

UNIVERSIDADE FEDERAL DE VIÇOSA

**Predição de mofo-branco do feijoeiro-comum utilizando técnicas de
aprendizagem de máquinas**

Flávio Souza Santos
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2025**

FLÁVIO SOUZA SANTOS

**Predição de mofo-branco do feijoeiro-comum utilizando técnicas de
aprendizagem de máquinas**

Tese apresentada à Universidade Federal
de Viçosa, como parte das exigências do
Programa de Pós-Graduação em
Engenharia Agrícola, para obtenção do
título de *Doctor Scientiae*.

Orientador: Daniel Marcal de Queiroz

Coorientadores: Domingos S. M. Valente
Francisco de A. de C.
Pinto
Jose E. de S. Carneiro

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S237p
2025

Santos, Flávio Souza, 1988-

Predição de mofo-branco do feijoeiro-comum utilizando técnicas de aprendizagem de máquinas / Flávio Souza Santos. – Viçosa, MG, 2025.

1 tese eletrônica (113 f.): il. (algumas color.).

Orientador: Daniel Marçal de Queiroz.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, 2025.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2026.556>

Modo de acesso: World Wide Web.

1. Feijão - Doenças e pragas. 2. *Phaseolus vulgaris*. 3. *Sclerotinia sclerotiorum*. 4. Sensoriamento remoto. 5. Aprendizado do computador. I. Queiroz, Daniel Marçal de, 1957-. II. Universidade Federal de Viçosa. Departamento de Engenharia Agrícola. Programa de Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 635.6529

FLÁVIO SOUZA SANTOS

**Predição de mofo-branco do feijoeiro-comum utilizando técnicas de
aprendizagem de máquinas**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 14 de fevereiro de 2025.

Assentimento:

Flávio Souza Santos
Autor

Daniel Marcal de Queiroz
Orientador

Essa tese foi assinada digitalmente pelo autor em 25/09/2026 às 14:43:50 e pelo orientador em 25/09/2026 às 14:57:58. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **EZGA.9WTW.8SEV** e clique no botão 'Validar documento'.

Aos meus pais,
José Flávio e Rosa Maria, por todo amor,
apoio e esforços a que me dedicaram.

AGRADECIMENTOS

Aos meus pais por terem proporcionado a mim oportunidades que me trouxeram crescimento e incentivo para concluir mais esta etapa.

À Universidade Federal de Viçosa e ao Departamento de Engenharia Agrícola pela oportunidade de realizar a pós-graduação.

Ao meu orientador Daniel Marçal de Queiroz, pela paciência, orientação, atenção, comprometimento e incentivo durante este trabalho.

Aos Doutores Rogério Faria Vieira e Pablo Henrique Teixeira, em nome da Epamig (Empresa de Pesquisa Agropecuária de Minas Gerais), por toda cooperação e apoio fundamental para a realização deste trabalho.

Ao Professor José Eustáquio de Souza Carneiro por ter aceitado a participar da coordenação, e agradecendo-o também em nome do Departamento de Agronomia, por ter cedido espaço para a implementação dos experimentos.

Aos coorientadores Francisco de Assis de Carvalho Pinto e Domingos Sárvio Magalhães Valente pelas considerações e contribuições para o trabalho.

Ao professor Marconi e ao Técnico Administrativo Vicente Firmino Filho, pela fundamental atenção dada em se disporem em me levar à Epamig na cidade de Oratórios para a realização das coletas de dados deste trabalho.

Ao professor Genelício Crusoé Rocha, sempre solícito e com presteza em me receber no Departamento de Solos e no Laboratório de Física de Solos, juntamente ao Cláudio Márcio Brustolini, técnico do laboratório.

Aos colegas e amigos do nosso grupo de pesquisa, Charles Cardoso Santana, Giovani Dambroz, Marcelo Portes, Miquéias Pereira, Matheus Assunção, Sarah Abade, Maria Luiza Marangon e Igor Araújo, e aos demais do Laboratório de Mecanização Agrícola, Bia, Igor Quintão, Jussara, Pedro, Poliana, Priscila, Rodrigo e Sérgio, pela ajuda e momentos de descontração nesta caminhada do doutorado.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“Toda pessoa que encontro me é superior em alguma coisa,
E, nesse particular, aprendo com ela.”
(Adaptado de Ralph Waldo Emerson)

RESUMO

SANTOS, Flávio Souza, D.Sc., Universidade Federal de Viçosa, fevereiro de 2025. **Predição de mofo-branco do feijoeiro-comum utilizando técnicas de aprendizagem de máquinas.** Orientador: Daniel Marcal de Queiroz. Coorientadores: Domingos Sarvio Magalhaes Valente, Francisco de Assis de Carvalho Pinto e Jose Eustaquio de Souza Carneiro.

