

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS

CÂMPUS DE JABOTICABAL

**Previsão e estimativa da produtividade de milho e cana-de-
açúcar utilizando Inteligência artificial e sensoriamento
remoto**

Antonio Alves Pinto

Engenheiro Agrônomo

2024

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS

CÂMPUS DE JABOTICABAL

**Previsão e estimativa da produtividade de milho e cana-de-
açúcar utilizando Inteligência artificial e sensoriamento
remoto**

Discente: Eng. Agrônomo. Antonio Alves Pinto

Orientador: Prof. Dr. Glauco de Souza Rolim

Coorientador: Prof. Dr. Cristiano Zerbato

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Agronomia (Produção Vegetal).

2024

P659p

Pinto, Antonio Alves

Previsão e estimativa da produtividade de milho e cana-de-açúcar utilizando Inteligência artificial e sensoriamento remoto / Antonio Alves Pinto. -- Jaboticabal, 2024
82 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Glauco de Souza Rolim
Coorientador: Cristiano Zerbato

1. Agricultura digital. 2. Sensoriamento orbital. 3. Modelagem da produtividade. 4. Agricultura de precisão. 5. PCA. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa

Esta pesquisa desempenha um papel importante no impulsionamento da utilização de técnicas de aprendizado de máquina na agricultura, ao apresentar novas abordagens metodológicas para a previsão da produtividade de milho (*Zea mays* L.) e estimativa da produtividade da cana-de-açúcar (*Saccharum officinarum* L.).

Potential impact of this research

This research plays an important role in boosting the use of machine learning techniques in agriculture, by presenting new methodological approaches for predicting corn (*Zea mays* L.) productivity and estimating sugarcane (*Saccharum officinarum*) productivity L.).



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



CERTIFICADO DE APROVAÇÃO

**TÍTULO DA TESE: PREVISÃO E ESTIMATIVA DA PRODUTIVIDADE DE MILHO E CANA-DE-AÇÚCAR
UTILIZANDO INTELIGÊNCIA ARTIFICIAL E SENSORIAMENTO REMOTO**

AUTOR: ANTONIO ALVES PINTO

ORIENTADOR: GLAUCO DE SOUZA ROLIM

COORIENTADOR: CRISTIANO ZERBATO

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia (Produção Vegetal), pela Comissão Examinadora:

Prof. Dr. GLAUCO DE SOUZA ROLIM (Participação Virtual)
Departamento de Engenharia e Ciências Exatas / FCAV UNESP Jaboticabal

Prof. Dr. LUAN PEREIRA DE OLIVEIRA (Participação Virtual)
Department of Horticulture / University of Georgia (UGA) - Tifton/EUA



Documento assinado digitalmente
LUAN PEREIRA DE OLIVEIRA
Data: 22/04/2024 22:29:03-0300
Verifique em <https://validar.it.gov.br>

Prof. Dr. LUCAS EDUARDO DE OLIVEIRA APARECIDO (Participação Virtual)
Departamento de Engenharia Agrícola/Agrometeorológica/Climatologia Agrícola / Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - Muzambinho/MG

Prof. Dr. EDUARDO NARDINI GOMES (Participação Virtual)
Departamento de Engenharia Agrônômica / FCAVR UNESP Registro



Documento assinado digitalmente
EDUARDO NARDINI GOMES
Data: 24/04/2024 11:14:13-0300
Verifique em <https://validar.it.gov.br>

Prof. Dr. LUIZ FABIANO PALARETTI (Participação Virtual)
Departamento de Engenharia e Ciências Exatas / FCAV UNESP Jaboticabal

Jaboticabal, 05 de abril de 2024



Documento assinado digitalmente
LUCAS EDUARDO DE OLIVEIRA APARECIDO
Data: 25/04/2024 15:00:42-0300
Verifique em <https://validar.it.gov.br>

DADOS CURRICULARES DO AUTOR

ANTONIO ALVES PINTO - Natural de Jardim - CE, nascido em 1º de julho de 1995, filho de Zenicleide Alves dos Santos e Luís Antonio Pinto. Formado em Engenharia Agrônômica pela Universidade Federal do Cariri - UFCA, no ano de 2018, desenvolveu durante a graduação trabalhos na área de fitotecnia, nutrição de plantas, máquinas e mecanização agrícola, sob orientação da Prof. Dr. Felipe Thomaz da Camara. Em março de 2019 iniciou o curso de mestrado no Programa de Pós-Graduação em Agronomia (Ciência do Solo), na Universidade Estadual Paulista “Júlio de Mesquita Filho” - Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal, São Paulo, desenvolvendo pesquisas na área de engenharia de água e solo, sob orientação do Prof. Dr. Afonso Lopes. A dissertação foi defendida em julho de 2020. No mês seguinte deu início ao doutorado em Agronomia – Produção vegetal pela Universidade Estadual Paulista “Júlio de Mesquita Filho” – Unesp, Câmpus Jaboticabal, sob a orientação do professor Dr. Glauco de Souza Rolim.

“Escale todas as montanhas, cruze todos os rios, siga todos os arco-íris, até encontrar seus sonhos”

Rodgers e Hammersein

Dedico...

A meus pais, meus irmãos e aos meus amigos pelo amor incondicional, e por todo o incentivo para chegar até aqui.

AGRADECIMENTOS

A Deus, primeiramente pelo dom da vida.

Aos meus pais Zenicleide e Luís que sempre estiveram presentes me apoiando e sonhando junto comigo.

Aos meus irmãos por sempre acreditarem em mim.

Aos meus amigos Luís Fernando, José Lucas, João de Deus, pelo apoio, amizade e por todos os momentos que passamos juntos.

À Tamires Doroteo pelo apoio e companheirismo.

Ao Professor Dr. Glauco de Souza Rolim, pela orientação, ensinamentos nas atividades desenvolvidas, amizade concedida e colaboração no trabalho.

Ao Professor Dr. Cristino Zerbato, pela coorientação, ensinamentos nas atividades desenvolvidas, apoio emocional e amizade concedida.

À Universidade Estadual Paulista “Júlio de Mesquita Filho” e à Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal – UNESP/FCAV, pela oportunidade de realização do curso.

Ao corpo docente do Programa de Pós-Graduação em Agronomia (Produção Vegetal), pelos conhecimentos transmitidos no decorrer do curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por conceder a bolsa de estudo.

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

A todos que de alguma forma contribuíram para a realização deste trabalho.

Obrigado!

SUMÁRIO

CAPÍTULO 1: CONSIDERAÇÕES GERAIS.....	11
1. INTRODUÇÃO	11
2. HIPÓTESE	12
3. OBJETIVOS	13
4. REVISÃO BIBLIOGRÁFICA	13
4.1. Importância econômica do milho	13
4.1.1. Fenologia do milho.....	14
4.1.2. Sensoriamento remoto no milho	15
4.1.3. Aprendizado de máquinas no milho	17
4.1.4. Importância econômica da cana-de-açúcar	20
4.1.5. Fenologia da cana-de-açúcar	20
4.1.6. Sensoriamento remoto na cana-de-açúcar	22
4.1.7. Aprendizado de máquinas na cana-de-açúcar	24
5. REFERÊNCIAS BIBLIOGRÁFICAS	25
CAPÍTULO 2: PREVISÃO DE PRODUTIVIDADE DO MILHO UTILIZANDO SENSORIAMENTO REMOTO, SOMA DE GRAUS-DIA E MODELOS DE APRENDIZADO DE MÁQUINA.....	32
1. INTRODUÇÃO	34
2. MATERIAL E MÉTODOS	37
2.1. Local de estudo	37
2.1. Obtenção e processamento de imagens de satélite	38
2.2. Variáveis de entradas dos modelos	41
2.3. Aquisição dos dados de produtividade	43
2.4. Modelos de aprendizado de máquina	44
2.5. Análise de desempenho dos modelos	44
3. RESULTADOS	45
4. DISCUSSÃO	49
5. CONCLUSÃO.....	52
6. REFERÊNCIAS.....	53

CAPÍTULO 3: ESTIMATIVA DA PRODUTIVIDADE DE CANA-DE-AÇÚCAR
UTILIZANDO APRENDIZADO DE MÁQUINA E ANÁLISE DE COMPONENTES
PRINCIPAIS DOS DADOS ESPECTRAIS E DO MOTOR.....57

1. INTRODUÇÃO	59
2. MATERIAL E MÉTODOS	60
2.1. Local de estudo	60
2.2. Obtenção e processamento de imagens de satélite	61
2.3. Aquisição dos dados de produtividade e do motor	63
2.4. Variáveis de entradas dos modelos	64
2.5. Modelos de aprendizado de máquina	65
2.6. Análise de desempenho dos modelos	66
3. RESULTADOS	67
3.1. Estatística descritiva dos dados da colhedora e produtividade.....	67
3.2. Seleção de variáveis preditoras PCAs.....	68
3.3. Avaliação de Precisão e Acurácia dos modelos.	70
4. DISCUSSÃO	72
5. CONCLUSÃO.....	74
6. REFERÊNCIAS.....	75

Previsão e estimativa da produtividade de milho e cana-de-açúcar utilizando Inteligência artificial e sensoriamento remoto

RESUMO

A necessidade de avanços nas técnicas de estimativa e previsão da produtividade de cultura agrícolas utilizando técnicas não destrutivas tem impulsionado o desenvolvimento de novas pesquisas utilizando técnicas de aprendizado de máquina associado a dados obtidos por sensoriamento remoto. A estimação é o cálculo de uma condição atual, enquanto, a previsão é a projeção de uma condição futura. Buscando contribuir com os avanços na estimativa e previsão da produtividade utilizando técnicas não destrutivas, foram desenvolvidos experimentos para os cultivos de milho e cana-de-açúcar. No primeiro capítulo apresenta-se o referencial teórico que aborda desde a importância econômica, fenologia, aplicações de sensoriamento remoto até o uso de aprendizado de máquina no cultivo de milho e na cana-de-açúcar. No segundo capítulo é apresentado os resultados obtidos a partir de uma abordagem baseada em classes de graus dias acumulados do cultivo de milho para organizar os dados das imagens de satélite em duas safras em que o modelo extreme gradient boosting (XGBoost), obteve acurácia superior ao k-nearest neighbors (KNN) na previsão da produtividade do cultivo de milho. No terceiro capítulo, foi realizado um experimento no cultivo de cana-de-açúcar com o objetivo de investigar a análise de componentes principais e integração de séries temporais de imagens do satélite Sentinel-2 e variáveis do motor, como entradas em modelos de aprendizado de máquina para estimar a produtividade da cultura, concluindo-se que o modelo Support Vector Machine (SVM), utilizando as Análise de Componentes Principais (PCAs) da integração de séries temporais de bandas espectrais e dados do motor apresentou maior acurácia, com RMSE de 4,20 t ha⁻¹.

Palavras-chave: Agricultura digital, sensoriamento orbital, modelagem da produtividade, agricultura de precisão, análise de componentes principais.

Forecasting and estimation of corn and sugarcane yield using Artificial Intelligence and remote sensing

ABSTRACT

The need for advanced techniques for estimating and forecasting agricultural crop productivity using non-destructive techniques has driven the development of new research using machine learning techniques associated with data obtained by remote sensing. Estimation is the calculation of a current condition, while forecasting is the projection of a future condition. Experiments were developed for corn and sugar cane crops to contribute to advances in productivity estimation and forecasting using non-destructive techniques. The first chapter presents the theoretical framework where we address the economic importance, phenology, remote sensing applications, and the use of machine learning in corn and sugar cane. The second chapter presents the results obtained from an approach based on classes of accumulated degree days of the corn crop to organize satellite image data in two harvests in which the XGBoost model achieved greater accuracy than KNN in predicting yield. In the third chapter, an experiment was carried out in the sugarcane crop to investigate the analysis of principal components and integration of time series of Sentinel-2 images and engine variables, as inputs in machine learning models. to estimate crop productivity, concluding that the SVM model using PCAs from the integration of time series of spectral bands and engine data presented greater accuracy, with RMSE of 4.20 t ha⁻¹.

Keywords: Digital agriculture, orbital sensing, productivity modeling, precision agriculture, principal component analysis.

CAPÍTULO 1: CONSIDERAÇÕES GERAIS

1. INTRODUÇÃO

Esta tese foi estruturada em quatro capítulos, abordando as técnicas aprendizado de máquina associado ao sensoriamento remoto orbital na previsão e estimativa da produtividade de cultivo de milho e cana-de-açúcar, respectivamente. Esta pesquisa desempenha um papel importante no impulsionamento da utilização de técnicas de aprendizado de máquina na agricultura, ao apresentar novas abordagens metodológicas para a previsão da produtividade de cultivo de milho e estimativa da produtividade da cana-de-açúcar.

Os avanços das técnicas não destrutivas utilizadas nas previsões e estimativas de fatores agronômicos das plantas são fundamentais no entendimento das relações entre esses fatores, nesse sentido, o sensoriamento possui uma diversidade de sensores e níveis de aquisição, além disso, não necessita de contato direto com os alvos e nem causa danos nos cultivos durante a obtenção dos dados. Dessa forma, pesquisadores têm buscado entender a relações das informações espectrais com os fatores agronômicos das plantas, utilizando aprendizado de máquina para lidar com o volume e diversidade dos dados provenientes do sensoriamento.

O Primeiro capítulo é constituído pelo referencial teórico que suporta as hipóteses de que o sensoriamento remoto associado a modelos de aprendizado de máquina tem potencial para gerar modelo preditivos acurados, esse capítulo aborda desde a importância econômica dos cultivos, passando pela fenologia, aplicações de sensoriamento remoto até o uso de aprendizado de máquina no cultivo de milho e na cana-de-açúcar.

O segundo capítulo apresenta os resultados obtidos a partir de uma abordagem baseada em classes de graus dias acumulados do cultivo de milho para organizar os dados das imagens de satélite em duas safras, utilizando posteriormente esses dados como variáveis de entradas em modelos de aprendizado de máquina para prever a produtividade do cultivo de milho antes da colheita.

No terceiro capítulo, utilizamos como variáveis de entradas em modelos de aprendizado de máquina para estimar a produtividade da área cultivada com cana-de-açúcar, uma abordagem baseada na análise de componentes principais e integração de séries temporais de imagens do Sentinel-2 com parâmetros do motor.

2. HIPÓTESE

O desenvolvimento desta tese baseou-se em duas hipóteses, descritas a seguir:

1. O agrupamento séries temporais de imagens do Sentinel-2 de duas safras com base nas classes de graus dias acumulados no cultivo de milho, associado a modelos de aprendizado de máquina, podem ser empregados para prever a produtividade.
2. Análise de componentes principais como estratégia de redução de dimensionalidade de dados oriundos da integração de séries temporais de imagens do Sentinel-2 e parâmetros do motor, como variáveis de entradas em modelos de aprendizado de máquina podem aumentar a acurácia na estimativa da produtividade de cana-de-açúcar.

3. OBJETIVOS

Prever, o mais antecipadamente possível, a produtividade do milho utilizando algoritmos de aprendizado de máquinas, associados a dados espectrais agrupados com base nos graus dias acumulados do milho.

Estimar a produtividade da cana-de-açúcar com base na análise de componentes principais de dados espectrais e parâmetros do motor.

4. REVISÃO BIBLIOGRÁFICA

4.1. Importância econômica do milho

A planta do milho (*Zea mays* L.) é uma das graminhas mais cultivadas no mundo, apresentando uma ampla utilização no consumo para alimentação animal bem com na alimentação humana. Dados da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) relatam que o milho consumido na cadeia produtiva de suínos e aves representa 70% do milho produzido no mundo e entre 70% e 80% do milho produzido no Brasil (Embrapa, 2021).

Os maiores produtores são os Estados Unidos, China, Brasil e União Europeia, somando mais de 70% da produção mundial de milho (CONAB, 2024a). No Brasil o milho é o segundo grão mais cultivado, com aproximadamente 22,2 milhões de hectares na safra 2022/23, produzindo cerca de 131,9 milhões de toneladas, representando aumento 16,6%, quando comparando à safra anterior, sendo a produtividade média de 5,6 toneladas por hectare, conforme estimado pela Companhia Nacional de Abastecimento (CONAB, 2024a).

Para a safra 2023/24 espera-se uma redução de 5,6% na área cultivada, 5,6% na produtividade e 10,9 % na produção total (Tabela 1), tais reduções deve-se

principalmente influência das condições climáticas desfavoráveis para a maioria das regiões do Brasil (CONAB, 2024a).

Tabela 1 - Balanço de oferta e demanda do milho em mil t.

	Safras				
	2019/20	2020/21	2021/22	2022/23	2023/24
Estoque Inicial	13.186,6	15.312,1	13.515,3	8.095,9	5.940,7
Produção	102.586,4	87.096,8	113.130,4	131.945,9	117.603,5
Importação	1.453,4	3.090,7	2.615,1	1.500,0	2.100,0
Suprimento	117.226,4	105.499,6	129.260,8	141.541,8	125.644,2
Consumo	67.021,4	71.168,6	74.534,6	79.601,1	84.367,4
Exportação	34.892,9	20.815,7	46.630,3	56.000,0	35.000,0
Estoque Final	15.312,1	13.515,3	8.095,9	5.940,7	6.276,8

Adaptado da CONAB, (2024).

Com base no resumo do balanço de oferta e demanda de milho no Brasil (Tabela 1), a Conab projeta um aumento 2,1 milhões de toneladas no volume de importação do milho na safra 2023/24, devido a perspectiva de redução da produção nacional. Já para as exportações, as projeções são de redução do volume em 37,5% quando comparado a safra 2022/23.

4.1.1. Fenologia do milho

O milho é nativo das Américas central, originado a partir da domesticação do teosinto a cerca de 10 mil anos atrás. A cultura pode ser cultivada nos mais variados ambientes e climas (Oliveira, 2021).

A fenologia do milho é dividida em duas fases (Tabela 2), a fase vegetativa corresponde ao período desde a emergência (VE), passando pelos estádios vegetativos de V1 a Vn e finalizado com o pendoamento VT. A segunda fase é a

reprodutiva, inicia com o estágio reprodutivo R1 correspondente ao aparecimento das bonecas do milho (posteriormente serão as espigas) e vai até o R6 correspondente ao período de maturação fisiológica.

Tabela 2 - Fenologia do milho, estádios vegetativos e reprodutivos.

Fase Vegetativa		Fase Reprodutiva	
Estágio	Descrição	Estágio	Descrição
VE	Emergência	R1	Embonecamento
V1	1ª Folha desenvolvida	R2	Grão Bolha d'água
V2	2ª Folha desenvolvida	R3	Grão leitoso
V3	3ª Folha desenvolvida	R4	Grão pastoso
V4	4ª Folha desenvolvida	R5	Formação de dente
V(n)	nª Folha desenvolvida	R6	Maturação fisiológica
VT	Pendoamento	-	-

Adaptado de (Embrapa, 2021)

O ciclo vegetativo do milho é variado, com genótipos extremamente precoces (superprecoce) em que pode ocorrer a polinização 30 dias após a emergência, até os tardios que podem alcançar 300 dias. No entanto, as cultivares utilizadas em território nacional normalmente variam entre 110 e 160 dias (Fancelli, 2015).

4.1.2. Sensoriamento remoto no milho

O sensoriamento remoto (SR) é uma ferramenta não destrutiva que faz uso de sensores para obter informações de um ou mais alvos sem a necessidade de contato físico direto entre o sensor e o alvo (WEISS et al., 2020). O sensoriamento remoto baseia-se em propriedades específicas que cada objeto tem ao interagir com a radiação eletromagnética (REM). Ao entrar em contato com o alvo, a REM pode ser absorvida, refletida ou transmitida (Molin et al., 2015). A interação com REM possibilita

que as propriedades específicas do objeto sejam visualizadas na forma de assinatura espectral, na figura 1 podemos observar um exemplo da assinatura espectral do solo, planta saudável e planta não saudável.

