	73
Figura 25 – Fluxograma que indica todas as etapas do desenvolvimento do modelo	75
Figura 26 – Configuração experimental	78
Figura 27 – RPA multirrotor	79
Figura 28 – Composição das cinco configurações de <i>features</i>	80
Figura 29 – Estrutura de validação cruzada	83
Figura 30 – Estrutura completa da validação cruzada <i>k-fold</i>	84
Figura 31 – Localização dos campos experimentais	85
Figura 32 – Desenho experimental contendo pontos de coleta de biomassa	86
Figura 33 – <i>Workflow</i> para a obtenção de uma estimativa de produtividade	94
Figura 34 – Resultados do regressor OLS. Fonte (SHENDRYK; DAVY; THOR- BURN, 2021)	95
Figura 35 – Regressores MLP, SVR, RF, AB e GB.	96

Figura 36 – Protocolo para a obtenção de uma estimativa de produtividade	101
Figura 37 – Visualização detalhada da etapa 5	102
Figura 38 – Processo de geração de MFs e de avaliação pelo metarregressor AdaBoost	104
Figura 39 – RMSE calculado para os modelos M1 (a) e M5 (b)	106
Figura 40 – Gráficos contendo os modelos com e sem MFs	106
Figura 41 – Extensão do processo de MFs com kNN	113
Figura 42 – Impacto do espaço de representação no desempenho do kNN	116
Figura 43 – Gráficos de MFs kNN sob o metarregressor Adaboost	118
Figura 44 – Gráficos de MFs kNN sob o metarregressor kNN	119
Figura 45 – Local #1 e Local #2 durante a segunda coleta de amostras	121
Figura 46 – Estrutura da validação cruzada temporal com janela expansiva	123
Figura 47 – Resultados comparativos das configurações de entrada do AdaBoost . .	125
Figura 48 – Projeção usando UMAP	127
Figura 49 – Os campos experimentais são apresentados como Local #1 e Local #2	134
Figura 50 – Exemplo de imagem aérea de uma parcela em um campo experimental	135
Figura 51 – <i>Workflow</i> de avaliação de resultados com reduções de dimensionalidade	135
Figura 52 – Uso e cobertura da terra no estado de São Paulo em 2024	145
Figura 53 – Estrutura do repositório de dados	146
Figura 54 – Mapas nacional, estadual e de áreas de cultivo de cana-de-açúcar . . .	149
Figura 55 – Posição dos talhões de cultivo de cana-de-açúcar	150
Figura 56 – Imagens da mesma área como exemplo de ocorrência e não de nuvens .	153
Figura 57 – Representação em série temporal de três conjuntos de imagens	153

Lista de tabelas

Tabela 1 – Variáveis com potencial de uso na estimativa de produtividade	33
Tabela 2 – Tabela de características utilizadas na estimativa da produtividade . .	61
Tabela 3 – Tabela de algoritmos de aprendizado de máquina	65
Tabela 4 – Tabela dos algoritmos de aprendizado de máquina profundo	66
Tabela 5 – Lista de artigos sobre estimativa de produtividade	67
Tabela 6 – Hiperparâmetros usados em cada regressor aplicado.	82
Tabela 7 – Índices vegetativos e suas equações	87
Tabela 8 – Medidas estatísticas calculadas a partir dos IVs	87
Tabela 9 – Medidas estatísticas calculadas a partir de dados LiDAR	87
Tabela 10 – Datas de coleta de amostras nos campos experimentais	88
Tabela 11 – Sumarização de médias de valores de $\overline{R}^2$ para os diferentes modelos . .	97
Tabela 12 – Correspondência entre os modelos e as características de que se baseiam.	103
Tabela 13 – Correspondência entre os modelos e as características de que se baseiam.	117
Tabela 14 – Comparação entre a linha de base e a abordagem do método proposto	120
Tabela 15 – Descrição de cada <i>feature</i> utilizada na predição.	134
Tabela 16 – Resultados de desempenho na predição de produtividade	141
Tabela 17 – Resultados de desempenho do extrator de características DINOv2 . . .	142
Tabela 18 – Características, descrição e tipo de dados	147
Tabela 19 – Características, descrição e tipo de dados do conjunto meteorológico. .	148
Tabela 20 – <i>Features</i> , descrição e tipo de dados	148
Tabela 21 – Identificação das fazendas	149
Tabela 22 – Acrônimo, nome correspondente e descrição dos tipos de solo	151
Tabela 23 – Ambientes de produção	151
Tabela 24 – Perguntas de pesquisa e respectivos capítulos.	156
Tabela 25 – Respostas às perguntas de pesquisa.	160
Tabela 26 – Publicações produzidas durante o doutorado.	165

Sumário

1	INTRODUÇÃO	19
1.1	Contextualização	19
1.2	Motivação e Desafios de Pesquisa	21
1.3	Perguntas de Pesquisa	23
1.4	Hipótese	24
1.5	Objetivos	25
1.6	Contribuições	27
1.7	Organização	30
2	FUNDAMENTAÇÃO TEÓRICA	32
2.1	Predição de Produtividade	32
2.2	Sensoriamento Remoto	36
2.3	Dados de Imageamento	39
2.4	Dados Agronômicos	39
2.5	Dados Meteorológicos	42
2.6	Extração de Características	43
2.7	Representação Temporal	46
2.8	Modelos de Simulação de Crescimento de Safra	47
2.9	Modelos de Aprendizado de Máquina	48
3	TRABALHOS RELACIONADOS	59
3.1	Visão Geral da Predição de Produtividade Agrícola	59
3.2	Tipos de Dados	60
3.3	Técnicas para Prever Produtividade de Culturas Agrícolas	62
3.4	Abordagens Baseadas em Sensoriamento Remoto	65
3.5	Modelos de Aprendizado de Máquina para Predição de Produtividade	65
3.5.1	Cana-de-açúcar	67
3.5.2	Outras Culturas	70
4	PROTOCOLO EXPERIMENTAL	77
4.1	Configuração Experimental	77
4.2	Conjuntos de Dados para Experimentos	82
4.3	Protocolo de Avaliação e Métricas	89
5	ESTUDO COMPARATIVO ENTRE REGRESSORES	92
5.1	Considerações Iniciais	92

5.2	Metodologia	93
5.3	Resultados	95
5.4	Discussão	96
5.5	Considerações Finais	97
6	MODELO DE <i>META-FEATURES</i> PARA COMBINAÇÃO DE DIFERENTES REGRESSORES	99
6.1	Considerações Iniciais	99
6.2	Geração de <i>Meta-Features</i> e Metarregressor	100
6.3	Avaliação Experimental e Resultados	105
6.4	Considerações Finais	107
7	MODELO DE <i>META-FEATURES</i> BASEADO NO REGRESSOR KNN	109
7.1	Considerações Iniciais	109
7.2	Metodologia	111
7.3	Avaliação Experimental e Resultados	117
7.4	Discussão	125
7.5	Considerações Finais	129
8	ESTUDO COMPARATIVO DE CARACTERÍSTICAS VISUAIS E REDUÇÃO DE DIMENSIONALIDADE	130
8.1	Considerações Iniciais	130
8.2	Metodologia Proposta	132
8.3	Extração de Características Visuais	136
8.4	Redução de Dimensionalidade	137
8.5	Resultados	139
8.6	Discussão e Limitações	139
8.7	Considerações Finais	140
9	UM CONJUNTO DE DADOS DE CANA-DE-AÇÚCAR DE CAMPOS COMERCIAIS EM SÃO PAULO	143
9.1	Considerações Iniciais	143
9.2	Descrição de Dados	145
9.3	Desenho Experimental, Materiais e Métodos	148
10	CONCLUSÕES	155
10.1	Contexto da Hipótese e Desafios	155
10.2	Síntese dos Resultados e Discussão	155
10.3	Respostas às Perguntas de Pesquisa	159
10.4	Contribuições Científicas	161
10.5	Limitações do Trabalho	162

10.6	Trabalhos Futuros	163
10.7	Publicações e Colaboração Internacional	165
10.8	Considerações Finais	166
	REFERÊNCIAS	168

1 Introdução

1.1 Contextualização

O Brasil é o maior produtor de cana-de-açúcar (*Saccharum officinarum*), responsável por 40% da produção mundial (FOOD; ORGANIZATION, 2022). O estado de São Paulo é responsável por 53,3% da produção brasileira de cana e por 47,5% da produção de etanol, pois é o estado com o maior índice de mecanização (98,5%) empregada nesta cultura agrícola (BOLETIM. . . ,). Essa é uma cultura semiperene, cultivada a partir de 4 a 6 colheitas do mesmo plantio, sendo a primeira chamada de colheita de cana-planta e as demais, de cana-soca, provenientes do rebrotamento da cana. O primeiro ciclo de cultivo pode levar de 12 a 18 meses, a depender da variedade cultivada, da época de plantio e das condições ambientais.

Para dimensionar a magnitude e a dinâmica deste setor no cenário nacional, a Figura 1 apresenta a evolução da área plantada de cana-de-açúcar no Brasil e a participação do estado de São Paulo entre 2006 e 2025, com base em dados do Acompanhamento da Safra Brasileira de Cana-de-açúcar da Companhia Nacional de Abastecimento (CONAB). É possível observar que, embora a área plantada total no Brasil tenha apresentado um crescimento expressivo até o final da década passada, seguido de um período de estabilização, a participação de São Paulo manteve-se expressiva, respondendo por uma parcela majoritária da área cultivada.

A análise da participação percentual do estado de São Paulo na área plantada nacional, detalhada na Figura 2, revela que após um período de crescimento contínuo, observa-se um ponto de inflexão por volta de 2018. A partir dali, a participação relativa de São Paulo se estabiliza e apresenta uma leve tendência de redução. Este fenômeno está associado à expansão do cultivo da cana-de-açúcar para novas fronteiras agrícolas, principalmente na região Centro-Oeste do Brasil, impulsionada por fatores logísticos e pela disponibilidade de terras.

Este cenário de concentração regional e, ao mesmo tempo, de diversificação geográfica da produção, aliado à complexidade dos fatores que influenciam a produtividade (clima, solo, manejo), estabelece o contexto do problema central investigado neste trabalho. A necessidade de ferramentas de suporte à decisão capazes de integrar dados heterogêneos — como imagens de satélite e dados agronômicos — para estimar a produtividade de forma antecipada e precisa torna-se ainda mais crítica para a gestão eficiente dos canaviais, tanto em São Paulo quanto nas novas regiões produtoras.

Existem hoje formas de coletar remotamente dados para estimar a produtividade

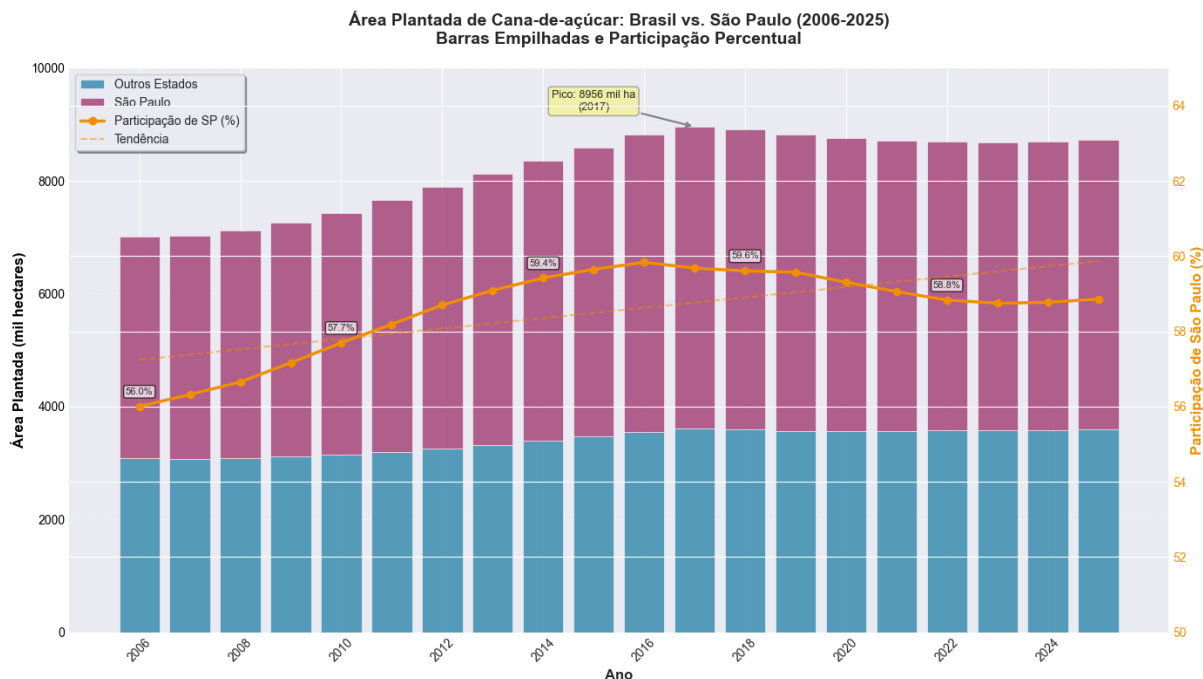


Figura 1 – Evolução da área plantada de cana-de-açúcar no Brasil (barras empilhadas) e da participação percentual do estado de São Paulo (linha) entre 2006 e 2025. As barras azuis representam outros estados e as barras roxas, São Paulo, totalizando a área do Brasil. A linha contínua laranja e a linha tracejada mostram, respectivamente, a participação real e a tendência linear de São Paulo. Baseado em dados de (Companhia Nacional de Abastecimento (CONAB), 2025).

da cana com base em dados meteorológicos e em modelos de crescimento de plantas (TEMPOCAMPO, 2021), em dados de sensoriamento remoto (FERNANDES; EBECKEN; ESQUERDO, 2017), e nos tipos de dados supramencionados, somados aos dados agrônômicos, fazendo uso de Aprendizado de Máquina (AM) ao nível de talhão¹ (LUCIANO *et al.*, 2021); bem como na observação realizada por especialistas com experiência no cultivo, por meio de visita a campo realizada poucos dias antes da colheita, que coletam amostras.

Tais possibilidades, com coletas remotas, são viabilizadas pelo uso de dados georreferenciados a partir de imagens. Embora dados de satélite sejam utilizados há décadas na pesquisa agrícola e em estudos governamentais (DORIGO *et al.*, 2007), a recente disponibilidade de imagens públicas de alta resolução espacial, espectral e temporal tornou a informação significativamente mais rica para os produtores. Além disso, estimativas mais fidedignas podem ser obtidas quando as abordagens são combinadas com dados coletados no campo (*on farm*), tais como informações agrônômicas da cultura, dados ambientais coletados por estações meteorológicas e avaliações biométricas no campo, como medidas de comprimento e diâmetro de colmos², número de colmos, quantidade de açúcar no colmo³,

¹ Porção de terreno de cultivo agrícola.

² Cada segmento do caule da cana-de-açúcar.

³ Brix: indica a qualidade da cana colhida.

