

UNIVERSIDADE ESTADUAL DE GOIÁS
CAMPUS DE CIÊNCIAS EXATAS E TECNOLÓGICAS – HENRIQUE SANTILLO
MESTRADO EM ENGENHARIA AGRÍCOLA

DIOGO SANTOS SOUSA

**APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL E DADOS DE
SENSORIAMENTO REMOTO POR IMAGEM DIGITAL PARA PREDIÇÃO DA
PRODUTIVIDADE DE SOJA (*GLYCINE MAX L.*) EM REGIÃO DO CERRADO**

ANÁPOLIS - GO
2023

DIOGO SANTOS SOUSA

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL E DADOS DE SENSORIAMENTO REMOTO POR IMAGEM DIGITAL PARA PREDIÇÃO DA PRODUTIVIDADE DE SOJA (*GLYCINE MAX L.*) EM REGIÃO DO CERRADO.

Dissertação apresentada à Universidade Estadual de Goiás - UEG, Campus Central-Sede Anápolis CET (Ciências Exatas e Tecnológicas) como parte das exigências do Programa de Pós-Graduação Stricto Sensu em Engenharia Agrícola – Tecnologia da Produção Agrícola, para obtenção do título de MESTRE em Engenharia Agrícola.

Orientador: Prof. Dr. Elton Fialho dos Reis

ANÁPOLIS – GO
2023

TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL (BDTD)

Na qualidade de titular dos direitos de autor, autorizo a Universidade Estadual de Goiás a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UEG), regulamentada pela Resolução, **CsA n.1087/2019** sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

Dados do autor (a)

Nome Completo: Diogo Santos Sousa

E-mail: diogo.s.sousa.x@gmail.com

Dados do trabalho

Título: Aplicação de técnicas de inteligência artificial e dados de sensoriamento remoto por imagem digital para predição da produtividade de soja (*Glycine max* L.) em região do cerrado

Tipo

☐ Tese ☒ Dissertação

Programa de Pós-Graduação *Stricto sensu* em Engenharia Agrícola

Concorda com a liberação documento

☐ SIM

☒ NÃO **Casos de impedimento:** Período de embargo por até um ano a partir da data de defesa. A extensão deste prazo necessita de justificativa junto à coordenação do Programa.

☐ Solicitação de registro de patente;

☒ Submissão de artigo em revista científica;

☐ Publicação como capítulo de livro;

☐ Publicação da dissertação/tese em livro.

Anápolis,
Local

31/03/2023
Data



Documento assinado digitalmente
DIOGO SANTOS SOUSA
Data: 01/06/2023 17:44:51-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do autor



Documento assinado digitalmente
ELTON FIALHO DOS REIS
Data: 01/06/2023 18:11:01-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do orientador

SSO72 Sousa, Diogo Santos
5a APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA
ARTIFICIAL E DADOS DE SENSORIAMENTO REMOTO POR
IMAGEM DIGITAL PARA PREDIÇÃO DA PRODUTIVIDADE
DE SOJA (*GLYCINE MAX L.*) EM REGIÃO DO CERRADO
/ Diogo Santos Sousa; orientador Elton Fialho
dos Reis. -- Anápolis, 2023. 45 p.
Dissertação (Mestrado - Programa de Pós-
Graduação
Mestrado Acadêmico em Engenharia Agrícola) -
- Câmpus Central - Sede: Anápolis - CET,
Universidade Estadual de Goiás, 2023.

1. Agricultura de precisão. 2. Processamento
Digital de Imagem. 3. Mapa de produtividade.
4. Aprendizado de máquina. 5. Redes Neurais
Artificiais. I. dos Reis, Elton Fialho,
orient. II. Título.

DIOGO SANTOS SOUSA

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL E DADOS DE SENSORIAMENTO REMOTO POR IMAGEM DIGITAL PARA PREDIÇÃO DA PRODUTIVIDADE DE SOJA (GLYCINE MAX L.) EM REGIÃO DO CERRADO

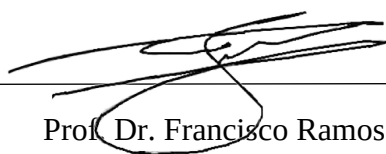
Dissertação apresentada à Universidade Estadual de Goiás - UEG, Campus Central-Sede: Anápolis CET (Ciências Exatas e Tecnológicas) como parte das exigências do Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola – Tecnologia da Produção Agrícola, para obtenção do título de MESTRE.

Aprovada em: 31, de Março, de 2023, pela banca examinadora constituída pelos professores:

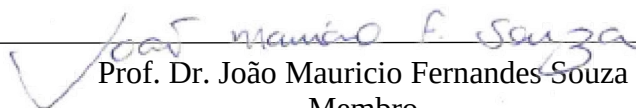


Documento assinado digitalmente
ELTON FIALHO DOS REIS
Data: 01/06/2023 18:09:41-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Elton Fialho dos Reis
Orientador
UEG/CCET



Prof. Dr. Francisco Ramos de Melo
Membro
UEG/CCET



Prof. Dr. João Mauricio Fernandes Souza
Membro
UniEVANGÉLICA

ANÁPOLIS - GO
2023

AGRADECIMENTOS

Em primeiro lugar, à Deus, por sua orientação, proteção e bênçãos durante todo o processo de pesquisa e escrita. Acredito que sem a sua sabedoria e bondade, não seria possível ter concluído esta importante etapa acadêmica.

Também quero agradecer à minha família, cujo incentivo foram fundamentais para que eu pudesse alcançar este objetivo. Sou imensamente grato pela confiança e pelo apoio incondicional que recebi de meus pais, irmãos e demais familiares.

Ao meu professor orientador e amigo, Dr. Elton Fialho dos Reis, cuja competência, comprometimento e generosidade foram essenciais para o sucesso deste projeto. Sua orientação acadêmica e profissional, bem como suas sugestões críticas e construtivas, foram de grande valia para o meu desenvolvimento como pesquisador.

Aos amigos que fiz durante o curso de mestrado, os quais contribuíram significativamente para a minha formação pessoal e intelectual. Sou grato pela amizade, pelo diálogo franco e enriquecedor, e pela colaboração mútua que tivemos ao longo desse período.

À Universidade Estadual de Goiás, ao Programa de Pós-Graduação Stricto Sensu em Engenharia Agrícola e aos professores do Mestrado em Engenharia Agrícola, pelo suporte e por toda a contribuição para o meu crescimento pessoal e profissional.

RESUMO

A soja (*Glycine max* L.) é uma das culturas mais importantes na produção agrícola do mundo e essencial no contexto da segurança alimentar. A demanda mundial por alimentos cresce cada vez mais, sendo necessária a adoção de novas tecnologias que permitam aumentar essa produção de forma sustentável. Nesse sentido, o objetivo deste trabalho foi verificar o potencial da utilização de modelos de inteligência artificial para estimar a produtividade da soja com base em dados de sensoriamento remoto (SR) obtidos por meio de plataformas aérea e orbital. Para isso utilizou-se uma área de aproximadamente 100 hectares cultivada com soja durante a safra 2021/2022 que foi subdividida em parcelas experimentais, totalizando um grupo amostral de 5011 parcelas destinadas ao estudo. Foram mensurados dados da resposta espectral do dossel da cultura em diferentes estádios fenológicos, e, por fim, feito o levantamento de produtividade com uma colhedora automotriz de grãos. Esses dados espectrais foram combinados, e em conjunto com índices de vegetação, utilizados para prever a produtividade da soja por meio de modelos de Redes Neurais Artificiais (RNAs). Foram modeladas duas arquiteturas de RNAs, sendo a primeira utilizando imagens do Satélite Planet Scope capturadas em 5 diferentes estádios fenológicos da cultura da soja, e a segunda a partir imagens capturadas por sensor acoplado em VANT no estágio R5. Os ajustes dos modelos foram avaliados para a parcela dos dados de teste por meio de índices estatísticos de coeficiente de correlação de Pearson (r), Erro Médio Absoluto (MAE), Erro Percentual Médio Absoluto (MAPE) e coeficiente de determinação (R^2). Para o modelo de predição baseado em imagens de satélite os melhores resultados encontrados foram no estágio vegetativo R5, tendo valores de correlação 0.71, erro médio absoluto de 175.45 kg ha⁻¹ representando um erro percentual de 3.68% e R^2 de 0.5. Já para o a modelagem a partir de dados do VANT foram encontrados valores de correlação de Pearson de 0.72, 172.94 kg ha⁻¹ para MAE, representando um MAPE de 3.66%, e R^2 de 0.51. Ambos os modelos não apresentaram tendência de agrupamento de erros espacialmente, tendo índice global de Moran próximo de 0. Os resultados indicam que a utilização tanto de imagens de satélite quanto de VANT, modelados por RNAs, podem fornecer uma predição, com precisão e robustez, da produtividade da soja.

Palavras-chave: Agricultura de precisão. Processamento digital de imagem. Mapa de produtividade. Aprendizado de máquina. Redes Neurais Artificiais.

ABSTRACT

Soybean (*Glycine max* L.) is one of the most important crops in the world's agricultural production and essential in the context of food security. The global demand for food is growing more and more, and the adoption of new technologies that allow this production to increase in a sustainable way is necessary. In this context, the objective of this work was to verify the potential of using artificial intelligence models to estimate soybean yields based on remote sensing (RS) data obtained through aerial and orbital platforms. For this study, an area of approximately 100 hectares cultivated with soybeans during the 2021/2022 harvest was used, which was subdivided into experimental plots, totaling a sample group of 5011 plots intended for the study. The spectral response of the crop canopy at different phenological stages was measured, and finally a yield survey was conducted with a grain harvester. These spectral data were combined, and together with vegetation indexes, used to predict soybean yields through Artificial Neural Network (ANN) models. Two ANN architectures were modeled, the first using data from the Planet Scope Satellite captured at five different phenological stages of the soybean crop, and the second from UAV data at the R5 stage. The fits of the models were evaluated for the plot of test data using the statistical indices of Pearson's correlation coefficient (r), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2). For the prediction model based on satellite images, the best results found were in the R5 vegetative stage, with correlation values of 0.71, mean absolute error of 175.45 kg ha⁻¹ representing a percentage error of 3.68% and R^2 of 0.5. As for modeling from UAV data, Pearson correlation values of 0.72, 172.94 kg ha⁻¹ for MAE, representing an MAPE of 3.66%, and R^2 of 0.51, were found. Both models showed no tendency to cluster errors spatially, having a global Moran's index of around 0. The results indicate that the use of both satellite and UAV imagery, modeled by ANNs, can provide accurate and robust prediction of soybean yields.

Keywords: Precision farming. Digital image processing. Yield map. Machine learning. Artificial Neural Networks.

SUMÁRIO

1	INTRODUÇÃO	10
2	REVISÃO BIBLIOGRÁFICA.....	13
2.1	CULTURA DA SOJA	13
2.2	AGRICULTURA DE PRECISÃO	14
2.3	MAPAS DE PRODUTIVIDADE	15
2.4	SENSORIAMENTO REMOTO	16
2.5	INTELIGÊNCIA ARTIFICIAL	17
2.6	REDES NEURAIS ARTIFICIAIS	18
3	MATERIAL E MÉTODOS.....	21
3.1	LOCALIZAÇÃO E CARACTERIZAÇÃO DA ÁREA DE ESTUDO	21
3.2	AQUISIÇÃO DE DADOS DE PRODUTIVIDADE	21
3.3	AQUISIÇÃO DE IMAGENS DE SATÉLITE.....	22
3.4	AQUISIÇÃO DE IMAGENS POR VANT.....	23
3.5	ANÁLISE ESTATÍSTICA E MODELAGEM	23
3.5.1	Redes Neurais Artificiais.....	23
3.5.2	Validação das Redes Neurais Artificiais	24
4	RESULTADOS E DISCUSSÃO	26
4.1	ESTATÍSTICA DESCRITIVA PARA OS DADOS DE PRODUTIVIDADE	26
4.2	MODELO DE PREDIÇÃO DE PRODUTIVIDADE COM USO DE SATELITE.....	27
4.3	MODELO DE PREDIÇÃO DE PRODUTIVIDADE UTILIZANDO DADOS VANT	30
5	CONCLUSÃO.....	34
	REFERÊNCIAS	35
	APÊNDICES	43

1. INTRODUÇÃO

A soja (*Glycine max* L.) é uma das culturas mais importantes na produção agrícola mundial e fator importante também no contexto da segurança alimentar, devido a sua alta qualidade nutricional e versatilidade industrial. No cenário mundial, o Brasil tem grande participação na produção de soja, onde o país se destaca por ser o maior produtor e exportador do grão no mundo, com projeção para a safra 2022/23, de produzir aproximadamente 150,36 milhões de toneladas do grão, tendo uma área plantada de 42,4 milhões de hectares (CONAB, 2022).