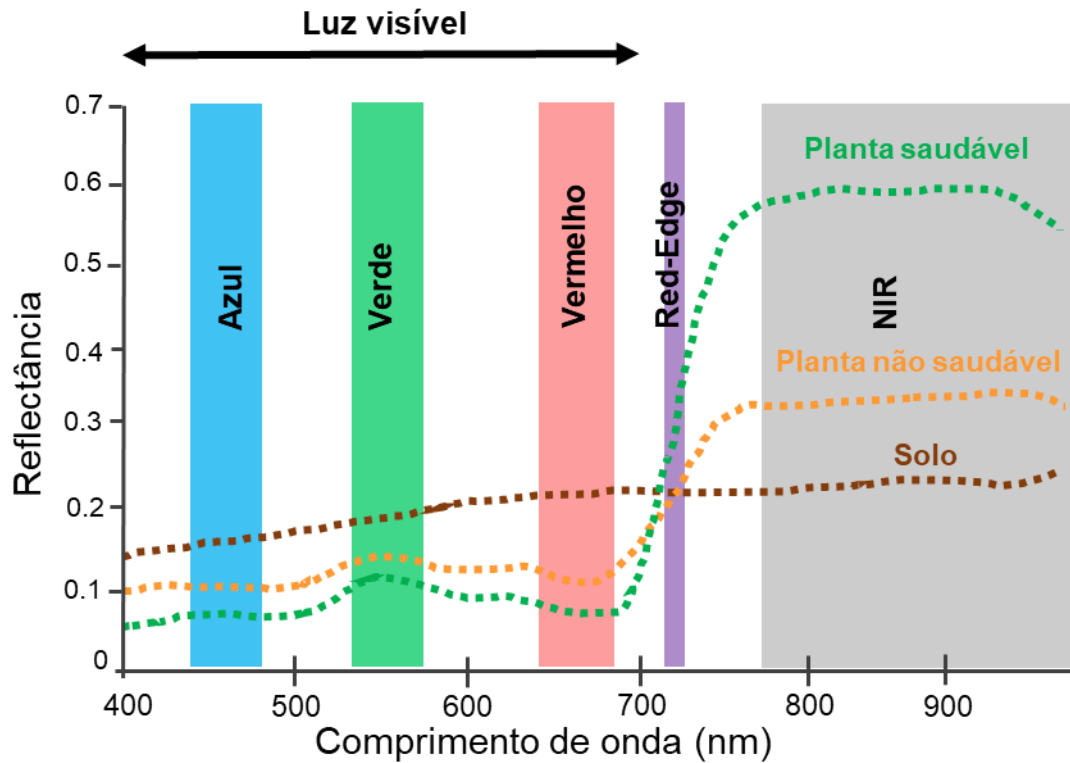


Figura 1 - Reflectância típica do solo, planta saudável e planta sob estresse. Adaptado de Burud et al., (2017).

Na agricultura as principais aplicações do SR são no monitoramento de uso do solo, otimização da produção, prestação de serviços ecossistêmicos relacionados aos recursos hídricos e solo, monitoramento e conservação da biodiversidade, previsão e estimativa da produção agrícola, estimativa de parâmetros biofísicos como altura de planta e diâmetro do dossel. Mais recentemente o SR tem sido empregado no auxílio ao desenvolvimento de novas cultivares por meio da fenotipagem de alto rendimento (WEISS et al., 2019).

Dentre as aplicações do SR encontrado na literatura, para a cultura do milho temos alguns trabalhos que utilizaram dados espectrais sensores proximal, imagens

de satélite e/ou veículo aéreo não tripulado (VANT) e índices de vegetação topográficos para estimativa e previsão da produtividade de milho, relatando êxito nos resultados obtidos (Abbas et al., 2020; Barzin et al., 2020; Oliveira et al., 2022; Pinto et al., 2022).

Outros exemplos que podem ser encontrados na literatura sobre a aplicação SR no milho é a detecção de doenças (Barbedo, 2018; Xhang et al., (2018); Wu et al., 2019; Darwish et al., 2020), manejo de cultura com a detecção de ervas daninhas (Gao et al., 2018; Louargant et al., 2018; Champ et al., 2020), reconhecimento e classificação da cultura (Ji et al., 2018; Zhong et al., 2019), e qualidade de sementes (Romero et al., 2018).

4.1.3. Aprendizado de máquinas no milho

Os avanços tecnológicos ocorridos no campo nas últimas décadas resultaram no aumento massivo da quantidade de dados coletados e armazenados (*big data*), contudo, isso também impulsionou o uso de novas técnicas capazes de analisar e aprender com tais dados, extraindo informações que podem auxiliar na cadeia produtiva das mais diversas culturas. Uma dessas técnicas é o Aprendizado de Máquinas ou *Machine Learning* (ML), que pode ser definido como ramo da inteligência artificial com a capacidade de aprender sem serem estritamente programados (Liakos, et al., 2018).

Conforme a habilidade de aprendizagem, os algoritmos de ML podem ser classificados como aprendizagem supervisionada, não supervisionada e por reforço. Na aprendizagem supervisionada podemos destacar os modelos que realizam a classificação e regressão (Bali e Singla, 2022), como exemplo de algoritmos podemos

destacar a Regressão Linear Multipla (RLM), Random Forest (RF), Support Vector Machine (SVM) e k-nearest neighbors (KNN), (Figura 2).

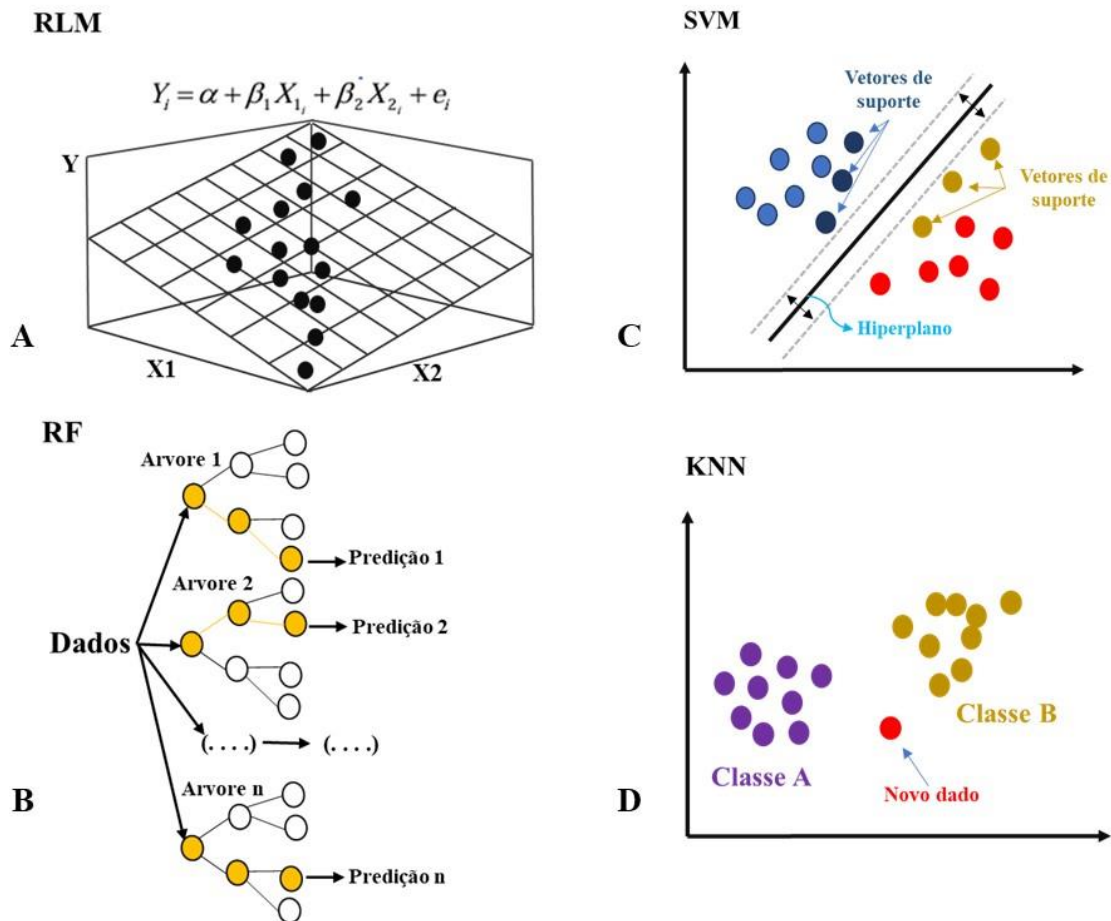


Figura 2 – Estrutura de alguns algoritmos aprendizado de máquina; A - RLM, B - RF, C - SVM, D - KNN. Símbolos: RLM, Regressão Linear Múltipla, RF, Random Forest, SVM, Support Vector Machine, KNN, K-Nearest Neighbors.

Na aprendizagem não supervisionada temos como exemplo Algoritmos de clusterização e redução de dimensionalidade com os algoritmos K-means e análise de componentes principais, respectivamente. Já no aprendizado por reforço, modelos de decisões em tempo real são agrupados nessa categoria (Bali e Singla, 2022).

No milho os modelos de aprendizado de máquinas foram aplicados nas mais diversas áreas como detecção de doenças na cultura, detecção de ervas daninhas e qualidade de sementes (Darwish et al., 2020; Champ et al., 2020; Romero et al., 2018).

Outra aplicação dos modelos de ML no milho é estudo da variabilidade espaço-temporal da produtividade, associando dados espectrais como variáveis preditoras nos modelos de ML (Kayad et al., 2019; Abbas et al., 2020; Pinto et al., 2022).

Ao estudarem o uso de imagens espectrais de um veículo aéreo não tripulado com variáveis preditoras em sete modelos de ML, Ramos et al, (2020) concluíram que o algoritmo RF obteve bons resultados na estimativa da produtividade do milho, sendo o modelo mais acurado com R^2 de 0,78 e MAE de 853,11 kg ha⁻¹.

Embora o uso de imagens espectrais de VANT associadas aos modelos de ML sejam promissoras, na modelagem da produtividade esse ainda é um recurso limitado para poucos pesquisadores, bem como apresenta limitações para utilização em grandes áreas, sendo necessário recorrer aos dados espectrais de satélites (Kayad et al., 2019; Pinto et al., 2022).

Vale ressaltar que, o uso de imagens de satélites para estimativa ou previsão de parâmetros agronômicos apresentam algumas limitações, dentre elas as principais são a ocorrência de nuvens no momento do imageamento e a dificuldade de organização de dados de series temporais usando critérios como dias após o plantio ou estádios fenológicos da cultura. O satélite Sentinel-2 tem variabilidade temporal de cinco dias, porém, caso esteja nublado no momento do imageamento será necessário esperar o próximo imageamento que também estará sujeito a perda por problemas com nuvem. Para culturas de ciclo curto a perda de uma imagem pode significar a obtenção de dados espectrais em estádios fenológicos diferentes entre uma safra e outra.

4.1.4. Importância econômica da cana-de-açúcar

A cana-de-açúcar é uma gramínea do gênero *Saccharum* utilizada como matéria prima para diversos produtos, tais como açúcar, etanol, bioenergia com a queima do bagaço e/ou a produção de biogás, bem como fertilizantes orgânicos a partir da torta de filtro e vinhaça (Oliveira et al., 2022).

O Brasil é o maior produtor mundial de cana-de-açúcar, com aproximadamente 8,4 milhões de hectares na safra 2023/24, com uma produtividade média de 81,1 toneladas por hectare e produzindo cerca de 677,6 milhões de toneladas, que foram usadas como matéria prima para a produção de aproximadamente 46,9 milhões de toneladas de açúcar e 27, 7 bilhões de litros de etanol, conforme estimado pela Companhia Nacional de Abastecimento (CONAB, 2024b).

Quanto a exportação a CONAB relata que de abril a outubro de 2023 foram exportados 18,8 milhões de toneladas de açúcar representando um aumento de 9,7% em comparação com o exportado na safra anterior no mesmo período, sendo que mais de 27% do total exportado foram para três países, sendo em primeiro lugar a China com uma compra de US\$ 1,2 bilhão em açúcar, seguido da Índia e Arábia Saudita com US\$ 809 e 564 milhões, respectivamente (CONAB, 2024b).

4.1.5. Fenologia da cana-de-açúcar

A cana-de-açúcar é uma planta semiperene, originária das regiões temperadas quentes e tropicais do Sudeste Asiático e da Nova Guiné, sendo atualmente produzida em regiões tropicais e subtropicais como no Brasil (Som-ard et al., 2021).

Em relação a arquitetura da planta, a cana-de-açúcar apresenta um caule fibroso, articulado e robusto que acumula sacarose nos entrenós com altura podendo

variar de 1,8 a 5 m, dependendo de fatores como variedade, manejo, clima, solo e outros (Som-ard et al., 2021).

Os estádios fenológicos da cana-de-açúcar são divididos em quatro fases (Figura 3), a Inicial (brotação e emergência), perfilhamento, desenvolvimento (crescimento dos colmos) e maturação dos colmos (fase de maturação inicial, terço médio e maturação final antes da colheita), conforme descrito por Molijn et al., (2019).

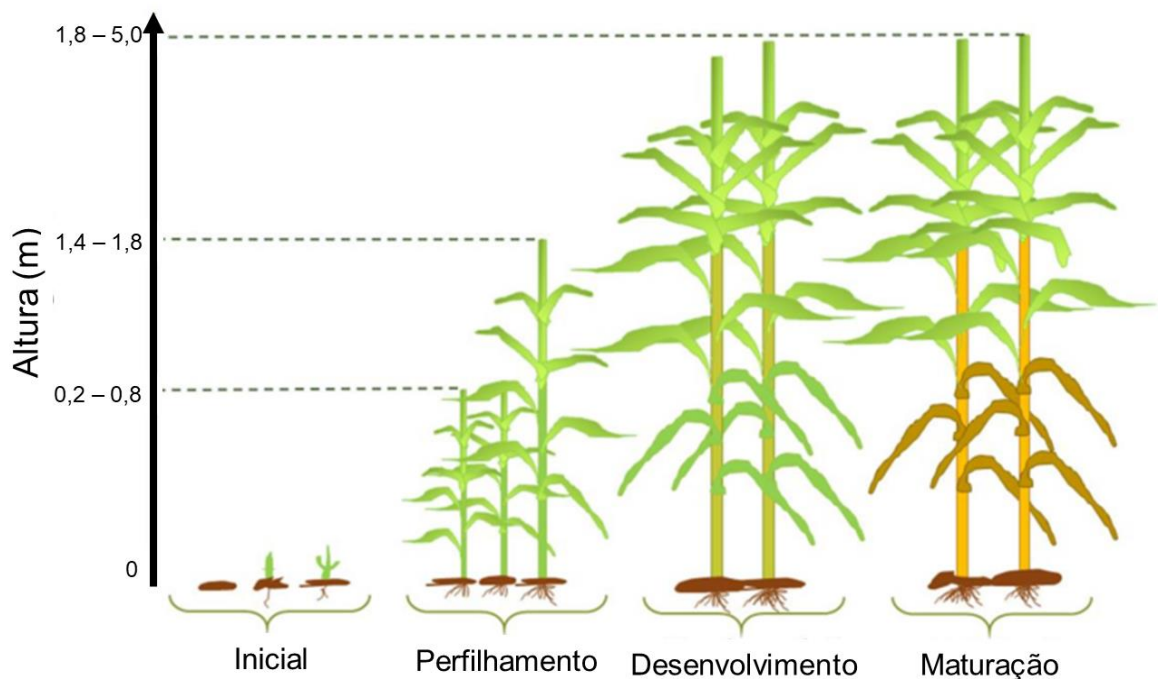


Figura 3. Estádios fenológicos da cana-de-açúcar. Adaptado de Molijn et al., (2019).

O ciclo da cultura pode variar de 12 a 22 meses, dependendo da variedade, das condições de cultura local e dos parâmetros geográficos, podendo ser classificada como cana de ano e meio (14 a 22 meses), cana de ano e cana soca (12 meses), e cana de inverno com duração de 12 a 16 meses (Prasara et al., 2016; Sukyai et al., 2016; Som-ard et al., 2021).

4.1.6. Sensoriamento remoto na cana-de-açúcar

Na cana-de-açúcar a assinatura espectral (Figura 4) é determinada pelos mesmos fatores que impulsionam as propriedades espectrais de outros tipos de vegetação; refletância solo ao fundo, morfologia da planta e coeficientes de absorção, dispersão e transmissões foliares (Everingham et al., 2007; Hamzeh et al., 2016; Somard et al., 2021).

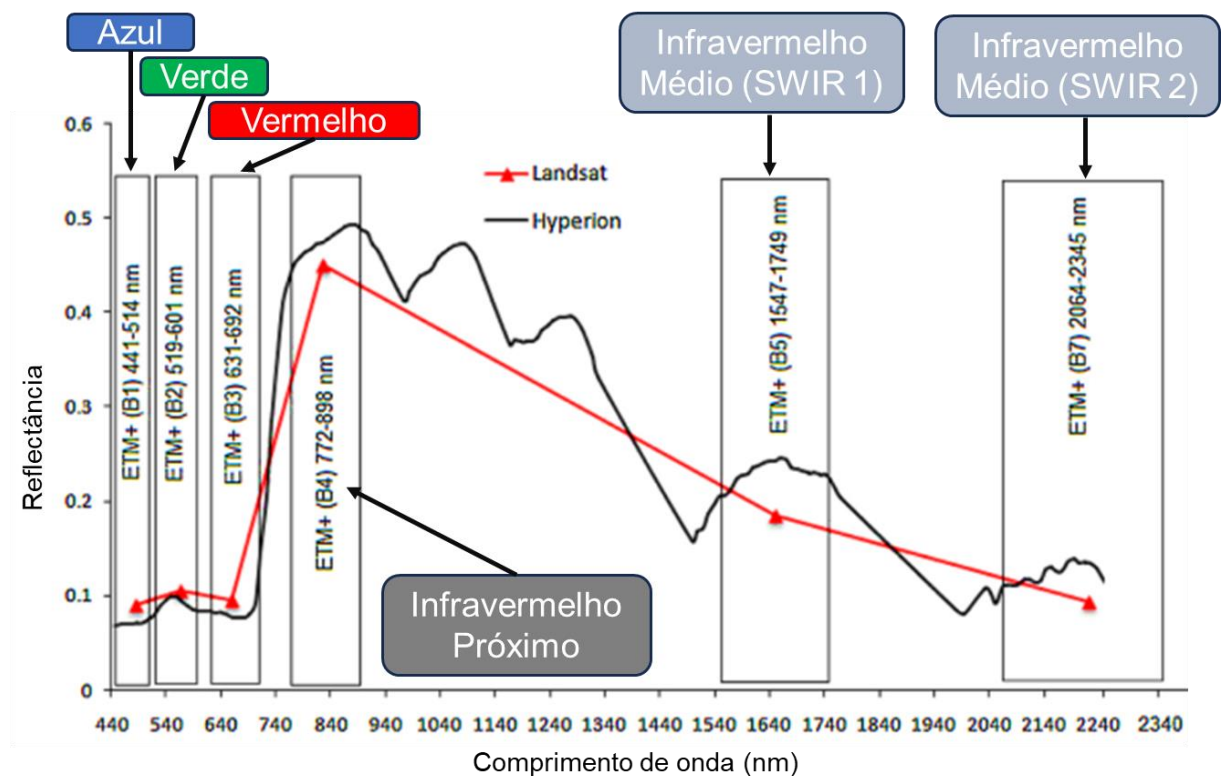


Figura 4. Assinatura espectral da cana-de-açúcar, comparação entre imagens Hyperion e Landsat 7 ETM+. Adaptado de Hamzeh et al., 2016. Símbolos: ETM+B1, Banda do azul, ETM+B2 Banda do verde, ETM+B3, Banda do vermelho, ETM+B4, Banda do infravermelho Próximo, ETM+B5, Banda do infravermelho médio (SWIR 1), ETM+B7, Banda do infravermelho médio (SWIR 2).

A assinatura espectral da cana-de-açúcar originada a partir de satélite, veículos aéreos não tripulados e sensores proximais, fornece informações que permitem analisar ocorrência de doenças, estresse, desenvolvimento de biomassa, pigmentos foliares, conteúdo de clorofila e manejo da cultura (Hamzeh et al., 2016;

Vargas et al., 2016). Ao estudarem o comportamento da assinatura espectral da cana-de-açúcar quando a acometida por *Diatraea saccharalis* Vargas et al., (2016), relataram que na região do visível as plantas com *Diatraea saccharalis* apresentaram maior percentual de refletância.

