UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE ENGENHARIA
CAMPUS DE ILHA SOLTEIRA

CELÍ SANTANA SILVA

USO DE VARIÁVEIS MULTI E HIPERESPECTRAIS PARA PREDIÇÃO DE CARACTERES AGRONÔMICAS EM SOJA EM DIFERENTES ESTÁGIOS FENOLÓGICOS

Ilha Solteira
2024

PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CELÍ SANTANA SILVA

USO DE VARIÁVEIS MULTI E HIPERESPECTRAIS PARA PREDIÇÃO DE VARIÁVEIS AGRONÔMICAS EM SOJA EM DIFERENTES ESTÁGIOS FENOLÓGICOS

Tese apresentada à Faculdade de Engenharia de Ilha Solteira – Unesp como parte dos requisitos para obtenção do título de Doutor. Especialidade: Produção vegetal.

Paulo Eduardo Teodoro
Orientador

Ilha Solteira
2024

FICHA CATALOGRÁFICA
Desenvolvida pela Diretoria Técnica de Biblioteca e Documentação

Silva, Celí Santana.
S586u Uso de variáveis multi e hiperespectrais para predição de caracteres agronômicas em soja em diferentes estágios fenológicos / Celí Santana Silva Silva. -- Ilha Solteira: [s.n.], 2024
56 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Sistemas de Produção, 2024

Orientador: Paulo Eduardo Teodoro
Inclui bibliografia

1. Sensores remotos. 2. Bandas espectrais. 3. Inteligência computacional. 4. Melhoramento de plantas.

Elaborada por Raiane da Silva Santos - CRB-8/9999

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Ilha Solteira

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: USO VARIÁVEIS MULTI E HIPERESPECTRAIS PARA PREDIÇÃO DE CARACTERES AGRONÔMICOS EM SOJA EM DIFERENTES ESTÁGIOS FENOLÓGICOS

AUTOR: CELÍ SANTANA SILVA

ORIENTADOR: PAULO EDUARDO TEODORO

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia, especialidade: Sistemas de Produção pela Comissão Examinadora:

Prof. Dr. PAULO EDUARDO TEODORO (Participaçao Virtual)
Universidade Federal de Mato Grosso do Sul - UFMS

Profa. Dra. RITA DE CÁSSIA FÉLIX ALVAREZ (Participaçao Virtual)
Departamento de Fitotecnia / Universidade Federal de Mato Grosso do Sul - UFMS

Prof. Dr. CARLOS ANTONIO DA SILVA JUNIOR (Participaçao Virtual)
Departamento de Geografia / Universidade do Estado de Mato Grosso - UNEMAT

Prof. Dr. JORGE GONZÁLEZ AGUILERA (Participaçao Virtual)
Departamento de Fitotecnia / Universidade Estadual do Mato Grosso do Sul - UEMS

Prof. Dr. ALAN MARIO ZUFFO (Participaçao Virtual)
Departamento de Fitotecnia / Universidade Estadual do Maranhão - UEMA

Ilha Solteira, 15 de outubro de 2024

gov.br Documento assinado digitalmente
PAULO EDUARDO TEODORO
Data: 15/10/2024 12:58:32-0300
Verifique em https://validar.iti.gov.br

DEDICATÓRIA

Dedico esta obra a Jesus Cristo, meu único Senhor e Salvador da minha vida a quem tenho fé e esperança de que um dia voltará para nos buscar com poder e grande glória e morarmos eternamente juntos na Nova Jerusalém Celestial, aquém preparou para os que O amam. Dedico a minha amada esposa Aline Carvalho de Menezes Santana, que sempre esteve ao meu lado nesse processo evolutivo de minha vida. Dedico também a minha filha linda, Débora Menezes Silva Santana, herança do Senhor Jesus em nossas vidas. Dedico aos meus pais, João Neto da Silva e Maria Aparecida Santana por toda educação e ensinamentos que tem me transmitido nesta jornada da vida. Dedico ao professor doutor Paulo Eduardo Teodoro por todos os ensinamentos com maestria, dedicação e atenção nesse processo intelectual de minha vida.

AGRADECIMENTO

Agradeço a Jesus Cristo pela vida eterna ao seu lado, pela vida abundante em Sua presença, por toda proteção, livramentos e por todas as promessas que tem feito, faz e irá fazer em nossas vidas. Agradeço a minha esposa amada Aline Carvalho de Menezes Santana que sempre me incentivou desde o início desta caminhada mesmo em meio as dificuldades em um período que o mundo parou devido a pandemia mesmo estando gravida de nossa princesa Débora. Agradeço a nossa filha Débora Menezes Silva Santana que é herança de Jesus em nossas vidas, que veio nesse mundo por um propósito de Deus nos alegrando e nos ensinando o que é amar, alegria e cuidado. Agradeço aos meus pais João Neto da Silva e Maria Aparecida Santana por toda educação, amor, direção, cuidado, carinho em todos os processos de minha vida que lhes tem me concebido. Agradeço ao professor Paulo Eduardo Teodoro por toda confiança ao aceitar ser meu orientador nessa jornada evolutiva, pela sua motivação incansável e por tudo que faz, por todo conselho que tem me ajudado e minha família, seus ensinamentos com maestria e direcionamentos decisivos nessa trajetória. Agradeço aos nossos pastores Denner Freitas Marques e Adriana Miranda Marques por todo amor e carinho para conosco, por todo encorajamento, orações e direções sobre minha vida e da minha família. Agradeço aos meus companheiros nessa jornada do doutorado, agradeço a Dthenifer Cordeiro Santana por todos os insights e ajuda nessa caminhada. Agradeço a Izabela Cristina de Oliveira por toda parceria e ajuda nessa trajetória. Agradeço também a Tatiane Deoti, entre todos os outros companheiros nesta caminhada. Agradeço a Unesp e todos os professores que tive, meu carinhoso agradecimento. Que Deus abençoe cada um poderosamente! Peço a Deus que os guarde e que direcione em cada propósito e projeto em que forem desenvolver, que a Mão poderosa de Deus os conduza. “Consagre tudo o que você faz ao Senhor, e os seus planos serão bem-sucedidos.

“Que a Shalom de Adonai seja sobre Israel, que haja Paz e prosperidade em Jerusalém!”

“Esqueçam o que se foi; não vivam no passado. Vejam, estou fazendo uma coisa nova! Ela ja está surgindo. Vocês não a reconhecem? Até no deserto vou abrir um caminho e riachos no ermo.”

Isaías 43:18,19

RESUMO

A caracterização de variáveis agronômicas com base nos sensores remotos utilizando as bandas espectrais com aprendizagem de máquina (AM) pode ser usado na fenotipagem. Essa tese foi dividida em dois capítulos. No primeiro, o objetivo foi avaliar o melhor AM e as configurações de entradas na classificação das variáveis agronômicas em diferentes estágios fenológicos. O segundo capítulo teve como objetivo utilizar o sensor hiperespectral para avaliação dos caracteres agronômicos em diversos estágios fenológicos da cultura da soja e identificar o melhor AM nas diferentes entradas testados. O experimento foi conduzido na safra 2021/2022 na área experimental da Universidade Federal de Mato Grosso do Sul, localizada no Município de Chapadão do Sul-MS. A aquisição das bandas espectrais do capítulo 1 foi realizada em três estágios fenológicos dos genótipos de soja: V8, R1 e R5. O sensor multiespectral Sequoia fez a aquisição da reflectância nos comprimentos de onda do azul (450 nm, proveniente do sensor RGB), verde (550 nm), vermelho (660 nm), borda do vermelho (735 nm) e infravermelho próximo (790 nm) possuindo um sensor de luminosidade que permite a calibração radiométrica radio métrica dos valores adquiridos. O sensor hiperespectral utilizado no capitulo 2 permitiu capturar a refletância entre os comprimentos de ondas entre 450 e 824 nm tendo como objetivo a avaliação dos caracteres agronômicos em diferentes estágios fenológicos da cultura. Os dados agronômicos e espectrais foram submetidos a análises estatísticas. A primeira análise feita foi a separação dos genótipos em grupos por meio do algoritmo k-means. Os modelos de AM de máquinas utilizados para classificação foram, REPTree (DT), Árvores de decisão J48 (J48), Floresta Aleatória (RF), Redes Neurais Artificiais (RNA), Regressão Logística (RL) e Máquina de Vetor Suporte (SVM). O desempenho de cada algoritmo foi mensurado por meio de duas métricas de acurácia: porcentagem de classificações corretas (CC) e F-score. Foram construídos boxplots para expressar acurácia dos modelos e dos inputs para cada métrica de acurácia. No capítulo 1 os modelos J48, REPtree e RNA podem ser utilizados para classificação de genótipos de soja quanto as características agronômicas. Os estágios vegetativos para aquisição dos dados espectrais podem ser V8 e R1. Com o uso de variáveis hiperespectrais, os algoritmos RL, SVM, J48 e REPTree podem ser aplicados para classificação dos genótipos de soja referente aos caracteres agronômicos nos estágios fenológicos V8 e R1.

Palavras-chave: sensores remotos; bandas espectrais; inteligência computacional; melhoramento de plantas.

ABSTRACT

The characterization of agronomic variables based on remote sensors using spectral bands used with machine learning (ML) can be used in phenotyping. This thesis was divided into two chapters. The first aimed to evaluate the best ML and input configurations in the classification of agronomic variables at different phenological stages. The second chapter aimed to use the hyperspectral sensor to evaluate agronomic traits at different phenological stages of soybean crops and identify the best ML in different inputs tested. The experiment was conducted in the 2021/2022 harvest in the experimental area of the Federal University of Mato Grosso do Sul, located in the municipality of Chapadão do Sul-MS. The acquisition of the spectral bands in chapter 1 was performed at three phenological stages of soybean genotypes: V8, R1 and R5. The Sequoia multispectral sensor acquired reflectance at wavelengths of blue (450 nm, from the RGB sensor), green (550 nm), red (660 nm), red edge (735 nm) and near infrared (790 nm), and has a luminosity sensor that allows radiometric calibration of the acquired values. The hyperspectral sensor used in chapter 2 allowed capturing reflectance between wavelengths between 450 and 824 nm, with the aim of evaluating agronomic traits at different phenological stages of soybean crops. The agronomic and spectral data were subjected to statistical analysis. The first analysis performed was the separation of genotypes into groups using the k-means algorithm. The machine learning models used for classification were REPTree (DT), J48 Decision Trees (J48), Random Forest (RF), Artificial Neural Networks (ANN), Logistic Regression (RL) and Support Vector Machine (SVM). The performance of each algorithm was measured using two accuracy metrics: percentage of correct classifications (CC) and F-score. Boxplots were constructed to express the accuracy of the models and inputs for each accuracy metric. In chapter 1, the J48, REPtree and ANN models can be used to classify soybean genotypes according to agronomic characteristics. The vegetative stages for acquiring spectral data can be V8 and R1. Using hyperspectral variables, the RL, SVM, J48 and REPTree algorithms can be applied to classify soybean genotypes according to agronomic traits in the V8 and R1 phenological stages.

Keywords: remote sensors; spectral bands; computational intelligence; plant breeding.

SUMÁRIO

1 INTRODUÇÃO GERAL 10

2 REVISÃO BIBLIOGRÁFICA 12

2.1 FENOTIPAGEM DE ALTA PRECISÃO 12

2.2 SENSORES MULTIESPECTRAIS 13

2.3 SENSORES HIPERESPECTRAIS 16

REFERÊNCIAS 18

3 CAPÍTULO 1 – FENOTIPAGEM DE ALTA PRECISÃO EM SOJA APLICANDO VARIÁVEIS MULTIESPECTRAIS ADQUIRIDAS EM DIFERENTES ESTÁGIOS FENOLÓGICOS 24

3.1 INTRODUÇÃO 25

3.2 MATERIAL E MÉTODOS 27

3.2.1 Condução do experimento 27

3.2.2 Aquisição das variáveis espectrais 27

3.2.3 Avaliação dos caracteres agronômicos 28

3.2.4 Modelos de Aprendizagem de Máquina e Análises Estatísticas 28

3.3 RESULTADOS 29

3.4 DISCUSSÃO 37

3.5 CONCLUSÃO 38

REFERÊNCIAS 39

4 CAPÍTULO 2 – FENOTIPAGEM DE ALTA PRECISÃO EM SOJA APLICANDO VARIÁVEIS HIPERESPECTRAL ADQUIRIDAS EM DIFERENTES ESTÁGIOS FENOLÓGICOS 42

4.1 INTRODUÇÃO 43

4.2 MATERIAL E MÉTODOS 45

4.3 RESULTADOS 47

4.4 DISCUSSÃO 51

4.5 CONCLUSÃO 52

REFERÊNCIAS 53

1 INTRODUÇÃO GERAL

A soja [*Glycine max* (L.) Merril], caracterizada seu centro de origem na China, é uma leguminosa pertencente à família Fabaceae (Câmara, 2015). Consideravelmente, é uma das principais commodities, gerando uma produção nacional na safra 2023/2024 de 147.684,8 milhões de toneladas, em área total cultivada de 45.7 milhões de hectares (CONAB, 2024).

O melhoramento genético da cultura de soja se caracteriza pela importância econômica visando a concentração nutritiva nos grãos como óleo, ácidos graxos, aminoácidos, açúcares, riqueza em proteínas, o favorecimento na alimentação humana, animal e no processo industrial (Rahman *et al.*, 2023). Por se destacar referente as suas características nutritivas, ao longo das décadas o melhoramento genético e o manejo cultural da soja obtiveram uma visão mais centrada aos aspectos ambientais adversos para um aumento significativo de sua produção (Mwiinga *et al.*, 2020). Algumas restrições quanto ao melhoramento genético da soja visando sua produtividade estão relacionados a demora dos ciclos das culturas. Com isso, a fenotipagem de alta precisão se torna importante na antecipação desse ciclo caracterizado pela potencialidade dos genótipos (Yuan *et al.*, 2019).

A importância do melhoramento genético de plantas visando o avanço do cultivo está relacionado a uma estratégia sustentável, ecologicamente equilibrada na aquisição de novos genótipos e ao aumento de produtividade. O ganho genético é determinado por período por meio de seleção artificial (Xu *et al.*, 2017). Além do tempo para maximizar e gerar esse ganho, também são adicionados controle parental de macho e fêmea, variância genética aditiva e intensidade de seleção, onde quanto menor a diferença de seleção, menor sua resposta no que sugere o melhoramento genético (Chawade *et al.*, 2019).

Condições ambientais diversificadas como consequência do estresse abiótico, biótico e características do solo são fatores não genéticos que afetam as características genotípicas da soja e das demais culturas, e são fatores predominantes para resposta fenotípica que pode potencializar a produtividade (Khaki *et al.*, 2020). Referente a essa alternância genética, necessita-se de uma seleção precisa e fenotipagem, na busca por acrescentar com regularidade alelos favoráveis. A fenotipagem tradicional consiste na avaliação de características agronômicas diversas aferidas com equipamento e mão de obra qualificada. Pesquisas têm se destacado por meio de técnicas precisas com o objetivo de determinar em condições ambientais adversas as características fenotípicas baseado nas condições biológicas das plantas (Santos; Yassitepe, 2014).

O uso de sensores remotos na aquisição de características das plantas tem se tornado técnicas novas de fenotipagem no que visa caracterizar variáveis agronômicas em todo período regular do ciclo fenológico principalmente pela sua forma não destrutiva (Choudhury *et al.*, 2019). O uso de aeronaves remotamente pilotadas (ARP) (Jang *et al.,* 2020) em melhoramento genético de culturas comerciais é importante com o uso de inteligência computacional, robótica, computadores e sensores visando a acurácia da fenotipagem em busca de informações de produtividade (Song *et al.,* 2020).

O ganho referente ao sistema de manejo produtivo da soja na caracterização de variáveis agronômicas com base nos sensores remotos utilizando as bandas espectrais à aprendizagem de máquina se dá pela aquisição por monitoramento de ARP e uso de satélites, tornando-se alvo de bastantes estudos e pesquisas (Silva *et al.,* 2020; Teodoro *et al.,* 2021). As análises de dados coletados pelos sensores multiespectrais é um referencial de precisão na mensuração e monitoramento de áreas (O'Connel *et al.,* 2014). Os sensores hiperespectrais é responsável por captar dados espectrais em diversas bandas estreitas e contínuas, mensurando múltiplas bandas de absorção, das mais amplas às mais sutis, tornando-se uma evolução dos sensores multiespectrais (Chlingaryan *et al.,* 2018). Esse trabalho tem como objetivo geral avaliar o uso de variáveis multi e hiperespectrais coletadas em diferentes estágios fenológicos para a predição de variáveis agronômicas em genótipos de soja e produtividade.

2 REVISÃO BIBLIOGRÁFICA

2.1 FENOTIPAGEM DE ALTA PRECISÃO

Devido ao seu recente avanço, a fenotipagem de alta precisão ainda se encontra em uma expectativa elevada, principalmente pela escassez de seus protocolos quanto ao seu registro. Para obter elevada produtividade, a aquisição de novos conhecimentos e desenvolvimento há uma limitação quanto a técnica de fenotipagem (Chawade *et al.,* 2019). O método de fenotipagem constitui-se em buscar cultivares com adaptabilidade aos estresses abióticos e bióticos (Holman *et al.,* 2016). Há uma consistência quanto a fenotipagem na caracterização de interesse dos cultivares superiores selecionados em suas características fenotípicas, buscando recursos genéticos superiores na utilização do método visual e na mensuração manual (Santana *et al.,* 2024).

O sensoriamento remoto é um método atual usado em pesquisas, que tem sua definição por obter informações equivalentes, por meio de montagem de um sensor que está relacionado ao satélite ou Aeronave remotamente pilotada, que visa por meio da captação de informações emitidas pelas plantas os caracteres da fenotipagem de alta precisão (Weiss *et al.,* 2020). Esta técnica tem como objetivo a obtenção de imagens no que visa mensurar as variáveis das plantas de forma regular no decorrer do ciclo, com alta precisão sem causar danos (Choudhury *et al.,* 2019). Essa plataforma é a mais viável com relação ao melhoramento de cultura comerciais em elevada escala (Song *et al*., 2021).

