UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICBAL

ABORDAGEM PARA PREDIÇÃO DE SOJA POR ÍNDICES DE VEGETAÇÃO E MODELOS DE MACHINE LEARNING

Túlio Mazetti Marra

Engenheiro de biossistemas

2023

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICBAL

ABORDAGEM PARA PREDIÇÃO DE SOJA POR ÍNDICES DE VEGETAÇÃO E MODELOS DE MACHINE LEARNING

Túlio Mazetti Marra

Orientador: Prof. Dr. Carlos Eduardo Angeli Furlani

Coorientador: Prof. Dr. Rouverson Pereira da Silva

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).

2023

M358a

Marra, Túlio Mazetti
Abordagem para predição de soja por índices de vegetação e modelos de machine learning / Túlio Mazetti Marra. -- Jaboticabal, 2023
53 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Carlos Eduardo Angeli Furlani
Coorientador: Rouverson Pereira da Silva

1. Soja. 2. Predição de produtividade. 3. PlanetScope. 4. Redes Neurais. 5. Random forest. I. Título.

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

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa

Ao prever a produtividade da soja com imagens de satélite, esse estudo otimiza agricultura, impulsiona segurança alimentar, minimiza desperdícios e orienta decisões econômicas e ambientais, contribuindo para um futuro agrícola mais eficiente e sustentável.

Potential impact of this research

By predicting soybean yield with satellite images, this study optimizes agriculture, boosts food security, minimizes waste and guides ecological and environmental decisions, hopefully for a more efficient and sustainable agricultural future.

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: ABORDAGEM PARA PREDIÇÃO DE SOJA POR INDICES DE VEGETAÇÃO E MODELOS DE MACHINE LEARNING

AUTOR: TÚLIO MAZETTI MARRA
ORIENTADOR: CARLOS EDUARDO ANGELI FURLANI
COORIENTADOR: ROUVERSON PEREIRA DA SILVA

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

Prof. Dr. CARLOS EDUARDO ANGELI FURLANI (Participaçao Virtual)
Departamento de Engenharia e Ciencias Exatas DECEx / FCAV UNESP Jaboticabal

Profa. Dra. TATIANA FERNANDA CANATA (Participaçao Virtual)
Departamento de Engenharia e Ciências Exatas / FCAV - UNESP Jaboticabal

Prof. Dr. DANILO TEDESCO DE OLIVEIRA (Participaçao Virtual)
Kansas State University / Manhattan/Kansas - EUA

Jaboticabal, 14 de julho de 2023

gov.br Documento assinado digitalmente
DANILO TEDESCO DE OLIVEIRA
Data: 14/07/2023 11:28:36-0300
Verifique em https://validar.iti.gov.br

DADOS CURRICULARES DO AUTOR

TÚLIO MAZETTI MARRA, filho de Fernando Marra e Sueli de Faria Mazetti Marra, nasceu em Ituverava – SP no dia 12 de junho de 1996, município localizado no interior do Estado de São Paulo. Cursou ensino fundamental e médio no município de Ituverava – SP, no período entre 2003 e 2013. Em 2014 iniciou o curso de bacharelado em Engenharia de Biossistemas na Universidade de São Paulo – USP, Câmpus de Pirassununga. Durante o curso foi membro do Laboratório de Bioquímica Aplicada, desenvolvendo atividades de pesquisa aplicada e extensão em degradação de corantes alimentícios artificias e antioxidantes para biodiesel. Foi bolsista no Programa Unificado de Bolsas de Estudos para Apoio à Permanência e Formação de Estudantes de Graduação (PUB-USP) por três anos, nas áreas de bioquímica aplicada, sob a orientação do Prof. Dr. Fernando Gustavo Tonim, agricultura de precisão, sob a orientação do Prof. Dr. Murilo Mesquita Baesso, e Zootecnia de Precisão, sob a orientação da Profa. Dra. Luciane Silva Martello. Obteve título de Engenheiro de Biossistemas no ano de 2020, ingressando no ano seguinte no Programa de Pós-graduação em Agronomia (Produção Vegetal) na Universidade Estadual Paulista "Júlio de Mesquita Filho" Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal – UNESP/FCAV, com bolsa de estudos concedida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), sob orientação do Prof. Dr. Carlos Eduardo Angeli Furlani. É integrante do Laboratório de Máquinas e Mecanização Agrícola (LAMMA), participando de projetos relacionados à agricultura de precisão, agricultura digital e sensoriamento remoto, com ênfase na utilização de imageamento aéreo e orbital, para obtenção de informações precisas de culturas agrícolas.

‘‘Que não me falte serenidade para aceitar as coisas que não posso mudar, coragem para mudar as que posso e sabedoria para distinguir entre elas’’.

Reinhold Niebuhr

Aos meus pais, Fernando Marra
e Sueli de Faria Mazetti Marra.
DEDICO!

AGRADECIMENTOS

À minha família, em nome dos meus pais, Fernando Marra e Sueli de Faria Mazetti Marra.

Ao meu orientador, Prof. Dr. Carlos Eduardo Angeli Furlani, e coorientador, Prof. Dr. Rouverson Pereira da Silva, que tenho grande admiração e consideração pelo seu trabalho, pela confiança e por possbilitar meu crescimento profissional e pessoal.

À Universidade de São Paulo, pelas condições dadas para a minha formação profissional (Engenheiro de Biossistemas) e pessoal, principalmente por todos os valores adquiridos.

Aos professores Fernando Gustavo Tonim, Murilo Mesquita Baesso e Luciane Silva Martello, pela orientação durante a graduação, disponibilidade para o desenvolvimento de trabalhos, pela paciência, compreensão e confiança.

À 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), pela oportunidade de poder aprimorar meus conceitos e adquirir novos conhecimentos em uma instituição de alto nível.

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

Aos integrantes e parceiros do Laboratório de Máquinas e Mecanização Agrícola – LAMMA, onde tive oportunidade de construir amizades que contribuiram muito para minha formação e crescimento, tanto na parte profissional como pessoal.

A todos que, de alguma maneira, contribuíram para a realização e conclusão deste trabalho. Minha mais sincera gratidão. Obrigado.

SUMÁRIO

LISTA DE TABELAS 1
LISTA DE FIGURAS 2
CAPÍTULO 1 – CONSIDERAÇÕES GERAIS 3
1.1. INTRODUÇÃO 3
1.2. REVISÃO BIBLIOGRÁFICA 4
1.2.1. A cultura da soja 4
1.2.2. Imageamento orbital na agricultura 6
1.2.3. Índices de vegetação na agricultura 8
1.2.4. Processamento de imagens 10
1.2.5. Uso de algoritmos para a predição de produtividade 12
1.3. Referências 13
CAPÍTULO 2 – Abordagem para a predição da soja por índices de vegetação e modelos de aprendizado de máquina 20
2.1. INTRODUÇÃO 20
2.2. MATERIAL E MÉTODOS 22
2.2.1. Área de estudo 22
2.2.2. Avaliações de campo 24
2.2.3. Aquisição das imagens PlanetScope 24
2.2.4. Índices de vegetação 25
2.2.5. Descrição dos modelos de predição de produtividade 26
2.2.6. Treinamento, validação e desempenho 27
2.3. RESULTADOS 28
2.3.1. Análise descritiva 28
2.3.2 Algoritmos preditivos de produtividade de soja 29
2.4. DISCUSSÃO 32
2.5. CONCLUSÕES 35
2.6. Referências 36
CAPÍTULO 3 - CONSIDERAÇÕES FINAIS 41

Abordagem para predição de soja por índices de vegetação e modelos de machine learning

RESUMO

A predição de produtividade da soja é uma informação importante para dar suporte a gestão de culturas, determinando o momento e local de colheita, agregando valor ao produto colhido. No entanto, esta tarefa é um desafio, uma vez que a combinação perfeita de características espectrais e algoritmos de Aprendizado de Máquina (ML) para predizer a produtividade das culturas não é amplamente conhecida. Para resolver esse quebra-cabeça, o objetivo deste estudo foi trabalhar com imagens multiespectrais de satélite de alta resolução para identificar algumas das melhores características para a predição de produtividade da soja, tais como estádios fenológicos para a coleta de imagens, algoritmos de predição e índices de vegetação (VI's) que compõe tais algoritmos. A metodologia foi explorada como forma de análise não-invasiva e não-destrutiva. A base de dados foi composta por um conjunto de imagens temporais e dados de produtividade de três cultivares de soja, produzidas em duas safras diferentes. Primeiramente, as imagens foram processadas e calculados os VI's para cada estádio fenológico de coleta das imagens, que foram de R4 a R7. Os VI's utilizados foram NDVI, GNDVI, VARI, SAVI e EVI. Em seguida, os VI's foram aplicados em modelos de regressão linear, *Random Forest* (RF) e Redes Neurais Artificiais (ANN) para predizer a produtividade da soja. MAE, MAPE e RMSE foram utilizados como métricas de acurácia. Nossos resultados mostraram que o modelo que apresentou melhor desempenho em relação à precisão e acurácia para prever a produtividade da soja foi baseado em ANN usando NDVI e GNDVI nos estádios fenológicos R5 e R4, respectivamente, provando assim ser uma alternativa precoce para apoiar decisões precisas nos campos de soja.

PALAVRAS-CHAVES: *Glycine max*, Predição de produtividade, PlanetScope, Redes Neurais, Random forest.

Approach to soybean prediction by vegetation indices and machine learning models

ABSTRACT

Predicting soybean yield is crucial task to support crop management, such as when and where to harvest, thus adding value to the harvested product. However, this task is challenging as the perfect combination of spectral features and Machine Learning (ML) algorithms to predict crop yield are not widely known. To solve this puzzle, the goal of this paper was to work with high-resolution multispectral satellite imagery to identify some of the best features for soybean yield forecasting, such as phenological stages for image collection, forecasting algorithms and vegetation indices (VI's) that compose such algorithms. The methodology was exploited as a form of non-invasive and non-destructive analysis. The database consisted of a set of temporal images and yield data of three soybean cultivars, produced in two different harvests. First, the images were processed and calculated VI's were calculated for each phenological stage of image collection, which ranged from at R4 to R7. The VI's used were NDVI, GNDVI, VARI, SAVI and EVI. Then, the VI's were applied in linear regression models, Random Forest (RF) and Artificial Neural Networks (ANN) to predict soybean yield. MAE, MAPE and RMSE were used as the accuracy metrics. Our results showed that the model that performed best with respect to accuracy to predict soybean yield was based on ANN using NDVI and GNDVI at R5 and R4 phenological stages, respectively, thus proving to be an early alternative to support accurate decisions in soybean fields.

KEYWORDS: *Glycine max*, Yield prediction, PlanetScope, Neural Network, Random Forest

LISTA DE TABELAS

Tabela 1. Produção, área plantada e produtividade dos principais estados produtores de soja do Brasil – outubro/21 6

Tabela 2. Índices de vegetação usados no estudo. 26

Tabela 3. Algoritmos utilizados na predição de produtividade da soja. 26

Tabela 4. Resultados das melhores estimativas obtidas pelos algoritmos de regressão linear em cada estádio fenológico. 29

Tabela 5. Resultados das melhores estimativas obtidas pelos algoritmos de RF em cada estádio fenológico. 31

Tabela 6. Resultados das melhores estimativas obtidas pelos algoritmos de Redes Neurais Artificiais em cada estádio fenológico. 32

LISTA DE FIGURAS

Figura 1. Etapas básicas de processamento de imagens 11

Figura 2. Condições climáticas das colheitas A) 2017/18 e B) 2020/21 22

Figura 3. Representação da A) primeira área experimental e B) segunda área experimental.....23

Figura 4. Análise descritiva de produtividade por cultivar.....29

Figura 5. Regressão linear com o GNDVI como variável preditora com dados de ambas as safras para o estádio fenológico R7.....30

Figura 6. Random Forest com o GNDVI como variável preditora com dados de ambas as safras para o estádio fenológico R4. 31

Figura 7. Redes neurais com A) perfil MLP e NDVI como variável preditora com dados de ambas as safras para o estádio fenológico R5 e B) perfil RBF com NDVI como variável preditora com números de ambas as colheitas para o estádio fenológico R7. 32

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

1.1. INTRODUÇÃO

Quanto mais informações dinâmicas e localizadas sobre uma cultura o produtor dispuser, maior será sua capacidade em otimizar um parâmetro ou sanar um problema de maneira mais precisa e individual. Essas informações adquiridas e geradas por meio do uso de diferentes equipamentos podem estar relacionadas aos mais diversos aspectos de desenvolvimento da cultura, e entre um dos mais importantes em relação a soja, podemos citar a produtividade, por estar diretamente atrelada ao lucro revertido ao produtor.