O feijão é uma importante fonte alimentar em várias regiões do mundo, fornecendo proteínas, carboidratos e minerais essenciais. No entanto, os produtores enfrentam desafios na condução das lavouras, sendo a doença denominada mofo-branco (causado pelo fungo *Sclerotinia sclerotiorum*), um dos problemas mais sérios. Esta doença pode resultar em perdas de até 90% na produtividade. A detecção precoce do mofo-branco é dificultada pelas características tanto do fungo quanto da cultura. Técnicas emergentes, como o uso de aprendizado de máquina, têm se mostrado promissoras para a identificação dessa doença. Essas técnicas têm sido eficazes em várias aplicações agrícolas, incluindo a previsão de condições ambientais, contagem automática de frutos e avaliação da severidade de doenças em culturas. Os objetivos deste estudo foram: (i) detectar a incidência e severidade do mofo-branco, além de prever a produtividade do feijoeiro-comum em condições de infecção, utilizando imagens multiespectrais obtidas por câmeras acopladas a Aeronaves Remotamente Pilotadas (RPAs) e técnicas de aprendizado de máquina; e (ii) desenvolver um modelo preditivo para estimar o risco de ocorrência do mofo-branco, com base em dados de campo sobre a doença e a produtividade da cultura. Foram aplicados quatro algoritmos de aprendizado de máquina: Cubist, Kernel K Nearest Neighbors (KNN), Random Forest (RF) e Support Vector Machine (SVM). Os modelos foram treinados com dados espectrais obtidos de câmeras multiespectrais acopladas a RPAs em experimentos de campo, utilizando 51 índices de vegetação calculados a partir das imagens adquiridas, além de considerar o espaçamento da cultura em cada experimento. Esses dados foram utilizados para prever o risco de ocorrência do mofo-branco e a produtividade do feijoeiro-comum. Os algoritmos Cubist e KNN se destacaram nas previsões, com o Cubist mostrando melhor desempenho na detecção da incidência e severidade do mofo-branco, utilizando principalmente as refletâncias nas bandas verde e azul aos 38, 52 e 59 dias após a emergência (DAE), além dos índices de vegetação MSR (Razão Simples Modificada) e EVI (Índice de Vegetação Melhorado). Na predição de produtividade, o KNN, juntamente com o RF, proporcionou bons resultados, sendo o modelo de predição de produtividade mais preciso do que os modelos de predição do mofo-branco. Além disso, foram aplicados modelos de aprendizado de máquina

para prever a produtividade e o risco de ocorrência do mofo-branco, com base em variáveis como densidade de plantio, tipo de cultivar, fungicida utilizado, precipitação e temperatura (mínima, média e máxima). Os algoritmos SVM, RF e Linear Model (LM) apresentaram desempenhos promissores.

Palavras-chave: agricultura digital; sensoriamento remoto; *Phaseolus vulgaris*; *Sclerotinia sclerotiorum*; dados multiespectrais

ABSTRACT

SANTOS, Flávio Souza, D.Sc., Universidade Federal de Viçosa, February, 2025.
Prediction of white mold in common bean using machine learning techniques.
Adviser: Daniel Marcal de Queiroz. Co-advisers: Domingos Sarvio Magalhaes Valente, Francisco de Assis de Carvalho Pinto and Jose Eustaquio de Souza Carneiro.

Common bean is a major food source in several regions of the world, providing proteins, carbohydrates, and essential minerals. However, growers face significant challenges in crop management, with white mold (caused by the fungus *Sclerotinia sclerotiorum*) being one of the most serious problems. This disease can lead to yield losses of up to 90%. Early detection of white mold is hindered by the biological characteristics of both the pathogen and the crop. Emerging techniques, such as the use of machine learning, have shown promise for disease identification. These approaches have proven effective in various agricultural applications, including forecasting environmental conditions, automatic fruit counting, and assessing disease severity in crops. The objectives of this study were: (i) to detect the incidence and severity of white mold and to predict common bean yield under infection conditions using multispectral images acquired by cameras mounted on remotely piloted aircraft (RPAs) combined with machine learning techniques; and (ii) to develop a predictive model to estimate the risk of white mold occurrence based on field data on disease incidence and crop yield. Four machine learning algorithms were applied: Cubist, Kernel K Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). The models were trained with spectral data obtained from multispectral cameras mounted on RPAs in field experiments, using 51 vegetation indices calculated from the acquired images, along with crop spacing in each experiment. These data were used to predict both the risk of white mold occurrence and common bean yield. Among the algorithms, Cubist and KNN stood out in prediction accuracy. Cubist performed best in detecting white mold incidence and severity, primarily using reflectance in the green and blue bands at 38, 52, and 59 days after emergence (DAE), as well as the Modified Simple Ratio (MSR) and Enhanced Vegetation Index (EVI). For yield prediction, KNN and RF produced robust results, with the yield prediction model proving more accurate than the white mold prediction models. Additionally, machine learning models were applied to predict both yield and white mold risk based on variables such as planting density, cultivar type, fungicide use, precipitation, and temperature (minimum, mean, and maximum). The SVM, RF, and Linear Model (LM) algorithms also showed promising performance.

Keywords: digital agriculture; remote sensing; *Phaseolus vulgaris*; *Sclerotinia sclerotiorum*; multispectral data.