Pesquisadores como Apan et al., (2004) e Bégué et al., (2010), estabeleceram uma relação entre a assinatura espectral de plantas afetadas por ferrugem alaranjada e plantas saudáveis, utilizando a análise discriminante nos estudos eles relataram que as plantas acometidas pela ferrugem alaranjada apresentaram menor refletância na região do vermelho e NIR, concluindo que assinatura espectral pode ser utilizada na detecção de anomalias nas plantas causadas pela ferrugem alaranjada.

Outras aplicações de sucesso da assinatura espectral da cana-de-açúcar foram na identificação de estresse salino (Hamzeh et al., 2013; Hamzeh et al., 2016), desenvolvimento de biomassa e status de nitrogênio (Abdel-Rahman et al., 2010; Lofton et al., 2012; Amaral et al., 2015 e Miphokasap e Wannasiri 2018) e mapeamento da produtividade conforme estudo conduzido por Molijn et al., (2019), ao utilizarem as bandas C e L do SAR (Synthetic Aperture Radar) e imagens ópticas de satélite no mapeamento da produtividade da cana-de-açúcar. Outro estudo conduzido por Pinheiros et al., (2018), estimou a produtividade da cana-de-açúcar utilizando como base a concentração de nutrientes no tecido foliar e NDVI obtido com sensor proximal.

Diante da crescente acessibilidade a dados provenientes de satélites, destacando-se pelo aumento da resolução temporal, espacial e espectral, nos deparamos com dados mais precisos, que podem fornecer informações essenciais

para auxiliar manejo das culturas agrícolas, buscando a melhoria da produção e da produtividade (Som-ard et al., 2021).

4.1.7. Aprendizado de máquinas na cana-de-açúcar

Na cana-de-açúcar as principais aplicações dos modelos de aprendizado de máquinas foram para mapeamento da cultura em comparação com outras formas de uso do solo, com 52 artigos publicados de 2006 a 2020 abordando o tema classificação, conforme levantamento realizado por Som-ard et al., 2021.

Pesquisadores como Luna et al., (2016); Cevallos et al., (2019); Jiang et al., (2019) Wang et al., (2019) e Wang et al., (2020), são exemplos de alguns dos estudos que utilizaram sensoriamento remoto e inteligência artificial na classificação da cana-de-açúcar em comparação com outras formas de uso do solo, usando como classificadores os algoritmos Random Forest, Support Vector Machine, Artificial Neural Network, Convolutional Neural Network, Decision Tree e Extreme Gradient Boosting.

Além da classificação, outras aplicações do aprendizado de máquina em cana-de-açúcar encontrados na literatura foram na estimativa de parâmetros biométricos (número de perfilho, altura de planta e diâmetro do colmo), conteúdo de clorofila do dossel e parâmetros de qualidade (Brix e pureza), utilizando dados espectrais coletados com VANT, trabalhos desenvolvidos por Oliveira et al., (2022), Narmilan et al., (2022) e Barbosa Júnior et al., (2023), respectivamente.

Em relação a estimativa da produtividade destacamos os trabalhos desenvolvido com dados coletados com sensor LiDAR (Light Detecting and Ranging) e sensores espectrais acoplados em VANT (Xu et al., 2020; Wang et al., 2022 e Akbarian et al., 2023), dados fenológico, climáticos e espectrais de satélite a nível de

bloco (Dimov et al., 2022; Nihar et al., 2022; Zhu et al., 2023 e Alemán-Montes et al., 2023), parâmetros do motor como variáveis preditoras (Lima et al., 2021; Maldaner et al., 2021). Já a nível de campo e com dados espectrais de satélite, Canata et al., (2021), destaca-se como sendo a pioneira nessa abordagem de estimativa de produtividade da cana-de-açúcar.

Quando comparamos a cana-de-açúcar com outros tipos de cultivos, observamos que a cana-de-açúcar possui padrão de desenvolvimento e fenologia diferente, com alta concentração de biomassa por hectares e altura podendo variar de 1,8 a 5 m, tais características dificultam o desenvolvimento de estudos que necessitam de dados coletados em contato com a cultura no campo, como por exemplo coletar a mostra de cana para estimativa de produtividade e altura (Som-ard et al., 2021). Dessa forma, um emprego de dados de sensoriamento remoto aliados a dados coletados dos equipamentos envolvidos na colheita da cana-de-açúcar podem ser uma alternativa viável no entendimento da variabilidade da produtividade dos canaviais.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. **Agronomy** 10: 1046. <https://doi.org/10.3390/agronomy10071046>.

Abdel-Rahman EM, Ahmed FB, Berg, MVD. (2021) Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy. **International Journal of Applied Earth Observation and Geoinformation**, 12, S52–S57. <https://doi.org/10.1016/j.jag.2009.11.003>.

Alemán-Montes B, Zabala A, Henríquez C, Serra P (2023) Modelling Two Sugarcane Agro-Industrial Yields Using Sentinel/Landsat Time-Series Data and Their Spatial Validation at Different Scales in Costa Rica. **Remote Sensing**, 15, 5476. <https://doi.org/10.3390/rs15235476>

Akbarian S, Jamnani MR, Xu C, Wang W, Lim S (2023) Plot level sugarcane yield estimation by machine learning on multispectral images: A case study of Bundaberg,

Australia. **Information Processing in Agriculture**, 1-11. <https://doi.org/10.1016/j.inpa.2023.06.004>.

Amaral LR, Molin JP, Portz G, Finazzi FB, Cortinove L (2015) Comparison of crop canopy reflectance sensors used to identify sugarcane biomass and nitrogen status. **Precision Agriculture**, 16, 15–28. <https://doi.org/10.1007/s11119-014-9377-2>

Apan A, Held A, Phinn S, Markley J (2004) Detecting sugarcane ‘orange rust’ disease using EO-1 Hyperion hyperspectral imagery. **International Journal of Remote Sensing**, 25, 489–498. <https://doi.org/10.1080/01431160310001618031>.

Bali N, Singla A (2022) Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey. **Archives of Computational Methods in Engineering**, 29, 95–112. <https://doi.org/10.1007/s11831-021-09569-8>.

Barzin R, Pathak R, Lotfi H, Varco J, Bora GC (2020) Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn. **Remote Sensing** 12: 2392. DOI: <https://doi.org/10.3390/rs12152392>.

Barbedo JGA (2018) Factors influencing the use of deep learning for plant disease recognition. **Biosystems Engineering**, 172, 84–91. <https://doi.org/10.1016/j.biosystemseng.2018.05.013>.

Barbosa Júnior MR, Moreira BRA, Oliveira RP, Shiratsuchi LS, Silva RP (2023) UAV imagery data and machine learning: A driving merger for predictive analysis of qualitative yield in sugarcane. **Frontiers Plant Science**, 14, 1114852. <https://doi.org/10.3389/fpls.2023.1114852>.

Bégué A, Lebourgeois V, Bappel E, Todoroff P, Pellegrino A, Baillarin, F, Siegmund B (2010) Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. **International Journal of Remote Sensing**, 31, 5391–5407. <https://doi.org/10.1080/01431160903349057>.

Burud I, Lange G, Lillemo M, Bleken E, Grimstad L, From PJ (2017) Exploring robots and UAVs as phenotyping tools in plant breeding. **IFAC PapersOnLine**, 50, 11479–11484. <https://doi.org/10.1016/j.ifacol.2017.08.1591>.

Canata TF, Wei MCF, Maldaner LF, Molin JP (2021) Sugarcane Yield Mapping Using High-Resolution Imagery Data and Machine Learning Technique. **Remote Sensing**, 13, 232. <https://doi.org/10.3390/rs13020232>.

Cevallos JC, Villagomez JA, Andryshchenko I (2019) Convolutional neural network in the recognition of spatial images of sugarcane crops in the troncal region of the coast of Ecuador. **Procedia Computer Science**, 150, 757–763. <https://doi.org/10.1016/j.procs.2019.02.001>

Champ J, Mora-Fallas A, Goëau H, Mata-Montero E, Bonnet P, Joly A (2020) Instance segmentation for the fine detection of crop and weed plants by precision agricultural

robots. **Applications in Plant Sciences**, 8, 11373. <https://doi.org/10.1002/aps3.11373>.

CONAB (Companhia Nacional de Abastecimento). (2024a) **Acompanhamento da safra brasileira grãos safra 2023/24**. Disponível online: <https://www.conab.gov.br/info-agro/safras>. Acessado em 24 de janeiro de 2024.

CONAB (Companhia Nacional de Abastecimento). (2024b) **Acompanhamento da safra brasileira Cana-de-açúcar 3º levantamento da safra 2023/24**. Disponível online: <https://www.conab.gov.br/info-agro/safras>. Acessado em 24 de janeiro de 2024.

Darwish A, Ezzat D, Hassanien AE (2020) An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. **Swarm and Evolutionary Computation**. 52, 100616. <https://doi.org/10.1016/j.swevo.2019.100616>

Dimov D, Uhl JH, Low F, Seboka GN, (2022) Sugarcane yield estimation through remote sensing time series and phenology metrics. **Smart Agricultural Technology**, 2, 100046. <https://doi.org/10.1016/j.atech.2022.100046>

Embrapa (Empresa Brasileira de Pesquisa Agropecuária). (2021) **Características da planta e Importância Socioeconômica do milho**. Disponível online: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/milho>. Acessado em 24 de janeiro de 2024.

Everingham YL, Lowe KH, Donald DA, Coomansa DH, Markley J, J (2007) Advanced satellite imagery to classify sugarcane crop characteristics. *Agronomy for Sustainable Development*, 27, 111–117 (2007). <https://doi.org/10.1051/agro:2006034>

FANCELLI, AL (2015) Manejo baseado na fenologia aumenta a eficiência de insumos e produtividade. **Visão Agrícola**, 13.

Gao J, Nuytens D, Lootens P, He Y, Pieters JG (2018) Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. **Biosystems Engineering**, 170, 39–50. <https://doi.org/10.1016/j.biosystemseng.2018.03.006>.

Hamzeh S, Naseri AA, AlaviPanah SK, Bartholomeus H, Herold M (2016) Assessing the accuracy of hyperspectral and multispectral satellite imagery for categorical and quantitative mapping of salinity stress in sugarcane fields. **International Journal of Applied Earth Observation and Geoinformation**, 52, 412–421. <https://doi.org/10.1016/j.jag.2016.06.024>

Hamzeh S, Naseri AA, Alavipanah SK, Mojaradi B, Bartholomeus HM, Clevers JG, Behzad M (2013) Estimating salinity stress in sugarcane fields with spaceborne hyperspectral vegetation indices. **International Journal of Applied Earth**

Observation and Geoinformation, 21, 282–290. <https://doi.org/10.1016/j.jag.2012.07.002>

Jiang H, Li D, Jing W, Xu J, Huang J, Yang J, Chen S (2019) Early Season Mapping of Sugarcane by Applying Machine Learning Algorithms to Sentinel-1A/2 Time Series Data: A Case Study in Zhanjiang City, China. **Remote Sensing**, 11, 861. <https://doi.org/10.3390/rs11070861>

Ji S, Zhang C, Xu A, Shi Y, Duan Y. (2018) 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. **Remote Sensing**.10, 75. <https://doi.org/10.3390/rs10010075>

Kayad A, Sozzi M, Gatto S, Marinello F, Pirotti, F (2019). Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. **Remote Sensing** 11, 2873. <https://doi.org/10.3390/rs11232873>.

Lima JJA, Maldaner LF, Molin, JP, (2021) Sensor Fusion with NARX Neural Network to Predict the Mass Flow in a Sugarcane Harvester. **Sensors**, 21, 4530. <https://doi.org/10.3390/s21134530>

Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. (2018) Machine Learning in Agriculture: A Review. **Sensors**. 18, 2674. <https://doi.org/10.3390/s18082674>

Louargant M, Jones G, Faroux R, Paoli JN, Maillot T, Gée C, Villette S, (2018) Unsupervised Classification Algorithm for Early Weed Detection in Row-Crops by Combining Spatial and Spectral Information. **Remote Sensing**, 10, 761. <https://doi.org/10.3390/rs10050761>.

Lofton J, Tubana BS, Kanke Y, Teboh J, Viator H, Dalen M. 2012. "Estimando o potencial de rendimento da cana-de-açúcar usando uma determinação na estação do índice vegetativo de diferença normalizada" **Sensores** 12, 6, 7529-7547. <https://doi.org/10.3390/s120607529>.

Luna I, Lobo A (2016) Mapping Crop Planting Quality in Sugarcane from UAV Imagery: A Pilot Study in Nicaragua. **Remote Sensing**, 8, 500. <https://doi.org/10.3390/rs8060500>.

Maldaner LF, Corrêdo LP, Canata TF, Molin JP (2021) Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. **Computers and Electronics in Agriculture**, 181,105945. <https://doi.org/10.1016/j.compag.2020.105945>

Miphokasap P, Wannasiri W (2018) Estimations of nitrogen concentration in sugarcane using hyperspectral imagery. **Sustainability**, 10, 1266. <https://doi.org/10.3390/su10041266>

Molijn RA, Iannini L, Rocha JV, Hanssen R.F (2019) Sugarcane productivity mapping through C-band and L-band SAR and optical satellite imagery. **Remote Sensing**, 11: 1109. <https://doi.org/10.3390/rs11091109>

Molin Jp, Amaral Lr, Colaço A (2015) **Agricultura de precisão**. Oficina de textos. 232 p.

Narmilan A, Gonzalez F, Salgadoe ASA, Kumarasiri UWLM, Weerasinghe HAS, Kulasekara BR, (2022) Predicting Canopy Chlorophyll Content in Sugarcane Crops Using Machine Learning Algorithms and Spectral Vegetation Indices Derived from UAV Multispectral Imagery. **Remote Sensing**, (2022), 14, 1140. <https://doi.org/10.3390/rs14051140>.

Nihar A, Patel NR, Danodia A (2022) Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, **India Journal of the Indian Society of Remote Sensing**, 50, 1519–1530. <https://doi.org/10.1007/s12524-022-01549-0>

Oliveira RP, Barbosa Júnior MR, Pinto AA, Oliveira JLP, Zerbato C, Furlani CEA. (2022) Predicting Sugarcane Biometric Parameters by UAV Multispectral Images and Machine Learning. **Agronomy** 12, 1992. <https://doi.org/10.3390/agronomy12091992>.

Oliveira MFD, Ortiz BV, Morata GT, Jiménez AF, Rolim GS, Silva RP (2022) Training Machine Learning Algorithms Using Remote Sensing and Topographic Indices for Corn Yield Prediction. **Remote Sensing**. 14, 6171. <https://doi.org/10.3390/rs14236171>.

Oliveira, MF (2021) **Previsão e estimativa de variáveis de cultivo utilizando níveis e formas de sensoriamento remoto e técnicas de aprendizado de máquina**. 123 f. Tese (Doutorado em Agronomia) – Unesp, Jaboticabal.

Prasara AJ, Gheewala SH, (2016) Sustainability of sugarcane cultivation: Case study of selected sites in north-eastern Thailand. **Journal of Cleaner Production**, 134, 613–622. <https://doi.org/10.1016/j.jclepro.2015.09.029>.

Pinto AA, Zerbato C, Rolim GS, Barbosa Júnior MR, Silva, LFV, Oliveira RP (2022). Corn grain yield forecasting by satellite remote sensing and machine learning models. **Agronomy journal**, 1-20. <https://doi.org/10.1002/agj2.21141>.

Pinheiro IL, Damian JM, Cherubin MR, Barros PPS, Fiorio PR, Cerri CC, Cerri CEP (2018) Prediction of Sugarcane Yield Based on NDVI and Concentration of Leaf-Tissue Nutrients in Fields Managed with Straw Removal, **Agronomy** 8, 9, 196. <https://doi.org/10.3390/agronomy8090196>.

Romero M, Luo Y, Su B (2018) Fuentes, S. Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management. **Computers Electronics in Agriculture**, 147, 109–117. <https://doi.org/10.1016/j.compag.2018.02.013>.

Som-ard J, Atzberger C, Izquierdo-Verdiguier E, Vuolo F, Immitzer M (2021) Remote Sensing Applications in Sugarcane Cultivation: A Review. **Remote Sensing**, 13, 4040. <https://doi.org/10.3390/rs13204040>

Sukyai P, Yingkamhaeng N, Lam NT, Tangsatianpan V, Watcharinrat C, Vanitjinda G, Vanichsiratana W, Sriroth K (2016) Research and Development Prospects for Sugarcane and Sugar Industry in Thailand. **Sugar Tech** 18, 583–587. <https://doi.org/10.1007/s12355-016-0495-1>.

Ramos APM, Osco LP, Furuya DEG, Gonçalves WN, Santana DC, Teodoro LPR Silva Junior CA, Capristo-Silva GF, Li J, Baio FHR, Marcato Junior J, Teodoro PE, Pistori H (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. **Computers and Electronics in Agriculture** 178: 105791. DOI: <https://doi.org/10.1016/j.compag.2020.105791>.

Vargas LAO, Mendoza GG, Gómez RA, Rivero NA (2016) Espinosa, L.Y. Characterization of *Diatraea saccharalis* in Sugarcane (*Saccharum officinarum*) with Field Spectroradiometry. **International Journal of Environmental & Agriculture Research**, 2, 1–6.

Wang J, Xiao X, Liu L, Wu X, Qin Y, Steiner JL, Dong J (2020) Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. **Remote Sensing of Environment**, 247, 111951. <https://doi.org/10.1016/j.rse.2020.111951>.

Wang M, Liu Z, Baig MHA, Wang Y, Li Y, Chen Y (2019) Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. **Land Use Policy**, 88, 104190. <https://doi.org/10.1016/j.landusepol.2019.104190>.

Wang Z, Lu Y, Zhao G, Sol C, Zhang F, He S. (2022) Sugarcane Biomass Prediction with Multi-Mode Remote Sensing Data Using Deep Archetypal Analysis and Integrated Learning. **Remote Sensing**, 14, 4944. <https://doi.org/10.3390/rs14194944>

WEISS M, JACOB F, DUVEILLER G (2020) Remote sensing for agricultural applications: A meta-review. **Remote Sensing of Environment**, 236, 111402. <https://doi.org/10.1016/j.rse.2019.111402>.

Wu H, Wiesner-Hanks T, Stewart EL, DeChant C, Kaczmar N, Gore MA, Nelson, RJ, Lipson H (2019) Autonomous Detection of Plant Disease Symptoms Directly from Aerial Imagery. **Plant Phenome Journal**, 2, 1–9. <https://doi.org/10.2135/tppj2019.03.0006>.

Xu JX, Ma J, Tang YN, Wu WX, Shao JH, Wu WB, Wei SY, Liu YF, Wang YC, Guo HQ. (2020) Estimation of Sugarcane Yield Using a Machine Learning Approach Based on UAV-LiDAR Data. **Remote Sensing**, 12, 2823. <https://doi.org/10.3390/rs12172823>

Zhang X, Qiao Y, Meng F, Fan C, Zhang M (2018) Identification of maize leaf diseases using improved deep convolutional neural networks. **IEEE Access**, 6, 30370–30377. <https://doi.org/10.1109/ACCESS.2018.2844405>.

Zhong L, Hu L, Zhou H (2019) Deep learning based multi-temporal crop classification. **Remote Sensing Environment**, 221, 430–443. <https://doi.org/10.1016/j.rse.2018.11.032>.