A demanda mundial por alimentos cresce cada vez mais, formando um cenário onde a agricultura futura precisará aumentar a produção, cumprindo essa demanda de uma forma sustentável, extraíndo o máximo de produtividade sem a abertura de novas áreas de cultivo (SRIVASTAVA et al., 2019). Nesse sentido, torna-se de fundamental importância a adoção de novas tecnologias que permitam o uso do solo de uma forma mais sustentável e produtiva (NOETZOLD et al., 2019).

Em busca por uma produção mais sustentável, a agricultura de precisão (AP) surge na década de 1980 e vêm sendo estudada nos últimos anos com o objetivo de identificar variações no ambiente produtivo, auxiliando tanto na tomada de decisão quanto no uso mais eficiente de insumos e equipamentos (FINGER, 2019). Para Molin (2015) a AP pode ser definida como uma forma de gerenciamento agrícola que considera a variabilidade espacial e temporal da unidade produtiva visando uma maior produtividade, rentabilidade econômica e redução de impactos ambientais.

A AP envolve o estudo da variabilidade espacial de aspectos como: solo, variação de relevo, vegetação e o histórico de produtividade da área para tentar solucionar problemas e gerenciar de forma localizada (BERNARDI et al., 2014). Para este gerenciamento e manejo de forma localizada, geralmente o ciclo de obtenção de dados da AP se inicia pelo mapa de produtividade, essencial para se conhecer o potencial produtivo da área cultivada (MOLIN, 2015).

Os dados de produtividade podem ser levantados por meio de colhedoras, de forma automática, porém há a necessidade de alguns dispositivos implantados nas máquinas além de mão de obra especializada para calibrar os sensores. Outro método tradicionalmente utilizado é a obtenção da produtividade por meio da coleta manual, entretanto por causa da laboriosidade e custo envolvido na coleta dos pontos amostrais, geralmente é feita de forma pontual, dificultando a análise da variabilidade espacial acurada (HEINEMANN; STONE; SILVA, 2010).

Para que se tenha um bom levantamento de variabilidade espacial é necessário uma grande quantidade de dados da área. Surge, nesse sentido, a possibilidade de obtenção desses dados a partir de sensoriamento remoto (SR). O SR se mostra como uma fonte de informação importante, capaz de auxiliar na identificação de variabilidade espacial de atributos de solo e planta, podendo fornecer indiretamente dados de produção. O SR tem como fonte principal de informação a radiação eletromagnética emitida ou refletida pelos alvos estudados, nessa técnica os dados podem ser obtidos por meio de sensores inseridos em bases orbitais, aéreas ou terrestres (LI et al., 2019).

Tem se tornado muito comum a utilização de modelos de Inteligência Artificial capazes de estimar o crescimento e avaliar o desenvolvimento de áreas produtivas, contribuindo para acompanhamento da produtividade agrícola e a compreensão dos fatores que estão envolvidos (ANAR et al., 2019). A aplicação de modelos baseados no aprendizado de máquina vem se mostrando como uma alternativa para a modelagem de dados complexos, como os provenientes do sensoriamento remoto (CATALÃO et al., 2011).

Um dos modelos de aprendizado de máquina mais difundido nos últimos tempos foram as redes neurais artificiais (RNAs). As RNAs são definidas como unidades de processamento interconectadas baseadas nas estruturas biológicas do cérebro humano (KHOSHROO et al., 2018). Esse método tem sido aplicado na resolução de problemas de grande complexidade, em que sua grande vantagem é a resolução de maneira facilitada de problemas não lineares, que geralmente não é encontrada em algoritmos de metodologias convencionais (SAMBORSKA et al., 2014).

Devido ao grande potencial das RNAs para resolução de problemas, estas tem sido aplicadas em diversos campos da pesquisa, inclusive na agricultura. Saaed e Wang (2019) aplicaram aprendizado de máquina para prever a produtividade em milhos híbridos e constataram que a implementação desses algoritmos possui grande potencial para a previsão de safras. Liu e Tian (2013) utilizaram RNAs para prever a produção de milho usando dados do solo, clima e manejo, tendo uma precisão na predição da produtividade em torno de 80%. Michelon et al., (2018) empregaram RNAs para estimar a produtividade de soja e milho por meio de leituras de clorofila e concluíram que seria necessário incluir mais variáveis de entrada para uma melhor estimativa da produtividade.

O potencial preditivo possibilitado pela utilização de dados do sensoriamento remoto aliado a modelagem através das RNAs poderá causar efeito positivo no monitoramento da produtividade agrícola de maneira localizada. A viabilização de mapas de produtividade é de suma importância para que o produtor possa gerir a sua propriedade de maneira precisa,

conhecendo áreas de maior ou menor potencial produtivo, afetando assim a aplicação de insumos, dimensionamento de estoques e a sustentabilidade da propriedade como um todo. Nesse sentido o objetivo deste trabalho foi verificar o potencial da utilização de modelos de inteligência artificial para estimar a produtividade da soja com base em dados de diferentes plataformas de sensoriamento remoto, sendo a aérea e orbital.

2. REVISÃO BIBLIOGRÁFICA

2.1 CULTURA DA SOJA

A soja (*Glycine max* L.) é uma cultura oleaginosa originária da China. Devido aos seus altos teores de proteína e óleo a planta passou a ter importância econômica durante o século XX, gerando grande interesse pelo setor industrial, sendo inicialmente no mercado asiático e logo depois se expandindo para o mundo (FEHR, 1980). O Brasil, Estados Unidos e Argentina dominam o mercado mundial do grão, sendo responsáveis por 80% da produção (TANWAR e GOYAL, 2021). No Brasil, a soja é cultivada em todas as regiões, tendo destaque dentro da cadeia do agronegócio ocupando 50% da área cultivada do país (DEFANTE et al., 2018).

Pertencente a família Fabaceae, conhecida popularmente como leguminosas, a soja é uma planta herbácea, de reprodução autógama, possuindo ciclo anual, podendo ter crescimento morfológico determinado ou indeterminado, com altura média variando de 0,3 a 2,0 metros e um ciclo de 75 a 200 dias, dependendo das condições ambientais e da cultivar (SEDIYAMA, 2013).

O ciclo fenológico da soja foi estudado por Fehr e Caviness (1977), que descreveram os estádios fenológicos da cultura em uma escala alfanumérica, subdividida em duas grandes fases, sendo a vegetativa representada pela letra V e reprodutiva representada pela letra R (tabela 1).

Tabela 1 - Descrição dos estádios fenológicos propostos por Fehr e Caviness (1977).

Estádios Vegetativos	Estádios Reprodutivos
VE - Cotilédones acima da superfície do solo	R1 - Primeira flor aberta
VC - Cotilédones totalmente abertos	R2 - Florescimento pleno
V1 - Primeiro par de folhas aberto	R3 - Início da Formação das vagens
V2 - Primeiro nó com folha trifoliolada aberta	R4 - Formação completa das vagens
V3 - Segundo nó com folha trifoliolada aberta	R5 - Processo de Enchimento de grãos
.	R6 - Grãos completamente formados
.	R7 - Início da maturidade fisiológica (folhas e vagens amareladas)
.	R8 - Final da maturidade fisiológica (desfolha natural)
V(n) - Enésimo nó com trifoliolada aberta	

Fonte: FEHR e CAVINESS, (1977);

Os estádios vegetativos, representados pela letra V, compreendem todo o processo inicial de desenvolvimento da planta, passando pela formação de nós, ramos e folhas. Já os estádios reprodutivos representados pela letra R, são relacionados ao período de florescimento

e maturação finalizando no estágio R8. Estes podem ser ainda subdivididos em quatro fases diferentes, sendo: (R1 e R2) florescimento; (R3 e R4) desenvolvimento da vagem; (R5 e R6) desenvolvimento de grão e (R7, R8) maturação da planta (tabela 1).

2.2 AGRICULTURA DE PRECISÃO

A Agricultura de Precisão (*Precision Agriculture* ou *Precision Farming*) é definida pela Sociedade Internacional de Agricultura de Precisão (ISPA, 2022) como uma estratégia de gerenciamento agropecuário que coleta, processa e analisa dados temporais e espaciais do local de produção, e, em conjunto com outras informações, guia a tomada de decisão do produtor. A AP torna possível a gestão agrícola considerando as dimensões de espaço e tempo, permitindo com que o produtor trate o local de produção como uma entidade heterogênea (FINGER, et al., 2019).

Apesar de ser comumente tratado como um conceito recente, a AP tem seus primeiros registros em estudos feitos por Linsley e Bauer (1929), acerca da variabilidade espacial do solo, no qual os autores recomendavam diferentes quantidades de calcário, com base em análises de pH do solo amostradas em um grid de 0,4 ha. Contudo a sua popularização só ocorreu a partir dos anos 2000 com a difusão de receptores de navegação, causada pelo desligamento da degradação do sinal GPS (Global Positioning System) pelo governo americano, viabilizando a gestão de informações de forma localizada (MOLIN, et al., 2015).

A variabilidade espacial consequentemente gera uma heterogeneidade da produtividade da lavoura, que pode ser influenciada por diversos fatores que interagem entre si, de forma complexa, interferindo no potencial econômico da área. Portanto, é de fundamental importância o estudo da variabilidade espacial, tanto de atributos ligados ao solo como a planta, para que se possa entender quais são os fatores-chaves que interferem na produtividade da cultura, desenvolvendo assim, manejos agrícolas em áreas específicas, visando a máxima lucratividade da cultura subdividindo em unidades de manejo diferenciado (VIAN et al., 2016).

A AP envolve um ciclo de atividades que podem ser resumidas em três etapas. A primeira consiste na identificação, coleta de dados e compreensão das diferenças espaciais e temporais de atributos que interferem na produtividade. Em uma segunda etapa são realizadas as análises desses dados coletados, transformando-os em informações e recomendações úteis, para que haja uma tomada de decisão assertiva no campo. Por fim é realizada a aplicação das recomendações de forma localizada (PIERCE e NOWAK, 1999).

Dessa forma, ao longo desse ciclo pode-se observar diversas aplicações para a AP, que podem englobar, mapeamento de produtividade, mapeamento de atributos físicos e químicos

do solo, zoneamento de plantas daninhas, identificação localizada de pragas e doenças, caracterização do estado nutricional das plantas, entre diversos outros monitoramentos e intervenções da lavoura. Para a construção dessas informações, são utilizadas tecnologias complementares: como de posicionamento, através dos Sistemas de Navegação Global por Satélite (GNSS); de aquisição de dados, a partir de imagens adquiridas por sensoriamento remoto e até sensores proximais acoplados em máquinas; e de processamento por meio de Sistemas de Informações Geográficas dedicadas (SIGs) e máquinas inteligentes (MOLIN, 2011).

2.3 MAPAS DE PRODUTIVIDADE

O processo de gerenciamento da lavoura considerando a sua variabilidade espacial é composto por uma sequência de tarefas, que consequentemente devem ter um ponto de partida. É comumente difundido na literatura e por usuários, que o mapa de produtividade é a informação mais adequada para este ponto de partida, permitindo mostrar a resposta real da cultura ao ambiente e manejo em que cada planta foi submetida (MOLIN; AMARAL; COLAÇO, 2015).

Um mapa de produtividade é basicamente uma representação da produção da lavoura por meio de um conjunto de pontos georreferenciados. Essa informação pode ser levantada de diversas formas, porém a mais difundida é por meio de sensores de fluxo equipados em colhedoras. Existem diferentes formas de se medir o fluxo de sólidos, no entanto, essa medição em condições dinâmicas como é a da colheita se torna uma tarefa complexa. Desde os anos 80, quando surgiram as primeiras tentativas de se medir o fluxo de grãos, vários produtos foram lançados, onde a maioria consiste em mecanismos compostos por sensores acoplados no elevador de grãos da colhedora, podendo possuir diferentes princípios de funcionalidade, sendo os primeiros desenvolvidos baseados em raios gama e atualmente os mais utilizados são os baseados em placas de impacto (MACHADO et al., 2004).

Com base nas características que os mapas de produtividade possuem, é possível identificar várias utilidades, onde três se destacam. A primeira delas é entender a dinâmica do processo produtivo, o porquê de haver diferentes produtividades através de um contexto de causa e efeito. A segunda é em relação a aplicação de fertilizantes embasada em metodologias que consideram a quantidade de nutrientes exportada na safra anterior, aplicando uma taxa variável por relação direta com o que foi produzido e retirado do solo. A terceira é o auxílio na identificação de zonas que podem ser tratadas de forma diferenciada, por apresentar

consistentes diferenças de produtividade quando comparados dados de uma sequência de safras (MOLIN et al., 2015).