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO GERAL	11
1.1 IMPORTÂNCIA DA DETECÇÃO DO MOFO-BRANCO NA CULTURA DO FEIJÃO	11
1.2. OBJETIVOS	14
1.3. ESTRUTURA DA TESE	15
1.4. REFERÊNCIAS.....	15
CAPÍTULO 2 - DETECÇÃO DE MOFO-BRANCO (<i>SCLEROTINIA SCLEROTIORUM</i>) NO FEIJOEIRO COM BASE EM IMAGENS MULTIESPECTRAIS E TÉCNICAS DE APRENDIZADO DE MÁQUINA	19
2.1. INTRODUÇÃO	20
2.2. MATERIAL E MÉTODOS.....	22
2.3. RESULTADOS.....	37
2.5. DISCUSSÃO	59
2.6. CONCLUSÃO	67
2.7. AGRADECIMENTOS	67
2.8. REFERÊNCIAS.....	68
CAPÍTULO 3 - DESENVOLVIMENTO DE MODELO DE PREDIÇÃO DA SEVERIDADE DO MOFO-BRANCO EM FEIJOEIRO-COMUM COM BASE EM DADOS CLIMÁTICOS UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINAS.....	77
3.1. INTRODUÇÃO	78
3.2. MATERIAL E MÉTODOS.....	80
3.3. RESULTADOS.....	87
3.4. DISCUSSÃO	104
3.5. CONCLUSÕES	108
3.6. AGRADECIMENTOS	109
3.7. REFERÊNCIAS.....	109
CAPÍTULO 4 - CONCLUSÕES GERAIS	113

CAPÍTULO 1 - INTRODUÇÃO GERAL

1.1 IMPORTÂNCIA DA DETECÇÃO DO MOFO-BRANCO NA CULTURA DO FEIJÃO

Para atender ao consumo do feijão (*Phaseolus vulgaris*) em vários países, incluído o Brasil, é necessário manter otimizada a eficiência da cadeia de produção desse produto. Além de sua função como fonte alimentar, o feijão tem significativa influência nos setores econômico, social e cultural do país, além de impactar positivamente na saúde física e mental da população (Conab, 2018). O feijão pode ser cultivado em todo o território do Brasil, desde que respeitadas as exigências climáticas específicas da espécie. Em Minas Gerais, a produtividade média de feijão foi de 1.707 kg/ha na safra de 2022/2023 e, atualmente, o estado é o segundo maior produtor nacional com 18,21%, da produção brasileira (Conab, 2023). A produção mineira foi de 553,1 mil de toneladas de feijão na safra 2022/2023.

O feijão é produzido em três safras: a semeada entre outubro e novembro (denominada safra das “águas”), a semeada de fevereiro a março (denominada safra da “seca”) e a semeada de abril a julho (denominada safra de outono-inverno). Essas safras também são denominadas de primeira, segunda e terceira safras, respectivamente. O feijão de outono-inverno também é conhecido como feijão irrigado, por ser indispensável o uso da irrigação em seu cultivo (Leite, 2005). Sob a ótica do uso da terra, o feijão irrigado possibilita um melhor uso das áreas de cultivo da propriedade, aumentando a geração de renda. No entanto, a irrigação, necessária para a boa produção da lavoura, eleva os níveis de umidade no solo. A elevada umidade do solo somada às temperaturas amenas do outono-inverno favorece o desenvolvimento de doenças fúngicas como o mofo-branco, causado pelo fungo *Sclerotinia sclerotiorum* (Lib.) de Bary.

O fungo *S. sclerotiorum* possui ampla gama de hospedeiros, perfazendo 408 espécies dentre 278 gêneros e 75 famílias de plantas (Boland & Hall, 1994). A infecção pelo fungo geralmente coincide com o fechamento do dossel e a época de floração da lavoura (Canteri et al., 1999). Ela é mais severa com temperaturas moderadas, entre 15 e 25 °C, elevados teores de água do solo (Paula Júnior et al.,

2006) e elevada umidade relativa do ar (Bardin; Huang, 2001). A fonte primária de inóculo do mofo-branco é a liberação de ascósporos pelos apotécios, que surgem da germinação das estruturas de resistência, os escleródios (Clarkson et al., 2003). Quando infectada, a planta manifesta sintomas que podem ocorrer em flores, folhas, hastes e vagens, como lesões encharcadas, de coloração parda e consistência mole, além de micélio branco de aspecto cotonoso (Leite, 2005; Paula Júnior et al., 2006). Esses sintomas causam mudanças nas respostas espectrais das plantas, conforme verificado por Machado et al. (2015b).

Uma vez presente numa área, o mofo-branco pode permanecer no solo por até oito anos na forma de escleródios (Adams; Ayers, 1979). A germinação do escleródio é influenciada por diversos fatores ambientais. Mesmo que os principais fatores edafoclimáticos para a germinação esclerótica sejam a luz, a temperatura e a umidade do solo (Abawi; Grogan, 1979; Letham, 1975), outros fatores também exercem influência na germinação, como o estado nutricional do substrato, pH e aeração do solo, idade e profundidade dos escleródios no solo (Willettts; Wong, 1980). Assim, para o desenvolvimento de sistemas de previsões de ocorrência da doença, é necessário que sejam quantificadas as interações que afetam o ciclo de vida do mofo-branco (Sun; Yang, 2000).

