

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

**PREVISÃO DE PARÂMETROS BIOMÉTRICOS DE CANA-
DE-AÇÚCAR POR IMAGENS DE VEÍCULO AÉREO NÃO
TRIPULADO**

Romário Porto de Oliveira

Engenheiro Agrônomo

2022

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

**PREVISÃO DE PARÂMETROS BIOMÉTRICOS DE CANA-
DE-AÇÚCAR POR IMAGENS DE VEÍCULO AÉREO NÃO
TRIPULADO**

Discente: Romário Porto de Oliveira

Orientador: Prof. Dr. Carlos Eduardo Angeli Furlani

Coorientador: Dr. Cristiano Zerbato

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

O48p	Oliveira, Romário Porto de Previsão de Parâmetros Biométricos de Cana-de-açúcar por Imagens de Veículo Aéreo Não Tripulado / Romário Porto de Oliveira. -- Jaboticabal, 2022 52 p. : il., tabs. Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal Orientador: Carlos Eduardo Angeli Furlani Coorientador: Cristiano Zerbato 1. Agricultura digital. 2. Altura de planta. 3. Bandas espectrais. 4. Diâmetro de colmo. 5. Número de perfilhos. I. Título.
------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor.

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: PREVISÃO DE PARÂMETROS BIOMÉTRICOS DE CANA-DE-AÇÚCAR
POR IMAGENS DE VEÍCULO AÉREO NÃO TRIPULADO

AUTOR: ROMÁRIO PÔRTO DE OLIVEIRA

ORIENTADOR: CARLOS EDUARDO ANGELI FURLANI

COORIENTADOR: CRISTIANO ZERBATO

Aprovado como parte das exigências para obtenção do Título de Mestre em AGRONOMIA
(PRODUÇÃO VEGETAL), pela Comissão Examinadora:



Prof. Dr. CARLOS EDUARDO ANGELI FURLANI (Participação Virtual)
Departamento de Engenharia e Ciências Exatas DECEX / FCAV UNESP Jaboticabal



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

Mailson Freire de Oliveira
Pós-doutorando MAILSON FREIRE DE OLIVEIRA (Participação Virtual)
Auburn University-College of Agriculture / Estados Unidos

Jaboticabal, 28 de julho de 2022

DADOS CURRICULARES DO AUTOR

ROMÁRIO PORTO DE OLIVEIRA, nascido em Curimatá – PI em 30 de março de 1994, filho de Maria da Glória dos Santos e Gesivaldo Carvalho de Oliveira. Coursou o ensino fundamental na Unidade Escolar Florentino Camelo e ensino médio na Unidade Escolar Benedito Oliveira (2009 – 2012), localizadas no município de Júlio Borges – PI. Ingressou na graduação em 2014, no curso de bacharelado em Engenharia Agrônômica na Universidade Federal do Piauí, Câmpus Professora Cinobelina Elvas. Durante a graduação foi bolsista de Iniciação Científica (CNPq) durante dois anos, trabalhando em projeto na área de caracterização e estimativa dos atributos do solo, sob orientação do Prof. Dr. Ronny Sobreira Barbosa. Participou também do grupo de pesquisa SOLOCER: Solos, Produção Vegetal e Qualidade Ambiental na Região dos Cerrados. O título de Engenheiro Agrônomo foi conquistado no ano de 2019. Ingressou no Mestrado pelo Programa de Pós-Graduação em Agronomia (Produção Vegetal) no ano de 2020, com linha de pesquisa em Agricultura Digital, na Universidade Estadual Paulista “Júlio de Mesquita Filho” Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal – UNESP/FCAV. Integrante do Laboratório de Máquinas e Mecanização Agrícola (LAMMA), localizado no Departamento de Engenharia e Ciências Exatas, desenvolveu trabalhos de pesquisa nas áreas de sensoriamento remoto e agricultura de precisão, com ênfase na utilização de veículos aéreos não tripulados na obtenção de variáveis agrícolas.

À minha família, minha base, meu
alicerce.

DEDICO!

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, que iluminou e abençoou o meu caminho durante este percurso.

A minha Mãe, Maria da Glória dos Santos, pelo amor e carinho, o qual me deu forças para seguir em frente, independente dos obstáculos encontrados.

A minha Avó, Rosa Porto Camêlo, pelo seu carinho, apoio e todos ensinamentos e conselhos concedidos durante essa longa caminhada.

Aos meus queridos Irmãos, Maiara Barbosa dos Santos, Cleiton Guimarães dos Santos e Paulo Roberto Barbosa dos Santos, por sempre estarem ao meu lado em todos os momentos da minha vida.

A minha Madrinha, Inaísa Porto dos Santos, tios e primos, Silvio Porto dos Santos, Idaíza Francisca dos Santos, Maria Gloriete dos Santos Pagadigorria, Manoel Porto dos Santos, José Antônio Borges, Rodrigo Rocha Porto, Bruna Santos Pagadigorria, Andréia Rocha Porto e Emerson Gonsalves, pelo carinho, apoio e força concedido.

A minha família Paterna, agradeço pelo apoio concedido durante esse trajeto.

A Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal, em especial ao Programa de Pós-Graduação em Agronomia (Produção Vegetal).

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

Ao meu Orientador Prof. Dr. Carlos Eduardo Angeli Furlani, pelo apoio, confiança e ensinamentos compartilhados.

Aos meu Coorientador Dr. Cristiano Zerbato, pela oportunidade concedida, confiança, conhecimento e experiências trocadas.

Ao Prof. Dr. Rouverson e demais docentes e funcionários do Departamento de Engenharia e Ciências Exatas, deixo meus agradecimentos.

Aos meus colegas de trabalho e amigos do Laboratório de Máquinas e Mecanização Agrícola – LAMMA. Pelos momentos de felicidades e aflições compartilhadas.

Por fim, agradeço a todos que de alguma forma, contribuíram para que fosse possível a concretização dessa etapa em minha formação.

SUMÁRIO

RESUMO.....	i
ABSTRACT.....	ii
LISTA DE TABELAS	iii
LISTA DE FIGURAS.....	iv
CAPÍTULO 1 – Considerações gerais.....	1
1. Introdução	1
1.1. Revisão de literatura	2
1.1.1. Cultura da cana-de-açúcar	2
1.1.2. Sensoriamento Remoto.....	4
1.1.3. Algoritmos de Machine Learning	7
1.2. Referências	10
CAPÍTULO 2 - Previsão de Parâmetros Biométricos de Cana-de-açúcar por Imagens Multiespectrais de VANT e Machine Learning	19
2.1. Introdução	19
2.2. Material e Métodos	21
2.2.1. Área de estudo e dados biométricos	21
2.2.2. Coleta de dados baseada em VANT	22
2.2.3. Modelos de previsão.....	23
2.2.3.1. Filtragem de dados.....	23
2.2.3.2. Análise de dados	24
2.3. Resultados	25
2.3.1. Seleção das variáveis preditoras	25
2.3.2. Predições dos parâmetros biométricos.....	26
2.4. Discussão.....	31
2.5. Conclusões	34
2.6. Referências	34
CAPÍTULO 3 – Considerações finais	40

PREVISÃO DE PARÂMETROS BIOMÉTRICOS DE CANA-DE-AÇÚCAR POR IMAGENS DE VEÍCULO AÉREO NÃO TRIPULADO

RESUMO – Sensores multiespectrais a bordo de veículos aéreos não tripulados (VANT's) são precisos e rápidos para prever produtividade de cana-de-açúcar. Neste estudo propõe-se prever parâmetros biométricos de cana-de-açúcar usando algoritmos de *machine learning* e dados multitemporais, por meio de análise de imagens multiespectral de VANT. A pesquisa foi realizada em cinco cultivares de cana-de-açúcar, como forma de construir modelos mais generalistas. Imagens multiespectrais foram coletadas a cada 40 dias e os parâmetros biométricos avaliados foram: número de perfilhos, altura da planta e diâmetro do colmo. Dois algoritmos de *machine learning* foram utilizados: modelo de *multiple linear regression* e modelo de *random forest*. Os resultados mostraram que os modelos de previsão foram acurados. As bandas espectrais Blue, Green e NIR forneceram melhor desempenho em prever os parâmetros biométricos da cana-de-açúcar. Resultados para melhorar a previsão de produtividade de cana-de-açúcar são fornecidos até sete meses antes da colheita. Estas descobertas expandem as possibilidades da utilização de imagem multiespectral de VANT na previsão de produtividade de cana-de-açúcar, particularmente pela inclusão de parâmetros biofísicos.

Palavras-chave: Agricultura digital, altura de planta, bandas espectrais, diâmetro de colmo, número de perfilhos

PREDICTING SUGARCANE BIOMETRIC PARAMETERS BY UNMANNED AERIAL VEHICLE IMAGES

ABSTRACT – Multispectral sensors on board unmanned aerial vehicles (UAVs) are accurate and fast to predict sugarcane yields. In this study we propose to predict sugarcane biometric parameters using machine learning algorithms and multitemporal data, through multispectral UAV image analysis. The research was conducted on five sugarcane cultivars, as a way to build more generalistic models. Multispectral images were collected every 40 days and the biometric parameters evaluated were: number of tillers, plant height, and cane diameter. Two machine learning algorithms were used: multiple linear regression model and random forest model. The results showed that the prediction models were accurate. The Blue, Green and NIR spectral bands provided better performance in predicting the biometric parameters of sugarcane. Results for improving sugarcane yield prediction are provided up to seven months before harvest. These findings expand the possibilities of using multispectral UAV imagery in sugarcane yield prediction, particularly by including biophysical parameters.

Keywords: Digital agriculture, plant height, spectral bands, stalk diameter, number of tillers

LISTA DE TABELAS

Tabela	página
1. Comprimento de onda para as faixas espectrais presentes no sensor MicaSense RedEdge-M.	23
2. Índices de vegetação utilizados no modelo de seleção de variáveis por stepwise. Estes índices são úteis para analisar as características biofísicas.	23
3. Variáveis selecionadas por stepwise.....	25
4. Resultados dos algoritmos de RF e MLR para previsão do NP.	27
5. Resultados dos algoritmos de RF e MLR para previsão de AP.....	28
6. Resultados dos algoritmos de RF e MLR para previsão de DC.	30

LISTA DE FIGURAS

Figura	página
1. Exemplo ilustrativo para a coleta de dados (voos, graus-dia acumulados e dados biométricos) ao longo dos dias após o transplântio.....	22
2. Fluxo de trabalho do processo de aquisição, processamento e modelagem dos dados.	25
3. Gráfico de desempenho da seleção das variáveis. (a) número de perfilhos – NP; (b) altura de planta – AP; (c); diâmetro de colmo - DC.	26
4. Gráficos de dispersão dos valores de NP observado versus predito usando três variáveis pela validação cruzada dos algoritmos de RF (A) e MLR (B). Combinação das imagens 349 GDA e 397 GDA.....	28
5. Gráficos de dispersão dos valores de AP observado versus predito usando três variáveis pela validação cruzada dos algoritmos de RF (A) e MLR (B). Combinação das imagens 349 GDA e 397 GDA.....	29
6. Gráficos de dispersão dos valores de DC observado versus predito usando três variáveis pela validação cruzada dos algoritmos de RF (A) e MLR (B). Combinação das imagens 349, 397 e 410 GDA.	31

CAPÍTULO 1 – Considerações gerais

1. Introdução

A determinação da produtividade da cana-de-açúcar é um processo que exige coleta de informações de desenvolvimento da cultura. Alguns parâmetros agrônômicos são bastante utilizados, como número de colmos, altura da planta e diâmetro do colmo. Contudo, o método mais utilizado pelos agricultores para quantificar estes parâmetros é baseado em contagem e medição manual em campo.

Como alternativa, a tecnologia de sensoriamento remoto possibilita obter informações das plantas de maneira eficiente e segura. A utilização de VANTS, por exemplo, é uma tecnologia que está revolucionando o sensoriamento remoto, possibilitando monitorar grandes áreas agrícolas, gerando muitos dados úteis à tomada de decisão pelos produtores.

Dentro desse contexto, com os avanços obtidos em áreas de processamento de dados, novas técnicas analíticas vêm sendo aplicadas, por terem capacidade de realizar complexas análises, como por exemplo, técnicas de *Machine Learning* (ML). Estas técnicas apresentam inúmeras vantagens, podendo superar limitações encontradas em métodos estatísticos tradicionais.

Sendo assim, esta pesquisa apresenta abordagem para previsão de parâmetros biométricos da cana-de-açúcar utilizando imagens aéreas obtidas por VANT e técnicas de ML. Esses conhecimentos contribuem para elaboração de estratégia rápida, prática e não destrutiva de previsão de produtividade de cana-de-açúcar.