Li *et al.* (2021) relatam uma divisão quanto a realização das plataformas de fenotipagem de alta precisão em duas categorias; morfológicas ou fisiológicas. Estas características estão ligadas ao melhoramento genético principalmente pela melhoria das plantas em situações de estresse abiótico em encontrar caracteres resilientes no estado de baixa disponibilidade de água, nutriente e atuando eficientemente no uso desses inputs (Santana *et al.,* 2024). As características fisiológicas como a clorofila, a capacidade hídrica da planta e teor de nutrientes foliar estão relacionados devido a emissão pela refletância extraída pelo dossel da planta pelo comprimento de onda (Burud *et al.,* 2017).

Também na identificação de seus caracteres morfológicos relaciona-se à altura de plantas e produtividade. Em um contexto geral, o uso de sensores por meio da fenotipagem nos dá essa possibilidade de elevadas informações sobre as características das plantas e referente ao aspecto econômico (Van Eeuwijk *et al.,* 2019; Burud *et al.,* 2017). A fenotipagem de alta precisão por meio dessa dinâmica promete uma acessibilidade maior quanto essa caracterização

principalmente pela utilização de sensores pelo auxílio de veículos aéreos não tripulados e robôs (Roth *et al.,* 2022).

Na utilização desses sensores pelas Aeronaves remotamente pilotadas em análises de fenotipagem em culturas, as câmeras conhecidas como RGB são as mais utilizadas. As características acessíveis estão relacionadas por sua leveza, o processamento de dados é simples e é um sensor de baixo custo. As bandas espectrais além do RGB trazem conhecimento específicos sobre a marca espectral da vegetação, porém trazendo algumas limitações das características fisiológicas (Tait *et al.,* 2019). As plataformas de Aeronave remotamente pilotada são de asas rotativas ou fixa obtendo os sensores de imagens que são usados para análises de fenotipagem onde captam o comprimento de ondas no espectro visível de RGB, as câmeras multiespectrais, sensores hiperespectrais e os sensores infravermelho (Santana *et al.,* 2024).

O método de fenotipagem de alta precisão no uso de sensores remotos com Aeronave remotamente pilotada são importantes no quesito em que haja uma rápida demonstração dos dados em tempo real (Santana *et al.,* 2022). Em centros de pesquisas de melhoramento de soja esse método tem um papel essencial ao avaliar os caracteres das plantas. Esses caracteres relacionam-se a refletância emitidas pelos sensores que fornecem informações da soja como, estresse no estágio inicial da planta, diferenciação de clorofila, déficit hídrico, altura, diâmetro da haste principal, altura de inserção da primeira vargem, número de ramificações e previsão da produtividade da cultura (Santana *et al.,* 2023). Outro aspecto importante está relacionado a necessidade dessas ferramentas para aquisição dos índices de vegetação que se caracterizam por ser modelos matemático para variedades de comprimento de ondas (Silva *et al.,* 2020).

2.2 SENSORES MULTIESPECTRAIS

Os sensores remotos mais utilizados na agricultura, são os conjuntos multiespectrais, principalmente por apresentar baixo custo comparado aos hiperespectrais, são compostos por câmeras fotográficas profissionais, conciliados para efetuar em diferentes bandas espectrais eletromagnética (He *et al.,* 2016). Alguns modelos de sensores multiespectrais relacionados a baixo custo estão os RedEdge, de resolução Distância da amostra do solo otimizada, sensor de luz DLS 2, gerador de imagens coloridas RGB, e flexível ao mercado. Esse sensor captura bandas espectrais visíveis e invisíveis (Dong *et al.,* 2019). Outro modelo de sensores disponíveis é o ALTUM MicaSense, que tem como capacidade a captura das faixas; verde, vermelho, azul, vermelho-limítrofe, térmica e infravermelho próximo, são capturadas

simultaneamente e os dados podem ser analisados com mais acurácia. Por essa flexibilidade na atribuição de seis bandas, os dados ATUM podem ser utilizados para uma diversificação de análises desde índices como o Normalized Difference Vegetation Index, inclusive a outros índices mais aprimorados como o Normalized difference red edge index (Hutton *et al.*, 2020). Já o sensor multiespectral Parrot Sequóia é caracterizado por ser de alta acurácia com aplicação direta a agricultura, de fácil integração a qualquer veículo aéreo não tripulado, possui quatro sensores espectrais autocalibrados, e possui um sensor RGB, de 16 mega pixel (Parrot, 2017).

Usando variáveis multiespectrais coletadas com Aeronave remotamente pilotada é possível caracterizar variáveis agronômicas tais como, teor de nutriente, ciclo fenológico, produtividade, altura, dias de maturidade, diagnósticos de doenças e pragas, dentre outros caracteres (Teodoro *et al.*, 2021). Contudo, aumentar a acurácia desses modelos tem sido uma questão desafiadora, gerando enormes quantidades de dados a serem estudados nos últimos anos. Devido a quantidade de múltiplas imagens, visando a captação de informações espectrais das plantas, os sensores multiespectrais trabalham em ângulos definidos admitindo a derivação por meio de junções matemáticas entre a variação de imagens bidimensionais 2D e na faixa visível RGB (Santana *et al.*, 2024).

Os sensores multiespectrais não estão ligados diretamente a variáveis agronômicas com relação ao seu interesse, mesmo gerando uma vasta quantidade de dados espectrais. Para que ocorra uma interação a esses dados espectrais, o uso de Aprendizagem de máquinas tem sido extremamente importante referente aos seus algoritmos na identificação de genótipos de soja (Medeiros *et al.*, 2020). Trabalho realizado por Teodoro *et al*. (2021), utilizando dados multiespectrais de um sensor conectado em Aeronave remotamente pilotada avaliaram características agronômicas como altura da planta, dias de maturação, e produtividade utilizando Aprendizagem de máquina.

As bandas espectrais e o índice de vegetação, se destaca em pesquisas como destaque na cultura da soja, como meio de se obter produtividades elevadas (Jensen, 2009). São modelos matemáticos para diferentes comprimentos de onda com objetivo de caracterizar e avaliar refletância a cobertura vegetal baseadas no sensoriamento remoto, que são baseados no comportamento espectral identificando vigor e dinâmica de crescimento, estado nutricional, dentre as demais aplicações (Xue; Su, 2017). O sensoriamento remoto caracteriza-se por adquirir dados espectrais correlacionados com variáveis agronômicas que definem a variabilidade espacial, são definidos por cálculos e convertidos em índice de vegetação (Abrahão *et al.*, 2009).

O favorecimento das plantas saudáveis e nutritivas se destaca referente sua visibilidade foliar para refletância das folhas verde e na absorção da luz azul e vermelha. Com relação ao comprimento de ondas existe uma intensa interação com pigmentos de clorofila α e β, presença de xantofilas e carotenos (Din *et al.,* 2017). Sendo essencial para fotossínteses referentes ao comprimento de ondas, a faixa de 475 nm que está ligada na parte azul do espectro, captura luz pelos pigmentos que confiam as cores verdes e amarelas das folhas. A regulagem de fotoperiodismo pelo processo de fotossíntese pelo fitocromos, identifica a faixa de comprimento de onda 660 nm que é encontrado na área vermelha que é absorvida pela clorofila, sendo importante essa atividade. Outras faixas importantes que se diferenciam pelos processos ligados ao crescimento de tecidos e fotoperiodismo são RedEdge e infravermelho próximo (Bergamaschi; Bergonci, 2017).

Dentre os índices de vegetação que podem ser calculados se destacam: o Índice de Vegetação Diferença Normalizada (NDVI). Semelhante aos demais índices, como Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Resistência à Atmosfera na Região Visível (VARI), é o que mais se destaca dentre os modelos espectrais, devido favorecer ao bom indicador de biomassa da vegetação, monitoramento na detenção de déficit hídrico, danos de pragas, estimativa de produtividade, onde se encontra boa cobertura vegetal sobre o solo (Menezes *et al.,* 2012).

O NDVI Normalizada está propício a não fornecer informações exatas sobre algumas variações de cobertura vegetal, isso está relacionado a quantidade elevada de massa vegetal no que tende a saturar e causar esse fator limitante (Gitelson *et al.,* 2002). Com o índice SAVI, bastante conhecido, não há problemas encontrado referente a saturação, principalmente em estágio de desenvolvimento inicial das culturas onde não ocorre neste período o fechamento das folhas, dando uma abertura suave ao solo e com isso fornecendo dados mais detalhados e confiáveis (Braz *et al.,* 2015). O Índice de Vegetação Melhorado (EVI), é praticamente destinado em regiões especificas que concentram grandes diversidades e quantidades de biomassa vegetal, com elevados teores de clorofila, geralmente referentes as florestas tropicais (Baio *et al.,* 2018). Para vegetações que tem dificuldades de iluminação e fenômenos atmosféricos como, neblina, chuva, fumaça, poluição do ar e poeira o uso do VARI foi formado para atender de forma suave as dificuldades para esse tipo de vegetação (Earth Observing System, 2022; Silva Junior; Pacheco, 2021). O índice Diferença Normalizada na Banda de Borda Vermelha (NDRE) se caracteriza por fornecer maior eficiência em suas análises e um maior aprofundamento no dossel da planta quando comparado ao NDVI. O Índice de Diferença Normalizada da Água (NDWI), se destaca pelo comportamento transformador na água e no

mesófilo da vegetação. É originário do Infravermelho próximo e Infravermelho de ondas curtas (Moreira; Nuñez, 2023).

2.3 SENSORES HIPERESPECTRAIS

Os sensores remotos têm sido uma estratégia eficiente permitindo que a agricultura de precisão tenha um monitoramento sobre as culturas de um modo geral, onde vemos diversas culturas com grandes prejuízos causado por danos foliares por pragas, estresse abiótico, déficit nutricional e perca de produtividade (Otone *et al.,* 2024). Com relação a grandes quantidades de áreas cultivadas, trazendo informações precisas das características das plantas os sensores remotos contribuem como uma excelente ferramenta precisa de análises com rápidas informações de dados e em grandes quantidades (Teodoro *et al.,* 2024).

Os conjuntos de sensores hiperespectrais possuem uma imagem mais tecnificada e sofisticada onde os dados de refletância obtidos possuem uma larga faixa espectral que vai de 350 a 2500nm, são mais recentes e promissores comparado a imagem RGB, porém com algumas desvantagens tendo um custo mais elevado comparado aos sensores multiespectrais. (Farber *et al.,* 2019). Porém permitem uma captação mais detalhada e confiável da variabilidade de uma área caracterizada de estudo como por exemplo, estado nutricional da planta, predição de produtividade, ciclo fenológico, entre outros fatores determinantes (Chlingaryan *et al.,* 2018). Especificamente ao teor de nutrientes da planta, e as predições agronômicas, são questões que entra em comum acordo quanto ao monitoramento para quantidade ideal de adubação para suprir sua necessidade, onde são utilizadas por meios de avaliações de métodos agronômicos tradicionais em períodos chaves no gerenciamento da fertilização agrícola e das avaliações das variáveis agronômicas (Li *et al.,* 2019).

Em resposta a pesquisas recentes ao uso de sensores hiperespectrais ao monitoramento de condições nutricionais das plantas e folhas, trouxeram significativos avanços em estimativa do teor de nitrogênio, sendo esses dados correlacionados estatisticamente as características espectrais, que são denominadas tipicamente de bandas espectrais finas, de alta dimensão, que estão altamente interligadas entre si (Delloye *et al.,* 2019). Para a realização destes métodos utiliza-se a coleta de elevados números de folhas visando a caracterização das análises química dos tecidos foliares (Osco *et al.,* 2020; Malmir *et al.,* 2019). As características espectrais destes elementos por suas ações diretas e indiretas referente em suas análises podem estar relacionados também a matéria orgânica da folha (Ling *et al.,* 2019).

Vários métodos de classificação foram desenvolvidos em décadas recentes para o processamento de dados adquiridos de sensores multi e hiperespectral (Gaci *et al.,* 2023). Referente a melhora do processamento dos dados dos sensores, os algoritmos de Aprendizagem de máquina se diferenciam nas análises de imagens. Isto está referido ao desenvolvimento dos algoritmos que são permitidos pelo Aprendizagem de máquina no uso da aquisição de elevados conjuntos de dados e na complexibilidade referente a informações que requerem a junção entre eles, exemplo como a imagens de dados espectrais (Teodoro *et al.,* 2024).

REFERÊNCIAS

ABRAHÃO, S. A.; PINTO, F. A. C.; DE QUEIROZ, E. M.; SANTOS, M. T.; GLERIANI, J. M.; ALVES, E. A. Índices de vegetação de base espectral para discriminar doses de nitrogênio em capim-tanzânia. **Revista Brasileira de Zootecnia,** Viçosa, MG, v. 38, n. 9, p.1637-1644, 2009.

ARAUJO, L. S.; KELLER, M.; D'OLIVEIRA, M. V. N.; VICTORIA, D. C.; MORTON, D.; TAKEMURA, C. M.; HUNTER, M. O. Dados LiDAR e análise orientada a objeto no monitoramento de manejo florestal. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16., 2013, Foz do Iguaçu. **Anais** [...] São José dos Campos: INPE, 2013. Disponível em: https://www.researchgate.net/profile/Adriana-Araujo-6/publication/273885694_Techne_2013_194_Monitoramento_da_corrosao_Sensor_galvanico/links/550f2a3f0cf2752610a00b97/Techne-2013-194-Monitoramento-da-corrosao-Sensor-galvanico.pdf. Acesso em: 30 abr. 2022.

BAIO, F.H.R.; NEVES, D.C.; CAMPOS, C.N.S.; TEODORO, P.E. Relationship between cotton productivity and variability of NDVI obtained by landsat images. **Bioscience Journal**, Uberlândia, v. 34, n. 6, p. 197–205, 2018.

BERGAMASCHI, H.; BERGONCI, J. I. **As plantas e o clima**: princípios e aplicações. Guaíba: Agrolivros, 2017.

BERNARDES, T.; MOREIRA, M. A.; ADAMI, M.; GIAROLLA, A.; RUDORFF, B. F. T. Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. **Remote Sensing**, São José dos Campos, v. 4, n. 9, p. 2492-2509, 2012.

BRAZ, A. M.; ÁGUAS, T. A.; GARCIA, P. H. M. Análise de índices de vegetação NDVI e SAVI e Índice de Área Foliar (IAF) para a comparação da cobertura vegetal na bacia hidrográfica do córrego Ribeirãozinho, município de Selvíria–MS. **Revista Percurso**, Maringá, v. 7, n. 2, p. 5–22, 2015.

BURUD, I.; LANGE, G.; LILLEMO, M.; BLEKEN, E.; GRIMSTAD, L.; JOHAN FROM, P. Exploring robots and UAVs as phenotyping tools in plant breeding. **IFAC-Papers OnLine**, Amsterdam, v. 50, n. 1, p. 11479–11484, 2017.

CÂMARA, G. M. S. (Novembro de 2015) Introdução ao Agronegócio Soja. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4484506/mod_resource/content/0/LPV%200584%202017%20-%20REVISAO%20Soja%20Apostila%20Agronegocio%20%282%29.pdf. Acesso em: 01 abr. 2022.

CARASKI, B. G.; MINGOTTE, F. L. C. Agricultura de precisão no cultivo de cana-de-açúcar. *In*: SIMPÓSIO DE TECNOLOGIA DA FATEC - SIMTEC, 3., 2015, Taquaritinga. **Anais** [...] Taquaritinga: [*s. n.*], 2015.

CHAWADE, A.; VAN HAM, J.; BLOMQUIST, H.; BAGGE, O.; ALEXANDERSSON, E.; ORTIZ, R. High-throughput field-phenotyping tools for plant breeding and precision agriculture. **Agronomy**, Basel, v. 9, n. 5, p. 258, 2019.

CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review computer electron and agricultural. **Computers and Electronics in Agriculture**, Amsterdam, v. 151, p. 61-69, 2018.

CHOUDHURY, S.; SAMAL, A.; AWADA, T. Leveraging image analysis for high-throughput plant phenotyping. **Frontiers in plant science**, Lausanne, v. 10, p. 508, 2019.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Monitoramento Agrícola.** Disponível online: file:///C:/Users/User/Downloads/E-book_BoletimZdeZSafrasZ Z8Zlevantamento%20(1).pdf. Acesso em: 30 jun 2024.

CUNLIFFE, A. M.; BRAZIER, R. E.; ANDERSON, K. Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. **Remote Sensing of Environment,** Amsterdam, v.183, p.129–143, 2016.

DELLOYE, C.; WEISS, M.; DEFOURNY, P. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. **Remote Sensing of Environment**, Amsterdam, v. 216, p. 245-261, 2018.

DIN, M.; ZHENG, W.; RASHID, M.; WANG, S.; SHI, Z. Evaluating hyperspectral vegetation indices for leaf area index estimation of Oryza sativa L. at diverse phenological stages. **Frontiers in plant science**, Lausanne, v. 8, p. 820, 2017.

DONG, T.; LIU, J.; SHANG, J.; QIAN, B.; MA, B.; KOVACS, J. M.; WALTERS, D.; JIAO, X.; GENG, X.; SHI, Y. Assessment of red-edge vegetation indices for crop leaf area index estimation. **Remote Sensing of Environment,** Amsterdam, v. 222, p. 133-43, 2019.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Embrapa Soja. Produção de Soja**. Londrina, PR. Embrapa Soja, 2021. Disponível em: https://www.embrapa.br/soja/cultivos/soja1/dados-economicos. Acesso em: 01 abr. 2022.

EARTH OBSERVING SYSTEM. Eos. **Índices de vegetação para impulsionar soluções digitais**. 2022. Agricultura Digital. Disponível em: Acesso em: 02 set. 2024.

FARBER, C.; MAHNKE, M.; SANCHES, L.; KUROUSKI, D. Advanced spectroscopic techniques for plant disease diagnostics. A review. **Trends in Analytical Chemistry,** Amsterdam, v. 118, p. 43-49, 2019.