O uso de imageamento aéreo ou orbital na agricultura do Brasil está diretamente relacionada a diversidade de cultivos e necessidade de monitorar e gerenciar de forma eficiente vastas extensões territoriais do país destinadas as atividades agrícolas. A necessidade de otimizar a produção agrícola e mitigar impactos ambientais torna o uso de sensoriamento remoto uma ferramenta valiosa na agricultura do país.

Por meio de tecnologioas de sensoriamento remoto, como Veículos Aéreos Não Tripulados (VANTs) ou satélites, imagens da área de cultivo de soja ao longo do ciclo de crescimento da cultura podem ser coletadas em diferentes espectros (por exemplo, infravermelho próximo). Através do calculo de Índices de Vegetação (VI's), como o Índice de Vegetação por Diferença Normalizada (NDVI) ou o Índice de Vegetação Ajustado ao Solo (SAVI), a partir das imagens processadas é possível medir a saúde e o vigor das plantas com base na absorção e reflexão da luz.

O emprego de mecanismos tecnlógicos sofisticados que realizam uma estimação acurada da produtividade da cultura da soja faz parte de uma evolução importante nos campos. A obtenção de maneira prévia/antecipada da produtividade de uma cultura por meio da análise de imagens de alta resolução obtidas via satélite vem sendo estudada pelo fato de possibilitar a análise da variabilidade produtiva de maneira não invasiva, menos subjetiva e relativamente simples.

A utilização de sensoriamento remoto para predição de atributos de cultivo auxilia as tomadas de decisão pelos produtores em relação ao manejo de safras, como

previsão e definição de melhores épocas de semeadura ou colheita. Por outro lado, ainda permite um planejamento mais otimizado da frota de maquinário utilizado, o que evita excesso de massa no solo e consequentes problemas, como compactação, além de tornar mais eficiente a escolha de cultivares com ciclos de desenvolvimento adequado ao ambiente (Xu, Meng and Quackenbush, 2019). Todos esses benefícios agregam valor ao produto final, revertendo maior lucratividade ao produtor.

Apesar do desenvolvimento de vários estudos acerca da estimativa de produtividade de soja, ainda existem muitas dúvidas em relação a quais índices de vegetação (VI's) podem ser utilizados e quais modelos matemáticos mais adequados para atingir esse objetivo (Shammi e Meng, 2021; Zhou et al., 2021). Além disso, a aplicação e estudo de algoritmos diversificados, que levam em consideração diferentes VI's como parâmetro preditor, é essencial para que seja possível identificar métodos que geram estimativas de produtividade com respostas mais precisas e antecipadas, em diversas situações, o que foi a temática central do trabalho.

Nesse contexto apresentou-se diferentes modelos de *Machine Learning* (ML) que utilizaram diferentes VI's como fatores de predição, a fim de explorar a possibilidade de utilização dos algoritmos abordados. Acreditamos que a abordagem apresentada esclareça as possibilidades quanto ao uso de cada algoritmo, em cada um dos momentos de desenvolvimento da cultura explorados, expondo resultados em relação a acurácia de estimativa dos valores de produtividade de soja, visando aplicação a nível comercial.

A presente dissertação é dividida em 3 capítulos para melhor compreensão do trabalho, onde o primeiro capítulo é constituído pelo referencial de embasamento teórico-científico para posterior análise dos resultados obtidos no capítulo seguinte. No capítulo 2, são apresentados e discutidos os resultados dos modelos gerados pelos algoritimos abordados. Por fim, no capítulo 3, tem-se as considerações finais do trabalho.

1.2. REVISÃO BIBLIOGRÁFICA

1.2.1. A cultura da soja

A soja (*Glycine max*) é uma cultura herbácea da classe *Rosideae*. Originária da costa leste da Ásia, se transformou no grão que conhecemos hoje por meio de cruzamentos naturais e melhoramento genético conduzido por cientistas na China (EMBRAPA, 2021a). Dependendo da variedade, o crescimento pode variar entre indeterminado, determinado ou semideterminado. Trata-se de uma planta de dias curtos, que retarda seu florescimento e prolonga seu ciclo em caso de regimes de dias longos em excesso (Nepomuceno, Farias and Neumaier, 2021).

A soja desempenha um papel fundamental no cenário global da agroindústria (Gaso, Berger and Ciganda, 2019). Essa relevância é justificada devido alto valor nutricional da composição do grão, atingindo até 43% de proteína, 18 a 20% de carboidratos e até 25% de óleo. Este óleo é usado não apenas na alimentação humana e animal, mas tabém na produção de biocombustíveis (Chaplygin, Tronev and Davydova, 2021). Além disso, a proteína da soja, rica em aminoácidos essenciais, é frequentemente usada na dieta humana como substituta da carne (Tamagno et al., 2017).

Na década de 1960, a soja começou a ganhar importância comercial no Brasil, quando houve aumento na produção de suínos e aves, resultando na necessidade de uma produção agrícola voltada para a alimentação animal (EMBRAPA, 2021a). Desde então, a produção de soja vem apresentando crescimento exponencial mundialmente, devido fatores além de sua importância como fonte vegetal de proteína para alimentação humana e animal, como o desenvolvimento de um mercado internacional sólido e estruturado, a relevância do complexo produtivo da soja, entre outras razões (Hirakuri e Lazzarotto, 2014). As variedades mais usadas comercialmente atualmente apresentam características que previnem o acamamento e facilitam a colheita mecanizada.

Atualmente, Brasil e EUA concentram 37 e 30% da produção mundial de soja, respectivamente (EMBRAPA, 2021b). Trata-se da quarta maior cultura em termos de produção mundial, atrás apenas de culturas como trigo, milho e arroz, destacando-se entre as leguminosas (Chaplygin, Tronev and Davydova, 2021). As primeiras projeções para a produção total de soja no Brasil para a safra 2022/23 apontam para uma colheita de cerca de 150 milhões de toneladas, enquanto EUA a produção chegará a 123 milhões de toneladas (CONAB, 2022).

Fatores como produtividade e avanço tecnológico têm impulsionado o crescimento da produção de soja em todo mundo. A área destinada ao plantio de soja tem aumentado, principalmente em países emergentes como Brasil, Argentina e Índia. A produtividade também vem aumentando devido ao desenvolvimento e à disponibilidade de tecnologias relacionadas à cultura, embora este seja um parâmetro fortemente influenciado por fatores climáticos (Hirakuri e Lazzarotto, 2014).

No Brasil, o principal estado produtor de soja é o Mato Grosso, seguido por Rio Grande do Sul e Paraná. Na Tabela 1 são apresentados as produções totais, áreas destinadas ao plantio e produtividade desses três estados que apresentam os melhores números nacionais quanto a produção de soja.

Tabela 1. Produção, área plantada e produtividade dos principais estados produtores de soja do Brasil – outubro/21

Maiores Estados Produtores	**Produção (mil kg)**	**Área plantada (mil ha)**	**Produtividade (kg ha^{-1})**
Mato Grosso	35.886	10.297	3.485,09
Rio Grande do Sul	20.788	6.055	3.433,20
Paraná	19.880	5.624	3.534,85

Fonte: (CONAB, 2021)

A soja, uma das culturas mais importantes no cenário agrícola mundial, tem experimentado uma evolução constante em termos de técnicas de cultivo e gerenciamento. Seja em relação ao seu valor nutricional ou à sua crescente produção, especialmente em países como o Brasil, a soja desempenha um papel de destaque em nossas vidas diárias. Ao entendermos melhor suas características e as complexidades de sua produção, conseguimos vislumbrar as inúmeras possibilidades para melhorar ainda mais sua produtividade e eficiência. Uma dessas possibilidades é o uso crescente do sensoriamento remoto e da análise de imagens de satélite na agricultura, ferramentas essas que estão revolucionando o manejo das culturas. Na próxima seção, abordaremos em profundidade como essas tecnologias estão sendo aplicadas na agricultura, especialmente na cultura da soja.

1.2.2. Imageamento orbital na agricultura

A utilização de sensoriamento remoto por meio de plataformas terrestres, aéreas ou orbitas, como sensores manuais, imagens capturadas por VANTs e satélites, tem se mostrando cada vez mais importante para obtenção de dados e, sem dúvidas, é um componente crucial da agricultura de precisão (Anastasiou et al., 2018; Khaki, Pham and Wang, 2021).

A agricultura de precisão, em definição, se trata de uma estratégia de gestão onde dados e informações são coletados, processados e analisados de maneira espacial, temporal e individual (Amaral et al., 2020). A adoção desse tipo de medida nas propriedades rurais vem mudando a maneira como o produtor lida com o campo com uso de técnicas exploratórias e mais assertivas, afim de substituir análises convencionais e subjetivas por ferramentas e soluções sofisticadas de monitoramento de cultura e produtividade com melhor custo benefício (Valente et al., 2020).

Análise de dados baseado em imagens de satélite começaram a ser utilizadas na década de 1970 e, a partir disso, muitos trabalhos começaram a ser publicados nas décadas seguintes fundamentados pela combinação de dados amostrais, frutos de observação de imagens de satélite. A partir da crise de 2008, em que o preço das commodities sofreu com alta volatilidade, aumentando o preço final dos alimentos, a utilização de sensoriamento remoto para monitoramento de culturas se tornou peça chave para o sucesso das lavouras (Gallego, Carfagna and Baruth, 2008).

A utilização de imagens de satélite vem se popularizando devido ao fato de sua capacidade de fornecer informações importantes sobre a cultura trabalhada e possibilitar uma administração mais otimizada dos processos que envolvem toda a produção agrícola, como monitoramento de seca e gerenciamento de irrigação, aplicação de fertilizantes e fungicidas, avaliação de impactos do clima na vegetação, definição de datas de semeadura e colheita, além da análise e previsão de produtividade e qualidade da cultura (Becker et al., 2021; Falco et al., 2021; Khaki, Pham and Wang, 2021; Nolasco et al., 2021).

Os dados obtidos via satélite, vem se tornando cada vez mais acessíveis, fornecendo imagens com altas resoluções, tanto espacial (menor que um metro), como temporal (imagens diárias) e espectral, fornecendo alternativas em relação ao manejo das lavouras (Schwalbert et al., 2020; Khaki, Pham and Wang, 2021; Skakun et al., 2021). Com imagens de satélite é possível determinar e extrair parâmetros

importantes para caracterização de culturas, como VI's (Nolasco et al., 2021).

Para finalizar, é importante destacar que a evolução da tecnologia de sensoriamento orbital está moldando a forma como a agricultura é praticada, tornando-a mais eficiente e sustentável. O acesso a dados de alta qualidade fornecidos pelo imageamento orbital é revolucionário, permitindo o monitoramento contínuo dos cultivos e uma análise mais precisa de sua saúde e produtividade. À medida que avançamos, essas ferramentas orbitais não serão apenas utilizadas para coletar dados, mas também para aplicar práticas agrícolas avançadas, como o uso de Índices de Vegetação. Estes, que serão abordados na próxima seção, trazem um novo nível de compreensão sobre a saúde das culturas e sua gestão, oferecendo um avanço significativo para a agricultura de precisão.