No primeiro capítulo é apresentado o referencial teórico que aprofunda o conhecimento sobre o tema, proporcionando embasamento dos resultados obtidos nos capítulos posteriores. No segundo capítulo é apresentado o artigo científico sobre aplicabilidade de dados obtidos por sensor multiespectral embarcado em VANT e algoritmos de ML, para previsão de perfilhamento, altura de planta e diâmetro de colmo. No terceiro capítulo são descritas as considerações gerais do trabalho.

1.1. Revisão de literatura

1.1.1. Cultura da cana-de-açúcar

A cultura da cana-de-açúcar (*Saccharum* spp.) é uma gramínea semi-perene que pertence ao gênero *Saccharum* L., oriunda de climas tropicais do sul e sudeste da Ásia (Hu et al., 2019) e atualmente é cultivada em mais de 100 países (FAOSTAT, 2021). No Brasil, os canaviais começaram ser implantados no litoral sudeste, na região conhecida como Ilha de São Vicente, a partir de 1568, a produção começou entrar em crise, devido limitações para expansão dos canaviais, perdendo espaço para a região nordeste, onde foram os maiores produtores de açúcar entre os anos de 1580 e 1950 (Figueiredo, 2008; Rodrigues e Ross, 2020).

Na região paulista, a partir de 1750, o açúcar começou a ser o produto mais relevante para a economia, o que levou ao crescimento da população e da economia local (Rodrigues e Ross, 2020). Atualmente, o Estado de São Paulo é considerado maior produtor nacional de cana-de-açúcar, sua expansão é atribuída principalmente ao forte desenvolvimento de pesquisas na região; terras agricultáveis com boas condições de solo, relevo, clima; e devido elevada procura de bioetanol, como fonte de bioenergia sustentável, por causa do interesse em reduzir a dependência pelos combustíveis fósseis (Garofalo et al., 2020). A expansão também ocorreu em outros estados, como Goiás, Minas Gerais, Mato Grosso do Sul e Paraná (CONAB, 2021).

A cana-de-açúcar é um dos produtos que mais agrega com o valor do Produto Interno Bruto (PIB) do agronegócio, sendo que o agronegócio contribui em torno de 26% no PIB brasileiro, influenciando diretamente na economia nacional (Cruz et al., 2022). Para isso, a produção de cana-de-açúcar colhida na safra de 2020/21, obteve valores de 654,5 milhões de toneladas (CONAB, 2021), representando em torno de 35% de toda produção mundial (FAOSTAT, 2021).

O açúcar é considerado valioso subproduto da cana-de-açúcar. O Brasil é o maior produtor e exportador mundial, com volume de exportação em torno de 32,15 milhões de toneladas em 2020/2021 (CONAB, 2021). A produção mundial em 2019/2020 foi estimada em 179 milhões de toneladas de açúcar. Os dois maiores produtores, Brasil e Índia, corresponderam no período por mais de 35% da produção mundial de açúcar (FAOSTAT, 2021).

Outro subproduto de elevada relevância é o etanol, e assim como a produção de açúcar, o Brasil situa-se como segundo maior produtor, ficando atrás somente dos Estados Unidos. A demanda pelo etanol em nível mundial tem perspectiva de aumentar significativamente e o Brasil é representado como importante contribuidor na produção (Van der Hilst et al., 2018). A energia elétrica é outro potencial produto derivada do processamento de subprodutos como do bagaço da cana-de-açúcar, que é um dos resíduos mais abundante resultante do processamento da cana-de-açúcar. A produção de eletricidade é obtida com a queima do bagaço por meio do processo de cogeração, permitindo que usinas de cana-de-açúcar, além de produzirem sua própria energia, possam comercializar o excesso para a rede externa (Hiloidhari et al., 2021).

Tomados em conjunto, esses subprodutos da cana-de-açúcar são de suma importância para economia brasileira. Boas colocações do Brasil no ranking de produção mundial da cana-de-açúcar estão diretamente relacionadas com elevados índices de produção e produtividade da cultura, os quais tendem aumentar a cada ano, principalmente devido ao incremento de tecnologias no processo de produção.

A área plantada de cana-de-açúcar no Brasil entre os anos de 1977 e 2021, passou de 2,3 milhões de hectares para 8,6 milhões de hectares, a qual corresponde produção em torno de 654,5 milhões de toneladas de cana-de-açúcar. A produtividade apresentou aumento de 43%, passando de 53 t ha⁻¹ em 1977 para 75.9 toneladas/ha em 2021 (CONAB, 2021).

A produtividade da cana-de-açúcar é condicionada por diversas características biofísicas e bioquímicas das plantas. Uma dessas características é o elevado potencial de perfilhamento. O perfilhamento é caracterizado como a ramificação das plantas gramináceas. Um perfilho é uma unidade funcional e estrutural de uma planta de cana-de-açúcar (Shrivastava et al., 2015). A taxa de perfilhamento é fundamental para produção de cana-de-açúcar, pois determina o número de canas usináveis, componente principal da produtividade (Zhao et al., 2017). Outros componentes que também condicionam a produtividade são o tamanho e diâmetro (Zhao et al., 2019). Segundo Shrivastava et al. (2015), a produtividade da cana-de-açúcar é caracterizada pela contribuição de aproximadamente 70% pelo número de canas usináveis, 27% pelo comprimento da cana e 3% pelo diâmetro da cana.

Os métodos tradicionais de estimativa de produtividade de cana-de-açúcar podem diferir de acordo a região, mas geralmente utilizam dados de números de colmos, diâmetro, altura e peso dos colmos (Marafon, 2012; Somard et al., 2018; Sumesh et al., 2021). O monitoramento convencional desses parâmetros é realizado de forma manual nos campos de produção, gerando muito trabalho e tempo para obtenção dos dados. Em contrapartida, várias tecnologias vêm sendo utilizadas na obtenção de dados das culturas agrícolas, como por exemplo, o sensoriamento remoto (SR). Técnicas de SR são consideradas fonte de informação precisa que pode ser utilizada para obtenção de vários tipos de informações da vegetação, tornando-se ferramenta de grande relevância para a agricultura moderna.

1.1.2. Sensoriamento Remoto

O sensoriamento remoto (SR) é definido como a obtenção de informações de um objeto ou fenômeno sem que ocorra contato com o mesmo e se baseia na interação da radiação eletromagnética com o alvo (Weiss et al., 2020). Os primeiros sensores de SR foram desenvolvidos para fins militares e foi utilizado com sucesso durante a primeira guerra mundial (Kaplan et al., 2022). As aplicações na agricultura iniciaram em 1972, com o primeiro lançamento do satélite Landsat Multispectral Scanner System, que foi usado para classificar áreas agrícolas na região dos Estados Unidos (Khanal et al., 2020).

Em geral, os sistemas de SR na agricultura podem ser classificados de acordo a plataforma e tipo de sensor. As plataformas podem ser classificadas em orbital, aérea e terrestre. A plataforma orbital inclui sensores embarcados em satélites, a aérea pode ser definida pelo uso de aeronaves tripuladas ou remotamente pilotadas. O sistema terrestre é chamado de SR proximal devido à proximidade de sua localização com o alvo, nesse tipo de categoria, os sensores podem ser portáteis e embarcados em maquinário agrícola (Sishodia et al., 2020). Quanto ao sensor, podem ser classificados em ativos ou passivos. Sensores ativos possuem fonte de luz própria, em contrapartida, sensores passivos não possuem fonte de luz própria, necessitando de radiação externa, como por exemplo, a radiação solar (Jasim et al., 2020).

Os sensores também diferem em relação a resolução espectral, espacial e temporal. A resolução espectral dos sensores é definida pela quantidade de bandas espectrais capturada em determinada faixa do espectro eletromagnético (Colomina e Molina, 2014). A resolução espacial está relacionada ao tamanho do pixel das imagens obtidas dos alvos (Sishodia et al., 2020). A resolução temporal é relacionada ao tempo levado para que as informações de um alvo possam ser coletadas, por exemplo, a resolução espacial do satélite é definida pelo tempo utilizado para revisitar uma mesma área de coleta de dados, ou seja, a realização de uma órbita completa (Yi et al., 2019).

O SR por satélites possibilita monitoramento de grandes áreas e é bastante aplicado em diversas áreas de pesquisa. No entanto, essa plataforma é suscetível a interferências ambientais como, por exemplo, nebulosidade. O menor tempo de revisita também é um fator limitante, pois o período de visita de determinado local é fixo (Hall et al., 2018). Outra questão é a baixa resolução espacial, que dependendo do nível de detalhamento exigido do alvo, as aplicações desses dados podem ser limitadas (Berni et al., 2009).

Com o rápido desenvolvimento tecnológico na área agrícola, os VANT's têm sido amplamente utilizados em diversas áreas (Tožička e Komenda, 2016). O SR com VANT's se originou principalmente devido ao avanço da robótica e miniaturização de sensores (Colomina e Molina, 2014). Com a ampla aplicação dessa plataforma aliada ao desenvolvimento de várias tecnologias de processamentos de dados, o SR apresentou progresso significativo. O uso de VANT's possibilitou obtenção de dados com elevada qualidade e baixo custo, apresentando alta aplicabilidade para monitoramento de campos agrícolas (Perz e Wronowski, 2019). Sensores embarcados nessas plataformas geralmente capturam imagens com resolução espacial de centímetros, além de oferecer alta flexibilidade de coleta de dados e pouca influência ambiental (Jiménez-Brenes et al., 2019).

Geralmente, dois tipos de VANT's têm sido usados para no monitoramento agrícola, asa fixa e multirrotor (Delavarpour et al., 2021; García et al., 2020). VANT de asa fixa apresentam maior autonomia de voo em relação aos multirrotor e maior capacidade de cobertura de área (Zhang et al., 2014). Por outro lado, os multirrotor apresentam estruturas de voo mais robustas, proporcionando capacidade de voar em

qualquer local no campo sem exigir requisitos de decolagem e pouso, além de permitir voo mais robusto e preciso contra perturbação do vento (Delavarpour et al., 2021).

A maioria dos sensores embarcados em VANT na área agrícola capturam informações nas faixas espectrais do Visível (400-700 nm), infravermelho próximo (700-1300 nm) e infravermelho de ondas curtas (1300-2500 nm). Com base nas bandas espectrais obtidas, são classificados em RGB, multiespectral e hiperespectral (Khan et al., 2022; Pádua et al., 2022; Zheng et al., 2020). Os sensores RGB coletam informações apenas do espectro do visível, correspondente as bandas Red (vermelho), Green (verde) e Blue (azul). Os sensores multiespectrais coletam dados de pelo menos um filtro além da faixa espectral do visível, geralmente no espectro do infravermelho próximo (NIR) (Xie e Yang, 2020). Além disso, estão sendo cada vez mais equipados com cinco bandas, incluindo a região da borda do vermelho (RedEdge) (Zheng et al., 2020). Em relação aos sensores hiperespectrais, sua principal diferença dos sensores descritos anteriormente está relacionada a capacidade de capturar separadamente centenas de bandas estreitas (Khan et al., 2022). O custo de aquisição desses sensores varia de acordo a quantidade de bandas obtidas, sendo os sensores RGB de menor custo, multiespectrais de valor intermediário e hiperespectrais de maior custo (Xie e Yang, 2020).

A reflectância da radiação solar da vegetação depende das características morfológicas e químicas das plantas (Sishodia et al., 2020). De modo geral, para extração das informações da vegetação, imagens obtidas por esses sensores são utilizadas para calcular índices de vegetação (IVs) (Zhang et al., 2019). Os IVs são expressões matemáticas que utiliza combinação de diferentes bandas do sensor, dependendo das características de absorção e reflexão da radiação, para produzir um valor que possibilita avaliar várias propriedades da vegetação (Rhyma et al., 2020).

Alguns IVs são comumente utilizados para obtenção de informações das culturas, como o índice de vegetação por diferença normalizada (NDVI) (Rouse et al., 1974), índice de vegetação ajustado ao solo (SAVI) (Huete, 1988), índice de RedEdge por diferença normalizada (NDRE) (Barnes et al., 2000), índice de vegetação por diferença normalizada do verde (GNDVI) (Gitelson et al., 1995). O princípio por trás da relação entre esses IVs e variáveis das plantas é que a absorção da radiação pelas plantas saudáveis na região do visível é elevada, principalmente das bandas azul e

vermelho, enquanto que na banda verde ocorre um pico de reflectância (Yang et al., 2017). Essa elevada absorção das bandas azul e vermelho é fortemente influenciada pelos pigmentos fotossintéticos presentes na folha, especialmente clorofila e carotenoides (Rabatel et al., 2014). Por outro lado, na faixa espectral NIR, ocorre elevada reflectância nas plantas saudáveis, principalmente devido influências das estruturas internas das folhas (Ollinger, 2011).