Zhu L, Liu X, Wang Z, Tian L (2023) High-precision sugarcane yield prediction by integrating 10-m Sentinel-1 VOD and Sentinel-2 GRVI indexes. **European Journal of Agronomy**, 149, 126889. <https://doi.org/10.1016/j.eja.2023.126889>

CAPÍTULO 2: PREVISÃO DE PRODUTIVIDADE DO MILHO UTILIZANDO SENSORIAMENTO REMOTO, SOMA DE GRAUS-DIA E MODELOS DE APRENDIZADO DE MÁQUINA²

² Artigo formatado para submissão na revista Theoretical and applied climatology

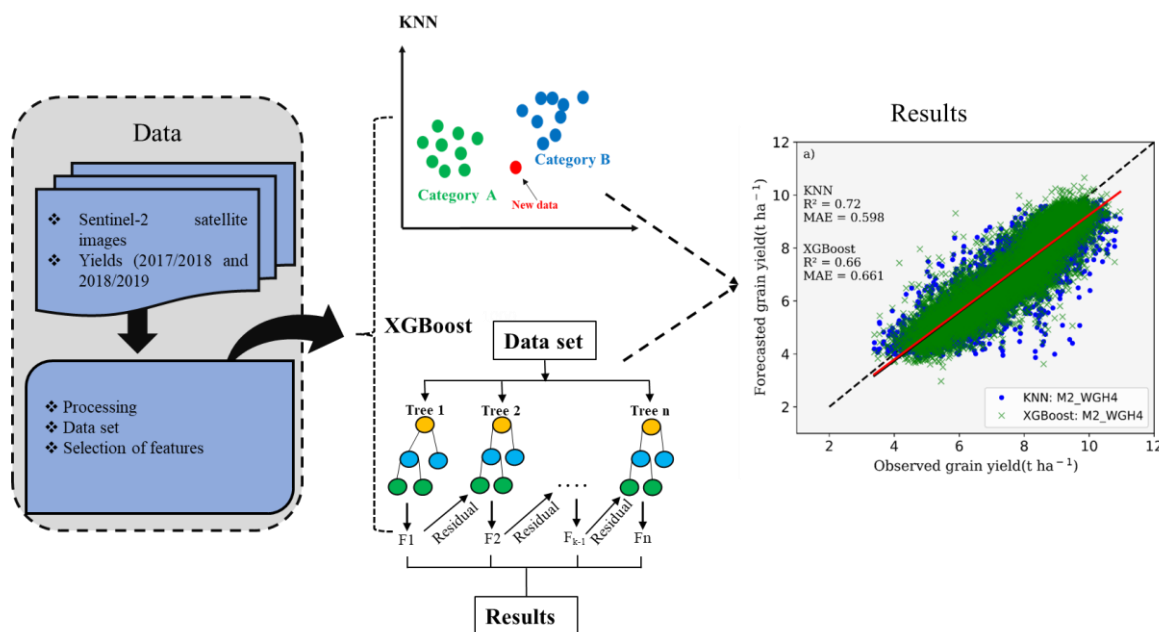


Figura 1. Resumo gráfico exemplificando as etapas de obtenção de dados, geração dos modelos e resultados.

RESUMO

Com o custo de produção cada vez mais elevado, a previsão de produtividade antes da colheita é um importante aliado nas tomadas de decisões buscando ganhos econômicos e sustentáveis. No presente estudo propomos uma abordagem baseada na soma de graus-dia da cultura do milho relacionada com as datas das imagens de satélite para analisar os dados de duas safras gerando modelos de previsão com os algoritmos *k*-nearest neighbors (KNN) e extreme gradient boosting (XGBoost). O estudo foi desenvolvido em área comercial nas safras 2017/18 e 2018/19. Os dados espectrais foram obtidos de imagens do satélite Sentinel-2, e posteriormente as imagens foram processadas e obtidos os valores das bandas espectrais e índices de

vegetação. Para o desenvolvimento dos modelos, as imagens foram divididas em três classes de semanas antes da colheita, com base na soma de graus-dia (WBH10: 0 – 800; WBH4: 801 – 1200 e WBH1 > 1200 graus-dias). Adotamos 12 combinações de entradas de dados e conseguimos prever a produtividade com aproximadamente de 1050, 760 e 140 graus-dia antes da colheita, para as WBH10, WBH4 e WBH1, respectivamente. Os modelos mais acurados na previsão da produtividade foram com os dados das bandas espectrais do vermelho, azul, verde e infravermelho próximo coletadas de 801 a 1200 graus-dia (WBH4). Para os modelos desenvolvidos combinando as WBHs para a previsão de produtividade com erro médio de 0,503 t ha⁻¹, sendo a maior acurácia com o uso de todas as variáveis e a combinação das três classes de graus-dia.

Palavras-chave: Agricultura digital, modelagem agrícola, inteligência artificial, python.

Corn yield forecasting using remote sensing, degree-days, and machine learning models.

ABSTRACT

With production costs increasingly high, forecast yield before harvesting is an important ally in decision-making seeking economic and sustainable gains. In the present study, we propose an approach based on the sum of degree days of the corn crop related to the dates of satellite images to analyze data from two harvests, generating forecast models with the k-nearest neighbors (KNN) and extreme gradient boosting algorithms. (XGBoost). The study was carried out in a commercial area in the 2017/18 and 2018/19 harvests. The spectral data were obtained from images from the Sentinel-2 satellite, and the images were subsequently processed, and the values of the spectral bands and vegetation indices were obtained. To develop the models, the images were divided into three classes of weeks before harvest, based on the sum of degree days (WBH10: 0 – 800; WBH4: 801 – 1200 and WBH1 > 1200-degree days). We adopted 12 combinations of data inputs and were able to forecast yield at approximately 1050, 760, and 140 degree-days before harvest, for WBH10, WBH4, and WBH1, respectively. The most accurate models in predicting productivity were with

data from the red, blue, green, and nir spectral bands collected from 801 to 1200-degree days (WBH4). For the models developed combining WBHs to forecast yield with an average error of 0.503 t ha⁻¹, the highest accuracy being with the use of all variables and the combination of the three classes of degree days.

Keywords: Digital agriculture, agricultural modeling, artificial intelligence, python.

1. INTRODUÇÃO

Acontecimentos como a crise do COVID-19, aliado aos problemas geopolíticos entre países dos continentes europeu e asiático, contribuem para a elevação dos custos de produção agrícola em todo o mundo. No Brasil a situação também não é diferente, pois não somos auto-suficientes em todos os insumos agrícolas que são essenciais para obtenção de altas produtividades. Nesse contexto, a cooperação entre as esferas da pesquisa, extensão e produtores rurais torna-se essencial nas tomadas de decisões buscando ganhos econômicos e sustentáveis (Tedesco et al., 2021b).

O Brasil é destaque mundial na produção de grãos, sendo o terceiro maior produtor de milho (*Zea mays* L.). No país o milho é o segundo grão mais cultivado, com aproximadamente 22,2 milhões de hectares na safra 2022/23, produzindo cerca de 131,9 milhões de toneladas, conforme estimado pela Companhia Nacional de Abastecimento (CONAB, 2024). Considerando a importância do milho no agronegócio brasileiro, acompanhar o desenvolvimento e a produtividade desse cultivo auxiliam os agentes da cadeia produtiva no aumento da produtividade sem necessariamente precisarem aumentar a área cultivada.

Na tentativa de acompanhar o desenvolvimento e a produtividade de diversos cultivos agrícolas como milho, trigo e cana-de-açúcar, pesquisadores têm buscado

cada vez mais o uso de tecnologias que possibilitam auxiliá-los nessa tarefa, considerando-se, que as áreas cultivadas com essas culturas são extensas para fazer somente monitoramento no nível de campo. Neste sentido, as imagens de satélite vêm sendo uma alternativa viável utilizada no monitoramento das culturas, tornando-se mais fundamentais nas tomadas de decisões em campo, bem como, na análise de retornos econômicos (Venancio et al., 2019; Barzin et al., 2020; Zhou et al., 2021).

Embora as imagens de satélite possam fornecer suporte no monitoramento das culturas, elas por si só, não são o suficiente para previsão ou estimativa da produtividade, sendo assim, é necessário estabelecer relações entre as informações de reflectância obtidas nas imagens e a produtividade observada em campo. Nesse sentido, a modelagem agrícola tem sido fundamental, permitindo estabelecimento dessas relações em diversas culturas como algodão (Haghverdi et al., 2018), trigo (Hunt et al., 2019 e Zhou et al., 2021), cana-de-açúcar (Canata et al., 2021 e Maldaner et al., 2021b) e milho (Ramos et al., 2020).

Na modelagem agrícola as técnicas de aprendizado de máquina (ML) têm ganhado destaque devido a capacidade com que esses modelos resolvem problemas complexos (Kayad et al., 2019). Nos modelos de ML o *k*-Nearest Neighbors (KNN), é um dos modelos de ML mais simples, depois da regressão linear, esse modelo é utilizado em problemas de classificação e regressão. São considerados os valores dos vizinhos na definição dos pesos, sendo que, pontos mais próximos contribuem mais com os valores previstos (Abbas et al., 2020).

Outro modelo de ML é o extreme gradient boosting (XGBoost) de Chen & Guestrin (2016), sendo considerado um modelo relativamente novo em comparação com os demais modelos, porém já foi relatado que ele apresenta alta capacidade de

resolver problemas complexos na área agrícola (Folberth et al., 2019; Shanhosseini et al., 2020). O XGBoost é um modelo baseado em árvores, em que, cada nova árvore aprende com os erros das árvores anteriores (Folberth et al., 2019). Na modelagem de produtividade o XGBoost foi utilizado em alguns trabalhos de estimativa de produtividade, como por exemplo Folberth et al., (2019) e Shanhosseini et al., (2020), que estimaram a produtividade do milho utilizando dados de clima e solo e Kamir et al., (2020) ao estimar a produtividade do trigo utilizando dados climáticos e séries temporais de satélite.

Embora o KNN e o XGBoost já tenham sido utilizados na modelagem da produtividade agrícola, até o presente momento pouco são os estudos referentes à previsão da produtividade, sendo observado que tais modelos foram mais utilizados em estudos de estimativas de produtividade, reduzindo a possibilidade de tomada de decisão em campo, corroborando com o relatado por Benos et al., (2021). Estimação é a atividade de estimar algo naquele momento, enquanto previsão é uma estimativa para o futuro.

O fato de a literatura apresentar mais trabalhos com estimativa da produtividade do milho está diretamente relacionada as variáveis que são utilizadas para gerar os modelos, na maioria das vezes essas variáveis não possibilitam previsão, até mesmo quando são utilizadas imagens de satélites são encontrados problemas sobre como organizar as informações de mais de uma safra, tendo em vista que existe discrepâncias entre as imagens coletadas entre uma safra e outra devido a resolução temporal dos satélites, bem como, a ocorrência de nuvens que impossibilita o uso dessas imagens (Wang et al., 2019).

Nesse trabalho o objetivo foi avaliar uma abordagem baseada em classes de soma de graus dias da cultura do milho para organizar os dados das imagens de satélite em duas safras, utilizando posteriormente esses dados como variáveis de entradas em modelos de machine learning (KNN e XGBoost), para prever a produtividade de milho antes da colheita.

2. MATERIAL E MÉTODOS

2.1. Local de estudo

O estudo foi desenvolvido em duas áreas comerciais de 50,7 e 25,7 hectares localizadas no município de Primavera do Leste, Mato Grosso, durante as safras 2017/18 e 2018/19 (Figura 2a).

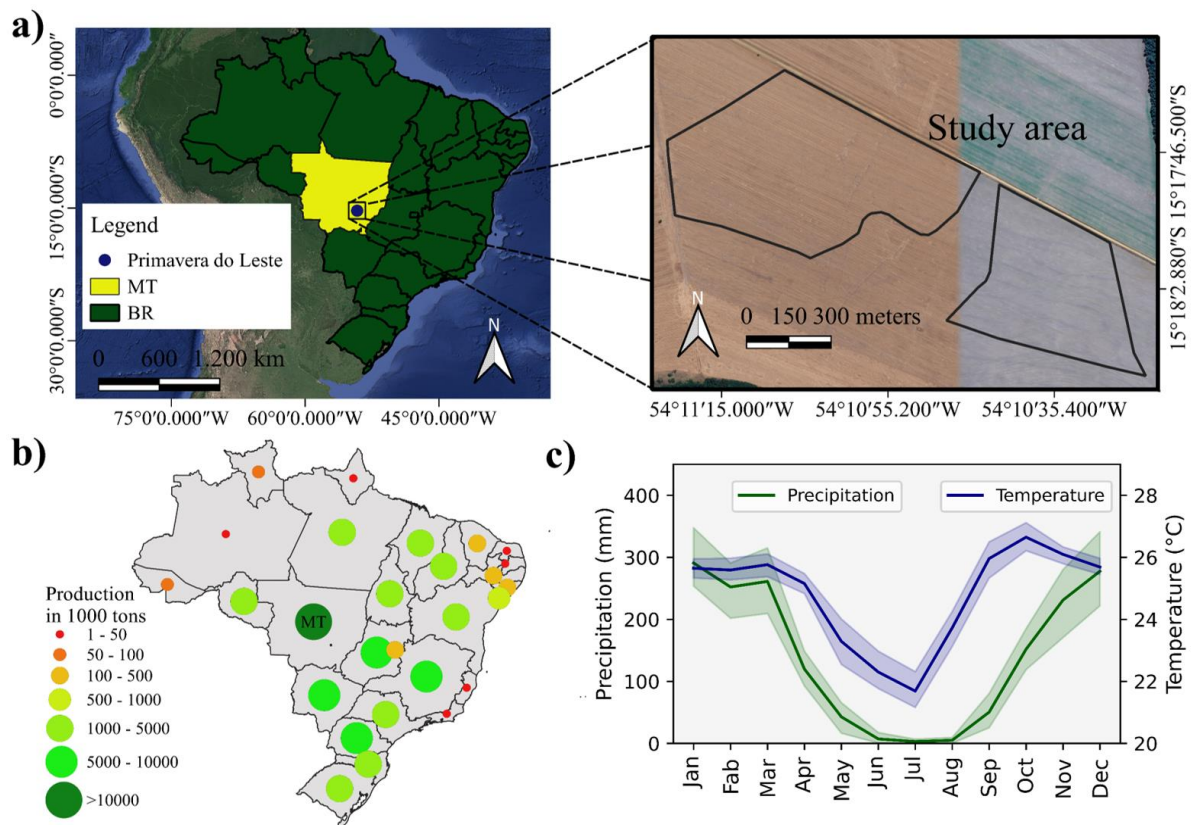


Figura 2. a – Localização da área de estudo no Brasil; b – Produção estimada de milho no Brasil na safra 2022/2023. Adaptado da CONAB 2024; c – Precipitação e temperatura mensais de 2000 a 2021. Adaptado do INMET 2023.

O município de Primavera do Leste está localizado em uma das mais importantes áreas de produção de milho no Brasil (Figura 2b). O clima na região é tropical quente e úmido, que inclui seca no inverno e chuvas no verão de outubro a abril, com baixa variabilidade climática de acordo com os dados de precipitação e temperatura (Figura 2c). A precipitação média anual da série histórica variou de aproximadamente 1140 mm a 2060 mm, e a temperatura média anual da série histórica para a região varia entre 24,1 °C e 25,5 °C.

Em ambas as safras foi utilizado o híbrido DKB 390, esse tem tolerância a temperaturas altas e veranico com ampla adaptação a diferentes solos, manejo e época de plantio e é ideal para cultivos de safrinha. A safra 2017/18 foi semeada dia 28 de fevereiro de 2018 e colhida no dia 20 de julho de 2018 totalizando 1720 graus-dias após a semeadura. Já a safra 2018/19 foi semeada no dia 08 de fevereiro de 2019 e colhida no dia 23 de junho de 2019 totalizando 1829 graus-dias após a semeadura.

2.1. Obtenção e processamento de imagens de satélite

No estudo foram utilizadas imagens do satélite Sentinel-2, durante a obtenção das imagens identificamos que as imagens correspondentes ao início do desenvolvimento da cultura para ambos as safras apresentaram problema com cobertura de nuvens (Wang et al., 2019). Assim, nove imagens do Sentinel-2 com baixa cobertura de nuvens (<1%) foram obtidas na safra 2017/18 e cinco na safra 2018/19. Todas as imagens foram obtidas do United States Geological Survey (USGS) por meio da plataforma Earth Explorer (earthexplorer.usgs.gov), posteriormente as imagens foram corrigidas utilizando o plugin de código aberto Semi-Automatic

Classification Plugin, versão 7.8.1 (Congedo, 2020) no Software Quantum GIS (QGIS), versão 3.16. Essa correção consiste em converter os números digitais em dados de reflectância, com base na metodologia de correção atmosférica Dark Object Subtraction (DOS1). Após a correção, um buffer de cinco metros foi criado no entorno da área de interesse e em seguida as imagens foram recortadas para a extração dessa área.

Logo após a correção atmosférica das imagens calculamos a soma de graus-dia (SDD) para todas as imagens das duas safras a partir de 28 de fevereiro de 2018 e a partir de 8 de fevereiro de 2019. Para determinar a SDD, utilizamos dados de uma estação meteorológica automática do Instituto nacional de meteorologia (portal.inmet.gov.br) que coleta dados a cada hora (ID: A923), localizada aproximadamente a 30 km da região dos campos de produção. O cálculo foi realizado de acordo com a data de aquisição de cada imagem de satélite, seguindo o método de Arnold (1959):

$$SDD = \sum \left(\frac{T_{max} + T_{min}}{2} \right) - T_{base} \quad (1)$$

Em que, SDD é a soma de graus-dia, T_{max} é a temperatura máxima do ar (°C), T_{min} é a temperatura mínima do ar (°C), e T_{base} é a temperatura base mais baixa para o desenvolvimento da cultura (10 °C) seguindo Jiang et al., (2018).

Considerando que as imagens correspondentes ao início do desenvolvimento da cultura em ambos as safras apresentaram problema com cobertura de nuvens, optamos por dividir essas imagens em três classes de soma de graus-dia, de 0 – 800 SDD corresponde a pelo menos dez semanas antes da colheita (WBH10); 801 – 1200 SDD quatro semanas antes da colheita (WBH4) e > 1200 SDD uma semana antes da colheita (WBH1).

A partir dessas classes (WBH10, WGH4 e WBH1), obtivemos a média aritmética dos dados espectrais de cada classe (Figura 3a e Figura 3b).

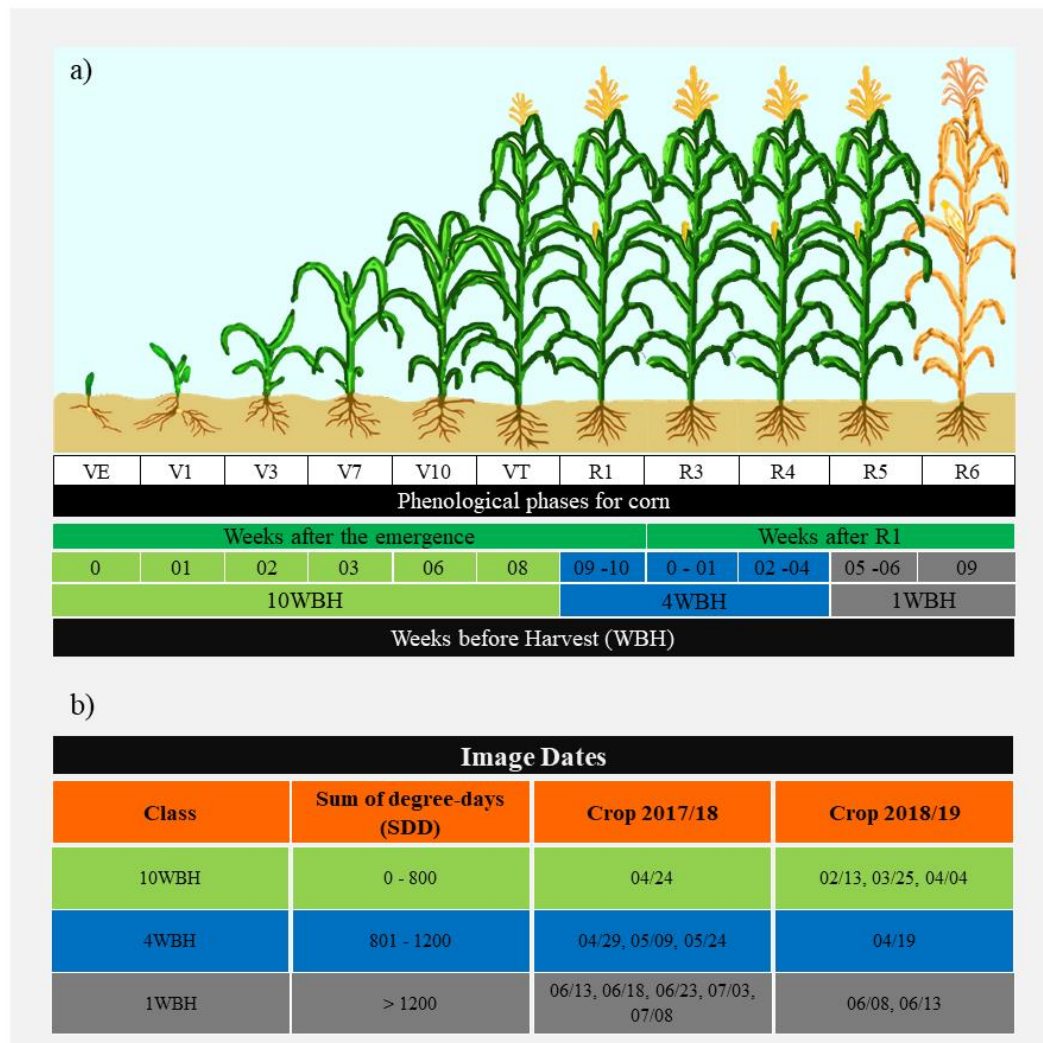


Figura 3 a - Divisão da soma graus-dia em função da fisiologia do milho, adaptado de Fancelli (1986), b - Datas das imagens em cada WBH. Fase vegetativa (VE: Emergência, V3: Três folhas totalmente emergidas, V7: Sete folhas totalmente emergidas, V10: Dez folhas totalmente emergidas, VT: Pendão), Fase reprodutiva (R1: Embonecamento, R3: Grão leitoso, R4 Grão pastoso, R5: Formação do dente, R6: Maturação fisiológica).