Apesar da grande quantidade de benefícios proporcionados pelos mapas de produtividade, para que eles se concretizem de forma eficiente e confiável, é preciso entender os mecanismos e equipamentos utilizados no processo de obtenção desses dados. Erros podem surgir devido a uma série de fatores, incluindo características do terreno, variações climáticas, erros do sensor, bem como procedimentos de calibração e validação dos dados (GEBBERS et al, 2010). No entanto, é importante destacar que tais erros não necessariamente descredibilizam a informação obtida, mas sim apontam para a necessidade de melhorias no processo de obtenção e processamento dos dados para garantir a confiabilidade dos resultados (ALMEIDA et al., 2017). A aplicação de técnicas de sensoriamento remoto pode fornecer informações precisas e valiosas para o mapeamento da produtividade agrícola, desde que os procedimentos de coleta, processamento e análise dos dados sejam conduzidos com rigor e precisão

2.4 SENSORIAMENTO REMOTO

No âmbito da agricultura, práticas envolvendo Sensoriamento remoto (SR) tem sido utilizadas desde o início da década de 1970, tendo os Estados Unidos como pioneiros no mundo em realização de pesquisas, possibilitadas pelo lançamento do primeiro satélite Landsat que entrou em órbita em 1972 (MACDONALD et al., 1975). Seguindo essa mesma linha, também desde a década de 1970, o INPE (Instituto Nacional de Pesquisas Espaciais) realiza estudos envolvendo SR a nível nacional, consolidando a tecnologia e destacando o Brasil como pioneiro na América do Sul (FORMAGGIO e SANCHES, 2017).

O SR está diretamente ligado à mensuração, por meio de sensores, da radiação eletromagnética emitida ou refletida em diferentes comprimentos de ondas por objetos e a sua disponibilização de forma passível de ser interpretada e tratada (MOREIRA, 2011). O uso destes sensores acoplados em plataformas orbitais, aéreas e terrestres proporciona o estudo eficiente da variabilidade espaço-temporal de áreas agrícolas, de forma rápida e com baixo custo (CASTRIGNANO, 2020).

As plataformas de SR mais utilizadas para o monitoramento agrícola são os satélites e os veículos aéreos não tripulados (VANTs). A qualidade dos sensores está relacionada com a capacidade destes em se obter medidas detalhadas da energia eletromagnética. Pode ser classificada em função de características ligadas a resolução espacial, temporal, espectral e radiométrica. A resolução espacial está relacionada ao tamanho dos alvos que o sensor consegue distinguir, enquanto a resolução temporal determina o tempo necessário para que o

sensor capte dados de um mesmo alvo. Do ponto de vista espectral está ligado a largura de espectro e quantidade de bandas em que esse sensor opera. Por fim, a resolução radiométrica diz respeito a capacidade do sensor em distinguir e captar sinais e ruídos vindos da energia eletromagnética refletida pelos objetos (MORAES, 2002).

No campo agrícola, diversos comprimentos de ondas do espectro eletromagnético têm sido aplicados para estudo da vegetação, as faixas mais utilizadas são as do espectro visível e do infravermelho próximo (MOLIN et al., 2015). A radiação incidente na superfície interage de modo diferente para cada tipo de objeto, no caso da vegetação, a região do visível (vermelho, verde e azul) é um indicativo importante pelo fato de que a maior parte da radiação emitida nesses comprimentos de onda são absorvidas pelas clorofilas “a” e “b” para realização de atividade fotossintética. Já na região do infravermelho o comportamento da radiação é em função de estruturas celulares internas, onde há uma baixa absorção e uma alta refletância, podendo chegar a mais de 50% (TUCKER e GARRATT, 1977).

As contribuições mais relevantes do SR para a agricultura envolvem o monitoramento e a estimativa de parâmetros biofísicos das culturas agrícolas. Dessa forma, Índices de vegetação (IV) podem ser utilizados para monitorar o vigor vegetativo e a dinâmica temporal. Os IVs são equações matemáticas que utilizam os valores de refletância de bandas espectrais específicas para realçar a resposta da vegetação e minimizar efeitos de solo e da variação da irradiação solar. A utilização de índices tem o objetivo de melhorar informações e evidenciar aspectos das plantas como, propriedade bioquímicas, biomassa, área foliar, teor de clorofila, produtividade, severidade de doenças, entre outros (KHANAL et al., 2017).

2.5 INTELIGÊNCIA ARTIFICIAL

As relações entre características de vegetação e sua resposta espectral tem sido estudadas por diversos tipos de modelagem. Modelos empíricos empregados na estimativa de parâmetros biofísicos, de um modo geral alcançam boas correlações, porém não podem ser generalizados pois são limitados a relações locais específicas da situação avaliada (REINERMANN e ASAM, 2020). Já os métodos não paramétricos de Inteligência artificial (IA) podem ter uma aplicação mais eficaz para predição com base em dados de SR, uma vez que envolvem pressupostos e modelos previamente estabelecidos para produzir resultados mais generalizáveis (POLEY e MCDERMID, 2020).

A inteligência artificial é um termo amplamente utilizado para se referir a métodos, ferramentas e máquinas que tem alta capacidade de reconhecer padrões e solucionar problemas de forma autônoma. Dentre esses métodos, o aprendizado de máquina (machine learning) é um

subdomínio da IA caracterizado por aplicar abordagens matemáticas para construir algoritmos que possuem a habilidade de trazer soluções para problemas complexos, onde a solução envolve muitas variáveis não sendo facilmente previsível (Jha et al., 2019).

Os métodos de aprendizado de máquina analisam relações não lineares entre variáveis independentes e dependentes, evidenciando os efeitos destas variáveis através de respostas adquiridas em treinamentos realizados com dados fornecidos previamente. Diversos métodos e lógicas tem sido desenvolvidos a partir do aprendizado de máquina (Lógica Fuzzy, Redes Neurais Artificiais (RNA's), Sistemas Especializados, Random Forest, entre outros) em busca de melhorar e trazer soluções mais assertivas para diversos tipos de problemas (SHRESTHA et al., 2021).

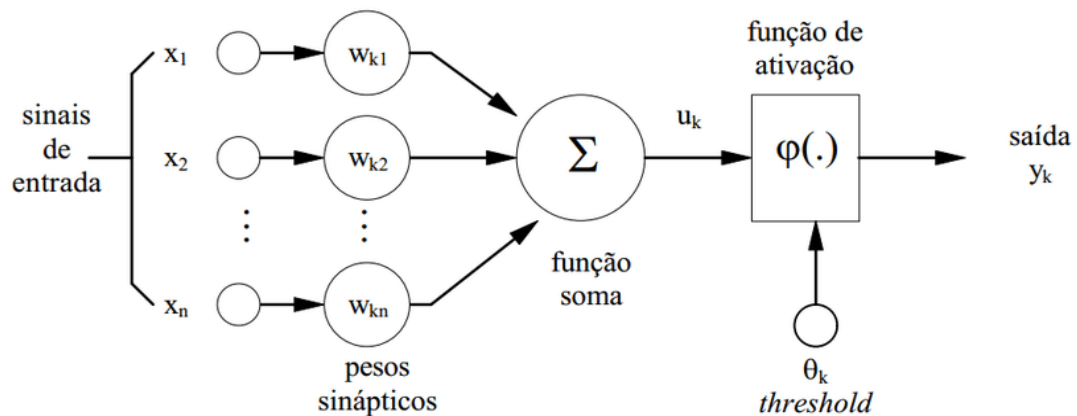
Dentre esses modelos, as Redes Neurais Artificiais têm sido amplamente utilizadas na agricultura por apresentarem alta precisão, capacidade de lidar com dados ruidosos, ser menos sensíveis a problemas como o overfitting e não demandar seleção de atributos. Diversas são as aplicações encontradas na literatura, passando desde determinação de características químicas do solo, estimativa de atributos físicos do solo, classificação de doenças, identificação de plantas daninhas até predição de produtividade (FERNANDES et al., 2019; FERNANDES et al., 2020; MONTEIRO et al., 2021; DYRMANN et al., 2016; ALVES et al., 2018)

2.6 REDES NEURAIIS ARTIFICIAIS

Inspiradas no sistema nervoso dos seres vivos, as Redes Neurais Artificiais (RNAs) são algoritmos computacionais que possuem a capacidade de aprender através de experiencias fornecidas (HAYKIN, 1999). De forma análoga a neurônios biológicos, as redes neurais são compostas por unidades de processamentos que são interligadas atuando simultaneamente em um processo de aprendizagem realizado por meio de algoritmos de treinamento (SOARES, 2021).

Em 1943, foi proposto por Warren McCulloch e Walter Pitts, o primeiro modelo básico de neurônio artificial. Como visto na Figura 1, o modelo é composto por sinais de entrada, que seguem para as conexões sinápticas, onde para cada conexão é dada um peso que é ajustado à medida que o conjunto de dados é treinado. No passo seguinte os valores passam por uma função soma, que tem o papel de unificar os valores de entrada conforme seus respectivos pesos no modelo. E por último, os valores passam pela função de ativação, que fornece o valor de saída de um neurônio, ativando ou não seu valor, conforme sua relevância dentro do modelo (LIN e SHEN, 2018).

Figura 1 - Modelo de Neurônio artificial proposto por McCulloch e Pitts em 1943.



Fonte: Haykin, 2001.

Dentro do modelo de rede neural, os neurônios são organizados formando uma arquitetura, de modo que o resultado é altamente dependente da definição dos parâmetros que envolvem essa arquitetura (FLECK et al., 2016). Os principais atributos a serem observados são relacionados ao número de camadas e as conexões entre elas, número de neurônios presentes em cada camada e a função de ativação atribuída a cada neurônio (Haykin, 2001).

Atualmente a arquitetura de RNA mais utilizada é estruturada em múltiplas camadas, sendo denominada Multilayer Perceptron (MLP), que possibilita o aprendizado de problemas mais complexos (Khoshroo et al., 2018). Sua estrutura é composta basicamente por: Uma camada de entrada, contendo informações das variáveis preditoras; n camadas intermediárias ou ocultas, contendo neurônios; e a camada de saída composta por neurônios que apresentam os resultados finais (Haykin, 2001).

Neste sentido, as RNAs MLP têm sido aplicadas utilizando diversos tipos de atributos em diferentes ambientes afim de estimar a produtividade da soja com precisão. Alves et al. (2018) realizou um estudo onde as variáveis de estrada se baseavam em classes de hábito de crescimento, densidade populacional e características agronômicas mensuradas manualmente, com base nos resultados o autor concluiu que é possível mensurar a produtividade com base nas características utilizadas.

Já Sagan et al. (2021), utilizaram modelo de RNA com dados de sensoriamento remoto vindos de satélites como entrada para previsão da produtividade de soja, onde concluíram que o aprendizado profundo foi capaz de prever a produtividade usando imagens de satélite no estado de Missouri no Estados Unidos.

Eugenio et al. (2020) construíram modelo de RNA para prever a produtividade com base em dados de sensores acoplados em VANT's, onde, apesar de considerar a produtividade como um parâmetro complexo de se prever, tiveram resultados satisfatórios. Utilizando uma

combinação multiplataforma de sensores, termal, na região do visível e multiespectral, Maimaitijiang et al. (2020) conseguiram obter previsões de produtividade precisas por meio de RNA's, constatando que esses modelos de aprendizado de máquina são eficientes para lidar com grande quantidade de variáveis de entrada.

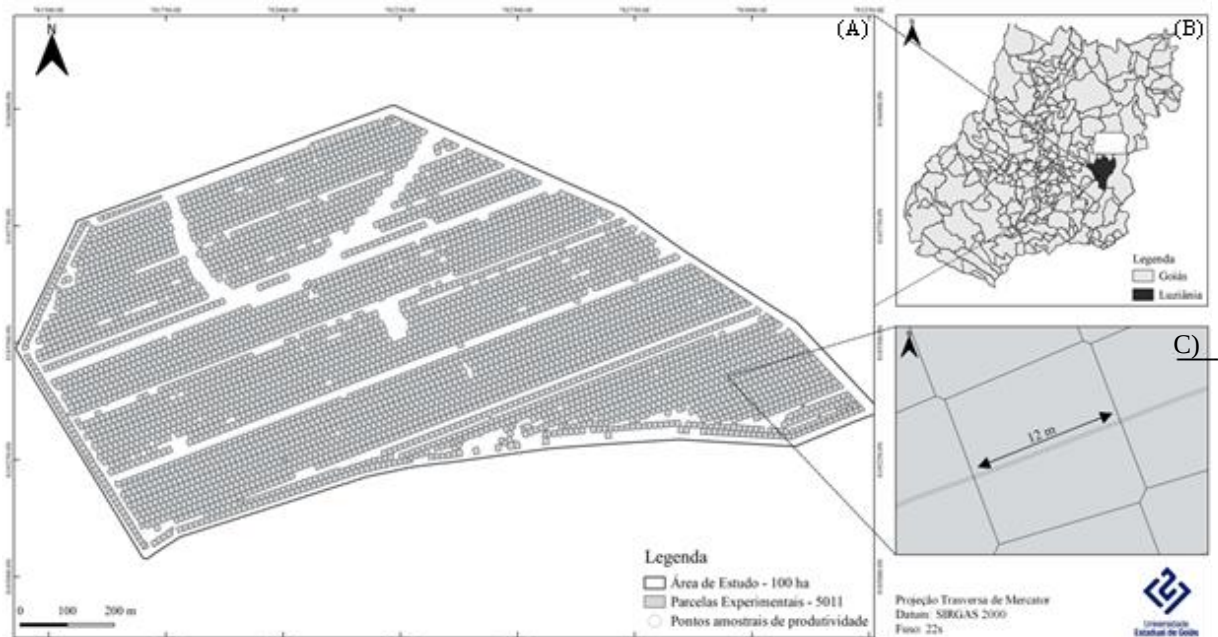
3. MATERIAL E MÉTODOS

3.1 LOCALIZAÇÃO E CARACTERIZAÇÃO DA ÁREA DE ESTUDO

O local de estudo pertence a uma fazenda comercial, localizada no município de Luziânia, estado de Goiás, Brasil (Figura 2). Segundo a classificação de Köppen (1948), o clima da região é caracterizado como do tipo Aw, com temperatura média de 22,7 °C e pluviosidade média anual próximo de 1.500 mm. O solo é classificado como Latossolo Vermelho com textura franco argilo arenosa.