GACI, B.; ABDELGHAFOUR, F.; RYCKEWAERT, M.; MAS-GARCIA, S.; LOUARGANT, M.; VERPONT, F.; LALOUM, Y.; MORONVALLE, A.; BENDOULA, R. ROGER, J. M. Visible: near infrared hyperspectral dataset of healthy and infected apple tree leaves images for the monitoring of apple fire blight. **Data in Brief,** Amsterdam, v. 50, p. 109532, 2023.

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

HE, L.; SONG, X.; FENG, W.; GUO, B.B.; ZHANG, Y.S.; WANG, Y.H.; WANG, C.Y.; GUO, T.C. Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data. **Remote Sensing Environment**, Amsterdam, v. 174, p.122–133, 2016.

HOLMAN, F. H.; RICHE, A. B.; MICHALSKI, A.; CASTLE, M.; WOOSTER, M. J.; HAWKESFORD, M. J. Hight Throughput Field Phenotyping of Wheat Plant Height and Growth Rate in Field Plot Trials Using UAV Based. **Remote Sensing**, Basel, v. 8, p. 1031, 2016.

HUTTON, J. J.; LIPA, G.; BAUSTIAN, D.; SULIK, J.; BRUCE, R. W. "High accuracy direct georeferencing of the altum multi-spectral uav camera and its application to high throughput plant phenotyping." *International Archives of the Photogrammetry,* **Remote Sensing and Spatial Information Sciences**. XLIII-B1-2020 (2020): 451-56. Web.

JANG, G.; KIM, J.; YU, J.-K.; KIM, H.-J.; KIM, Y.; KIM, D.-W.; KIM, K.-H.; LEE, C. W.; CHUNG, Y. S. Cost-Effective Unmanned Aerial Vehicle (UAV) Platform for Field Plant Breeding Application. **Remote Sensing**, Basel, v. 12, n. 6, p. 998, 2020.

JENSEN, J. R.; EPIPHANO, J. C. **Remote sensing of the environment: an earth resource perspective**. São José dos Campos: Parêntese, 2011.

JÚNIOR, J. C. A.; NUÑEZ, D. N. C. O uso de drones na agricultura 4.0. **Brasilian Journal of Science**, Rio Verde, v. 3, n. 1, p. 1-13, 2023.

KHAKI, S.; KHALILZADEH, Z.; WANG, L. Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach. **PLoS ONE**, San Francisco, v. 15, n. 5, p. 1-13, 2020.

LI, F.; LI, D.; ELSAYED, S.; HU, Y.; SCHMIDHALTER, U. Using optimized three-band spectral indices to assess canopy N uptake in corn and wheat. **European Journal of Agronomy**, Amsterdam, v. 127, p. 126286, 2021.

LI, F.; WANG, L.; LIU, J.; WANG, Y.; CHANG, Q. Evaluation of leaf N concentration in winter wheat based on discrete wavelet transform analysis. **Remote Sensing**, Basel, v. 11, p. 11, 2019.

LING, B.; GOODIN, D.G.; RAYNOR, E.J.; JOERN, A. Hyperspectral analysis of leaf pigments and nutritional elements in tallgrass prairie vegetation. **Frontiers in Plant Science**, Lausanne, v. 10, p. 1-13, 2019.

MALMIR, M.; TAHMASBIAN, I.; XU, Z.; FARRAR, M.B.; BAI, S.H. Prediction of macronutrients in plant leaves using chemometric analysis and wavelength selection. **Journal of Soils and Sediments,** Heidelberg, v. 20, p. 249-259, 2019.

MANTOVANI, E. C. Agricultura de precisão e sua organização no Brasil. In: BORÉM, A. **Agricultura de precisão**. Viçosa: UFV, 2000. p. 77-108.

MEDEIROS, A. D.; CAPOBIANGO, N. P.; DA SILVA, J. M.; DA SILVA, L. J.; DA SILVA, C. B.; DIAS, D. C. F. S. Interactive machine learning for soybean seed and seedling quality classification. **Scientific Reports**, London, v. 10, p. 11267, 2020.

MENESES, P. R.; ALMEIDA, T.; ROSA, A. N. C. S.; SANO, E. E.; SOUZA, E. B.; BAPTISTA, G. M. M.; BRITES, R. S. **Introdução ao processamento de imagens de sensoriamento remoto.** Brasília, DF: UnB, p. 266, 2012.

MOREIRA, V.; NUÑEZ, D. N. C. Estimativa de produtividade da soja na safra 2021/2022: Índice de Vegetação e Machine Learning. **Brazilian Journal of Science**, Rio Verde, v. 2, n. 1, p. 7-15, 2023.

OLIVEIRA, A. J.; DA SILVA, E.E.; BAIO, F.H.R.; TEODORO, L.P.R.; JUNIOR, C.A.D.S.; BORGES, R.S.; TEODORO, P.E. UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. **Remote Sensing Application. Society Environment**, Amsterdam, v.18, p.100-318, 2020.

OSCO, L.; JUNIOR, J.; RAMOS, A.; FURUYA, D.; SANTANA, D.; TEODORO, L.; GONÇALVES, W.; BAIO, F.; PISTORI, H.; JUNIOR, C.; ET AL. Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. **Remote Sensing,** Basel**,** v. 12, p. 3237, 2020.

OTONE, J. D. Q; THEODORO, G. D. F.; SANTANA, D. C.; TEODORO, L. P. R.; DE OLIVEIRA, J. T.; DE OLIVEIRA, I. C.; DA SILVA JUNIOR, C. A.; TEODORO, P. E.; BAIO, F. H. R. Hyperspectral Response of the Soybean Crop as a Function of Target Spot (Corynespora cassiicola) Using Machine Learning to Classify Severity Levels. **AgriEngineering**, Basel, v. 6, p. 330–343, 2024.

PARROT. Sequoia. **User guide**. [*S. l.: s. n.*], 2017.

ROTH, L.; BARENDREGT, C.; BÉTRIX, C. A.; HUND, A.; WALTER, A. High-throughput field phenotyping of soybean: Spotting an ideotype. **Remote Sensing of Enviornment**, Amsterdam, v. 269, n. 11, p. 2797, 2022.

SANTANA, D. C.; DE OLIVEIRA, I. C.; CAVALHEIRO, S. B.; DAS CHAGAS, P. H. M.; FILHO, M. C. M. T.; DELLA-SILVA, J. L.; TEODORO, L. P. R.; CAMPO, C. N. S.; BAIO, F. H. R.; JUNIOR, C. A. S.; TEODORO, P. E. Classification of Soybean Genotaypes as to Calcium, Magnesium, and Sulfur Content Using Machine Learning Models and UAV–Multispectral Sensor. **AgriEngineering**, Basel, v. 6, p. 1581-1593, 2024.

SANTANA, D.C.; THEODORO, G.D.F.; GAVA, R.; DE OLIVEIRA, J.L.G.; TEODORO, L.P.R.; DE OLIVEIRA, I.C.; BAIO, F.H.R.; DA SILVA JUNIOR, C.A.; DE OLIVEIRA, J.T.; TEODORO, P.E. A New Approach to Identifying Sorghum Hybrids Using UAV Imagery Using Multispectral Signature and Machine Learning. **Algorithms,** Basel, v. 17, p. 23, 2024.

SANTANA, D.C.; TEIXEIRA FILHO, M.C.M.; DA SILVA, M.R.; CHAGAS, P.H.M.D.; DE OLIVEIRA, J.L.G.; BAIO, F.H.R.; CAMPOS, C.N.S.; TEODORO, L.P.R.; DA SILVA JUNIOR, C.A.; TEODORO, P.E.; SHIRATSUCHI, L. S. Machine Learning in the Classification of Soybean Genotypes for Primary Macronutrients' Content Using UAV–Multispectral Sensor. **Remote Sensing**, Basel, v. 15, p. 1457, 2023.

SANTANA, D. C.; CUNHA, M. P. O.; DOS SANTOS, R. G.; COTRIM, M. F.; TEODOR, L. P. R.; JUNIOR, C. A. S.; BAIO, F. H. R.; TEODORO, P. E. High-throughput phenotyping allows the selection of soybean genotypes for earliness and high grain yield. **Plant Methods**, London, v. 18, p. 1-13, 2022.

SANTI, A. L.; AMADO, T. J. C.; EITELWEIN, M. T.; CHERUBIN, M. R.; DA SILVA, R. F.; DA ROS, C. O. Definição de zonas de produtividade em áreas manejadas com agricultura de precisão. **Revista Brasileira de Ciências Agrárias**, Recife, v. 8, n. 3, p. 510-515, 2013.

SANTOS, T. T.; YASSITEPE, J. E. de C. T. Fenotipagem de plantas em larga escala: um novo campo de aplicação para a visão computacional na agricultura. **Embrapa Informática Agropecuária-Capítulo em livro científico (ALICE)**, [*s. l.*], 2014.

SILVA, E. E.; BAIO, F. H. R.; TEODORO, L. P. R.; JUNIOR, C. A. S.; BORGES, R. S.; TEODORO, P. E. UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. **Remote Sensing Applications: Society and 18 Environment**, Amsterdam, v. 18, n. 4, 2020.

SILVA JUNIOR, J. A.; PACHECO, A. P. Avaliação de incêndio em ambiente na caatinga a partir de imagens Landsat-8, índice de vegetação realçado e análise por componentes principais. **Ciência Florestal**, Santa Maria, v. 31, n.1, p. 417-439, 2021.

SONG, H.; AHMAD, A.; FAROOQ, F.; OSTROWSKI, K. A.; MAŚLAK, M.; CZARNECKI, S.; ASLAM, F. Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. **Asian Journal of Civil Engineering**, Cham, v. 308, p. 125021, 2021.

TAIT, L.; BIND, J.; CHARAN-DIXON, H.; HAWES, I.; PIRKER, J.; SCHIEL, D. Unmanned aerial vehicles (UAVs) for monitoring macroalgal biodiversity: Comparison of RGB and multispectral imaging sensors for biodiversity assessments. **Remote Sensing**, Basel, v. 11, n. 19, p. 2332, 2019.

TEODORO, L. P. R.; ESTEVÃO, R.; SANTANA, D.C.; OLIVEIRA, I.C.D.; LOPES, M.T.G.; AZEVEDO, G.B.D.; BAIO, F.H. R.; JUNIOR, C.A. S; TEODORO, P.E. Eucalyptus Species Discrimination Using Hyperspectral Sensor Data and Machine Learning. **Forests**, Basel, v. 15, p. 39, 2024.

TEODORO, L. P. R.; ESTEVÃO, R.; SANTANA, D. C.; DE OLIVEIRA, I. C.; LOPES, M. T. G.; DE AZEVEDO, G. B.; BAIO, F. H. R.; JUNIOR, C. A. S.; TEODORO, P. E. Eucalyptus Species Discrimination Using Hyperspectral Sensor Data and Machine Learning. **Forests,** Basel, v. 15, p. 39, 2023.

TEODORO, P.E.; TEODORO, L.P.R.; BAIO, F.H.R.; DA SILVA JUNIOR, C.A.; DOS SANTOS, R.G.; RAMOS, A.P.M.; PINHEIRO, M.M.F.; OSCO, L.P.; GONÇALVES, W.N.; CARNEIRO, A.M.; JUNIOR, J. M.; PISTORI, H.; SIRATSUCHI, L. S. Predicting Days to Maturity, Plant Height, and Grain Yield in Soybean: A Machine and Deep Learning Approach Using Multispectral Data. **Remote Sensing,** Basel, v. 13, p. 4632, 2021.

USDA. 2022. World **Agricultural Production. United States Department of Agriculture-USDA**, Circular Series. Acessado em: https:// https://apps.fas.usda.gov/psdonline/circulars/production.pdf. Acesso em: 01 abr. 2022.

VAN EEUWIJK, F. A.; BUSTOS-KORTS, D.; MILLET, E. J.; BOER, M. P.; KRUIJER, W.; THOMPSON, A.; MALOSETTI, M.; IWATA, H.; QUIROZ, R.; KUPPE, C.; MULLER, O.; BLAZAKIS, K. N.; YU, K.; TARDIEU, F.; CHAPMAN, S. C. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. **Plant Science**, Shannon, v. 282, p. 23–39, 2019.

WEISS, M.; JACOB, F.; DUVEILLER, G. Remote sensing for agricultural applications: A meta-review. **Remote Sensing of Environment**, Amsterdam, v. 236, p. 111402, 2020.

XU, Y.; LI, P.; ZOU, C.; LU, Y.; XIE, C.; ZHANG, X.; PRASANNA, B. M.; OLSEN, M. S. Enhancing genetic gain in the era of molecular breeding. **Journal of Experimental Botany**, Oxford, v. 68, n. 11, p. 2641–2666, 2017.

XUE, J.; SU, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. **Journal of Sensors,** London, p.1-17, 2017.

ZHU, X.; CHEN, L.; XU, H.; ZHANG, X. Application of agriculture 4.0 in China: **A Review. IEEE Access**, Piscataway, v. 9, p. 81981-81991, 2021.

3 CAPÍTULO 1 – FENOTIPAGEM DE ALTA PRECISÃO EM SOJA APLICANDO VARIÁVEIS MULTIESPECTRAIS ADQUIRIDAS EM DIFERENTES ESTÁGIOS FENOLÓGICOS

RESUMO

A soja se destaca por ser a oleaginosa de maior importância econômica do mundo. As técnicas de sensoriamento remoto estão sendo analisadas por meio de pesquisas em diferentes regiões agrícolas como sistema tecnológico visando produtividade e possíveis baixo custo. Os métodos de Aprendizagem de máquinas (AM) juntamente com o advento de demanda por aeronave remotamente pilotada disponíveis no mercado em década recente tem sido propício aos processos de dados de sensoriamento remoto. O objetivo deste trabalho foi avaliar o melhor AM e as configurações de entradas na classificação das variáveis agronômicas em diferentes estágios fenológicos. As variáveis espectrais foram obtidas em três estágios fenológicos dos genótipos de soja: V8 (aos 45 dias após a emergência – DAE), R1 (60 DAE) e R5 (80 DAE). Foi utilizada Aeronave remotamente pilotada (ARP) de asa fixa Sensefly eBee equipado com o sensor multiespectral Parrot Sequoia acoplada junto ao sensor RGB. O sensor multiespectral Sequoia com o sensor RGB fez a aquisição da reflectância nos comprimentos de onda do azul (450 nm), verde (550 nm), vermelho (660 nm), infravermelho próximo (735 nm) e infravermelho (790 nm). Os seguintes caracteres agronômicos foram avaliados: dias de maturação, número de ramificações, produtividade, altura de plantas, altura da inserção da primeira vagem e diâmetro da haste principal. Os genótipos foram separados em dois grupos: um grupo com caracteres agronômicos desejáveis e outro com desempenho agronômico inferior. Os modelos J48, REPtree e RNA podem ser utilizados para classificação de genótipos de soja referente aos caracteres agronômicos. Os estágios vegetativos para aquisição dos dados multiespectrais podem ser V8 e R1.

Palavras-chave: sensor multiespectral; bandas espectrais; aprendizagem de máquina; caracteres agronômicos.

ABSTRACT

Soybean stands out for being the most economically important oilseed in the world. Remote sensing techniques are being analyzed through research in different agricultural regions as a technological system aiming at productivity and low-cost reduction. Machine learning (ML) methods, together with the advent of demand for remotely piloted aircraft available on the market in the recent decade, have been conducive to remote sensing data processes. The objective of this work was to evaluate the best ML and input configurations in the classification of agronomic variables in different phenological stages. The spectral variables were obtained in three phenological stages of soybean genotypes: V8 (at 45 days after emergence - DAE), R1 (60 DAE) and R5 (80 DAE). A Sensefly eBee fixed-wing RPA equipped with the Parrot Sequoia multispectral sensor coupled to the RGB sensor was used. The Sequoia multispectral sensor with RGB sensor acquired reflectance at wavelengths of blue (450 nm), green (550 nm), red (660 nm), near infrared (735 nm) and infrared (790 nm). The following were used to evaluate the agronomic traits: days to maturity, number of branches, productivity, plant height, height of the first pod insertion and diameter of the main stem. The J48, REPtree and RNA models can be used to classify soybean genotypes according to agronomic traits. The vegetative stages for acquiring multispectral data can be V8 and R1.

Keywords: multispectral sensor; spectral bands; machine learning; agronomic traits.

3.1 INTRODUÇÃO

Por ser a oleaginosa de maior importância econômica no mundo, a soja (*Glycine max* L. Merril) se destaca por vários quesitos. A identificação de suas características como, ciclo fenológico, dias de maturidade, produtividade, teor de nutriente e altura de plantas podem contribuir estrategicamente para manejos futuros da cultura trazendo benefícios (ZHOU *et al.*, 2019). O Brasil é o maior produtor mundial de soja, mesmo com queda de produção devido ao déficit hídrico em algumas regiões. Em alguns estados brasileiros em anos recentes, obteve aumento de área de aproximadamente 4,1%, alcançando 40.804,9 milhões hectares na safra 2021/2022 com uma produtividade média 3.000kg/ha¹ com produção total de 122,43 milhões de toneladas (CONAB, 2022).

As características genotípicas da soja e das demais culturas, juntamente com diversos fatores ambientais como: abióticos, bióticos e características do solo, são fatores predominantes

para potencialização da produtividade (Khaki *et al.,* 2020). As técnicas de sensoriamento remoto por meio de pesquisas em diversas áreas agricultáveis com fatores tecnológicos dão ênfase ao: diagnóstico de doenças e pragas, aplicações de adubos em taxas variáveis e produtividade (Weiss *et al.,* 2020).

Em programas de melhoramento, o uso de técnicas de sensoriamento remoto é útil, onde são avaliados centenas de genótipos anualmente, principalmente devido as dificuldades que se encontram nas análises de caracteres agronômicos. Com o advento de demanda por ARP disponíveis no mercado, há um enorme incentivo a aplicações no campo devido ao seu rápido mapeamento detalhado de grandes áreas agrícolas (Osco *et al.,* 2020). Variáveis multiespectrais coletadas com ARP é possível caracterizar variáveis agronômicas tais como, teor de nutriente, ciclo fenológico, altura de plantas, dias de maturidade, diagnósticos de doenças e pragas, produtividade, dentre outros caracteres (Teodoro *et al.,* 2021). Contudo, aumentar a acurácia desses modelos tem sido uma questão desafiadora, gerando enormes quantidades de dados a serem estudados nos últimos anos.