Ainda que a identificação e monitoramento da ocorrência de fitopatógenos sejam realizados visualmente em campo, essa atividade também pode ser realizada por meio de sensoriamento remoto (SR), utilizando as respostas espectrais das plantas. O SR é uma técnica que envolve a obtenção de dados de objetos na superfície terrestre sem que o dispositivo de sensoriamento esteja em contato direto com esses objetos (Moreira, 2001). Diversos trabalhos utilizaram o SR com base em índices de vegetação na identificação de doenças em várias culturas, como em trigo (Franke; Menz, 2007; HE et al., 2020), arroz (Qin; Zhang, 2005; Wang; Zhang; Zhou, 2017), amendoim (Omran, 2017; Chen et al., 2019), algodão (Wang et al., 2020), soja (Bajwa; Rupe; MASON, 2017) tomate (Prince; Clarkson; Rajpoot, 2015; Abdulridha et al., 2020), citros (Abdulridha, Batuman; Ampatzidis, 2019; Deng et al., 2020), feijão (Boechat et al., 2014; Machado et al., 2015a), o que ratifica o uso potencial do SR como ferramenta na identificação de doenças em plantas.

A aquisição de dados por SR pode ser feita por meio, entre outros, de satélites (nível orbital) ou por Aeronaves Remotamente Pilotadas (RPAs), em nível suborbital. Por mais que o uso de satélites permita mapear grandes áreas, eles proporcionam baixa resolução espacial, podendo ainda sofrer restrições pela presença de nuvens. A baixa resolução temporal é outro problema quando se quer acompanhar as fases fenológicas da cultura (Matese et al., 2015). Já a utilização de RPAs na agricultura apresenta algumas vantagens em relação ao sensoriamento orbital: flexibilidade de obtenção de imagens, facilidade de operação, elevada resolução espacial e a possibilidade de aquisição de dados no momento ideal para se realizar o monitoramento (Deng et al., 2018). Em consequência dessas vantagens dos RPAs no SR, Wang et al. (2020) e Thomasson et al. (2018) obtiveram sucesso na identificação e classificação entre presença e ausência da doença da podridão da raiz no algodoeiro.

No que tange ao uso do SR na identificação do mofo-branco na cultura do feijão, há poucos trabalhos realizados. Os trabalhos de Machado et al. (2015a), Machado et al. (2015b), Boechat et al. (2014) e Hughes (2019) utilizaram técnicas espectrais na detecção do mofo-branco em feijoeiro. Esses autores demonstraram que o uso das técnicas espectrais na detecção do mofo-branco foi eficiente. No entanto, a metodologia pode ser melhorada com a otimização do uso de vários índices espectrais. Por isso, é justificável realizar mais pesquisas para atender a constante demanda de manejo mais eficiente da produção e melhor entendimento da dinâmica da ocorrência do mofo-branco no feijoeiro. O acompanhamento em campo e a contínua aquisição de imagens são importantes no monitoramento de doenças na lavoura. Modelos desenvolvidos com base em técnicas de aprendizagem de máquinas podem ser utilizados nas predições de risco de doenças.

A aprendizagem de máquinas é um tipo de inteligência artificial que envolve o desenvolvimento de modelos com base em dados. Para isso, são usados algoritmos sem depender de equações matemáticas escritas para representar o sistema em análise (Mekonnen et al., 2019). Entre os modelos de aprendizado de máquinas que se destacam estão o *Support Vector Machine* (SVM), *Random Forest* (RF), Cubist e *K-Nearest-Neighbours* (KNN). Esses modelos têm sido utilizados com sucesso na

identificação de padrões de comportamento de crescimento de roseiras por meio da previsão de dados do ambiente agrícola (Rodríguez; Gualotuña; Grilo, 2017), na contagem de frutos em ramos de café (Ramos et al., 2017), na identificação de doença na cultura do dendê (Santoso; Tani; Wang, 2017), na aceleração de fenotipagem e diagnóstico de doenças em plantas (Yamamoto; Togami; Yamaguchi, 2017) e em medições de incidência e severidade de doenças de plantas com base em imagens de folhas (Mwebaze; Owomugisha, 2016). Logo, o aprendizado de máquinas, dentre outras técnicas de inteligência artificial, pode ajudar os agricultores a maximizar o retorno financeiro da propriedade e permitir o uso de recursos de forma sustentável (Mekonnen et al., 2019).

Devido à elevada severidade do mofo-branco na agricultura brasileira, o Ministério da Agricultura, Pecuária e Abastecimento (Mapa) o incluiu na Portaria N°5, de 21 de agosto de 2015, que trata da busca por soluções de novos produtos e tecnologias para o controle de pragas e doenças de maior risco fitossanitário e importância econômica. Atualmente, é rotina a aplicação preventiva de fungicidas no feijoeiro quando a lavoura está no início do florescimento, mesmo sem sinais e sintomas do mofo-branco. Conforme exposto anteriormente, pesquisas já realizadas indicam ser possível o desenvolvimento de um modelo de predição de mofo-branco gerado por meio de imagens espectrais. Esse modelo poderia ser usado para detectar o início da presença do mofo-branco no feijoeiro, e uma vez detectada a infestação em determinado local da lavoura, a aplicação localizada, ao invés de uma aplicação em área total, poderia ser realizada. Caso o sistema verifique a não existência do fungo em toda a área, a aplicação preventiva poderia ser evitada. Dessa forma, o emprego do modelo possibilitaria a otimização da utilização de fungicidas e a redução dos possíveis impactos ambientais, econômicos e sociais causados pelo seu uso, juntamente com o aumento de produtividade.