A utilização de bandas espectrais e IVs para obtenção de características agronômicas pode ser feita por abordagens empíricas, que referem-se a algoritmos que buscam relacionar entradas e saídas por meio de métodos estatísticos, como por exemplo, as técnicas de regressão, que são utilizadas para encontrar relações entre dados espectrais e variáveis biofísicas (Weiss et al., 2020). Técnicas avançadas incluem algoritmos complexos de *machine learning* (Guo et al., 2021). Em comparação com métodos estatísticos tradicionais, essas técnicas possibilitam caracterizar relações complexas entre variáveis independentes e dependentes, fornecendo maior precisão e desempenho mais robusto (Kayad et al., 2019; Mahmood et al., 2022), sendo de grande interesse para aplicações agrícolas.

1.1.3. Algoritmos de *Machine Learning*

O termo *machine learning* pode ser definido como a capacidade de aprendizagem dos computadores com base em dados. É uma técnica em que o conhecimento da máquina é aprimorado ou expandido a partir de uma pesquisa, sem necessidade de programação explícita (Mohinur Rahaman e Azharuddin, 2022). Consiste em algoritmos que realizam complexa análise de dados para encontrar padrões e criar regras de decisão, fornecendo para um software a capacidade de obter resultados. O objetivo dos métodos de *machine learning* é desenvolver modelos com fórmulas estatísticas para estimar ou prever saídas, por meio da entrada de novos dados no sistema (Arumugam et al., 2022).

As técnicas de *machine learning* podem ser divididas principalmente em duas categorias, aprendizado supervisionado e não supervisionado. No aprendizado supervisionado, o resultado final dos dados é conhecido, ou seja, para cada amostra de entrada, é fornecido ao algoritmo o resultado final, conhecido também como rótulo

(Ayoub Shaikh et al., 2022). Para o aprendizado não supervisionado, apenas os dados de entrada são fornecidos aos algoritmos, sem rótulos de saída. Nesse processo, os algoritmos buscam encontrar padrões ocultos existentes nos dados (agrupamentos) (Mahmood et al., 2022).

Os algoritmos de aprendizado supervisionado são classificados tipicamente em regressão e classificação. Algoritmos de regressão buscam gerar previsões de valores contínuos, enquanto classificação gera resultados categóricos (Borém et al., 2022). Vários desses algoritmos são utilizados para análise de dados na agricultura, como *Multiple Linear Regression* (MLR), *Decision Tree* (DT), *Random Forest* (RF), Redes Neurais Artificiais (RNA), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (K-NN), etc (Akbarian et al., 2022; Jiang et al., 2020; Mariano e Mónica, 2021; Parra et al., 2022; Qiu et al., 2021).

A MLR é uma técnica de *machine learning* popularmente utilizada em diversas áreas da ciência. São usadas para modelar a relação linear entre variáveis independentes (explicativas) e variáveis dependentes (resposta) (Meerasri e Sothornvit, 2022). Geralmente, duas ou mais variáveis preditoras são utilizadas para estimar ou prever a variável de resposta (Dhulipala e Patil, 2020; Jain et al., 2022). É regularmente utilizada como modelo empírico ou função de aproximação que possibilita estabelecer um modelo matemático para descrever um fenômeno do mundo real (Lee et al., 2019). É um algoritmo bastante aplicado na agricultura, devido apresentar estrutura simples, com fácil cálculo e interpretação (Korkmaz, 2021).

Decision Tree é um algoritmo que soluciona problemas de classificação e regressão. Apresenta estrutura semelhante a árvore e consiste em um conjunto regras que são usadas para vincular características a uma zona de interesse específico (Shirahatti e Patil, 2018). Possui também dois tipos de nós, nó de decisão e nó folha. O nó de decisão realiza opções de escolha, enquanto que o nó folha encontra o resultado da árvore (Mohinur Rahaman e Azharuddin, 2022). Esse algoritmo apresenta forma de análise de dados relativamente simples e pode ser usada para vários tipos de dados (Ayoub Shaikh et al., 2022). O algoritmo de RF apresenta o princípio de funcionamento semelhante ao DT, a principal diferença está relacionada ao número de árvores que são combinadas. RF é construído por várias árvores de decisão individuais e o resultado é obtido pela votação ou média de todas as árvores

(Jagtap et al., 2021). RF tem vantagens de ser um processo de fácil modelagem, rápida velocidade de cálculo e custo computacional mínimo (Lin e Liu, 2022). Esse algoritmo pode ser aplicado para diversos problemas de regressão e classificação.

Os algoritmos de RNA são inspirados nos neurônios biológicos humanos. É composta por grande número de neurônios, que realizam o processamento dos dados. Conhecida também como sistema conexionista, cada conexão recebe um peso, com responsabilidade de transmitir o sinal (Jagtap et al., 2021). O aprendizado da rede é obtido com ajustes dos pesos e a força do sinal aumenta ou diminui à medida que o aprendizado progride. Os algoritmos de RNA apresentam elevados desempenho na análise de dados complexos e não lineares (Güner et al., 2022).

Support Vector Machine é um dos algoritmos do método supervisionado que é utilizado para encontrar um vetor de suporte ideal para categorizar os dados. Esse algoritmo realiza o aprendizado supervisionado encontrando um hiperplano que possa classificar os rótulos em suas classes correspondentes. São adequados para análise de conjuntos de dados de alta dimensão (Arumugam et al., 2022; Kok et al., 2021). Em relação ao K-NN, o método de aprendizado é baseado em instâncias e às vezes identificado como aprendizado baseado em memória. A classificação dos dados é realizada de acordo as distâncias entre as amostras de treinamento e teste. A distância euclidiana é uma métrica de distância comumente usada. O K-NN pode ser usado para solucionar problemas de regressão e classificação (Guerrero et al., 2022; Saleem et al., 2019).

Na categoria de aprendizado não supervisionado, alguns exemplos são os algoritmos de agrupamento K-Means, Hierárquico e Fuzzy C-means. A análise de cluster é a técnica de aprendizagem não supervisionada mais predominante (Roter et al., 2022). Nesse método, os registros não rotulados da coleta de dados devem ser previstos. A principal metodologia de agrupamento consiste em classificar os grupos de acordo à similaridade dos dados. Na abordagem básica na formação do cluster, pontos de registro inteiros são agrupados como um único cluster e seu conjunto de dados é dividido à medida que a distância é maximizada (Sakthipriya e Chandrakumar, 2022).

Essas técnicas de *machine learning* têm sido utilizadas na agricultura para realizar diversos tipos de análises, tais como predição de produtividade (Mariano e

Mónica, 2021; Marques Ramos et al., 2020), biomassa (Akbarian et al., 2022), status nutricional de culturas (Jasim et al., 2020; Qiu et al., 2021), identificação de doenças (Javidan et al., 2022; Jiang et al., 2020) e monitoramento de pragas (Manrique-Silupu et al., 2021; Rouabah et al., 2022).

O avanço da agricultura digital, proporcionou acesso à grande quantidade e variedades de dados do sistema agrícola (Kayad et al., 2022). As técnicas de *machine learning* utilizam esses dados para extração de informações úteis, possibilitando tomada de decisão mais assertiva (Borém et al., 2022). O objetivo principal é proporcionar produção agrícola mais eficiente, maximizando os lucros e minimizando as despesas, respeitando o equilíbrio ambiental.

1.2. Referências

Akbarian S, Xu C, Wang W, Ginns S, Lim S (2022) An investigation on the best-fit models for sugarcane biomass estimation by linear mixed-effect modelling on unmanned aerial vehicle-based multispectral images: A case study of Australia. **Information Processing in Agriculture**. doi:10.1016/j.inpa.2022.03.005.

Arumugam K, Swathi Y, Sanchez DT, Mustafa M, Phoemchalard C, Phasinam K, Okoronkwo E (2022) Towards applicability of machine learning techniques in agriculture and energy sector. **Materials Today: Proceedings** 51:2260–2263. doi:10.1016/j.matpr.2021.11.394.

Ayoub Shaikh T, Rasool T, Rasheed Lone F (2022) Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. **Computers and Electronics in Agriculture** 198:107119. doi:10.1016/j.compag.2022.107119.

Barnes EM, Clarke TR, Richards SE et al. (2000) Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In: **Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA** 1619:6.

Berni JAJ, Zarco-Tejada PJ, Suárez L, Fereres E (2009) Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial

vehicle. **IEEE Transactions on Geoscience and Remote Sensing** 47:722–738. doi:10.1109/TGRS.2008.2010457.

Borém A, Queiroz DM, Valente DSM, Pinto FAC (2022) Agricultura Digital. São Paulo: Oficina de Textos.

Colomina I, Molina P (2014) Unmanned aerial systems for photogrammetry and remote sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing** 92:79–97. Elsevier B.V. doi:10.1016/j.isprsjprs.2014.02.013.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO (2021) Acompanhamento da Safra Brasileira de cana-de-açúcar. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar> (acessado 26 novembro 2021).

Cruz JE, Medina G da S, Júnior JR de O (2022) Brazil's Agribusiness Economic Miracle: Exploring Food Supply Chain Transformations for Promoting Win–Win Investments. **Logistics** 6:23. doi:10.3390/logistics6010023.

Delavarpour N, Koparan C, Nowatzki J, Bajwa S, Sun X (2021) A technical study on UAV characteristics for precision agriculture applications and associated practical challenges. **Remote Sensing** 13:1204. doi:10.3390/rs13061204.

Dhulipala S, Patil GR (2020) Freight production of agricultural commodities in India using multiple linear regression and generalized additive modelling. **Transport Policy** 97:245–258. doi:10.1016/j.tranpol.2020.06.012.

FAOSTAT - Food and Agriculture Organization of the United Nation (2021). Disponível em: <http://www.fao.org/faostat/en/?#data/QC> (acessado 27 dezembro 2021).

Figueiredo P (2008) **Breve história da cana-de-açúcar e do papel do Instituto Agrônômico no seu estabelecimento no Brasil**. Cana-de-açúcar. Campinas: Instituto Agrônômico, p. 31-44.

García L, Parra L, Jimenez JM, Lloret J, Mauri P V., Lorenz P (2020) DronAway: A proposal on the use of remote sensing drones as mobile gateway for wsn in precision agriculture. **Applied Sciences (Switzerland)** 10:6668. doi:10.3390/APP10196668.

Garofalo D, Packer A, Ramos N, Kondo V, Matsuura MDS, Cabral O (2020) **Dinâmica**

na cultura da cana-de-açúcar no Brasil: 1990 a 2018. Jaguariúna: Embrapa: CNPMA, 41 p. (Embrapa: CNPMA. Documentos, 124).

Gitelson AA, Kaufman YJ, Merzlyak MN, Blaustein J (1995) Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. **Remote Sens. Environ.** 58:289-298.

Guerrero A, Javadi SH, Mouazen AM (2022) Automatic detection of quality soil spectra in an online vis-NIR soil sensor. **Computers and Electronics in Agriculture** 196:106857. doi:10.1016/j.compag.2022.106857.

Güner ŞT, Diamantopoulou MJ, Poudel KP, Çömez A, Özçelik R (2022) Employing artificial neural network for effective biomass prediction: An alternative approach. **Computers and Electronics in Agriculture** 192:106596. doi:10.1016/j.compag.2021.106596.

Guo Y, Fu Y, Hao F, Zhang X, Wu W, Jin X, Robin Bryant C, Senthilnath J (2021) Integrated phenology and climate in rice yields prediction using machine learning methods. **Ecological Indicators** 120:106935. doi:10.1016/j.ecolind.2020.106935.

Hall O, Dahlin S, Marstorp H, Bustos MFA, Öborn I, Jirstrom M (2018) Classification of maize in complex smallholder farming systems using UAV imagery. **Drones** 2:22. doi:10.3390/drones2030022.

Hiloidhari M, Haran S, Banerjee R, Rao AB (2021) Life cycle energy–carbon–water footprints of sugar, ethanol and electricity from sugarcane. **Bioresource Technology** 330:125012. doi:10.1016/j.biortech.2021.125012.

Hu S, Shi L, Huang K, Zha Y, Hu X, Ye H, Yang Q (2019) Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. **Field Crops Research** 232:49–61. doi:10.1016/j.fcr.2018.12.009.

Huete AR (1988) A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment** 25:295–309. doi:10.1016/0034-4257(88)90106-X.

Jagtap ST, Phasinam K, Kassanuk T, Jha SS, Ghosh T, Thakar CM (2021) Towards application of various machine learning techniques in agriculture. In: **Materials Today: Proceedings** 51:793–797. doi:10.1016/j.matpr.2021.06.236.

Jain S, Rathee S, Kumar A, Sambasivam A, Boadh R, Choudhary T, Kumar P, Kumar Singh P (2022) Prediction of temperature for various pressure levels using ANN and multiple linear regression techniques: A case study. **Materials Today: Proceedings** 56:194–199. doi:10.1016/j.matpr.2022.01.067.