1.2.3. Índices de vegetação na agricultura

Os Índices de Vegetação (VI, do inglês Vegetation Index) são funções matemáticas fundamentadas derivadas da reflectância do dossel das plantas. A extração desses índices descritivos pode ser realizada de maneira precisa e não destrutiva por sensores embarcados em plataformas de sensoriamento remoto, como VANTs e satélites, sendo uma abordagem válida para aquisição de informações relacionadas ao monitoramento e mapeamento da variabilidade de atributos biofísicos de diversas culturas, como soja (Nolasco et al., 2021; Sagan et al., 2021; Shammi and Meng, 2021a), cana-de-açucar (Sumesh, Ninsawat and Som-ard, 2021), milho (Petersen, 2018; Sagan et al., 2021), e amendoim (Santos et al., 2021).

Diversos VI's já foram desenvolvidos baseados nas diferentes necessidades dos produtores, equipamentos disponíveis para uso, bandas espectrais utilizadas e atributos que conseguem ser compreendidos por tais índices. O *Normalized Difference Vegetation Index* (NDVI) é um dos VI's mais utilizados, pois apresenta forte relação com o status da planta e sua produtividade, uma vez que este índice tem relação direta e quase linear com a Radiação Fotossinteticamente Ativa (PAR) (Gamon et al., 1995). Tal índice representa a diferença normalizada entre a faixa do espectro do infravermelho próximo (NIR) e do vermelho refletido pelo dossel da cultura (Rouse et al., 1976).

Huete (1988) desenvolveu o *Soil Adjusted Vegetation Index* (SAVI) com o objetivo de minimizar ruídos da interferência da presença de solo na caracterização do dossel das plantas. O índice leva em conta um fator de adaptação do solo ao NDVI, que varia dependendo da quantidade de vegetação da área analisada, podendo ser utilizado também para caracterização de culturas (Nagy et al., 2021).

Devido ao fato de o NDVI não levar em consideração a influência das condições atmosféricas e do fundo do dossel, o *Enhanced Vegetation Index* (EVI) (Justice et al., 1998) foi proposto baseado nesses conceitos, a fim de gerar um modelo mais aprimorado e com boa sensibilidade para regiões de alta biomassa (Matsushita et al., 2007).

O *Green Normalized Difference Vegetation Index* (GNDVI) (Gitelson, Kaufman and Merzlyak, 1996) é um índice mais sensível à concentração de clorofila que o NDVI, visto que ele utiliza a banda do verde visível e por conta dessa característica também pode ser útil para avaliação de biomassa e produtividade (Yawata et al., 2019).

Já em relação ao *Visual Atmospheric Resistance Index* (VARI), desenvolvido por Kaufman e Tanre (1992), foi um índice originalmente proposto com o objetivo de fazer parte de trabalhos que fizessem uso de imagens de satélite. Esse índice é bastante resistente a influências atmosféricas e mostra uma resposta linear em relação a fração vegetal, desde pequenas faixas até frações moderadas e altas (Schneider, Roberts and Kyriakidis, 2008).

Compreendendo o potencial dos Índices de Vegetação na agricultura, é necessário entender que a sua aplicação não é uma ciência isolada. Ao final desta seção, é evidente que a contribuição desses índices para a agricultura é significativa, permitindo uma compreensão detalhada da saúde vegetal e suas implicações para a produtividade. Entretanto, esses índices são apenas uma parte de um sistema mais amplo que é alimentado por uma série de outros processos. Um deles, e talvez um dos mais cruciais, é o processamento de imagens. Este é o assunto da nossa próxima seção, onde abordaremos como as técnicas de processamento de imagens, incluindo as avançadas como Inteligência Artificial, Machine Learning e Deep Learning, estão integradas à agricultura de precisão e como elas amplificam a eficácia dos Índices de Vegetação.

1.2.4. Processamento de imagens

O desenvolvimento de novos métodos de processamento de imagens, como Inteligência Artificial (IA), *Machine Learning* (ML) e *Deep Learning* (DL) vem avançando numa velocidade cada vez maior e, na mesma proporção, esses procedimentos vêm sendo cada vez mais integrados a agricultura de precisão (Sabzi, Abbaspour-Gilandeh and Javadikia, 2017; Zhuang, Wang and Jiang, 2020). Fatores como o aumento na disponibilidade de equipamentos e softwares, a diminuição do preço, melhora na resolução e aumento da capacidade de processamento dos dispositivos que captam as imagens a serem utilizadas são fatores importantes que vem colaborando fortemente com o desenvolvimento das técnicas de processamento (Mahajan, Das and Sardana, 2015).

Esse tipo de método vem se tornando cada vez mais relevante e útil em pesquisas na área de ciências agrárias e estudos de plantas, uma vez que permite a medida mais facilitada de variáveis cuja aquisição pode ser complexa e consumir tempo de maneira desnecessária, além de remover a subjetividade de uma análise humana (Pérez-Rodríguez e Gómez-García, 2019). A extração de características do dossel e da vegetação verde da cultura de interesse é parte primordial e um dos primeiros passos a serem dados no processamento de imagens para diferentes finalidades dentro da agricultura de precisão, como monitoramento de crescimento de cultura e produtividade, quantidade de biomassa, contagem de plantas, além de outros fatores (Hernández-Hernández et al., 2017; Zhuang et al., 2020).

O processamento inicial de imagens trata-se de uma técnica onde ocorre a utilização de imagens brutas como entradas para extração de informações e dados importantes, que serão usados posteriormente para identificação, reconhecimento e classificação de padrões (Chandel et al., 2020). Geralmente, o procedimento de processamento da imagem obedece as seguintes etapas (Figura 1).

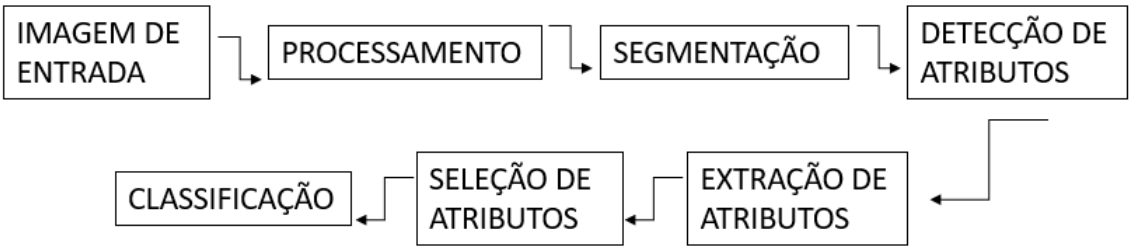

Figura 1. Etapas básicas de processamento de imagens

Inicialmente, ocorre o processamento dos dados brutos, principalmente para calibração, correção e remoção de ruídos (“granulação” da imagem que diminui sua qualidade). Na sequência, converte-se a imagem com três componentes (bandas vermelho, verde e azul) para uma imagem em escala de cinza. Após esse processo, ocorre a detecção, extração e seleção das caraterísticas desejáveis para obtenção do máximo de informações sobre a espécie vegetal de interesse. Esses processos facilitam a classificação e diferenciação entre espécies intrusas e infestantes e a cultura cultivada (Sohail et al., 2021).

O processamento de imagens tradicional já se mostra como uma técnica muito útil e estabelecida, porém a variação de certos fatores na imagem como oclusão, luminosidade, sombreamento e outras dificuldades na captação da imagem, acabam por aumentar a dificuldade no processamento, visto que uma abordagem padrão pode não ser suficiente, abrindo portas para utilização e desenvolvimento de técnicas mais robustas, como IA, ML e DL (Singh et al., 2018).

Para finalizar a seção sobre processamento de imagens, podemos afirmar que sua aplicação na agricultura de precisão trouxe enormes benefícios e continua a evoluir, impulsionada por avanços em tecnologias como Inteligência Artificial, Machine Learning e Deep Learning. As técnicas de processamento de imagens permitiram uma análise mais detalhada e objetiva das culturas, contribuindo para uma gestão agrícola mais eficiente. Contudo, as informações geradas por essas técnicas necessitam de uma interpretação e análise mais profundas. A etapa seguinte, o uso de algoritmos para a predição de produtividade, é o caminho natural que segue após a coleta e processamento de imagens. Vamos discutir como esses algoritmos são aplicados na agricultura e como eles transformam os dados de imagens em informações úteis para a tomada de decisão na próxima seção.

1.2.5. Uso de algoritmos para a predição de produtividade

O desenvolvimento de novas tecnologias e estudos voltados para a aquisição de informações dos processos agrícolas já são uma realidade dentro das propriedades rurais. Isso permite a criação de bases de dados robustas em relação aos parâmetros de produção de uma área agrícola. O uso de diferentes ferramentas e algoritmos que através de uma análise muitas vezes simples desses bancos de dados criados, fornecem informações de suma importância para o produtor no que tange o controle e gerenciamento de uma área de produção (Li, Ding and Xu, 2022; Nguyen, Robinson and Galpern, 2022).

Uma opção que já vem sendo bastante trabalhada e é bastante aplicável no campo é o uso de regressões lineares simples, onde se mede a força de correlação entre a variável predita, no caso a produtividade, e uma varável preditora, que podem ser os valores de dados espectrais de VI's, como o NDVI (Andrade et al., 2022). Porém, em alguns casos, a relação entre essas variáveis não obedece um comportamento linear, sendo necessário o uso de outros tipos de regressão (Burdett e Wellen, 2022), tornando-se cada vez mais recorrente o uso da IA na construção de modelos totalmente adequados ao *dataset* utilizado como *input* (Wei et al., 2020; Bojanowski et al., 2022; Rosado et al., 2022).

Dentro da IA existe um subconjunto de técnicas denominado ML que tem a capacidade de detectar efeitos de variáveis colineares e identificar relações não lineares entre parâmetros, o que naturalmente aumenta o desempenho de modelos criados a partir dessa técnica em relação a regressões lineares simples (da Silva et al., 2020). Um exemplo de técnica para predição de atributos, com intuito de gerar melhor resposta em comparação a regressões lineares é o *Random Forest* – Florestas Aleatórias. O *Random Forest* (RF) é um algoritmo de ML que pode ser utilizado tanto para classificação como para regressão (Breiman, 2001). Sendo um algoritmo de conjunto, constrói várias árvores de decisão independentes para a montagem de modelos. São selecionadas de maneira aleatória um determinado número de variáveis de entrada de cada nó e os subconjuntos são utilizados para calcular a melhor saída do modelo. O modelo de previsão com o máximo de votos torna-se o resultado final (Alabi et al., 2022).

Como subconjunto do ML, ainda podemos citar o DL, onde se encaixam as classes de Redes Neurais Artificiais (ANN's) (Souza, 2021). Devido às características de reconhecimento e aprendizado, as ANN's vem sendo usadas fortemente na agricultura, visto que estudos indicam que esse tipo de técnica apresenta melhor desempenho preditivo em relação a técnicas de regressão linear (Kumar, 2020). As ANN's diferenciam-se pela sua estrutura ou arquitetura, ou seja, pelas maneiras com que as conexões entre neurônios são realizadas. Dentre as arquiteturas mais utilizadas de ANN's podemos citar as redes Perceptron Multicamadas (MLP) e Redes de Ativação de Base Radial (RBF). A rede MLP é composta por uma camada de entrada que recebe o sinal, várias camadas ocultas, que realizam cálculos intermediários que auxiliam a rede a encontrar os valores finais e uma camada de saída que toma a decisão ou prevê um resultado a partir das entradas (Ghorbani et al., 2016). A rede RBF possui uma estrutura mais simples, com apenas uma camada oculta (Yoosefzadeh-Najafabadi et al., 2021).

Abordar os mais diversos tipos de abordagens preditivas se faz interessante, uma vez que, caso um modelo não opere bem para uma determinada área e/ou situação, com estudos variados em torno dessa questão, se torna possível encontrar um melhor caminho para atingir o objetivo de predizer a produtividade de uma lavoura com o menor erro possível. Essa abordagem será a temática central do trabalho.

1.3. Referências

Alabi TR, Abebe AT, Chigeza G and Fowobaje KR (2022) Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. **Remote Sensing Applications: Society and Environment** 27:100782.

Amaral LR do, Zerbato C, de Freitas RG, Júnior MRB and da Silva Simões IOP (2020) UAV applications in Agriculture 4.0. **Revista Ciencia Agronomica** 51:1–15.