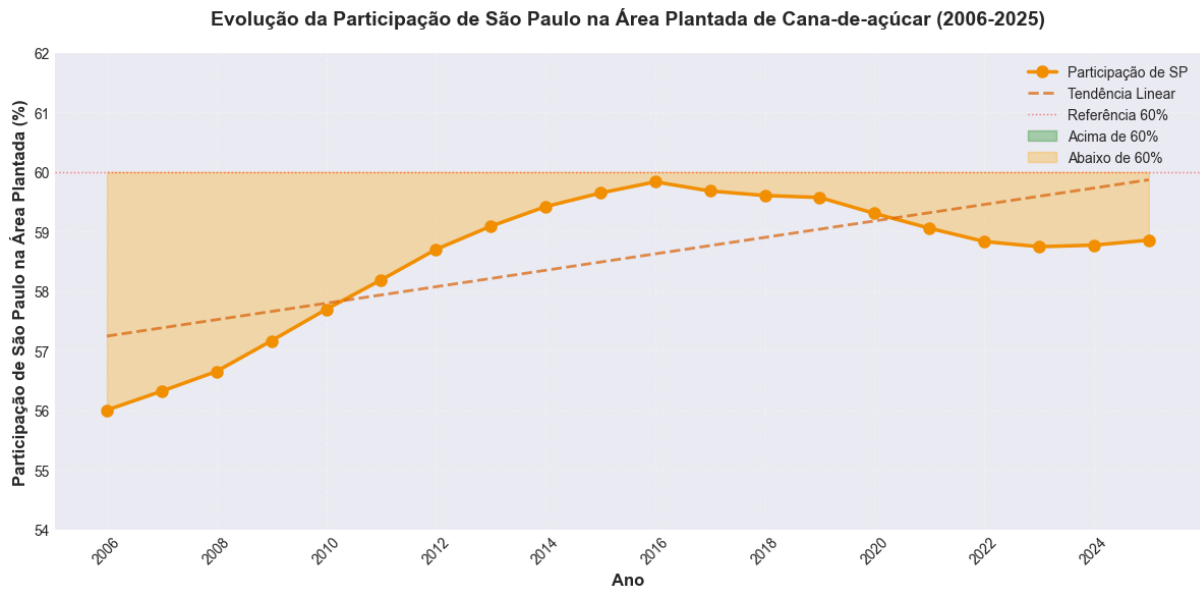


Figura 2 – Evolução da participação do estado de São Paulo na área total plantada de cana-de-açúcar no Brasil entre 2006 e 2025, com destaque para a linha de tendência linear e para a referência de 60%. Baseado em dados de (Companhia Nacional de Abastecimento (CONAB), 2025).

amostras de folhas para mensuração de nitrogênio, dentre outros (SILVA *et al.*, 2024).

A predição da safra é definida pela Organização das Nações Unidas para a Alimentação e a Agricultura (FAO) como a arte de prever a produção da safra antes da colheita, normalmente com alguns meses de antecedência (GOMMES; BERNARDI; PETRASSI, 1996).

1.2 Motivação e Desafios de Pesquisa

A predição da produtividade de culturas agrícolas constitui uma tarefa fundamental para os tomadores de decisão no âmbito da Agricultura de Precisão (AP) (van Klompenburg; KASSAHUN; CATAL, 2020). No contexto específico da cana-de-açúcar, essa estimativa adquire relevância ainda maior dada a magnitude do setor no Brasil, que responde por aproximadamente 40% da produção mundial (FOOD; ORGANIZATION, 2022), e a concentração da produção no estado de São Paulo, responsável por mais da metade da área cultivada nacionalmente, conforme ilustrado nas Figuras 1 e 2.

A capacidade de estimar, com meses de antecedência, quantas toneladas de cana por hectare (TCH) serão colhidas permite ao produtor elaborar um planejamento estratégico abrangente. Do ponto de vista comercial, a estimativa antecipada subsidia decisões de comercialização da produção, permitindo a negociação de contratos futuros em condições mais favoráveis e a gestão de riscos de mercado. No âmbito operacional, possibilita antecipar investimentos na aquisição de insumos para a safra seguinte — como fertilizantes, defensivos

e mudas —, otimizando o fluxo de caixa e reduzindo custos financeiros associados à compra tardia (SHENDRYK *et al.*, 2020).

Ainda no campo das decisões agronômicas, a estimativa de produtividade subsidia a determinação da proporção ideal de área a ser reformada após o ciclo produtivo da cana-soca, bem como o momento mais adequado para iniciar esse processo. A reforma de canaviais envolve custos significativos e impacta diretamente a produtividade das safras subsequentes.

Ademais, a modelagem desses dados por meio de técnicas de AM pode auxiliar canavicultores e usineiros a realizar estimativas mais precisas das safras futuras, mesmo diante da complexidade inerente ao sistema produtivo. A predição da produtividade da cana-de-açúcar é intrinsecamente desafiadora devido à dependência de múltiplos fatores inter-relacionados, incluindo condições climáticas, padrões meteorológicos, propriedades do solo, práticas de manejo e características genéticas das variedades cultivadas (SILVA *et al.*, 2024; van Klompenburg; KASSAHUN; CATAL, 2020; AMARASINGAM *et al.*, 2022; NOGUEIRA *et al.*, 2023; RIBEIRO *et al.*, 2024; ARAKAWA *et al.*, 2025). Nesse cenário, diferentes abordagens têm sido exploradas para lidar com essas complexidades, desde o uso de IVs obtidos por sensoriamento remoto até a integração de dados multiespectrais, de *Light Detection And Ranging* (LiDAR) e de variáveis agronômicas (LUCIANO *et al.*, 2021; SHENDRYK *et al.*, 2020).

A relevância da estimativa antecipada transcende o âmbito privado. A disponibilidade de metodologias robustas para predição de safras pode servir de base científica para orientar iniciativas de políticas governamentais no setor sucroalcooleiro. Órgãos como a CONAB e o Instituto Brasileiro de Geografia e Estatística (IBGE) realizam levantamentos sistemáticos da produção agrícola, e a incorporação de abordagens orientadas por dados pode aprimorar a precisão e a tempestividade dessas estimativas oficiais, beneficiando toda a cadeia produtiva e a formulação de políticas públicas.

Em suma, a estimativa antecipada da produtividade da cana-de-açúcar não se limita a uma ferramenta de gestão isolada; trata-se de uma capacidade estratégica que permeia decisões de natureza comercial, operacional, agrônômica e institucional. A diversidade e a profundidade dessas decisões justificam o investimento no desenvolvimento de modelos preditivos cada vez mais precisos e generalizáveis, capazes de explorar a crescente disponibilidade de dados provenientes de sensoriamento remoto, de estações meteorológicas e de registros de manejo agrícola.

Dada a relevância da tarefa de predição de produtividade no setor sucroalcooleiro, um amplo espectro de técnicas baseadas em AM tem sido desenvolvido e aplicado (SILVA *et al.*, 2024; van Klompenburg; KASSAHUN; CATAL, 2020; SHAWON *et al.*, 2025). Contudo, dentre um grande conjunto de técnicas e algoritmos, é um desafio pouco investigado em que situações se podem obter os melhores resultados de cada abordagem.

Como as formas existentes de estimativa de produtividade ainda não fornecem respostas com alta explicabilidade dos dados e baixo erro (SHENDRYK *et al.*, 2020), métodos de AM vêm contribuindo para que o produtor tenha informações relevantes sobre sua produção. Diante do exposto, identificam-se três limitações principais nas abordagens atuais para predição da produtividade da cana-de-açúcar:

- (i) a ausência de estudos comparativos sistemáticos entre diferentes regressores, que dificulta a seleção do modelo mais adequado para cada cenário;
- (ii) a falta de exploração de representações de dados mais ricas, como MFs derivadas de múltiplos regressores; e
- (iii) a integração limitada de modalidades heterogêneas de dados — incluindo características visuais extraídas de imagens, índices vegetativos e dados LiDAR — no mesmo arcabouço preditivo.

Essas lacunas motivam as investigações conduzidas nesta tese.

Nesse cenário, esta tese aborda o problema computacional da integração de múltiplas modalidades de dados heterogêneos para predição em contextos com pequeno número de observações e alta dimensionalidade. Tal problema surge com frequência em aplicações de sensoriamento remoto agrícola, nas quais diferentes fontes de informação — como imagens, índices espectrais e variáveis meteorológicas — devem ser combinadas para estimar variáveis agrônomicas de interesse. Além disso, avalia e combina diferentes regressores, avalia técnicas de redução de dimensionalidade, utiliza *features* visuais e propõe um conjunto de dados de *benchmark*.

Embora a literatura recente apresente avanços em modelos multimodais profundos, muitos desses métodos pressupõem a disponibilidade de grandes volumes de dados, o que limita sua aplicação em experimentos agrícolas de campo.

1.3 Perguntas de Pesquisa

1. Como diferentes algoritmos de regressão se comportam na predição da produtividade da cana-de-açúcar quando avaliados no mesmo conjunto de dados e sob o mesmo protocolo experimental?
2. A geração de MFs a partir das saídas de modelos de regressão previamente treinados pode melhorar o desempenho preditivo em comparação ao uso apenas das características originais, oferecendo uma abordagem simples e robusta de fusão de diferentes regressores?

3. De que maneira o uso de um metarregressor baseado em *k-Nearest Neighbors* pode explorar relações de similaridade entre amostras e contribuir para aprimorar a predição de produtividade?
4. A integração de dados provenientes de múltiplas fontes — incluindo imagens de sensoriamento remoto, índices vegetativos, dados agronômicos e meteorológicos — pode produzir modelos preditivos mais robustos e generalizáveis?
5. Características visuais extraídas automaticamente por redes neurais profundas pré-treinadas, como *Convolutional Neural Networks* (CNN) e *Vision Transformers* (ViT), podem complementar as características tradicionais utilizadas na estimativa da produtividade agrícola?
6. Métodos de redução de dimensionalidade aplicados às características visuais extraídas de imagens contribuem para melhorar o desempenho dos modelos preditivos?
7. A combinação de diferentes modalidades de dados (visuais, meteorológicas e agronômicas) permite capturar relações espaciais e temporais relevantes para a estimativa antecipada da produtividade da cana-de-açúcar?
8. A criação de novos conjuntos de dados provenientes de áreas comerciais de cultivo pode contribuir para avaliar a generalização e a aplicabilidade prática dos modelos propostos?

1.4 Hipótese

Em função da escassez de dados suficientes e da abrangência e diversidade dos dados para o estudo da predição da produtividade de cana-de-açúcar, foram exploradas diferentes abordagens guiadas por dados. A principal hipótese desse trabalho é definida como:

“Abordagens guiadas por dados podem melhorar o desempenho da predição da produtividade de cana-de-açúcar por meio: (i) da combinação de diferentes regressores em uma estrutura de *stacking ensemble*, na qual MFs geradas a partir de regressões aplicadas às características originais são utilizadas como entrada para metarregressores baseados em AdaBoost e *k-Nearest Neighbors*; (ii) da incorporação de características visuais extraídas de imagens de sensoriamento remoto utilizando rede neural profunda e mecanismo de atenção (CNN e ViT); e (iii) da construção e disponibilização de um novo conjunto de dados multimodal proveniente de campos comerciais, que permite avaliar a generalização dos modelos propostos.”

A validação da hipótese é realizada por meio de experimentos conduzidos nas diferentes abordagens mencionadas nesta tese.

1.5 Objetivos

Com base nas perguntas de pesquisa formuladas na Seção 1.3 e na hipótese apresentada na Seção 1.4, esta seção delinea os objetivos gerais e específicos que orientaram o desenvolvimento experimental da tese.

Objetivo Geral

Desenvolver e avaliar modelos de AM, guiados por diferentes abordagens de dados, para a predição da produtividade da cana-de-açúcar, integrando dados de múltiplas modalidades — incluindo IVs, dados LiDAR, características visuais de imagens, variáveis meteorológicas e informações agronômicas — e investigando estratégias de geração de MFs e de combinação de regressores por meio de *stacking ensemble* (MIENYE; SUN, 2022). Assim, desenvolver um modelo de AM baseado em diferentes abordagens de dados para estimar a produtividade esperada em talhões comerciais de cana-de-açúcar assume grande importância para a gestão e otimização das tomadas de decisão no setor sucroalcooleiro do Brasil.

Objetivos Específicos

Os objetivos específicos desta tese estão organizados em capítulos experimentais que compõem o trabalho.

Capítulo 5: Estudo Comparativo entre Regressores

- Realizar uma avaliação sistemática e comparativa de diferentes algoritmos de regressão — incluindo *Support Vector Regression* (SVR), *Random Forest* (RF), *Gradient Boosting* (GB), *AdaBoost* (AB), Regressão Linear e *Multi-Layer Perceptron* (MLP) — aplicados à predição da produtividade da cana-de-açúcar, utilizando o mesmo conjunto de dados e o mesmo protocolo experimental descritos em (SHENDRYK *et al.*, 2020).
- Identificar quais regressores apresentam melhor desempenho em diferentes configurações de *features* (média de NDVI, máximo de NDVI), componentes principais multiespectrais, componentes principais de LiDAR e a fusão destes).

- Comparar os resultados obtidos com os diferentes regressores aos originalmente reportados em (SHENDRYK *et al.*, 2020), que utilizaram exclusivamente regressão linear.

Capítulo 6: Modelo de *Meta-features* para Combinação de Diferentes Regressores

- Propor e avaliar uma estratégia de geração de MFs baseada em *stacking ensemble*, na qual cinco configurações de *features* originais são submetidas a cinco regressores base distintos (SVR, RF, GB, AB e Regressão Linear), gerando 25 vetores de MFs.
- Avaliar o desempenho preditivo do metarregressor AdaBoost quando aplicado a cada um dos 30 vetores de entrada disponíveis (5 vetores de *features* originais e 25 vetores de MFs).
- Investigar se as MFs geradas a partir de diferentes combinações de *features* e regressores capturam informações preditivas complementares às *features* originais.

Capítulo 7: Modelo de *Meta-features* Baseado no Regressor kNN

- Estender a abordagem de MFs incorporando o regressor *k-Nearest Neighbors* (kNN) como gerador adicional de MFs, produzindo cinco novos vetores de MFs (MF kNN M1 a MF kNN M5).
- Avaliar o desempenho do metarregressor AdaBoost sobre as MFs geradas pelo kNN.
- Investigar o uso do próprio kNN como metarregressor, aplicado tanto às *features* originais quanto às MFs geradas pelos diferentes regressores.
- Analisar a capacidade do kNN de explorar relações de similaridade entre amostras no espaço de MFs, por meio de visualizações com UMAP.
- Comparar o desempenho das configurações agregadas (30 MFs completas e fusão de *features* originais com MFs) com e sem redução de dimensionalidade (PCA e UMAP).

Capítulo 8: Estudo Comparativo de Características Visuais e Redução de Dimensionalidade

- Extrair características visuais (*visual features*) de imagens de sensoriamento remoto utilizando redes neurais profundas pré-treinadas: ResNet-101 (CNN) e DINOv2 (ViT).
- Avaliar diferentes estratégias de extração, incluindo imagens completas de parcelas e *patches* de 80×80 pixels (para ResNet-101).

- Aplicar técnicas de redução de dimensionalidade (PCA e UMAP) às características visuais extraídas, com o objetivo de selecionar os preditores mais informativos.
- Combinar as características visuais com IVs e dados LiDAR, avaliando se essa fusão multimodal proporciona ganhos de desempenho em relação ao uso isolado de cada modalidade.
- Comparar os resultados obtidos com base em características visuais com os reportados nos capítulos anteriores, que utilizaram exclusivamente IVs e dados de LiDAR.