A área destinada a pesquisa correspondeu a um talhão de 100 ha cultivado com soja (*Glycine max* L.), cultivar Voraz Ipro. A semeadura ocorreu no dia 12 de novembro de 2022 em sistema de plantio direto sob a palhada de milho com espaçamento de 0,45 m entre linhas e densidade populacional de 350.000 plantas por hectare em regime de sequeiro.

Figura 2 - Localização da área de estudo: A) Talhão estudado; B) Estado de Goiás; C) Parcela experimental.



Fonte: Autor (2022).

3.2 AQUISIÇÃO DE DADOS DE PRODUTIVIDADE

Para levantamento de informações de produtividade foram utilizados dados levantados provenientes de uma colhedora da marca John Deere®, modelo S790, equipada com display CommandCenter™ 4600, sensores de umidade e peso em seu sistema de transporte de grãos e receptor de sinal RTK (real-time kinematic) StarFire™ 6000, oferecendo precisão de localização de 2,5 cm.

Após coleta dos dados de produtividade foi realizada a identificação e remoção de valores anormais por meio de um processo de filtragem de 2 etapas, conforme Taylor et al.

(2007). Na primeira etapa foram removidos valores globais nitidamente fora dos padrões, onde possivelmente houve falha em algum sensor durante o processo, nesse sentido foram removidos valores abaixo de 100 kg ha^{-1} e acima de 12000 kg ha^{-1} , também foram removidos nessa etapa pontos onde o sensor de umidade apresentou valores abaixo de 5%, que são indicativos de erro durante o processo de mensuração.

Na segunda etapa foram removidos outliers locais, utilizando o software MapFilter (MOLIN, 2013). O software foi configurado para identificar pontos discrepantes, em um raio de 12m com valores que tinham variação superior a 25% em relação à mediana de seus vizinhos (MALDANER et al., 2021).

Após a remoção dos pontos discrepantes foi construída uma grade, no mesmo sentido das linhas de colheita, com parcelas amostrais de 12 metros de comprimento, conforme visualizado na figura 2. Para cada parcela amostral foi computada a média de produtividade com base nos pontos mensurados pela colhedora, totalizando um grupo amostral de 5011 parcelas.

3.3 AQUISIÇÃO DE IMAGENS DE SATÉLITE

Para imageamento via satélite, cinco conjuntos de imagens do satélite PlanetScope(PS), sem presença de nuvem, foram adquiridas entre o período de 12 de novembro de 2021 a 01 de março de 2022 . Essas imagens foram utilizadas como refletância de superfície para análise. As imagens PS possuem quatro bandas, incluindo B (485 nm), G (545 nm), R (630 nm) e NIR (820 nm), com um tamanho de pixel ortorretificado de 3,125 m e uma taxa de revista diária (HOUBORG e MCCABE, 2018).

As 5 datas escolhidas para aquisição de imagens correspondem a estádios vegetativos chave no desenvolver do ciclo da cultura: na primeira imagem adquirida a soja estava no estágio V3, caracterizado por conter dois trifólios; a segunda ocorreu no estágio R2, onde há flores completamente desenvolvidas na parte superior da planta; a terceira imagem foi no estágio R3, início do desenvolvimento das vagens; a quarta imagem ocorreu no estágio R5, final do processo de enchimento de grãos; e a quinta imagem no estágio R6 onde os grãos já estão completamente formados (FEHR; CAVINESS, 1977).

Utilizando as imagens, foram calculados os principais índices de vegetação, possíveis a partir das informações disponíveis no sensor utilizado, que são considerados correlatos com informações de produtividade (Apêndice A). Além dos índices de vegetação, foram extraídos valores de, média, mediana, desvio padrão e variância da refletância bruta das quatro bandas

do sensor (R, G, B e NIR) dentro de cada parcela experimental, totalizando 34 variáveis preditoras calculadas a partir das imagens PS (Apêndice A).

3.4 AQUISIÇÃO DE IMAGENS POR VANT

Os dados referentes ao sensoriamento aéreo foram obtidos utilizando um Veículo aéreo não tripulado (VANT) modelo eBee SQ da marca SenseFly®, equipado com um sensor multiespectral Parrot®, integrado ao sistema Sunshine® sensor de correção radiométrica, que capta a radiação em 4 bandas, sendo: G (550 nm); R (660 nm); Red Edge (735 nm); e NIR (790 nm).

A captura de imagens da área de estudo ocorreu no dia 15 de fevereiro de 2022, no momento em que a cultura se encontrava no estágio R5, momento final de enchimento de grãos, sendo período de grande interesse para estimativa de produtividade (FEHR et al., 1977). O voo foi realizado no período entre as 11:00 e 14:00 h. A altura de voo foi de 160 m, o que proporcionou um GSD (Ground Sample Distance) de aproximadamente 0,15 m. Os voos foram automatizados com um plano de passadas paralelas, garantindo uma sobreposição de imagens de 80% lateral e frontal (TMUŠIĆ et al., 2020).

O software Agisoft Metashape® foi empregado para ortorretificar e mosaicar as imagens, sendo calibradas radiometricamente usando a irradiância de valores capturados pelo sensor de luminosidade durante o voo. Os valores brutos da imagem foram convertidos em refletância, e a partir desses dados, dentro de cada parcela experimental, foram extraídos os valores de média, mediana, desvio padrão e variância de cada banda espectral. Foram calculados também os principais índices de vegetação, possíveis a partir das informações disponíveis no sensor utilizado, totalizando 34 variáveis preditoras calculadas a partir das imagens aéreas (Apêndice B).

3.5 ANÁLISE ESTATÍSTICA E MODELAGEM

3.5.1 Redes Neurais Artificiais

As RNAs foram implementadas utilizando ambiente de programação Python com TensorFlow (Abadi et al., 2016) e bibliotecas Keras (Keras, 2018). Das 5011 parcelas contidas na base de dados, foram divididas de forma estratificada em 70% (3507 parcelas) para treino, 15% (752 parcelas) para validação interna e 15% (752 parcelas) para teste externo.

A RNA proposta para os dados provenientes de imageamento orbital constituiu em uma estrutura composta por sub redes paralelas, contendo informações espectrais de cada estágio de coleta separadamente, nas quais cada sub rede é treinada numa estrutura individual

e é feita uma fusão de recursos a nível intermediário através de uma camada de concatenação (Apêndice C, a).

Entre as camadas intermediárias foi utilizada função de ativação denominada ReLU (Rectified Linear Unit), (LECUN et al., 2015), em conjunto com uma normalização em lote, para reduzir os efeitos de *overfitting* (IOFFE e SZEGEDY 2015). Entre a última camada oculta e a camada de saída, a função aplicada foi a Sigmoid, por ser semelhante a função ReLU, porém trabalhando melhor com camadas contendo poucos neurônios como na camada de saída (KERAS, 2022).

Para o modelo proposto para os dados de VANT foi testado um algoritmo de otimização Bayesiana para selecionar os valores ótimos dos hiperparâmetros (BROCHU et al., 2010). Foram testados diferentes números de camadas e neurônios, onde, para o número de camadas foi fornecido um campo amostral variando de 1 a 10, e para o número de neurônios foi fornecido valores de 1 a 1000. Foram treinadas 500 redes, e, conforme observado na Apêndice C-b, a otimização chegou a um modelo com 3 camadas ocultas, sendo a primeira com 910 neurônios, a segunda com 338 e a terceira com 13. Entre as camadas foram utilizadas as mesmas funções de ativação do modelo anterior.

3.5.2 Validação das Redes Neurais Artificiais

A avaliação dos modelos considerou em primeira instância o coeficiente de determinação R^2 (Equação 1), seguido pela correlação de Pearson (Equação 2) e análise dos erros MAE (Equação 3) e MAPE (Equação 4), utilizando a validação cruzada.

$$R^2 = r^2 \quad (1)$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n (x_i - y_i)}{n} \quad (3)$$

$$MAPE = \frac{MAE}{\bar{y}} * 100 \quad (4)$$

Em que:

x_i = valores estimados;

y_i = valores observados;

$\bar{x}$ = média dos valores estimados;

$\bar{y}$ = média dos valores observados;

n = número de parcelas.

A fim de verificar a adaptabilidade do modelo de previsão de produtividade de grãos no espaço, devido às várias condições ambientais geralmente encontradas no campo, foi calculado o Índice Global de Moran (GMI). O GMI fornece uma estimativa de incerteza relacionada ao padrão de distribuição espacial dos erros (IMRAN et al. 2015). O erro refere-se aos resíduos entre o rendimento medido e previsto de cada parcela, o índice varia de -1 a 1, demonstrando autocorrelação espacial negativa a positiva, respectivamente, com 0 indicando aleatoriedade (Moran, 1950), onde maior aleatoriedade indica melhor desempenho do modelo.

Para comparar estatisticamente os valores médios estimados pela RNA com os valores médios mensurados em campo, para ambos os modelos treinados, foi realizado o Teste-t.

4 RESULTADOS E DISCUSSÃO

4.1 ESTATÍSTICA DESCRITIVA PARA OS DADOS DE PRODUTIVIDADE

Os dados de produtividade da área de estudo foram submetidos a uma análise descritiva, no qual os resultados podem ser observados na Tabela 2. Conforme descrito na tabela, a área registrou valor médio de produtividade de 4942.70 kg ha⁻¹, se mostrando acima da média nacional, que registrou 3029 kg ha⁻¹ no ano de 2022 segundo dados da CONAB (2022). Em relação a dispersão relativa dos dados, o coeficiente de variação (CV%), se classificou como baixo (CV<12%), conforme Warrick e Nielsen (1980).

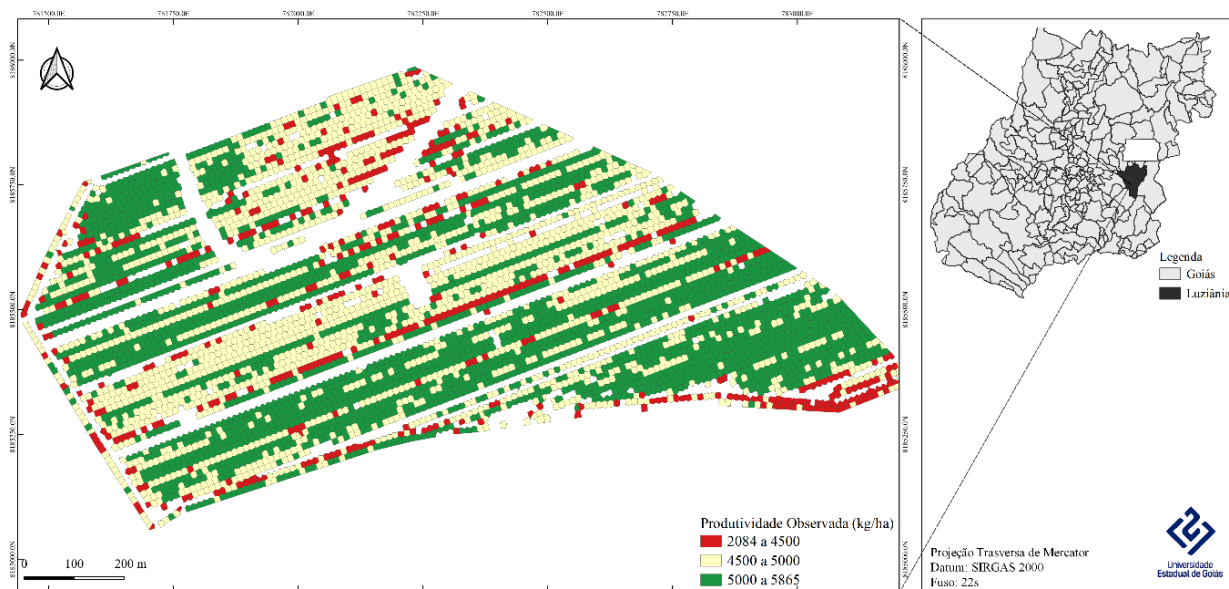
Tabela 2 - Estatística descritiva da produtividade mensurada em campo em tempo real com uso de uma colhedora de soja.