Jasim A, Zaeen A, Sharma LK, Bali SK, Wang C, Buzza A, Alyokhin A (2020) Predicting phosphorus and potato yield using active and passive sensors. **Agriculture (Switzerland)** 10:564. doi:10.3390/agriculture10110564.

Javidan SM, Banakar A, Vakilian KA, Ampatzidis Y (2022) Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. **Smart Agricultural Technology** 3:100081. doi:10.1016/j.atech.2022.100081.

Jiang F, Lu Y, Chen Y, Cai D, Li G (2020) Image recognition of four rice leaf diseases based on deep learning and support vector machine. **Computers and Electronics in Agriculture** 179:105824. doi:10.1016/j.compag.2020.105824.

Jiménez-Brenes FM, López-Granados F, Torres-Sánchez J, Peña JM, Ramírez P, Castillejo-González IL, de Castro AI (2019) Automatic UAV-based detection of *Cynodon dactylon* for site-specific vineyard management. **PLoS ONE** 14:e0218132. doi:10.1371/journal.pone.0218132.

Kaplan G, Rashid T, Gasparovic M, Pietrelli A, Ferrara V (2022) Monitoring war-generated environmental security using remote sensing: A review. **Land Degradation & Development** 33:1513-1526. doi:10.1002/ldr.4249.

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. doi:10.3390/rs11232873.

Kayad A, Sozzi M, Paraforos DS et al. (2022) How many gigabytes per hectare are available in the digital agriculture era? A digitization footprint estimation. **Computers and Electronics in Agriculture** 198:107080. doi:10.1016/j.compag.2022.107080.

Khan A, Vibhute AD, Mali S, Patil CH (2022) A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. **Ecological Informatics** 69:101678. doi:10.1016/j.ecoinf.2022.101678.

Khanal S, Kushal KC, Fulton JP, Shearer S, Ozkan E (2020) Remote sensing in agriculture—accomplishments, limitations, and opportunities. **Remote Sensing** 12:3783. doi:10.3390/rs12223783.

Kok ZH, Mohamed Shariff AR, Alfatni MSM, Khairunniza-Bejo S (2021) Support Vector Machine in Precision Agriculture: A review. **Computers and Electronics in Agriculture** 191:106546. doi:10.1016/j.compag.2021.106546.

Korkmaz M (2021) A study over the general formula of regression sum of squares in multiple linear regression. **Numerical Methods for Partial Differential Equations** 37:406–421. doi:10.1002/num.22533.

Lee Y, Jung C, Kim S (2019) Spatial distribution of soil moisture estimates using a multiple linear regression model and Korean geostationary satellite (COMS) data. **Agricultural Water Management** 213:580–593. doi:10.1016/j.agwat.2018.09.004.

Lin L, Liu X (2022) Mixture-based weight learning improves the random forest method for hyperspectral estimation of soil total nitrogen. **Computers and Electronics in Agriculture** 192:106634. doi:10.1016/j.compag.2021.106634.

Mahmood A, Tiwari AK, Singh SK, Udmale SS (2022) Contemporary machine learning applications in agriculture: Quo Vadis? **Concurrency and Computation: Practice and Experience** 34:e6940. doi:10.1002/cpe.6940.

Manrique-Silupu J, Campos JC, Paiva E, Ipanaqué W (2021) Thrips incidence prediction in organic banana crop with Machine learning. **Heliyon** 7:e08575. doi:10.1016/j.heliyon.2021.e08575.

Marafon AC (2012) **Análise quantitativa de crescimento em cana-de-açúcar: uma introdução ao procedimento prático**. Aracaju: Embrapa: CPATC, 31 p. (Embrapa: CPATC. Documentos, 168).

Mariano C, Mónica B (2021) A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. **Computers and Electronics in Agriculture** 184:106094. doi:10.1016/j.compag.2021.106094.

Marques Ramos AP, Prado Osco L, Elis Garcia Furuya D et al. (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:10.1016/j.compag.2020.105791.

Meerasri J, Sothornvit R (2022) Artificial neural networks (ANNs) and multiple linear regression (MLR) for prediction of moisture content for coated pineapple cubes. **Case Studies in Thermal Engineering** 33:101942. doi:10.1016/j.csite.2022.101942.

Mohinur Rahaman M, Azharuddin M (2022) Wireless sensor networks in agriculture through machine learning: A survey. **Computers and Electronics in Agriculture** 197:106928. doi:10.1016/j.compag.2022.106928.

Ollinger S V. (2011) Sources of variability in canopy reflectance and the convergent properties of plants. **New Phytologist** 189:375-394. doi:10.1111/j.1469-8137.2010.03536.x.

Pádua L, Matese A, Di Gennaro SF, Morais R, Peres E, Sousa JJ (2022) Vineyard classification using OBIA on UAV-based RGB and multispectral data: A case study in different wine regions. **Computers and Electronics in Agriculture** 196:106905. doi:10.1016/j.compag.2022.106905.

Parra JS, de Souza ZM, de Oliveira SRM, Farhate CVV, Marques J, Siqueira D (2022) Phosphorus adsorption prediction through Decision Tree Algorithm under different topographic conditions in sugarcane fields. **Catena** 213:106114. doi:10.1016/j.catena.2022.106114.

Perz R, Wronowski K (2019) UAV application for precision agriculture. **Aircraft Engineering and Aerospace Technology** 91:257–263. doi:10.1108/AEAT-01-2018-0056.

Qiu Z, Ma F, Li Z, Xu X, Ge H, Du C (2021) Estimation of nitrogen nutrition index in rice from UAV RGB images coupled with machine learning algorithms. **Computers and Electronics in Agriculture** 189:106421. doi:10.1016/j.compag.2021.106421.

Rabatel G, Gorretta N, Labbé S (2014) Getting simultaneous red and near-infrared band data from a single digital camera for plant monitoring applications: Theoretical and practical study. **Biosystems Engineering** 117:2–14. doi:10.1016/j.biosystemseng.2013.06.008.

Rhyima PP, Norizah K, Hamdan O, Faridah-Hanum I, Zulfa AW (2020) Integration of normalised different vegetation index and Soil-Adjusted Vegetation Index for mangrove vegetation delineation. **Remote Sensing Applications: Society and Environment** 17:100280. doi:10.1016/j.rsase.2019.100280.

Rodrigues GS de SC, Ross JLS (2020) A trajetória da cana-de-açúcar no Brasil: perspectivas geográfica, histórica e ambiental. EDUFU. 268 p. doi:10.14393/EDUFU/978-65-86084-00-9.

Roter B, Ninkovic N, Dordevic SV (2022) Clustering superconductors using unsupervised machine learning. **Physica C: Superconductivity and its Applications** 598:1354078. doi:10.1016/j.physc.2022.1354078.

Rouabah A, Meiss H, Villerd J, Lasserre-Joulin F, Tosser V, Chabert A, Therond O (2022) Predicting the abundances of aphids and their natural enemies in cereal crops: Machine-learning versus linear models. **Biological Control** 169:104866. doi:10.1016/j.biocontrol.2022.104866.

Rouse J, Haas RH, Schell JA, Deering DW (1974) Monitoring vegetation systems in the Great Plains with ERTS. **NASA special publication** 351:309-317.

Sakthipriya D, Chandrakumar T (2022) A study on agriculture engineering implements using machine learning. **Materials Today: Proceedings**. doi:10.1016/j.matpr.2022.04.604.

Saleem G, Akhtar M, Ahmed N, Qureshi WS (2019) Automated analysis of visual leaf shape features for plant classification. **Computers and Electronics in Agriculture** 157:270–280. doi:10.1016/j.compag.2018.12.038.

Shirahatti J, Patil R (2018) A Survey Paper on Plant Disease Identification Using Machine Learning Approach. **International Conference on Communication and Electronics Systems (ICCES)** 1171–1174. doi:10.1109/CESYS.2018.8723881.

Shrivastava AK, Solomon S, Rai RK, Singh P, Chandra A, Jain R, Shukla SP (2015) Physiological Interventions for Enhancing Sugarcane and Sugar Productivity. **Sugar Tech** 17:215-226. doi:10.1007/s12355-014-0321-6.

Sishodia RP, Ray RL, Singh SK (2020) Applications of remote sensing in precision

agriculture: A review. **Remote Sensing** 12:1–31. doi:10.3390/rs12193136.

Som-ard J, Hossain MD, Ninsawat S, Veerachitt V (2018) Pre-harvest Sugarcane Yield Estimation Using UAV-Based RGB Images and Ground Observation. **Sugar Tech** 20:645–657. doi:10.1007/s12355-018-0601-7.

Sumesh KC, Ninsawat S, Som-ard J (2021) Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. **Computers and Electronics in Agriculture** 180:105903. doi:10.1016/j.compag.2020.105903.

Tožička J, Komenda A (2016) Diverse planning for UAV control and remote sensing. **Sensors (Switzerland)** 16:2199. doi:10.3390/s16122199.

van der Hilst F, Verstegen JA, Woltjer G, Smeets EMW, Faaij APC (2018) Mapping land use changes resulting from biofuel production and the effect of mitigation measures. **GCB Bioenergy** 10:804–824. doi:10.1111/gcbb.12534.

Weiss M, Jacob F, Duveiller G (2020) Remote sensing for agricultural applications: A meta-review. **Remote Sensing of Environment** 236:111402. doi:10.1016/J.RSE.2019.111402.

Xie C, Yang C (2020) A review on plant high-throughput phenotyping traits using UAV-based sensors. **Computers and Electronics in Agriculture** 178:105731. doi:10.1016/j.compag.2020.105731.

Yang G, Liu J, Zhao C et al. (2017) Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. **Frontiers in Plant Science** 8:1111. doi:10.3389/fpls.2017.01111.

Yi C, Zhao YQ, Chan JCW (2019) Spectral Super-Resolution for Multispectral Image Based on Spectral Improvement Strategy and Spatial Preservation Strategy. **IEEE Transactions on Geoscience and Remote Sensing** 57:9010–9024. doi:10.1109/TGRS.2019.2924096.

Zhang C, Walters D, Kovacs JM (2014) Applications of low altitude remote sensing in agriculture upon farmers' requests-A case study in northeastern Ontario, Canada. **PLoS ONE** 9:e112894. doi:10.1371/journal.pone.0112894.

Zhang X, Zhang F, Qi Y, Deng L, Wang X, Yang S (2019) New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV). **International Journal of Applied Earth Observation and Geoinformation** 78:215–226. doi:10.1016/j.jag.2019.01.001.

Zhao D, Irey M, Laborde C, Hu CJ (2017) Identifying physiological and yield-related traits in sugarcane and energy cane. **Agronomy Journal** 109:927–937. doi:10.2134/agronj2016.10.0585.

Zhao D, Irey M, Laborde C, Hu CJ (2019) Physiological and yield characteristics of 18 sugarcane genotypes grown on a sand soil. **Crop Science** 59:2741–2751. doi:10.2135/cropsci2019.02.0107.

Zheng H, Zhou X, He J, Yao X, Cheng T, Zhu Y, Cao W, Tian Y (2020) Early season detection of rice plants using RGB, NIR-G-B and multispectral images from unmanned aerial vehicle (UAV). **Computers and Electronics in Agriculture** 169:105223. doi:10.1016/j.compag.2020.105223.

CAPÍTULO 2 - Previsão de Parâmetros Biométricos de Cana-de-açúcar por Imagens Multiespectrais de VANT e *Machine Learning*¹

2.1. Introdução

A cana-de-açúcar é uma gramínea semi-perene cultivada globalmente, principalmente em países tropicais e subtropicais, representada como produto de grande importância para o comércio agrícola internacional e matéria-prima relevante para agroindústria (FAOSTAT, 2021). Pode produzir massivamente açúcar, biocombustível e bioenergia (Barbosa Júnior et al., 2022). Portanto, a previsão precisa e oportuna da produtividade da cana-de-açúcar é valiosa para segurança alimentar nacional e para o desenvolvimento sustentável da agricultura (Abebe et al., 2022a). A previsão de produtividade de cana-de-açúcar não é tarefa fácil e, embora estudos anteriores tenham encontrado soluções para este enigma, desafios ainda precisam ser enfrentados. Portanto, o desenvolvimento de alternativas é necessário.