Capítulo 9: Um Conjunto de Dados de Cana-de-açúcar de Campos Comerciais no Estado de São Paulo

- Construir e disponibilizar publicamente um novo conjunto de dados de cana-de-açúcar proveniente de campos comerciais no nordeste do estado de São Paulo, integrando três modalidades de informação: (i) dados do ambiente de produção (variedade, solo, datas de plantio e colheita, produtividade observada); (ii) registros meteorológicos diários (precipitação, temperatura, evapotranspiração, umidade, radiação solar, velocidade do vento); e (iii) imagens de satélite Sentinel-2 (bandas R, G, B) organizadas em séries temporais.
- Organizar o *dataset* em um repositório de acesso aberto, com estrutura documentada e arquivo de metadados, para viabilizar sua reutilização por outros pesquisadores.
- Fornecer uma base de dados robusta (2.157 observações) que sustente a aplicação de algoritmos de AM em estudos futuros de predição da produtividade da cana-de-açúcar.

Diante desse cenário, esta tese investiga uma abordagem de expansão de características baseada em *stacking ensemble*, na qual MFs são geradas a partir das saídas de diferentes regressores aplicados aos dados originais. Diferentemente de formulações que combinam *features* originais e MFs em um único vetor de entrada, esta tese adota uma estratégia comparativa, avaliando separadamente a contribuição preditiva das *features* originais e das MFs quando utilizadas como entrada para o metarregressor. Essa estratégia permite analisar com mais precisão o ganho de informação proporcionado pelas representações de segundo nível, mantendo a flexibilidade e a robustez em contextos de dados limitados.

1.6 Contribuições

Esta tese propõe e avalia uma estratégia de geração de MFs, por meio de *stacking*, a partir das saídas de diferentes modelos de regressão previamente treinados. O processo é

estruturado em duas etapas. Na primeira etapa, cinco configurações de *features* originais são definidas a partir dos dados de sensoriamento remoto: (i) NDVI máximo; (ii) NDVI médio; (iii) componentes principais multiespectrais; (iv) componentes principais de LiDAR; (v) fusão dos componentes principais multiespectrais e de LiDAR. Cada uma dessas cinco configurações é submetida a cinco regressores base distintos (SVR, RF, GB, AB e Regressão Linear), o que resulta em 25 vetores de MFs. Na segunda etapa, um metarregressor é aplicado a cada um dos 30 vetores de entrada disponíveis: cinco vetores de *features* originais e 25 vetores de MFs. Essa estratégia permite avaliar, de forma sistemática, a contribuição preditiva de cada representação individualmente, bem como investigar se as MFs geradas a partir de diferentes combinações de *features* e regressores capturam informações complementares às *features* originais.

A tese também propõe e avalia uma abordagem de integração de dados provenientes de múltiplas fontes, incluindo o uso de sensores remotos de RPAs associados, IVs, características extraídas dos pixels (no trabalho desenvolvido em (BARBOSA *et al.*, 2025a)), LiDAR e dados agronômicos. Essa integração multimodal permite explorar as complementaridades entre as resoluções espacial, temporal e espectral, resultando em estimativas de produtividade mais robustas e informativas.

Do ponto de vista aplicado, a predição da produtividade da cana-de-açúcar por meio de algoritmos de AM, a partir de séries temporais de imagens orbitais (satélites) ou suborbitais (RPAs), aliada a dados do ambiente de produção, permite que canavicultores e usineiros disponham de uma plataforma que agregue dados relevantes para a tomada de decisão na gestão das safras atuais e futuras.

Os experimentos foram conduzidos com o conjunto de dados disponibilizado por (SHENDRYK *et al.*, 2020), que foi avaliado em regressões em (BARBOSA; PEDRONETTE; GUILHERME, 2023a) e, posteriormente, o mesmo conjunto de dados foi avaliado em (BARBOSA; PEDRONETTE; GUILHERME, 2023b), por meio de MFs extraídas dos resultados obtidos com os diferentes regressores utilizados em (BARBOSA; PEDRONETTE; GUILHERME, 2023a), a fim de obter melhores resultados. Nos trabalhos (BARBOSA; PEDRONETTE; GUILHERME, 2023a; BARBOSA; PEDRONETTE; GUILHERME, 2023b; BARBOSA *et al.*, 2025b), abordados nos Capítulos 5, 6 e 7, foram realizadas melhorias progressivas no desempenho das abordagens. Também em (BARBOSA *et al.*, 2025a), abordado no Capítulo 8, os dados utilizados nos capítulos anteriores foram combinados com outros obtidos por meio da extração de características de imagens por CNN e ViT pré-treinados, com e sem fusão, o que permitiu avaliar cenários em que é possível melhorar o desempenho na predição de produtividade. Por fim, a construção de um novo conjunto de dados, descrita em (BARBOSA *et al.*, 2026) e no Capítulo 9, disponibiliza à comunidade uma base multimodal com 2.157 observações de campos comerciais de cana-de-açúcar no estado de São Paulo, preenchendo uma lacuna de dados integrados para estudos de

predição de produtividade.

Do ponto de vista da Ciência da Computação, esta tese investiga o problema da fusão de dados multimodais para predição em cenários caracterizados por um pequeno número de observações, alta dimensionalidade e dependência espaço-temporal entre as amostras. Essa abordagem permite integrar diferentes modalidades de dados, sendo particularmente adequada para cenários com conjuntos de dados pequenos e heterogêneos, como os encontrados em aplicações de sensoriamento remoto agrícola. Nesse sentido, os dados agrícolas utilizados nesta tese são empregados como estudo de caso para investigar um problema mais geral de integração de dados multimodais no AM aplicado a dados espaço-temporais.

A Figura 3 ilustra as contribuições desta tese e como se relacionam entre si.

Visão Geral das Contribuições da Tese

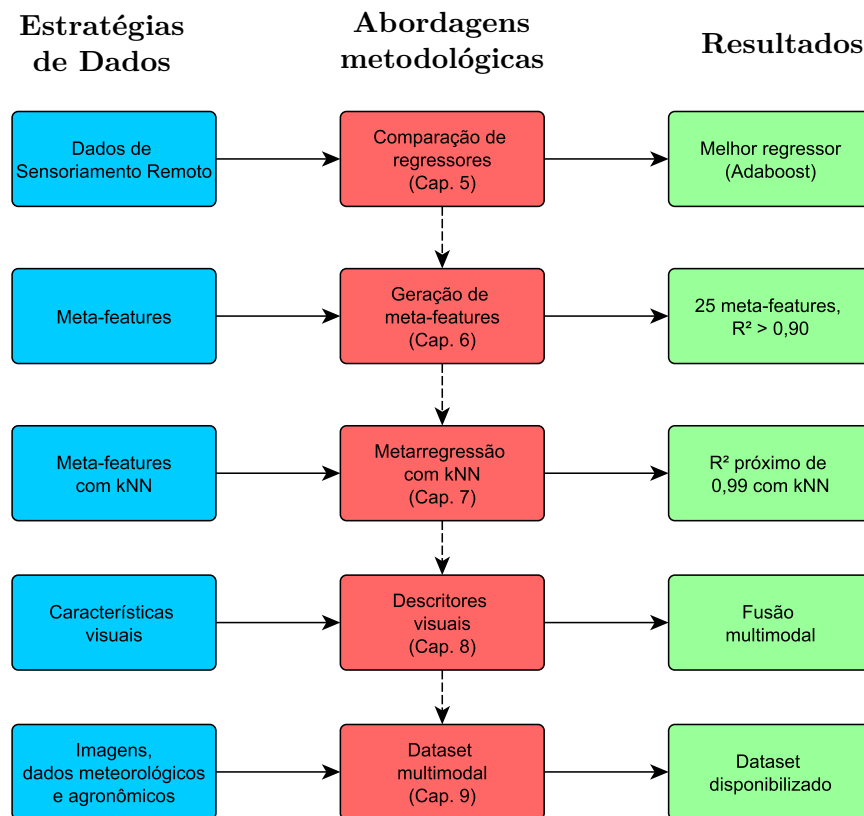


Figura 3 – Visão geral estruturada das contribuições da tese, organizada em três eixos principais: (i) fontes de dados e estratégias de extração de características; (ii) abordagens metodológicas desenvolvidas e avaliadas ao longo dos capítulos experimentais; (iii) principais resultados e produtos gerados. As setas tracejadas entre os blocos metodológicos indicam a evolução incremental das abordagens.

1.7 Organização

O restante da tese é organizado como se segue:

- o Capítulo 2 trata das técnicas empregadas na obtenção de estimativa de produtividade e dos tipos de dados utilizados para obtê-la;
- o Capítulo 3 apresenta os trabalhos realizados atualmente para se estimar a produtividade de culturas agrícolas;
- o Capítulo 4 apresenta o protocolo experimental utilizado nesta tese;
- o Capítulo 5 trata da comparação entre diferentes regressores, como forma de explorar melhor a capacidade de prever produtividade com o conjunto de dados utilizado ao longo dos experimentos (BARBOSA; PEDRONETTE; GUILHERME, 2023a);
- o Capítulo 6 discute a possibilidade de uso de resultados de predições como novas *features* (MFs) com o objetivo de refinar ainda mais o desempenho da predição de produtividade (BARBOSA; PEDRONETTE; GUILHERME, 2023b);
- o Capítulo 7 incorpora o kNN como gerador adicional de MFs e como metarregressor alternativo ao AdaBoost, investigando relações de similaridade no espaço de MFs (BARBOSA *et al.*, 2025b);
- o Capítulo 8 compara a combinação do uso de *features* utilizadas em (SHENDRYK *et al.*, 2020) com *features* extraídas de imagens, por meio de CNN e ViT pré-treinados (BARBOSA *et al.*, 2025a);
- o Capítulo 9 trata da criação de um novo *dataset* realizada no âmbito desta tese a partir de dados agronômicos, meteorológicos e de imagens de áreas de cultivo de cana-de-açúcar no estado de São Paulo (BARBOSA *et al.*, 2026);
- o Capítulo 10 discute resultados e contribuições, menciona publicações e colaboração internacional e avalia trabalhos futuros.

A Figura 4 apresenta como os dados utilizados, as contribuições, as publicações e os capítulos se relacionam ao longo da tese.

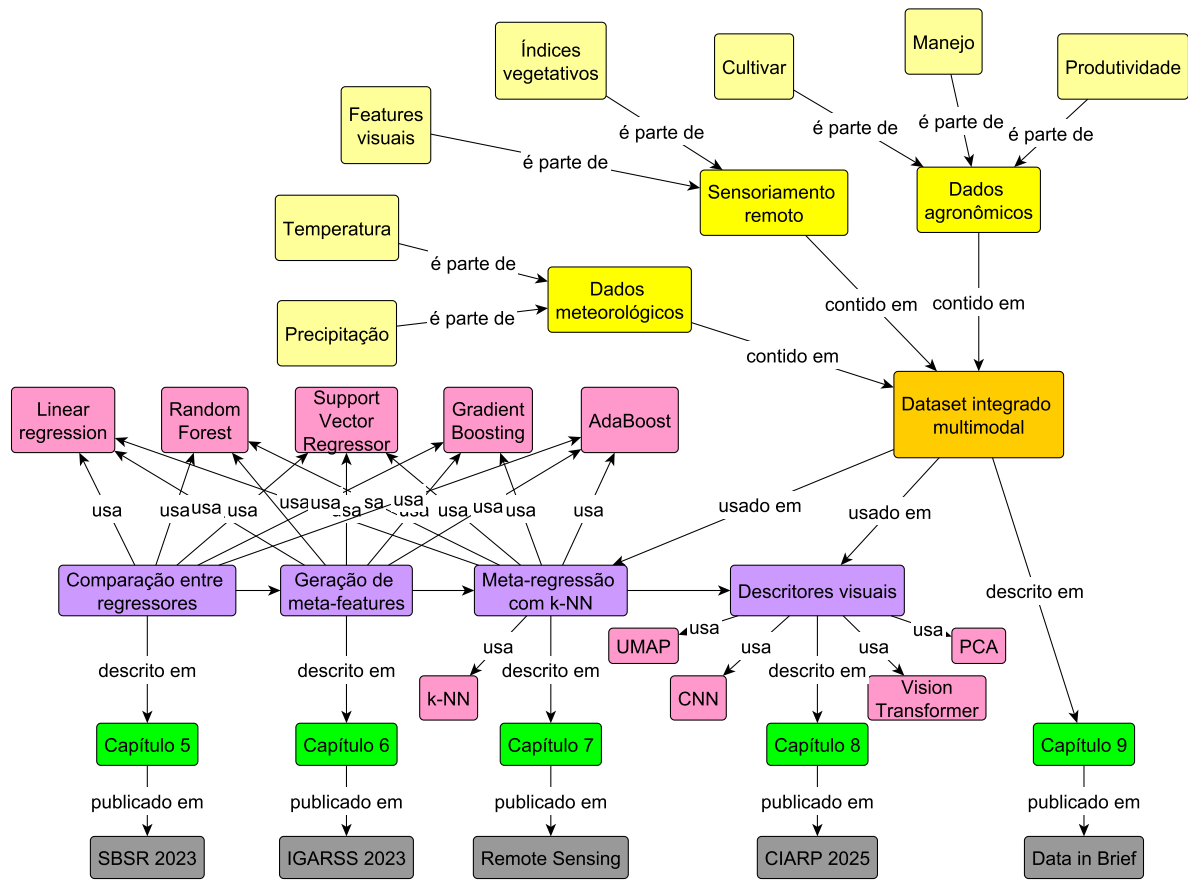


Figura 4 – Visão geral da estrutura conceitual da tese. O diagrama ilustra a relação entre os dados utilizados (laranja e tons de amarelo), as contribuições (roxo) desenvolvidas ao longo do trabalho, os algoritmos utilizados nelas (lilás), os capítulos (verde) da tese em que essas contribuições são apresentadas, bem como os artigos científicos (cinza) associados a cada etapa da pesquisa.

10 Conclusões

10.1 Contexto da Hipótese e Desafios

Esta tese foi motivada por uma questão central de caráter prático e científico: é possível estimar, com precisão e antecedência, a produtividade da cana-de-açúcar por meio de abordagens guiadas por dados, mesmo em cenários com número reduzido de observações e alta dimensionalidade das variáveis de entrada?

Para responder a essa questão, foi formulada a seguinte hipótese central, apresentada na Seção 1.4:

“Abordagens guiadas por dados podem melhorar o desempenho da predição da produtividade de cana-de-açúcar por meio: (i) da combinação de diferentes regressores em uma estrutura de *stacking ensemble*, na qual MFs geradas a partir de regressões aplicadas às características originais são utilizadas como entrada para metaregressores baseados em AdaBoost e *k-Nearest Neighbors*; (ii) da incorporação de características visuais extraídas de imagens de sensoriamento remoto utilizando redes neurais profundas e mecanismos de atenção (CNN e ViT); e (iii) da construção e disponibilização de um novo conjunto de dados multimodal proveniente de campos comerciais, que permite avaliar a generalização dos modelos propostos.”

Essa hipótese desdobrou-se em oito perguntas de pesquisa específicas (Seção 1.3), cujas respostas foram investigadas ao longo dos Capítulos 5 a 9. A Tabela 24 sintetiza essas perguntas e mapeia cada uma ao capítulo em que foi investigada, oferecendo uma visão panorâmica do programa de pesquisa e facilitando a navegação pelas respostas apresentadas nas seções seguintes.