Atributo	Med	MD	Máx	Mín	DP	CV (%)
Produtividade kg ha ⁻¹	4942.70	4982.15	5865.29	2083.86	345.20	6.98

Med= média; MD – mediana; Máx – valor máximo; Mín – valor mínimo; DP – desvio padrão; CV – coeficiente de variação;

Fonte: Autor (2022).

Na Figura 3, é possível observar a distribuição da produtividade na área de estudo ao longo das parcelas. Com produtividade variando dos 2080 aos 5865 kg ha⁻¹. Nota-se que os maiores valores se concentraram nas regiões noroeste e mais ao sul do mapa temático, com valores ultrapassando os 5000 kg ha⁻¹. É possível observar também que o fator produtividade segue um padrão não natural em formato de linhas muito definidas, o que pode ser explicado pelo fato de que as linhas de plantio estão posicionadas nesse mesmo sentido, e o tráfego de tratores para os tratos culturais também ocorreu de forma controlada nesse sentido, e consequentemente causou amassamento das plantas nesses locais, ocasionando blocos de linhas inteiras com produtividade abaixo dos padrões do talhão.

Figura 3 – Distribuição espacial dos dados de produtividade extraídos da colhedora.

Fonte: Autor (2022).

4.2 - MODELO DE PREDIÇÃO DE PRODUTIVIDADE COM USO DE SATELITE

São apresentados na Tabela 3 os resultados dos indicadores estatísticos para estimativa de produtividade da soja utilizando dados de satélite em diferentes estádios fenológicos. De maneira geral, é possível notar que os melhores valores de indicadores foram obtidos com a análise no estágio reprodutivo R5, resultados semelhantes foram encontrados em outros estudos que mostraram que o momento ideal para predição da produtividade vai da floração até o estágio de enchimento de grãos (EUGENIO et al., 2020; ZHANG et al., 2019; GAO et al., 2018).

Tabela 3 - Indicadores estatísticos para estimativa da produtividade em diferentes estádios de desenvolvimento da cultura da soja.

Estádio da análise	r			MAE (kg ha ⁻¹)			MAPE (%)			R ²		
	Tr	Val	Te	Tr	Val	Te	Tr	Val	Te	Tr	Val	Te
V3	0.60	0.52	0.52	207.84	227.65	226.02	4.32	4.83	4.73	0.36	0.27	0.27
R2	0.75	0.68	0.66	175.12	201.93	196.54	3.61	4.21	4.1	0.56	0.46	0.43
R3	0.73	0.69	0.69	181.03	201.74	192.22	3.73	4.17	3.99	0.53	0.46	0.48
R5	0.83	0.76	0.71	142.66	177.79	175.42	2.95	3.67	3.68	0.69	0.57	0.50
R6	0.83	0.74	0.71	144.41	183.45	180.36	2.99	3.82	3.78	0.68	0.54	0.50

Tr- treinamento; Val- validação interna; Te- Teste externo;

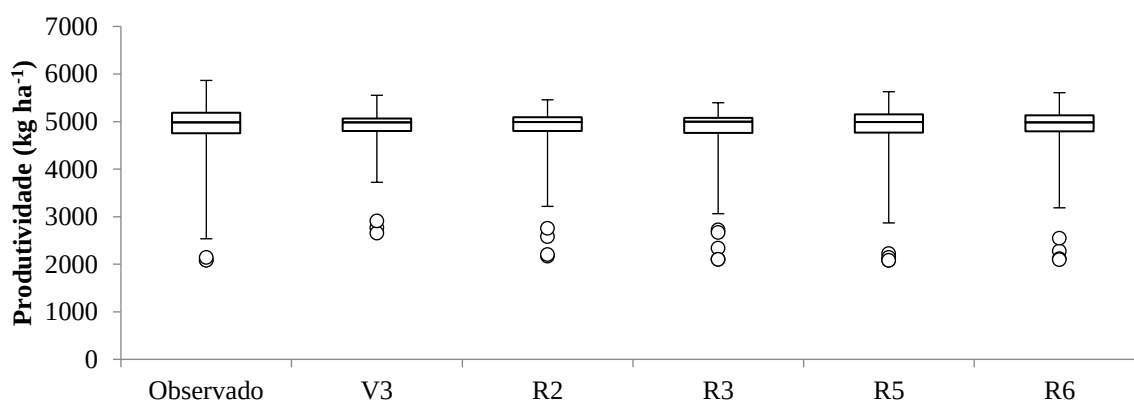
Fonte: Autor (2022).

Conforme mostrado na Tabela 3, os valores de correlação de Pearson e R^2 cresceram, enquanto os valores de MAE e MAPE diminuíram gradativamente com o acréscimo de informações referentes aos estágios fenológicos mais avançados, alcançando os melhores resultados no estágio R5. Isso provavelmente se deve ao fato de que a produtividade de uma lavoura de soja é resultado de um conjunto de fatores agrônômicos que podem se alterar ao longo do ciclo, como fatores climáticos destacados em estudos por BATTISTI (2018), e a sua interação com o manejo durante o desenvolvimento da cultura (EDREIRA et al., 2017).

Além disso, estudos feitos por Gaspar et al. (2017) evidenciam que a soja possui um pico de acúmulo de matéria seca próximo do estágio R4, tendo a partir do estágio R5 uma maior atividade de remobilização de nutrientes para os grãos. O que corrobora com os resultados obtidos, mostrando que o período entre esses dois estágios se configura como momento mais favorável para uma melhor estimativa de produtividade.

A Figura 4 mostra graficamente a distribuição dos valores de produtividade, observados em campo e preditos pelas redes neurais nos diferentes estágios fenológicos. Para todos os estágios, os modelos foram eficientes em prever valores próximos dos valores médios reais, porém é possível observar que o modelo que compreende o estágio R5 foi o que teve o range com valores máximos e mínimos mais próximos dos reais. Estudos anteriores utilizando redes neurais também evidenciaram que o acréscimo de informações tem forte influência na capacidade de generalização e credibilidade dos resultados (LIU et al., 2021; BAHRAMI et al., 2021).

Figura 4 - Produtividades observada e prevista pelos modelos.

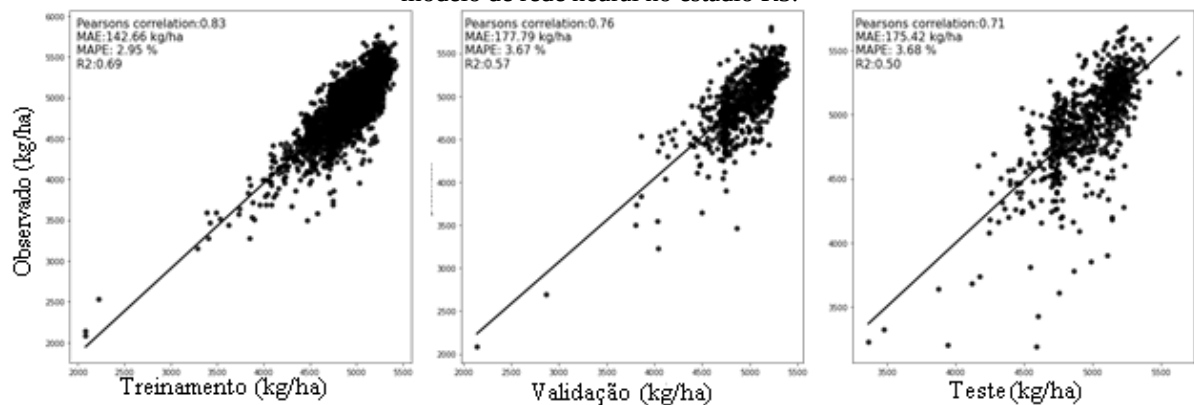


Fonte: Autor (2022).

A produtividade estimada pelo modelo do estágio R5, que teve os melhores indicadores, foi comparada com os valores observados de produtividade em campo utilizando gráfico de dispersão para os grupos de treinamento, validação interna e teste externo (Figura

5). É possível observar a capacidade de generalização do modelo, que mesmo trabalhando com conjunto de dados de teste, onde os dados não passaram pelo treinamento, houve comportamento similar aos do treinamento. Em estudos anteriores, Poley e Mcdermid (2020) destacam a importância do conjunto de dados externo para entender o quão robusto o modelo se apresenta, principalmente quando se deseja aplicá-lo em diferentes pontos do espaço e do tempo.

Figura 5 - Gráfico de dispersão para os dados de produtividade observados em relação aos previstos pelo modelo de rede neural no estágio R5.



Fonte: Autor (2022).

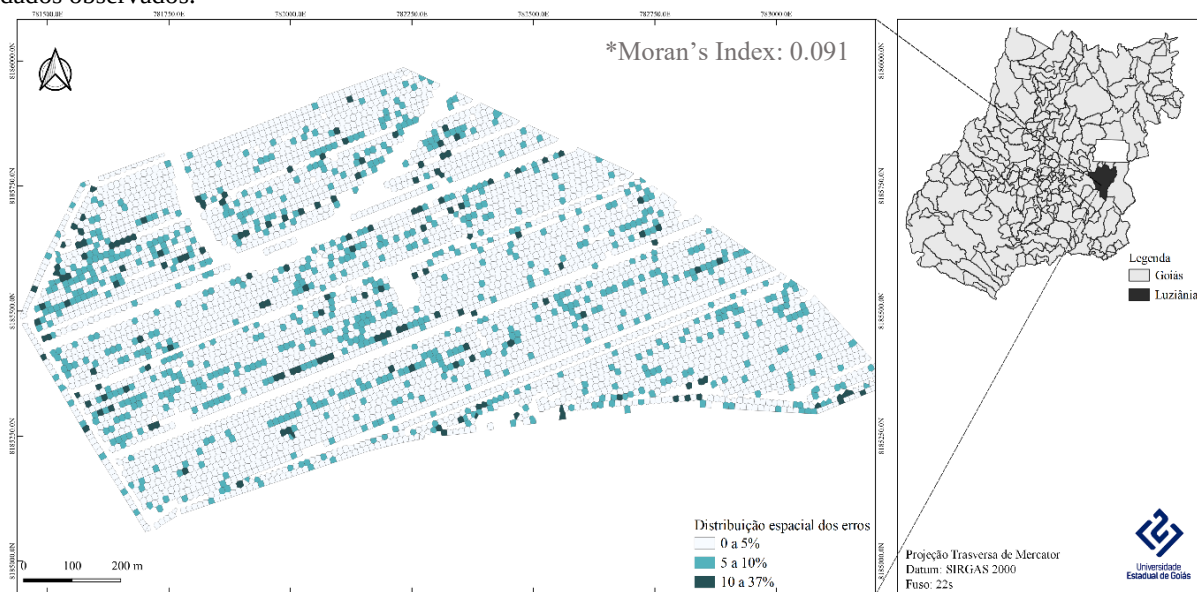
Para o grupo de teste foi encontrado valor de coeficiente de correlação R^2 de 0.50, classificado por Montgomery (2020) como valor de ajuste moderado. Em estudos envolvendo também imagens do satélite PlanetScope e técnicas de aprendizado profundo para estimar produtividade da soja, Sagan et al. (2021) encontraram melhores valores de R^2 , variando de 0,55 a 0,80 em diferentes modelos testados.

É possível observar uma alta concordância entre os valores observados e preditos pelo modelo, tendo em vista valores de correlação de Pearson superiores a 0.70, definida como uma correlação forte (CALLEGARI JACQUES, 2008). A disparidade entre os valores de R^2 e correlação de Pearson encontrada nos resultados pode ter ocorrido pela característica que o R^2 possui de ser enviesado pela não normalidade dos erros ou presença de outliers, comum em grandes conjuntos de dados (G. POLEY, 2020).

No geral o modelo resultou em valores de erro médio absoluto baixos na produtividade estimada, conforme pode ser observado na Figura 6, alcançando valores de MAE e MAPE, 175.42 kg ha⁻¹ e 3.68% respectivamente, para o conjunto de teste. Resultados semelhantes para estimativa de produtividade da soja foram encontrados utilizando sensores proximais, apresentando valores de MAE de 149.28 kg ha⁻¹ (YOOSEFZADEH-NAJAFABADI, 2021), e sensoriamento remoto, obtendo valores de MAPE entre 7% e 9% (LIU, 2021).

Outro fator importante para qualidade do modelo é a capacidade de lidar com a variação e heterogeneidade encontrada nos campos agrícolas devido a fatores ligados a declividade, fertilidade do solo, umidade e outros fatores que afetam a produtividade (RISCHBECK et al., 2016). Com base na figura 6 é possível observar visualmente um padrão de distribuição aleatório dos erros, o que é comprovado pelo Índice Global de Moran calculado de 0.091, o que mostra a alta aleatoriedade na distribuição dos erros.