A utilização de técnicas de sensoriamento remoto para a previsão da produtividade de culturas, é uma alternativa que possibilita o manejo oportuno das lavouras e melhor tomada de decisão pelos produtores, além de proporcionar rapidez e precisão na obtenção dos dados. Na literatura acadêmica é possível encontrar vários estudos sobre soluções utilizando sensoriamento remoto para estimativa de produtividade de cana-de-açúcar. Por exemplo, uma abordagem por satélite da Abebe et al. (2022b) combinou imagens Landsat 8 e Sentinel 2A para estimar produtividade de cana-de-açúcar de forma melhorada. A abordagem aplicou índices de vegetação (IVs) mais populares para 2^a e 3^a safras de cana-de-açúcar e usou a *Support Vector Regression* (SVR) como método estatístico. Os resultados mostraram ser úteis para estimativa de produtividade da cana-de-açúcar. No entanto, a produtividade da cultura depende do tipo de solo, condições climáticas, genótipo e práticas de manejo. Portanto, o uso de um tipo único de variável não é confiável para avaliar o verdadeiro estado do sistema agroecológico, causando desinformação da estimativa de produtividade da cultura (Wang et al., 2013). Yu et al. (2022) integraram múltiplas variáveis para estimativa de produtividade em cana-de-açúcar a campo. Os pontos

¹ Artigo publicado na revista Agronomy: <https://doi.org/10.3390/agronomy12091992>

mais fortes neste estudo incluíram as variáveis altura da planta, índice de área foliar e umidade do solo para melhorar a estimativa de produtividade da cultura. Entretanto, foram observadas incertezas, principalmente no modelo estatístico para caracterizar o crescimento potencial da altura da planta, o que restringe futuras aplicações. Outra importante colaboração de Tanut et al. (2021), em que foi usado uma nova geração de plataformas de sensoriamento remoto, o VANT. A abordagem propôs alta precisão utilizando métodos de mineração de dados. O modelo desenvolvido foi treinado na fase de colheita da cana-de-açúcar em 2019 e testado com o conjunto de dados da safra de 2020. Os resultados para a previsão de produtividade da cana-de-açúcar foram alcançados com precisão de 98,69%. O modelo dessa abordagem utilizou uma variedade de tipos de dados, tais como dados climáticos, cultivares de cana-de-açúcar, espaçamento de plantio, quantidade de cortes, tipos de solo e dados da taxa de defeitos da cana-de-açúcar no campo, obtidos através da análise de imagens RGB, como acamamento, baixo crescimento e outros possíveis problemas de cultivo. A utilização do VANT em determinadas situações e finalidades, tem algumas vantagens sobre os satélites, pode voar de forma flexível e capturam imagens com elevada resolução espacial (Tanut et al., 2021).

A densidade de copa das plantações de cana-de-açúcar é um desafio para mapear, resultando em maiores dificuldades na estimativa de produtividade. Essas questões muitas vezes resultam em estimativas de produtividade menos precisas e não confiáveis. A taxa de perfilhamento, altura de planta e diâmetro do colmo têm alta influência na produtividade da cana-de-açúcar, por isso são características agrônomicas amplamente utilizadas para estimar esta variável (Rossi Neto et al., 2018; Som-ard et al., 2018). O método mais utilizado para obter a taxa de perfilhamento e diâmetro do colmo é pela contagem e medição em campo para quantificá-los, uma tarefa trabalhosa e demorada. Em consequência disso, novas abordagens são necessárias. Por outro lado, os métodos estatísticos tradicionais assumem correlações lineares entre produtividade observada e prevista, o qual não poderiam refletir com precisão relações não lineares do sistema agrícola. Assim, para superar estes desafios, podem ser adotados algoritmos de *Machine Learning* (ML) (Guo et al., 2021). *Random forest* (RF) é um algoritmo de ML de árvore de decisão com alto potencial em previsões agrônomicas, tais como previsão de produtividade

(Jeong et al., 2016) e estimativa de biomassa (Yue et al., 2019). Outro algoritmo amplamente aplicado é o *Multiple linear regression* (MLR). Trata-se de um modelo de regressão de baixa complexidade e encontrou resultados de previsão confiáveis em estudos anteriores (Abebe et al., 2022b; Picoli et al., 2014; Simões et al., 2009; Xu et al., 2020).

Perante o exposto, buscando resultados na previsão remota da produtividade de cana-de-açúcar, foi analisado o desempenho dos modelos de *Random forest* e *Multiple linear regression* na previsão de parâmetros biométricos, usando imagens multiespectrais de VANT.

2.2. Material e Métodos

2.2.1. Área de estudo e dados biométricos

O estudo foi realizado em campo experimental localizado na Universidade Estadual Paulista, Campus de Jaboticabal, São Paulo, Brasil (21°14'57" S, 48°16'55" W, altitude de ~570 m). O clima local é Cwa, clima subtropical úmido com estação de inverno seca e temperatura média anual de 22 °C (Alvares et al., 2013). A estação chuvosa se concentra entre novembro e março, com média anual de 1.428 mm, o tipo de solo é Latossolo Vermelho Distrófico (Santos et al., 2018).

As cultivares, CTC4, CTC9001, IACSP95-5094, RB867515 e RB966928, foram transplantadas na área experimental. Para construir modelos mais generalista, os dados de todas as cultivares formaram um único conjunto de dados. Para coleta de dados biométricos, 40 parcelas foram analisadas. As parcelas foram representadas por área de 10 x 22 m. Nessas parcelas, seis pontos foram designados para coleta de dados biométricos, cada ponto correspondia a uma linha de cultivo de 10 m, totalizando 240 pontos. Foram analisados os parâmetros biométricos: número de perfilhos por metro (NP), altura média de planta (AP) e diâmetro médio de colmo (DC). NP e AP foram avaliados em 130 dias após o plantio (DAP), período próximo ao final do perfilhamento (Caetano e Casaroli, 2017; Inman-Bamber, 1994). O DC foi avaliado em 220 DAP (Figura 1). O graus-dia acumulado (GDA) correspondente ao período de coleta de dados foi determinado utilizando 20°C como temperatura basal inferior (Barbieri et al., 1979), para estabelecer relações entre as mudanças espectrais no

dossel e os componentes da produção (Figura 1). Os dados de temperatura foram obtidos de uma estação meteorológica automática localizada a aproximadamente 1,3 km da área de estudo. O GDA foi calculado de acordo metodologia proposta por Arnold (1959):

$$GDA = \left(\frac{T_{max} + T_{min}}{2} \right) - T_{base}$$

Em que:

GDA: graus-dia acumulado; Tmax: temperatura máxima do ar (°C); Tmin: temperatura mínima do ar (°C) e Tbase: temperatura de base inferior da cultura (20 °C).

2.2.2. Coleta de dados baseada em VANT

Foram realizados três voos a 80, 120 e 160 DAP (Figura 1). Para obtenção das imagens, utilizou-se câmera multiespectral (MicaSense RedEdge-M, MicaSense Inc., Seattle, Washington, EUA) a bordo de um VANT multirrotor (modelo). O sensor capturou cinco bandas espectrais: Blue, Green, Red, RedEdge e NIR (Tabela 1).

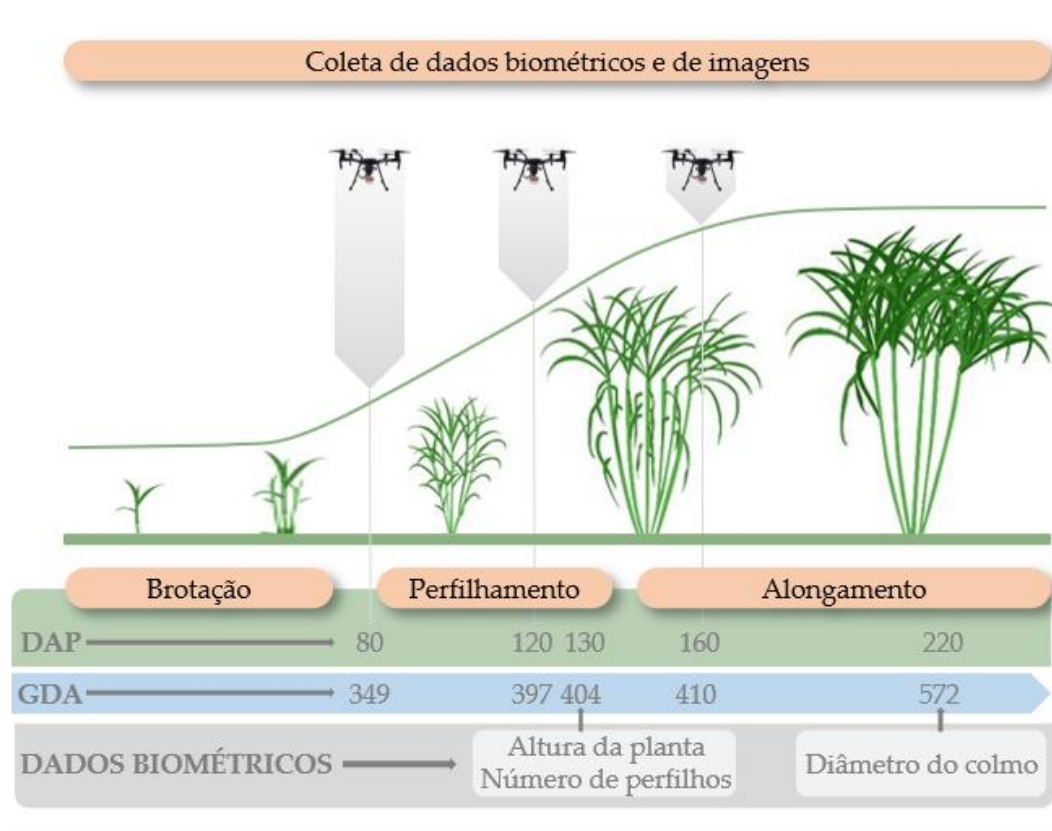


Figura 1. Exemplo ilustrativo para a coleta de dados (voos, graus-dia acumulados e dados biométricos) ao longo dos dias após o transplante.

Tabela 1. Comprimento de onda para as faixas espectrais presentes no sensor MicaSense RedEdge-M.

Banda	Comprimento de onda (nm)	Largura de banda FWHM (nm)
Blue	475	20
Green	560	20
Red	668	10
RedEdge	717	10
NIR	842	40

Os voos foram realizados a 30 m de altura, garantindo assim distância da amostra do solo (GSD) de 2,8 cm. Adotou-se sobreposição frontal de 75% e lateral de 70%. As imagens foram processadas para obtenção do ortomosaico usando o software (Agisoft Metashape Professional 1.5.5, Agisoft, São Petersburgo, Russia). Antes de aplicar a abordagem proposta, os ortomosaicos foram segmentados para remover a interferência do solo. Finalmente, as informações de banda espectrais e IVs (Tabela 2) foram obtidas utilizando as ferramentas de Máscara, dzetsaka: ferramenta de classificação e Estatísticas Zonais do software QGIS 3.10.9 (Free software Inc., Boston, EUA).

Tabela 2. Índices de vegetação utilizados no modelo de seleção de variáveis por stepwise. Estes índices são úteis para analisar as características biofísicas.

Índice de Vegetação	Equação	Referência
Índice de vegetação por diferença normalizada (NDVI)	$\frac{NIR - R}{NIR + R}$	Rouse et al. (1974)
Índice de RedEdge por diferença normalizada (NDRE)	$\frac{NIR - RedEdge}{NIR + RedEdge}$	Barnes et al. (2000)
Índice de vegetação ajustado ao solo (SAVI)	$\left(\frac{NIR - R}{NIR + R + 0.5} \right) (1 + 0.5)$	Huete (1988)

2.2.3. Modelos de previsão

2.2.3.1. Filtragem de dados

Os dados espectrais correspondentes a cada data de coleta de imagens foram avaliados separadamente e juntos para determinar qual deles apresenta o melhor desempenho na previsão dos parâmetros. A variável independente utilizada nos modelos foram valores de reflectância correspondentes às bandas espectrais (Tabela 1) e IVs (Tabela 2). Os parâmetros biométricos da cana-de-açúcar medidos no campo foram as variáveis dependentes. Para detectar e remover outliers do conjunto de dados, usamos o método Z-score (± 3 dp) para medir a tendência central e desvio (Urolagin et al., 2021). A fim de selecionar as melhores variáveis independentes a serem usadas nos modelos de previsão, foi realizada a análise *stepwise*. Este método procura determinar as variáveis que têm maior influência sobre o conjunto de saída, favorecendo maior eficiência do modelo ao reduzir a quantidade de dados de entrada. A métrica utilizada para selecionar os preditores de melhor desempenho por *stepwise* foi a raiz quadrada do erro quadrático médio (RMSE).

2.2.3.2. Análise de dados

Foram utilizados dois modelos preditivos de ML, o modelo MLR e o RF. Os modelos de regressão foram processados usando a biblioteca de código aberto de ML 'scikit-learn' em linguagem de programação Python v3.8.10, no ambiente Jupyter Notebook. Para RF, foi selecionado a melhor combinação de número de árvores (*n_estimators*) e profundidade máxima da árvore (*max_depth*) usando a ferramenta GridSearchCV. Foi utilizado o método de validação cruzada (CV), aplicou-se esse método de validação com $K = 5$. O conjunto de dados foi dividido em cinco subconjuntos (*fold*), em que 80% dos dados foram usados para treinamento e 20% para validação. Em cada repetição, o modelo foi treinado e validado em um conjunto de dados diferente. Como todas as amostras foram avaliadas manualmente em campo, o número total de amostras não foi suficientemente grande, portanto, o teste não foi realizado. Para desempenho dos modelos de MLR e RF foi avaliado precisão pelo coeficiente de determinação (R^2), acurácia pelo erro médio absoluto (MAE) e o RMSE. A figura 2 apresenta visão geral da metodologia.