Em pesquisas recentes os métodos de AM têm sido propícios aos processos de dados de sensoriamento remoto (Khaki *et al.,* 2020). Dentre as culturas relatadas, a soja devido sua diversidade genotípica e suas variáveis agronômicas, usando dados de sensoriamento remoto, tem sido algo desafiador, não só por suas características de variedades genotípicas, mas também pelas diferenças de áreas geográficas (Silva *et al.,* 2020).

Vários estudos estão relacionados aos métodos de AM, como *regression tree, random forest*, regressão multivariada, *association rule mining*, e redes neurais artificiais para predição de produtividade de lavouras. Os modelos de AM têm como as saídas dos dados referentes as produtividades das lavouras, função subentendida das variáveis de entrada, tendo as condições do clima e as condições do solo fatores que podem ser algo complexo e uma função não linear (Khaki *et al.,* 2020). Ainda se torna difícil o êxito da predição de produtividade da cultura referente a vários fatores climáticos e condições limitantes do solo. Aspectos determinantes também estão relacionados a interação entre genótipos e ambiente, fazendo com que haja maior desafio quanto ao sucesso da predição de produtividade da cultura (Khaki *et al.,* 2019).

Como variáveis multiespectrais, informações sobre os dados que foram usados como entrada nos algoritmos de AM, que são informações cruciais para melhorar a precisão do processamento de dados, para a avaliação dos caracteres agronômicos são essenciais. Este trabalho teve como objetivo avaliar o melhor AM e as configurações de entradas na classificação das variáveis agronômicas em diferentes estágios fenológicos.

3.2 MATERIAL E MÉTODOS

3.2.1 Condução do experimento

A condução do experimento foi realizada na safra 2021/2022 no campo experimental da Universidade Federal de Mato Grosso do Sul, situada no Município de Chapadão do Sul-MS (18°41’33’’S, 52°40’45’’W, com 810 m de altitude). Seguindo a classificação de Köppen e Geiger o clima da região é Tropical de Savana (Aw), a temperatura anual está entre 13 a 29 °C, a precipitação pluviométrica em média é de 1.850 mm, com concentração de chuva no verão e seca no inverno. A classificação do solo é Latossolo Vermelho Distrófico Argiloso, com características na camada de 0 – 0,20 m: pH (H_2O) = 6,2; Al trocável ($cmol_c\ dm^{-3}$) = 0,0; Ca+Mg ($cmol_c\ dm^{-3}$) = 4,31; P (mg dm^{-3}) = 41,3; K ($cmol_c\ dm^{-3}$) = 0,2; Matéria orgânica (g dm^{-3}) = 19,74; V (%) = 45; m (%) = 0,0; CTC ($cmol_c\ dm^{-3}$) = 5,1. Soma de bases ($cmol_c\ dm^{-3}$) = 2,3;

O delineamento utilizado foi o de blocos casualizados com 28 populações F_3 de soja, quatro tratamentos cultivares comerciais (BMX COMPACTA, TMG7061, TMG7062 e TMG7063) e duas repetições. Foram definidas as parcelas em 4 linhas com um 1 de comprimento, com espaçamento de 0,45 m entre fileiras e densidade de 15 plantas por metro.

Em outubro de 2021, foi estabelecida a semeadura, com preparo convencional do solo (aração e gradagem niveladora). Para o tratamento das sementes foram com fungicida (Piraclotrobina + Metil Tiofanato) e inseticida (Fipronil), na dose de 200 mL do produto comercial para cada 100 kg de sementes para proteção contra o ataque de pragas e fungos de solo. Para inoculação das sementes foram com bactérias do gênero *Bradyrhizobium*, utilizada a dose de 200 mL de inoculante líquido concentrado para cada 100 kg de sementes. Os tratos culturais foram realizados de acordo com a necessidade da cultura.

3.2.2 Aquisição das variáveis espectrais

Foram obtidas em três estágios fenológicos dos genótipos de soja: V8 (aos 45 dias após a emergência – DAE), R1 (60 DAE) e R5 (80 DAE). Foi utilizada a aeronave remotamente pilotada (ARP) de asa fixa Sensefly eBee RTK em cada voo, com controle autônomo de decolagem, plano de voo e aterrissagem. O eBee está equipado com o sensor multiespectral Parrot Sequoia, o qual também possui um sensor RGB. A aquisição da refletância nos comprimentos de onda do azul (450 nm, proveniente do sensor RGB), verde (550 nm),

vermelho (660 nm), borda do vermelho (735 nm) e infravermelho próximo (790 nm) foi pelo sensor multiespectral Sequoia, possuindo um sensor de luminosidade que permite a calibração radiométrica dos valores adquiridos. Pelo programa computacional Pix4Dmapper, as imagens foram mosaicadas e ortorretificadas. A acurácia posicional das ortoimagens será verificada com pontos de controle em campo, levantados com o RTK.

3.2.3 Avaliação dos caracteres agronômicos

Para os caracteres agronômicos foram avaliados, altura de inserção da primeira vagem (AIV, cm), altura de plantas (AP, cm), número de ramificações (NR), diâmetro da haste principal (DHP, cm), dias para maturação (DM) e produtividade de grãos (PROD, kg ha^{-1}). Os caracteres AIV, AP, DHP e NR foram avaliados em cinco plantas em cada parcela, sendo os dois primeiros avaliados com auxílio de uma fita métrica e o terceiro com auxílio de um paquímetro digital. Para os dias entre emergência e maturação, o DM correspondeu de mais de 50% de plantas em cada unidade experimental. Referente a PROD foi caracterizado pela colheita das fileiras centrais de cada parcela e avaliado. A correção de umidade foi estabelecida e corrigida a 13% de umidade e transformado para kg ha^{-1}.

3.2.4 Modelos de Aprendizagem de Máquina e Análises Estatísticas

Foi utilizado o algoritmo k-means para os dados obtidos a partir das informações espectrais e das avaliações agronômicas para agrupar os tratamentos em que os centroides estão aproximados, dividindo os genótipos em dois grupos chamados Cluster 1 e Cluster 2 de forma que havendo uma distância mínima na observação referente a qualquer centroide não ocorra uma variação significativa. Para expressar a separação dos clusters, contribuindo para formação de um biplot feito com ajuda do pacote “ggfortify” do software R (R Core Team, 2013), a análise de componentes principais (PCA) foi utilizada, contribuindo para esses fatores. Posteriormente foi construído boxplots para cada cluster segundo as informações dos dados agronômicos. A sujeição dos dados foi pela análise de correlação de Pearson, pelo software Rbio (Bhering, 2017), para cada época de avaliação.

Para variáveis de saída (output) foram utilizados os clusters formados, e referente a variáveis de entrada foram utilizados os dados espectrais, para os modelos de classificação nas análises de aprendizado de máquina (Tabela 2). As configurações testadas nos modelos foram as épocas de coleta dos dados espectrais. Foi utilizado a validação cruzada estratificada com k-

fold=10 e dez repetições para classificação dos clusters, com obtenção de 100 execuções para cada modelo. As medidas dos modelos estabelecidos foram em relação a configuração default do software Weka 3.8.5.

Tabela 2 - Aquisição de modelos de aprendizagem de máquinas para classificação

Sigla	Aprendizagem de máquinas	Referência
ANN	Redes neurais artificiais	(EGMONT-PETERSEN; DE RIDDER; HANDELS, 2002)
J48	Árvores de decisão J48	(QUINLAN, 1993)
LR	Regressão logística	(ŠTEPANOVSKÝ *et al.*, 2017)
DT	REPTree	(SNOUSY *et al.*, 2011)
RF	Floresta aleatória	(BELGIU; DRĂGUŢ, 2016)
SVM	Máquina de vetor suporte	(NALEPA; KAWULOK, 2019)

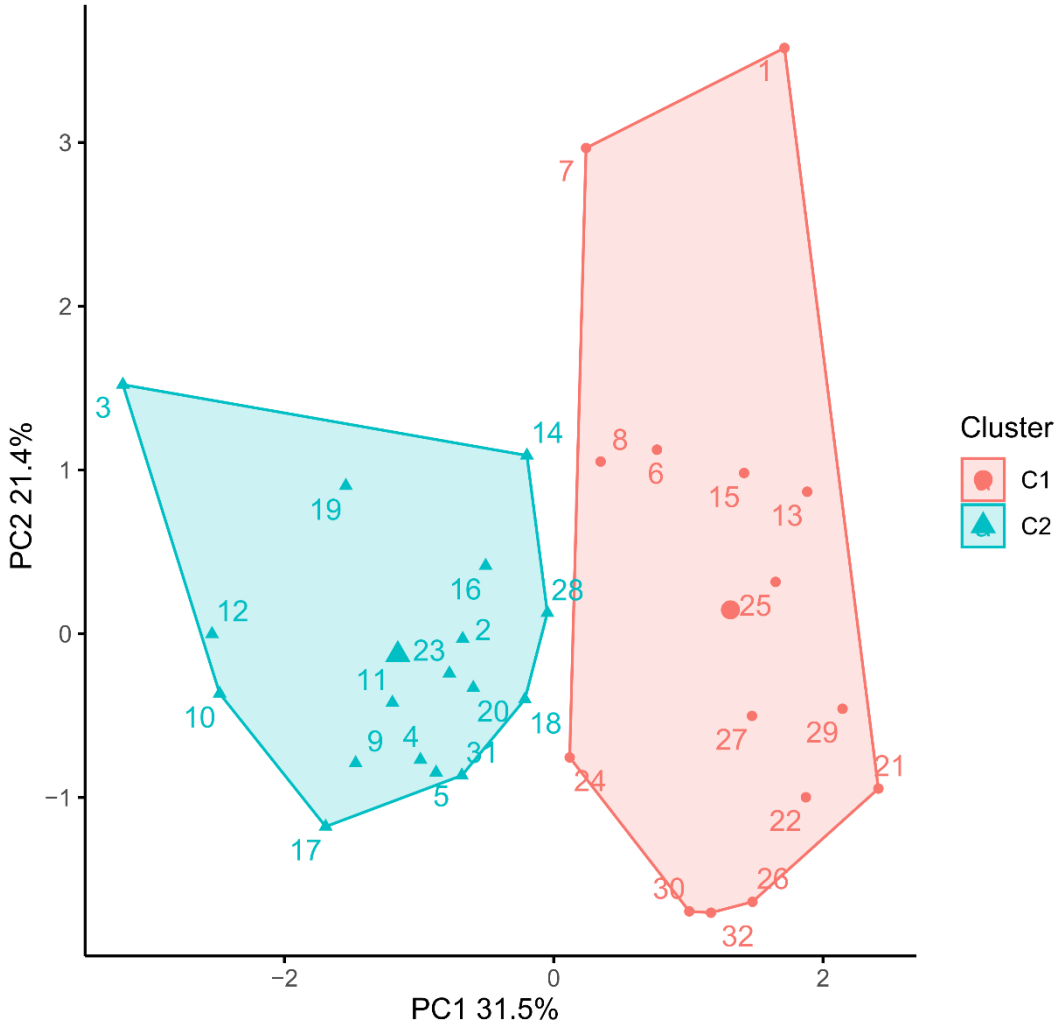
Fonte: Elaboração do autor.

As métricas de acurácia de porcentagem utilizados para o desempenho desses modelos foram de classificações corretas (CC), F-socre e coeficiente Kappa. A avaliação do desempenho dos modelos foi realizada segundo métricas de acurácia de porcentagem de classificações corretas (CC), F-score e coeficiente Kappa. A verificação do desempenho dos inputs, modelos de AM e interação entre eles foram feitas pela análise de variância com base nos modelos, gerando boxplots com as médias no caso de significância a de 5%, de acordo com o teste de (Scott; Knott, 1974) aplicando os pacotes ggplot2 e ExpDes.pt do software R (Belgiu; Drăguţ, 2016).

3.3 RESULTADOS

O algoritmo k-means foi associado a técnica de componentes principais para agrupar os genótipos de soja em dois grupos cujo centroides estão próximos (Figura 1) baseado nos caracteres agronômicos avaliados. Os genótipos de soja classificados dentro do Cluster 1, foram; 1, 6, 7, 8, 13, 15, 21, 22, 24, 25, 26, 27, 29, 30 e 32. Dentro do Cluster 2, foram; 2, 3, 4, 5, 9, 10, 11, 12, 14, 16, 17, 18, 19, 20, 23, 28 e 31.

Figura 1 - Análise de Componentes Principais (PC) para os 32 genótipos de soja avaliados e clusters formados pelo algoritmo k-means baseado nos caracteres agronômicos avaliados

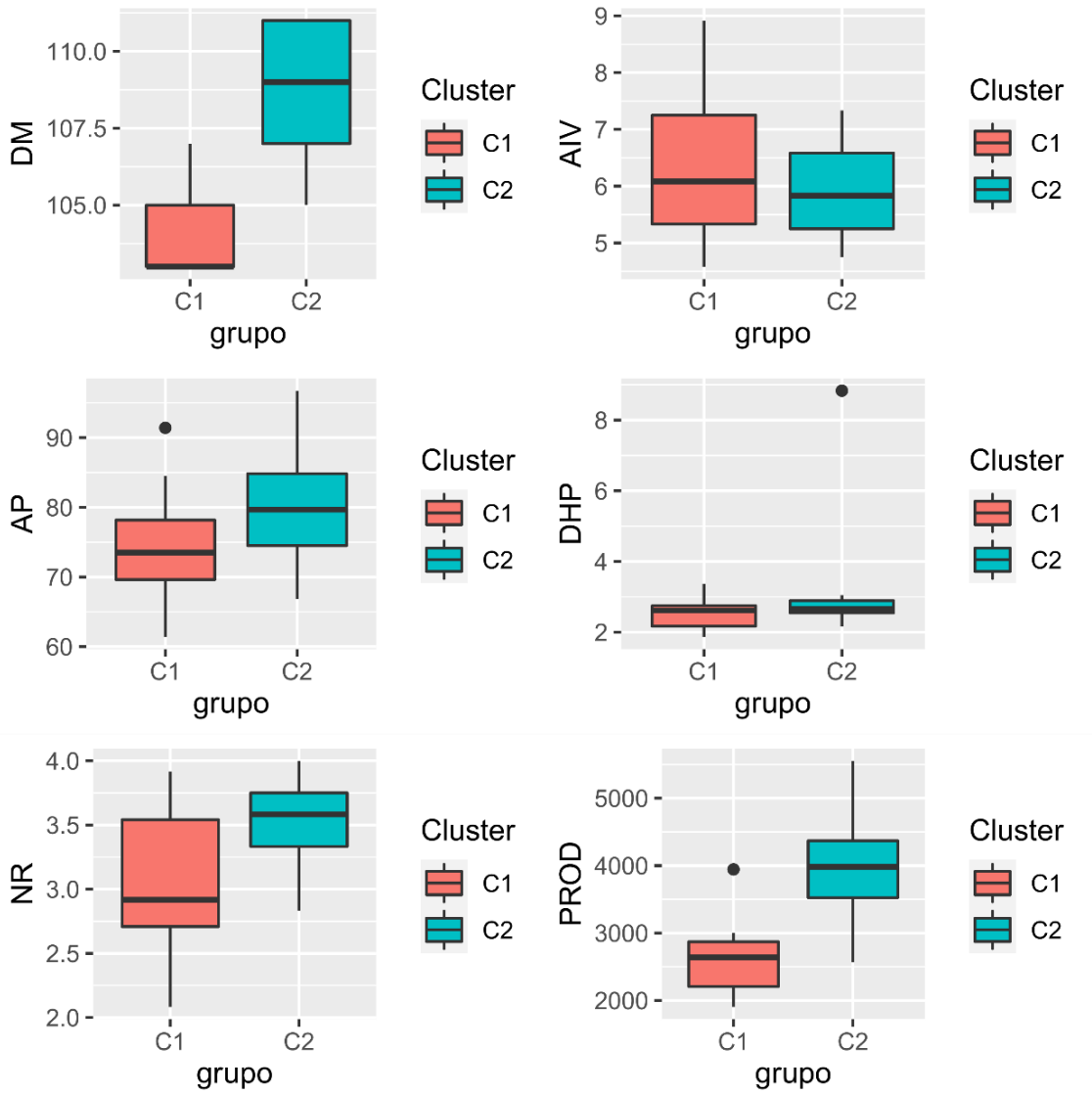


Fonte: Elaboração do autor.

Para DM, o Cluster 2 diferenciou-se do Cluster 1, obtendo um período maior de dias, com aproximadamente a 110 dias (Figura 2). A AIV, o Cluster 1 teve uma maior altura, com uma média de 6 a 7cm comparado ao Cluster 2. Para AP, o Cluster 2 obteve maior média quando comparado ao Cluster 1, com altura de aproximadamente 85 cm, que é uma altura excelente quando comparado as plantas comerciais de soja. Para o DHP, foi identificado que o Cluster 2 se diferenciou do Cluster 1, com uma média de diâmetro de 3,0 cm. O Cluster 2 demonstrou maior NR quando comparado ao Cluster 1, com uma variação de 3,5 a 4. Para a PROD, foi identificado que o Cluster 2 obteve maior número comparado ao Cluster 1, com uma média de 4000 kg ha¹. De forma geral, ao considerar todos os caracteres agronômicos, os genótipos de

soja dentro do grupo Cluster 2 obtiveram melhores resultados quando comparado aos do Cluster 1.

Figura 2 - Boxplot com as médias de cada Cluster formados para as variáveis agronômicas DM, AIV, cm, AP, cm, DHP, cm, (NR), PROD kg ha¹, avaliados em 32 genótipos de soja



Fonte: Elaboração do autor.