Figura 6 - Distribuição espacial dos erros de previsão (resíduos de regressão) do modelo R5 comparado com dados observados.



Fonte: Autor (2022).

A fim de verificar a precisão e acurácia do modelo, é apresentado na Tabela 4 os resultados do teste t, comparando valores médios observados em campo frente aos estimados pela rede. É possível concluir que não houve diferença significativa a 5% de probabilidade, apresentado médias muito próximas.

Tabela 4 - Teste-t para os valores observados e estimados da produtividade da soja para o modelo no estágio R5.5.

	Observado	Estimado
Média	4942.70	4949,02
Variância	119193.09	72593,28
Observações		5011
gl		10020
p-value		0.31

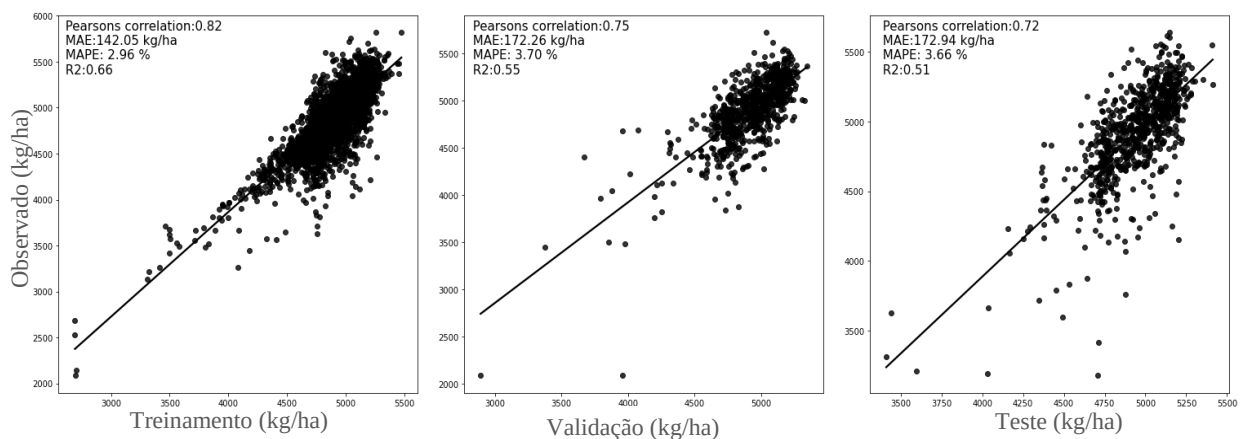
Gl: graus de liberdade; p-value: teste de t-student significativo a 5% de probabilidade.

Fonte: Autor (2022).

4.3 MODELO DE PREDIÇÃO DE PRODUTIVIDADE UTILIZANDO DADOS DE VANT

Os resultados obtidos pela melhor configuração de rede treinada para os dados das imagens de VANT são apresentados na Figura 7. Os padrões e valores dos indicadores estatísticos foram semelhantes, com pequenas melhorias, quando comparado aos encontrados pelo modelo que utilizou dados de satélite no período R5 (Figura 5). Isso pode ser explicado pela proximidade entre as larguras de bandas de ambos os sensores, e também pelo tamanho da parcela utilizada de 12 metros, que em alguns casos pode dissolver a diferença de dados captadas entre as diferentes resoluções espaciais, principal característica responsável pelas diferenças observadas entre plataformas orbitais e aéreas, conforme já evidenciado em estudos anteriores (NONNI et al., 2018; MESSINA et al., 2020).

Figura 7- Gráfico de dispersão para os dados de produtividade observados em relação aos previstos pelo modelo de rede neural com dados do VANT.



Fonte: Autor (2022).

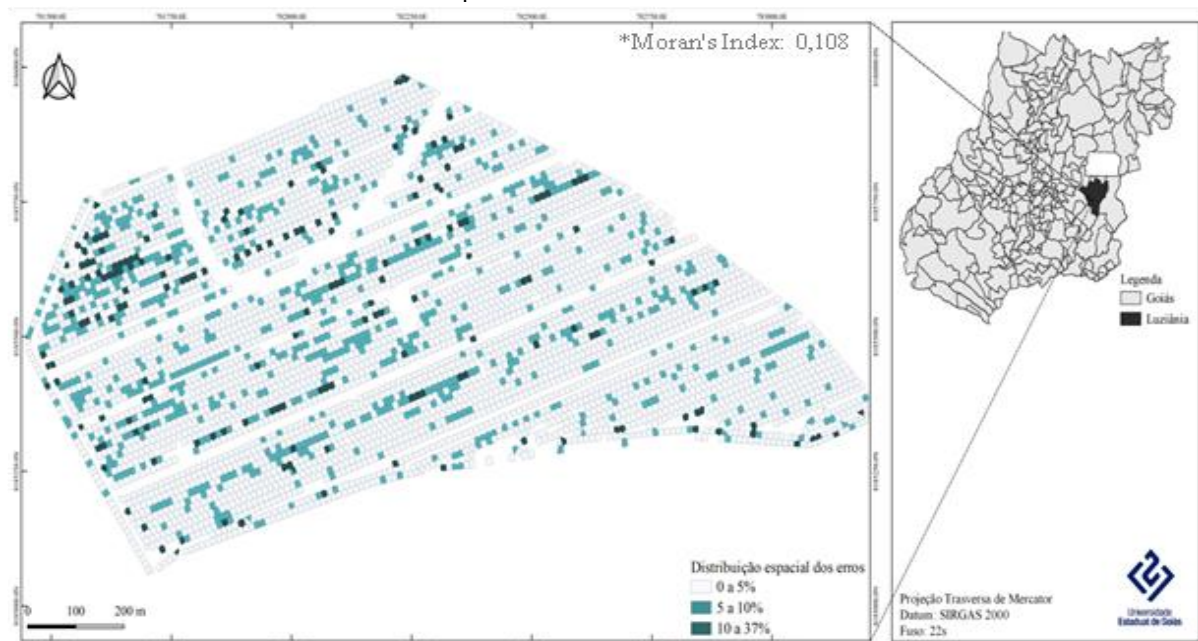
De forma geral os valores estatísticos do teste demonstram um bom ajuste do modelo conforme mostrado na Figura 7. Herrero-Huerta et al. (2020) utilizando dados de sensor multiespectral acoplado em VANT e diferentes modelos de aprendizado de máquina obtiveram resultados inferiores de R^2 , encontrando 0.38 para modelo Random Forest e 0.42 para XGBoost. Já Eugenio et al. (2020) utilizando o mesmo sensor e mesma cultura, porém modelados por redes neurais, tiveram resultados próximos, apresentado correlação de Pearson, 0.70 e 0.9 para os conjuntos de treino e teste respectivamente, entretanto com valores de erro médio absoluto maiores, apresentando $397.75 \text{ kg ha}^{-1}$ e $324.43 \text{ kg ha}^{-1}$ para os conjuntos treino e teste.

Estudos realizados por Maimaitijiang et al. (2020) avaliando modelos para prever produtividade da soja através de imagens de VANT, tiveram resultados superiores, com valores de R^2 variando de 0.51, para modelo utilizando apenas imagens RGB, a 0.72, com acréscimo de informações multiespectrais e de temperatura. Confirmando a capacidade das redes neurais

em trabalhar com grande volume de dados, onde o acréscimo de informações gerou melhorias no modelo. Outro fator que pode explicar o melhor desempenho encontrado no estudo foi a altura de captura de imagens de 30 m, que melhora a resolução espacial das imagens, mas por outro lado dificulta a replicação do modelo em áreas extensas, como é o caso da cultura da soja.

Conforme mostrado na figura 8, para os resultados de previsão de rendimento de soja utilizando VANT, o padrão de distribuição espacial dos erros se mostrou aleatório apresentando um Índice Global de Moran de 0,108. Indicando que o modelo é consistente espacialmente, não mostrando tendências de erros agrupados em determinados locais no espaço amostral.

Figura 8 - Distribuição espacial dos erros de previsão (resíduos de regressão) do modelo de rede neural comparado com dados observados.



Fonte: Autor (2022).

No comparativo entre as médias do valor estimado e o real, observa-se que não houve diferença estatística a 5% de probabilidade entre as mesmas, conforme mostrado na Tabela 5. O resultado demonstra a grande capacidade das redes neurais em correlacionar um grande número de variáveis e obter bons desempenhos na resolução de problemas complexos, apresentando médias do grupo estimado muito próximas do observado, similar ao que ocorreu com o modelo utilizando dados de satélite.

Tabela 5 - Teste-t para os valores observados e estimados da produtividade da soja pelo modelo utilizando imagens de VANT.

	Observado	Estimado
Média	4942.70	4952,96
Variância	119193.09	60176,51
Observações		5011
gl		10020
p-value		0.086

Gl: graus de liberdade; p-value: teste de t-student significativo a 5% de probabilidade.

Fonte: Autor (2022).

5 CONCLUSÃO

O estudo trouxe uma abordagem metodológica relacionada ao uso do sensoriamento remoto via satélite e VANT na modelagem preditiva da produtividade no cultivo de soja. Apesar da produtividade ser um parâmetro complexo para modelagem, a metodologia utilizada no presente trabalho apresentou resultados promissores, podendo fornecer estimativas de produtividade com precisão e robustez, utilizando diferentes plataformas de levantamento.

Os resultados dos indicadores estatísticos mostram que o estágio vegetativo R5 tem o melhor desempenho para prever a produtividade da soja.

Ambos os modelos, tanto utilizando imagens de satélite, como de VANT, não apresentaram indícios de *overfitting*, mostrando alta capacidade em trabalhar com dados externos ao treinamento, tendo boa estabilidade e generalização.

A modelagem feita a partir de imagens de satélite e VANT tiveram resultados muito próximos, tendo a possibilidade de escolha da plataforma com base em aspectos econômicos de contexto local, como por exemplo custos de levantamento, disponibilidade do equipamento, flexibilidade do clima no estágio ideal de análise, tamanho da área a ser levantada, entre outros.

REFERÊNCIAS

- ALMEIDA, L. S.; GUIMARÃES, E. C. Agricultura de precisão: um alerta à imprecisão – estudo de caso na cafeicultura. **Revista de Agricultura Neotropical**, Cassilândia, v. 4, n. 2, p. 41-48, 2017.
- ALVES, G.R.; TEIXEIRA, I.R.; MELO, F.R.; SOUZA, R.T.G.; SILVA, A.G. Estimating soybean yields with artificial neural networks. **Acta Scientiarum**, Maringá v. 40, e35250, 2018.
- ANAR, M.J.; LIN, Z.; HOOGENBOOM, G.; SHELIA, V.; BATCHELOR, W.D.; TEBOH J. M.; KHAN, M. Modeling growth, development and yield of Sugarbeet using DSSAT. **Agricultural systems**, Fargo, v. 169, p. 58-70, 2019.
- APAN, A.; HELD, A.; PHINN, S.; MARKLEY, J. Formulation and assessment of narrow-band vegetation indices from EO-1 Hyperion imagery for discriminating sugarcane disease. In: **Proceedings of the Spatial Sciences Institute Biennial Conference** (SSC 2003): Spatial Knowledge Without Boundaries. Spatial Sciences Institute, 2003. p. 1-13.
- BAHRAMI, H.; HOMAYOUNI, S.; ABDOLREZA, S.; MIRZAEI, S.; MAHDIANPARI, M.; REISI-GAHROUEI, O. Deep learning-based estimation of crop biophysical parameters using multi-source and multi-temporal remote sensing observations. **Agronomy**, v.11, n.7, p.1363, 2021.
- BANNARI, A.; MORIN, D.; BONN, F.; HUETE, A. A review of vegetation indices. **Remote sensing reviews**, v. 13, n. 1-2, p. 95-120, 1995.
- BATISTA, R.; SENTELHAS, P.C.; PASCOALINO, J.A.L.; SAKO, H.; DANTAS, J.P.S.; MORAES, M.F. Soybean yield gap in the areas of yield contest in Brazil. **International Journal of Plant Production**, v.12, n.3, p. 159-168, 2018.
- BENINCASA, P.; ANTOGNELLI, S.; BRUNETTI, L.; FABBRI, C.; NATALE, A.; SARTORETTI, V.; MODEO, G.; GUIDUCCI, M.; TEI, F.; VIZZARI, M. Reliability of NDVI derived by high resolution satellite and UAV compared to in-field methods for the evaluation of early crop N status and grain yield in wheat. **Experimental Agriculture**, v.54, n.4, p.604-622, 2018.
- BERNARDI, A. C. C.; NAIME, J. M.; RESENDE A. V.; BASSOI, L. H.; INAMASU, R. Y. **Agricultura de Precisão: Resultados de um novo olhar**. Brasília: Embrapa, 2014. 600p.
- BRAGA, A.P.; CARVALHO, A.P.L.F.; LUDERMIR, T.B. **Redes Neurais Artificiais: Teorias e Aplicações**. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., 2014. 226p.
- BROCHU, E.; CORA, V.M.; DE FREITAS, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. **arXiv preprint arXiv:1012.2599**, 2010.
- CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2008. 255 p.