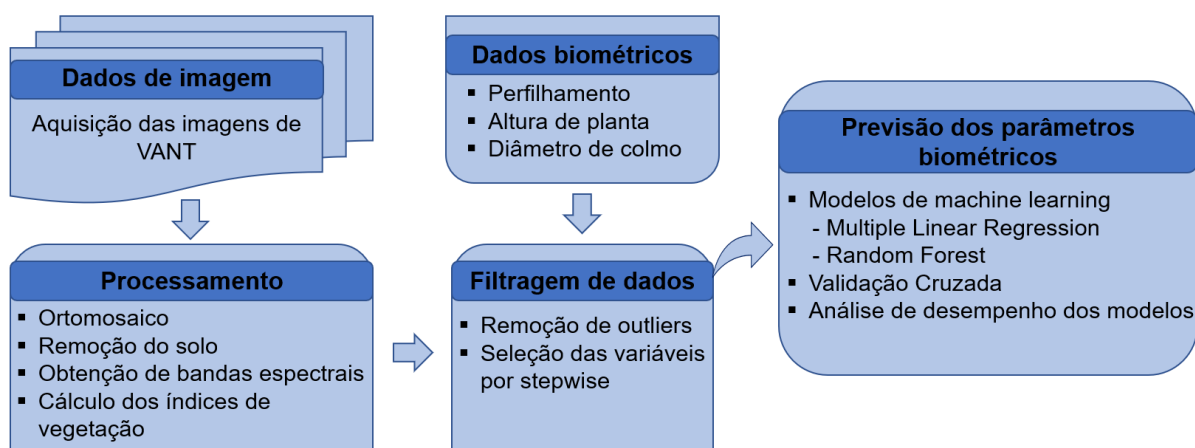


Figura 2. Fluxo de trabalho do processo de aquisição, processamento e modelagem dos dados.

2.3. Resultados

2.3.1. Seleção das variáveis preditoras

Para seleção das variáveis preditoras por *stepwise*, foram obtidas duas combinações preditoras com bom desempenho, compostas de duas e três variáveis. A primeira combinação selecionada, composta de duas variáveis, foi representada pelas bandas espectrais Blue e Green, a segunda combinação correspondeu às bandas Blue, Green e NIR (Tabela 3). A combinação de três variáveis obteve modelos mais robustos. Para NP, a combinação de duas variáveis apresentou acurácia correspondente ao RMSE = 3,9, com três variáveis o RMSE = 3,6. O desempenho dos modelos permaneceu estável até seis variáveis, depois o RMSE aumentou à medida que variáveis adicionais foram incluídas, o mesmo pode ser observado para AP e DC (Figura 3). Para AP, obtivemos RMSE = 20,4 cm com duas variáveis e 20,2 cm com três variáveis. As variáveis preditoras para DC mostraram acurácia de 2,54 mm e 2,52 mm para duas e três variáveis, respectivamente.

Tabela 3. Variáveis selecionadas por stepwise.

Combinações	Bandas
Duas variáveis	Blue, Green
Três variáveis	Blue, Green e NIR

Como as bandas da primeira combinação são todas da região espectral visível, os modelos foram avaliados usando as duas combinações de variáveis selecionadas, com duas e três variáveis.

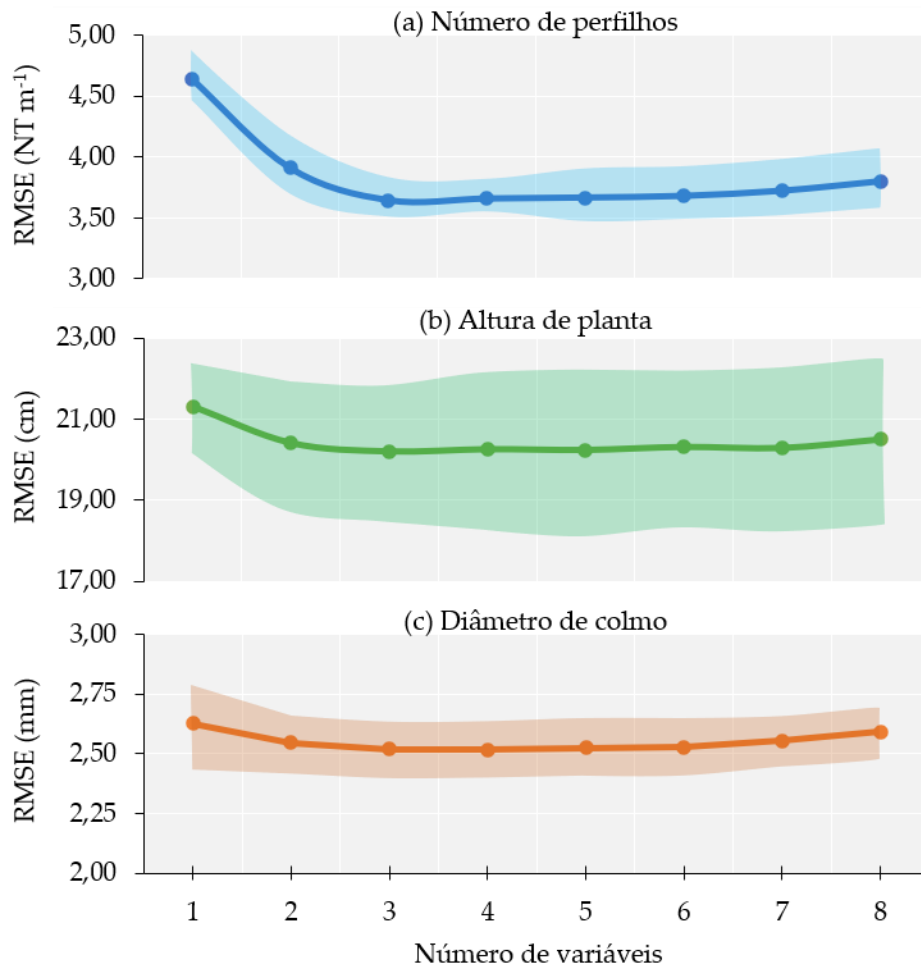


Figura 3. Desempenho da seleção das variáveis. (a) número de perfilhos – NP; (b) altura de planta – AP; (c); diâmetro de colmo - DC.

2.3.2. Predições dos parâmetros biométricos

Os modelos de previsão para NP mostraram melhor precisão e acurácia usando o algoritmo RF com três variáveis, de acordo com os valores de R^2 , MAE e RMSE na Tabela 4. Usando três variáveis, a fusão de dados de séries temporais de imagens (349;397) exibiu melhor precisão e acurácia, com valor de $R^2 = 0,91$, MAE = 1,04 e RMSE = 1,40. Quanto ao desempenho das previsões independentes de imagens, o GDA de 349 mostrou melhor precisão e acurácia (maior R^2 e menor MAE

e RMSE) em comparação com o GDA de 397. Ao utilizar duas variáveis, a precisão e acurácia dos modelos diminuiu. Os modelos que utilizavam ambos GDA apresentaram melhor desempenho em comparação com a previsão independente, com $R^2 = 0,88$, MAE = 1,29 e RMSE = 1,67. Com relação à previsão independente de imagens, o GDA de 349 mostrou maior desempenho, com $R^2 = 0,85$, MAE = 1,42 e RMSE = 1,83.

Tabela 4. Resultados dos algoritmos de RF e MLR para previsão do NP.

Algoritmos	Variáveis	Métricas	GDA		
			349	397	349+397
RF	Duas	R^2	0,85	0,82	0,88
		MAE	1,42	1,59	1,29
		RMSE	1,83	2,08	1,67
	Três	R^2	0,90	0,84	0,91
		MAE	1,13	1,41	1,04
		RMSE	1,50	1,87	1,40
MLR	Duas	R^2	0,56	0,48	0,57
		MAE	2,50	2,73	2,47
		RMSE	3,15	3,50	3,11
	Três	R^2	0,65	0,52	0,67
		MAE	2,16	2,61	2,11
		RMSE	2,78	3,34	2,73

O algoritmo MLR foi semelhante ao RF para prever NP. Entretanto, com valores de precisão e acurácia menores (Tabela 3). Usando três variáveis, a precisão e acurácia dos modelos superaram o desempenho dos modelos usando duas variáveis. Com três variáveis, a combinação de GDA = 349 e 397 mostrou precisão ($R^2 = 0,67$) e acurácia (MAE = 2,11 e RMSE = 2,73) ligeiramente superior à previsão independente GDA = 349, com valor de $R^2 = 0,65$, MAE = 2,16 e RMSE = 2,78. O mesmo pode ser observado usando duas variáveis, a combinação de GDA = 349 e 397 com melhor desempenho ($R^2 = 0,57$, MAE = 2,47 e RMSE = 3,11), seguido por previsões independentes de GDA = 349 ($R^2 = 0,56$, MAE = 2,50 e RMSE = 3,15) e GDA = 397 ($R^2 = 0,48$, MAE = 2,73 e RMSE = 3,50). Observa-se nos gráficos de dispersão as melhores correlações entre os parâmetros observados e preditos para o NP, com R^2 superior a 0,90 para RF e 0,67 para MLR (Figura 4).

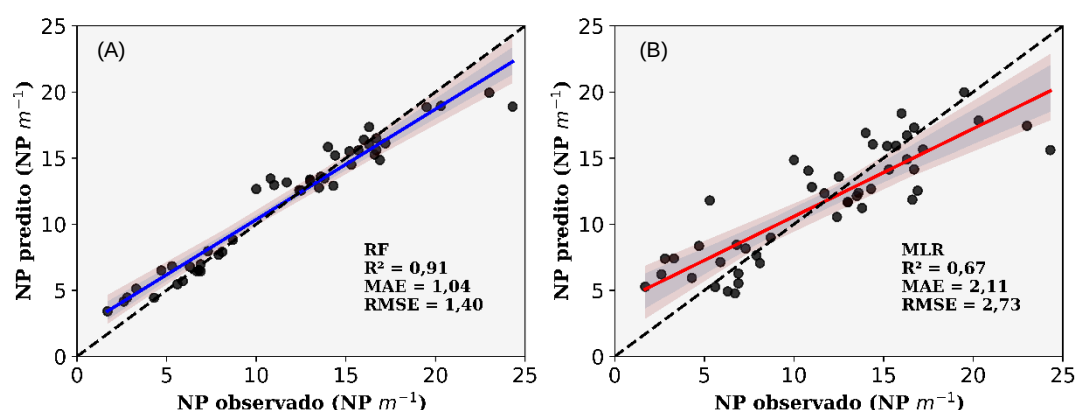


Figura 4. Gráficos de dispersão dos valores de NP observado versus predito usando três variáveis pela validação cruzada dos algoritmos de RF (A) e MLR (B). Combinação das imagens 349 GDA e 397 GDA.

O desempenho do algoritmo de RF superou o algoritmo de MLR para previsão de AP (Tabela 5). Usando três variáveis, o desempenho dos modelos foi maior em comparação aos modelos que usavam duas variáveis. A combinação de GDA = 349 e 397 revelou resultados mais precisos e acurados, com $R^2 = 0,88$, MAE = 6,97 cm e RMSE = 8,90 cm. Quanto ao desempenho de previsão independente das imagens, GDA = 349 mostrou resultados mais precisos e acurados em comparação com GDA = 397, com $R^2 = 0,84$, MAE = 7,77 cm e RMSE = 10,18 cm. Para duas variáveis, a combinação de GDA = 349 e 397 mostrou bom desempenho, com $R^2 = 0,86$, MAE = 7,54 cm e RMSE = 9,71 cm. GDA = 349 mostrou maior precisão e acurácia do que GDA = 397, com valores de $R^2 = 0,81$, MAE = 8,60 cm e RMSE = 11,16 cm (Figura 5).

Tabela 5. Resultados dos algoritmos de RF e MLR para previsão de AP.