10.2 Síntese dos Resultados e Discussão

Conforme identificado na Seção 1.2, três limitações principais caracterizavam o estado da arte na predição da produtividade da cana-de-açúcar:

- (i) a ausência de estudos comparativos sistemáticos entre diferentes regressores;
- (ii) a falta de exploração de representações de dados mais ricas, como MFs derivadas de múltiplos regressores; e

Tabela 24 – Perguntas de pesquisa e respectivos capítulos.

Nº	Pergunta de pesquisa	Capítulo
P1	Como diferentes algoritmos de regressão se comportam na predição de produtividade quando avaliados no mesmo conjunto de dados e sob o mesmo protocolo experimental?	5
P2	A geração de MFs a partir das saídas de modelos previamente treinados pode melhorar o desempenho preditivo em comparação ao uso apenas das características originais?	6
P3	De que maneira o uso de um metarregressor baseado em kNN pode explorar relações de similaridade entre amostras e contribuir para aprimorar a predição?	7
P4	A integração de dados de múltiplas fontes pode produzir modelos de predição mais robustos e generalizáveis?	8
P5	Características visuais extraídas por redes profundas pré-treinadas (CNN e ViT) podem complementar as características tradicionais na estimativa de produtividade?	8
P6	Métodos de redução de dimensionalidade aplicados às características visuais contribuem para melhorar o desempenho dos modelos?	8
P7	A combinação de diferentes modalidades de dados permite capturar relações espaciais e temporais relevantes para a estimativa antecipada da produtividade?	6–9
P8	A criação de novos conjuntos de dados de áreas comerciais contribui para avaliar a generalização e a aplicabilidade prática dos modelos propostos?	9

(iii) a integração limitada de modalidades heterogêneas de dados em um único arcabouço preditivo.

Os experimentos conduzidos nesta tese foram organizados em torno de três eixos temáticos progressivos, cada um endereçando diretamente uma das limitações: o Eixo I responde à limitação (i), o Eixo II à limitação (ii) e o Eixo III à limitação (iii).

A discussão a seguir segue essa estrutura para revelar a lógica científica subjacente ao programa de pesquisa.

Eixo I — Comparação e Seleção de Regressores

O ponto de partida desta tese baseou-se no trabalho de referência adotado ([SHENDRYK *et al.*, 2020](#)) que utilizava exclusivamente regressão linear (OLS) para estimar a produtividade a partir de IVs e de dados LiDAR obtidos por RPA. Embora esse trabalho tenha estabelecido um protocolo experimental rigoroso e um conjunto de dados valioso, a escolha de um único regressor deixou em aberto a questão de saber se outros algoritmos poderiam captar com maior fidelidade as relações não lineares entre as variáveis espectrais e a produtividade da cultura.

Para investigar essa questão, o Capítulo 5 conduziu uma avaliação sistemática de cinco regressores adicionais — SVR, *Random Forest*, *Gradient Boosting*, AdaBoost e MLP — sob o mesmo protocolo e o mesmo conjunto de dados de (SHENDRYK *et al.*, 2020). Os resultados (Tabela 11) demonstraram que o AdaBoost obteve os melhores valores médios de $\overline{R}^2$ e o menor valor de RMSE entre todos os regressores avaliados.

Essa evidência confirmou a pergunta P1 e estabeleceu uma conclusão metodológica relevante: a seleção do algoritmo de regressão é uma decisão de projeto com impacto mensurável no desempenho preditivo, e estudos futuros não deveriam restringir-se a um único modelo sem avaliação comparativa. Adicionalmente, o fato de o AdaBoost — um método de *ensemble* baseado em árvores de decisão fracas — ter superado regressores mais complexos, como *Random Forest* e *Gradient Boosting*, sugere que, em regimes de pequena amostra, a capacidade de adaptação iterativa ao erro residual pode ser mais vantajosa do que a diversidade de árvores base.

Eixo II — Enriquecimento de Representações via *Meta-features*

A segunda linha de investigação partiu de uma pergunta estruturalmente distinta: em vez de buscar o melhor regressor isolado, seria possível combinar as informações produzidas por múltiplos regressores para gerar representações de segundo nível — as MFs — que capturassem aspectos preditivos complementares às *features* originais?

O Capítulo 6 formalizou essa ideia por meio de uma estrutura de *stacking ensemble*: cinco regressores base (SVR, RF, GB, AB e OLS) foram aplicados a cada uma das cinco configurações de *features* originais (M1 a M5), gerando 25 vetores de MFs adicionais. O metarregressor AdaBoost, avaliado sobre esses vetores, produziu resultados superiores às *features* originais para os modelos M1 (NDVI médio) e M5 (MultispectralPCs + LiDARPCs). Esse resultado confirmou a pergunta P2: as representações de segundo nível capturam informações preditivas não presentes nas *features* originais, sem exigir a coleta de novos dados de campo.

O Capítulo 7 estendeu essa abordagem ao incorporar o kNN como gerador adicional de MFs e, em seguida, como metarregressor. A aplicação do kNN como metarregressor no espaço de MFs produziu os melhores resultados entre todas as configurações avaliadas, com $\overline{R}^2$ de até 0,9946, confirmando a pergunta P3. Visualizações por UMAP revelaram agrupamentos coerentes das amostras no espaço de MFs, fornecendo evidência qualitativa de que o kNN é capaz de explorar relações de proximidade entre parcelas com características espectrais e agronômicas semelhantes — um tipo de estrutura latente que regressores globais tendem a ignorar. Adicionalmente, a avaliação por *expanding window cross-validation*, que trata as 336 observações como uma série temporal contínua, não superou os resultados obtidos pela validação por data isolada, indicando que a estrutura temporal não agrega informação preditiva adicional à já capturada pelas MFs quando avaliadas por data.

Do ponto de vista prático, o resultado mais importante desse eixo é que a abordagem de MFs permitiu melhorar substancialmente o desempenho preditivo sem qualquer custo adicional de aquisição de dados. Isso tem implicações diretas para a escalabilidade da metodologia, especialmente em contextos em que a coleta de dados de campo é custosa e a logística é complexa.

Eixo III — Integração Multimodal e Características Visuais

O terceiro eixo investigou se as informações contidas nas imagens de sensoriamento remoto — além dos IVs derivados dessas imagens — poderiam enriquecer os modelos preditivos. Essa é uma pergunta relevante porque os IVs são, por construção, sumários escalares das bandas espectrais.

O Capítulo 8 investigou a extração de VFs por meio de duas arquiteturas profundas pré-treinadas: ResNet-101, da família de CNNs, e DINOv2, um ViT treinado por autossupervisão. Foram avaliadas estratégias de extração a partir da imagem completa da parcela e de *patches* de 80×80 pixels (para a ResNet-101), combinadas com técnicas de redução de dimensionalidade (PCA e UMAP). A combinação de VFs com os IVs e dados LiDAR resultou em melhorias nas configurações de maior riqueza espectral e estrutural, respondendo afirmativamente às perguntas P4, P5, P6 e P7.

Contudo, os ganhos foram mais modestos do que os obtidos com as abordagens de MFs. Essa assimetria tem uma explicação metodológica clara: redes neurais profundas pré-treinadas em grandes *datasets* de imagens de objetos (como o ImageNet) podem não transferir de forma ótima para imagens agrícolas multiespectrais de talhões de cana-de-açúcar, cujas características visuais diferem substancialmente das de objetos. Além disso, o número reduzido de observações no conjunto de dados original (SHENDRYK *et al.*, 2020) limita a capacidade do modelo de generalizar com base nas representações visuais aprendidas. Esses dois fatores, combinados, explicam por que as VFs *per se*, embora complementares, não superaram, de forma consistente, as MFs na métrica final de R^2 .

Essa análise não diminui a importância da contribuição: pelo contrário, delimita com precisão o regime de aplicabilidade de cada abordagem. Em cenários com *datasets* maiores — como o que foi construído no Capítulo 9 — é razoável esperar que as VFs apresentem maior vantagem competitiva sobre as MFs, pois as redes profundas têm capacidade de aprendizado que escala com o volume de dados.

Conjunto de Dados Aberto à Comunidade Científica

A ausência de conjuntos de dados abertos sobre a produtividade da cana-de-açúcar provenientes de campos comerciais brasileiros constitui um dos principais obstáculos ao avanço de métodos de aprendizado de máquina aplicados ao setor sucroalcooleiro nacional.

Sem dados representativos das condições operacionais reais — que envolvem diversidade de cultivares, manejos, solos e climas —, os modelos desenvolvidos pela comunidade científica permanecem restritos a *datasets* experimentais de pequena escala, limitando tanto a sua capacidade de generalização quanto a reprodutibilidade dos resultados. Essa carência tornou-se particularmente evidente ao longo desta tese: os resultados dos Capítulos 5 a 8, embora promissores, foram obtidos sobre um *dataset* experimental australiano com apenas 336 observações, e abordagens como a extração de características visuais (Capítulo 8) evidenciaram a necessidade de volumes substancialmente maiores de dados para que as técnicas de aprendizado profundo atinjam seu potencial. O Capítulo 9 responde diretamente a essa lacuna ao construir e disponibilizar publicamente o primeiro *dataset* de produtividade de cana-de-açúcar de campos comerciais no nordeste do estado de São Paulo, integrando imagens do satélite Sentinel-2, variáveis meteorológicas de alta resolução espacial (XAVIER *et al.*, 2022a), dados de insumos agrícolas e registros do ambiente de produção.

O *dataset*, publicado na revista *Data in Brief* (BARBOSA *et al.*, 2026), contém 2.157 observações distribuídas em 12 talhões de 6 fazendas no nordeste do estado de São Paulo. A amplitude das modalidades integradas — de imagens multitemporais com filtragem de cobertura de nuvens até a classificação do ambiente de produção pelo sistema CTC — confere ao conjunto de dados uma riqueza que vai além do que seria necessário para os experimentos desta tese, constituindo uma base para investigações futuras de maior escopo.

Em suma, o Capítulo 9 responde afirmativamente à pergunta P8 e, ao mesmo tempo, cria as condições para que todas as abordagens desenvolvidas nos Capítulos 5 a 8 sejam validadas em um contexto de maior escala e diversidade agronômica — o que representa o próximo passo natural desta linha de pesquisa.

10.3 Respostas às Perguntas de Pesquisa

A Tabela 25 consolida, de forma objetiva, as respostas obtidas para cada uma das oito perguntas de pesquisa formuladas na Seção 1.3, indicando o capítulo experimental correspondente, o principal resultado quantitativo ou qualitativo e a publicação associada.

Tabela 25 – Respostas às perguntas de pesquisa.

Nº	Pergunta (síntese)	Resposta	Cap.	Pub.
P1	Diferentes regressores produzem resultados distintos sob o mesmo protocolo?	Sim. AdaBoost apresentou desempenho superior ao OLS de referência em todos os modelos preditores.	5	(BARBOSA; PEDRONETTE; GUILHERME, 2023a)
P2	<i>Meta-features</i> melhoraram o desempenho em relação às <i>features</i> originais?	Sim. MF Grad Boost M1 e MF AdaBoost M5 superaram as <i>features</i> originais no metarregressor AdaBoost.	6	(BARBOSA; PEDRONETTE; GUILHERME, 2023b)
P3	O kNN como metarregressor explora similaridade entre amostras com ganho?	Sim. kNN, tanto como gerador de MFs adicionais quanto como metarregressor, produziu os melhores resultados ($\bar{R}^2$ de até 0,9946).	7	(BARBOSA <i>et al.</i> , 2025b)
P4	Integração de múltiplas fontes produzem modelos mais robustos?	Sim, com ressalva: os ganhos foram observados quando as VFs foram combinadas com IVs ou dados de LiDAR, e não quando utilizadas isoladamente.	8	(BARBOSA <i>et al.</i> , 2025a; BARBOSA; PEDRONETTE; GUILHERME, 2023b; BARBOSA <i>et al.</i> , 2025b)
P5	VFs por CNN e ViT complementam características tradicionais?	Parcialmente. Melhorias observadas em configurações de maior riqueza espectral; impacto limitado devido ao pequeno <i>dataset</i> .	8	(BARBOSA <i>et al.</i> , 2025a)
P6	Redução de dimensionalidade nas VFs melhora o desempenho?	Sim. PCA mostrou desempenho robusto.	8	(BARBOSA <i>et al.</i> , 2025a)
P7	Combinação de modalidades captura relações espaciais e temporais relevantes?	Sim. A fusão multimodal (espectral + LiDAR + visual) produziu as melhores configurações em cenários de alta dimensionalidade.	6–9	(BARBOSA <i>et al.</i> , 2025a; BARBOSA <i>et al.</i> , 2025b; BARBOSA <i>et al.</i> , 2026)
P8	Novos <i>datasets</i> de campos comerciais contribuem para a generalização?	Sim. O <i>dataset</i> de (BARBOSA <i>et al.</i> , 2026) disponibiliza 2.157 observações de 12 talhões comerciais, suprimindo uma lacuna reconhecida na literatura.	9	(BARBOSA <i>et al.</i> , 2026)

Destaca-se que nenhuma das perguntas de pesquisa ficou sem resposta — fato que, por si só, evidencia a completude do programa experimental conduzido. As respostas às perguntas P4 e P5 incluem ressalvas importantes, que são discutidas com maior profundidade na Seção 10.5. Essas ressalvas não invalidam as contribuições; ao contrário, delimitam com precisão o escopo de aplicabilidade de cada abordagem e abrem caminho para investigações futuras detalhadas na Seção 10.6.

10.4 Contribuições Científicas

Com base nos resultados sintetizados nas seções anteriores, as contribuições desta tese podem ser organizadas em duas categorias: contribuições metodológicas, que propõem novas abordagens ou estendem abordagens existentes; e contribuições de suporte à pesquisa, que produzem artefatos reutilizáveis pela comunidade.

Contribuições Metodológicas

Análise comparativa sistemática de algoritmos de regressão para a predição da produtividade agrícola: Foi realizada uma avaliação comparativa de múltiplos regressores de AM — SVR, RF, GB, AdaBoost e MLP — aplicados à predição da produtividade da cana-de-açúcar, com base em IVs e dados LiDAR obtidos por RPA, sob um protocolo experimental controlado. Essa contribuição preenche uma lacuna apontada em revisões recentes da área (van Klompenburg; KASSAHUN; CATAL, 2020; SHAWON *et al.*, 2025), que destacam a ausência de estudos comparativos sistemáticos entre regressores.

Proposição de um modelo de MFs por *stacking ensemble* para dados agrícolas de pequeno porte: A tese propõe e valida uma estratégia de geração de MFs baseada em *stacking*, na qual as saídas de múltiplos regressores base são utilizadas como entradas para um metarregressor, sendo bastante inovadora em estudos de predição da produtividade de culturas agrícolas e inédita para a cultura da cana-de-açúcar. A abordagem demonstrou ganhos de desempenho sem exigir a coleta adicional de dados, sendo especialmente adequada ao regime de pequena amostra, característico dos experimentos agrícolas de campo.

Desenvolvimento de uma abordagem de metarregressão com kNN explorando relações de similaridade: O uso do kNN como metarregressor no espaço de MFs representa uma contribuição original que explora a estrutura de vizinhança entre talhões com características espectrais e agronômicas semelhantes. A análise visual por UMAP proporcionou interpretabilidade adicional à abordagem, evidenciando a coerência dos agrupamentos no espaço de representação aprendido.