A Figura 3A demonstra as correlações no estágio V8, onde foi observado que os comprimentos de onda *blue and green* apresentaram alta correlação. Houve correlação positiva entre PROD e DM. Os comprimentos de onda *Red and RedEdge* obtiveram alta correlação negativa. Na Figura 3B, que relata o estágio reprodutivo R1, houve alta correlação entre DM e NR, que não foi observada no estágio vegetativo V8. Os comprimentos de ondas *blue* and *green*

também obtiveram alta correlação neste estágio, porém no estágio reprodutivo R1 isso não foi observado. O gráfico da figura 3C, demonstra correlações positivas para PROD e DM, onde voltaram a ter uma correlação mais destacada. Para os demais caracteres foram observadas baixas correlações das variáveis espectrais.

4 5

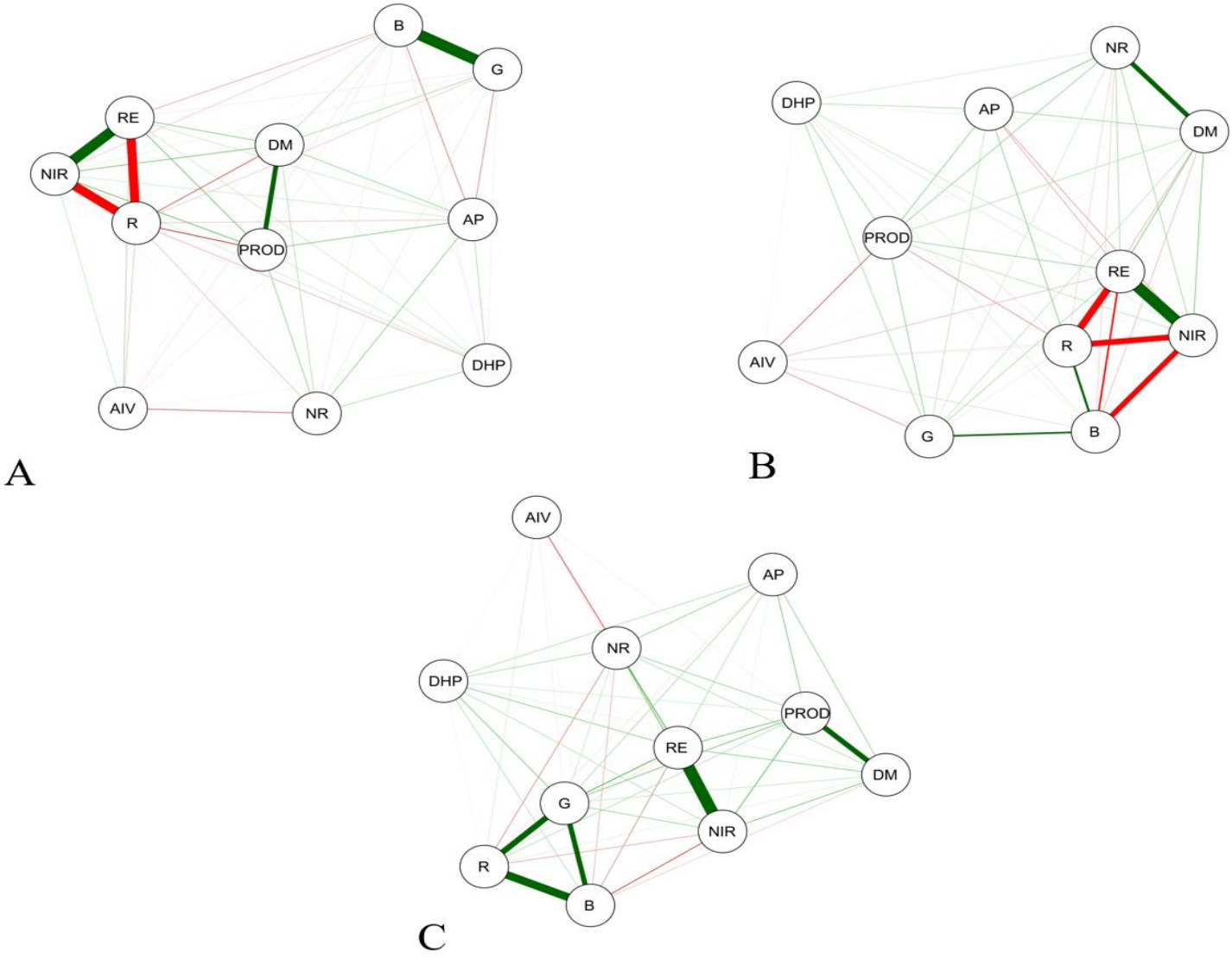


Figura 3 - Rede de correlações entre as variáveis agronômicas DM, PROD, NR, AP, AIV, DHP e as bandas espectrais comprimentos de onda do azul (450 nm), verde (550 nm), vermelho (660 nm), borda do vermelho (735 nm) e infravermelho próximo (790 nm) para as três épocas de avaliação (V8, R1 e R5). Linhas verdes e vermelhas indicam correlações positivas e negativas, respectivamente. A magnitude dessas relações é indicada pela espessura das linhas, sendo as linhas grossas representando as correlações acima de 0,60. Fonte: Elaboração do autor.

A Tabela 2 contém a análise de variância para CC, coeficiente Kappa e F-score para os modelos de AM testados usando dados espectrais coletados em diferentes estágios vegetativos da soja. Observa-se que houve interação significativa entre modelos de AM e estágios vegetativos para coleta dos dados espectrais em todas as métricas avaliadas.

Tabela 2 - Resumo da análise de variância para as variáveis CC, coeficiente Kappa e F-score para diferentes modelos de AM utilizando como entrada variáveis espectrais coletadas em diferentes estágios vegetativos da soja.

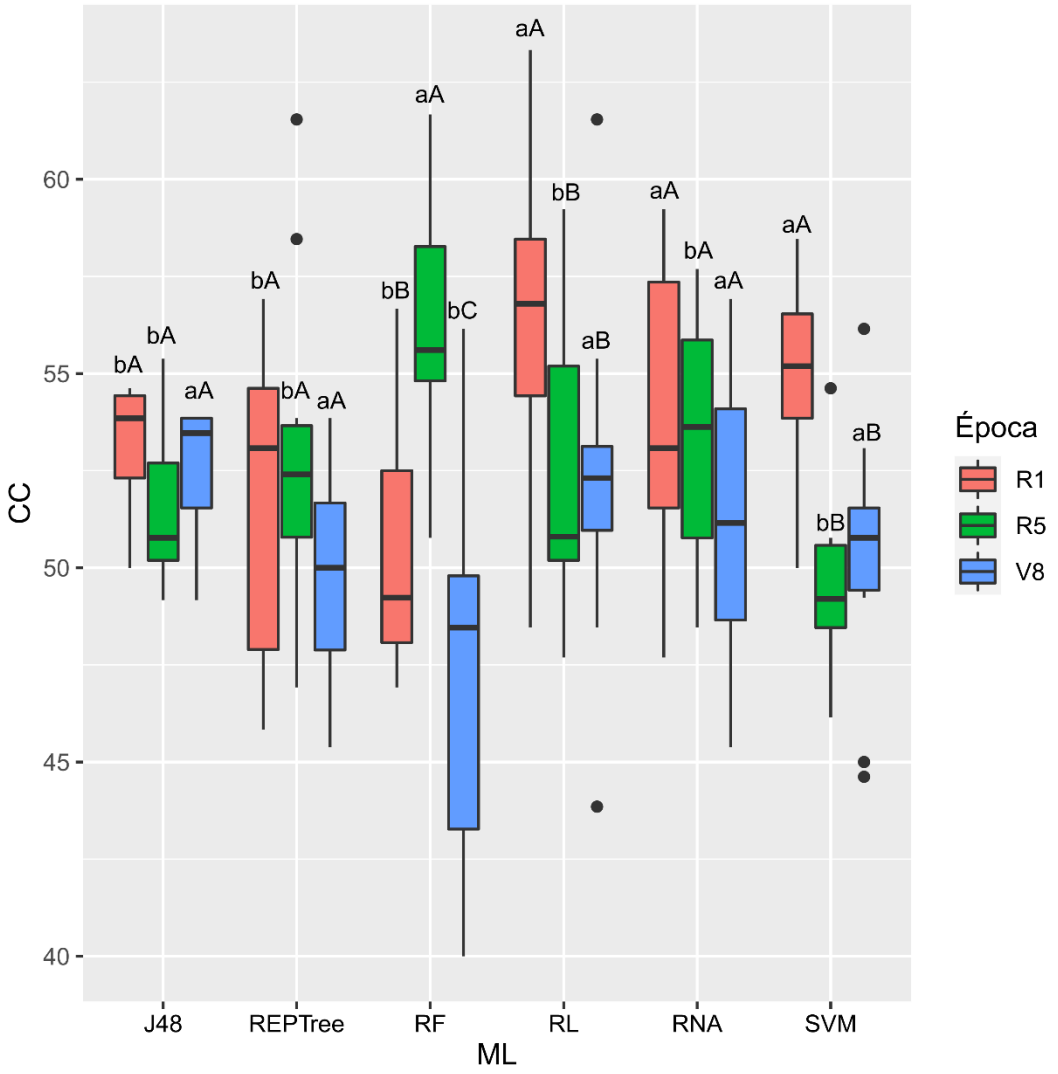
FV	GL	CC	Kappa	F-score
AM	5	26.136*	0.025986*	0.11173*
Estágio	2	131.485*	0.05219*	0.0312*
AM*Estágio	10	49.442*	0.020217*	0.032808*

Nota: *significativo a 5% e **significativo a 1% pelo teste F; ML: inteligência computacional.
Fonte: Próprio autor.

De acordo com CC dos genótipos de soja nos clusters formados (Figura 4), os algoritmos RL, RNA e SVM foram os que obtiveram melhor desempenho no estágio vegetativo V8 e estágio reprodutivo R1. O algoritmo RL obteve 52,5% em V8 e 57,5% em R1. Para RNA em V8, a média foi de 53% e em R1 57%. E SVM, 51,5% em V8 e R1 56%. Para o estágio reprodutivo R5, o modelo algoritmo que se destacou frente aos demais foi o RF, com porcentagem de classificação de 56,5%. Independente dos inputs testados, os modelos J48, REPtree e RNA obtiveram o melhor desempenho com base na classificação correta.

Figura 4 - Boxplot para porcentagem de CC para a interação significativa entre os modelos de aprendizado de máquinas e os inputs testados.

As médias seguidas pelas mesmas letras maiúsculas para as diferentes entradas e as mesmas letras minúsculas para os diferentes algoritmos de AM não diferem pelo teste de Scott-Knott a 5% de probabilidade

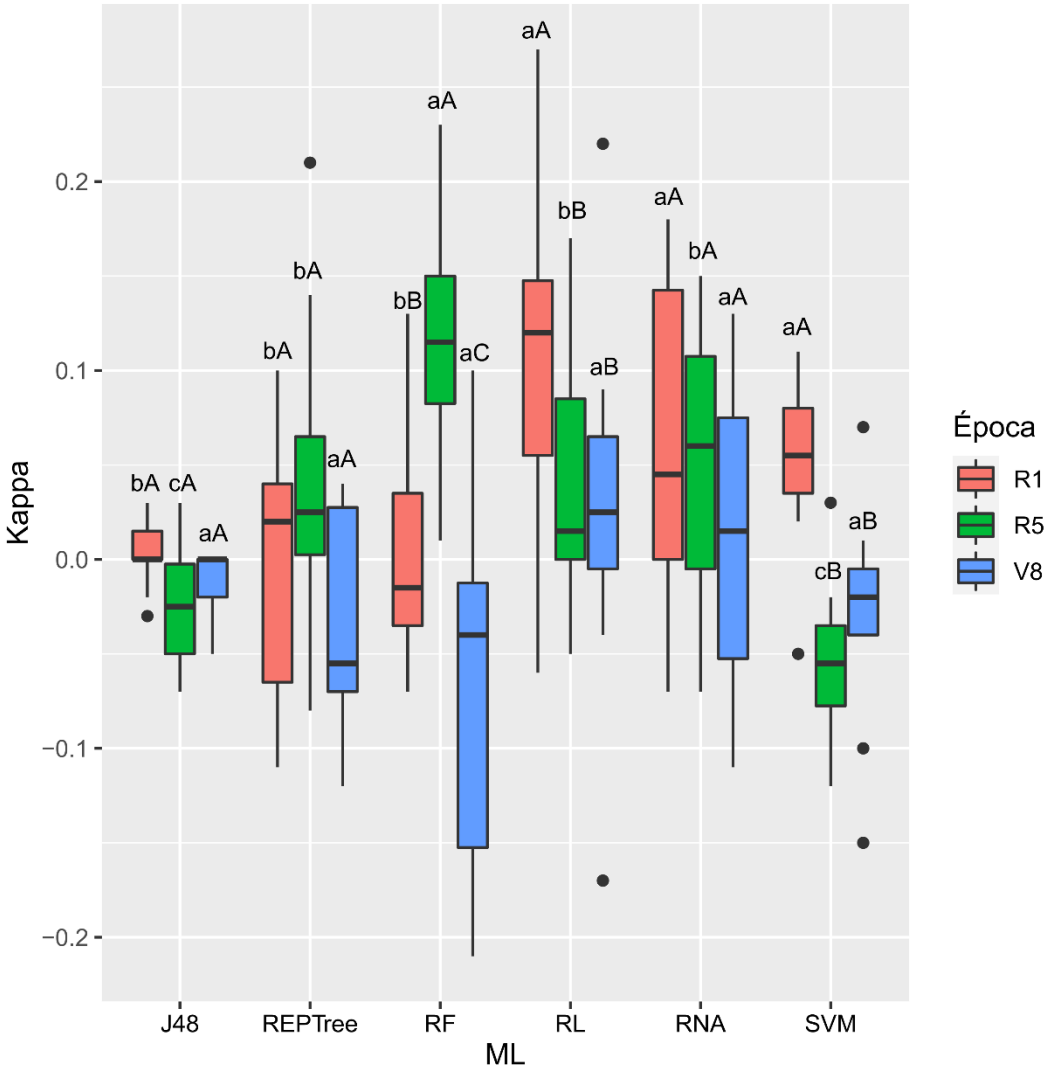


Fonte: Elaboração do autor.

Com base no coeficiente Kappa, os modelos de AM que se sobressaíram no estágio reprodutivo R1 foram RL, RNA e SVM. Para o estágio vegetativo V8, não houve diferença entre os algoritmos avaliados. Para o estágio reprodutivo R5, o algoritmo RF foi o único que obteve melhor desempenho frente aos demais modelos. Com relação aos inputs testados, os modelos J48, REPtree e RNA tiveram melhores resultados.

Figura 5 - Boxplot para coeficiente Kappa para a interação significativa entre os modelos de aprendizado de máquinas e os inputs testados.

As médias seguidas pelas mesmas letras maiúsculas para as diferentes entradas e as mesmas letras minúsculas para os diferentes algoritmos de AM não diferem pelo teste de Scott-Knott a 5% de probabilidade

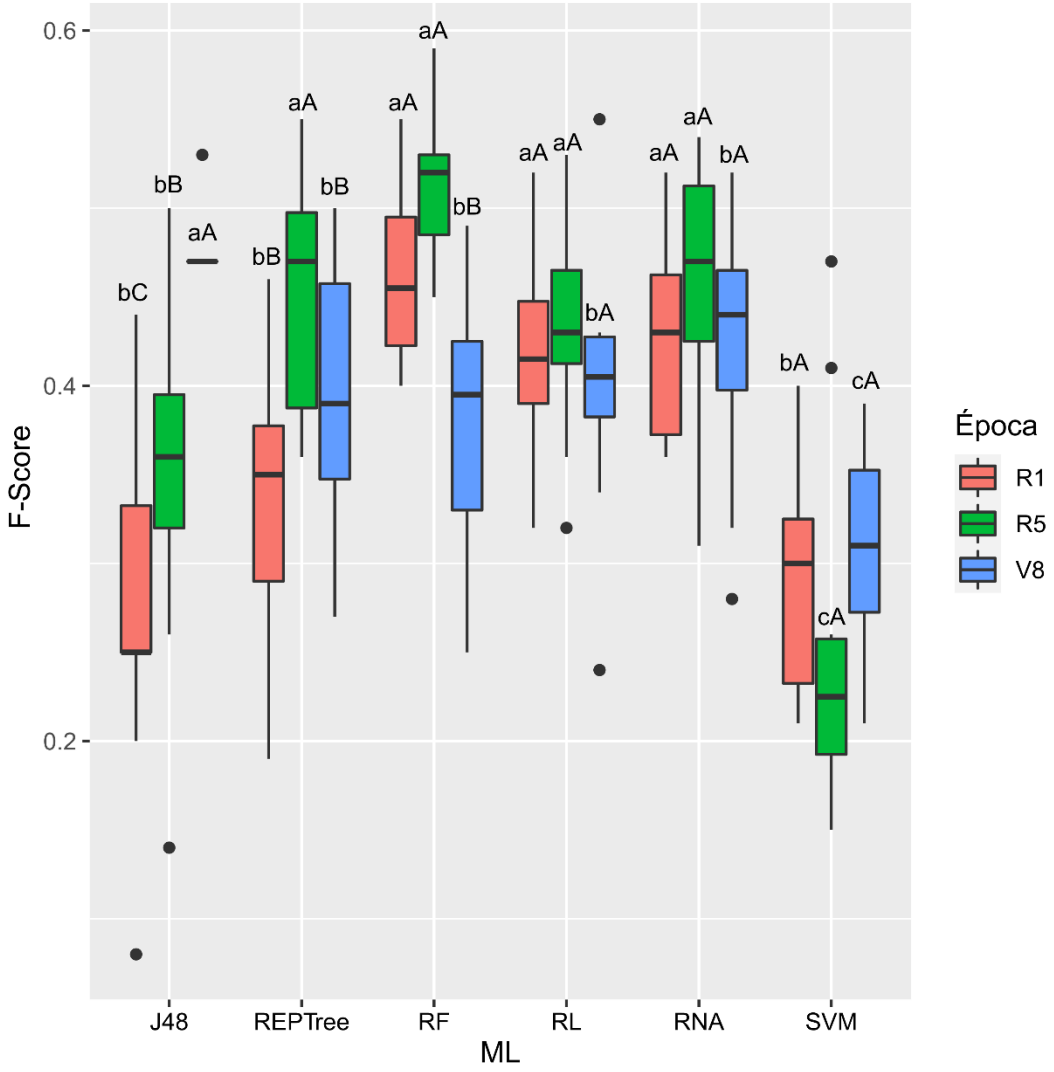


Fonte: Elaboração do autor.