Anastasiou E, Balafoutis A, Darra N, Psiroukis V, Biniari A, Xanthopoulos G and Fountas S (2018) Satellite and proximal sensing to estimate the yield and quality of table grapes. **Agriculture (Switzerland)** 8.

Andrade TG, Junior ASDA, Souza MO, Lopes JWB and Vieira PFDMJ (2022) Soybean Yield Prediction Using Remote Sensing in Southwestern Piauí State, Brazil. **Revista Caatinga** 35:105–116.

Becker WR, Silva LCDA, Richetti J, Ló TB and Johann JA (2021) Harvest date forecast for soybeans from maximum vegetative development using satellite images. **International Journal of Remote Sensing** 42:1121–1138.

Bojanowski JS, Sikora S, Musiał JP, Woźniak E, Dąbrowska-Zielińska K, Slesiński P, Milewski T and Łączyński A (2022) Integration of Sentinel-3 and MODIS Vegetation Indices with ERA-5 Agro-Meteorological Indicators for Operational Crop Yield Forecasting. **Remote Sensing** 14:1238.

Burdett H and Wellen C (2022) Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. **Precision Agriculture** [Preprint].

Chandel NS, Chakraborty SK, Rajwade YA, Dubey K, Tiwari MK and Jat D (2020) Identifying crop water stress using deep learning models. **Neural Computing and Applications** 4.

Chaplygin ME, Tronev S V. and Davydova SA (2021) Soybean harvesting using current dedicated headers and adapters. **IOP Conference Series: Earth and Environmental Science** 659.

CONAB (2021) *Portal de Informações Agropecuárias*. Available at: https://portaldeinformacoes.conab.gov.br/produtos-360.html.

CONAB (2022) *Safra 2022/23: Produção de grãos pode chegar a 308 milhões de t impulsionada pela boa rentabilidade de milho, soja e algodão*. Available at: https://www.conab.gov.br/ultimas-noticias/4731-safra-2022-23-producao-de-graos-pode-chegar-a-308-milhoes-de-toneladas-impulsionada-pela-boa-rentabilidade-de-milho-soja-e-algodao.

EMBRAPA (2021a) *História da soja*. Available at: https://www.embrapa.br/en/soja/cultivos/soja1/historia.

EMBRAPA (2021b) *Soja em números (safra 2020/21)*. Available at: https://www.embrapa.br/en/soja/cultivos/soja1/dados-economicos.

Falco N, Wainwright HM, Dafflon B, Ulrich C, Soom F, Peterson JE, Brown JB, Schaettle KB, Williamson M, Cothren JD, Ham RG, McEntire JA and Hubbard SS (2021) Influence of soil heterogeneity on soybean plant development and crop yield evaluated using time-series of UAV and ground-based geophysical imagery. **Scientific Reports** 11:1–17.

Gallego J, Carfagna E and Baruth B (2008) Accuracy , objectivity and efficiency of remote sensing for agricultural statistics.

Gamon JA, Field CB, Goulden ML, Griffin KL, Hartley E, Joel G, Peñuelas J and Valentini R (1995) NDVI canopy structure photosynthesis. **Ecological Applications** 5:28–41.

Gaso D V., Berger AG and Ciganda VS (2019) Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. **Computers and Electronics in Agriculture** 159:75–83.

Ghorbani MA, Zadeh HA, Isazadeh M and Terzi O (2016) A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. **Environmental Earth Sciences** 75:1–14.

Gitelson AA, Kaufman YJ and Merzlyak MN (1996) Use of a green channel in remote sensing of global vegetation from EOS- MODIS. **Remote Sensing of Environment** 58:289–298.

Hernández-Hernández JL, Ruiz-Hernández J, García-Mateos G, González-Esquiva JM, Ruiz-Canales A and Molina-Martínez JM (2017) A new portable application for automatic segmentation of plants in agriculture. **Agricultural Water Management** 183:146–157.

Hirakuri M and lazzarotto J (2014) O agronegócio da soja nos contextos mundial e brasileiro.

Huete A (1988) A Soil-Adjusted Vegetation Index (SAVI). **Remote Sensing of Environment** 25:295–309.

Justice CO, Vermote E, Townshend JRG, Defries R, Roy DP, Hall DK, Salomonson V V., Privette JL, Riggs G, Strahler A, Lucht W, Myneni RB, Knyazikhin Y, Running SW, Nemani RR, Wan Z, Huete AR, Van Leeuwen W, et al. (1998) The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. **IEEE Transactions on Geoscience and Remote Sensing** 36:1228–1249.

Kaufman Y and Tanre. D (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. **IEEE Transactions on Geoscience and Remote Sensing** 30:260–271.

Khaki S, Pham H and Wang L (2021) Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. **Scientific Reports** 11:1–14.

Kumar S (2020) Comparison of linear regression and artificial neural network technique for prediction of a soybean biodiesel yield. **Energy Sources, Part A: Recovery, Utilization and Environmental Effects** 42:1425–1435.

Li Z, Ding L and Xu D (2022) Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. **Science of the Total Environment** 815:152880.

Mahajan S, Das A and Sardana HK (2015) Image acquisition techniques for assessment of legume quality. **Trends in Food Science and Technology** 42:116–133.

Matsushita B, Yang W, Chen J, Onda Y and Qiu G (2007) Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to topographic effects: A case study in high-density cypress forest. **Sensors** 7:2636–2651.

Nagy A, Szabó A, Adeniyi OD and Tamás J (2021) Wheat yield forecasting for the tisza river catchment using landsat 8 ndvi and savi time series and reported crop statistics. **Agronomy** 11.

Nepomuceno AL, Farias JRB and Neumaier N (2021) *Características da soja*. Available at: https://www.agencia.cnptia.embrapa.br/gestor/soja/arvore/CONTAG01_24_271020069131.html

Nguyen LH, Robinson S and Galpern P (2022) Medium-resolution multispectral satellite imagery in precision agriculture: mapping precision canola (Brassica napus L.) yield using Sentinel-2 time series. **Precision Agriculture** 23:1051–1071.

Nolasco M, Ovando G, Sayago S, Magario I and Bocco M (2021) Estimating soybean yield using time series of anomalies in vegetation indices from MODIS. **International Journal of Remote Sensing** 42:405–421.

Pérez-Rodríguez F and Gómez-García E (2019) Codelplant: Regression-based processing of RGB images for colour models in plant image segmentation. **Computers and Electronics in Agriculture** 163:104880.

Petersen LK (2018) Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. **Remote Sensing** 10:1–31.

Rosado RDS, Penso GA, Serafini GAD, Magalhães dos Santos CE, Picoli EA de T, Cruz CD, Barreto CAV, Nascimento M and Cecon PR (2022) Artificial neural network as an alternative for peach fruit mass prediction by non-destructive method. **Scientia Horticulturae** 299.

Rouse JW, Haas Jr. RH, Schell JA and Deering DW (1976) Monitoring Vegetation Systems in the Great Plains with Erts. **NASA Spec. Publ.** 74:309–317.

Sabzi S, Abbaspour-Gilandeh Y and Javadikia H (2017) Machine vision system for the automatic segmentation of plants under different lighting conditions. **Biosystems Engineering** 161:157–173.

Sagan V, Maimaitijiang M, Bhadra S, Maimaitiyiming M, Brown DR, Sidike P and Fritschi FB (2021) Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. **ISPRS Journal of Photogrammetry and Remote Sensing** 174:265–281.

dos Santos AF, Corrêa LN, Lacerda LN, Tedesco-Oliveira D, Pilon C, Vellidis G and da Silva RP (2021) High-resolution satellite image to predict peanut maturity variability in commercial fields. **Precision Agriculture** 22:1464–1478.

Schneider P, Roberts DA and Kyriakidis PC (2008) A VARI-based relative greenness from MODIS data for computing the Fire Potential Index. **Remote Sensing of Environment** 112:1151–1167

Schwalbert RA, Amado T, Corassa G, Pott LP, Prasad PVV and Ciampitti IA (2020) Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. **Agricultural and Forest Meteorology** 284:107886.

Shammi SA and Meng Q (2021) Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. **Ecological Indicators** 121:107124.

da Silva EE, Rojo Baio FH, Ribeiro Teodoro LP, da Silva Junior CA, Borges RS and Teodoro PE (2020) UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. **Remote Sensing Applications: Society and Environment** 18.

Singh AK, Ganapathysubramanian B, Sarkar S and Singh A (2018) Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. **Trends in Plant Science** 23:883–898.

Skakun S, Kalecinski NI, Brown MGL, Johnson DM, Vermote EF, Roger JC and Franch B (2021) Assessing within-field corn and soybean yield variability from worldview-3, planet, sentinel-2, and landsat 8 satellite imagery. **Remote Sensing** 13:1–18.

Sohail R, Nawaz Q, Hamid I, Mushhad S, Gilani M, Mateen A and Chauhdary JN (2021) AN ANALYSIS ON MACHINE VISION AND IMAGE PROCESSING TECHNIQUES FOR WEED DETECTION IN AGRICULTURAL CROPS 58:187–204.

Souza JBC (2021) *TÉCNICAS DE AGRICULTURA DIGITAL PARA PREDIÇÃO DA MATURAÇÃO DO AMENDOIM*.

Sumesh KC, Ninsawat S and 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.

Tamagno S, Balboa GR, Assefa Y, Kovács P, Casteel SN, Salvagiotti F, García FO, Stewart WM and Ciampitti IA (2017) Nutrient partitioning and stoichiometry in soybean: A synthesis-analysis. **Field Crops Research** 200:18–27.

Valente J, Sari B, Kooistra L, Kramer H and Mücher S (2020) Automated crop plant counting from very high-resolution aerial imagery. **Precision Agriculture** 21:1366–1384.

Wei MCF, Maldaner LF, Ottoni PMN and Molin JP (2020) Carrot Yield Mapping: A Precision Agriculture Approach Based on Machine Learning. **Machine Learning** 1:229–241.

Xu J, Meng J and Quackenbush LJ (2019) Use of remote sensing to predict the optimal harvest date of corn. **Field Crops Research** 236:1–13.

Yawata K, Yamamoto T, Hashimoto N and Ishida R (2019) Mixed model estimation of rice yield based on NDVI and GNDVI using a satellite image. **Proceedings of SPIE - The International Society for Optical Engineering** 11149.

Yoosefzadeh-Najafabadi M, Earl HJ, Tulpan D, Sulik J and Eskandari M (2021) Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean. **Frontiers in Plant Science** 11:1–14.
Zhou Jing, Zhou Jianfeng, Ye H, Ali ML, Chen P and Nguyen HT (2021) Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network. **Biosystems Engineering** 204:90–103.

Zhuang S, Wang P, Jiang B, Li M and Gong Z (2017) Early detection of water stress in maize based on digital images. **Computers and Electronics in Agriculture** 140:461–468.,

Zhuang S, Wang P and Jiang B (2020) Vegetation extraction in the field using multi-level features. **Biosystems Engineering** 197:352–366.

CAPÍTULO 2 – Abordagem para a predição da soja por índices de vegetação e modelos de aprendizado de máquina

2.1. INTRODUÇÃO

A soja (*Glycine max*) é uma das culturas mais importantes dentro do cenário global da agroindústria por conta do seu alto teor proteico e da composição geral do grão, atingindo até 43% de proteína, 18 a 20% de carboidratos e até 25% de óleo, que é destinado, além da alimentação, humana e animal, para produção de biocombustíveis (Chaplygin, Tronev and Davydova, 2021). Atualmente, Brasil e EUA concentram cerca de 37 e 30% da produção mundial desse grão, respectivamente (EMBRAPA, 2021). As primeiras projeções para a produção total de soja no país para a safra 2022/23 apontam para uma colheita de 150,36 milhões de toneladas (CONAB, 2022).