CASTRIGNANO, A.; BUTTAFUOCO, G.; KHOSLA, R.; MOUAZEN, A.M.; MOSHOU, D.; NAUD, O. **Agricultural internet of things and decision support for precision smart farming**. Academic Press, 2020. 459p.

CATALÃO J.P.D.S.; POUSINHO H.M.I.; MENDES V.M.F. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. **Renewable Energy**, Covilha, v.36, n.4, p.1245-1251, 2011.

CONAB. Companhia Nacional de Abastecimento. **Boletim da safra de grãos**. 2022. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-degraos>>. Acesso em 12 de agosto de 2022.

CONAB. **Perspectivas para a agropecuária**. Companhia Nacional de Abastecimento, Brasília, v.10, p.12–12, 2022.

DE MORAES, E. C. **Capítulo 1 fundamentos de sensoriamento remoto**. 2002.

DEFANTE, L.R., VILPOUX, O.F., SAUER, L. Rapid expansion of sugarcane crop for biofuels and influence on food production in the first producing region of Brazil. **Food Policy**, v.79, p.121-131, 2018.

DYRMANN, M.; KARSTOFT, H.; MIDTIBY, H.S. Plant species classification using deep convolutional neural network. **Biosystems engineering**, v.151, p.72-80, 2016.

EDREIRA, J.I.R., MOURTZINIS, S., CONLEY, S.P., ROTH, A.C., CIAMPITTI, I.A., LICHT, M.A., KANDEL, H.; KYVERYGA, P.M.; LINDSY, L.E.; MUELLER, D.S.; NAEVE, S.L.; NAFZIGER, E.; SPECHT, J.E.; STANLEY, J.; STATON, M.J.; GRASSINI, P. Assessing causes of yield gaps in agricultural areas with diversity in climate and soils. **Agricultural and Forest Meteorology**, v.247, p.170–180, 2017.

EUGENIO, F.C.; GROHS, M.; VENÂNCIO, L.P.; SCHUH, M.; BOTTEGA, E.L.; RUOSO, R.; SCHONS, C.; MALLMANN, C.L.; BADIN, T.L.; FERNANDES, P. Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. Remote Sensing Applications: **Society and Environment**, v.20, p. 100397, 2020.

FEHR, W.R., CAVINESS, C.E. **Stages of soybean development**. Ames: Iowa State University of Science and Technology, p.15, 1977.

FEHR, W.R.; CAVINESS, C.E.; BURMOOD, D.T.; PENNINGTON, J.S. Stage of development descriptions for soybeans, Glycine Max (L.) Merrill 1. **Crop science**, v.11, n.6, p.929-931, 1971.

FERNANDES, M.M.H.; COELHO, A.P.; FERNANDES, C.; SILVA, M.F.; MARTA, C.C.D. Estimation of soil organic matter content by modeling with artificial neural networks. **Geoderma**, v.350, p.46-51, 2019.

FERNANDES, M.M.H.; COELHO, A.P.; SILVA, M.F.; BERTONHA, R.S.; QUEIROZ, R.F.; FURLANI, C.E.A.; FERNANDES, C. Estimation of soil penetration resistance with standardized moisture using modeling by artificial neural networks. **Catena**, Países Baixos, v.189, p.104-505, 2020.

FILHA, C.S.T.; WALTHALL, C.L.; KIM, M.S.; COLSTOUN, E.B.; MCMURTREY, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. **Remote sensing of Environment**, v. 74, n. 2, p. 229-239, 2000.

FINGER, R.; SWINTON, S. M.; BENNI, N.EL. Precision Farming at the Nexus of Agricultural Production and the Environment. **Annual Review of Resource Economics**, v.11, n.1, p.313-335, 2019.

FINGER, R.; SWINTON, S.M.; BENNI, N; WALTER, A. Precision farming at the nexus of agricultural production and the environment. **Annual Review of Resource Economics**, Zürich, v.11, p.313-335, 2019.

FLECK, L.; TAVARES, M.H.F.; EYNG, E.; HELMANN, A.C.; ANDRADE, M.A.M. Redes neurais artificiais: princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**, Medianeira, v.1, n.13, p.47-57, 2016.

FORMAGGIO, A.R.; SANCHES, I.D.A. **Sensoriamento remoto em agricultura**. Oficina de Textos, 2017. 49p.

GAO, F.; ANDERSON, M.; DAUDHTRY, C.; JOHNSON, D. Assessing the variability of corn and soybean yields in central Iowa using high spatiotemporal resolution multi-satellite imagery. **Remote Sensing**, v. 10, n.9, p. 1489, 2018.

GASPAR, A.P.; LABOSKI, C.A. M.; NAEVE, S.L.; CONLEY, S.P. Dry matter and nitrogen uptake, partitioning, and removal across a wide range of soybean seed yield levels. **Crop Science**, v. 57, n. 4, p. 2170–2182, 2017.

GEBBERS, R.; ADAMCHUK, V. I. Precision agriculture and food security. **Science**, v. 327, n. 5967, p. 828-831, 2010.

GITELSON, A.A.; KAUFMAN, Y.J.; STARK, R.; RUNDQUIST, D. Novel algorithms for remote estimation of vegetation fraction. **Remote sensing of Environment**, v. 80, n. 1, p. 76-87, 2002.

GITELSON, A.A.; KEYDAN, G.P.; MERZLYAK, M.N. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. **Geophysical research letters**, v.33, n.11, 2006.

GITELSON, A.A.; MERZLYAK, M.N. Remote estimation of chlorophyll content in higher plant leaves. **International Journal of Remote Sensing**, v.18, n.12, p.2691-2697, 1997.

HABOUDANE, D.; MILLER, J.R.; TREMBLAY, N.; ZARCO-TEJADA, P.J.; DEXTRAZE, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. **Remote sensing of environment**, v.81, n.2-3, p.416-426, 2002.

HASSAN, M.A.; YANG, M.; RASHEED, A.; JIN, X.; XIA, X.; XIAO, Y.; ELE, Z. Time series multispectral indices from unmanned aerial vehicle imagery reveal senescence rate in bread wheat. **Remote Sensing**, v.10, n.6, p.809, 2018.

HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre: Bookman, 2001, 900 p.

HAYKIN, S. Multilayer perceptrons. **Neural networks: a comprehensive foundation**, v.2, p. 156-255, 1999.

HEINEMANN, A.B.; STONE, L.F.; SILVA, S.C. Modelos de simulação do crescimento, desenvolvimento e produtividade na pesquisa agronômica. **Embrapa Arroz e Feijão**, Santo Antônio de Goiás, v.1, p. 26, 2010.

HERRERO-HUERTA, M.; RODRIGUEZ-GONZALVEZ, P.; RAINEY, K.M. Yield prediction by machine learning from UAS-based multi-sensor data fusion in soybean. **Plant Methods**, v.16, n.1, p. 1-16, 2020.

HERUBIN, M.R.; SANTI, A.L.; BASSO, C.J.; EITELWEIN, M.T.; VIAN, A.L. **Caracterização e estratégias de manejo da variabilidade espacial dos atributos químicos do solo utilizando a análise dos componentes principais**. Enciclopédia Biosfera, v.7, p.196-210, 2011.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.P.; GAO, X.; FERREIRA, LG. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote sensing of environment**, v. 83, n. 1-2, p. 195-213, 2002.

IMRAN, M.; STEIN, A.; ZURITA-MILLA, R. Using geographically weighted regression kriging for crop yield mapping in West Africa. **International journal of geographical information science**, v.29, n.2, p.234-257, 2015.

INTERNATIONAL SOCIETY FOR PRECISION AGRICULTURE - ISPA. **Precision Ag Definition**. 2019. Disponível em:< <https://www.ispag.org/about/definition>>. Acesso em: 02 outubro. 2022.

IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: **International conference on machine learning**. PMLR, p. 448-456, 2015.

JEONG, J.H.; RESOP, J.P.; MUELLER, N.D.; FLEISHER, D.H.; YUN, K.; BUTLER, E.E.; TIMLIN, D.J.; SHIM, K.M.; GERBER, J.S.; REDDY, V.R.; KIM, S.H. Random forests for global and regional crop yield predictions. **Plos One**, Seattle, v.10, n.621, p.51–59, 2019.

JHA, k.; DOSHI B, A.; PATEL, P.; SHAH, M. A comprehensive review on automation in agriculture using artificial intelligence. **Artificial Intelligence in Agriculture**, v.2, p.1-12, 2019.

JIANG, Z.; HUETE, A.R.; DIDAN, K.; MIURA, T. Development of a two-band Enhanced vegetation index without a blue band. **Remote sensing of Environment**, v. 112, n. 10, p. 3833-3845, 2008.

KERAS, 2022. **Keras: The Python Deep Learning Library**. Disponível em:< <https://keras.io/>. Acesso em: 2022

KHANAL, S.; FULTON, J.; SHEARER, S. An overview of current and potential applications of thermal remote sensing in precision agriculture. **Computers and Electronics in Agriculture**, v. 139, p. 22-32, 2017.

KHOSHROO A.; EMROUZNEJAD A.; GHAFARIZADEH A.; KASRAEI M.; OMID M. Sensitivity analysis of energy inputs in crop production using artificial neural networks. **Journal of Cleaner Production**, Yasouj, v.197, p.992-998, 2018.

KHOSHROO, A.; EMROUZNEJAD, A.; GHAFARIZADEH, A.; KASRAEI, M.; OMID, M. Sensitivity analysis of energy inputs in crop production using artificial neural networks. **Journal of cleaner production**, v.197, p. 992-998, 2018.

LECUN, Y.; BENGIO, Y.; HINTON, G.. Deep learning. **nature**, v.521, n.7553, p.436-444, 2015.

LI, H.; WAN, W.; FANG, Y.; ZHU, S.; CHEN, X.; LIU, B.; HONG, Y. A Google Earth Engine-enabled software for efficiently generating high-quality user-ready Landsat mosaic images. **Environmental Modelling & Software**, Beijing, v. 112, p. 16-22, 2019.

LINSLEY, C. M.; BAUER, F. C. **Test your soil for acidity**. Circular, University of Illinois, Agricultural Experiment Station; n.346, 1929.

LIU J.G.C.E.; TIAN, L. A neural network for setting target corn yields. **Transactions of the ASAE**, Setor Joseph, v.44, n.3, p.705, 2013.

LIU, Y.; KIM, J.; FLEISHER, D.H.; KIM, K.S. Analogy-Based Crop Yield Forecasts Based on Temporal Similarity of Leaf Area Index. **Remote Sensing**, v.13, n.16, p.3069, 2021.

MACDONALD, R. B.; HALL, F. G.; ERB, R. B. **The use of Landsat data in a large area crop inventory experiment (LACIE)**. LARS Symposia. 1975. 46p.

MACHADO, P.L.O.A.; BERNARDI, A.C.C.; SILVA, C.A. **Agricultura de precisão para o manejo da fertilidade do solo em sistema plantio direto**. Rio de Janeiro: Embrapa Solos, 2004. 209p.

MAIMAITIJIANG, M.; SAGAN, V.; SIDDIKE, P.; HARTLING, S.; ESPOSITO, F.; FRITSCHI, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote sensing of environment**, v. 237, p.111599, 2020.

MAIN, R.; CHO, M.A.; MATHIEU, R.; O'KENNEDY, M.M.; RAMOELO, A.; KOCH, S. Na investigation into robust spectral indices for leaf chlorophyll estimation. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 6, p. 751-761, 2011.

MALDANER, L.F.; MOLIN, J.P.; SPEKKEN, M. Methodology to filter out outliers in high spatial density data to improve maps reliability. **Scientia Agricola**, Piracicaba, v.79, n.1, 2021.
MCCULLOCH W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115-133, 1943.

MESSINA, G.; PEÑA, J.M.; VIZZARI, M.; MODICA, G. A comparison of UAV and satélites multispectral imagery in monitoring onion crop. An application in the 'Cipolla Rossa di Tropea' (Italy). **Remote Sensing**, v.12, n.20, p.3424, 2020.

MICHELON G. K.; ENEZES P.L.; BAZZI C.L.; JASSE E.P.; MAGALHÃES P.S.; BORGES L.F. Artificial neural networks to estimate the productivity of soybeans and corn by chlorophyll readings. **Journal of Plant Nutrition**, Medianeira, v.41, n.10, p.1285-1292, 2018.

MOLIN J.P.; AMARAL L.R.; COLAÇO, A.F. **Agricultura de precisão**. São Paulo: Oficina de Textos, 2015. p.119-153.