Com base no F-Score, para o estágio reprodutivo R1, os algoritmos RF, RL e RNA se diferenciaram dos demais modelos, obtendo melhores resultados (Figura 6). Para o estágio reprodutivo R5, apenas os algoritmos J48 e SVM que não obtiveram bons resultados comparado aos demais. Para os diferentes inputs testados, o algoritmo J48 obteve melhor resultado quando comparado ao estágio vegetativo V8. Para REPtree, diferenciou-se dos demais apenas no

estágio reprodutivo R5. Ao comparar o modelo algoritmo RF, foi o único que não obteve resultados relevante quando comparado aos outros modelos no estágio vegetativo V8.

Figura 6 - Boxplot para F-score para a interação significativa entre os modelos de aprendizado de máquinas e os inputs testados.



Fonte: Elaboração do autor.

As médias seguidas pelas mesmas letras maiúsculas para as diferentes entradas e as mesmas letras minúsculas para os diferentes algoritmos de AM não diferem pelo teste de Scott-Knott a 5% de probabilidade.

3.4 DISCUSSÃO

Os algoritmos de aprendizagem de máquina (AM), se destacam em processar dados não lineares e não paramétricos, como também os dados espectrais (OSCO *et al.,* 2020). Diante disso, os genótipos de soja foram separados em dois grupos com base no algoritmo K-means. Entretanto, ao separar os genótipos em clusters, nota-se que os genótipos pertencentes ao Cluster 2, possuem melhores desenvolvimentos de acordo com os caracteres avaliados, isso ocorre devido aos fatores genéticos da soja em suas características em absorção nutricional e metabólica (Chaney, 2022). A correlação entre as variáveis espectrais coletadas em diferentes estágios vegetativos e os caracteres agronômicos foi baixa. Mesmo na ausência de relação linear entre os caracteres agronômicos e as variáveis espectrais, é possível usar os modelos de AM para classificação dos genótipos de acordo com os clusters formados (Osco *et al.,* 2020). A finalidade tecnológica na utilização de sensoriamento remoto junto ao crescimento do uso computacional em análise de dados é uma maneira de ser constituída na agricultura, contribuindo aos dados analisados como espaço temporal de inúmeras características de plantas de forma rápida e econômica (Khechba *et al.,* 2021).

A aplicabilidade de técnicas de AM no seguimento dessas referencias é imprescindível os aspectos de genótipos analisados e a quantidade de números de dados espectrais ocasionado. A ocorrência da falta de dados lineares entre os dados de caracteres agronômicos e espectrais, são fatores que as técnicas de AM conseguem lidar com precisão (Osco *et al.,* 2020). Relacionado aos dados espectrais, a eficácia do uso de AM disponibiliza garantir importantes fontes de informações dos caracteres agronômicos, como exemplo as folhas, e todo o estágio fisiológico e nutricional da planta (Soba *et al.,* 2020). Referente a esses fatores tecnológicos é possível analisar vários caracteres agronômicos no que diz respeito a suas características genéticas e desenvolvimento, propiciando a escolha de uma melhor linhagem de soja.

Vários modelos de algoritmos inputs testados aos caracteres agronômicos obtiveram bons resultados. Os algoritmos J48, SVM e REPtree obtiveram interação significativa com os inputs para os coeficientes de acurácia. Em específico ao algoritmo J48, são extraídos resultados de classificação eficientemente e rápido de acordo com o processamento (Rossi *et al.,* 2017). Essas características estão relacionadas a vantagens como a demanda de tempo para o processamento de dados, havendo uma menor interferência humana na formação do algoritmo (Vieira *et al.,* 2012). Referente ao SVM e REPtree, obteve bom dinamismo na determinação dos dados espectrais e boa capacidade de disseminação e acurácia (Bigdeli *et al.,* 2013; Okwuashi, 2020; Mountrakis, 2011).

O uso de bandas espectrais como entrada para modelo de AM é mais favorável evitando a necessidade da utilização de cálculos para índice de vegetação (Gava *et al.,* 2022). Diversos fatores são determinados eficientemente ao uso de bandas espectrais como a identificação de várias características da planta, água na soja, estado nutricional referente a alguns nutrientes, faixa apropriada quanto aos caracteres agronômicos e fator determinante em relação a toxicidade no eucalipto (Braga *et al.,* 2021). Sendo assim, os modelos de AM juntamente com algoritmos de entrada no uso de banda espectrais permitem a detecção de vários aspectos determinantes referente a modificações no desenvolvimento dos caracteres agronômicos. O uso adequado das técnicas de AM permite agregar valores no que se refere a superar vários quesitos problemáticos de classificação (Barbedo, 2019).

Tais técnicas são promissoras, e necessitam de mais trabalhos visando sua eficácia. Um dos principais fatores ocasionados ao uso das técnicas de AM é o fator de auxiliar economicamente programas de melhoramento. As técnicas de AM e os dados multiespectrais podem trazer resultados e orientações com precisão ainda no campo, acarretando em várias informações de características das plantas (Brian *et al.,* 2022). Portanto, encontrar modelos de AM capazes de identificar genótipos de soja que pertençam a esse cluster pode ser útil para programas de melhoramento de soja, que avaliam centenas de genótipos anualmente.

3.5 CONCLUSÕES

Os modelos J48, REPtree e RNA podem ser utilizados para classificação de genótipos de soja quanto as características agronômicas.

Os estágios fenológicos mais promissores para aquisição dos dados multiespectrais foram V8 e R1.

REFERÊNCIAS

BARBEDO, J. G. A. Detection of nutrition deficiencies in plants using proximal images and machine learning: a review. **Computers and Electronics in Agriculture**, Amsterdam, v. 162, p. 482–492, 2019.

BELGIU, M.; DRĂGUŢ, L. Random Forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, Basel, v. 114, p. 24–31, 2016.

BHERING, L.L. Rbio: a tool for biometric and statistical analysis using the R Platform. Crop Breeding and Ap-plied Biotechnology, [*s. l.*] v. 17, p. 187–190, 2017.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Monitoramento Agrícola**. Disponível online: https://E-book_BoletimZdeZSafrasZ-Z7oZlevantamento-compactado.pdf

BRIAN, C.; SHI, H.; WU, S.; ZHANG, K.; WEI, M.; ZHAO, Y.; SUN, Y.; ZHUANG, H.; ZHANG, X.; CHEN, S. Prediction of field-scale wheat yield using machine learning method and multi-spectral UAV Data. **Remote Sensing Evironment**, Basel, v. 14, p. 1456-1474, 2022.

BIGDELI, B.; SAMADZADEGAN, F.; REINARTZ, P. A Multiple SVM system for classification of hyperspectral remote sensing data. **Journal of the Indian Society of Remote Sensing**, Basel, v. 41, p. 763–776, 2013.

BRAGA, P.; CRUSIOL, L. G. T.; NANNI, M. R.; CARANHATO, A. L. H.; FUHRMANN, M. B.; NEPOMUCENO, A. L.; NEUMAIER, N.; FARIAS, J. R. B.; KOLTUN, A.; GONÇALVES, L. S. A.Vegetation indices and NIR-SWIR spectral bands as a phenotyping tool for water status determination in soybean. **Precision Agriculture**, Viçosa, MG, v. *22*, p. 249–266, 2021.

CHANEY, R.L. Breeding soybeans to prevent mineral deficiencies or toxicities. *In:* WORLD SOYBEAN RESEARCH CONFERENCE, 3., 2022, [*s. l.*]. **Proceedings** [...] [*S. l.*]: CRC Press, 2022. v. 7, p. 453–459.

SILVA, E. E.; BAIO, F. H. R.; TEODORO, L. P. R.; JUNIOR, C. A. D. S.; BORGES, R. S.; TEODORO, P. E. UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. **Remote Sensing Applications: Society and Environment**, Amsterdam, v. 18, p. 100-318, 2020.

EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks a review. **Pattern Recognition**, Amsterdam, v. 35, n. 10, p. 2279–2301, 2002.

GAVA, R.; SANTANA, D.C.; COTRIM, M.F.; ROSSI, F.S.; TEODORO, L.P.R.; DA SILVA JUNIOR, C.A.; TEODORO, P.E. Soybean Cultivars Identification Using Remotely Sensed Image and Machine Learning Models. **Sustainability**, New Rochelle, v. 14, p. 7110-7125, 2022.

HE, L.; SONG, X.; FENG, W.; GUO, B.B.; ZHANG, Y.S.; WANG, Y.H.; WANG, C.Y.; GUO, T.C. Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data. **Remote Sensing Environment**, Amsterdam, v. 174, p. 122–133, 2016.

KHAKI, S.; WANG, L.; ARCHONTOULIS, S.V. A Cnn-rnn Framework for Crop Yield Prediction. Front. **Plant Science**, Shannon, v. 10, p. 1738-1750, 2019.

KHECHBA, K.; LAAMRANI, A.; DHIBA, D.; MISBAH, K.; CHEHBOUNI, A. Monitoring and analyzing yield gap in africa through soil attribute best management using remote sensing approaches: a review. **Remote Sensing Environment**, Basel, v. 13, p. 4583-4602, 2021.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, Basel, v. 66, p. 247–259, 2011.

NALEPA, J.; KAWULOK, M. Selecting training sets for support vector machines: a review. **Artificial Intelligence Review**, Dordrecht, v. 52, n. 2, p. 857–900, 2019.

OKWUASHI, O.; NDEHEDEHE, C.E. Deep Support Vector Machine for Hyperspectral Image Classification. **Pattern Recognit**, Amsterdam, v. 103, p. 107-298, 2020.

OSCO, L.; JUNIOR, J.; RAMOS, A.; FURUYA, D.; SANTANA, D.; TEODORO, L.; GONÇALVES, W.; BAIO, F.; PISTORI, H.; JUNIOR, C. Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. **Remote Sensing Environment**, Amsterdam, v. 12, p. 3220-3237, 2020.

OSCO, L. P.; RAMOS, A. P. M.; FAITA PINHEIRO, M. M.; MORIYA, É. A. S.; IMAI, N. N.; ESTRABIS, N.; IANCZYK, F.; ARAÚJO, F. F.; LIESENBERG, V.; JORGE, L. A. C. A. Machine Learning Framework to Predict Nutrient Content in Valencia-Orange Leaf Hyperspectral Measurements. **Remote Sensing Environment**, Basel, v. 12, p. 885-906, 2020.

QUINLAN, J. R. C4. 5: Programming for machine learning. **Morgan Kauffmann**, [*s. l.*] v. 38, n. 48, p. 49, 1993.

ROSSI NETO, J.; DE SOUZA, Z.M.; DE MEDEIROS OLIVEIRA, S.R.; KÖLLN, O.T.; FERREIRA, D.A.; CARVALHO, J.L.N.; BRAUNBECK, O.A.; FRANCO, H.C.J. Use of the Decision Tree Technique to Estimate Sugarcane Productivity Under Edaphoclimatic Conditions. **Sugar Tech**, New Delhi, v. 19, p. 662–668, 2017.

SNOUSY, M. B. AL et al. Suite of decision tree-based classification algorithms on cancer gene expression data. **Egyptian Informatics Journal**, Amsterdam, v. 12, n. 2, p. 73–82, 2011.

SOBA, D.; SHU, T.; RUNION, G.B.; PRIOR, S.A.; FRITSCHI, F.B.; ARANJUELO, I.; SANZ-SAEZ, A. Effects of Elevated [CO2] on Photosynthesis and Seed Yield Parameters in Two Soybean Genotypes with Contrasting Water Use Efficiency. **Environment Experimental Botany**, Oxford, v. 178, p. 104-154, 2020.

ŠTEPANOVSKÝ, M.; IBROVÁ, A.; BUK, Z.; VELEMÍNSKÁ, J. Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. **Forensic Science International**, Amsterdam, v. 279, p. 72–82, 2017.

TEAM, R.C. R: **A Language and Environment for Statistical Computing**. 2013.

TEODORO, P. E.; TEODORO, L. P. R.; BAIO, F. H. R.; SILVA JUNIOR, C. A.; SANTOS, R. G.; RAMOS, A. P. M.; PINHEIRO, M. M. F.; OSCO, L. P.; GONÇALVES, W. N.; CARNEIRO, A. M. Predicting days to maturity, plant height, and grain yield in soybean: a machine and deep learning approach using multispectral data. **Remote Sensing Environment**, Amsterdam, v. 13, p. 4617-4632, 2021.

VIEIRA, M. A.; FORMAGGIO, A. R.; RENNÓ, C. D.; ATZBERGER, C.; AGUIAR, D. A.; MELLO, M. P. Object Based Image Analysis and Data Mining Applied to a Remotely Sensed Landsat Time-Series to Map Sugarcane over Large Areas. **Remote Sensing Environment**, Amsterdam, v. 123, p. 553–562, 2012.

ZHOU, J.; YUNGBLUTH, D. C.; VONG, C. N.; SCABOO, A. M.; ZHOU, J. Estimation of maturity date of soybean breeding lines using UAV-based imagery. **Remote Sensing Environment**, Amsterdam, v. 11, p. 2058-2075, 2019.

WEISS, M.; JACOB, F.; DUVEILLER, G. Remote sensing for agricultural applications: a meta-review. **Remote Sensing Environment**, Amsterdam, v. 236, p. 111-402, 2020.

4 CAPÍTULO 2 – FENOTIPAGEM DE ALTA PRECISÃO EM SOJA COM VARIÁVEIS HIPERESPECTRAL AVALIADAS EM DIFERENTES ESTÁGIOS FENOLÓGICOS

RESUMO

A soja destaca seu cultivo mundialmente devido sua importância econômica e nutritiva. O sensoriamento remoto tem se destacado por técnicas em recursos para sua produção. Os índices de vegetação e bandas espectrais por meio do sensoriamento remoto inseridos ao aprendizado de máquinas (AM), tem sido importante na eficiência e a confiabilidade para predição dos caracteres agronômicos da soja. Assim, este trabalho teve como objetivo utilizar o sensor hiperespectral para avaliação dos caracteres agronômicos em diferentes estágios fenológicos da cultura e identificar o melhor AM em distintas entradas testados. A refletância capturada pelos comprimentos de onda foram 450 a 824 nm utilizado o sensor hiperespectral Ocean Opticis modelo STS-VIS-L-50-400-SMA da análise espectral de três amostras foliares escolhidas ao acaso da parte superior do dossel da planta de cada parcela. Os dados dos caracteres agronômicos e espectrais foram submetidos a análises estatísticas. A primeira análise feita foi a separação dos genótipos em grupos por meio do algoritmo k-means. Os modelos de AM utilizados para classificação foram, Árvores de decisão J48, REPTree, Floresta Aleatória, Redes Neurais Artificiail, Regressão Logística (RL), Máquina de Vetor Suporte (SVM). Inicialmente os genótipos foram separados em dois grupos. O desempenho de cada algoritmo em classificar os genótipos em cada grupo foi mensurado por meio de duas métricas de acurácia: porcentagem de classificações corretas (CC) e F-score. Foram construídos boxplots para expressar acurácia dos modelos e dos inputs para cada métrica de acurácia. O período de avaliação dos dados hiperespectrais depende do algoritmo de aprendizagem utilizado para classificação dos genótipos de soja quanto aos caracteres agronômicos. De forma geral, os estágios fenológicos mais indicados para obtenção dos dados hiperespectrais e classificação dos genótipos de soja são R1 e R5 associado ao uso dos modelos de RL e SVM.

Palavras-chave: sensor hiperespectral; bandas espectrais; aprendizagem de máquina; caracteres agronômicos.

ABSTRACT

Soybean cultivation stands out worldwide due to its economic and nutritional importance. Remote sensing has stood out as a resource for its production. Vegetation indices and spectral bands through remote sensing inserted into machine learning (ML) have been important in the efficiency and reliability of predicting soybean agronomic traits. Thus, this work aimed to use the hyperspectral sensor to evaluate agronomic traits at different phenological stages of the crop and identify the best ML in different inputs tested. The reflectance captured by the wavelengths were 450 to 824 nm using the Ocean Opticis hyperspectral sensor model STS-VIS-L-50-400-SMA from the spectral analysis of three leaf samples randomly chosen from the upper part of the plant canopy of each plot. The data on agronomic and spectral traits were subjected to statistical analysis. The first analysis performed was the separation of genotypes into groups using the k-means algorithm. The ML models used for classification were J48 decision trees, REPTree, Random Forest, Artificial Neural Networks, Logistic Regression (RL), and Support Vector Machine (SVM). Initially, the genotypes were separated into two groups. The performance of each algorithm in classifying the genotypes in each group was measured using two accuracy metrics: percentage of correct classifications (CC) and F-score. Boxplots were constructed to express the accuracy of the models and inputs for each accuracy metric. The evaluation period of the hyperspectral data depends on the learning algorithm used to classify the soybean genotypes according to the agronomic traits. In general, the most suitable phenological stages for obtaining hyperspectral data and classifying the soybean genotypes are R1 and R5 associated with the use of the RL and SVM models.

Keywords: hyperspectral sensor; spectral bands; machine learning; agronomic traits.

4.1 INTRODUÇÃO

Devido suas características econômicas, nutritivas, fontes de matéria prima, destinado a alimentação animal e humana, a cultura da soja [*Glicine max* (L) Marril] é a leguminosa mais cultivada no mundo (Zhou *et al.*, 2019). A soja se caracteriza importante no Brasil por ser a principal commoditie, e se destaca por ser o maior produtor mundial, mesmo com uma grande queda de produção nesta safra, recorrente do desastre climático ocorrido no estado do Rio Grande do Sul, um dos estados mais produtivos do país,

chegando a uma média de produção 147.684,8 milhões de toneladas, em uma área produzida de 45.733,2 milhões de hectares na safra 2023/24 (CONAB, 2024).