Por conta desse cenário, a informação antecipada da produtividade da soja é significativa, pois auxilia os produtores na tomada de decisões quanto ao manejo da cultura, como prever e definir as melhores épocas de semeadura ou colheita. Contribui também para o planejamento das frotas de máquinas utilizadas, além da escolha de cultivares com ciclos de desenvolvimento adequados ao ambiente (Xu, Meng and Quackenbush, 2019). Esses benefícios tornam a aquisição dessas informações extremamente importante para proporcionar o desenvolvimento de uma agricultura mais sustentável e promover maior segurança alimentar.

Buscando soluções para a predição de produtividade de soja na literatura, diversos estudos utilizam do sensoriamento remoto para tal função. Da Silva et al. (2020) utilizaram diversos índices de índices de vegetação (VI's) para observar a correlação dos mesmos com a produtividade de grãos de soja e atestaram que os VI's que mais se destacaram entre os analisados foram *Normalized Difference Vegetation Index* (NDVI) e *Soil Adjusted Vegetation Index* (SAVI). Eugenio et al. (2020) também testaram diferentes VI's para estimar a produtividade da cultura da soja e analisando parâmetros de correlação e erro, chegaram em modelos definidos como excelentes para predizer a produtividade da soja com precisão, reforçando a capacidade de combinação entre sensoriamento remoto e métodos de *Machine Learning* (ML) em aplicações destinadas à agricultura de precisão.

Apesar do desenvolvimento desses estudos, alguns gargalos permeiam a aplicação dessa técnica de gestão da produção por imagens aéreas a nível de campo, muitas vezes dificultando sua utilização. Dentre algumas limitações, quando a coleta de imagens é feita por meio de Veículo Aéreo Não Tripulado (VANT) pode-se citar questões que envolvem a bateria, ou seja, a autonomia de voo, o fato de os equipamentos e softwares utilizados terem custo considerável e a necessidade de um mínimo de experiência básica no manuseio e processamento das imagens, além das condições climáticas, uma vez que no Brasil, o cultivo da soja, na maioria dasvezes acontece em períodos chuvosos, o que pode diminuir a qualidade do resultado ou até mesmo inviabilizar a operação.

Para superar estas limitações, a aquisição de imagens de satélite através de plataformas em nuvem, como o Explorer da plataforma de satélites PlanetScope, tem se mostrado uma alternativa promissora. Vários estudos têm demonstrado consistentemente o potencial da utilização do imagemamento orbital, juntamente com a análise exploratória da utilização de técnicas de ML a nível global (Kuwata and Shibasaki, 2016; Sun et al., 2019; Edreira et al., 2020; Schwalbert et al., 2020), o que demonstra que a comunidade científica tem evidências de que é uma abordagem relevante para a predição de produtividade de culturas.

Apesar do desenvolvimento desses estudos, ainda existem muitas dúvidas no que diz respeito ao tema, principalmente em relação a quais VI’s podem ser úteis à estimativa de produtividade, a partir de qual momento do desenvolvimento da soja essa estimativa pode ocorrer e quais algoritmos mais adequados para tal função (Shammi e Meng, 2021; Zhou et al., 2021). Além disso, a aplicação e estudo de modelos de ML diversificados, que levam em consideração diferentes VI’s como parâmetro preditor e diferentes datas de coleta de imagem são essenciais para que seja possível identificar métodos que geram estimativas de produtividade com respostas mais precisas e antecipadas, em diversos estádios de desenvolvimento da cultura.

O objetivo deste trabalho é analisar os fatores que impactam a precisão dos modelos de previsão de produtividade da soja. Buscamos esclarecer a influência da data de coleta das imagens de satélite, que está intimamente ligada ao estágio de desenvolvimento da cultura, na eficácia desses modelos. Ademais, vamos avaliar a

eficiência dos algoritmos de previsão e a contribuição dos Índices de Vegetação (VI's) em suas composições. Pretendemos, assim, esclarecer questões sobre a aplicabilidade desta ferramenta e demonstrar como as imagens de satélite podem se mostrar uma alternativa superior aos Veículos Aéreos Não Tripulados (VANT's) na gestão da produção de soja.

2.2. MATERIAL E MÉTODOS

2.2.1. Área de estudo

Os experimentos foram instalados em dois campos de soja da Fazenda de Ensino, Pesquisa e Extensão – FEPE da Faculdade de Ciências Agrárias e Veterinárias (FCAV) – UNESP, de Jaboticabal, Estado de São Paulo. Os experimentos estiveram localizados próximos às coordenadas 21°14'52.47" de latitude sul e 48°16'50.52" de longitude oeste, com 564 m de altitude média (Figura 2). O clima da região é subtropical de inverno relativamente seco (Aloisi e Demattê, 1974), classificado como Cwa pelo Sistema Internacional de Classificação de Köppen. A temperatura média anual é 22°C e a pluviosidade média anual de 1424 mm (Estação Agroclimatológica, UNESP-Jaboticabal). O tipo de solo do local de estudo é Latossolo Vermelho Distrófico, com textura argilosa típica (Rossi, 2017). Como os dados das duas safras foram utilizados, a cultura foi submetida às diferentes condições climáticas presentes em cada um desses anos. A Figura 3 ilustra as condições climáticas durante os períodos de desenvolvimento de cada safra.

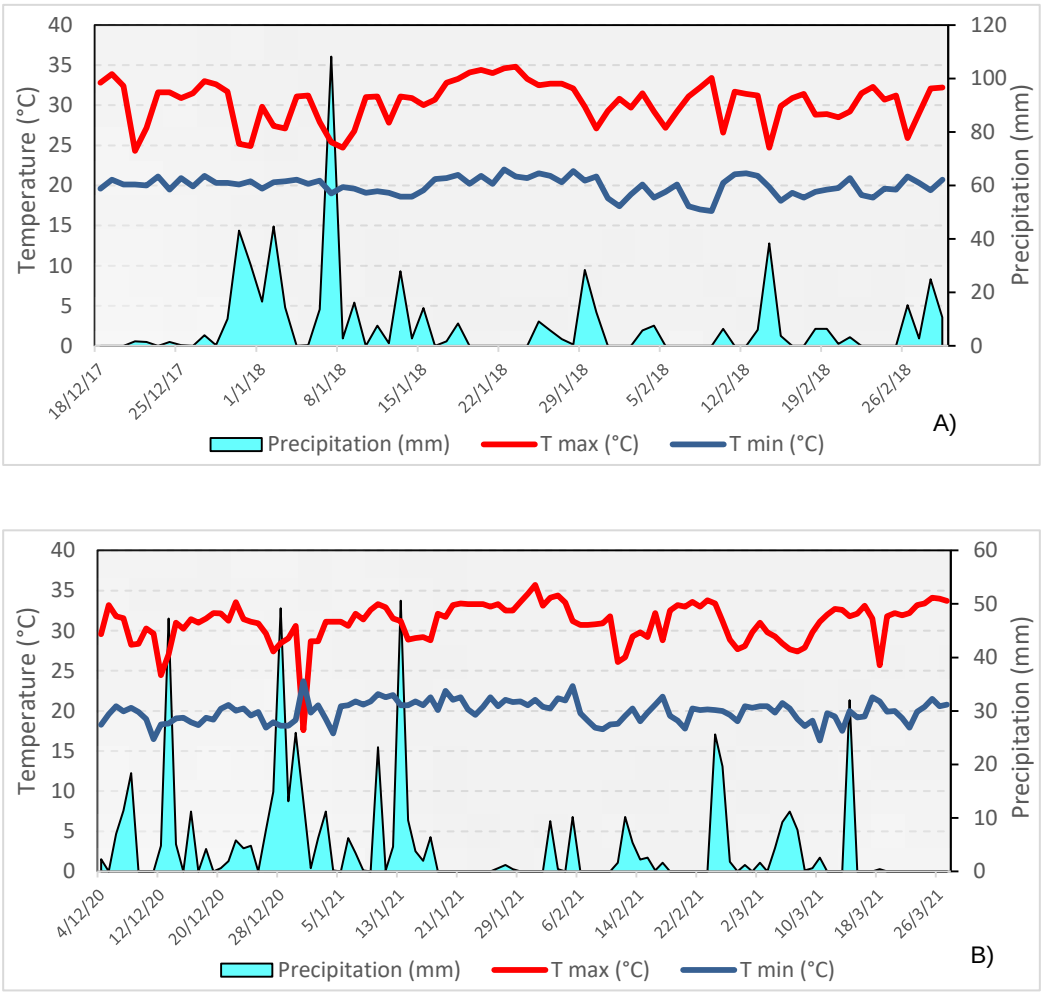

Figura 2. Condições climáticas das colheitas A) 2017/18 e B) 2020/21

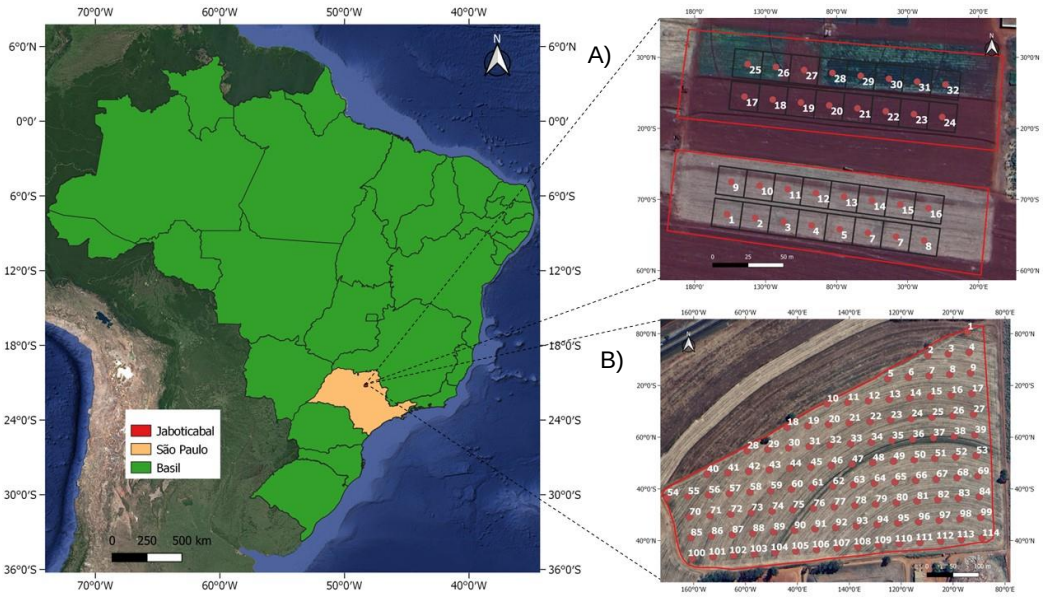

Figura 3. Representação da A) primeira área experimental e B) segunda área experimental

A primeira área experimental (Figura 3 - A) foi de pouco mais de 1,0 ha, com

as cultivares de soja 64H0114IPRO (Pirapó) e 66H0116IPRO (Tererê), sendo a primeira com grupo de maturidade 6.6 e a segunda 6.4, e ambas com hábito de crescimento indeterminado. Esta área foi semeada em 04/12/2020 e colhida em 27/03/2021, totalizando um tempo de ciclo de 113 dias. A semeadura foi realizada no espaçamento de 0,45 m entre linhas, com uma taxa de semeadura de 14 sementes m^{-1}. A área cultivada foi dividida em quatro parcelas principais, sendo duas dedicadas a cada cultivar. No total, foi analisado a massa de 32 pontos de 20x20 m, sendo 16 para cada cultivar.

A segunda área experimental (Figura 3 - B) possuía cerca de 20 ha e foi plantada em 24/10/2017 e colhida em 01/03/2018, com ciclo de 128 dias, onde foi cultivada a variedade de soja AS3797IPRO. Esta cultivar possui maturidade relativa 7,3, com crescimento indeterminado e porte ereto de haste única. A semeadura foi realizada com uma taxa de 19 sementes m^{-1} e espaçamento de 0,45m. Nesta área foram recolhidos 114 pontos para avaliação. As zonas numeradas assinaladas em cada área na Figura 1 representam os pontos de amostragem onde foram recolhidas as amostras de produção. A distribuição dos pontos foi baseada na forma e tamanho de cada área experimental.