Algoritmos	Variáveis	Métricas	GDA		
			349	397	349+397
RF	Duas	R^2	0,81	0,72	0,86
		MAE (cm)	8,60	10,24	7,54
		RMSE (cm)	11,16	13,50	9,71
	Três	R^2	0,84	0,83	0,88
		MAE (cm)	7,77	8,08	6,97
		RMSE (cm)	10,18	10,44	8,90
MLR	Duas	R^2	0,46	0,32	0,60
		MAE (cm)	14,60	16,42	12,89
		RMSE (cm)	18,81	21,02	16,32
	Três	R^2	0,49	0,53	0,65
		MAE (cm)	14,45	14,09	11,98
		RMSE (cm)	18,34	17,49	15,17

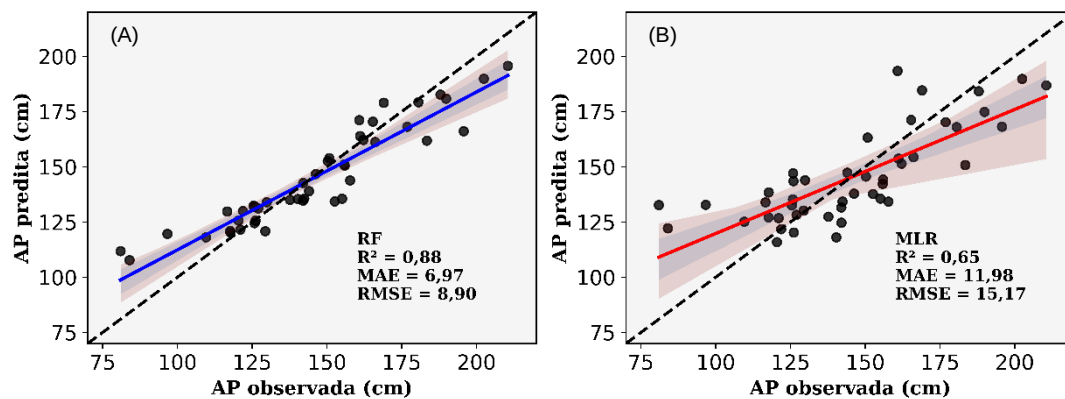


Figura 5. Gráficos de dispersão dos valores de AP observado versus predito usando três variáveis pela validação cruzada dos algoritmos de RF (A) e MLR (B). Combinação das imagens 349 GDA e 397 GDA.

A MLR exibiu acurácia relativamente alta para o AP. Entretanto, a precisão dos modelos variou de 0,32 a 0,65 considerando o uso de duas e três variáveis (Tabela 4). Com três variáveis, o melhor modelo correspondeu à combinação GDA = 349 e 397, que obteve $R^2 = 0,65$, MAE = 11,98 cm e RMSE = 15,17 cm. As previsões independentes de GDA = 349 e 397 exibiram desempenho moderado, com precisão inferior a 0,54. GDA = 397 apresentou melhor precisão e acurácia em comparação com GDA = 349, com $R^2 = 0,53$, MAE = 14,09 cm e RMSE = 17,49 cm. Os modelos com duas variáveis de entrada apresentaram desempenho inferior aos modelos com três variáveis. O melhor modelo de previsão correspondeu à combinação de GDA = 349 e 397 ($R^2 = 0,60$, MAE = 12,89 cm e RMSE = 16,32 cm), seguido da previsão independente de GDA = 349 ($R^2 = 0,46$, MAE = 14,60 cm e RMSE = 18,81 cm) e GDA = 397 ($R^2 = 0,32$, MAE = 16,42 cm e RMSE = 21,02 cm).

Os resultados de previsão para DC apresentaram modelos mais precisos e acurados utilizando o algoritmo RF (Tabela 6). Quanto ao uso de duas ou três variáveis, assim como para os parâmetros preditos anteriormente, os modelos com três variáveis apresentaram melhores desempenhos para ambos os algoritmos (RF e MLR).

Tabela 6. Resultados dos algoritmos de RF e MLR para previsão de DC.

Algoritmos	Variáveis	Métricas	GDA			
			349	397	349+397	349+397+410
RF	Duas	R ²	0,49	0,32	0,50	0,50
		MAE (mm)	1,54	1,78	1,53	1,52
		RMSE (mm)	1,97	2,27	1,95	1,93
	Três	R ²	0,51	0,39	0,52	0,52
		MAE (mm)	1,53	1,69	1,51	1,50
		RMSE (mm)	1,95	2,16	1,91	1,88
MLR	Duas	R ²	0,35	0,15	0,36	0,36
		MAE (mm)	1,72	2,00	1,70	1,68
		RMSE (mm)	2,22	2,55	2,21	2,19
	Três	R ²	0,38	0,34	0,40	0,43
		MAE (mm)	1,64	1,75	1,64	1,55
		RMSE (mm)	2,17	2,25	2,14	2,07

O algoritmo de RF com três variáveis apresentou melhor desempenho usando a combinação de GDA = 349 e 397, assim como a combinação de GDA = 349, 397 e 410, essas duas combinações apresentaram $R^2 = 0,52$. Para GDA = 349 e 397, obteve MAE = 1,51 mm e RMSE = 1,91 e para a combinação de GDA = 349, 397 e 410, foi obtido MAE = 1,50 mm e RMSE = 1,88 mm. Para modelos de previsão independentes, GDA = 349 foi mais preciso e acurado ($R^2 = 0,51$, MAE = 1,53 mm e RMSE = 1,95 mm) do que GDA = 397. O uso de duas variáveis mostrou resultados próximos aos valores usando três variáveis, as maiores diferenças foram para GDA = 397, com valores menores de R^2 , MAE e RMSE.

O uso de três variáveis no algoritmo MLR para previsão do DC, apresentou melhor desempenho com a combinação de GDA = 349, 397, e 410, com $R^2 = 0,43$, MAE = 1,55 mm e RMSE = 2,07 mm. Em contraste, combinando GDA = 349 e 397, as métricas foram $R^2 = 0,40$, MAE = 1,64 mm e RMSE = 2,14 mm (Tabela 5). Com relação às previsões independentes, GDA = 349 obteve resultados menos precisos e acurados em comparação com qualquer combinação de GDA, no entanto, com valores relativamente próximos ($R^2 = 0,38$, MAE = 1,64 mm e RMSE = 2,17 mm. Por outro lado, a menor precisão e acurácia, foram alcançadas em GDA = 397, com $R^2 = 0,15$, MAE = 1,75 e RMSE = 2,25 mm. O uso de duas variáveis mostrou melhor desempenho usando combinações de GDA = 349 e 397 e GDA = 349, 397 e 410. As duas combinações apresentaram $R^2 = 0,36$, usando GDA = 349, 397 e 410, obteve

MAE = 1,68 mm e RMSE = 2,19 mm. A previsão independente de GDA = 349 mostrou $R^2 = 0,35$, MAE = 1,72 mm e RMSE = 2,22 mm. Os valores métricos para GDA = 397 foram $R^2 = 0,15$, MAE = 2,00 mm e RMSE = 2,55 mm. Em resumo, os modelos RF e MLR apresentaram boa acurácia (Figura 6).

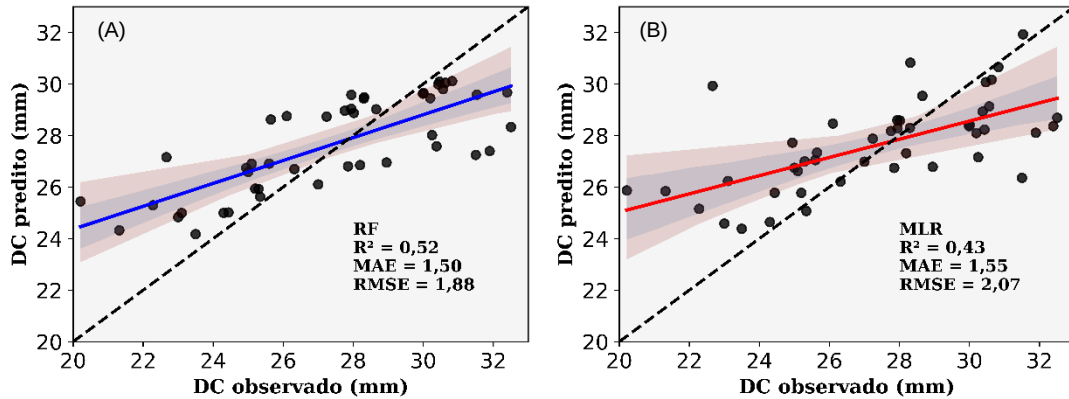


Figura 6. Gráficos de dispersão dos valores de DC observado versus predito usando três variáveis pela validação cruzada dos algoritmos de RF (A) e MLR (B). Combinação das imagens 349, 397 e 410 GDA.

2.4. Discussão

Neste estudo, foi analisado se é possível prever parâmetros biométricos de cana-de-açúcar utilizando imagens multiespectrais de VANT. Se confirmado, isso poderá servir como base para previsão acurada e oportuna de produtividade de cana-de-açúcar. É possível encontrar numerosos estudos que investigaram melhorias na previsão de produtividade de cana-de-açúcar (Abebe et al., 2022b; Tanut et al., 2021; Yu et al., 2022). No entanto, esta é uma abordagem que utilizou imagens multiespectral de VANT combinando algoritmos de ML para prever o número de perfilhos, a altura da planta e o diâmetro do colmo, particularmente, três parâmetros essenciais para prever produtividade de cana-de-açúcar.

O método *stepwise* aplicado em nossa abordagem apresentou as bandas Blue, Green e NIR como preditores de entrada para previsão do NP, AP e DC da cana-de-açúcar. A seleção das variáveis por *stepwise* mostrou estabilidade na precisão dos modelos a partir de três variáveis. O uso de dados combinados (séries temporais) e não combinados (previsão independente sem combinação de GDA) de imagens

multiespectrais permitiu obter resultados precisos e acurados para prever parâmetros biométricos da cana-de-açúcar até sete meses antes da colheita, que podem ser usados para melhor compreensão da variabilidade no campo e para o processo de tomada de decisão. Os modelos de RF obtiveram bons resultados na previsão dos parâmetros, com R^2 superiores a 0,70 para todos os modelos de NP e AP, exceto para DC, a qual a precisão foi inferior. Em relação as abordagens utilizando algoritmos de MLR, obteve-se desempenho inferior aos algoritmos de RF. Esse desempenho inferior pode estar relacionado com a incapacidade da MLR em processar relações não lineares entre as variáveis dependentes e independentes (Miphokasap e Wannasiri, 2018). Jeong et al. (2016) avaliaram modelos de RF e MLR para previsões de safras das culturas de milho, trigo e batata, obtiveram resultados em que os modelos de RF superaram os modelos de MLR para as culturas avaliadas. Neste trabalho, apesar dos baixos valores de R^2 de alguns modelos da MLR, os resultados podem ser considerados acurados, pois apresentaram baixos valores de RMSE.

Em relação a utilização de duas ou três variáveis preditoras de entrada, os modelos com três variáveis obtiveram resultados mais precisos e acurados, no entanto, com valores próximos dos modelos com duas variáveis. Os modelos com duas variáveis utilizaram as bandas Blue e Green. Essas bandas estão na faixa espectral visível, portanto, podem ser obtidas usando câmeras RGB, que geralmente têm custos inferiores comparado a câmeras multiespectrais e requerem menos procedimentos de calibração (Chianucci et al., 2016; Costa et al., 2020), isso possibilita eficiência e economia na obtenção desses dados e potencial uso dessas variáveis na previsão desses parâmetros. Alguns trabalhos de previsões da produção de culturas encontraram resultados semelhantes entre índice de vegetação calculado a partir das bandas Blue e Green e índice utilizando a banda NIR, com uso de modelos de RF e regressão linear, além disso, conseguiram encontrar melhores resultados utilizando a combinação desses índices (Sulik e Long, 2016; Wan et al., 2020).

Quanto a previsão do NP, foi observado que o uso de duas variáveis (Blue e Green) e três variáveis (Blue, Green e NIR) de entrada obtiveram resultados com elevado R^2 e que a combinação de GDA = 349 e GDA = 397 fornece resultados mais precisos e acurados. Esses resultados podem estar relacionados às características de reflectância do dossel. À medida que o NP da cana-de-açúcar aumenta, o índice

de área foliar (Santos e Diola, 2015) e a biomassa também aumentam (Sumesh et al., 2021). Esse crescimento na cobertura do dossel permite maior absorção da radiação eletromagnética visível, devido à forte influência dos pigmentos foliares (Hamzeh et al., 2016; Yang et al., 2017). Nesse caso, a banda Blue apresenta maior absorção da radiação em plantas com maiores índices de perfilhamento, pois a clorofila apresenta picos de absorção da radiação nos comprimentos de onda das faixas do azul e do vermelho (Homolová et al., 2013). Em relação a banda Green, o mesmo processo pode estar relacionado, no entanto, essa banda é caracterizada por acentuada redução da absorção da radiação em comparação às demais bandas da faixa espectral do visível, apresentando características de reflectância mais elevada do que a banda Blue, entretanto, apresenta padrões de reflectância bastante inferior em comparação ao NIR (Hamzeh et al., 2016; Yang et al., 2017). Sumesh et al. (2021) utilizaram IV obtido de imagens RGB (ExG - índice de excesso de verde médio) e regressão de mínimos quadrados ordinários (OLS) para estimar a densidade de colmos da cana-de-açúcar, obtendo resultados de R^2 de 0,72 e RMSE igual a 9,84 colmos. NIR é uma banda espectral que ocorre predominância de reflexão da radiação nas plantas saudáveis, reflectância essa que pode aumentar à medida em que o índice de área foliar aumenta. Nesse sentido, os bons resultados, incluindo a banda NIR na previsão do NP, podem estar relacionados à elevada sensibilidade dessa banda às variações na biomassa (Asner, 1998; Fortes e Demattê, 2006; Yang et al., 2017). Outros estudos de estimativa de parâmetros da cana-de-açúcar por imagens multiespectrais obtiveram bons resultados utilizando IVs derivados do NIR e de uma banda do espectro do visível, como o NDVI (Lisboa et al., 2018; Luciano et al., 2021).