A produção de soja tem aumentado significativamente em décadas recentes, fazendo com que, programas de melhoramento genéticos e tecnológicos têm sido importantes para inteiração e adaptação genotípica, fenotípica, ambiental, e na caracterização de óleo, teor de proteína e carboidratos (Gupta; Manjaya, 2022). Esses fatores destacam-se não só devidamente a expansão de área em anos recentes, mas principalmente aos procedimentos de novas tecnologias como agricultura de precisão, mecanização, transgenia, melhoramento genético e técnicas avançadas, manejo de pragas e doenças (Hirakuri; Lazzarotto, 2014).

Em busca de maior produtividade e qualidade, a característica fisiológica da soja em correlação a sua taxa fotossintética é muito importante (Karyawati *et al.*, 2015). Isso requer a captação desses caracteres fisiológicos em forma de dados, principalmente ao difícil acesso aos métodos, mesmo disponível, existe uma demanda de tempo, e por precisar de mão de obra especializada referente a complexidade na utilização e calibração dos aparelhos (Capelin *et al.*, 2017).

O sensoriamento remoto tem se destacado em recursos para produção de soja, fazendo o monitoramento referentes a imagens por meio de aeronave remotamente pilotada (ARP) e via satélite (Gao *et al.*, 2017). Com a diversidade de estudos que demonstre a eficiência e a confiabilidade para predição dos caracteres agronômicos da cultura da soja, os índices de vegetação e as bandas espectrais por meio do sensoriamento remoto introduzidos ao AM tem sido importante nesses aspectos (Teodoro *et al.*, 2021).

O uso da técnica de sensoriamento remoto, por vias tecnológicas sensoriais terrestres hiperespectrais, tem sido atribuída em estudos e pesquisas além de menor demanda de custo, para adquirir redução no tempo e no monitoramento dos plantios agrícolas, possibilitando a identificação de plantas para predições de variáveis agronômicas, como; índices de vegetação, taxas fotossintéticas, estado nutricional das plantas, fertilidade, condições de estresse hídrico, além de outros danos abióticos e bióticos que podem ocorrer ao longo do ciclo fenológico da cultura e produtividade (Jesen; Epiphanio, 2011; Formaggio; Sanches, 2017).

Para a realização destes métodos utiliza-se a coleta de elevados números de folhas visando a caracterização das análises química dos tecidos foliares. Este método exige uma tarefa minuciosa, demorada e trabalhosa (Osco *et al.*, 2019; Malmir *et al.*, 2019). Para uma análise nutricional da planta quanto ao teor de nutriente, ciclo fenológico, predição de rendimento, altura, com mais precisão, visando uma alternativa eficaz, o sensoriamento remoto hiperespectral pode ser vantajoso comparado aos métodos tradicionais (Ling *et al.*, 2019;

Román *et al.,* 2019). As características espectrais destes elementos por suas ações diretas e indiretas referente em suas análises podem estar relacionados também a matéria orgânica da folha (Ling *et al.,* 2019). Com isso, este trabalho teve por objetivo utilizar o sensor hiperespectral para avaliação dos caracteres agronômicos em diferentes estágios de desenvolvimento da cultura da soja.

4.2 MATERIAL E MÉTODOS

A Universidade Federal de Mato Grosso do Sul, se localiza no Município de Chapadão do Sul-MS (18°41’33’’S, 52°40’45’’W, com 810 m de altitude) onde o experimento foi conduzido, na safra 2021/2022. Segundo a classificação de Köppen e Geiger o clima da região é Tropical de Savana (Aw), a temperatura anual está entre 13 a 29 °C, a precipitação pluviométrica em média é de 1.850 mm, com concentração de chuva no verão e seca no inverno. O solo da área foi identificado como Latossolo Vermelho Distrófico Argiloso. A Tabela 1 contém a análise química na camada de 0-20 cm.

O delineamento utilizado foi o de blocos casualizados sendo duas repetições e 28 populações F_3 de soja e quatro tratamentos cultivares comerciais (BMX COMPACTA, TMG7061, TMG7062 e TMG7063). As parcelas consistiram em 4 linhas com um 1 de comprimento, com espaçamento de 0,45 m entre fileiras e densidade de 15 plantas por metro.

Tabela 1 - Análise química do solo na área experimental na camada de 0-20 cm.

Atributo	Valor
pH (H_2O	6,2
Al trocável (cmolc dm-3)	0
Ca+Mg ($cmol_c\ dm^{-3}$)	4,31
P ($mg\ dm^{-3}$)	41,3
K ($cmol_c\ dm^{-3}$)	0,2
M.O ($g\ dm^{-3}$)	19,74
V (%)	45
m (%)	0
S.B ($cmol_c\ dm^{-3}$)	2,3
CTC ($cmol_c\ dm^{-3}$)	5,1

Nota: pH (H_2O); Al trocável (cmolc dm-3): alumínio trocável; Ca: Cálcio; Mg: Magnésio; P: Fósforo; K: Potássio; M.O: Matéria orgânica; V: Índice de saturação por bases; CTC: Capacidade de troca catiônica.
Fonte: Próprio autor.

O preparo do solo para a semeadura foi convencional (aração e gradagem niveladora) em outubro de 2021. Os fungicidas (Piraclotrobina + Metil Tiofanato) e inseticida (Fipronil), foram utilizados para o tratamento de sementes contra os ataques de pragas e fungos de solo, na dose de 200 mL do produto comercial para cada 100 kg de sementes. As bactérias do gênero *Bradyrhizobium*, foram utilizadas para inoculação das sementes com dosagem de 200 mL de inoculante líquido concentrado para cada 100 kg. Os tratos culturais realizados foram referentes a necessidade da cultura.

A análise espectral de três amostras foliares foram escolhidas ao acaso da parte superior do dossel da planta de cada parcela. Foi utilizado o sensor hiperespectral Ocean Opticis modelo STS-VIS-L-50-400-SMA, que permitiu capturar a refletância entre os comprimentos de onda entre 450 e 824 nm. O sensor foi calibrado antes da leitura das amostras de cada bloco, utilizando-se uma placa de sulfato de bário para a calibração do branco e o isolamento do sensor para a calibração no escuro. A captura das refletâncias foi realizada em laboratório utilizando-se iluminação artificial por duas lâmpadas halógenas de 120 W apontadas para o ponto de leitura. O sensor estava conectado a um computador para os registros de cada leitura por meio de um software próprio do sensor chamado OceanView, que registra as leituras feitas pelo equipamento com resolução espectral de 1.5 nm.

As avaliações dos caracteres agronômicos destacam-se, altura de inserção da primeira vagem (AIV, cm), diâmetro da haste principal (DHP, cm), dias para maturação (DM), altura de plantas (AP, cm), número de ramificações (NR) e produtividade de grãos (PROD, kg ha^{-1}). Os caracteres, AP, DHP, AIV e NR foram analisados em cinco plantas relacionado a cada parcela, tornando-se os dois primeiros avaliados com auxílio de uma fita métrica e o terceiro de um paquímetro digital. Para A PROD houve uma correção para 13% de umidade e transformado para kg ha^{-1}. Foi avaliada pela colheita das fileiras centrais de cada parcela. Em cada unidade experimental o DM caracterizou-se relacionando aos dias entre emergência e maturação acima de 50% de plantas.

Os dados agronômicos e espectrais foram submetidos a análises estatísticas. A primeira análise feita foi a separação dos genótipos em grupos por meio do algoritmo k-means. O algoritmo k-means depende do valor de k que sempre precisa ser especificado antes da execução de qualquer análise de agrupamento (Ahmed *et al.*, 2020) em que no trabalho em questão o k que melhor separou os grupos de acordo com as características agronômicas utilizando a separação em dois grupos. Os genótipos em questão foram divididos em dois grupos denominados Cluster 1 e Cluster 2.

A classificação dos modelos de aprendizagem de máquinas utilizados se destaca: RePTree (DT) (Snousy *et al.*, 2011), Árvores de decisão J48 (J48) (Quinlan, 1993), Floresta aleatória (RF) (BELGIU; DRĂGUŢ, 2016), Redes neurais artificiais (RNA), (Egmont-Petersen; Ridder; Handels, 2002), Regressão logísitica (RL) (Štepanovský *et al.*, 2017), Máquina de vetor suporte (SVM) (Nalepa; Kawulok, 2019).

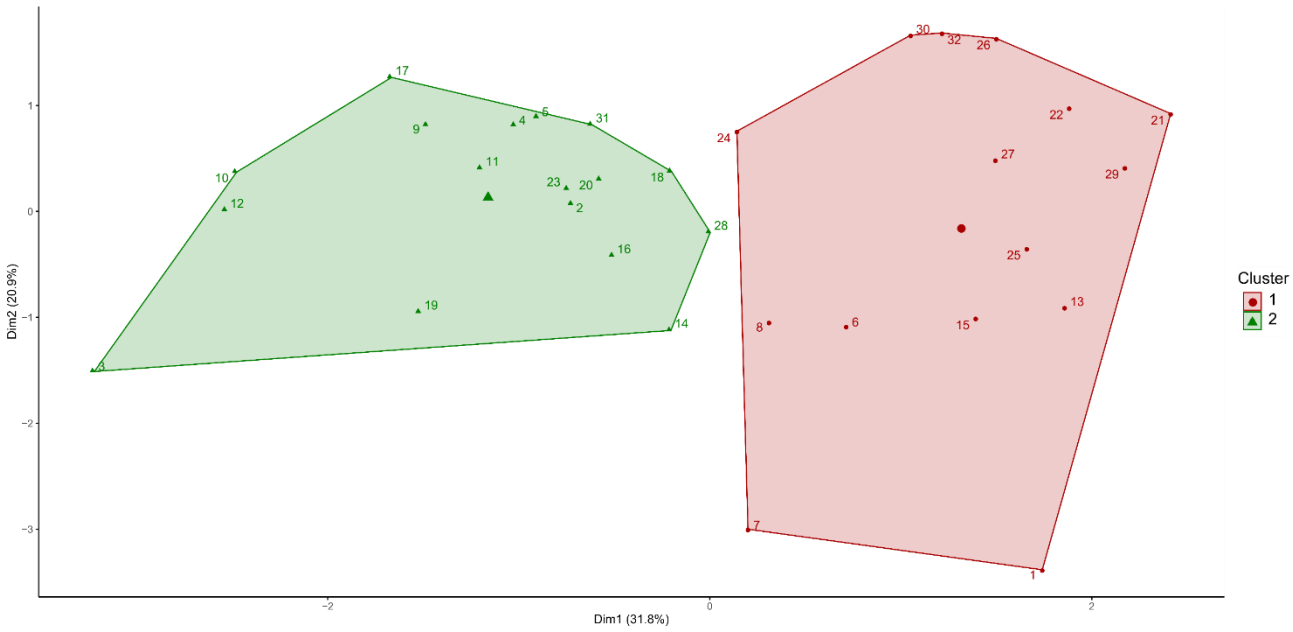
A validação cruzada estratificada acompanhada do k-fold = 10 e dez repetições (100 execuções por modelo) foi utilizado para classificação dos clusters. Somente para Redes Neurais Artificiais utilizou-se duas camadas juntamente a 10 neurônios cada, e para todos os outros os parâmetros dos modelos seguiram as configurações padrão do software Weka 3.8.5.

O desempenho de cada algoritmo foi mensurado por meio de duas métricas de acurácia: porcentagem de classificações corretas (CC) e F-score. Foram construídos boxplots para expressar acurácia dos modelos e dos inputs para cada métrica de acurácia. Pelo teste de Scott-Knott (Scott; Knott, 1974) foi feito o agrupamento de médias a 5% de significância. A análise PCA-Kmeans, as correlações de Pearson, os gráficos e as análises de médias foram conduzidas pelo software R (TEAM, 2013) e utilizando os pacotes ggplot2 e ExpDes.pt e Metan.

4.3 RESULTADOS

O algoritmo k-means foi utilizado, associado aos componentes principais para agrupar os genótipos de soja. Os genótipos de soja classificados em agrupamento destacam-se em Cluster 1 e Cluster 2 (Figura 1). Com base nos clusters formados, foram construídos boxplots considerando a média dos genótipos agrupados em cada cluster para as variáveis avaliadas (Figura 2).

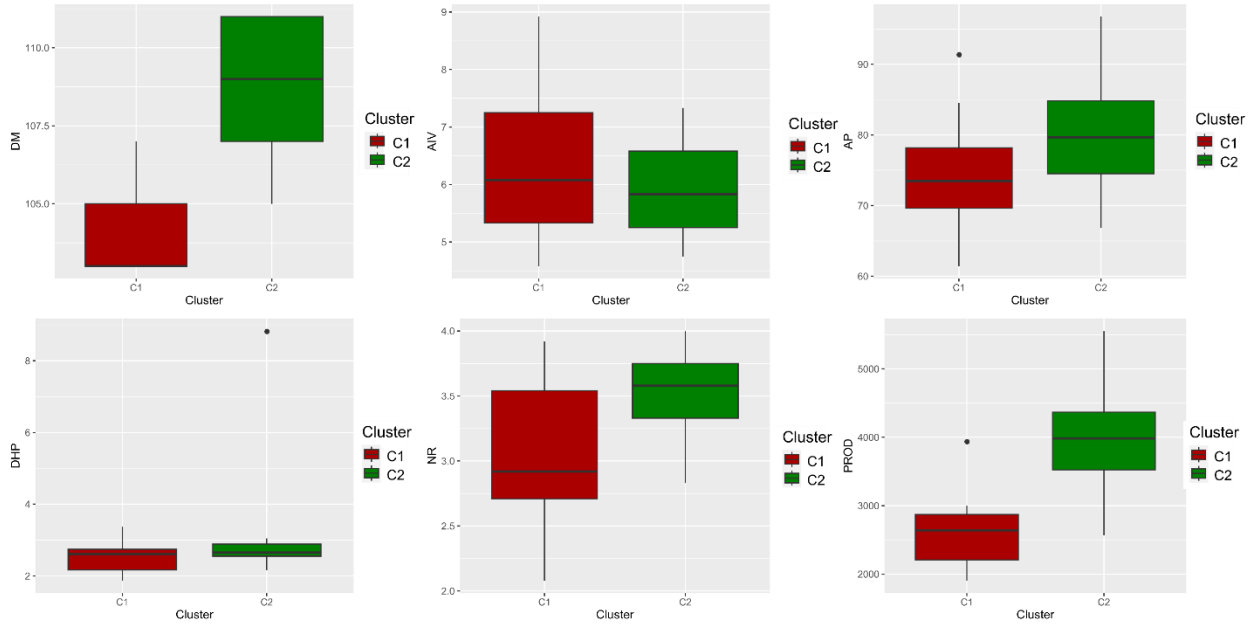
Figura 1 - Análise de PC para os 32 genótipos de soja avaliados e clusters formados pelo algoritmo k-means baseado nos caracteres agronômicos avaliados



Fonte: Próprio autor.

O Cluster 2 diferenciou-se do Cluster 1 em um período maior de dias, com uma média de 110 dias (Figura 2). Para AIV o Cluster 1 obteve maior altura, com uma média de 6 a 7cm. O Cluster 2, quando comparado à AP se diferenciou do Cluster 1 com uma média de 85cm de altura. Para DHP, o Cluster 2 diferencia-se do Cluster 1 com maior diâmetro, de 3,0 cm. O NR do Cluster 2 obteve uma média maior de 3,5 a 4. Com relação a produtividade, o Cluster 2 obteve maior resultado comparado ao Cluster 1, com uma média de PROD de 4.400 kg ha¹. Os genótipos de soja dentro do grupo Cluster 2, de uma forma geral obteve melhores resultados comparado ao Cluster 1 referente a todos os caracteres agronômicos.

Figura 2 - Boxplot com as médias de cada Cluster formados para as variáveis agronômicas DM, AIV, cm, AP, cm, DHP, cm, NR, PROD kg ha¹, avaliados em 32 genótipos de soja

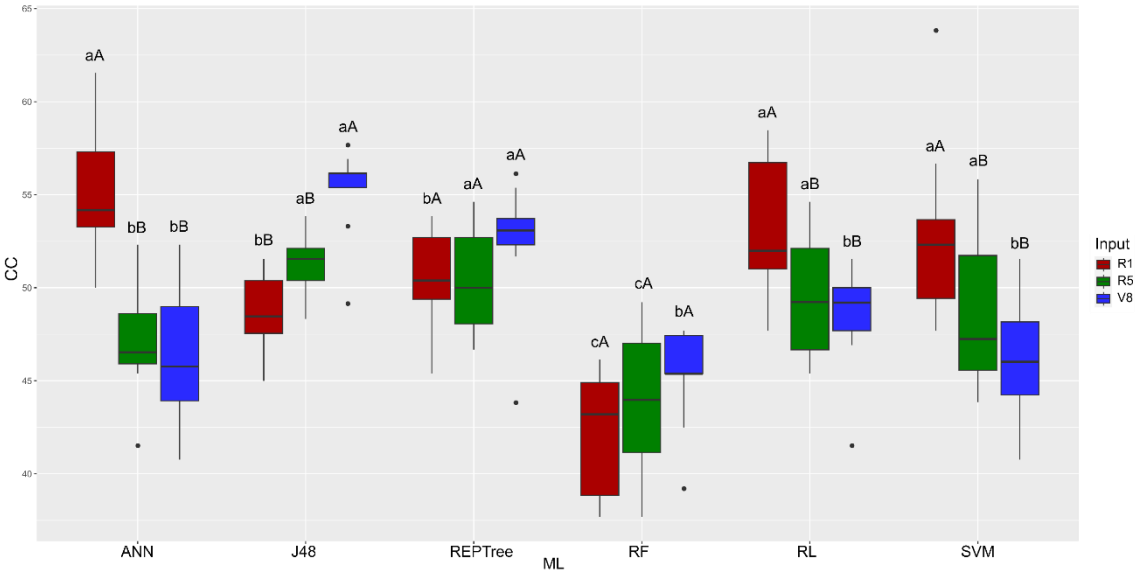


Fonte: Próprio autor.