2.2.2. Avaliações de campo

Os pontos de amostragem da soja foram coletados manualmente em cada um dos campos avaliados. Para padronizar a coleta de dados nas lavouras, determinou-se em cada ponto de amostragem três linhas paralelas de cultivo com comprimento de 2,0 m, e como o espaçamento entre linhas foi de 0,45 m, isso acabou gerando uma área coletada de 1,8 m^2. Após a colheita das plantas, os grãos foram separados mecanicamente por um trilhador, pesados e secos em estufa, corrigindo sua umidade para próximo de 13%. Os valores corrigidos, junetamente com a área de coleta foram utilizados para converter a unidade de medida em kg ha^{-1}.

2.2.3. Aquisição das imagens PlanetScope

A plataforma escolhida para a aquisição das imagens remotas foi o satélite PlanetScope. Os fatores que levaram à escolha da utilização deste satélite foram algumas das suas principais características, como a captação de imagens diárias e a possibilidade de trabalhar com imagens de quatro bandas espectrais: vermelho, verde, azul e infravermelho próximo, fornecendo comprimentos de onda no espectro visível e não visível.

A constelação de satélites PlanetScope fornece imagens de alta qualidade, mesmo sob condições atmosféricas específicas, minimizando possíveis ruídos nas respostas espectrais, temporais e espaciais. O produto fornecido pelo PlanetScope é uma imagem GeoTIFF de 16 bits com valores de reflectância escalados para 10.000, com correcção de imagem baseada na reflectância do topo da atmosfera (TOA),

As imagens foram recolhidas em quatro fases diferentes de crescimento da cultura - R4 a R7 - para análise comparativa dos resultados, a fim de proporcionar melhor compreensão da forma como as imagens de satélite podem ser utilizadas para efeitos de estimativa de produtividade e em que estádios fenológicos faz sentido utilizar este tipo de metodologia. Ao trabalhar com a classificação por estádios de crescimento da cultura, torna-se possível fornecer informações relevantes ao produtor, de forma a tornar a aplicação do método o mais antecipada possível, permitindo melhor gestão da área cultivada.

2.2.4. Índices de vegetação

Neste estudo foram utilizados cinco VI's (Tabela 2) como *inputs* para criação de modelos lineares e não lineares de predição de produtividade da soja. As diferentes bandas que compõe o espectro, como vermelho, azul, verde e infravermelho próximo, e que poderiam ser usadas para compor os algoritmos foram testadas, mas não tiveram o desempenho esperado, de modo que apenas os VI's compostos por tais bandas foram selecionados para o estudo. Todos esses índices foram escolhidos por terem aplicações promissoras na previsão de parâmetros agronômicos em diversas culturas, como soja (Nolasco et al., 2021; Sagan et al., 2021; Shammi and Meng, 2021b), cana-de-açúcar (Sumesh, Ninsawat

and Som-ard, 2021), milho (Petersen, 2018) e amendoim (Dos Santos, Dos Santos and Rolim, 2021).

Tabela 2. Índices de vegetação usados no estudo.

IV	Abreviação	Equação	Referência
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - RED}{NIR + RED}$	(Rouse et al., 1976)
Soil Adjusted Vegetation Index	SAVI	$\frac{(1.5) \cdot (NIR - RED)}{0.5 + NIR + RED}$	(Huete, 1988)
Enhanced Vegetation Index	EVI	$\frac{2.5 \cdot (\text{NIR} - \text{RED})}{\text{NIR} + 6.0 \cdot \text{RED} - 7.5 \cdot \text{BLUE} + 1}$	(Justice et al., 1998)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR - GREEN}{NIR + GREEN}$	(Gitelson e Merzlyak, 1996)
Visual Atmospheric Resistance Index	VARI	$\frac{GREEN - RED}{GREEN + RED - BLUE}$	(Kaufman e Tanre., 1992)

2.2.5. Descrição dos modelos de predição de produtividade

Para construir os modelos de estimativa de produtividade foram utilizados, além de regressões lineares simples para efeitos de comparação, o RF e dois modelos ANN: Radial Basis Function (RBF) e Multilayer Perceptron (MLP) (Tabela 3).

Tabela 3. Algoritmos utilizados na predição de produtividade da soja.

Algoritmos	Referência
Linear Regressions	(Islam et al., 2021; Matese and Di Gennaro, 2021)
Random Forest	(Breiman, 2001; Alabi et al., 2022)
Multilayer Perceptron (MLP)	(Santos et al., 2022; Souza et al., 2022)
Radial Basis Function (RBF)	(Haykin, 1994; Bishop, 1995; Oliveira et al., 2021)

Foram utilizados diferentes algoritmos com a intenção de explorar a complexidade de cada um e analisar até que ponto essa complexidade conduz realmente ao resultado mais confiável.

2.2.6. Treinamento, validação e desempenho

O conjunto de dados foi construído a partir do processamento de imagens capturadas em diferentes estádios fenológicos da cultura da soja pela constelação de satélites PlaneScope. O software QGIS foi utilizado para o processamento dessas imagens, onde foi possível calcular os VI's utilizados para cada ponto amostral nos respectivos estádios fenológicos analisados para ambas as safras. Com o *dataset* completo com os valores dos VI's e a produtividade de cada ponto amostral, os modelos puderam ser construídos.

A base de dados foi dividida em 80% para treino e 20% para teste dos modelos. Os procedimentos de treinamento e teste dos modelos baseados em ANN foram implementados utilizando o pacote de software de análise estatística de dados Neural Networks (Statistica 7.0, Statsoft Inc., Tulsa, OK, EUA), enquanto os procedimentos para regressão linear simples e RF foram implementados utilizando o ambiente de linguagem R (RStudio). Foram usados dois softwares para algoritmos mais simples, como foi o caso da regressão linear e do RF, a implementação com a linguagem R foi mais simplificada. Para as ANN's, um software mais completo acabou por melhorar a experiência no desenvolvimento do algoritmo.

Os valores preditos de produtividade foram avaliados pela acurácia por meio de métricas de erro como *Mean Absolute Error* (MAE), o *Root Mean Squared Error* (RMSE) e *Mean Absolute Percentage Error* (MAPE) (Equações (1), (2) e (3), respectivamente), visando expressão dos valores de erro absoluto, erro médio do modelo preditivo em relação aos dados originais e da acurácia do erro em percentagem,

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |Yi - \hat{Y}\, i|, \qquad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^{N}(Yi - \hat{Y}i)^2}, \quad (2)$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{(Yi-\hat{Y}i)}{Yi}\right|, \quad (3)$$

Em que: N é o número total de observações e Yi e $\hat{Y}i$ são as produtividades observadas e estimadas, respectivamente, para o ponto i. Para seleção dos melhores modelos para predizer a produtividade da soja, procurou-se modelos que apresentassem os menores valores de MAE, MAPE e RMSE.

2.3. RESULTADOS

2.3.1. Análise descritiva

Os resultados obtidos evidenciam diferenças marcantes na produtividade da soja ao longo de duas safras agrícolas distintas, utilizando-se dados coletados de três cultivares distintas. Em 2017, a produtividade média alcançou pouco mais de 2000 kg ha-1, enquanto em 2020 observou-se um aumento expressivo nessa métrica, com valores superiores a 3000 kg ha-1.

Para visualização e interpretação mais efetiva dos dados, foi construído um gráfico de box-plot (Figura 4) para apresentar a distribuição dos valores de produtividade por cultivar. Destaca-se que as cultivares 64H0114 IPRO e 66H0116 IPRO foram associadas à safra de 2020/21, enquanto a cultivar AS3797IPRO fez parte da safra 2017/18.

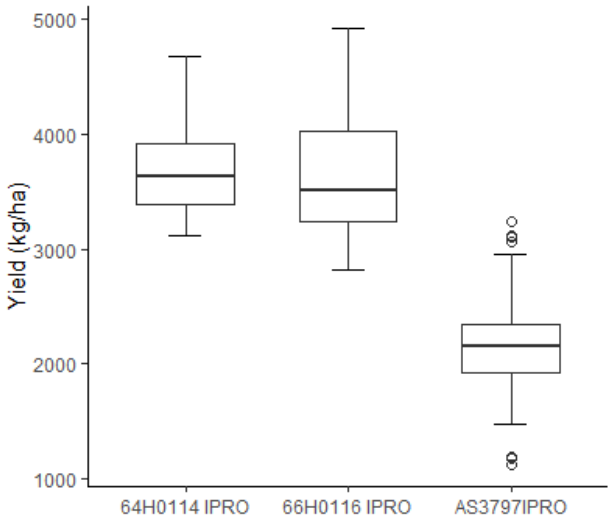

Figura 4. Análise descritiva de produtividade por cultivar

2.3.2 Algoritmos preditivos de produtividade de soja

Os modelos preditivos de produtividade da soja foram estruturados para cada estádio fenológico analisado (R4 a R7), utilizando o índice de vegetação (VI) mais adequado para cada estágio. A avaliação da precisão dos algoritmos foi realizada por meio dos parâmetros de acurácia MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) e RMSE (Root Mean Square Error). A análise inicial envolveu a implementação de algoritmos de regressão linear (Tabela 4).

Tabela 4. Resultados das melhores estimativas obtidas pelos algoritmos de regressão linear em cada estádio fenológico.

Phenological Stage	VI	MAE (kg ha^{-1})	MAPE	RMSE (kg ha^{-1})
R4	GNDVI	496.67	0.22	663.17
R5	SAVI	471.38	0.22	637.01
R6	GNDVI	679.64	0.3	852.1
R7	GNDVI	475.3	0.21	623.81

O modelo mais eficiente foi obtido no estágio fenológico R7 (Figura 5), o qual apresentou um coeficiente de determinação (R^2) próximo de 0,60, indicando um ajuste moderado. Ainda assim, os valores de erro associados (MAPE ≤ 0,30) em todos os modelos indicam que a regressão linear poderia prever a produtividade a partir do estágio R4 com razoável precisão. Notavelmente, o GNDVI foi o índice de vegetação que apresentou maior relevância nas análises nos estágios R4, R6 e R7, enquanto o SAVI prevaleceu na análise do estágio R5.

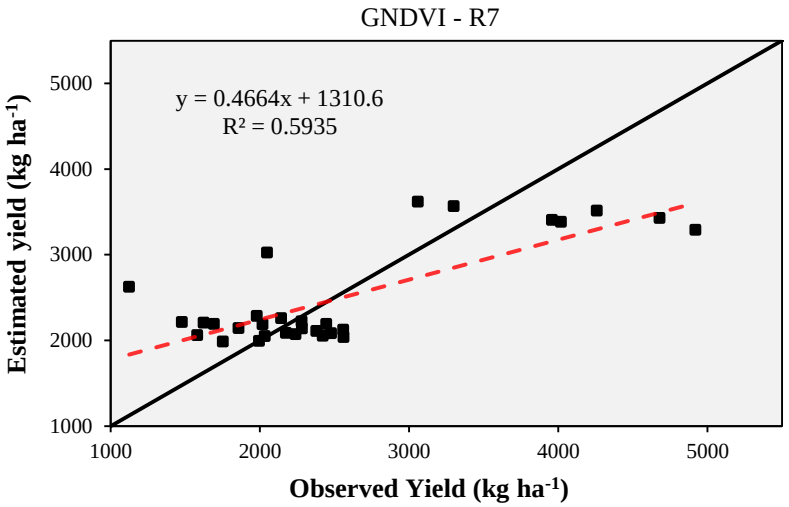

Figura 5. Regressão linear com o GNDVI como variável preditora com dados de ambas as safras para o estádio fenológico R7.

Foi também observado uma tendência dos modelos de regressão linear a superestimar as produtividades menores que 2800 kg ha-1, e, inversamente, a subestimar quando a produtividade observada excedeu 3000 kg ha-1. Esta característica poderia ser atribuída à presença de diferenças substanciais entre as produtividades das duas safras distintas.