MOLIN, J. P.; AMARAL, L. R.; COLAÇO, A. F. **Agricultura de Precisão**. São Paulo: Oficina de Textos, 2015. 238p.

MOLIN, J.P. Agricultura de precisão. In: MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. F. **Agricultura de precisão: o gerenciamento da variabilidade**. Piracicaba: O Autor, 2001. 83p.

MONTEIRO, A.L.; SOUZA, M.F.; LINS, H.A.; TEÓFILO, T.M.S.; BARROS JÚNIOR, A.P.; SILVA, V.D.; MENDONÇA, V. A new alternative to determine weed control in agricultural systems based on artificial neural networks (ANNs). **Field Crops Research**, v.263, p.108075, 2021.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. John Wiley & Sons, Ed.10, 2020.

MORAN, P.A.P. Notes on continuous stochastic phenomena. **Biometrika**, v.37, n. 1/2, p.17-23, 1950.

MOREIRA, M. A. (2011). **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 4 ed. atual. e ampl. Viçosa, MG: Ed. UFV. 422p.

NOETZOLD, R.; CARVALHO A.M.; JÚNIOR, M.M.G.; GOUSSAIN, R.D.C.S. Variabilidade espacial da eficiência do uso de potássio e fósforo na cultura da soja. **Revista Engenharia na Agricultura**, Viçosa, v.27, n.6, p.529- 541, 2019.

NONNI, F.; MALACARNE, D.; PAPPALARDO, S.; CODATO, D.; MEGGIO, F.; DE MARCHI, M. Sentinel-2 Data Analysis and Comparison with UAV Multispectral Images for Precision Viticulture. **GI Forum**, p. 105-116, 2018.

PIERCE, F.J.; NOWAK, P. Aspects of precision agriculture. **Advances in Agronomy**, v.67, p. 1-85, 1999.

POLEY, G.L.; J. MCDERMID, G. A systematic review of the factors influencing the estimation of vegetation aboveground biomass using unmanned aerial systems. **Remote Sensing**, v.12, n.7, p.1052, 2020.

RAPER, T.B.; VARCO, J.J. Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status. **Precision Agriculture**, v.16, n.1, p.62-76, 2015.

RAYMOND HUNT JR., E.; DAUGHTRY, C.S.T.; EITEL, J.U.H.; LONG, D.S. Remote sensing leaf chlorophyll content using a visible band index. **Agronomy journal**, v.103, n.4, p.1090-1099, 2011.

RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. **Remote sensing of environment**, v. 55, n. 2, p. 95-107, 1996.

ROUSE, J. W., HAAS, R. H., SCHELL, J. A., & DEERING, D. W. Monitoring vegetation systems in the great plains with ERTS. **NASA Spec. Publ**, v. 351, n. 1, p. 309, 1974.

SAGAN, V.; MAIMAITIJANG, M.; BHADRA, S.; MAIMAITIYIMING, M.; BRWN, D.R.; SIDIKKEE, P.; FRITSCHI, F.B. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. **ISPRS journal of photogrammetry and remote sensing**, v. 174, p. 265-281, 2021.

SAMBORSKA, I.A.; ALEXANDROV, V.; SIECZKO, L.; KORATOWSKA, B.; GOLTSEV, V.; CETNER, M.D.; KALAJI, H.M. Artificial neural networks and their application in biological and agricultural research. **Jornal NanoPhotoBioSciences**, Warsaw, v.2, p.14-30, 2014.

SEDIYAMA, T. **Tecnologias de produção de sementes de soja**. Londrina: Mecenass, p.9, 2013.

SHRESTHA, Y.R.; KRISHNA, V.; KROGH, G.V. Augmenting organizational decision making with deep learning algorithms: Principles, promises, and challenges. **Journal of Business Research**, New York v.123, p.588-603, 2021.

SOARES, G.M.; SILVA, L.D.; HIGA, S.A.R.; SIMON, A.A.; SÃO JOSÉ, F.B. Artificial neural networks (ANN) for height estimation in a mixed-species plantation of *Eucalyptus globulus* labill and *Acacia mearnsii* de wild. **Revista Árvore**, v.45, p.4512, 2021.

SPEKKEN, M.; ANSELM, A.A.; MOLIN, J.P. **A simple method for filtering spatial data. European Conference of Precision Agriculture, Leida**. Spain: Precision Agriculture 13, Wageningen, 2013. p. 259-266.

SRIVASTAVA P.; SINGH R.; BHADOURIA R.; TRIPATHI S.; SINGH H. Understanding Soil Aggregate Dynamics and Its Relation With Land Use and Climate Change. Em: **Climate Change and Agricultural Ecosystems**, Ed. Woodhead Publishing, p. 331-354, 2019.

TANWAR, B., GOYAL, A. **Oilseeds: Health Attributes and Food Applications**. Springer, 2021.

TAYLOR, J.A.; MCBRATNEY, A.B.; WHELAN, B.M. Establishing management classes for broadacre agricultural production, **Agronomy Journal**, Madison, v.99, n.5, p.1366-1376, 2007.

TMUŠIĆ, G.; MANFREDA, S.; AASEN, H.; JAMES, M. R.; GONÇALVES, G.; BEN-DOR, E.; BROOK, A.; POLINOVA, M.; ARRANZ, J. J.; MÉSZÁROS, J.; ZHUANG, R.; JOHANSEN, K.; MALBETEAU, Y.; LIMA, I. P.; DAVIDS, C.; HERBAN, S.; MCCABE, M. F. Current Practices in UAS-based Environmental Monitoring. **Remote Sensing**, Basel, v.12, n.6, p.1001-1035, 2020.

TORRES-SÁNCHEZ, J.; PEÑA, J.M.; CASTRO, A.I.; LÓPEZ-GRANADOS, F. Multi temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. **Computers and Electronics in Agriculture**, v. 103, p. 104-113, 2014.

TUCKER, C.J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote sensing of Environment**, v.8, n.2, p.127-150, 1979.

TUCKER, C.J.; GARRATT, M.W. Leaf optical system modeled as a stochastic process. **Applied optics**, v.16, n.3, p.635-642, 1977.

VIAN, A. L.; SANTI, A. L.; AMADO, T. J. C.; CHERUBIN, M. R.; SIMON, D. H.; DAMIAN, J.M.; BREDEMEIER, C. Variabilidade espacial da produtividade de milho irrigado e sua correlação com variáveis explicativas de planta. **Ciência Rural**, v.46, n.3, p.464-471, 2016.

WARRICK, A.W.; NIELSEN D.R. Spatial variability of soil physical properties in the field. In: HILLEL, D. **Applications of soil physics**. New York: Academic Press, p. 319-344, 1980.

YANG, X. K.; QU, D.; ZHANG, W.L.; ZHANG, W.Q. An adapted data selection for deep leaning-based audio segmentation in multi-genre broadcast channel. **Digital Signal Processing**, Kidlington, v.81, p.8-15, 2018.

YOOSEFZADEH-NAJAFABADI, M.; TULPAN, D.; ESKANDARI, M. Using hybrid artificial intelligence and evolutionary optimization algorithms for estimating soybean yield and fresh biomass using hyperspectral vegetation indices. **Remote Sensing**, v.13, n.13, p. 2555, 2021.

ZHANG, X.; ZHAO, J.; YANG, G.; LIU, J.; CAO, J.; LI, C.; ZHAO, X.; GAI, J.. Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. **Remote Sensing**, v.11, n.23, p. 2752, 2019.

APÊNDICES

Apêndice A - Lista de variáveis de entrada utilizadas para predição da produtividade, a partir de dados de satélite.

Feature	Fórmula	Referência
Reflectância bandas R, G, B, N	média, mediana, Desvio Padrão e Variância	//
Visible Atmospherically Resistant Index Green	$VARI = (G-R)/(G+R-B)$	(GITELSON et al., 2004)
Enhanced Vegetation Index	$EVI = 2.5 * (NIR - R) / (1 + NIR - 2.4 * R)$	(HUETE et al., 2002)
Enhanced Vegetation Index 2	$EVI2 = 2.5 * (NIR - R) / (NIR + 2.4 * R + 1)$	(JIANG et al., 2008)
Green chlorophyll index	$GCI = (NIR/G) - 1$	(GITELSON et al., 2005)
Green Normalized Difference Vegetation Index	$GNDVI = (NIR - G) / (NIR + G)$	(HUNT JR et al., 2011)
Simple Ratio NIR/G Green Ratio Vegetation Index	$GRVI = (G - R) / (G + R)$	(TUCKER 1979)
Difference NIR/Green Green Difference Vegetation Index	$GDVI = N - G$	(TUCKER 1979)
Simple Ratio 550/680 Disease-Water Stress Index 4	$DSWI-4 = G/R$	(APAN et al., 2003)
Green-Blue NDVI	$GBNDVI = Nir - (G + B)$	(WANG et al., 2007)
Chlorophyll vegetation index	$CVI = N * (R / (G^2))$	(HUNT JR et al., 2011)
Normalized difference vegetation index	$NDVI = (NIR - R) / (NIR + R)$	(ROUSE et al., 1974)
Optimized soil adjusted vegetation index	$OSAVI = (NIR - R) / (NIR - R + L) (L = 0.16)$	(RONDEAUX et al., 1996)
Modified NDVI	$mNDVI = (N - R) / (N + R - 2 * B)$	(MAIN et al., 2011)
Simple Ratio 800/670 Ratio Vegetation Index	$RVI = NIR/R$	(BANNARI et al., 1995)
mCRIG	$mCRIG = (B^{(-1)}) - (G^{(-1)}) * N$	(GITELSON et al., 2011)
Normalized Difference 550/450 Plant pigment ratio	$PPR = (G - B) / (G + B)$	(GITELSON et al., 2011)
Green leaf index	$GLI = ((2 * G) - R - B) / ((2 * G) + R + B)$	(HUNT JR et al., 2011)
Wide Dynamic Range Vegetation Index	$WDRVI = (a * NIR - R) / (a * NIR + R) (a = 0.12)$	(GITELSON et al., 2004)
Altitude	//	//

Apêndice B - Lista de variáveis de entrada utilizadas para predição da produtividade, a partir de dados de VANT.

Feature	Formula	Referencia
Reflectância bandas R,G,N, RE	Média, mediana, Desvio Padrão e Variância	//
Area Foliar	$AF = (N^{\circ} \text{ de pixels soja}) / (\text{número de pixels total})$	(TORRES-SANCHEZ et al., 2014)
Enhanced Vegetation Index	$EVI = 2.5 * (NIR - R) / (1 + NIR - 2.4 * R)$	(HUETE et al., 2002)
Enhanced Vegetation Index 2	$EVI2 = 2.5 * (NIR - R) / (NIR + 2.4 * R + 1)$	(JIANG et al., 2008)
Wide Dynamic Range Vegetation Index	$WDRVI = (a * NIR - R) / (a * NIR + R) (a = 0.12)$	(GITELSON et al., 2004)
Red-edge chlorophyll index	$RECI = (NIR/RE) - 1$	(GITELSON et al., 2005)
Green chlorophyll index	$GCI = (NIR/G) - 1$	(GITELSON et al., 2005)
Green Normalized Difference Vegetation Index	$GNDVI = (NIR - G) / (NIR + G)$	(HUNT JR et al., 2011)
Simple Ratio NIR/G Green Ratio Vegetation Index	$GRVI = (G - R) / (G + R)$	(TUCKER 1979)
Normalized difference vegetation index	$NDVI = (NIR - R) / (NIR + R)$	(ROUSE et al., 1974)
Normalized difference red-edge	$NDRE = (NIR - RE) / (NIR + RE)$	(GITELSON and MERZLYAK, 1997)
Normalized difference red-edge index	$NDREI = (RE - G) / (RE + G)$	(HASSAN et al., 2018)
Simplified canopy chlorophyll content index	$SCCCI = NDRE / NDVI$	(RAPER and VARCO, 2015)
Optimized soil adjusted vegetation index	$OSAVI = (NIR - R) / (NIR - R + L) (L = 0.16)$	(RONDEAUX et al., 1996)
Simple Ratio Vegetation Index	$RVI = NIR/R$	(BANNARI et al., 1995)
Transformed chlorophyll absorption in reflectance index	$TCARI = 3 * ((RE - R) - 0.2 * (RE - G) * (RE/R))$	(HABOUDANE et al., 2002)
Modified chlorophyll absorption in reflectance index	$MCARI = ((RE - R) - 0.2 * (RE - G)) * (RE/R)$	(DAUGHTRY et al., 2000)
MCARI/OSAVI	$MO = MCARI/OSAVI$	(DAUGHTRY et al., 2000)
TCARI/OSAVI	$TO = TCARI$	(DAUGHTRY et al., 2000)
Altitude	//	//

Apêndice C - Arquitetura da rede neural utilizada para modelagem: estrutura da rede utilizada para imagens de satélite (a); estrutura da rede utilizada para imagens de VANT (b).