Incorporação de características visuais profundas para a estimativa de produtividade: A investigação do uso de CNNs (ResNet-101) e ViTs (DINOv2) para a

extração de características visuais a partir de imagens de sensoriamento remoto, combinada com técnicas de redução de dimensionalidade, amplia o escopo das representações utilizadas em modelos de predição agrícola, sendo esta a maior contribuição deste trabalho e a de maior potencial de inovação a ser trabalhada futuramente pelo autor. O estudo identifica as condições em que VFs trazem ganhos e as em que o tamanho amostral limita sua eficácia.

Contribuições de Suporte à Pesquisa

Construção e disponibilização de um conjunto de dados multimodal de campos comerciais de cana-de-açúcar: O *dataset* publicado em (BARBOSA *et al.*, 2026) integra imagens multitemporais do Sentinel-2, variáveis meteorológicas diárias, dados de insumos agrícolas e registros de ambiente de produção de doze talhões comerciais no nordeste do estado de São Paulo. Com 2.157 observações e documentação detalhada nos metadados, constitui o primeiro conjunto de dados aberto sobre a produtividade da cana-de-açúcar proveniente de fazendas comerciais naquela região que combina essas modalidades.

Produção científica indexada associada às contribuições da tese: As contribuições metodológicas e de suporte à pesquisa resultaram em cinco publicações em conferências e periódicos internacionais de alto impacto nas áreas de sensoriamento remoto, aprendizado de máquina e visão computacional, detalhadas na Seção 10.7.

10.5 Limitações do Trabalho

A apresentação honesta das limitações de uma tese é um componente essencial da maturidade científica e contribui para a correta interpretação dos resultados pela comunidade científica. As principais limitações desta tese são discutidas a seguir.

L1 — Tamanho e origem do conjunto de dados experimental (Capítulos 5–8): Os experimentos dos Capítulos 5 a 8 foram conduzidos com o *dataset* de Shendryk *et al.* (SHENDRYK *et al.*, 2020), que contém um número reduzido de observações coletadas em condições experimentais controladas com RPA, em uma única região e safra. Embora esse *dataset* seja o mais completo disponível com as modalidades necessárias para os experimentos propostos, seu tamanho limita a generalização direta dos resultados quantitativos (valores de R^2 e RMSE) para outros contextos agrícolas, regiões ou variedades de cana.

L2 — Ausência de validação empírica no *dataset* multimodal do Capítulo 9: O conjunto de dados de (BARBOSA *et al.*, 2026) foi construído e disponibilizado, mas experimentos de predição de produtividade utilizando esse *dataset* — que representa condições comerciais brasileiras — não foram conduzidos no escopo desta tese, uma vez que a construção do *dataset* foi concluída na etapa final do doutorado. Essa validação

constitui o próximo passo mais relevante da linha de pesquisa e é detalhada como trabalho futuro prioritário na Seção 10.6.

L3 — Dependência de imagens obtidas por RPA com LiDAR nos Capítulos 5–8: Os dados LiDAR utilizados nos experimentos dos Capítulos 5 a 8 foram obtidos com equipamentos de alto custo, o que limita a replicação desses experimentos em escala comercial. As abordagens de MFs e de VFs, por outro lado, são aplicáveis a imagens de satélite de livre acesso (como as do Sentinel-2), o que amplia seu potencial de escalabilidade.

L4 — Transferibilidade dos modelos de VFs pré-treinados: As redes ResNet-101 e DINOv2 utilizadas no Capítulo 8 foram pré-treinadas em *datasets* de imagens de objetos (ImageNet), cujas características visuais diferem das de imagens agrícolas multiespectrais de talhões de cana-de-açúcar. A ausência de *fine-tuning* específico para o domínio agrícola pode ter limitado o potencial das VFs, especialmente devido à falta de dados suficientes para adaptar os modelos ao novo domínio.

É importante ressaltar que as Limitações L1 e L2 são complementares e não circulares: a L1 motiva a construção do novo *dataset* (que resolve parcialmente os problemas de escala e de origem comercial), e a L2 indica que a validação nesse *dataset* é o próximo passo lógico. As limitações L3 e L4 apontam para direções metodológicas específicas a serem exploradas em trabalhos futuros.

10.6 Trabalhos Futuros

Com base nas contribuições desta tese e nas limitações identificadas na seção anterior, delineiam-se, a seguir, as investigações futuras mais relevantes, organizadas por dependência metodológica.

Validação em *Dataset* Multimodal

A validação das abordagens de MFs e de VFs desenvolvidas nos Capítulos 5 a 8, utilizando o *dataset* de (BARBOSA *et al.*, 2026), constitui a extensão mais direta e necessária desta tese. Esses experimentos permitirão avaliar: (i) se os ganhos obtidos pelas MFs se generalizam para um contexto de maior escala e diversidade agrônômica; (ii) se as VFs extraídas de imagens Sentinel-2 — em contraste com imagens de RPA — apresentam maior impacto quando o tamanho amostral é substancialmente maior; e (iii) se a combinação de modalidades (espectral, meteorológica, agrônômica) melhora a predição de forma consistente entre fazendas e safras distintas.

Avaliação Experimental da Arquitetura MMST-ViT

Uma direção promissora consiste na implementação e avaliação da arquitetura MMST-ViT (LIN *et al.*, 2023), um modelo baseado em atenção (*self-attention*) projetado especificamente para a predição de produtividade agrícola de grãos a partir da fusão de séries temporais de imagens de satélite com variáveis meteorológicas. Essa arquitetura é particularmente relevante para a continuidade desta tese porque opera com as mesmas modalidades de dados disponíveis no *dataset* do Capítulo 9: imagens multitemporais do Sentinel-2, variáveis meteorológicas diárias e registros de produtividade. A investigação permitirá avaliar o potencial de arquiteturas baseadas em mecanismos de atenção para culturas tropicais semiperenes.

Estudos de Ablação em Arquiteturas Multimodais

Uma vez estabelecido o protocolo experimental com o MMST-ViT, torna-se possível conduzir estudos de ablação sistemáticos para quantificar a contribuição individual de cada componente: (i) ablação de modalidade — comparação entre modelos treinados apenas com imagens, apenas com variáveis meteorológicas e com a fusão completa; (ii) ablação temporal — avaliação do impacto da janela de observação climática (curto vs. longo prazo); e (iii) ablação arquitetural — investigação de diferentes estratégias de fusão de modalidades e profundidades de rede.

Extensões Metodológicas de Médio Prazo

Entre as extensões metodológicas de médio prazo mais relevantes, destacam-se:

- o uso de pré-treinamento autossupervisionado específico para imagens agrícolas, visando reduzir a dependência de grandes volumes de dados rotulados e melhorar a transferibilidade dos modelos para novos contextos;
- a integração de dados LiDAR obtidos por satélite (e.g., GEDI (DUBAYAH *et al.*, 2020)), ampliando a acessibilidade da modalidade estrutural para além dos cenários experimentais com RPA de alta precisão;
- a investigação de técnicas de aprendizado por transferência entre culturas, aproveitando representações aprendidas em *datasets* de maior escala (soja, milho, trigo) para inicializar modelos de predição de cana-de-açúcar;
- a incorporação de variáveis explicativas adicionais disponíveis no *dataset* de (BARBOSA *et al.*, 2026), como o histórico detalhado de insumos agrícolas e a classificação do ambiente de produção, em modelos que permitam analisar a importância das *features*, a fim de subsidiar decisões de manejo. Bem como a utilização de técnicas

de explicabilidade mapeando o comportamento do modelo, a fim de identificar ao gestor o que ocorre e onde atuar para promover a melhoria do manejo, de modo que a produção aumente, sendo, possivelmente, mais importante do que o próprio resultado da predição.

10.7 Publicações e Colaboração Internacional

Publicações

As investigações conduzidas ao longo deste doutorado resultaram em cinco publicações em conferências e periódicos internacionais (Tabela 26).

Tabela 26 – Publicações produzidas durante o doutorado.

Referência	Título	Veículo	Ano
(BARBOSA; PEDRO-NETTE; GUILHERME, 2023a)	Estudo Comparativo entre Diferentes Regressores para Estimar a Produtividade da Cana-de-açúcar	SBSR	2023
(BARBOSA; PEDRO-NETTE; GUILHERME, 2023b)	<i>A meta-feature model for exploiting different regressors to estimate sugarcane crop yield</i>	IGARSS (IEEE)	2023
(BARBOSA et al., 2025b)	<i>Meta-Features Extracted from Use of kNN Regressor to Improve Sugarcane Crop Yield Prediction</i>	<i>Remote Sensing</i> (MDPI)	2025
(BARBOSA et al., 2025a)	<i>Comparative Study of Image Descriptors and Dimensionality Reduction for the Prediction of Sugarcane Crop Yield</i>	CIARP	2025
(BARBOSA et al., 2026)	<i>A dataset of sugarcane crop yield, production environment, meteorological records, and satellite images of commercial fields in the northeast of São Paulo State, Brazil</i>	<i>Data in Brief</i> (Elsevier)	2026

Colaboração Internacional

Durante o doutorado, o autor recebeu bolsa de estudos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por meio do Projeto Unesp de Internacionalização (CAPES-PrInt-Unesp), concessão nº 88887.899737/2023-00, com permanência de seis meses no *Institut Agro Montpellier* (de dezembro de 2023 a maio de 2024). A pesquisa foi conduzida sob supervisão do Prof. Dr. Bruno Tisseyre, da unidade ITAP (*Technologies and Methods for the Agriculture of Tomorrow*), Unidade de pesquisa conjunta entre o INRAe (*Institut National de Recherche pour l'Agriculture, l'alimentation et l'environnement*) e o *Institut Agro Montpellier*. Essa colaboração resultou diretamente nas publicações (BARBOSA *et al.*, 2025a; BARBOSA *et al.*, 2025b) e contribuiu de forma significativa para o desenvolvimento das abordagens de MFs com kNN e de descritores visuais apresentadas nos Capítulos 7 e 8.

10.8 Considerações Finais

Esta tese partiu de um problema prático de grande relevância econômica e estratégica para o Brasil — a estimativa antecipada da produtividade da cana-de-açúcar — e construiu um programa de pesquisa coerente que progressivamente ampliou a compreensão de como abordagens guiadas por dados podem contribuir para esse desafio.

O conjunto das contribuições aqui apresentadas demonstra que: (i) a escolha criteriosa do algoritmo de regressão já produz ganhos mensuráveis; (ii) a combinação inteligente de múltiplos regressores por meio de MFs amplifica esses ganhos sem custos adicionais de coleta de dados; (iii) descritores visuais profundos complementam as abordagens tradicionais, especialmente em cenários de maior volume de dados; e (iv) a disponibilização de um *dataset* multimodal de campos comerciais cria as condições para que toda essa linha de pesquisa avance em direção à validação operacional.

Do ponto de vista do campo científico, esta tese contribui para a consolidação de uma agenda de pesquisa que integra sensoriamento remoto, aprendizado de máquina e agricultura de precisão — três áreas que, em conjunto, têm potencial de transformar a gestão de culturas de grande escala. A cana-de-açúcar, como cultura semiperene de ciclos longos e de enorme importância no contexto energético e alimentar brasileiro, constitui um objeto de estudo que desafia os métodos estabelecidos e demanda soluções adaptadas às suas especificidades agronômicas e temporais.

Os trabalhos futuros delineados na Seção 10.6 não representam lacunas desta tese — representam a fronteira que ela ajudou a definir.

Referências

- ABBAS, F. *et al.* Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, v. 10, n. 7, 2020. ISSN 2073-4395. Available in: <<https://www.mdpi.com/2073-4395/10/7/1046>>.
- ABDULLAHI, M. *et al.* A systematic literature review of visual feature learning: deep learning techniques, applications, challenges and future directions. *Multimedia Tools and Applications*, Springer, v. 84, n. 19, p. 20439–20496, 2025.
- ADAMI, M. *et al.* Sugarcane yield estimation using satellite remote sensing data in empirical or mechanistic modeling: A systematic review. *Remote Sensing*, MDPI, v. 16, n. 5, p. 863, 2024.
- AFRAEI, S.; SHAHRIAR, K.; MADANI, S. H. Statistical assessment of rock burst potential and contributions of considered predictor variables in the task. *Tunnelling and Underground Space Technology*, Elsevier, v. 72, p. 250–271, 2018.
- AGHIGHI, H. *et al.* Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 oli. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 11, n. 12, p. 4563–4577, 2018.
- AHMAD, M. *et al.* A fast and compact 3-d cnn for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, p. 1–5, 2020.
- AKBARIAN, S. *et al.* An investigation on the best-fit models for sugarcane biomass estimation by linear mixed-effect modeling on unmanned aerial vehicle-based multispectral images: A case study of australia. *Information Processing in Agriculture*, 2022.
- ALEXANDRE, M. L. d. S. *et al.* Sugarcane (*saccharum officinarum*) productivity estimation using multispectral sensors in rpas, biometric variables, and vegetation indices. *Agronomy*, v. 15, n. 9, p. 2149, 2025.
- ALLEN, R. G. *et al.* Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, v. 300, n. 9, p. D05109, 1998.
- AMARASINGAM, N. *et al.* A review of uav platforms, sensors, and applications for monitoring of sugarcane crops. *Remote Sensing Applications: Society and Environment*, Elsevier, v. 26, p. 100712, 2022.
- AMID, E.; WARMUTH, M. K. Trimap: Large-scale dimensionality reduction using triplets. *arXiv preprint arXiv:1910.00204*, 2019.
- ANBANANTHEN, K. S. M. *et al.* An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms. *F1000Research*, Faculty of 1000 Ltd, v. 10, 2021.
- ANTHONY, D. *et al.* On crop height estimation with UAVs. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. [S.l.: s.n.], 2014. p. 4805–4812.

- ARAKAWA, K. *et al.* Sugar cane yield prediction using drone data processed by lstm algorithm. In: IEEE. *2025 3rd International Conference on Mechatronics, Control and Robotics (ICMCR)*. [S.l.], 2025. p. 75–79.
- ASLAN, M. F.; SABANCI, K.; ASLAN, B. Artificial intelligence techniques in crop yield estimation based on sentinel-2 data: A comprehensive survey. *Sustainability*, Mdpi, v. 16, n. 18, p. 8277, 2024.
- AWAD, M.; KHANNA, R. Support vector machines for classification. In: *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA: Apress, 2015. p. 39–66. Available in: https://doi.org/10.1007/978-1-4302-5990-9_3.
- BAKER, J. C.; O'CONNELL, K. M.; WILLIAMSON, e. R. *Commercial Observation Satellites: At the Leading Edge of Global Transparency*. Santa Monica, CA: RAND Corporation, 2001.
- BANSAL, Y.; LILLIS, D.; KECHADI, M.-T. A neural meta model for predicting winter wheat crop yield. *Machine Learning*, Springer, v. 113, n. 6, p. 3771–3788, 2024.
- BARBOSA, A. *et al.* Modeling yield response to crop management using convolutional neural networks. *Computers and Electronics in Agriculture*, v. 170, p. 105197, 2020. ISSN 0168-1699. Available in: <https://www.sciencedirect.com/science/article/pii/S0168169919308543>.
- BARBOSA, A. de M. *et al.* Energy conversion efficiency in sugarcane cultivars as a function of production environments in brazil. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 150, p. 111500, 2021.
- BARBOSA, L. A. F. *et al.* Comparative study of image descriptors and dimensionality reduction methods for the prediction of sugarcane crop yield. In: SPRINGER. *Iberoamerican Congress on Pattern Recognition*. [S.l.], 2025. p. 16–29.
- BARBOSA, L. A. F. *et al.* Meta-features extracted from use of knn regressor to improve sugarcane crop yield prediction. *Remote Sensing*, MDPI, v. 17, n. 11, p. 1846, 2025.
- BARBOSA, L. A. F. *et al.* *Sugarcane dataset northeast Sao Paulo state, Brazil (Revision d064a64)*. Hugging Face, 2025. Available in: https://huggingface.co/datasets/lafbarbosa/sugarcane_dataset_northeast_sao_paulo_state_brazil.
- BARBOSA, L. A. F. *et al.* A dataset of sugarcane crop yield, production environment, meteorological records, and satellite images of commercial fields in the northeast of são paulo state, brazil. *Data in Brief*, p. 112549, 2026. ISSN 2352-3409. Available in: <https://www.sciencedirect.com/science/article/pii/S2352340926001022>.
- BARBOSA, L. A. F.; PEDRONETTE, D. C. G.; GUILHERME, I. R. Estudo comparativo entre diferentes regressores para estimar produtividade de cana-de-açúcar. *SBSR*, 2023.
- BARBOSA, L. A. F.; PEDRONETTE, D. C. G.; GUILHERME, I. R. A meta-feature model for exploiting different regressors to estimate sugarcane crop yield. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. [S.l.: s.n.], 2023.