1.2. OBJETIVOS

O objetivo geral deste trabalho foi desenvolver modelos de aprendizado de máquina para estimar a intensidade do mofo-branco e predizer seus efeitos sobre a produtividade do feijoeiro-comum. Para isso, foram estabelecidos os seguintes

objetivos específicos: i) estimar a incidência e a severidade do mofo-branco e prever a produtividade do feijoeiro-comum a partir de imagens multiespectrais; e ii) estimar a severidade do mofo-branco e prever a produtividade do feijoeiro-comum a partir de dados meteorológicos.

1.3. ESTRUTURA DA TESE

Este trabalho foi dividido em quatro capítulos. No primeiro capítulo é apresentada uma introdução geral. No segundo, é apresentada a proposta de aplicação de modelos de aprendizado de máquinas para prever a incidência e severidade do mofo-branco no feijoeiro-comum usando sensoriamento remoto, assim como prever a produtividade em consequência da infecção do mofo-branco. No terceiro, é apresentado o uso da aprendizagem de máquinas aplicado a um banco de dados climáticos e de densidades de plantio na previsão da severidade do mofo-branco no feijoeiro-comum. Por fim, no quarto capítulo são apresentadas as conclusões gerais e considerações finais da tese.

1.4. REFERÊNCIAS

- ABAWI, G. S.; GROGAN, R. G. Epidemiology of diseases caused by *Sclerotinia* species. **Phytopathology**, v. 69, n. 8, p. 899-904, 1979.
- ABDULRIDHA, J.; ABDULRIDHA, J.; AMPATZIDIS, Y.; KAKARLA, S. C.; ROBERTS, P. Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. **Precision Agriculture**, v. 21, n. 5, p. 955-978, 2020.
- ABDULRIDHA, J.; BATUMAN, O.; AMPATZIDIS, Y. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. **Remote Sensing**, v. 11, n. 11, p. 1373, 2019.
- ADAMS, P. B.; AYERS, W. A. Ecology of *Sclerotinia* species. **Phytopathology**, v. 69, n. 8, p.896-898, 1979.
- BAJWA, S. G.; RUPE, J. C.; MASON, J. Soybean disease monitoring with leaf reflectance. **Remote Sensing**, v. 9, n. 2, p. 127, 2017.
- BARDIN, S.D.; HUANG, H.C. Research on biology and control of *Sclerotinia* diseases in Canadá. **Canadian Journal of Plant Pathology**, v. 23, p. 88-98, 2001.

BOECHAT, L. T.; PINTO, F. A. C.; PAULA JÚNIOR, T. J.; QUEIROZ, D. M.; TEIXEIRA, H. Detecção do mofo-branco no feijoeiro, utilizando características espectrais. **Revista Ceres**, v. 61, n. 6, p. 907-915, 2014.

BOLAND, G. J.; HALL, R. Index of plant hosts of *Sclerotinia sclerotiorum*. **Canadian Journal of Plant Pathology**, v. 16, n.1, p. 93–108, 1994.

CANTERI, M. G.; DALLA PRIA, M.; SILVA, O. C. **Principais doenças fúngicas do feijoeiro: orientações para manejo econômico e ecológico**. Ponta Grossa: UEPG, 1999. 178 p.

CHEN, T.; ZHANG, J.; CHEN, Y.; WAN, S.; ZHANG, L. Detection of peanut leaf spots disease using canopy hyperspectral reflectance. **Computers and electronics in agriculture**, v. 156, p. 677-683, 2019.

CLARKSON, J. P.; STAVELEY, J.; PHELPS, K.; YOUNG, C. S.; WHIPPS, J. M. Ascospore release and survival in *Sclerotinia sclerotiorum*. **Mycological research**, v. 107, n. 2, p. 213-222, 2003.

CONAB - Companhia Nacional de Abastecimento. A cultura do feijão. Brasília: Conab, 2018. 244 p. Disponível também em: <<http://www.conab.gov.br>>.

CONAB - Companhia Nacional de Abastecimento. Boletim da Safra de Grãos – 3º Levantamento, Safra 2023/2024 - Dezembro, 2023. Disponível também em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos/item/download/50528_3c2beae44d782e81ef5d4e976d06955b>.

DENG, L.; MAO, Z.; LI, X.; HU, Z.; DUAN, F.; YAN, Y. UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 146, p. 124-136, 2018.

DENG, X.; ZHU, Z.; YANG, J.; ZHENG, Z.; HUANG, Z.; YIN, X.; WEI, S.; LAN, Y. Detection of Citrus Huanglongbing Based on Multi-Input Neural Network Model of UAV Hyperspectral Remote Sensing. **Remote Sensing**, v. 12, n. 17, p. 2678, 2020.

FRANKE, J.; MENZ, G. Multi-temporal wheat disease detection by multi-spectral remote sensing. **Precision Agriculture**, v. 8, n. 3, p. 161-172, 2007.