Na safra 2017/18 na primeira classe (WBH10), obtivemos apenas uma imagem durante o período de 0 a 800 SDD, então adotamos os valores das bandas espectrais

desta imagem como sendo os valores da classe WBH10 para cada banda espectral nesta safra, já na safra 2018/19 obtivemos três imagens no período de 0 a 800 SDD, então obtivemos os valores médios das três imagens para cada banda espectral, adotando esses valores como sendo os valores da classe WBH10 na safra 2018/19 (Figura 3a e Figura 3b). Mais detalhes sobre a soma dos graus-dia de cada imagem podem ser encontrados no Apêndice I.

2.2. Variáveis de entradas dos modelos

No estudo foram consideradas quatro bandas espectrais do Sentinel-2 para compor parte das variáveis de entrada dos modelos, a banda do vermelho, verde, azul e infravermelho próximo (Tabela 1).

Tabela 1. Especificações das bandas espectrais do Sentinel-2 e Índices de vegetação (VIs)

Bandas	Comprimento de onda central (nm)	Resolução		
		Espacial (m)	Temporal (Dias)	Radiométrica (Bits)
Blue	490	10	5	12
Green	560			
Red	665			
Near-infrared	842			
Índices de vegetação	Equação		Referência	
GNDVI	$\left(\frac{p^{NIR} - p^{GREEN}}{p^{NIR} + p^{GREEN}}\right)$		Gitelson et al., (1996)	
NDVI	$\left(\frac{p^{NIR} - p^{RED}}{p^{NIR} + p^{RED}}\right)$		Rouse et al., (1974)	
SAVI	$\left(\frac{p^{NIR} - p^{RED}}{p^{NIR} + p^{RED} + 0.5}\right) \cdot (1.5)$		Huete (1988)	

Além das bandas espectrais, foram obtidos das imagens três índices de vegetação, Green Normalized Difference Vegetation Index (GNDVI); Normalized Difference Vegetation Index (NDVI); Soil-Adjusted Vegetation Index (SAVI), tais

índices são citados na literatura no monitoramento da produtividade de grãos de diversas culturas (Barzin et al., 2020; Ramos et al., 2020; Zhou et al., 2021).

Nesse estudo, adotamos 12 combinações de entradas de dados (Tabela 2), para desenvolvimento dos modelos de previsão da produtividade, utilizando as bandas do RGB (vermelho, verde e azul), NIR (infravermelho próximo) e os índices de vegetação (IVs) como variáveis para essas combinações, denominadas de M1 a M12, em que, M1 corresponde a combinação RGB para cada uma das classes de WBH, ou seja, os dados RGB1 correspondem aos dados das bandas espectrais RGB somente da classe WBH10 (0-800 SDD), enquanto, RGB2 correspondem aos dados das bandas espectrais RGB somente da classe WBH4 (801-1200 SDD).

Tabela 2. Combinações de entradas de dados para os modelos.

MODELO	WBH10	WBH4	WBH1
M1	RGB1	RGB2	RGB3
M2	RGB1+NIR1	RGB2 + NIR2	RGB3 + NIR3
M3	RED1+NIR1	RED2 + NIR2	RED3 + NIR3
M4	IVs1	IVs2	IVs3
M5	RGB1+ IVs1	RGB2+ IVs2	RGB3 + IVs3
M6	RGB1 + NIR1+ IVs1	RGB2 + NIR2 + IVs2	RGB3 + NIR3+ IVs3
MODELO	WBH10+4	WBH4+1	WBH10+4+1
M7	RGB1+RGB2	RGB2+RGB3	RGB1+RGB2+RGB3
M8	RGB1+NIR1+RGB2+NIR2	RGB2+NIR2+RGB3+NIR3	RGB1+NIR1+RGB2+NIR2+RGB3 + NIR3
M9	RED1+NIR1+RED2+NIR2	RED2+NIR2+RED3+NIR3	RED1+NIR1+RED2+NIR2+RED3 + NIR3
M10	IVs1+IVs2	IVs2+IVs3	IVs1+IVs2+IVs3
M11	RGB1+IVs1+RGB2+IVs2	RGB2+ IVs2+RGB3 + IVs3	RGB1+ IVs1+RGB2+ IVs2+RGB3+IVs3
M12	RGB1+NIR1+IVs1+RGB2 +NIR2+IVs2	RGB2+NIR2+IVs2+RGB3+ NIR3+IVs3	RGB1+NIR1+IVs1+RGB2 +NIR2+IVs2+RGB3+NIR3+IVs3

Analizamos também a combinação das variáveis de uma classe WBH com as demais classes, por exemplo, a combinação das variáveis de entrada RGB1 da classe WBH10 com as variáveis de entrada do RGB2 da classe WBH4 formando o M7 e assim sucessivamente, sendo denominadas de WBH10+4 (WBH10 + WBH4), WBH4+1 (WBH4 + WBH1) e WBH10+4+1 (WBH10 + WBH4 + WBH1), conforme Tabela 2.

2.3. Aquisição dos dados de produtividade

Os dados de produtividade do milho nas duas safras foram obtidos por sensores embarcados na colhedora, esses sensores coletam aproximadamente 600 pontos de georreferenciados por hectare, medindo o fluxo de grãos (massa ou volume) para converter em produtividade.

Antes do processo de colheita os sensores foram calibrados a fim de minimizar os erros de medição em campo. Após a colheita, os dados foram processados removendo os outliers identificados no processamento de dados conforme metodologia descrita por Maldaner et al., (2021a) que leva em consideração a variabilidade local e global no processamento dos dados.

Após a filtragem dos dados, foram gerados os mapas de produtividade das duas safras utilizando o interpolador IDW (Inverse Distance Weighted), o tamanho do pixel foi de 10 m equivalente ao tamanho do pixel das imagens do Sentinel-2, permitindo que tais dados fossem utilizados para o treinamento e teste dos modelos de predição. O software utilizado na interpolação e geração dos mapas foi Quantum GIS, versão 3.16.

2.4. Modelos de aprendizado de máquina

O modelo KNN leva em consideração os vizinhos mais próximos para definição dos pesos, assim, os pontos mais próximos contribuem mais para a média dos valores previstos (Abbas et al., 2020). No KNN o k é um parâmetro ajustado com papel importante na obtenção de previsão precisa, no estudo utilizou k igual a 15.

O modelo extreme gradient boosting (XGBoost) é um algoritmo de aprendizagem de conjunto baseado em árvores construindo o modelo sequencialmente, em que, cada árvore se ajusta a uma versão modificada do conjunto de dados de treinamento original ou seja, cada nova árvore aprende com os erros das árvores anteriores (Folberth et al., 2019). No XGBoost os parâmetros ajustados foram Learning rate (0.098), número de estimadores (1483 árvores), subamostras (0.213) e o número máximo de nós por árvore (5).

A seleção do melhor ajuste dos parâmetros para cada modelo foi por meio do modelo de busca exaustiva GridSearchCV da biblioteca scikit-learn, esse modelo testa todos as combinações dos parâmetros passados de forma ordenada, buscando o mínimo geral da função da soma de quadrados entre os dados observados e previstos pelos modelos. Os dados foram divididos em treinamento e teste, sendo os dados da primeira safra usado para treinamento e os dados da segunda safra usados para teste.

2.5. Análise de desempenho dos modelos

Foram obtidos os valores do coeficiente de determinação (R^2) e o erro absoluto médio (MAE) como métricas de precisão e acurácia, respectivamente. A precisão está relacionada à repetição em relação aos dados observados e previstos, já a acurácia

está relacionada ao desempenho geral dos modelos com relação aos valores previstos em relação à reta 1:1.

3. RESULTADOS

Os valores mínimo e máximo de produtividade do milho na safra 2017/2018 foi de 2,62 a 9,61 t ha⁻¹, com CV de 19% (tabela 3). Na safra 2018/2019 os valores da produtividade variaram de 3,37 a 11,75 t ha⁻¹, com CV de 15%. O DP observado para os dados de produtividade das safras 2017/2018 e 2018/2019 foram 1,13 e 1,22 t ha⁻¹.

Tabela 3. Estatística descritiva dos dados de produtividade do milho considerando as duas safras.

Safra	Máximo	Mínimo	Média	SD	CV
	t ha ⁻¹				%
2017/2018	9,61	2,62	5,81	1,13	19
2018/2019	11,75	3,37	8,02	1,22	15

SD: Desvio-padrão. CV: coeficiente de variação.

Para cada classe de WBH é apresentado a correlação de Pearson da produtividade de grãos com as quatro bandas espectrais e três índices de vegetação (Figura 4a e Figura 4b). A correlação permite inferir o comportamento de cada WBH na previsão da produtividade. Desse modo, nota-se que para as bandas espectrais as correlações com a produtividade variaram de - 0,39 na banda espectral do blue da classe três (BLUE3) á - 0,86 na banda do verde da classe dois (GREEN2). Já para os IVs variaram de 0,4 (GNDVI3) a 0,88 (GNDVI2).

Ressaltamos que, a correlação de todas a variáveis com a produtividade é maior para os dados da classe WBH4 (801-1200) com -0,86; -0,81; -0,8 e 0,74 para as bandas espectrais do green, red, blue e nir, respectivamente (Fig.4a), enquanto, para os IVs GNDVI, NDVI e SAVI foi de 0,88, 0,84 e 0,83 (Figura 4b), esse resultado

permite identificar uma relação entre os dados espectrais e a produtividade com uma antecipação média de 30 dias antes da colheita.

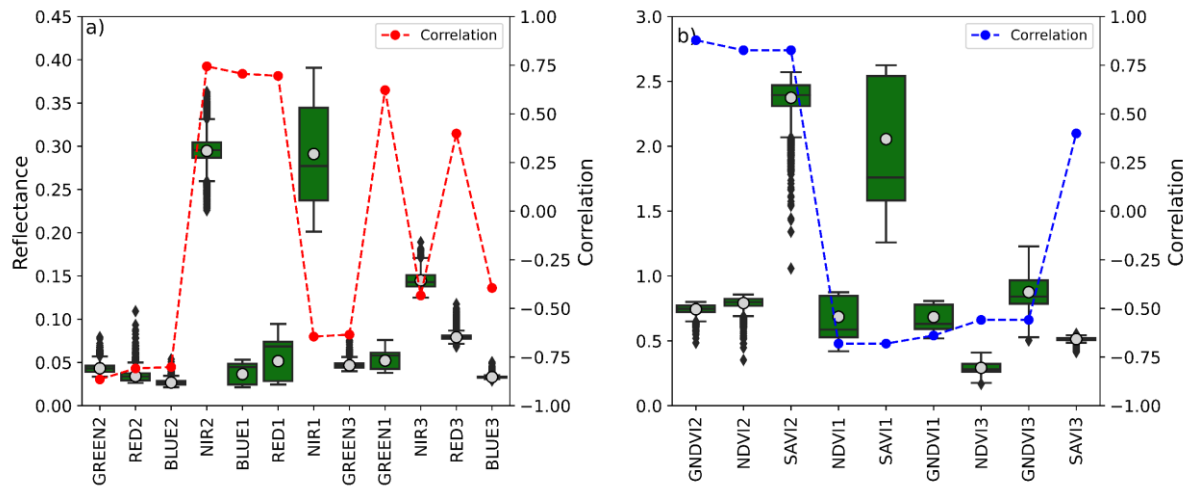


Figura 4 Correlação das bandas espectrais (a) e índices de vegetação (b), com a produtividade. Os símbolos RED, GREEN, BLUE, NIR (Near-infrared), são bandas espectrais e GNDVI, NDVI e SAVI são IVs, bandas espectrais e IVs que terminam em 1(um) correspondem entre 0 e 800 SDD; 2 (dois) entre 801 e 1200 e 3 (três) maiores que 1200 SDD.

A variabilidade da acurácia dos modelos nas três classes de WBH foi similar nos dois algoritmos (Figura 5). Nota-se que o M2 (RGB+NIR) obteve a maior precisão e acurácia nas três classes de WBH para os dois algoritmos, sendo a classe WBH4 superior às demais com R^2 0.72 e MAE 0.598 t ha⁻¹ no KNN e no XGBoost R^2 0.66 e MAE 0.661 t ha⁻¹.

Para ambos os modelos o M3 (RED+NIR) foi o modelo que apresentou menor precisão e acurácia, além da maior variação entre as classes de WBH para essa combinação de variáveis de entrada de dados com R^2 variando de 0,40 a 0,62 e MAE de 0.729 a 0.877 t ha⁻¹ no KNN (Figura 5a e Figura 5b). Enquanto, no XGBoost o R^2 variou de 0.24 a 0.56 e MAE 0.781 a 0.982 t ha⁻¹ (Figura 5c e Figura 5d).

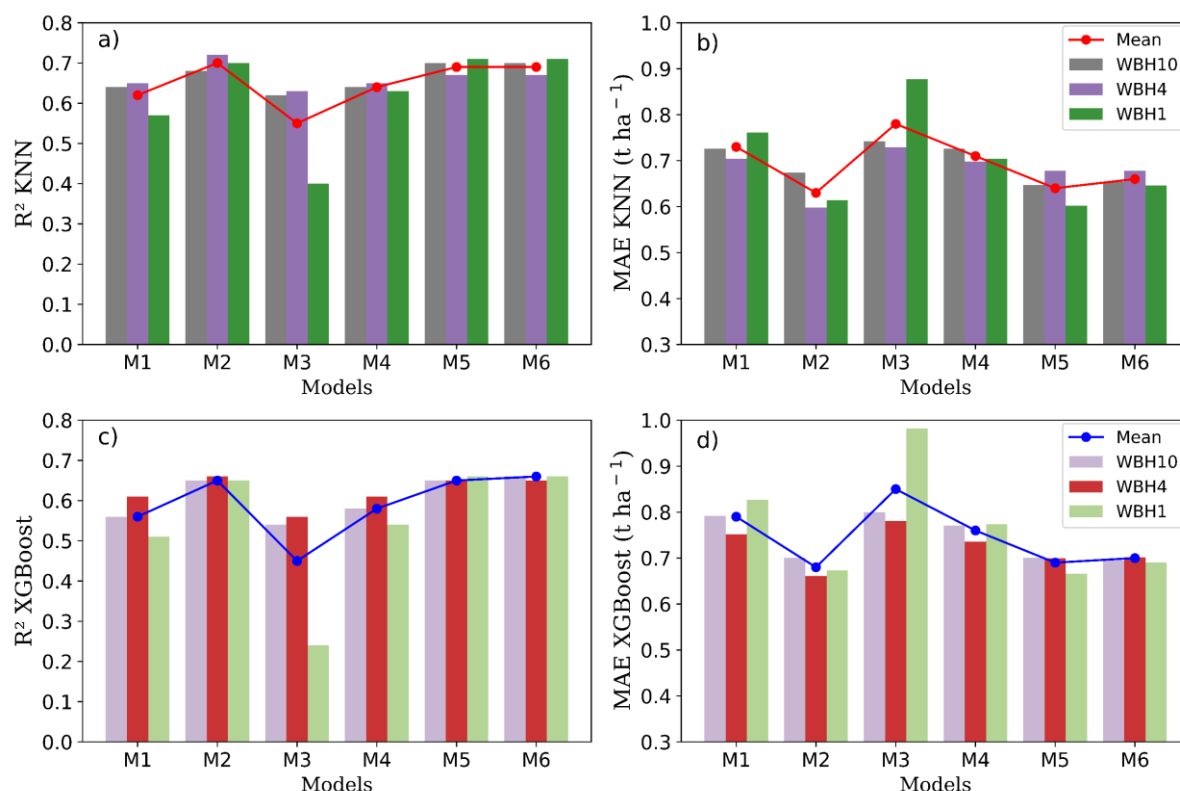


Figura 5 Desempenho de classes WBH individuais para cada modelo, (a) R^2 KNN, (b) MAE KNN, (c) R^2 XGBoost e (d) MAE XGBoost.

Para os modelos desenvolvidos com as combinações das três classes de WBH (WBH10+4, WBH4+1 e WBH10+4+1) a média de precisão e acurácia nos dois modelos foram similares (Figura 6). Para o KNN e XGBoost o M8 obteve a maior precisão e acurácia nas três combinações das classes de WBH com a combinação WBH10+4+1 sendo superior as demais com $R^2=0,81$ e $MAE=0,503 \text{ t ha}^{-1}$ no KNN e $R^2 0,80$ e $MAE 0,525 \text{ t ha}^{-1}$ no XGBoost.

Nota-se ainda que, nos modelos desenvolvidos com as combinações dos WBH (Figura 6), a variável de entrada do modelo tem menor variação no desempenho dos modelos quando comparado os modelos desenvolvidos para cada classe de WBH separadamente (Figura 5), em que a variável de entrada do modelo causa alta variação no desempenho dos modelos.

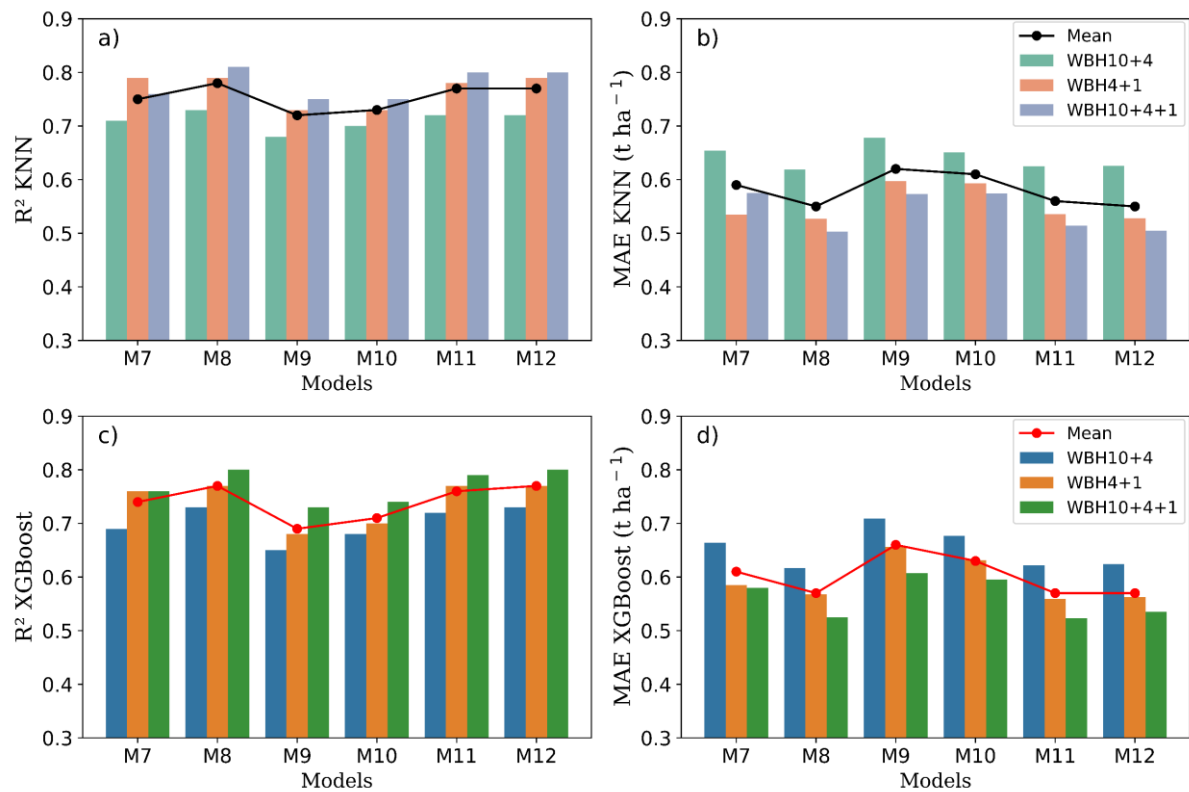


Figura 6. Desempenho das combinações de classes WBH para cada modelo, (a) R^2 KNN, (b) MAE KNN, (c) R^2 XGBoost e (d) MAE XGBoost.