- BARNES, E. *et al.* Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In: *Proceedings of the fifth international conference on precision agriculture, Bloomington, MN, USA*. [S.l.: s.n.], 2000. v. 1619.
- BENDIG, J. *et al.* Combining uav-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 39, p. 79–87, 2015.
- BESSA, A. D. *Explaining and Identifying Data to Support Data-driven Analyses*. Tese (Doutorado) — New York University Tandon School of Engineering, 2020.
- BOCCA, F. F.; RODRIGUES, L. H. A. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and electronics in agriculture*, Elsevier, v. 128, p. 67–76, 2016.
- BOLETIM da Safra de Cana-de-Açúcar - SAFRA 2023/2024 - Vol. 11 Nº 1.
<https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>.
- BOSE, P. *et al.* Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, n. 11, p. 6563–6573, 2016.
- BOX, G. E. *et al.* *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.
- BREIMAN, L. Random forests. *Machine Learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, L. *et al.* *Classification And Regression Trees*. [S.l.]: Routledge, 1984.
<<https://doi.org/10.1201/9781315139470>>.
- BROWN, T. *et al.* Language models are few-shot learners. In: LAROCHELLE, H. *et al.* (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2020. v. 33, p. 1877–1901. Available in: <<https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>>.
- BU, Y. *et al.* Resnet incorporating the fusion data of rgb & hyperspectral images improves classification accuracy of vegetable soybean freshness. *Scientific reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 2568, 2024.
- CAMARGO, G.; KEMANIAN, A. Six crop models differ in their simulation of water uptake. *Agricultural and Forest Meteorology*, v. 220, p. 116–129, 2016. ISSN 0168-1923. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168192316300156>>.
- CANATA, T. F. *et al.* Sugarcane yield mapping using high-resolution imagery data and machine learning technique. *Remote Sensing*, v. 13, n. 2, 2021. ISSN 2072-4292. Available in: <<https://www.mdpi.com/2072-4292/13/2/232>>.
- CAO, H. *et al.* Trends in crop yield estimation via data assimilation based on multi-interdisciplinary analysis. *Field Crops Research*, Elsevier, v. 322, p. 109745, 2025.
- CAO, J. *et al.* Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches. *Agricultural and Forest Meteorology*, v. 297, p. 108275, 2021. ISSN 0168-1923. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168192320303774>>.

- CHHIKARA, P. *et al.* Data dimensionality reduction techniques for industry 4.0: Research results, challenges, and future research directions. *Software: Practice and Experience*, Wiley Online Library, v. 52, n. 3, p. 658–688, 2022.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *Peerj computer science*, PeerJ Inc., v. 7, p. e623, 2021.
- CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, v. 151, p. 61–69, 2018. ISSN 0168-1699. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168169917314710>>.
- CIOS, K.; PEDRYCZ, W.; SWINIARSKI, R. W. *Data Mining Methods for Knowledge Discovery*. USA: Kluwer Academic Publishers, 1998. ISBN 0792382528.
- CLAVERIE, M. *et al.* Maize and sunflower biomass estimation in southwest france using high spatial and temporal resolution remote sensing data. *Remote Sensing of Environment*, v. 124, p. 844–857, 2012. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/S0034425712001678>>.
- COIFMAN, R.; LAFON, S. Diffusion maps: Applied and computational harmonic analysis. *Applied and Computational Harmonic Analysis*, v. 21, n. 1, p. 5–14, 2004.
- COLWELL, R. N. R. N. *Manual of remote sensing*. 2nd. ed. Falls Church, Va.: American Society of Photogrammetry, 1983. ISBN 0937294411.
- Companhia Nacional de Abastecimento (CONAB). *Acompanhamento da Safra Brasileira de Cana-de-açúcar - 4^o Levantamento*. Brasília: [s.n.], 2025. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana>. Acesso em: 20 mar. 2026.
- CORLEY, I. *et al.* Revisiting pre-trained remote sensing model benchmarks: resizing and normalization matters. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2024. p. 3162–3172.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, Springer, v. 20, n. 3, p. 273–297, 1995.
- CUADRA, S. V. *et al.* A biophysical model of sugarcane growth. *Global Change Biology Bioenergy*, v. 4, n. 1, p. 36–48, 2012. Available in: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1757-1707.2011.01105.x>>.
- de Wit, A. *et al.* 25 years of the WOFOST cropping systems model. *Agricultural Systems*, v. 168, p. 154–167, 2019. ISSN 0308-521X. Available in: <<https://www.sciencedirect.com/science/article/pii/S0308521X17310107>>.
- DEVIJVER, P. A.; KITTLER, J. *Pattern Recognition: A Statistical Approach*. [S.l.]: Prentice Hall, 1982. 448 p. ISBN 978-0136542360.
- DEVLIN, J. *et al.* Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 4171–4186. Available in: <<https://aclanthology.org/N19-1423>>.

- DIAS, H. B. *et al.* Sugarcane yield future scenarios in Brazil as projected by the APSIM-Sugar model. *Industrial Crops and Products*, v. 171, p. 113918, 2021. ISSN 0926-6690. Available in: <<https://www.sciencedirect.com/science/article/pii/S0926669021006828>>.
- DONAHUE, J. *et al.* Decaf: A deep convolutional activation feature for generic visual recognition. In: XING, E. P.; JEBARA, T. (Ed.). *Proceedings of the 31st International Conference on Machine Learning*. Beijing, China: PMLR, 2014. (Proceedings of Machine Learning Research, v. 32), p. 647–655. Available in: <<https://proceedings.mlr.press/v32/donahue14.html>>.
- DORIGO, W. *et al.* 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, 2007. ISSN 0303-2434. Advances in airborne electromagnetics and remote sensing of agro-ecosystems. Available in: <<https://www.sciencedirect.com/science/article/pii/S0303243406000201>>.
- DOSOVITSKIY, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- DOSOVITSKIY, A. *et al.* *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021.
- DU, W.-Y. *et al.* Utilization of thermal infrared image for inversion of winter wheat yield and biomass. *Spectroscopy and Spectral Analysis*, Editorial Office, Spectroscopy and Spectral Analysis, v. 31, n. 6, p. 1476–1480, 2011.
- DUBAYAH, R. *et al.* The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. *Science of remote sensing*, Elsevier, v. 1, p. 100002, 2020.
- ELAVARASAN, D. *et al.* Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and electronics in agriculture*, Elsevier, v. 155, p. 257–282, 2018.
- ELSAYED, S. *et al.* Thermal imaging and passive reflectance sensing to estimate the water status and grain yield of wheat under different irrigation regimes. *Agricultural Water Management*, Elsevier, v. 189, p. 98–110, 2017.
- EPIPHANIO, J. C. N. Cbers-3/4: características e potencialidades. *Simpósio Brasileiro de Sensoriamento Remoto – SBSR*, Anais XV, p. 9009–9016, 2011. Available in: <<http://marte.sid.inpe.br/attachment.cgi/dpi.inpe.br/marte/2011/07.14.19.08/doc/p1222.pdf>>.
- ERMOLIEVA, T. *et al.* A novel robust meta-model framework for predicting crop yield probability distributions using multisource data. *Cybernetics and Systems Analysis*, Springer, v. 59, n. 5, p. 844–858, 2023.
- EVETT, I. *et al.* The impact of the principles of evidence interpretation on the structure and content of statements. *Science & Justice*, v. 40, n. 4, p. 233–239, 2000. ISSN 1355-0306. Available in: <<https://www.sciencedirect.com/science/article/pii/S1355030600719939>>.
- FEI, Y. *et al.* A survey of geometric optimization for deep learning: From euclidean space to riemannian manifold. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 57, n. 5, jan. 2025. ISSN 0360-0300. Available in: <<https://doi.org/10.1145/3708498>>.

- Felipe Maldaner, L. *et al.* Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. *Computers and Electronics in Agriculture*, v. 181, p. 105945, 2021. ISSN 0168-1699. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168169920331501>>.
- FERNANDES, J. L.; EBECKEN, N. F. F.; ESQUERDO, J. C. D. M. Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *International Journal of Remote Sensing*, Taylor & Francis, v. 38, n. 16, p. 4631–4644, 2017. Available in: <<https://doi.org/10.1080/01431161.2017.1325531>>.
- FERREIRA, K. R. *et al.* Earth observation data cubes for brazil: Requirements, methodology and products. *Remote Sensing*, MDPI, v. 12, n. 24, p. 4033, 2020.
- FIEUZAL, R.; BAUP, F. Forecast of wheat yield throughout the agricultural season using optical and radar satellite images. *International Journal of Applied Earth Observation and Geoinformation*, v. 59, p. 147–156, 2017. ISSN 0303-2434. Available in: <<https://www.sciencedirect.com/science/article/pii/S0303243417300739>>.
- FOOD; ORGANIZATION, A. Faostat: Value of agricultural production. <https://www.fao.org/faostat/en>, 2022.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, Elsevier, v. 55, n. 1, p. 119–139, 1997.
- FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189–1232, 2001.
- GADE, S. A. *et al.* Advancements in uav remote sensing for agricultural yield estimation: A systematic comprehensive review of platforms, sensors, and data analytics. *Remote Sensing Applications: Society and Environment*, Elsevier, v. 37, p. 101418, 2025.
- GEIPEL, J.; LINK, J.; CLAUPEIN, W. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 6, n. 11, p. 10335–10355, 2014.
- GEOSYSTEMS, L. *Leica Viva GS16–Self-Learning GNSS Smart Antenna*. 2019.
- GEWERS, F. L. *et al.* Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 54, n. 4, p. 1–34, 2021.
- GILBERT, R. *et al.* The effect of genotype, environment and time of harvest on sugarcane yields in florida, usa. *Field Crops Research*, v. 95, n. 2, p. 156–170, 2006. ISSN 0378-4290. Available in: <<https://www.sciencedirect.com/science/article/pii/S0378429005000602>>.
- GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from eos-modis. *Remote Sensing of Environment*, v. 58, n. 3, p. 289–298, 1996. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/S0034425796000727>>.

- GITELSON, A. A. *et al.* Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction. *International Journal of Remote Sensing*, Taylor & Francis, v. 23, n. 13, p. 2537–2562, 2002.
- GLENN, E. P. *et al.* Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors*, v. 8, n. 4, p. 2136–2160, 2008. ISSN 1424-8220. Available in: <<https://www.mdpi.com/1424-8220/8/4/2136>>.
- GOEL, M.; PANDEY, M. Crop yield prediction using ai: A review. In: IEEE. *2024 2nd international conference on disruptive technologies (ICDT)*. [S.l.], 2024. p. 1547–1553.
- GOMMES, R.; BERNARDI, M.; PETRASSI, F. Agrometeorological crop forecasting. *Sustainable Development Dimensions, FAO website, www.fao.org/waicent/faoinfo/sustdev/Eidirect/AGROMET/FORECAST. HTM*, 1996.
- GOODFELLOW, I. *et al.* *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1.
- GREAVES, H. E. *et al.* Estimating aboveground biomass and leaf area of low-stature arctic shrubs with terrestrial LiDAR. *Remote Sensing of Environment*, Elsevier, v. 164, p. 26–35, 2015.
- GUO, J. *et al.* Evaluation of the grain yield and nitrogen nutrient status of wheat (*triticum aestivum* l.) using thermal imaging. *Field Crops Research*, Elsevier, v. 196, p. 463–472, 2016.
- GÉRON, A. *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. Sebastopol, CA: O'Reilly Media, 2017. ISBN 978-1491962299.
- HAGA, T.; KAN, T.; FUJIKOSHI, Y. Variable selection criteria based on multiple correlation coefficient in regression model. *Journal of Statistical Applications*, Japan Statistical Society, v. 4, n. 2-3, p. 265–279, 2009.
- HAGHIGHATTALAB, A.; PÉREZ, L. G.; AL., S. M. *et al.* Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods*, v. 12, n. 1, 2016. Available in: <<https://doi.org/10.1186/s13007-016-0134-6>>.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. ed. New York: Springer, 2009.
- HE, J. *et al.* Identifying irrigation and nitrogen best management practices for sweet corn production on sandy soils using ceres-maize model. *Agricultural Water Management*, v. 109, p. 61–70, 2012. ISSN 0378-3774. Available in: <<https://www.sciencedirect.com/science/article/pii/S0378377412000625>>.
- HE, K. *et al.* Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778.
- HE, K. *et al.* Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

- HEDLEY, C. The role of precision agriculture for improved nutrient management on farms. *Journal of the Science of Food and Agriculture*, v. 95, n. 1, p. 12–19, 2015. Available in: <<https://onlinelibrary.wiley.com/doi/full/10.1002/jsfa.6734>>.
- HESS, K. W. *et al.* Tidal constituent and residual interpolation (tcari): A new method for the tidal correction of bathymetric data. *NOAA Technical Report NOS CS4*, 1999.
- HIREMANI, V. *et al.* Federated learning for crop yield prediction: A comprehensive review of techniques and applications. *MethodsX*, Elsevier, v. 14, p. 103408, 2025.
- HORIE, T.; YAJIMA, M.; NAKAGAWA, H. Yield forecasting. *Agricultural Systems*, v. 40, n. 1, p. 211–236, 1992. ISSN 0308-521X. Systems approaches for agricultural. Available in: <<https://www.sciencedirect.com/science/article/pii/0308521X9290022G>>.
- HUBER, F. *et al.* Integrating process-based models and machine learning for crop yield prediction. In: *Proceedings of the 1st Workshop on Synergy of Scientific and Machine Learning Modeling (SynS & ML)*, ICML. Honolulu, Hawaii, USA: [s.n.], 2023.
- HUETE, A. A soil-adjusted vegetation index (savi). *Remote Sensing of Environment*, v. 25, n. 3, p. 295–309, 1988. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/003442578890106X>>.
- HUETE, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, v. 83, n. 1, p. 195–213, 2002. ISSN 0034-4257. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring. Available in: <<https://www.sciencedirect.com/science/article/pii/S0034425702000962>>.
- HUETE, A.; JUSTICE, C.; LIU, H. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, v. 49, n. 3, p. 224–234, 1994. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/0034425794900183>>.
- HUNT, L.; WHITE, J.; HOOGENBOOM, G. Agronomic data: advances in documentation and protocols for exchange and use. *Agricultural Systems*, v. 70, n. 2, p. 477–492, 2001. ISSN 0308-521X. Available in: <<https://www.sciencedirect.com/science/article/pii/S0308521X01000567>>.
- INAM, S. A.; RAJPUT, H.; UMER, S. A hierarchical ensemble approach for multi-country pm10 forecasting using lightgbm and residual neural network. *Discover Atmosphere*, Springer, v. 4, n. 1, p. 1, 2026.
- IPPOLITO, P. P. Hyperparameter tuning: The art of fine-tuning machine and deep learning models to improve metric results. In: *Applied data science in tourism: Interdisciplinary approaches, methodologies, and applications*. [S.l.]: Springer, 2022. p. 231–251.
- JABED, M. A.; MURAD, M. A. A. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, Elsevier, v. 10, n. 24, 2024.
- JAMES, G. *et al.* *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112.
- JIA, W. *et al.* Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, Springer, v. 8, n. 3, p. 2663–2693, 2022.