HE, L.; QI, S. L.; DUAN, J. Z.; GUO, T. C.; FENG, W.; HE, D. X. Monitoring of Wheat Powdery Mildew Disease Severity Using Multiangle Hyperspectral Remote Sensing. **IEEE Transactions on Geoscience and Remote Sensing**, 2020.

HUGHES, E. W. **Spatially-explicit Snap Bean Flowering and Disease Prediction Using Imaging Spectroscopy from Unmanned Aerial Systems**. Rochester Institute of Technology, 2019.

LEITE, R. M. V. B. C. Ocorrência de doenças causadas por *Sclerotinia sclerotiorum* em girassol e soja. **Embrapa Soja. Comunicado Técnico**, 2005.

LETHAM, D. B. Simulation of light of apothecial initial development of *Sclerotinia sclerotiorum* in soil. **Soil Biol. Biochem**, v. 8, p. 385-389, 1975.

MACHADO, M. L.; PINTO, F. A. C.; QUEIROZ, D. M.; PAULA JÚNIOR, T. J.; VIEIRA, R. F. Estimativa de severidade do mofo-branco em lavouras de feijão utilizando-se sensores hiper e multiespectral. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 5, p. 426-432, 2015a.

MACHADO, M. L.; PINTO, F. A. C.; PAULA JÚNIOR, T. J.; QUEIROZ, D. M.; CERQUEIRA, O. D. A. White mold detection in common beans through leaf reflectance spectroscopy. **Engenharia Agrícola**, v. 35, n. 6, p. 1117-1126, 2015b.

MATESE, A.; TOSCANO, P.; DI GENNARO, S. F.; GENESIO, L.; VACCARI, F. P.; PRIMICERIO, J.; BELLI, C.; ZALDEI, A.; BIANCONI, R.; GIOLI, B. Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. **Remote Sensing**, v. 7, n. 3, p. 2971-2990, 2015.

MAYNARD, D. C.; ANJOS, H. A.; MAGALHÃES, A. C. V.; GRIMES, L. N.; COSTA, M. G. O.; SANTOS, R. B. Food consumption and anxiety among the adult population during the COVID-19 pandemic in Brazil. **Research, Society and Development**, [S. l.], v. 9, n. 11, p. e4279119905, 2020. DOI: 10.33448/rsd-v9i11.9905. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/9905>. Acesso em: 6 jan. 2021.

MEKONNEN, Y.; NAMUDURI, S.; BURTON, L.; SARWAT, A.; BHANSALI, S. Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture. **Journal of the Electrochemical Society**, v. 167, n. 3, p. 037522, 2019.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. São José dos Campos: 2001. Instituto Nacional de Pesquisas Espaciais (INPE).

MWEBAZE, E.; OWOMUGISHA, G. Machine learning for plant disease incidence and severity measurements from leaf images. In: **2016 15th IEEE international conference on machine learning and applications (ICMLA)**. IEEE, 2016. p. 158-163.

OMRAN, E. S. E. Early sensing of peanut leaf spot using spectroscopy and thermal imaging. **Archives of Agronomy and Soil Science**, v. 63, n. 7, p. 883-896, 2017.

PAULA JÚNIOR, T. J.; MORANDI, M. A. B.; LOBO JUNIOR, M.; VIEIRA, R. F.; ZAMBOLIM, L. Controle alternativo do mofo-branco do feijoeiro. **Embrapa Meio Ambiente-Capítulo em livro científico (ALICE)**, 2006.

PRINCE, G.; CLARKSON, J. P.; RAJPOOT, N. M. Automatic detection of diseased tomato plants using thermal and stereo visible light images. **PloS One**, v. 10, n. 4, p. e0123262, 2015.

QIN, Z.; ZHANG, M. Detection of rice sheath blight for in-season disease management using multispectral remote sensing. **International Journal of Applied Earth Observation and Geoinformation**, v. 7, n. 2, p. 115-128, 2005.

RAMOS, P. J.; PRIETO, F. A.; MONTOYA, E. C.; OLIVEROS, C. E. Automatic fruit count on coffee branches using computer vision. **Computers and Electronics in Agriculture**, v. 137, p. 9-22, 2017.

RODRÍGUEZ, S.; GUALOTUÑA, T.; GRILO, C. A system for the monitoring and predicting of data in precision agriculture in a rose greenhouse based on wireless sensor networks. **Procedia Computer Science**, v. 121, p. 306-313, 2017.

SANTOSO, H.; TANI, H.; WANG, X. Random Forest classification model of basal stem rot disease caused by *Ganoderma boninense* in oil palm plantations. **International Journal of Remote Sensing**, v. 38, n. 16, p. 4683-4699, 2017.

SUN, P.; YANG, X. B. 2000. Light, temperature, and moisture effects on apothecium production of *Sclerotinia sclerotiorum*. **Plant Disease**, v. 84, n. 12, p. 1287-1293, 2000.

THOMASSON, J. A.; WANG, T.; WANG, X.; COLLETT, R.; YANG, C.; NICHOLS, R. L. Disease detection and mitigation in a cotton crop with UAV remote sensing. In: **Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping III**. International Society for Optics and Photonics, 2018. p. 106640L.