Na dispersão da produtividade observada e prevista para os modelos com maior precisão e acurácia nos dois modelos (Figura 7a), nota-se que, ambos os modelos treinados para os dois algoritmos apresentam tendência de previsão abaixo dos valores observados de produtividade, esse comportamento também foi notado nos modelos mais precisos e acurados dos dois modelos para a produtividade observada e prevista nos modelos combinados das classes de WBH para produtividade acima de 6 t ha^{-1} (Figura 7b).

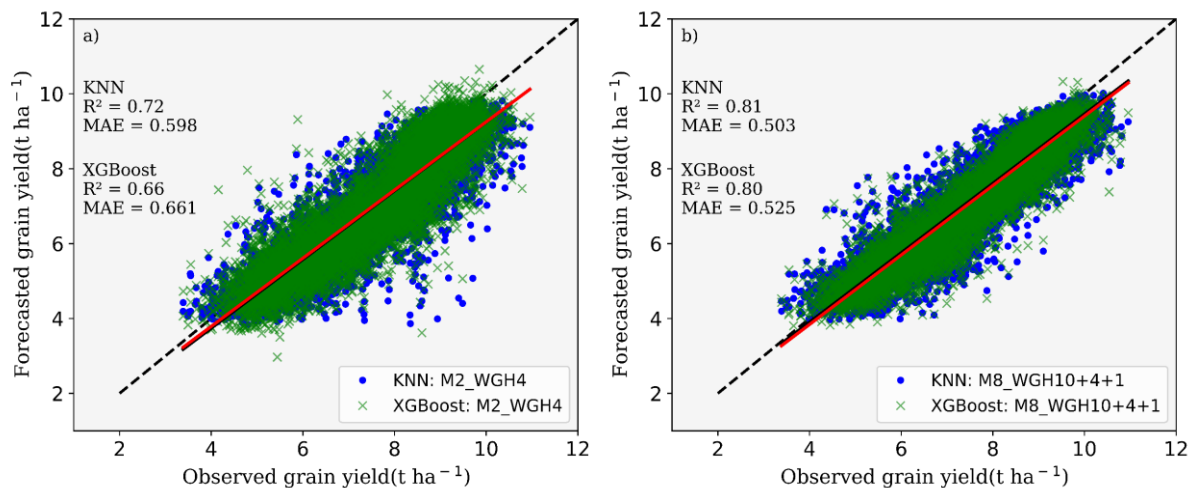


Figura 7. Produtividade observada e prevista dos modelos KNN e XGBoost para os modelos com maior acurácia no WBH individual (a) e na combinação de classes (b).

4. DISCUSSÃO

As variáveis com maior correlação com a produtividade do milho (Figura 4a e Figura 4b), apresentaram a banda espectral do verde como característica comum entre si (GREEN e GNDVI). Conforme, Gitelson et al., (2002) a faixa de reflectância entre 520 e 630 nm (banda do verde) é mais sensível a clorofila, assim o GNDVI e a banda do verde tendem a ser sensíveis a uma ampla gama de variações da concentração de clorofila, possibilitando que tais variáveis tenham maior correlação com a produtividade quando comparada às demais variáveis (Pinto et al., 2022). Os resultados corroboram com o descrito por Kayad et al., (2019) que relataram o GNDVI e o GARVI (Green Atmospherically Resistant Vegetation Index) com maior correlação com a produtividade.

As maiores correlações da produtividade com as variáveis de entrada são correspondentes a segunda classe de soma de graus dias (WBH4), sendo essas variáveis as bandas espectrais do GREEN2, RED2, BLUE2 e NIR2 e os índices de

vegetação GNDVI2, NDVI2 e SAVI2 (Figura 4a e Figura 4b). A segunda classe de WBH corresponde aos estádios fenológicos reprodutivos denominados R1, R2, R3 e R4 (Figura 3a) corroborando com Kayad et al., (2019) e Barzin et al., (2020) ao relatarem comportamento similar com as maiores correlações das variáveis com a produtividade para os dados coletados na transição da fase vegetativa para a reprodutiva. Esse comportamento coincide com a mudança de fase na cultura do milho em cada um das WBHs, uma vez que no WBH4 as plantas já cessaram o acúmulo de nutrientes, acúmulo de peso seco e tamanho das espigas já foram definidos, nessa fase as plantas iniciam a fertilização do óvulos e posteriormente o enchimento dos grãos, mudando assim a fonte dreno das folhas para as espigas. Além disso, nesta fase a produtividade já está definida, sendo comprometida apenas em caso específicos como estresse hídrico (Silva et al., 2021).

Quando os WBHs foram utilizados separadamente para o desenvolvimento dos modelos (Figura 5 e Figura 6), as variáveis de entrada que proporcionaram os modelos mais precisos e acurados foram a combinação das bandas do visível (RGB) e do infravermelho (NIR), corroborando com Khanal et al., (2018) ao estudarem o desempenho de modelos de ML na estimativa da produtividade de milho, relataram maior importância das bandas NIR, GREEN, BLUE e RED com 100, 70, 60 e 58 respectivamente, em uma escala de 0 a 100, no desempenho do modelo de ANN.

Outro ponto a ser ressaltado é a variação que ocorre na precisão e acurácia dos modelos M3 quando a variável de entrada é alterada para WBH10 e WBH1 em relação a WBH4 (Figura 7), nesses modelos as variáveis foram constituídas apenas pela banda do Red e do NIR, desse modo, essa variação ocorrida pode estar relacionada com a ausência da banda do verde, a qual conforme já mencionada tem maior

sensibilidade a concentração de clorofila (Gitelson et al., 2002), no WBH10 a concentração de clorofila tende a aumentar com o desenvolvimento das plantas, enquanto que, no WBH1 essa concentração tende a reduzir, pois a cultura inicia o processo de senescência para a colheita (Fancelli, 1986).

As representações gráficas da produtividade observada e prevista (Figura 7a e Figura 7b) expressaram os erros ocorridos no modelo de previsão mais preciso e acurado para cada modelo. Porém, esses erros são comuns quando trabalhamos com modelagem agrícola conforme observado em diversos trabalhos de modelagem agrícola presente na literatura (Barzin et al., 2020; Canata et al., 2021; Haghverdi et al., 2018; Hunt et al., 2019; Maldaner et al., 2021b; Ramos et al., 2020; Saranya & Nagarajan. 2020; Tedesco et al., 2021a; Tedesco et al., 2021b; Venancio et al., 2019; Zhou et al., 2021).

Segundo Kayad et al., (2019) os erros ocorridos nos modelos que trabalham com dados de produtividade observada coletada por sensores podem ser impulsionados pelas distorções que ocorrem na coleta dos dados, como imprecisões de GPS (Global Positioning System) e sensores de rendimento ou até mesmo erros de calibração, explicando assim a diferença geral na precisão da previsão ou estimativa. Contudo, destacamos que para os nossos dados de produtividade das duas safras utilizamos a metodologia proposta por Maldaner et al., (2021a) que leva em consideração a variabilidade local e global no processo de remoção dos dados outliers, reduzindo as variações discrepantes entre os valores dos dados, ocasionadas por falhas que ocorrem durante a coleta dos dados.

Os resultados obtidos no presente estudo evidenciam que é possível prever a produtividade do milho utilizando a soma de graus dias como estratégia de seleção

do período de coleta dados de imagens de satélite. Contudo, destacamos que mesmo com resultados positivos obtidos no nosso estudo faz-se necessário o desenvolvimento de novos estudos com novas variedades com ciclos precoce, médio e tardios, bem como um maior número de imagens, permitindo assim uma divisão maior do número de classes de WBH e consequentemente identificando novos padrões de relação entre as informações espectrais da cultura coletados em uma faixa de soma de graus dias e a produtividade observada em campo.

Recomendamos também novos estudos que visem entender melhor as relações existentes entre a banda espectral do verde e a produtividade da cultura resultando em uma modelagem mais acurada e precisa. Outro ponto que poderá ser explorado em novos estudos é como as variáveis de entrada podem influenciar na acurácia e precisão dos modelos, de acordo com o período de soma de graus-dia em que os dados espectrais foram coletados.

5. CONCLUSÃO

No nosso trabalho propomos uma abordagem baseada na soma de graus dias da cultura do milho relacionada com as datas das imagens de satélite para organizar os dados de duas safras gerando modelos previsão com os modelos KNN e XGBoost. Conseguimos prever a produtividade com aproximadamente de 1050, 760 e 140 graus-dia antes da colheita, para as WBH10, WBH4 e WBH1, respectivamente. Os modelos mais acurados foram quando os dados utilizados como variáveis de entrada foram as bandas espectrais do vermelho, verde, azul e infravermelho próximo quando coletadas no intervalo de 801 a 1200 graus-dia da cultura (WBH4). Para os modelos desenvolvidos combinando a WBH foi possível prever a produtividade com erro médio

de 0,503 t ha⁻¹, sendo que a maior acurácia ocorreu com o uso das bandas espectrais do vermelho, verde, azul e infravermelho próximo.

AGRADECIMENTOS

Os autores agradecem o apoio da UNESP (Universidade Estadual Paulista) e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Código Financeiro 001) e a empresa VSL pelo fornecimento dos dados de produtividade.

Os autores agradecem a empresa VSL consultorias pelo fornecimento dos dados de produtividade.

6. REFERÊNCIAS

Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy* 10: 1046. <https://doi.org/10.3390/agronomy10071046>.

Arnold, C.Y. 1995. The determination and significance of the base temperature in a linear heat unit system. *Proceedings of the American Society for Horticultural Science* 74,430–450.

Barzin R, Pathak R, Lotfi H, Varco J, Bora GC (2020) Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn. *Remote Sensing* 12: 2392. DOI: <https://doi.org/10.3390/rs12152392>.

Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. 2021. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* 2021, 21, 3758. <https://doi.org/10.3390/s21113758>.

Canata TF, Wei MCF, Maldaner LF, Molin JP (2021) Sugarcane Yield Mapping Using High-Resolution Imagery Data and Machine Learning Technique *Remote Sensing* 13: 232. <https://doi.org/10.3390/rs13020232>.

Congedo L (2020) Semi-Automatic Classification Plugin Documentation. Release 4: 29 DOI: <http://dx.doi.org/10.13140/RG.2.2.29474.02242/1>. Accessed 19 Jun 2021.

CONAB (Companhia Nacional de Abastecimento). (2024) Acompanhamento da safra brasileira grãos safra 2023/24. Available online: <https://www.conab.gov.br/info-agro/safras>. Acessado em 24 de janeiro de 2024.

Chen T. & Guestrin C. (2016). Xgboost: a scalable tree boosting system. In Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <http://dx.doi.org/10.1145/2939672.2939785>

Folberth C, Baklanov A, Balkovic J, Skalský R, Khabarov N, Obersteiner M. (2019) Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. *Agricultural and Forest Meteorology* 264: 1–15. <https://doi.org/10.1016/j.agrformet.2018.09.021>.

Fancelli, AL. Plantas Alimentícias: guia para aula, estudo e discussão. Piracicaba: USP/ eSAIQ, 1986. 131 p.

Gitelson AA, Stark R, Grits U, Rundquist D, Kaufman Y, Derry D (2002) Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction. *International Journal of Remote Sensing* 23 (13): 2537-2562. <https://doi.org/10.1080/01431160110107806>.

Gitelson AA, Kaufman YJ, Merzlyak MN (1996) Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* 58: 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7).

Haghverdi A, Allen RAW, Leib BG, (2018) Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Computers and Electronics in Agriculture* 152:186–197. <https://doi.org/10.1016/j.compag.2018.07.021>.

Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3): 295–309. DOI: [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).

Hunt ML, Blackburn GA, Carrasco L, Redhead JW, Rowland CS, (2019) High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment* 233: 111410. <https://doi.org/10.1016/j.rse.2019.111410>.

INMET (Instituto Nacional de Meteorologia) (2023) Dados meteorológicos. Available online: <https://bdmep.inmet.gov.br/>. Accessed on 18 January 2023.

Kamir, E.; Waldner, F.; Hochman, Z. (2020) Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal Photogramm. Remote Sensing* 160, 124–135. <https://doi.org/10.1016/j.isprsjprs.2019.11.008>.

Kayad A, Sozzi M, Gatto S, Marinello F, Pirotti, F (2019). Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. *Remote Sensing* 11, 2873. <https://doi.org/10.3390/rs11232873>.

Khanal S, Fulton J, Klopfenstein A, Douridas N, Shearer S (2018) 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* 153, 213–225. <https://doi.org/10.1016/j.compag.2018.07.016>.

Maldaner LF, Molin JP, Spekken M (2021a) Methodology to filter out outliers in high spatial density data to improve maps reliability. *Scientia Agricola* 79(1): e20200178. <http://doi.org/10.1590/1678-992X-2020-0178>.

Maldaner LF, Corrêdo LP, Canata TF, Molin JP (2021b) Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. *Computers and Electronics in Agriculture* 181:105945. <https://doi.org/10.1016/j.compag.2020.105945>

Pinto A A, Zerbato C, Rolim G S, Barbosa Júnior M R Silva, L F V, Oliveira R P (2022). Corn grain yield forecasting by satellite remote sensing and machine learning models. *Agronomy journal*, 1-20. <https://doi.org/10.1002/agj2.21141>.

Ramos APM, Osco LP, Furuya DEG, Gonçalves WN, Santana DC, Teodoro LPR Silva Junior CA, Capristo-Silva GF, Li J, Baio FHR, Marcato Junior J, Teodoro PE, Pistori H (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture* 178: 105791. DOI: <https://doi.org/10.1016/j.compag.2020.105791>.

Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS. In *Proceedings of the Earth Resources Technology Satellite—1 Symposium*, Washington, DC, USA, 10–14 December 1974; pp. 309–317.

Silva DFD, Garcia PHDM, Santos GCL, Farias IMSC, Pádua GVG, Pereira PHB, Silva FE, Batista RF, Gonzaga Neto S, Cabral AMD (2021). Morphological characteristics, genetic improvement and planting density of sorghum and corn crops: a review. *Research, Society and Development*, 10(3) e12310313172. <https://doi.org/10.33448/rsd-v10i3.13172>.

Sonobe R, Yamaya, Y, Tani H, Wang X, Kobayashi N, Mochizuki K. (2018) Crop classification from Sentinel-2- derived vegetation indices using ensemble learning. *Journal of Applied Remote Sensing*, 12 (2), 026019 (2018). <https://doi.org/10.1117/1.JRS.12.026019>.

Tedesco D, Moreira ABR, Barbosa Júniora MR, Papa JP, Silva RP. (2021a) Predicting on multi-target regression for the yield of sweet potato by the market class of its roots upon vegetation indices. *Computers and Electronics in Agriculture* 191 106544. <https://doi.org/10.1016/j.compag.2021.106544>.

Tedesco D, Oliveira MF, Santos AD, Silva EHC, Rolim GS, Silva RP (2021b). Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons. *European Journal of Agronomy* 129 126337. <https://doi.org/10.1016/j.eja.2021.126337>.

Venancio LP, Mantovani EC, Amaral CH, Neale CMU, Gonçalves IZ, Filgueiras R, Campos I (2019) Forecasting corn yield at the farm level in Brazil based on the FAO-66 approach and soil-adjusted vegetation index (SAVI). *Agricultural Water Management* 225 (20): 105779. <https://doi.org/10.1016/j.agwat.2019.105779>.

Wang T, Shi J, Letu H, Ma Y, Li X, & Zheng, Y. (2019). Detection and removal of clouds and associated shadows in satellite imagery based on simulated radiance fields. *Journal of Geophysical Research: Atmospheres*, 124, 7207– 7225. <https://doi.org/10.1029/2018JD029960>.

Jiang Z, Liu C, Hendricks N P, Ganapathysubramanian B, Hayes D J, Sarkar S (2018) Predicting County Level Corn Yields Using Deep Long Short Term Memory Models. *arXiv*. <https://doi.org/10.48550/arXiv.1805.12044>

Zhou X, Kono Y, Win A, Tsutomu Matsui T, Tanaka TST (2021) Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. *Plant Production Science*, 24(2): 137-151. <https://doi.org/10.1080/1343943X.2020.1819165>.

CAPÍTULO 3: ESTIMATIVA DA PRODUTIVIDADE DE CANA-DE-AÇÚCAR UTILIZANDO APRENDIZADO DE MÁQUINA E ANÁLISE DE COMPONENTES PRINCIPAIS DOS DADOS ESPECTRAIS E DO MOTOR DA COLHEDORA

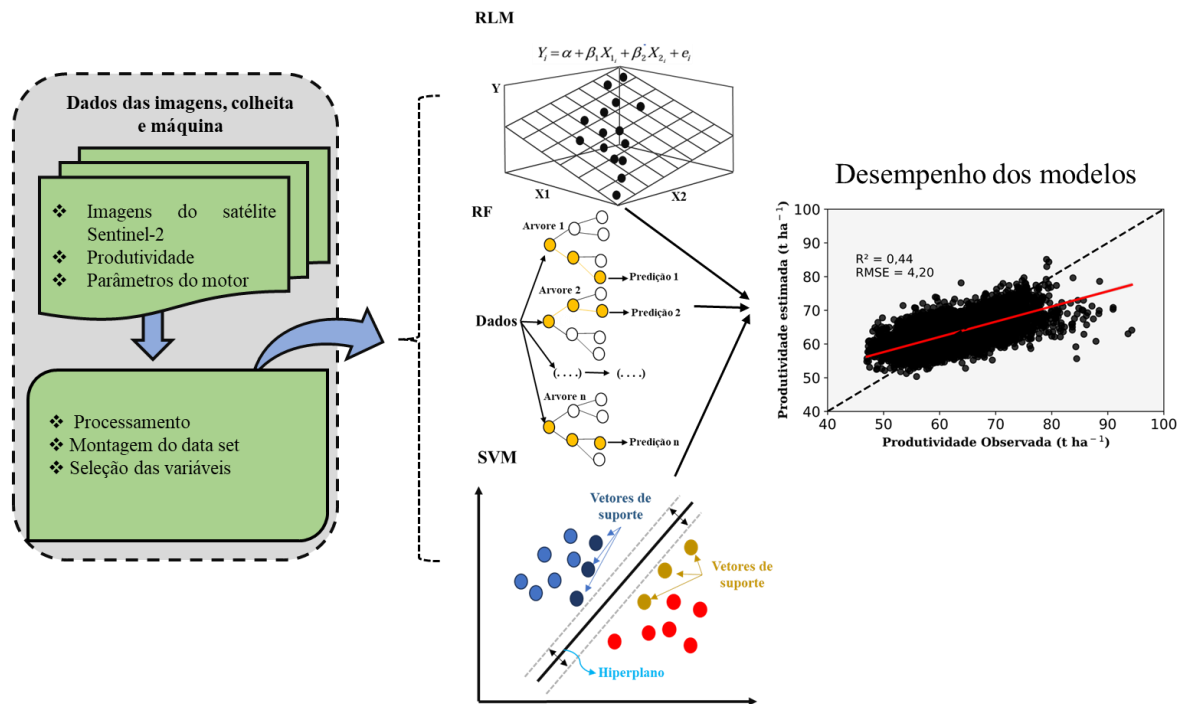


Figura 1. Resumo gráfico exemplificando as etapas de obtenção dos dados, geração dos modelos e resultados.