Em relação à AP e DC, foram obtidos modelos de previsão acurados. AP e DC são características que se correlacionam diretamente com a biomassa (Bendig et al., 2014; Yu et al., 2020; Zhang et al., 2021). Nesse sentido, vários trabalhos encontraram relações entre bandas espectrais, IVs e parâmetros biofísicos das culturas, especialmente biomassa, devido ao fato de que estes parâmetros influenciam a intensidade da reflectância das plantas (Shendryk et al., 2020; Yue et al., 2019; Zhang et al., 2021). Os resultados mostraram que os modelos de previsão do NP, AP e DC da cana-de-açúcar através de séries temporais e os algoritmos MLR e RF apresentaram previsões acuradas. Portanto, o uso de sensores multiespectrais a

bordo de VANT pode ser considerado técnica viável para avaliação da cultura da cana-de-açúcar, sendo que apresentam capacidade de previsão de parâmetros biométricos em grandes áreas de produção.

2.5. Conclusões

Os modelos treinados foram acurados para a previsão de produtividade de cana-de-açúcar até sete meses antes da colheita. Este trabalho trouxe uma novidade para a previsão da produtividade. O uso de bandas espectrais obtidas por sensor multiespectral a bordo de VANT associadas a algoritmos de *Machine Learning* obtiveram resultados precisos e acurados para previsão do número de perfilhos, altura da planta e diâmetro do colmo da cana-de-açúcar. Os modelos de *Random forest* apresentaram precisão e acurácia superior aos modelos de *Multiple linear regression*. As bandas espectrais Blue, Green e NIR forneceram elevado desempenho para serem utilizadas na previsão de parâmetros biométricos da cana-de-açúcar. Os modelos construídos com as bandas Blue e Green também apresentaram elevada precisão nas previsões, possibilitando economia, pois essas bandas podem ser obtidas usando sensores de baixo custo. Os dados espectrais combinados (série temporal) melhoram ainda mais a precisão dos modelos de previsão dos parâmetros biométricos.

2.6. Referências

- Abebe G, Tadesse T, Gessesse B (2022a) Assimilation of leaf Area Index from multisource earth observation data into the WOFOST model for sugarcane yield estimation. **International Journal of Remote Sensing** 43. Taylor & Francis:698–720. doi:10.1080/01431161.2022.2027547.
- Abebe G, Tadesse T, Gessesse B (2022b) Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia. **Journal of the Indian Society of Remote Sensing** 50:143–157. doi:10.1007/s12524-021-01466-8.
- Alvares CA, Stape JL, Sentelhas PC, De Moraes Gonçalves JL, Sparovek G (2013) Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift** 22:711–

728. doi:10.1127/0941-2948/2013/0507.

Arnold CY. The determination and significance of the base temperature in a linear heat unit system. In: *Proceedings of the american Society for horticultural Science* 1959, 430-445.

Asner GP (1998) Biophysical and Biochemical Sources of Variability in Canopy Reflectance. **Remote Sensing of Environment** 64:234–253. doi:10.1016/S0034-4257(98)00014-5.

Barbosa Júnior MR, Moreira BR de A, Brito Filho AL de, Tedesco D, Shiratsuchi LS, Silva RP da (2022) UAVs to Monitor and Manage Sugarcane: Integrative Review. **Agronomy** 12:661. doi:10.3390/agronomy12030661.

Bendig J, Bolten A, Bennertz S, Broscheit J, Eichfuss S, Bareth G (2014) remote sensing Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. **Remote Sens.** 6:10395–10412. doi:10.3390/rs61110395.

Caetano JM, Casaroli D (2017) Sugarcane yield estimation for climatic conditions in the state of Goiás. **Revista Ceres** 64:298–306. doi:10.1590/0034-737x201764030011.

Chianucci F, Disperati L, Guzzi D, Bianchini D, Nardino V, Lastrì C, Rindinella A, Corona P (2016) Estimation of canopy attributes in beech forests using true colour digital images from a small fixed-wing UAV. **International Journal of Applied Earth Observation and Geoinformation** 47:60–68. doi:10.1016/J.JAG.2015.12.005.

Costa L, Nunes L, Ampatzidis Y (2020) A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. **Computers and Electronics in Agriculture** 172:105334. doi:10.1016/J.COMPAG.2020.105334.

FAOSTAT - Food and Agriculture Organization of the United Nation (2021). Disponível em: <http://www.fao.org/faostat/en/?#data/QC> (acessado 27 dezembro 2021).

Fortes C, Demattê JAM (2006) Discrimination of sugarcane varieties using Landsat 7 ETM+ spectral data. **International Journal of Remote Sensing** 27:1395–1412.

doi:10.1080/01431160500383863.

Guo Y, Fu Y, Hao F, Zhang X, Wu W, Jin X, Robin Bryant C, Senthilnath J (2021) Integrated phenology and climate in rice yields prediction using machine learning methods. **Ecological Indicators** 120:106935. doi:10.1016/j.ecolind.2020.106935.

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. doi:10.1016/J.JAG.2016.06.024.

Homolová L, Malenovský Z, Clevers JGPW, García-Santos G, Schaepman ME (2013) Review of optical-based remote sensing for plant trait mapping. **Ecological Complexity** 15:1–16. doi:10.1016/J.ECOCOM.2013.06.003.

Inman-Bamber NG (1994) Temperature and seasonal effects on canopy development and light interception of sugarcane. **Field Crops Research** 36:41–51. doi:10.1016/0378-4290(94)90051-5.

Jeong JH, Resop JP, Mueller ND et al. (2016) Random forests for global and regional crop yield predictions. **PLoS ONE** 11:e0156571. doi:10.1371/journal.pone.0156571.

Lisboa IP, Damian M, 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** 8:196. doi:10.3390/agronomy8090196.

Luciano AC dos S, Picoli MCA, Duft DG, Rocha JV, Leal MRLV, le Maire G (2021) Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. **Computers and Electronics in Agriculture** 184:106063. doi:10.1016/j.compag.2021.106063.

Miphokasap P, Wannasiri W (2018) Estimations of Nitrogen Concentration in sugarcane using hyperspectral imagery. **Sustainability (Switzerland)** 10:1266. doi:10.3390/su10041266.

Picoli MCA, Lamparelli RAC, Sano EE, Rocha J V (2014) THE USE OF ALOS/PALSAR DATA FOR ESTIMATING SUGARCANE PRODUCTIVITY. **Eng. Agríc** 34:1245–1255.

Rossi Neto J, de Souza ZM, Kölln OT et al. (2018) The Arrangement and Spacing of Sugarcane Planting Influence Root Distribution and Crop Yield. **Bioenergy Research** 11:291–304. doi:10.1007/s12155-018-9896-1.

Santos F, Diola V (2015) Physiology. In: **Sugarcane: Agricultural Production, Bioenergy and Ethanol**. Academic Press, p. 13–33. doi:10.1016/B978-0-12-802239-9.00002-5.

Santos HG, Jacomine PKT Dos, Anjos LHC et al. (2018) *Sistema brasileiro de classificação de solos*. Brasília: Embrapa, 356 p.

Shendryk Y, Sofonia J, Garrard R, Rist Y, Skocaj D, Thorburn P (2020) Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. **International Journal of Applied Earth Observation and Geoinformation** 92:102177. doi:10.1016/j.jag.2020.102177.

Simões MDS, Rocha JV, Lamparelli RAC (2009) Orbital spectral variables, growth analysis and sugarcane yield. **Sci. Agric** 66:451–461. doi:10.1590/S0103-90162009000400004

Som-ard J, Hossain MD, Ninsawat S, Veerachitt V (2018) Pre-harvest Sugarcane Yield Estimation Using UAV-Based RGB Images and Ground Observation. **Sugar Tech** 20:645–657. doi:10.1007/s12355-018-0601-7.

Sulik JJ, Long DS (2016) Spectral considerations for modeling yield of canola. **Remote Sensing of Environment** 184:161–174. doi:10.1016/j.rse.2016.06.016.

Sumesh KC, Ninsawat S, Som-ard J (2021) Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. **Computers and Electronics in Agriculture** 180:105903. doi:10.1016/j.compag.2020.105903.

Tanut B, Waranusast R, Riyamongkol P (2021) High Accuracy Pre-Harvest Sugarcane

Yield Forecasting Model Utilizing Drone Image Analysis, Data Mining, and Reverse Design Method. **Agriculture** 11:682. doi:10.3390/AGRICULTURE11070682.

Urolagin S, Sharma N, Datta TK (2021) A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting. **Energy** 231:120963. doi:10.1016/j.energy.2021.120963.

Wan L, Cen H, Zhu J et al. (2020) Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – a case study of small farmlands in the South of China. **Agricultural and Forest Meteorology** 291:108096. doi:10.1016/j.agrformet.2020.108096.

Wang J, Li X, Lu L, Fang F (2013) Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. **European Journal of Agronomy** 49:126–140. doi:10.1016/j.eja.2013.03.005.

Xu JX, Ma J, Tang YN et al. (2020) Estimation of sugarcane yield using a machine learning approach based on uav-lidar data. **Remote Sensing** 12:2823. doi:10.3390/rs12172823.

Yang G, Liu J, Zhao C et al. (2017) Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. **Frontiers in Plant Science** 8:1111. doi:10.3389/fpls.2017.01111.

Yu D, Zha Y, Shi L, Jin X, Hu S, Yang Q, Huang K, Zeng W (2020) Improvement of sugarcane yield estimation by assimilating UAV-derived plant height observations. **European Journal of Agronomy** 121:126159. doi:10.1016/j.eja.2020.126159.

Yu D, Zha Y, Shi L, Ye H, Zhang Y (2022) Improving sugarcane growth simulations by integrating multi-source observations into a crop model. **European Journal of Agronomy** 132:126410. doi:10.1016/j.eja.2021.126410.

Yue J, Yang G, Tian Q, Feng H, Xu K, Zhou C (2019) Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. **ISPRS Journal of Photogrammetry and Remote Sensing** 150:226–244. doi:10.1016/j.isprsjprs.2019.02.022.

Zhang Y, Xia C, Zhang X, Cheng X, Feng G, Wang Y, Gao Q (2021) Estimating the maize biomass by crop height and narrowband vegetation indices derived from UAV-based hyperspectral images. **Ecological Indicators** 129:107985. doi:10.1016/j.ecolind.2021.107985.

CAPÍTULO 3 – Considerações finais

Neste trabalho, de acordo com os resultados obtidos, foi demonstrado que a utilização de bandas espectrais obtidos por sensor embarcado em VANT é viável para previsão de perfilhamento, diâmetro e altura de cana-de-açúcar. Pelo método convencional, os dados desses parâmetros são coletados de forma manual nos campos de produção, que geralmente apresentam extensas áreas, gerando muito trabalho e tempo. Desta forma, utilizar sensores embarcados em VANT é excelente alternativa para o desenvolvimento de modelo de previsão de produtividade em cana-de-açúcar. É uma tecnologia rápida, precisa e não destrutiva que proporciona tomada de decisão mais assertiva quanto ao manejo da cultura.

A determinação desses parâmetros biométricos nas condições estudadas não soluciona todos os problemas do processo de previsão de produtividade de cana-de-açúcar. É importante investigar o comportamento dos dados espectrais no decorrer do desenvolvimento da cultura. Por isso, em futuras pesquisas será avaliado a utilização de dados espectrais para a previsão dos parâmetros biométricos em diferentes estádios fenológico da cana-de-açúcar. Pois podem ocorrer variações nas características de reflectância da radiação pela vegetação de acordo o estágio de desenvolvimento.

Também serão avaliados o uso de bandas espectrais e índices de vegetação na estimativa de fatores que podem influenciar na produtividade, como o status nutricional das plantas. A qualidade do estado nutricional é um dos principais fatores para o desenvolvimento da cultura.

Esses conhecimentos proporcionarão meios para elaboração de mapas nutricionais e de produtividade, que poderão ser utilizados para aplicação localizada de fertilizantes, conforme necessidade específica das plantas em cada ponto da lavoura. Possibilitando redução nos custos de produção e aumento da produtividade da cana-de-açúcar, respeitando o equilíbrio ambiental.