WANG, T.; THOMASSON, J. A.; ISAKEIT, T.; YANG, C.; NICHOLS, R. L. A Plant-by-Plant Method to Identify and Treat Cotton Root Rot Based on UAV Remote Sensing. **Remote Sensing**, v. 12, n. 15, p. 2453, 2020.

WANG, T.; THOMASSON, J. A.; YANG, C.; ISAKEIT, T.; NICHOLS, R. L. Automatic Classification of Cotton Root Rot Disease Based on UAV Remote Sensing. **Remote Sensing**, v. 12, n. 8, p. 1310, 2020.

WANG, X.; ZHANG, X.; ZHOU, G. Automatic detection of rice disease using near infrared spectra technologies. **Journal of the Indian Society of Remote Sensing**, v. 45, n. 5, p. 785-794, 2017.

WILLETTS, Haydn J.; WONG, J. A. L. The biology of *Sclerotinia sclerotiorum*, *S. trifoliorum*, and *S. minor* with emphasis on specific nomenclature. **The Botanical Review**, v. 46, p. 101-165, 1980.

YAMAMOTO, K.; TOGAMI, T.; YAMAGUCHI, N. Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. **Sensors**, v. 17, n. 11, p. 2557, 2017.

CAPÍTULO 2 - DETECÇÃO DE MOFO-BRANCO (*SCLEROTINIA SCLEROTIORUM*) NO FEIJOEIRO COM BASE EM IMAGENS MULTIESPECTRAIS E TÉCNICAS DE APRENDIZADO DE MÁQUINA

Resumo

Para otimizar os sistemas de produção de feijão é essencial aprimorar a eficiência dos processos que compõem a cadeia produtiva. É necessário procurar meios mais sustentáveis de atender, tanto a demanda nutricional da cultura, quanto o controle de doenças como o mofo-branco. O objetivo deste trabalho foi aplicar técnicas de *machine learning* em imagens multiespectrais capturadas por Aeronaves Remotamente Pilotadas (RPAs) e identificar os melhores índices de vegetação para estimar a incidência (IMB) e severidade (ISMB) do mofo-branco, além da produtividade na presença da doença. Foram realizados três ensaios em dois campos experimentais. Nesses ensaios, as imagens foram capturadas a 40 metros de altura acima da cultura com resolução espacial de 0,03 m. A linhagem usada nos ensaios foi a CNFC 10720, com 7, 10, 13 ou 16 plantas por metro, e os espaçamentos entre linhas de 0,25 e 0,50 m. Foram calculados 51 índices de vegetação no dossel, para os quais foram usadas as reflectâncias nas bandas do azul, verde, vermelho, borda do vermelho e infravermelho. As covariáveis selecionadas pelos modelos foram diferentes para a predição do IMB, ISMB e produtividade. Entre as covariáveis selecionadas para os modelos, tanto índices espectrais quanto bandas isoladas mostraram potencial para prever IMB, ISMB e produtividade. O Cubist apresentou melhor predição do IMB e ISMB, em que se destacaram as reflectâncias nas bandas do verde e azul, e os índices Razão Simples Modificada *NIR/RED* (MSR) e Índice de Vegetação Melhorado (EVI) na modelagem do IMB, enquanto para a predição do ISMB destacaram-se as bandas verde e azul, e o índice MSR. Já para a produtividade o melhor algoritmo foi o KNN, com as bandas do verde e azul, e o Índice de Coloração, o Índice de Clorofila da Vegetação e Índice da Diferença Normalizada da Vegetação Utilizando a Banda do Verde.

2.1. INTRODUÇÃO

A cultura do feijão (*Phaseolus vulgaris*) é importante na alimentação humana, sendo cultivada e consumida em todo o mundo. No entanto, existem diversas ameaças que comprometem a produção de feijão, dentre elas estão as doenças fúngicas, como a do mofo-branco (MB). Essa doença ocorre tanto nas regiões tropicais quanto nas temperadas (Lehner et al., 2017). O MB é causado pelo fungo *Sclerotinia sclerotiorum* (Lib.) de Bary, que possui ampla gama de hospedeiros, perfazendo 408 espécies dentre 278 gêneros e 75 famílias de plantas (Boland; Hall, 1994). O MB infecta flores, folhas, hastes e vagens (Leite, 2005; Paula Júnior et al., 2006), podendo levar a perdas de 90% da produção (Singh; Schwartz, 2010). No Brasil, as perdas por essa doença são estimadas em 40% a 70% da produtividade (Nasser et al., 1995). Portanto, o mofo-branco é um problema relevante na produção de feijão em todo o mundo (Hind; Ash; Murray, 2003).

Na fase final do desenvolvimento, o MB se mantém na área por meio de estrutura de resistência, o escleródio. Uma vez que o MB infecta uma lavoura, essa área pode apresentar escleródios viáveis por até 11 anos (Görgen et al., 2009, 2010). Sabe-se que na maioria das vezes, o cultivo de feijão se dá em áreas já infestadas com escleródios, por isso, a identificação antecipada é importante para o manejo da lavoura, como por exemplo, saber onde tem a doença e evitar tráfegar inicialmente nessas áreas com inóculo da doença e não levar inóculo para áreas não infectadas. A identificação precoce de doenças, como o mofo-branco, é importante, visto a possibilidade de manejo antecipado e redução da proliferação da doença no campo (Zhang, 2021).