Resumo

O mapeamento da variabilidade espacial da produtividade fornece informações essenciais para um manejo eficiente das culturas agrícolas permitindo melhores tomadas de decisões e uso de práticas de agricultura de precisão. Na cana-de-açúcar a estimativa espacializada da produtividade ainda é desafiador e pouco explorado quando comparado a culturas anuais como milho e soja. Propomos nesse trabalho uma abordagem baseada na análise de componentes principais da integração de séries temporais de imagens do Sentinel-2 associada a dados do motor obtidas em campo, como rotação do motor por minuto, velocidade de colheita e de consumo de combustível. Os objetivos específicos foram comparar o desempenho dos MLR

(Multiple Linear Regression), RF (Random Forest) e SVM (Support Vector Machine) e selecionar o conjunto de variáveis que apresenta melhor acurácia nos modelos de ML avaliados. Foi utilizado um cultivo comercial de cana-de-açúcar para a obtenção dos dados de produtividade e variáveis do motor, nessa área também foram adquiridas as imagens do Sentinel-2 das duas safras estudadas. O estudo foi desenvolvido em uma área comercial. Utilizamos três cenários nesse estudo, no primeiro cenário a produtividade foi estimada apenas com os dados das bandas espectrais mais os do motor; o segundo cenário os índices de vegetação mais os dados do motor e o terceiro cenário todas as bandas, índices de vegetação e dados do motor. Esses dados foram as variáveis de entrada em modelos de aprendizado de máquina para estimar a produtividade da cana-de-açúcar. O modelo SVM utilizando as PCAs da integração de séries temporais de bandas espectrais e dados do motor apresentou maior acurácia com RMSE de 4,20 t ha⁻¹.

Palavras-chave: Imagens orbitais, índices de vegetação, PCA, agricultura de precisão, agricultura digital.

Estimation of sugar cane productivity using machine learning and principal component analysis of spectral and harvesting motor data

ABSTRACT

Mapping the spatial variability of yield provides essential information for efficient management of crops, allowing for better decision-making and the use of precision agriculture practices. In sugarcane, spatialized productivity estimation is still challenging and little explored when compared to annual crops such as corn and soybeans. In this work, we propose an approach based on the principal components analysis of the integration of time series of Sentinel-2 images associated with engine data obtained in the field, such as engine rotation per minute, harvesting speed, and fuel consumption. The specific objectives were to compare the performance of MLR (Multiple Linear Regression), RF (Random Forest), and SVM (Support Vector Machine) and select the set of variables that presents the best accuracy in the ML models evaluated. A commercial sugarcane crop was used to obtain productivity data

and engine variables. Sentinel-2 images of the two crops analyzed were also acquired in this area. We used three scenarios in these studies. In the first scenario, productivity was estimated only with data from the spectral bands plus the engine; in the second scenario the vegetation indices plus engine data and in the third scenario all bands, vegetation indices, and engine data. These data were the input variables in machine learning models to estimate sugarcane productivity. The SVM model using PCAs from the integration of time series of spectral bands and engine data showed greater accuracy with RMSE of 4.20 t ha⁻¹.

Keywords: Orbital images, vegetation indices, PCA, precision agriculture, digital agriculture.

1. INTRODUÇÃO

A cana-de-açúcar (*Saccharum officinarum* L.) é cultivada em países com clima tropicais e subtropicais. A cultura é matéria prima para diversos produtos, tais como açúcar e etanol, bem como fertilizantes orgânicos a partir da torta de filtro e vinhaça (Oliveira et al., 2022). No Brasil a cana-de açúcar é amplamente cultivada, com grande relevância no comércio agrícola do país.

Considerando a importância econômica que a cultura da cana-de-açúcar tem para o país, a cultura é bastante estudada por instituições do setor público e privado, porém um dos grandes desafios é a estimação da variabilidade espacial da produtividade (Canata et al., 2021), surgindo assim a necessidade de desenvolvimento de estudos que busquem alternativas para facilitar a avaliação da variabilidade espacial da produtividade, com resultados significativos (Somard et al., 2023).

Revisando a literatura sobre estudos de estimação de produtividade da cana-de-açúcar, podemos encontrar trabalhos com resultados promissores utilizando dados de

sensoriamento remoto espectral e modelos de aprendizado de máquina, seja estimativa a nível de bloco (Dimov et al., 2022; Nihar et al., 2022; Zhu et al., 2023 e Alemán-Montes et al., 2023) ou de campo (Canata et al., 2021), Já Maldaner et al., (2021a) utilizou uma abordagem baseada no uso de aprendizado de máquina associado a parâmetros do motor disponíveis nas próprias máquinas para estimar a produtividade, obtendo raiz do erro quadrático médio (RMSE) de 7.6 t ha⁻¹ para uma rede neural artificial.

Tais estudos, proporcionam uma melhoria por alternativas para facilitar a avaliação da variabilidade espacial da produtividade, contudo acreditamos que o uso de apenas um tipo de variável de entrada nos modelos pode não representar toda a complexidade existentes no sistema agrícola.

Diante do problema apresentado acima propomos nesse trabalho a estimação da produtividade de cana-de-açúcar com modelos de machine learning, regressão linear múltipla, Random forest, Support Vector Machine a partir de dados temporais de imagens do Sentinel-2 e do motor, como rotação do motor por minuto, velocidade de colheita e de consumo de combustível.

2. MATERIAL E MÉTODOS

2.1. Local de estudo

O estudo foi desenvolvido em uma área comercial de 51,2 hectares, em campo localizado na região centro-oeste do estado de Mato Grosso do sul (22°03'25"S, 54°00'51"W; 523 m de altitude), durante as safras 2020/21 e 2021/22. O clima na região de estudo é classificado como Cwa mesotérmico úmido com verão quente e

invernos seco, A precipitação média anual é de aproximadamente 1400 mm e temperatura média anual 23,6 °C (Fietz et al., 2017).

Em ambas as safras foi utilizado a variedade RB855156 material de maturação precoce com colheita de abril a julho. A safra 2020/21(quarta soca) foi colhida no dia 15 de abril de 2021, com a colheita anterior sendo realizado na data de 20 de abril de 2020, totalizando 360 dias após a colheita (DAC). Já a safra 2021/22 (quinta soca) foi colhida no dia 10 de maio de 2022, totalizando 390 DAC.

2.2.Obtenção e processamento de imagens de satélite

Foram utilizadas 38 imagens orbitais do Sentinel-2 considerando as safras de cana-de-açúcar 2020/2021 e 2021/2022, sendo 18 imagens na primeira safra e 20 imagens na segunda. Todas as imagens foram obtidas da plataforma Sentinel Hub EO Browser (sentinel-hub.com), que disponibiliza imagens com as correções atmosféricas realizadas. Após baixar todas as imagens com baixa cobertura de nuvens (<1%), realizamos a projeção na zona WGS/UTM 21S, criando em seguida um buffer de cinco metros entorno da área de interesse e recortando para extração dessa área.

No estudo selecionamos cinco bandas espectrais do Sentinel-2, Red, Green, Blue, Near-infrared e Red-Edge (Tabela 1). Além das bandas espectrais, foram obtidos das imagens cinco índices de vegetação, Green Normalized Difference Vegetation Index (GNDVI); Normalized Difference Vegetation Index (NDVI); Normalized Difference Red-Edge Index (NDRE), Enhanced Vegetation Index (EVI) e Enhanced Vegetation Index 2 (EVI2), tais índices são citados na literatura no

monitoramento da produtividade de diversas culturas (Barzin et al., 2020; Ramos et al., 2020; Canata et al., 2021; Zhou et al., 2021).

Tabela 1. Especificações das bandas espectrais do Sentinel-2 utilizado nesse estudo.

Bandas	Comprimento de onda central (nm)	Resolução		
		Espacial (m)	Temporal (Dias)	Radiométrica (Bits)
B2 - Blue	490			
B3 - Green	560	10	5	12
B4 - Red	665			
B8 - Near-infrared	842			
B5 – Red - Edge	705	20		

A soma de graus dia (SGD) para todas as imagens das duas safras foram calculadas a partir de 21 de abril de 2020 e a partir de 16 de abril de 2021. Para determinar a SDD, utilizamos dados de uma estação meteorológica automática do Instituto nacional de meteorologia (portal.inmet.gov.br) que coleta dados a cada hora, localizada aproximadamente a 35 km da região dos campos de produção. O cálculo foi realizado de acordo com a data de aquisição de cada imagem de satélite, seguindo o método de Arnold (1959),

$$SDD = \sum \left(\frac{T_{max} + T_{min}}{2} \right) - T_{base} \quad (1)$$

Em que: SDD é a soma de graus-dia, Tmax é a temperatura máxima do ar (°C), Tmin é a temperatura mínima do ar (°C), e Tbase é a temperatura base mais baixa para o desenvolvimento da cultura (10 °C) como sugerido por Nava et al., (2016).

Padronizamos os períodos (Tabela 2) para processamento dos dados das duas safras, com intervalos de 500 graus-dia devido a desigualdade na disponibilidade de

imagens orbitais entre uma safra e outra, utilizamos o valor médio dos dados de imagens orbitais que compuseram cada intervalo de soma de graus-dias. Por exemplo, na safra 2020/2021, o valor definido para o SGD1 foi o valor médio das imagens das datas 10/05, 30/05 e na safra 2021/2022 foram as imagens de 10/05, 25/05.

Tabela 2. Datas das imagens orbitais e respectivos estádios fenológicos da cana-de-açúcar.

ID	Estádios fenológicos	SGD	Datas das imagens orbitais (dia/mês)	
			Safra	Safra
			2020/2021	2021/2022
SGD1	Inicial (0-60 dias)	0-500	10/05, 30/05	10/05, 25/05
SGD2	Perfilhamento	501-1000	19/06, 24/07	14/06, 09/07
SGD3	(60-150 dias)	1001-1500	08/08, 28/08	18/08, 07/09
SGD4	Desenvolvimento (150-240 dias)	1501-2000	12/09, 07/10	27/09, 12/10
SGD5		2001-2500	17/10, 06/11	22/10, 16/11
SGD6		2501-3000	21/11	21/11, 06/12
SGD7	Fase de Maturação	3001-3500	26/12, 10/01	26/12, 05/01
SGD8	inicial e terço médio	3501-4000	19/02	25/01, 09/02
SGD9	(240-330 dias)	4001-4500	11/03, 21/03	24/02, 16/03
SGD10	Maturação Final (>330 dias)	>4500	31/03 10/04	05/04, 30/04

ID: identificação; SGD: Soma de graus-dia: Estádios fenológicos adaptado de Molijn et al., (2019)

2.3. Aquisição dos dados de produtividade e do motor

Os dados de produtividade de cana-de-açúcar para ambas as safras foram obtidos de três colhedoras John Deere modelo CH570, equipadas com sinal de

correção diferencial Real-Time Kinematic (RTK) e um sistema comercial de sensores que se baseia no peso total de cana colhida distribuídas espacialmente no campo, considerando as diferenças das pressões hidráulicas do corte de base, mais detalhes sobre esse sistema foi descrito por Maldaner et al., (2022).

Os dados rotação do motor por minuto (RPM), Velocidade (km h^{-1}) e consumo de combustível (l h^{-1}), foram obtidos a partir de sensores disponíveis nas colhedoras que auxiliam no funcionamento da máquina, sendo transmitidos via CAN (Controller Area Network) para o computador de bordo da colhedora que integra esses dados com as informações de produtividade e coordenadas corrigidas do RTK, disponibilizando essa informações em uma ambiente on-line para consumo.

Após a colheita, os dados foram processados removendo os outliers identificados no processamento de dados conforme metodologia descrita por Maldaner et al., (2021b) que leva em consideração a variabilidade local e global no processamento dos dados. Após a filtragem dos dados, foram gerados os mapas de produtividade e do motor das duas safras, utilizando o interpolador IDW (Inverse Distance Weighted), o tamanho do pixel foi equivalente ao tamanho do pixel das imagens do Sentinel-2 (10 m), permitindo que tais dados fossem utilizados para o treinamento e teste dos modelos de estimação. O software utilizado na interpolação e geração dos mapas foi Quantum GIS, versão 3.16.

2.4. Variáveis de entradas dos modelos

Os dados foram organizados em três cenários, sendo o primeiro cenário os dados das bandas espectrais mais os do motor; o segundo cenário os índices de vegetação mais os dados do motor e o terceiro cenário todas as bandas, índices de

vegetação e dados do motor. Todas as variáveis de cada cenário foram submetidas a análise de componentes principais (PCA), sendo o número total de variáveis em cada cenário considerado como o número máximo de componentes. Nos modelos utilizamos os próprios valores das PCAs como variáveis de entradas dos modelos. Para a seleção do número de variáveis de entradas dos modelos em cada cenário usamos como base na regra do cotovelo em que o valor de k (variância explicada na PCA) é o ponto após o qual a inclinação entre dois pontos esteja próxima ou na horizontal, conforme sugerido por (Yang et al., 2020).

2.5. Modelos de aprendizado de máquina

A RLM é um dos algoritmos mais usados em ML. O algoritmo busca ajustar uma equação linear para minimizar a soma residual dos quadrados entre os dados observados e os previstos pela aproximação linear (Kannangara et al., 2018).

O algoritmo RF uma técnica de aprendizagem que faz uso de um conjunto árvores de decisão, com cada árvore sendo dependente dos valores de um vetor aleatório. No RF os dados são divididos recursivamente em unidades mais homogêneas em cada árvore (nodes), dessa forma, os valores previstos são a média da resposta ajustada de todas as árvores individuais originárias de cada amostra inicializada (Hariharan, 2021).

O SVM é uma das técnicas de aprendizado, que busca minimizar o erro adicionando ao hiperplano e maximizar a margem entre os valores previstos e observados (Zhou et al., 2021).

Para os modelos, RF e SVM a seleção do melhor ajuste dos parâmetros (Tabela 3) ocorreu por meio do algoritmo de otimização Bayesiano, implementado usando a biblioteca Scikit-optimize de Python, usando o método `gp_minimize` do módulo `skopt`.

Tabela 3. Valores dos parâmetros utilizados pelos modelos de Machine Learning em cada cenário.

Modelo	Parâmetros	Cenário 1	Cenário 2	Cenário 3
RF	n_estimators	242	247	243
	max_depth	29	26	26
	min_samples_split	2	3	2
	min_samples_leaf	1	2	1
SVM	Kernel	RBF	RBF	RBF
	Gamma	auto	auto	auto
	C	20	23	15
	Épsilon	0,85	0,94	0,99

Cenário1: PCA bandas + parâmetros do motor; cenário 2 PCA IVs + parâmetros do motor; cenário 3 PCA todas as variáveis + parâmetros do motor.

Os dados foram divididos em treinamento e teste, sendo os dados da primeira safra usado para treinamento e os dados da segunda safra usados para teste.

2.6. Análise de desempenho dos modelos

Foram obtidos os valores do coeficiente de determinação (R^2), o erro absoluto médio (MAE) e raiz do erro quadrático médio (RMSE) como métricas de precisão e acurácia, respectivamente. A acurácia está relacionada à repetição em relação aos dados observados e previstos, já a precisão está relacionada ao desempenho geral dos modelos com relação aos valores previstos em relação à reta 1:1.

3. RESULTADOS

3.1. Estatística descritiva dos dados da colhedoras e produtividade

Nos dados de consumo de combustível de safra 2020/2021 a variação foi de 35,8 a 69,9 l h⁻¹ com desvio padrão de 5,7 l h⁻¹, já na safra 2021/2022 a variação foi de 36,6 a 71,5 l h⁻¹ com desvio padrão de 5,7 l h⁻¹ (Tabela 4).

Tabela 4. Estatística descritiva dos dados do motor e da produtividade.

Safra 2020/2021				
	Consumo (l h ⁻¹)	RPM	Velocidade (k h ⁻¹)	Produtividade (t ha ⁻¹)
Máximo	69,9	1896,0	6,7	100,0
Média	53,6	1802,4	5,4	63,9
Mínimo	35,8	1701,0	2,0	30,0
DP	5,7	21,3	0,7	9,0
25%	50,3	1792,0	5,0	59,7
50%	53,2	1801,0	5,5	63,8
75%	56,3	1810,0	5,9	67,8
CV (%)	10,6	1,2	13,9	14,4
Safra 2021/2022				
Máximo	71,5	1892,0	6,9	95,9
Média	54,3	1803,2	4,9	63,3
Mínimo	36,6	1703,0	1,0	45,0
DP	5,7	21,5	0,8	7,9
25%	50,7	1793,0	4,4	55,7
50%	53,6	1801,0	5,1	63,5
75%	57,2	1810,0	5,6	70,6
CV (%)	10,5	1,2	15,7	12,6

RPM: Rotação por minuto do motor; DP: Desvio Padrão; CV: Coeficiente de variação.

Nos dados de RPM e velocidade de deslocamento na colheita a variação foi de 1701 a 1896 rpm e de 02 a 6,7 k h^{-1} , respectivamente na safra 2020/2021, já na safra 2021/2022 a variação foi de 1703 a 1892 rpm para o RPM e de 01 a 6,9 para a velocidade, com desvio padrão de 21,3 e 21,5 rpm no RPM e 0,7 e 0,8 k h^{-1} nas safras 2020/2021 e 2021/2022, respectivamente. Os coeficientes de variação (CV) dos dados de parâmetros do motor podem ser classificados como baixo para RPM e médio nos dados de consumo de combustível e velocidade, conforme classificação de Pimentel Gomes (2009).

Na produtividade a média foi de 63,9 t ha^{-1} , com desvio padrão de 9,1 t ha^{-1} e variação de 30,0 a 100 t ha^{-1} na safra 2020/2021, enquanto na safra seguinte a produtividade média foi de 63,3, com desvio padrão de 7,9 t ha^{-1} com mínima e máxima de 45 e 95,9 t ha^{-1} , respectivamente. O CV dos dados produtividade pode ser classificados médio em ambas as safras.

3.2. Seleção de variáveis preditoras PCAs

A seleção do número de PCAs para cada cenário ocorreu com base na regra do cotovelo, em que foi considerado o número ideal de variáveis quele no qual a variância explicada na PCA apresenta-se inclinação entre dois pontos próxima horizontal, sendo o cenário um (Figura 2 A), composto pelas bandas espectrais do azul (B2), verde (B3), vermelho (B4), infravermelho próximo (B8) e borda do vermelho (B5) mais os dados de parâmetros do motor totalizando 53 variáveis inseridas na análise de componentes principais, nesse cenário o número de PCAs selecionadas foram cinco com 83,8% da variância explicada (Tabela 5).

No cenário dois o número de PCAs selecionadas foram quatro com 86,19% da variância explicada, esse cenário foi constituído pelos índices de vegetação GNDVI,

NDVI, NDRE, EVI e EVI2, juntamente com os dados de parâmetros do motor totalizando 53 variáveis inseridas na análise de componentes principais desse cenário. Já o cenário três resultou a fusão dos dados das bandas espectrais, índices de vegetação e parâmetros do motor totalizando 103 variáveis inseridas na análise de componentes principais desse cenário, sendo cinco o número de PCAs selecionadas com 85,23 % da variância explicada.