- JOHNSTONE, I. M. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, Institute of Mathematical Statistics, v. 29, n. 2, p. 295–327, 2001.
- JONES, C. *et al.* Epic: An operational model for evaluation of agricultural sustainability. *Agricultural Systems*, v. 37, n. 4, p. 341–350, 1991. ISSN 0308-521X. Application of Crop Simulation Models. Available in: <<https://www.sciencedirect.com/science/article/pii/0308521X9190057H>>.
- JONES, J. *et al.* The dssat cropping system model. *European Journal of Agronomy*, v. 18, n. 3, p. 235–265, 2003. ISSN 1161-0301. Modelling Cropping Systems: Science, Software and Applications. Available in: <<https://www.sciencedirect.com/science/article/pii/S1161030102001077>>.
- JONES, J. W. *et al.* Decision support system for agrotechnology transfer: DSSAT v3. In: TSUJI, G. Y.; HOOGENBOOM, G.; THORNTON, P. K. (Ed.). *Understanding Options for Agricultural Production*. Dordrecht: Springer Netherlands, 1998. p. 157–177. ISBN 978-94-017-3624-4. Available in: <https://doi.org/10.1007/978-94-017-3624-4_8>.
- JONES, J. W. *et al.* The dssat cropping system model. *European journal of agronomy*, Elsevier, v. 18, n. 3-4, p. 235–265, 2003.
- JOSHAGHANI, M. *et al.* Retail time series forecasting using an automated deep meta-learning framework. *Available at SSRN 4393300*, 2023.
- JR, E. R. H. *et al.* Remote sensing leaf chlorophyll content using a visible band index. *Agronomy journal*, Wiley Online Library, v. 103, n. 4, p. 1090–1099, 2011.
- JR, J. W. R. *et al.* *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*. [S.l.], 1973.
- JUNIOR, G. A. L. *et al.* Nitrogen uptake and physiological use efficiency in brazilian sugarcane (saccharum × spp.) cultivars: A gist of current understanding. *South African Journal of Botany*, Elsevier, v. 178, p. 101–112, 2025.
- KAMIR, E.; WALDNER, F.; HOCHMAN, Z. Estimating wheat yields in australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 160, p. 124–135, 2020. ISSN 0924-2716. Available in: <<https://www.sciencedirect.com/science/article/pii/S092427161930262X>>.
- KANEKO, A. *et al.* Deep learning for crop yield prediction in africa. *International Conference on Machine Learning AI for Social Good Workshop*, 2019. Available in: <https://aiforsocialgood.github.io/icml2019/accepted/track1/pdfs/20_aig_icml2019.pdf>.
- KANG, Y. *et al.* Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US midwest. *Environmental Research Letters*, IOP Publishing, v. 15, n. 6, p. 064005, may 2020. Available in: <<https://doi.org/10.1088/1748-9326/ab7df9>>.
- KEENAN, T. F. *et al.* Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change*, v. 4, p. 598 – 604, 2014/07// 2014. Available in: <<http://www.nature.com/nclimate/journal/v4/n7/full/nclimate2253.html>>.

- KERN, A. *et al.* Statistical modelling of crop yield in central europe using climate data and remote sensing vegetation indices. *Agricultural and Forest Meteorology*, v. 260-261, p. 300–320, 2018. ISSN 0168-1923. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168192318302004>>.
- KHAN, A. *et al.* *PCA vs t-SNE vs UMAP: Comparison of dimensionality reduction algorithms*. 2025. GitHub repository. Comparação prática mostrando que PCA é mais rápido e preserva melhor a estrutura de dados pequenos. Available in: <<https://github.com/ayeshuukhan/PCA-vs-tSNE-vs-UMAP>>.
- KHANAL, S. *et al.* Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, v. 153, p. 213–225, 2018. ISSN 0168-1699. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168169918300334>>.
- KIM, H. G.; PARK, J.-S.; LEE, D.-H. Potential of Unmanned Aerial Sampling for Monitoring Insect Populations in Rice Fields. *Florida Entomologist*, Florida Entomological Society, v. 101, n. 2, p. 330 – 334, 2018. Available in: <<https://doi.org/10.1653/024.101.0229>>.
- KUMAR C. S. MURTHY, M. V. R. S. S. M. N.; ROY, P. S. On the use of standardized precipitation index (SPI) for drought intensity assessment. *METEOROLOGICAL APPLICATIONS*, n. 16, p. 381 – 389, 2009. Available in: <<https://rmets.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/met.136>>.
- KUMARI, M.; PRASAD, D. *et al.* Crop yield prediction using remote sensing: a review. In: IEEE. *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)*. [S.l.], 2024. v. 1, p. 547–552.
- LAWES, R.; LAWN, R. Applications of industry information in sugarcane production systems. *Field Crops Research*, Elsevier, v. 92, n. 2-3, p. 353–363, 2005.
- LECUN, Y. *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.
- LI, W. *et al.* Combined use of airborne lidar and satellite gf-1 data to estimate leaf area index, height, and aboveground biomass of maize during peak growing season. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 8, n. 9, p. 4489–4501, 2015.
- LIAGHAT, S.; BALASUNDRAM, S. K. A review: The role of remote sensing in precision agriculture. *American Journal of Agricultural and Biological Sciences*, Science Publications, v. 5, n. 1, p. 50–55, Mar. 2010. Available in: <<https://thescipub.com/abstract/ajabssp.2010.50.55>>.
- LIN, F. *et al.* Mmst-vit: Climate change-aware crop yield prediction via multi-modal spatial-temporal vision transformer. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2023. p. 5774–5784.
- LIU, Z. *et al.* In-season crop yield prediction: State of the art and future research direction. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 146, p. 105129, 2026.

- LOBELL, D. B. *et al.* A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, v. 164, p. 324–333, 2015. ISSN 0034-4257. Available in: <https://www.sciencedirect.com/science/article/pii/S0034425715001637>.
- LORENA, A. C.; CARVALHO, A. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. ISSN 21752745. Available in: https://seer.ufrgs.br/rita/article/view/rita_v14_n2_p43-67.
- LU, J. *et al.* Deep learning for multi-source data-driven crop yield prediction in northeast china. *Agriculture*, MDPI, v. 14, n. 6, p. 794, 2024.
- LUCIANO, A. C. dos S. *et al.* Empirical model for forecasting sugarcane yield on a local scale in Brazil using landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture*, v. 184, p. 106063, 2021. ISSN 0168-1699. Available in: <https://www.sciencedirect.com/science/article/pii/S0168169921000818>.
- LUZ, B. R. da; CROWLEY, J. K. Identification of plant species by using high spatial and spectral resolution thermal infrared (8.0–13.5 μm) imagery. *Remote Sensing of Environment*, Elsevier, v. 114, n. 2, p. 404–413, 2010.
- MAIMAITIJIANG, M. *et al.* Unmanned aerial system (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 134, p. 43–58, 2017.
- MAIMAITIJIANG, M. *et al.* Vegetation index weighted canopy volume model (cvmvi) for soybean biomass estimation from unmanned aerial system-based rgb imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 151, p. 27–41, 2019.
- MAIMAITIJIANG, M. *et al.* Soybean yield prediction from uav using multimodal data fusion and deep learning. *Remote Sensing of Environment*, v. 237, p. 111599, 2020. ISSN 0034-4257. Available in: <https://www.sciencedirect.com/science/article/pii/S0034425719306194>.
- Maya Gopal, P.; BHARGAVI, R. A novel approach for efficient crop yield prediction. *Computers and Electronics in Agriculture*, v. 165, p. 104968, 2019. ISSN 0168-1699. Available in: <https://www.sciencedirect.com/science/article/pii/S0168169919311093>.
- MCCOWN, R. *et al.* Apsim: a novel software system for model development, model testing and simulation in agricultural systems research. *Agricultural Systems*, v. 50, n. 3, p. 255–271, 1996. ISSN 0308-521X. Available in: <https://www.sciencedirect.com/science/article/pii/0308521X9400055V>.
- MCCOWN, R. L. *et al.* Apsim: a novel software system for model development, model testing and simulation in agricultural systems research. *Agricultural systems*, Elsevier, v. 50, n. 3, p. 255–271, 1996.
- MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, Taylor & Francis, v. 17, n. 7, p. 1425–1432, 1996. Available in: <https://doi.org/10.1080/01431169608948714>.
- MCINNES, L.; HEALY, J.; MELVILLE, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

- MCKEE, N. J. D. T. B.; KLEIST, J. *The relationship of drought frequency and duration to time scales*. 1993. Available in: <<https://climate.colostate.edu/pdfs/relationshipofdroughtfrequency.pdf>>.
- MCKINNEY, N.; JR, W. S.; KANEMASU, E. Canopy temperature, seed yield, and vapor pressure deficit relationship in soybean. *Crop science*, Wiley Online Library, v. 29, n. 4, p. 1038–1041, 1989.
- MEGHRAOUI, K. *et al.* Applied deep learning-based crop yield prediction: A systematic analysis of current developments and potential challenges. *Technologies*, MDPI, v. 12, n. 4, p. 43, 2024.
- MIENYE, I. D.; SUN, Y. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *Ieee Access*, IEEE, v. 10, p. 99129–99149, 2022.
- MITCHELL, T. M. Machine learning. *Computer*, IEEE, v. 30, n. 11, p. 4–9, 1997.
- MOON, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, Nature Publishing Group US New York, v. 37, n. 12, p. 1482–1492, 2019.
- MURTHY, V. R. K. Crop growth modeling and its applications in agricultural meteorology. In: SIVAKUMAR, M. *et al.* (Ed.). *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*. World Meteorological Organisation, 2004. Proceedings of the Training Workshop, p. 235–261. Available in: <https://www.unisdr.org/preventionweb/files/1682_9970.pdf#page=236>.
- MURUGANANTHAM, P. *et al.* A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, MDPI, v. 14, n. 9, p. 1990, 2022.
- NASSIF, D. S. P. *et al.* Parametrização e avaliação do modelo dssat/canegro para variedades brasileiras de cana-de-açúcar. *Pesquisa Agropecuária Brasileira*, v. 47, p. 311–318, 2012. Available in: <<https://www.scielo.br/j/pab/a/SmCFRPFYTHMDS6XmvFM5zBx/?lang=pt>>.
- NEVAVUORI, P.; NARRA, N.; LIPPING, T. Crop yield prediction with deep convolutional neural networks. *Computers and Electronics in Agriculture*, v. 163, p. 104859, 2019. ISSN 0168-1699. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168169919306842>>.
- NOGUEIRA, E. A. *et al.* Deep learning for super resolution of sugarcane crop line imagery from unmanned aerial vehicles. In: SPRINGER. *International Symposium on Visual Computing*. [S.l.], 2023. p. 597–609.
- OIKONOMIDIS, A.; CATAL, C.; KASSAHUN, A. Deep learning for crop yield prediction: a systematic literature review. *New Zealand Journal of Crop and Horticultural Science*, Wiley Online Library, v. 51, n. 1, p. 1–26, 2023.
- OQUAB, M. *et al.* DINOv2: Learning Robust Visual Features without Supervision. 2024. Available in: <<https://arxiv.org/abs/2304.07193>>.
- PAIDIPATI, K. K. *et al.* Forecasting of sugarcane productivity estimation in india - a comparative study with advanced non-parametric regression models. *JOURNAL OF ALGEBRAIC STATISTICS*, v. 13, p. 760–778, 2022.

- PAUDEL, D. *et al.* Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, v. 187, p. 103016, 2021. ISSN 0308-521X. Available in: <https://www.sciencedirect.com/science/article/pii/S0308521X20308775>.
- PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, v. 12, p. 2825–2830, 2011.
- PIGNÈDE, E. *et al.* Sugarcane yield forecast in Ivory Coast (West Africa) based on weather and vegetation index data. *Atmosphere*, MDPI, v. 12, n. 11, p. 1459, 2021.
- PRADO, H. do. *Ambientes de produção de cana-de-açúcar na região centro-sul do brasil*. 2005. São Paulo: Ribeirão Preto.
- PRIESTLEY, M. *Spectral Analysis and Time Series*. [S.l.]: Academic Press, 1981.
- QIAO, M. *et al.* Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3d convolutional neural networks. *International Journal of Applied Earth Observation and Geoinformation*, v. 102, p. 102436, 2021. ISSN 0303-2434. Available in: <https://www.sciencedirect.com/science/article/pii/S0303243421001434>.
- QIAO, M. *et al.* Exploiting hierarchical features for crop yield prediction based on 3-d convolutional neural networks and multikernel gaussian process. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 14, p. 4476–4489, 2021.
- QUINLAN, J. R. *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0. Available in: <http://portal.acm.org/citation.cfm?id=152181>.
- RAHMAN, A. J. R. M. M. A novel approach for sugarcane yield prediction using Landsat time series imagery: A case study on Bundaberg Region. *Advances in Remote Sensing*, v. 5, n. 2, 2016. Available in: <https://www.scirp.org/journal/paperinformation.aspx?paperid=67130>.
- REINECKE, M.; PRINSLOO, T. The influence of drone monitoring on crop health and harvest size. In: *2017 1st International Conference on Next Generation Computing Applications (NextComp)*. [S.l.: s.n.], 2017. p. 5–10.
- RIBEIRO, J. B. *et al.* Automated detection of sugarcane crop lines from uav images using deep learning. *Information Processing in Agriculture*, Elsevier, v. 11, n. 3, p. 385–396, 2024.
- RISCHBECK, P. *et al.* Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. *European Journal of Agronomy*, Elsevier, v. 78, p. 44–59, 2016.
- RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. *Remote sensing of environment*, Elsevier, v. 55, n. 2, p. 95–107, 1996.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.

RUSSAKOVSKY, O. *et al.* Imagenet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 2014. Cite arxiv:1409.0575Comment: 43 pages, 16 figures. v3 includes additional comparisons with PASCAL VOC (per-category comparisons in Table 3, distribution of localization difficulty in Fig 16), a list of queries used for obtaining object detection images (Appendix C), and some additional references. Available in: <http://arxiv.org/abs/1409.0575>.

SAKAMOTO, T. Incorporating environmental variables into a modis-based crop yield estimation method for united states corn and soybeans through the use of a random forest regression algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 160, p. 208–228, 2020. ISSN 0924-2716. Available in: <https://www.sciencedirect.com/science/article/pii/S0924271619303065>.