No geral, os modelos de regressão linear mostraram um desempenho satisfatório, com R^2 ~ 0,55 e MAPE ~ 0,20, à exceção do modelo aplicado ao estágio R6. Este último apresentou uma precisão e acurácia significativamente reduzidas (R^2 < 0,2, MAE > 600 kg ha-1, MAPE = 0,3 e RMSE > 800 kg ha-1), indicando uma eficácia limitada nesse estágio específico.

A avaliação de produtividade de soja continuou com a implementação do algoritmo de Florestas Aleatórias (RF) (Tabela 5). As imagens de R4 geraram o modelo mais eficaz (Figura 6), superando a regressão linear em desempenho com um

MAPE de 0,18 e um MAE inferior a 450 kg ha-1. O GNDVI novamente se destacou como variável preditora em R4 e R7, gerando os modelos mais precisos e acurados com o algoritmo RF. Os índices NDVI e EVI também foram marcantes em R5 e R6, respectivamente. A tendência de superestimação para produtividades inferiores a 3000 kg ha-1 e subestimação para valores superiores persistiu com este algoritmo.

Tabela 5. Resultados das melhores estimativas obtidas pelos algoritmos de RF em cada estádio fenológico.

Phenological Stage	VI	MAE (kg ha^{-1})	MAPE	RMSE (kg ha^{-1})
R4	GNDVI	434.91	0.18	584.04
R5	NDVI	507.25	0.23	625.11
R6	EVI	471.84	0.20	609.52
R7	GNDVI	472.97	0.22	603.47

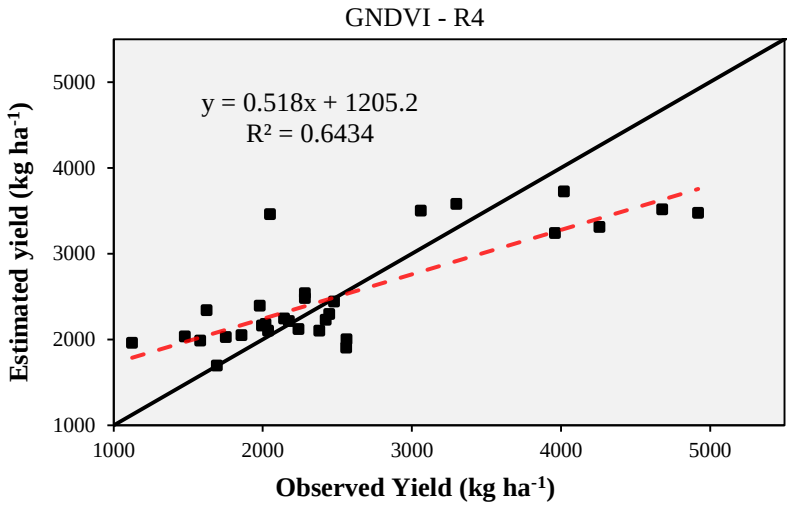

Figura 6. Random Forest com o GNDVI como variável preditora com dados de ambas as safras para o estádio fenológico R4.

As Redes Neurais Artificiais (ANNs) foram posteriormente testadas para a função de previsão de produção (Tabela 6). Neste contexto, os modelos mais precisos e acurados foram obtidos em comparação com os outros algoritmos. No estágio R5 (Figura 7), foram gerados os modelos mais exatos da avaliação. Os modelos para R6 e R7 mantiveram uma precisão comparável à da regressão linear simples. No entanto, os erros associados a todos os modelos baseados em ANNs

foram inferiores aos das outras avaliações. O NDVI emergiu como uma variável crítica na construção das redes, servindo como variável preditora nos modelos R5 e R7. O GNDVI e o SAVI também foram relevantes nos modelos dos outros dois estádios fenológicos. O perfil de rede neural que gerou a maioria dos melhores modelos foi o Perceptron Multicamada (MLP), com a exceção do modelo R7, onde a Função de Base Radial (RBF) foi utilizada. Notavelmente, a tendência de superestimação para produtividades até 2800 kg ha-1 e subestimação para valores mais elevados foi menos pronunciada em todos os modelos baseados em redes neurais.

Tabela 6. Resultados das melhores estimativas obtidas pelos algoritmos de ANN's em cada estádio fenológico.

Phenological Stage	Profile	VI	MAE (kg ha^{-1})	MAPE	RMSE (kg ha^{-1})
R4	MLP	GNDVI	264.93	0.13	380.78
R5	MLP	NDVI	295.74	0.12	372.84
R6	MLP	SAVI	392.64	0.17	522.78
R7	RBF	NDVI	341.30	0.15	475.36

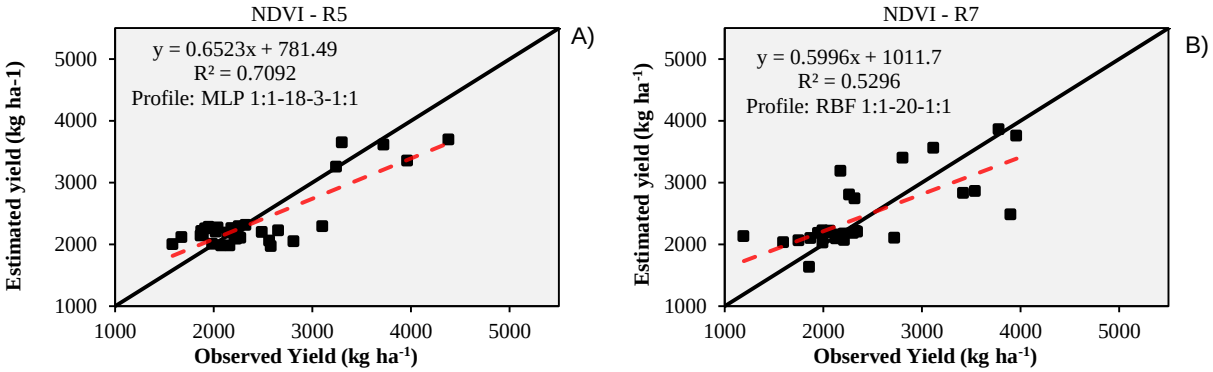

Figura 7. Redes neurais com A) perfil MLP e NDVI como variável preditora com dados de ambas as safras para o estádio fenológico R5 e B) perfil RBF com NDVI como variável preditora com números de ambas as colheitas para o estádio fenológico R7.

2.4. DISCUSSÃO

Com os resultados obtidos neste trabalho, foi possível esclarecer dúvidas quanto ao uso de alguns algoritmos para a estimativa de produtividade da soja e

em quais períodos o uso destes algoritmos geraria respostas confiáveis. Embora não tenha sido possível definir com precisão o estádio fenológico ideal em que as respostas tivessem destaque absoluto em relação aos valores de erro, foi possível identificar que os períodos de desenvolvimento discutidos trouxeram boas possibilidades quanto ao uso da metodologia para obtenção de informações de produtividade. Além disso foi possível traçar paralelos de comparação entre os algoritmos utilizados, onde as ANN's obtiveram destaque.

As medições realizadas pelo sensoriamento remoto por meio de imagens coletadas por sensores embarcados em VANT's são capazes de fornecer informações importantes sobre a cultura da soja, tanto em termos de desenvolvimento quanto de conhecimento sobre o potencial produtivo dessa cultura (Campos et al., 2018). No entanto, o estádio de desenvolvimento da cultura da soja em que as imagens para análise são capturadas é um fator determinante no sucesso da previsão de informações para aquela área e, conforme relatado em (Eugenio et al., 2020; Maimaitijiang et al., 2020), diferentes estudos mostram maior poder preditivo em relação à janela de estádios fenológicos da cultura da soja para coleta de imagens quando em R2 a R5, ou seja, floração e enchimento inicial de grãos. Outros estudos também destacam especificamente o estádio fenológico R5 como o melhor para fornecer imagens para fins de estimativa de produtividade (Ma et al., 2001; K. F. Zhang et al., 2019). E mesmo abordando estádios fenológicos como R6 e R7, nosso trabalho atestou o que foi apontado anteriormente. Para os três algoritmos utilizados, os períodos de desenvolvimento R4 e R5 sempre aparecem em destaque, sendo que o estádio R7 também apresentou resultados satisfatórios (para regressões lineares foi o período que mais se destacou, por exemplo).

Com o aumento das pesquisas em agricultura de precisão e sensoriamento remoto, a limitação sofrida pelos primeiros modelos desenvolvidos vêm sendo superadas, como a baixa precisão, acurácia e assertividade dos modelos. Isso ocorre principalmente quando são utilizados modelos de regressão linear, pois estes apresentam desempenho inferior na identificação de comportamentos não lineares quando comparados a algoritmos mais complexos. Kumar (2020) utilizou regressões lineares, comparadas com ANN's para estimar a produtividade da

cultura da soja, chegando a resultados semelhantes aos encontrados neste trabalho (regressões lineares com valores de R^2 ~ 0,5 e ANN's com valores de R^2 acima de 0,7). Alguns trabalhos como os de Júnior et al. (2022) e Andrade et al. (2022) conseguiram alcançar métricas melhores do que as encontradas no presente trabalho para regressões lineares (R^2 ~ 0,7, RMSE < 300 kg ha^{-1}). Acredita-se que essas respostas tenham sido obtidas devido ao uso de regressões múltiplas, que levaram em consideração bandas nas regiões do infravermelho próximo (NIR), vermelho e red-edge (Red-Edge) para expressar a resposta espectral da soja. Seguindo a tendência da pesquisa desenvolvida, outros autores, como Dos Santos e Rolim (2021) também encontraram no uso de redes neurais MLP os melhores resultados de predição comparados a outras estruturas (R^2 > 0,8, RMSE < 200 kg ha^{-1}), mostrando a evolução no aumento da precisão e acurácia ao utilizar modelos que conseguem identificar padrões não lineares de comportamento, o que confere um ganho quanto à assertividade dos modelos.

Apesar dessa evolução, as informações que serão utilizadas para abastecer as entradas desses modelos ainda são aspectos desafiadores. Na tentativa de encontrar respostas para esse tipo de problema, (Kuwata and Shibasaki, 2016; Ji et al., 2018; Davarzani et al., 2020) já utilizaram dados como condições climáticas, além de análises e ocupação do solo na composição de modelos de previsão de produtividade, com a intenção de trazer maior robustez aos mesmos. No entanto, informações confiáveis e precisas sobre clima, propriedades e ocupação do solo não são tão comumente disponíveis. Dados de sensoriamento remoto, por outro lado, estão disponíveis globalmente e são relativamente baratos, em muitos casos até gratuitos (You et al., 2017). Neste trabalho, mesmo utilizando apenas as bandas espectrais das áreas e informações de produtividade, foi possível obter bons resultados. Um estudo realizado por (Davarzani et al., 2020), investigando a influência de propriedades do solo e metodologias de tratamento no aumento da produtividade média da soja, obteve valores de R^2 moderadamente superiores aos encontrados no presente trabalho (R^2 de regressões lineares entre 0,62 e 0,78 e de ANN's entre 0,82 e 0,93), porém a relação custo-benefício proporcionada pelo uso de imagens e softwares gratuitos, faz com que os modelos desenvolvidos por nós se mostrem como alternativas viáveis.

Diferentes VI's foram utilizados como entrada para os três tipos de algoritmos de predição utilizados e os melhores foram selecionados para cada data. No presente trabalho, o GNDVI fez parte da maioria dos modelos selecionados, porém com o uso de algoritmos mais robustos, outros índices fizeram parte de modelos mais precisos e acurados, como é o caso do NDVI que apareceu como a melhor entrada em dois dos quatro modelos construídos com ANN's e em um modelo construído com RF (algoritmos com melhor desempenho). O NDVI também fez parte dos modelos com melhor desempenho em trabalhos como o de X. Zhang et al. (2019).

Devido a limitações relacionadas à alta densidade da vegetação, que resulta em alta absorção de radiação no comprimento de onda vermelho pelos pigmentos de clorofila, alguns trabalhos buscam utilizar o NDRE, quando há a possibilidade de trabalhar com a banda Red-Edge, devido a sua maior sensibilidade quando em vegetação com alta densidade (Oliveira et al., 2021; Souza et al., 2022), sendo uma alternativa na busca de modelos mais eficientes.