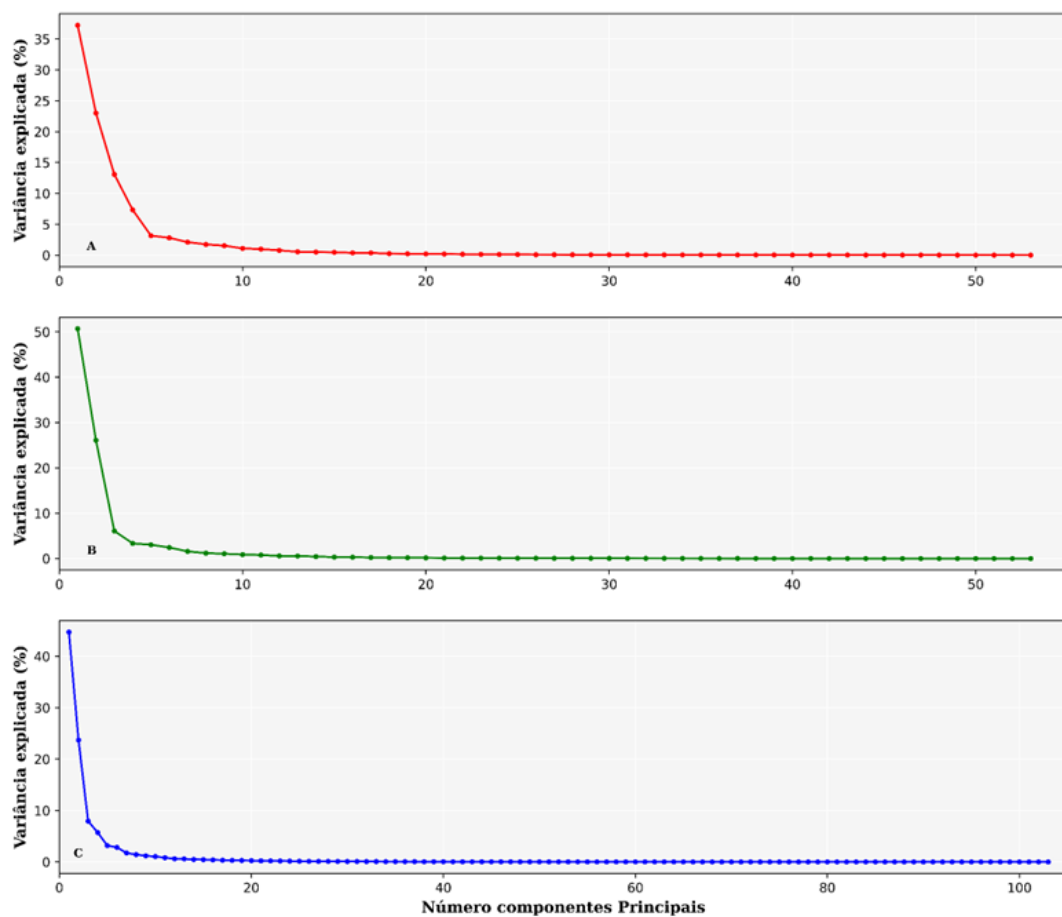


Figura 2. Variância explicada pelo número de componentes principais utilizados (Regrado Cotovelo): A- Cenário1: PCA bandas + parâmetros do motor; B - cenário 2 PCA IVs + parâmetros do motor; C - cenário 3 PCA todas a variáveis.

Tabela 5. Variância explicada por cada variável de PCA selecionada nos cenários um, dois e três.

Número de PCA	Variância explicada (%)		
	Cenário1	Cenário2	Cenário3
1	37,22	50,70	44,71
2	23,00	26,07	23,70
3	13,08	6,08	7,94
4	7,34	3,34	5,71
5	3,16	-	3,17
Total	83,8	86,19	85,23

Cenário1: PCA bandas + parâmetros do motor; cenário 2 PCA IVs + parâmetros do motor; cenário 3 PCA todas as variáveis + parâmetros do motor.

3.3. Avaliação de Precisão e Acurácia dos modelos.

O modelos MLR, RF e SVM tiveram diferentes acurácias em seus desempenhos para os diferentes cenários (Tabela 6). Nos modelos MRL, RF e SVM os modelos mais acurados foram no cenário um (PCA bandas e dados do motor), com R^2 de 20, MAE de 3,89 t ha⁻¹ e RMSE 5,00 t ha⁻¹, para o MRL, R^2 de 0,42, MAE de 3,10 t ha⁻¹ e RMSE 4,25 t ha⁻¹, para o RF e R^2 de 0,44, MAE de 3,00 t ha⁻¹ e RMSE 4,20 t ha⁻¹ no SVM, analisando o conjunto de dados de teste.

Para os cenários dois e três a precisão e acurácia foram inferiores aos cenário um, com R^2 variando de 16 a 0,42 e RMSE de 4,26 a 5,12 t ha⁻¹, considerando os dados de teste de ambos os cenários (Tabela 6). Nesse estudo as bandas espectrais mais os dados de parâmetros do motor foram mais eficazes com maior R^2 e o menor MAE e RMSE na estimativa da produtividade da cana-de-açúcar.

Tabela 6. Desempenho dos algoritmos RLM, RF e SVM na estimativa de produtividade nos cenários um, dois e três. Simbologia: RLM- Regressão Linear Múltipla, RF- Random Forest, SVM- Support Vector Machine, R^2 – coeficiente de determinação (precisão), MAE – erro absoluto médio (acurácia) e RMSE- Raiz do Erro quadrático médio (acurácia).

Cenários	RLM		
	R^2	MAE	RMSE
Cenário1	0,20	3,89	5,00
Cenário2	0,16	4,04	5,12
Cenário3	0,18	3,95	5,07
RF			
Cenário1	0,42	3,10	4,25
Cenário2	0,38	3,21	4,39
Cenário3	0,41	3,15	4,30
SVM			
Cenário1	0,44	3,00	4,20
Cenário2	0,39	3,15	4,36
Cenário3	0,42	3,08	4,26

Cenário1: PCA bandas e parâmetros do motor; cenário 2 PCA IVs e parâmetros do motor; cenário 3 PCA todas as variáveis e parâmetros do motor.

Na dispersão da produtividade observada e prevista para os modelos com maior precisão e acurácia (cenário um no modelo SVM), nota-se que, o modelo treinado apresentou tendência de estimativa abaixo dos valores observados de produtividade, quando os valores de produtividade estão acima de 65 t ha⁻¹ (Figura 3), sendo os maiores erros observados para produtividades superiores a 80 t ha⁻¹. Enquanto valores inferiores a 65 t ha⁻¹ a tendência erro de valores estimados acima da produtividade observada.

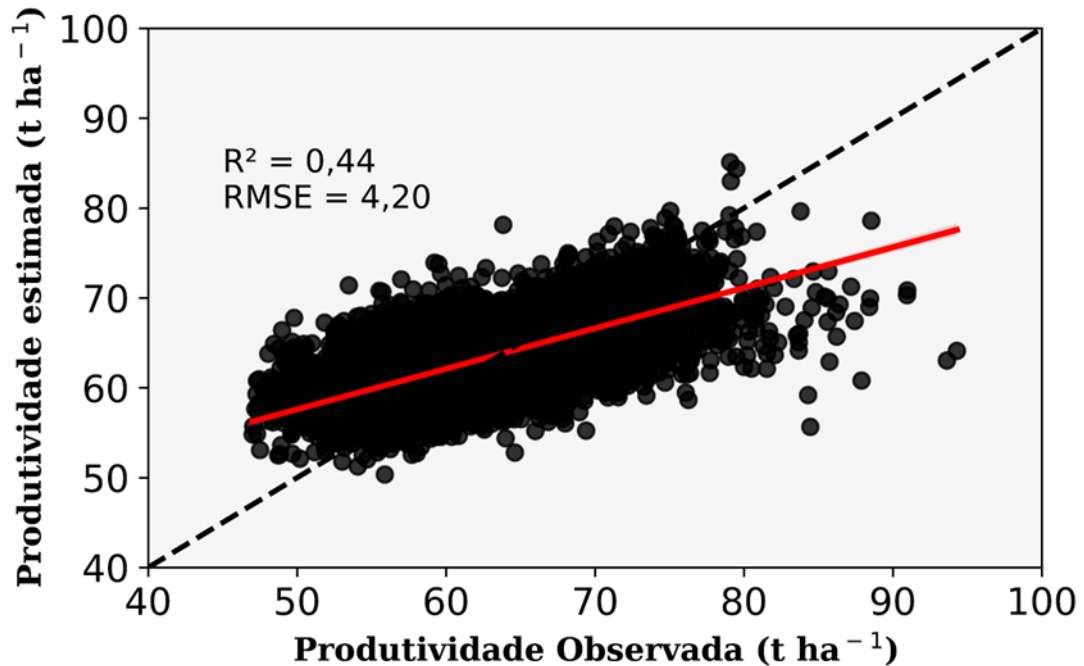


Figura3. Produtividade observada e estimada para o modelo SVM no cenário 1.

4. DISCUSSÃO

Nesse estudo utilizamos a integração de séries temporais de imagens do Sentinel-2 e parâmetros do motor, ambos os tipos de variáveis preditoras (dados espectrais e parâmetros do motor) já foram usadas separadamente na literatura para estabelecer relação com a produtividade da cana de açúcar (Zhao et al., , 2016; Dubey et al., , 2018; Canata et al., 2021; Maldaner et al., 2021a), contudo, esse estudo é o pioneiro a utilizar uma abordagem baseada na análise de componentes principais da integração de séries temporais de imagens do Sentinel-2 e parâmetros do motor, como variáveis de entradas em modelos de aprendizado de máquina para estimar a produtividade da cana-de-açúcar.

A análise de PCA das séries temporais das bandas espectrais em conjunto com os dados de parâmetros do motor possibilitaram o desenvolvimento de modelos mais

acurados quando comparado ao uso de índices de vegetação que são oriundos de equações matemáticas envolvendo as bandas espectrais, até mesmo a junção de todos os dados (bandas, IVs e parâmetros do motor) não foram capazes de superar a acurácia dos modelos envolvendo somente as bandas espectrais. Canata et al., (2021), em um estudo de mapeamento da produtividade da cana-de-açúcar usando dados de imagens de alta resolução e técnica de aprendizado de máquina a nível de campo, também relataram que as bandas espectrais foram superiores nos modelos RF (RMSE de 4,63 t ha⁻¹) e MLR (RMSE 3,46 t ha⁻¹), em relação ao uso de IVs como variáveis preditoras.

Conforme observado em outros trabalhos o RLM obteve as menores acurácias em todos os cenários, sendo o modelo com maior desempenho no cenário1 com R² e RMSE de 0,21 e 5,00 t ha⁻¹, respectivamente, contudo tais valores foram inferiores no R², porém superiores no RMSE aos observado por Canata et al., (2021) com R² 0,47 e RMSE 6,11 t ha⁻¹, ao trabalharem com bandas espectrais como variáveis preditoras. Ao estudarem potencial de produtividade da cana-de-açúcar em vários genótipos usando a refletância do dossel com o modelo MLR com bandas espectrais e regressão linear simples com NDVI na estimativa da produtividade da cana-de-açúcar Zhao et al., (2016), relataram que a MLR não obteve ganhos significativos de acurácia quando comparado à regressão linear simples com o NDVI.

O modelo RF que normalmente apresenta bons resultados quando comparados com outros algoritmos de predição em outras culturas (Ramos et al., 2020; Zhou et al., 2020; Pinto et al., 2022), no presente estudo foi o segundo melhor algoritmo, sendo o SVM no cenário1 o modelo que obteve melhor desempenho (maior R² e menor MAE

e RMSE). Guo et al., (2020) relataram que o SVM obteve melhor desempenho ao estimarem a produtividade da cultura do milho, assim como no presente trabalho.

A análise de componentes principais da integração de séries temporais de imagens do Sentinel-2 e parâmetros do motor permitiu o desenvolvimento de uma metodologia baseada nas técnicas de aprendizado de máquina para auxiliar o avanço do entendimento da variabilidade espaço-temporal da cana-de-açúcar. O nosso estudo considerou duas safras (dois anos) e um local único de estudo, contudo, para o desenvolvimento de um modelo robusto, capaz de ser aplicado nas mais diversas condições são necessários novos estudos, incluindo também variáveis ambientais, diversas variedades de cana-de-açúcar, tipos de solo, práticas de gestão agrícola e outros fatores que afetem a produtividade da cultura. Conforme relatado por Canata et al., (2021) as condições ideais para refinar a modelagem devem envolver múltiplas estações e vários campos comerciais, obtendo assim as mais variadas condições que afetam a variabilidade espaço-temporal da cana-de-açúcar.

5. CONCLUSÃO

A abordagem baseada na análise de componentes principais da integração de séries temporais de imagens do Sentinel-2 e dados do motor, como variáveis de entradas em modelos de aprendizado de máquina para estimar a produtividade da cana-de-açúcar, permitiu o desenvolvimento de modelos de estimativa de produtividade na área estudada. O modelo SVM utilizando as PCAs da integração de séries temporais de bandas espectrais e dados parâmetros dos motor apresentou maior precisão e acurácia (R^2 de 0,44 e RMSE 4,20 t ha⁻¹).

AGRADECIMENTOS

Os autores agradecem o apoio da UNESP (Universidade Estadual Paulista) e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Código Financeiro 001).

6. REFERÊNCIAS

Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. **Agronomy** 10, 1046. <https://doi.org/10.3390/agronomy10071046>.

Alemán-Montes B, Zabala A, Henríquez C, Serra P (2023) Modelling Two Sugarcane Agro-Industrial Yields Using Sentinel/Landsat Time-Series Data and Their Spatial Validation at Different Scales in Costa Rica. **Remote Sensing**, 15, 5476. <https://doi.org/10.3390/rs15235476>

Arnold, C.Y. 1995. **The determination and significance of the base temperature in a linear heat unit system**. Proceedings of the American Society for Horticultural Science 74,430–450.

Barzin R, Pathak R, Lotfi H, Varco J, Bora GC (2020) Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn. **Remote Sensing**, 12, 2392. DOI: <https://doi.org/10.3390/rs12152392>.

Canata TF, Wei MCF, Maldaner LF, Molin JP (2021) Sugarcane Yield Mapping Using High-Resolution Imagery Data and Machine Learning Technique. **Remote Sensing**, 13, 232. <https://doi.org/10.3390/rs13020232>.

Dimov D, Uhl JH, Low F, Seboka GN, (2022) Sugarcane yield estimation through remote sensing time series and phenology metrics. **Smart Agricultural Technology**, 2, 100046. <https://doi.org/10.1016/j.atech.2022.100046>

Dubey SK, Gavli AS, Yadav SK, Sehgal S, Ray SS (2018) Remote Sensing-Based Yield Forecasting for Sugarcane (*Saccharum officinarum* L.) Crop in India. Journal of the Indian Society. **Remote Sensing**, 46, 1823–1833. <https://doi.org/10.1007/s12524-018-0839-2>

FIETZ CR., FISCH GF, COMUNELLO E, FLUMIGNAN DL. (2017) **O clima da região de Dourados, MS**. 3. ed. rev. e atual. Dourados: Embrapa Agropecuária Oeste, 2017. 34 p. (Embrapa Agropecuária Oeste. Documentos, 138).

Guo Y, Wang H, Wu Z, Wang S, Sun H, Senthilnath J, Wang J, Bryant CR, Fu, Y. (2020). Modified Red Blue Vegetation Index for Chlorophyll Estimation and Yield

Prediction of Maize from Visible Images Captured by UAV. *Sensors*, 20, 5055; doi:10.3390/s20185055.

Hariharan R (2021). Random forest regression analysis on combined role of meteorological indicators in disease dissemination in an Indian city: A case study of New Delhi. *Urban Climate*, 36, 100780. <https://doi.org/10.1016/j.uclim.2021.100780>.

Kannangara M, Dua R, Ahmadi L, Bensebaa F (2018) Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Manag*, 74,3 -15.

Maldaner LF, Canata TF, Molin, JP (2022a). An Approach to Sugarcane Yield Estimation Using Sensors in the Harvester and ZigBee Technology. *Sugar Tech*, 24, 813–821. <https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s12355-021-01050-x>

Maldaner LF, Corrêdo LP, Canata TF, Molin JP (2021b) Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. *Computers and Electronics in Agriculture* 181, 105945. <https://doi.org/10.1016/j.compag.2020.105945>

Maldaner LF, Molin JP, Spekken M (2021b) Methodology to filter out outliers in high spatial density data to improve maps reliability. *Scientia Agricola*, 79, e20200178. <http://doi.org/10.1590/1678-992X-2020-0178>.

Molijn RA, Iannini L, Rocha JV, Hanssen R.F (2019) Sugarcane productivity mapping through C-band and L-band SAR and optical satellite imagery. *Remote Sensing*, 11, 1109. <https://doi.org/10.3390/rs11091109>

Nava SC, Huerta AJ, Cruz JMPDL, Rodríguez, EM. (2016) Leaf Growth and Canopy Development of Three Sugarcane Genotypes under High Temperature Rainfed Conditions in Northeastern Mexico. *International Journal of Agronomy*, 2561026, 7 <http://dx.doi.org/10.1155/2016/2561026>.

Nihar A, Patel NR, Danodia A (2022) Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, *India Journal of the Indian Society of Remote Sensing*, 50, 1519–1530. <https://doi.org/10.1007/s12524-022-01549-0>

Oliveira RP, Barbosa Júnior MR, Pinto AA, Oliveira JLP, Zerbato C, Furlani CEA. (2022) Predicting Sugarcane Biometric Parameters by UAV Multispectral Images and Machine Learning. *Agronomy* 12, 1992. <https://doi.org/10.3390/agronomy12091992>.

Pinto AA, Zerbato C, Rolim G S, Barbosa Júnior MR Silva, LFV, Oliveira RP (2022). Corn grain yield forecasting by satellite remote sensing and machine learning models. *Agronomy journal*, 1-20. <https://doi.org/10.1002/ajq2.21141>.

Pimentel Gomes F (2009) **Curso de estatística experimental**. 15. ed. FEALQ, Piracicaba.

Ramos APM, Osco LP, Furuya DEG, Gonçalves WN, Santana DC, Teodoro LPR Silva Junior CA, Capristo-Silva GF, Li J, Baio FHR, Marcato Junior J, Teodoro PE, Pistori H (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. **Computers and Electronics in Agriculture** 178, 105791. <https://doi.org/10.1016/j.compag.2020.105791>.

Som-ard J, Atzberger C, Izquierdo-Verdiguier E, Vuolo F, Immitzer M, (2021) Remote Sensing Applications in Sugarcane Cultivation: A Review. **Remote Sensing**, 13, 4040. <https://doi.org/10.3390/rs13204040>.

Yang W, Zhao Y, Wang D, Wu H, Lin A, He L (2020) Using Principal Components Analysis and IDW Interpolation to Determine Spatial and Temporal Changes of Surface Water Quality of Xin'anjiang River in Huangshan, China. **International Journal of Environmental Research and Public Health**, 17, 2942. <https://doi.org/10.3390/ijerph17082942>.

Zhao D, Gordon VS, Comstock JC, Glynn NC, Johnson RM (2016). Assessment of Sugarcane Yield Potential across Large Numbers of Genotypes using Canopy Reflectance Measurements. **Crop Science**, 56, 1747–1759. <https://doi.org/10.2135/cropsci2015.12.0747>

Zhou X, Kono Y, Win A, Tsutomu Matsui T, Tanaka TST (2021) Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. **Plant Production Science**, 24, 137-151. <https://doi.org/10.1080/1343943X.2020.1819165>

Zhu L, Liu X, Wang Z, Tian L (2023) High-precision sugarcane yield prediction by integrating 10-m Sentinel-1 VOD and Sentinel-2 GRVI indexes. **European Journal of Agronomy**, 149, 126889. <https://doi.org/10.1016/j.eja.2023.126889>

Anexos

Anexo I. Data das imagens e respectivo soma de graus-dia em cada imagem das safras 2017/18 e 2018/19.

Soma de graus- dia (SDD)	Símbolo	Datas da imagens da safra 2017/18	SDD em cada imagem da safra 2017/18	Datas da imagens da safra 2018/19	SDD em cada imagem da safra 2018/19
0 - 800	WBH10	04/24	771	02/13	85
				03/25	637
				04/04	776
800 - 1200	WBH4	04/29	844	04/19	980
		05/09	978		
		05/24	1147		
>1200	WBH1	06/13	1370	06/08	1633
		06/18	1412	06/13	1698
		06/23	1476	-	-
		07/03	1607	-	-
		07/08	1675	-	-