De acordo com os Clusters formados (Figura 3) dos genótipos de soja, os algoritmos RL e SVM obtiveram melhores resultados para o estágio reprodutivo R1 e R5. O algoritmo RL obteve 57% em R1 e 52% em R5. O algoritmo SVM obteve uma média de 54% em R1 e 52% para R5. Para o estágio vegetativo V8, os modelos algoritmos que se destacaram foram o J48, 57,5% e REPTree, 54%. Invariavelmente para os inputs testados, os modelos REPTree e RF, tiveram os melhores desempenhos para classificação correta (CC).

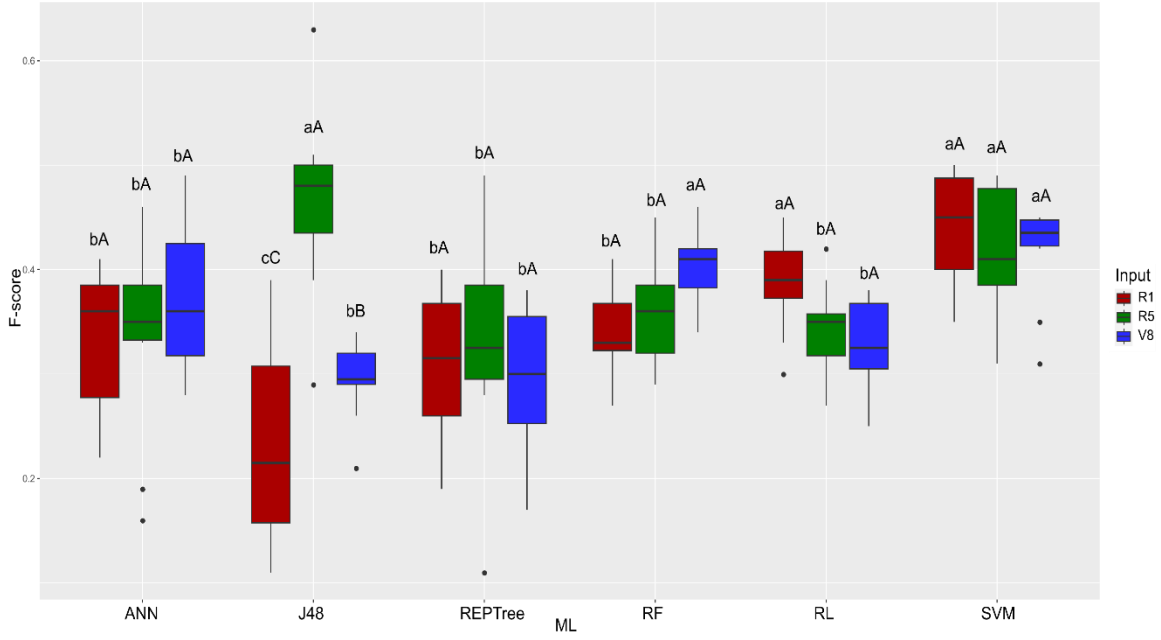
Para F-score testado nos estágios fenológicos, o algoritmo SVM se destacou em todos os parâmetros avaliados (Figura 4), R1, R5 e V8 apresentando boa performance. Porém o algoritmo J48 para o estágio fenológico R5 se diferenciou dos demais inputs avaliados.

Figura 3 - Boxplot para porcentagem CC para a interação significativa entre os modelos de aprendizado de máquinas e os inputs testados. As médias seguidas pelas mesmas letras maiúsculas para as diferentes entradas e as mesmas letras minúsculas para os diferentes algoritmos de AM não diferem pelo teste de Scott-Knott a 5% de probabilidade



Fonte: Próprio autor.

Figura 4 - Boxplot para coeficiente F-Score para a interação significativa entre os modelos de aprendizado de máquinas e os inputs testados. As médias seguidas pelas mesmas letras maiúsculas para as diferentes entradas e as mesmas letras minúsculas para os diferentes algoritmos de AM não diferem pelo teste de Scott-Knott a 5% de probabilidade



Fonte: Próprio autor.

Para a CC os algoritmos RL e SVM obtiveram melhores resultados para o estágio reprodutivo R1 e R5. Para o estágio vegetativo V8 os modelos REPTree e RF, tiveram os melhores desempenhos. O algoritmo SVM se destacou em todos os parâmetros avaliados para F-score, R1, R5 e V8 apresentando boa performance. J48 para o estágio fenológico R5 se destacou dos demais inputs avaliados.

4.4 DISCUSSÃO

Por captar dados espectrais em diversas bandas estreitas e contínuas, mensurando múltiplas bandas de absorção, das mais amplas às mais sutis, os sensores hiperespectrais é responsável por exercer essa função, tornando-se uma evolução dos sensores multispectrais. Vários trabalhos científicos têm demonstrado que devido ao rápido avanço tecnológico nos sistemas de sensoriamento remoto, tem contribuído economicamente e eficientemente para melhores práticas de tomada de decisão quanto aos caracteres agronômicos e manejo de culturas em anos decorrentes sucessores (Chlingaryan *et al.,* 2018). A forma mais eficiente e mais rápida, a refletância capturada do dossel da planta ao utilizar a fenotipagem é um excelente precursor para auxiliar a seleção de genótipos de soja (Santana *et al.,* 2022; Yoosefzadeh-Najafabadi *et al.,* 2021).

Referente ao que foi observado, os algoritmos de aprendizagem de máquinas, no que visa regular esse problema de não linearidade das variáveis espectrais juntamente com os caracteres agronômicos, podem ser utilizados. Devido a eficiência desses algoritmos, há uma robustez na resolução dos problemas de não linearidade das variáveis (Schwalbert *et al.,* 2020). Trazendo respostas conclusivas e produtivas deste estudo com informações concernentes a diversos aspectos, é devido praticamente da junção desta técnica de inteligência computacional, junto com sensoriamento remoto e dados capturados do campo.

Os clusters gerados são fundamentais na separação dos genótipos de soja. Quando separados, vemos em destaque os genótipos dentro do Cluster 2, em que sua diferenciação está relacionada de acordo com os caracteres genotípicos avaliados no que se refere a absorção de nutrientes e envolvendo todo processo metabólico. As variáveis espectrais e os caracteres agronômicos mesmo no proceder não linear é utilizável o AM para sua classificação genotípica referente aos Clusters gerados (Chaney, 2022; Osco *et al.,* 2020).

O procedimento de escolha do algoritmo de AM é muito importante no aspecto de complexidade na geração dos dados em correlação a fenotipagem (Singh *et al.,* 2016). Os algoritmos que se diferenciaram dos demais referentes aos caracteres agronômicos avaliados

deste estudo foram RL, SVM, J48 e REPTree. Em questão de eliminar a parcialidade e redução em erros de respostas e na contribuição de imagens automatizada a escolha pelo algoritmo J48 se destaca referente a melhoria de classificação. Mesmo com a quantidade de ruídos de seus dados e a complexidade e não linearidade, o algoritmo J48 se destaca referente as ANN. Ao comparar RF, devido sua sensibilidade de conjunto de dados o algoritmo J48 alcança melhores resultados (Aljawarneh *et al.,* 2019; Rossi *et al.,* 2020; Benzer *et al.,* 2022; Massawe *et al.,* 2018).

O método algoritmo SVM é eficaz na classificação de imagens para ocorrências de pragas e doenças na soja, e outras ocasiões, (Tetila *et al.,* 2020; Gui *et al.,* 2021). Corrales *et al.,* (2022) determinam que seu uso também pode ser eficiente na previsão de produtividade de soja. Ju *et al.,* (2021) declara também que para ter uma previsão de produtividade de soja baseado em IVs o SVM também é um excelente método. Para medição do teor de clorofila na soja utilizando as imagens RGB Hassanijalilian *et al.,* (2020) definem que o SVM é uma excelente técnica de aprendizado de máquina.

As bandas espectrais referentes aos inputs testados mostraram relevantes em todos os algoritmos de AM principalmente devido sua praticidade, exemplo da não necessidade de cálculos matemáticos como o de índice de vegetação. Há, porém, uma sensibilidade das bandas quanto a dimensão de lesões em folhas danificadas por insetos, classificação quanto aos genótipos de soja, porém não destrutiva, e a acurácia na distinção de variáveis de soja (Alves *et al.,* 2019; Silva Junior *et al.,* 2020; Luna *et al.,* 2013).

4.5 CONCLUSÕES

O período de avaliação dos dados hiperespectrais depende do algoritmo de aprendizagem utilizado para classificação dos genótipos de soja quanto aos caracteres agronômicos.

De forma geral, os estágios fenológicos mais indicados para obtenção dos dados hiperespectrais e classificação dos genótipos de soja são R1 e R5 associado ao uso dos modelos de regressão logística e máquina de vetor suporte.

REFERÊNCIAS

ALJAWARNEH, S.; YASSEIN, M. B.; ALJUNDI, M. An enhanced J48 classification algorithm for the anomaly intrusion detection systems. **Cluster Computing**, New York, v. 22, n. 5, p. 10549–10565, 2019.

ALVES, T. M.; MOON, R. D.; MACRAE, I. V; KOCH, R. L. Optimizing band selection for spectral detection of Aphis glycines Matsumura in soybean. **Pest Management Science**, Oxford, v. 75, n. 4, p. 942–949, 2019.

BENZER, S.; GARABAGHI, F. H.; BENZER, R.; MEHR, H. D. Investigation of some machine learning algorithms in fish age classification. **Fisheries Research**, [*s. l.*], v. 245, p. 106-151, 2022.

CAPELIN, D.; DANELUZZI, G. S.; OLIVEIRA, R. F.; BRESSAN, D. F.; CORREA, C. V.; JOCA, T. A. C.; BORETTO, F. **Utilização do IRGA**: analisador de gases por infravermelho para avaliação de trocas gasosas em plantas: check list de preparação. O Estresse das Plantas - Teoria & Prática, [*s. l.*] p. 187–194, 2017.

CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. MACHINE. Learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. **Computers and Electronics in Agriculture**, Amsterdam, v. 151, p 61–69, 2018.

COMPANHIA NACIONAL DE ABSTECIMENTO - CONAB. **Monitoramento Agrícola.** Disponível online: file:///C:/Users/User/Downloads/E-book_BoletimZdeZSafrasZ Z8Zlevantamento%20(1).pdf. Acesso em: 30 jun. 2024.

CORRALES, D. C.; SCHOVING, C.; RAYNAL, H.; DEBAEKE, P.; JOURNET, E.-P.; CONSTANTIN, J. A. Surrogate model based on feature selection techniques and regression learners to improve soybean yield prediction in southern France. **Computers and Electronics in Agriculture**, Amsterdam, v. 192, p. 106-578, 2022.

FORMAGGIO, A. R.; SANCHES I, D. **Sensoriamento remoto em agricultura**. São Paulo: Oficina de Textos, 2017.

GAO, F.; ANDERSON, M. C.; ZHANG, X.; YANG, Z.; ALFIERI, J. G.; KUSTAS, W. P.; MUELLER, R.; JOHNSON, D. M.; PRUEGER, J. H. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. **Remote Sensing of Environment**, Amsterdam, v. 188, p. 9–25, 2017.

GUI, J.; FEI, J.; WU, Z.; FU, X.; DIAKITE, A. Grading method of soybean mosaic disease based on hyperspectral imaging technology. **Information Processing in Agriculture**, [*s. l.*], v. 8, n. 3, p. 380–385, 2021.

GUPTA, S. K.; MANJAYA, J. G. Advances in improvement of soybean seed composition traits using genetic, genomic and biotechnological approaches. **Euphytica**, [*s. l.*], v. 218, n. 7, p. 99, 2022.

HASSANIJALILIAN, O.; IGATHINATHANE, C.; DOETKOTT, C.; BAJWA, S.; NOWATZKI, J.; HAJI ESMAEILI, S. A. Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. **Computers and Electronics in Agriculture**, [*s. l.*], v. 174, p. 105433, 2020.

HIRAKURI, M. H.; LAZZAROTTO, J. J. **O agronegócio da soja nos contextos mundial e brasileiro**. Londrina: Embrapa Soja, 2014.

JENSEN, J. R.; EPIPHANO, J. C. Remote sensing of the environment: an earth resource perspective. São José dos Campos: **Parêntese**, [*s. l.*], p. 604, 2011.

LING, B.; GOODIN, D.G.; RAYNOR, E.J.; JOERN, A. Hyperspectral analysis of leaf pigments and nutritional elements in tallgrass prairie vegetation. **Frontiers in Plant Science**, Lausanne, v. 10, p. 1–13, 2019.

JU, S.; LIM, H.; MA, J. W.; KIM, S.; LEE, K.; ZHAO, S.; HEO, J. Optimal county-level crop yield prediction using MODIS-based variables and weather data: A comparative study on machine learning models. **Agricultural and Forest Meteorology**, [*s. l.*], v. 307, p. 108-530, 2021.

LIU, K. Soybeans: chemistry, technology, and utilization. [*s.l.*] : **Springer**, 2012.

LUNA, A. S.; DA SILVA, A. P.; PINHO, J. S. A.; FERRÉ, J.; BOQUÉ, R. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, [*s. l.*], v. 100, p. 115–119, 2013.

KARYAWATI, A. S.; . SITOMPUL, S. M.; BASUKI, N.; SITAWATI. Combining Ability Analysis for Physiological Characters of Soybean (*Glycine max* L. Merrill). **International Journal of Plant Research**, [*s. l.*] v. 5, n. 5, p. 113–121, 2015.

MALMIR, M.; TAHMASBIAN, I.; XU, Z.; FARRAR, M.B.; BAI, S.H. Prediction of macronutrients in plant leaves using chemometric analysis and wavelength selection. J. **Soils Sediments,** [*s. l.*], v. 20, p. 249-259, 2020.

MASSAWE, B. H. J.; SUBBURAYALU, S. K.; KAAYA, A. K.; WINOWIECKI, L.; SLATER, B. K. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. **Geoderma**, [*s. l.*], v. 311, p. 143–148, 2018.

OSCO, L.; JUNIOR, J.; RAMOS, A.; FURUYA, D.; SANTANA, D.; TEODORO, L.; GONÇALVES, W.; BAIO, F.; PISTORI, H.; JUNIOR, C.; ET AL. Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. **Remote Sensing**, [*s. l.*], v. 12, n. 19, p. 3237, 2020.

ROMÁN, J. R.; RODRÍGUEZ-CABALLERO, E.; RODRÍGUEZ-LOZANO, B.; RONCERO-RAMOS, B.; CHAMIZO, S.; ÁGUILA-CARRICONDO, P.; Cantón, Y. Spectral response analysis: An indirect and non-destructive methodology for the chlorophyll quantification of biocrusts. **Remote Sensing,** [*s. l.*], v. 11, n. 11, p. 1350, 2019.

ROSSI, F. S.; SILVA JUNIOR, C. A. Da; OLIVEIRA-JÚNIOR, J. F. De; TEODORO, P. E.; SHIRATSUCHI, L. S.; LIMA, M.; TEODORO, L. P. R.; TIAGO, A. V.; CAPRISTO-SILVA, G. F. Identification of tillage for soybean crop by spectro-temporal variables, GEOBIA, and decision tree. **Remote Sensing Applications: Society and Environment**, [*s. l.*], v. 19, p. 100356, 2020.

SANTANA, D. C.; DE OLIVEIRA CUNHA, M. P.; DOS SANTOS, R. G.; COTRIM, M. F.; TEODORO, L. P. R.; DA SILVA JUNIOR, C. A.; BAIO, F. H. R.; TEODORO, P. E. High-throughput phenotyping allows the selection of soybean genotypes for earliness and high grain yield. **Plant Methods**, [*s. l.*], v. 18, n. 1, p. 13, 2022.

SCHWALBERT, R. A.; AMADO, T.; CORASSA, G.; POTT, L. P.; PRASAD, P. V. V.; CIAMPITTI, I. A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. **Agricultural and Forest Meteorology**, [*s. l.*], v. 284, p. 107886, 2020.

SILVA JUNIOR, C. A. DA; TEODORO, L. P. R.; TEODORO, P. E.; BAIO, F. H. R.; PANTALEÃO, A. de A.; CAPRISTO-SILVA, G. F.; FACCO, C. U.; OLIVEIRA-JÚNIOR, J. F. De; SHIRATSUCHI, L. S.; SKRIPACHEV, V.; LIMA, M.; NANNI, M. R. Simulating multispectral MSI bandsets (Sentinel-2) from hyperspectral observations via spectroradiometer for identifying soybean cultivars. **Remote Sensing Applications: Society and Environment**, [*s. l.*], v. 19, p. 100328, 2020.

TEODORO, P. E.; TEODORO, L. P. R.; BAIO, F. H. R.; SILVA JUNIOR, C. A.; SANTOS, R. G.; RAMOS, A. P. M.; PINHEIRO, M. M. F.; OSCO, L. P.; GONÇALVES, W. N.; CARNEIRO, A. M.; JUNIOR, J. M.; PISTORI, H.; SHIRATSUCHI, L. S. Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. **Remote Sensing of Environment**, Amsterdam, v. 13, n. 22, p. 4632, 2021.

TETILA, E. C.; MACHADO, B. B.; ASTOLFI, G.; BELETE, N. A. de S.; AMORIM, W. P.; ROEL, A. R.; PISTORI, H. Detection and classification of soybean pests using deep learning with UAV images. **Computers and Electronics in Agriculture**, [*s. l.*], v. 179, p. 105836, 2020.

YOOSEFZADEH-NAJAFABADI, M.; EARL, H. J.; TULPAN, D.; SULIK, J.; ESKANDARI, M. Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean. **Frontiers in Plant Science**, Lausanne, v. 11, p. 1-14, 2021.

ZHOU, J.; YUNGBLUTH, D.C.; VONG, C.N.; SCABOO, A.M.; ZHOU, J. Estimation of maturity date of soybean breeding lines using UAV-based imagery. **Remote Sensing of Environment**, Amsterdam, v. 11, p. 2058-2075, 2019.