SANTOS, H. d. dos *et al.* Brazilian soil classification system. *Technical Report Embrapa*, 2018.

SARKAR, S. *et al.* Crop yield prediction using multimodal meta-transformer and temporal graph neural networks. *IEEE Transactions on AgriFood Electronics*, IEEE, 2024.

SCHWALBERT, R. A. *et al.* Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern brazil. *Agricultural and Forest Meteorology*, v. 284, p. 107886, 2020. ISSN 0168-1923. Available in: <https://www.sciencedirect.com/science/article/pii/S0168192319305027>.

SCREPNIK, C. R.; ZAMUDIO, E.; GIMENEZ, L. I. Artificial intelligence in agriculture: a systematic review of crop yield prediction and optimization. *IEEE Access*, IEEE, v. 13, p. 70691–70697, 2025.

SHAHHOSSEINI, M. *et al.* Coupling machine learning and crop modeling improves crop yield prediction in the us corn belt. *Scientific Reports*, Springer Science and Business Media LLC, v. 11, n. 1, Jan 2021. ISSN 2045-2322. Available in: <http://dx.doi.org/10.1038/s41598-020-80820-1>.

SHARKEY, T. D. Effects of moderate heat stress on photosynthesis: importance of thylakoid reactions, rubisco deactivation, reactive oxygen species, and thermotolerance provided by isoprene. *Plant, Cell & Environment*, Wiley Online Library, v. 28, n. 3, p. 269–277, 2005.

SHAWON, S. M. *et al.* Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology*, Elsevier, v. 10, p. 100718, 2025.

SHEHAB, A. *et al.* Explainable optimized lstm model for sustainable wind power forecasting in smart cities. *International Journal of Intelligent Systems*, Wiley Online Library, v. 2026, n. 1, p. 6851438, 2026.

SHENDRYK, I. *et al.* Data for: Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. 2020. Available in: <https://data.mendeley.com/datasets/wck5rrxy7/1>.

SHENDRYK, Y.; DAVY, R.; THORBURN, P. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*, Elsevier, v. 260, p. 107984, 2021.

- SHENDRYK, Y. *et al.* Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. *International Journal of Applied Earth Observation and Geoinformation*, v. 92, p. 102177, 2020. ISSN 0303-2434. Available in: <<https://www.sciencedirect.com/science/article/pii/S0303243420303457>>.
- SHI, Y. *et al.* Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PloS one*, Public Library of Science San Francisco, CA USA, v. 11, n. 7, p. e0159781, 2016.
- SHUI, J. *et al.* Entity semantic feature fusion network for remote sensing image-text retrieval. In: SPRINGER. *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*. [S.l.], 2024. p. 130–145.
- SILVA, N. R. de França e *et al.* Sugarcane yield estimation using satellite remote sensing data in empirical or mechanistic modeling: A systematic review. *Remote Sensing*, MDPI, v. 16, n. 5, p. 863, 2024.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015.
- SIMS, D. A.; GAMON, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote sensing of environment*, Elsevier, v. 81, n. 2-3, p. 337–354, 2002.
- SIMÕES, M. dos S.; ROCHA, J. V.; LAMPARELLI, R. A. C. Orbital spectral variables, growth analysis and sugarcane yield. *Scientia Agricola*, v. 66, n. 4, p. 451–461, 2009. Available in: <<https://www.scielo.br/j/sa/a/S93KQMgM7xYLdK6966KprHS/?lang=en>>.
- SOFONIA, J. J. *et al.* Modelling the effects of fundamental uav flight parameters on lidar point clouds to facilitate objectives-based planning. *ISPRS journal of photogrammetry and remote sensing*, Elsevier, v. 149, p. 105–118, 2019.
- SOILLE, P. *et al.* A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Generation Computer Systems*, v. 81, p. 30–40, 2018. ISSN 0167-739X. Available in: <<https://www.sciencedirect.com/science/article/pii/S0167739X1730078X>>.
- SON, N. *et al.* A comparative analysis of multitemporal modis evi and ndvi data for large-scale rice yield estimation. *Agricultural and Forest Meteorology*, v. 197, p. 52–64, 2014. ISSN 0168-1923. Available in: <<https://www.sciencedirect.com/science/article/pii/S016819231400149X>>.
- STEELE, M. R. *et al.* Nondestructive estimation of anthocyanin content in grapevine leaves. *American Journal of Enology and Viticulture*, American Journal of Enology and Viticulture, v. 60, n. 1, p. 87–92, 2009.
- STÖCKLE, C. O.; DONATELLI, M.; NELSON, R. Cropsyst, a cropping systems simulation model. *European Journal of Agronomy*, v. 18, n. 3, p. 289–307, 2003. ISSN 1161-0301. Modelling Cropping Systems: Science, Software and Applications. Available in: <<https://www.sciencedirect.com/science/article/pii/S1161030102001090>>.

- SZEGEDY, C. *et al.* Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI Press, 2017. (AAAI'17), p. 4278–4284.
- TAHASEEN, M.; MOPARTHI, N. R. An assessment of the machine learning algorithms used in agriculture. In: IEEE. *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. [S.l.], 2021. p. 1579–1584.
- TAKARA, L. de A. *et al.* Optimizing multi-step wind power forecasting: Integrating advanced deep neural networks with stacking-based probabilistic learning. *Applied Energy*, Elsevier, v. 369, p. 123487, 2024.
- TARAVAT, A. *et al.* Estimation of sugarcane yield using multi-temporal sentinel-2 satellite imagery and random forest regression. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. [S.l.: s.n.], 2024. (ISPRS Archives, XLVIII-4/W9), p. 357–362.
- TEDESCO, M. Remote sensing and the cryosphere. *Remote Sensing of the Cryosphere*, Wiley Online Library, p. 1–16, 2015.
- TEENBAUM, J.; SILVA, D.; LANGFORD, J. Global geometric framework for nonlinear dimensionality reduction [j]. *Science*, v. 290, n. 5500, p. 2319–2323, 2000.
- TEMPOCAMPO. Sistema de monitoramento do panorama agrometeorológico de culturas. <https://tempocampo.org>, 2021.
- TOWNSHEND, J. e. a. Integrated global observation of the land: an IGOS-P theme. *FAO IGOL*, GTOS 54, n. 8, 2008. Available in: <http://www.fao.org/3/i0536e/i0536e.pdf>.
- TUNIO, M. H. *et al.* Meta-knowledge guided bayesian optimization framework for robust crop yield estimation. *Journal of King Saud University-Computer and Information Sciences*, Elsevier, v. 36, n. 1, p. 101895, 2024.
- UNION, E. Article 67. *THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 December 2013 on the financing, management and monitoring of the common agricultural policy and repealing Council Regulations*, Chapter II, n. 165/94, 2013.
- van Klompenburg, T.; KASSAHUN, A.; CATAL, C. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, v. 177, p. 105709, 2020. ISSN 0168-1699. Available in: <https://www.sciencedirect.com/science/article/pii/S0168169920302301>.
- VAPNIK, V. *The Nature of Statistical Learning Theory*. [S.l.]: Springer, New York, 1995.
- VAPNIK, V. N.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, v. 16, n. 2, p. 264–280, 1971.
- VASCONCELOS, J. C. S. *et al.* Predicting sugarcane yield through temporal analysis of satellite imagery during the growth phase. *Agronomy*, v. 15, n. 4, p. 793, 2025.
- VELODYNE, L. I. Vlp-16 user manual. <https://velodynelidar.com/wp-content/uploads/2019/12/63-9243-Rev-E-VLP-16-User-Manual.pdf>, 2019.

- VENKATESH, T.; LM, J. L.; RAJKUMAR, S. A predictive model for marine debris prediction (a comparative case study). In: IEEE. *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*. [S.l.], 2024. p. 1–7.
- VILLARES, M.; SAUNDERS, C. M.; FEY, N. Comparison of dimensionality reduction techniques for the visualisation of chemical space in organometallic catalysis. *Artificial Intelligence Chemistry*, Elsevier, v. 2, n. 1, p. 100055, 2024.
- VOCATURO, E.; PERNA, D.; ZUMPANO, E. Machine learning techniques for automated melanoma detection. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.: s.n.], 2019. p. 2310–2317.
- WANG, A. X. *et al.* Deep transfer learning for crop yield prediction with remote sensing data. In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. New York, NY, USA: Association for Computing Machinery, 2018. (COMPASS '18). ISBN 9781450358163. Available in: <<https://doi.org/10.1145/3209811.3212707>>.
- WANG, C. *et al.* Estimating the biomass of maize with hyperspectral and lidar data. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 9, n. 1, p. 11, 2017.
- WANG, Y. *et al.* Ssl4eo-s12: A large-scale multimodal, multitemporal dataset for self-supervised learning in earth observation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, IEEE, v. 11, n. 3, p. 98–106, 2023.
- WANG, Y. *et al.* Understanding how dimension reduction tools work: An empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization. *Journal of Machine Learning Research*, v. 22, n. 201, p. 1–73, 2021. Available in: <<http://jmlr.org/papers/v22/20-1061.html>>.
- WEI, M. C. F.; MOLIN, J. P.; LONGCHAMPS, L. Predictive power vs interpretability: Machine learning approaches to unravel sugarcane yield drivers. *Computers and Electronics in Agriculture*, Elsevier, v. 243, p. 111353, 2026.
- WEISS, C. H. *Discrete-valued time series*. [S.l.]: MDPI, 2023. 1576 p.
- WEISS, M.; JACOB, F.; DUVEILLER, G. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, v. 236, p. 111402, 2020. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/S0034425719304213>>.
- WOOD, S. *et al.* Calibrating dimension reduction hyperparameters in the presence of noise. *PLOS Computational Biology*, Public Library of Science, v. 20, n. 9, p. e1012427, 2024. Demonstra que UMAP e t-SNE tendem a overfit quando hiperparâmetros são calibrados para capturar toda a variação dos dados, incluindo ruído.
- WOODCOCK, C. E. *et al.* Free access to landsat imagery. *Science*, v. 320, n. 5879, p. 1011–1011, 2008. Available in: <<https://www.science.org/doi/abs/10.1126/science.320.5879.1011a>>.
- XAVIER, A. C. *et al.* New improved brazilian daily weather gridded data (1961–2020). *International Journal of Climatology*, Wiley Online Library, v. 42, n. 16, p. 8390–8404, 2022.

XAVIER, A. C. *et al.* New improved brazilian daily weather gridded data (1961–2020). *International Journal of Climatology*, Wiley Online Library, v. 42, n. 16, p. 8390–8404, 2022.

XU, X. *et al.* Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. *Ecological Indicators*, v. 101, p. 943–953, 2019. ISSN 1470-160X. Available in: <<https://www.sciencedirect.com/science/article/pii/S1470160X19300767>>.

XUE, J.; SU, B. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, Hindawi, v. 2017, p. 1353691, 2017.

YANG, M.-D. *et al.* Adaptive autonomous uav scouting for rice lodging assessment using edge computing with deep learning edanet. *Computers and Electronics in Agriculture*, v. 179, p. 105817, 2020. ISSN 0168-1699. Available in: <<https://www.sciencedirect.com/science/article/pii/S0168169920319542>>.

YANG, Q. *et al.* Deep convolutional neural networks for rice grain yield estimation at the ripening stage using uav-based remotely sensed images. *Field Crops Research*, v. 235, p. 142–153, 2019. ISSN 0378-4290. Available in: <<https://www.sciencedirect.com/science/article/pii/S037842901831390X>>.

YENKIKAR, A. *et al.* Explainable forecasting of air quality index using a hybrid random forest and arima model. *MethodsX*, Elsevier, p. 103517, 2025.

YOU, J. *et al.* Deep gaussian process for crop yield prediction based on remote sensing data. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI Press, 2017. (AAAI'17), p. 4559–4565.

ZENG, L. *et al.* A hybrid approach for detecting corn and soybean phenology with time-series modis data. *Remote Sensing of Environment*, v. 181, p. 237–250, 2016. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/S0034425716301432>>.

ZHANG, J. K. C. The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, v. 13, n. 6, p. 693–712, 2012. Available in: <<https://link.springer.com/article/10.1007/s11119-012-9274-5>>.

ZHANG, Y.; LIU, J.; SHEN, W. A review of ensemble learning algorithms used in remote sensing applications. *Applied Sciences*, MDPI, v. 12, n. 17, p. 8654, 2022.

ZHONG, L.; HU, L.; ZHOU, H. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, v. 221, p. 430–443, 2019. ISSN 0034-4257. Available in: <<https://www.sciencedirect.com/science/article/pii/S0034425718305418>>.

DADOS CURRICULARES

IDENTIFICAÇÃO	
	18/02/1974
Nacionalidade	Brasileiro
Nome em citações bibliográficas	L. A. F. Barbosa BARBOSA, L. A. F.
Currículo Lattes	http://lattes.cnpq.br/5524528660042920
ORCID	http://orcid.org/0000-0002-1711-6339
FORMAÇÃO ACADÊMICA	
30/11/2003	Bacharelado em Ciências da Computação Universidade Estadual Paulista “Júlio de Mesquita Filho” - Instituto de Geociências e Ciências Exatas - Campus de Rio Claro
PRODUÇÃO BIBLIOGRÁFICA	
BARBOSA, A. et al. Modeling yield response to crop management using convolutional neural networks. Computers and Electronics in Agriculture, v. 170, p. 105197, 2020. ISSN 0168-1699. Available in: <https://www.sciencedirect.com/science/article/pii/S0168169919308543>. BARBOSA, L. A. F. PEDRONETTE, D. C. G.; GUILHERME, I. R. A meta-feature model for exploiting different regressors to estimate sugarcane crop yield. In: International Geoscience and Remote Sensing Symposium (IGARSS). [S.l.: s.n.], 2023. BARBOSA, L. A. F. et al. Comparative study of image descriptors and dimensionality reduction methods for the prediction of sugarcane crop yield. In: SPRINGER. Iberoamerican Congress on Pattern Recognition. [S.l.], 2025. p. 16–29. BARBOSA, L. A. F. et al. Meta-features extracted from use of knn regressor to improve sugarcane crop yield prediction. Remote Sensing, MDPI, v. 17, n. 11, p. 1846, 2025. BARBOSA, L. A. F. et al. Sugarcane dataset northeast Sao Paulo state, Brazil (Revision d064a64). Hugging Face, 2025. Available in: <https://huggingface.co/datasets/lafbarbosa/sugarcane_dataset_northeast_sao_paulo_state_brazil>. BARBOSA, L. A. F. et al. A dataset of sugarcane crop yield, production environment, meteorological records, and satellite images of commercial fields in the northeast of são paulo state, brazil. Data in Brief, p. 112549, 2026. ISSN 2352-3409. Available in: <https://www.sciencedirect.com/science/article/pii/S2352340926001022>.	
PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS	
XX Simpósio Brasileiro de Sensoriamento Remoto (SBSR 2023). Apresentação de Trabalho. Florianópolis, SC, Brasil, 02 de abril a 05 de abril 2023. 2023 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2023). Apresentação de Trabalho. Pasadena, CA, Estados Unidos, 16 de julho a 21 de julho 2023. 28º Congresso Ibero-Americano de Reconhecimento de Padrões (CIARP 2025). Apresentação de Trabalho. Bogotá, CUN, Colômbia, 25 de novembro a 28 de novembro 2025.	