Alguns outros gargalos ainda evidenciam as vantagens da utilização de imagens de satélite sobre os VANT's, além do fato de ainda ser uma ferramenta de custo significativo e trabalhosa. No Brasil, a época de semeadura da soja pode ser marcada por chuvas fortes e tempo nublado, o que é de grande interesse quando se trata de gerenciar o uso de VANT's ou imagens de satélite. Com o uso de imagens de satélite, é possível selecionar adequadamente o dia e a hora, bem como a imagem que será usada. Na mesma situação, as condições climatéricas podem até mesmo impossibilitar o voo de um VANT em condições de chuva ou mesmo diminuir a qualidade das imagens devido à presença de neblina ou sombras provocadas pelas nuvens. Quando este serviço é contratado de empresas terceiras, o problema burocrático pode tornar-se ainda maior devido ao preço do serviço e a necessidade de agendamento de voos.

2.5. CONCLUSÕES

A predição da produtividade agrícola, particularmente da soja, é um elemento-chave na tomada de decisão e gerenciamento de práticas agrícolas. Este

estudo explorou a eficácia dos Índices de Vegetação (VI's), derivados de imagens de satélite, em associação com diversos algoritmos de regressão para prognosticar a produtividade da soja. Em consonância com os objetivos propostos, este trabalho avaliou com sucesso a competência dos modelos de previsão na estimação da produtividade da soja a partir de dados de sensoriamento remoto, contribuindo assim para o aperfeiçoamento e avanço de métodos capazes de fornecer informações acuradas sobre o estado da cultura antecipadamente. As Redes Neurais Artificiais (ANN's) destacaram-se pela superior acurácia e precisão na predição, sendo os índices NDVI e GNDVI as variáveis preditoras mais frequentemente empregadas pelos modelos. Portanto, este estudo atesta a viabilidade e eficiência do uso de imagens de satélite na elaboração de modelos de estimativa de produtividade da soja, oferecendo insights significativos para futuras pesquisas e aplicações práticas.

2.6. Referências

Alabi TR, Abebe AT, Chigeza G and Fowobaje KR (2022) Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. **Remote Sensing Applications: Society and Environment** 27:100782.

Aloisi RR and Demattê JLI (1974) Levantamento de solos da área onde se localiza a Faculdade de Medicina Veterinária e Agronomia de Jaboticabal. Científica. **Científica** 2:123–126.

Andrade TG, Junior ASDA, Souza MO, Lopes JWB and Vieira PFDMJ (2022) Soybean Yield Prediction Using Remote Sensing in Southwestern Piauí State, Brazil. **Revista Caatinga** 35:105–116.

Bishop CM (1995) *Neural Networks for Pattern Recognition*.

Breiman L (2001) Random Forests. **Machine Learning** 45:5–32.

Campos I, González-Gómez L, Villodre J, González-Piqueras J, Suyker AE and Calera A (2018) Remote sensing-based crop biomass with water or light-driven crop growth models in wheat commercial fields. **Field Crops Research** 216:175–188.

Chaplygin ME, Tronev S V. and Davydova SA (2021) Soybean harvesting using current dedicated headers and adapters. **IOP Conference Series: Earth and Environmental Science** 659.

CONAB (2022) *Safra 2022/23: Produção de grãos pode chegar a 308 milhões de t impulsionada pela boa rentabilidade de milho, soja e algodão*. Available at: https://www.conab.gov.br/ultimas-noticias/4731-safra-2022-23-producao-de-graos-pode-chegar-a-308-milhoes-de-toneladas-impulsionada-pela-boa-rentabilidade-de-milho-soja-e-algodao.

Davarzani S, Nagahi M, Tidwell M and Smith BK (2020) Pattern recognition using machine learning for corn and soybean yield prediction. **Proceedings of the 2020 IISE Annual Conference** 1467–1472.

Edreira JIR, Mourtzinis S, Azzari G, Andrade JF, Conley SP, Specht JE and Grassini P (2020) Combining field-level data and remote sensing to understand impact of management practices on producer yields. **Field Crops Research** 257:107932.
EMBRAPA (2021) *Soja em números (safra 2020/21)*. Available at: https://www.embrapa.br/en/soja/cultivos/soja1/dados-economicos.

Eugenio FC, Grohs M, Venancio LP, Schuh M, Bottega EL, Ruoso R, Schons C, Mallmann CL, Badin TL and Fernandes P (2020) Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. **Remote Sensing Applications: Society and Environment** 20.

Gitelson AA and Merzlyak MN (1996) Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. **Journal of Plant Physiology** 148:494–500.

Haykin S (1994) Department of Technology and Cognition University of North Texas 696.

Huete A (1988) A Soil-Adjusted Vegetation Index (SAVI). **Remote Sensing of Environment** 25:295–309.

Islam MM, Matsushita S, Noguchi R and Ahamed T (2021) Development of remote sensing-based yield prediction models at the maturity stage of boro rice using parametric and nonparametric approaches. **Remote Sensing Applications: Society and Environment** 22:100494.

Ji S, Zhang C, Xu A, Shi Y and Duan Y (2018) 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. **Remote Sensing** 10.

Júnior ASDA, Silva SP, Setúbal IS, Souza HA De, Vieira PFDMJ and Casari RACN (2022) Predicting soybean grain yield using aerial drone images. **Revista Brasileira de Engenharia Agrícola e Ambiental** 26:466–476.

Justice CO, Vermote E, Townshend JRG, Defries R, Roy DP, Hall DK, Salomonson V V., Privette JL, Riggs G, Strahler A, Lucht W, Myneni RB, Knyazikhin Y, Running SW, Nemani RR, Wan Z, Huete AR, Van Leeuwen W, et al. (1998) The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. **IEEE Transactions on Geoscience and Remote Sensing** 36:1228–1249.

Kaufman Y and Tanre. D (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. **IEEE Transactions on Geoscience and Remote Sensing** 30:260–271.

Kumar S (2020) Comparison of linear regression and artificial neural network technique for prediction of a soybean biodiesel yield. **Energy Sources, Part A: Recovery, Utilization and Environmental Effects** 42:1425–1435.

Kuwata K and Shibasaki R (2016) Estimating Corn Yield in the United States With Modis Evi and Machine Learning Methods. **ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences** III–8:131–136.

Ma BL, Dwyer LM, Costa C, Cober ER and Morrison MJ (2001) Early prediction of soybean yield from canopy reflectance measurements. **Agronomy Journal** 93:1227–1234.

Maimaitijiang M, Sagan V, Sidike P, Hartling S, Esposito F and Fritschi FB (2020) Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote Sensing of Environment** 237:111599.

Matese A and Di Gennaro SF (2021) Beyond the traditional NDVI index as a key factor to mainstream the use of UAV in precision viticulture. **Scientific Reports** 11:1–13.

Nolasco M, Ovando G, Sayago S, Magario I and Bocco M (2021) Estimating soybean yield using time series of anomalies in vegetation indices from MODIS. **International Journal of Remote Sensing** 42:405–421.

Oliveira MF de, Santos AF dos, Kazama EH, Rolim G de S and Silva RP da (2021) Determination of application volume for coffee plantations using artificial neural networks and remote sensing. **Computers and Electronics in Agriculture** 184:1–7.

Petersen LK (2018) Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. **Remote Sensing** 10:1–31.

Rossi M (2017) *Mapa pedológico do Estado de São Paulo: revisado e ampliado*. **Instituto Florestal (Florestal Institute) - IF**. Available at: http://correiadasilvaadvogados.com/pdf/art_amb/art_amb2.pdf.

Rouse JW, Haas Jr. RH, Schell JA and Deering DW (1976) Monitoring Vegetation Systems in the Great Plains with Erts. **NASA Spec. Publ.** 74:309–317.

Sagan V, Maimaitijiang M, Bhadra S, Maimaitiyiming M, Brown DR, Sidike P and Fritschi FB (2021) Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. **ISPRS Journal of Photogrammetry and Remote Sensing** 174:265–281.

Santos AF, Lacerda LN, Rossi C, Moreno LDA, Oliveira MF, Pilon C, Silva RP and Vellidis G (2022) Using UAV and Multispectral Images to Estimate Peanut Maturity Variability on Irrigated and Rainfed Fields Applying Linear Models and Artificial Neural Networks. **Remote Sensing** 14.

Dos Santos VB, Dos Santos AMF and Rolim G de S (2021) Estimation and forecasting of soybean yield using artificial neural networks. **Agronomy Journal** 113:3193–3209.

Schwalbert RA, Amado T, Corassa G, Pott LP, Prasad PVV and Ciampitti IA (2020) Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. **Agricultural and Forest Meteorology** 284:107886.

Shammi SA and Meng Q (2021a) Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. **Ecological Indicators** 121:107124.

Shammi SA and Meng Q (2021b) Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. **Ecological Indicators** 121:107124.

da Silva EE, Rojo Baio FH, Ribeiro Teodoro LP, da Silva Junior CA, Borges RS and Teodoro PE (2020) UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. **Remote Sensing Applications: Society and Environment** 18.

Souza JBC, de Almeida SLH, Freire de Oliveira M, Dos Santos AF, Filho AL de B, Meneses MD and Silva RP da (2022) Integrating Satellite and UAV Data to Predict Peanut Maturity upon Artificial Neural Networks. **Agronomy** 351.

Sumesh KC, Ninsawat S and 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.
Sun J, Di L, Sun Z, Shen Y and Lai Z (2019) County-level soybean yield prediction using deep CNN-LSTM model. **Sensors (Switzerland)** 19:1–21.

Xu J, Meng J and Quackenbush LJ (2019) Use of remote sensing to predict the optimal harvest date of corn. **Field Crops Research** 236:1–13.

You J, Li X, Low M, Lobell D and Ermon S (2017) Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. in **Thirty-First AAAI Conference on Artificial Intelligence** 4559–4565.

Zhang KF, Zhang Z, Zhang YH and Li H (2019) Experimental study of single-rotor UAV on droplet deposition distribution in soybean field. **Applied Ecology and Environmental Research** 17:13833–13844.

Zhang X, Zhao J, Yang G, Liu J, Cao J, Li C, Zhao X and Gai J (2019) Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. **Remote Sensing** 11.

Zhou Jing, Zhou Jianfeng, Ye H, Ali ML, Chen P and Nguyen HT (2021) Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network. **Biosystems Engineering** 204:90–103.

CAPÍTULO 3 - CONSIDERAÇÕES FINAIS

Fornecer valores de produtividade de maneira antecipada a colheita é de suma importância para o gerenciamento de propriedades agrícolas. Com esse estudo, conseguiu-se evidenciar fatos importantes para a sequência do melhoramento e desenvolvimento de novas técnicas e metodologias de estimativa de atributos produtivos. É possível predizer a produtividade da cultura da soja de forma precisa e acurada, explorando ferramentas que ganham, cada dia mais visibilidade com o desenvolvimento tecnológico voltado para o campo. O presente trabalho trouxe diferentes critérios comparativos – fases de desenvolvimento da cultura da soja, VI's para *input* e algoritimos para construção dos modelos - que enobrecem e clareiam os caminhos futuros para estudo e aplicação dessas ferramentas.

Em estudos futuros, pode-se estabelecer comparações entre imagens obtidas via satélite e VANT, por ser uma ferramenta de sensoriamento remoto e análise multiespectral que vem ganhando espaço nos campos e tornando-se mais acessível, além de possibilitar a análise de uma cultura dentro de estádios fenológicos específicos. Pode-se ainda acrescentar mais áreas produtivas as análises para avaliar o ganho em robustez de cada um dos modelos, sempre mantendo em vista o custo-benefício de cada operação e suas variáveis.

O potencial de implementação desse tipo de plataforma de fenotipagem de alto rendimento é de muita utilidade, uma vez que permite a exploração do desenvolvimento de culturas, não só da soja. O aumento da resolução das imagens fornecidas por satélites e o uso de sensores espectrais embarcados em VANT's poderão ser utilizados para avaliação do estado de diversas culturas, gerando uma resposta evidente em relação ao estado do crescimento e desenvolvimento fenológico das plantas.