No campo, a identificação do MB é realizada de maneira visual. Entretanto, além dessa identificação ser uma etapa demorada, ela exige técnicos especializados, o que pode elevar os custos de produção. Métodos com base nas características espectrais da cultura têm sido usados em alguns trabalhos para estimar a severidade do MB no feijoeiro utilizando espectrorradiômetro (Boechat et al., 2014; Machado et al., 2015a; Machado et al., 2015b). Os resultados desses autores permitem identificar possíveis locais de ocorrência do MB com base na assinatura espectrais das plantas. No entanto, o espectrorradiômetro possui baixa capacidade operacional e elevado custo, além da

necessidade de mão de obra especializada para uso do equipamento e interpretação dos dados coletados.

Para aumentar a capacidade operacional pode-se utilizar técnicas de sensoriamento remoto (SR). O SR pode ser aplicado em nível orbital, por meio de satélite, e em nível sub-orbital, por meio de RPAs (*Remotely Piloted Aircraft*), popularmente conhecidos como drones ou Vants (Veículos Aéreos Não Tripulados). Câmeras acopladas aos RPAs são utilizadas na agricultura para mapear a infestação de plantas daninhas, monitorar a saúde, o desenvolvimento e produtividade das lavouras, e auxiliar no manejo da irrigação e aplicar pesticidas (Tsouros; Bibi; Sarigiannidis, 2019). No entanto, o uso de imagens para fins de identificação de doenças exige cuidados que vão desde a aquisição de imagens até o processamento e interpretação dos dados espectrais. Para contornar esses potenciais problemas de interpretação dos dados coletados, uma alternativa tem sido combinar as imagens/dados espectrais adquiridas pelos RPAs com técnicas de aprendizagem de máquina. A união dessas duas técnicas para o manejo de doença já foi usada nas culturas dos citros (Abdulridha; Batuman; Ampatzidis, 2019), soja (Negrisoli et al., 2022), milho (Zhang et al., 2023), arroz (Pallathadka et al., 2023), entre outros.

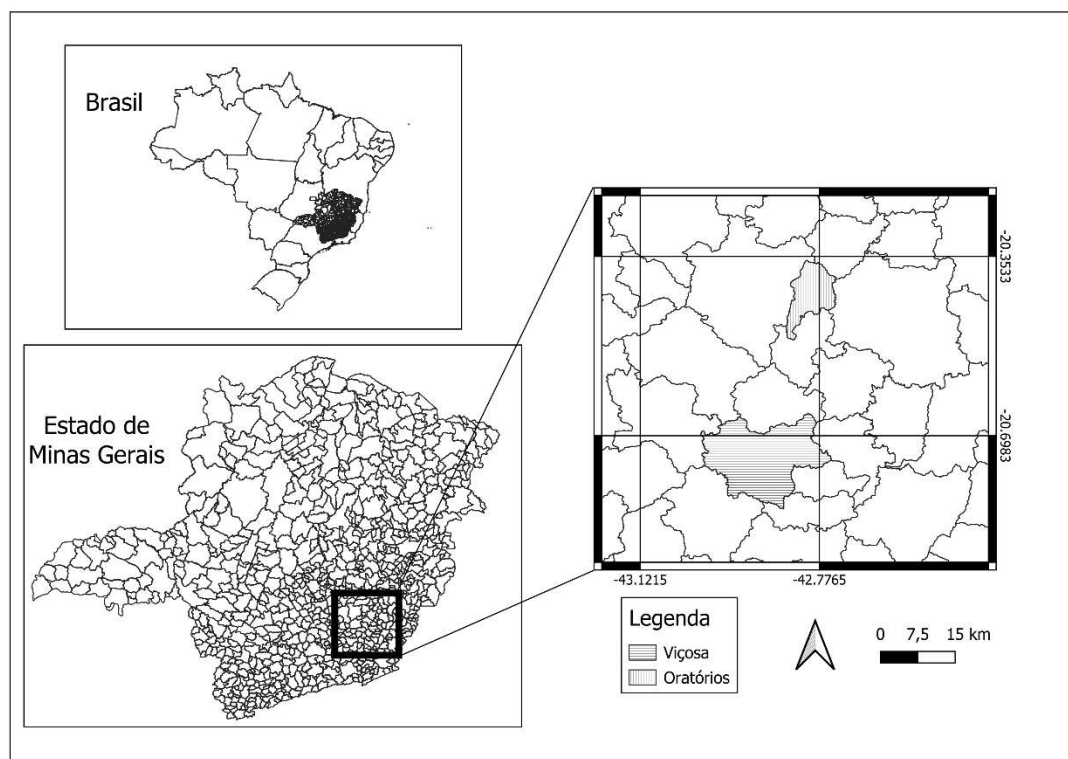
Mesmo que sejam encontradas publicações que envolvam a aplicação de SR na identificação do MB, essa tecnologia não está plenamente estabelecida. Ainda são necessários a realização de estudos que envolvam uma análise espacial mais detalhada em nível de parcela, em que se conheça a resposta de toda a população e não apenas por amostragem desta, utilizando RPA. Mesmo que os trabalhos já publicados tenham mostrado bom desempenho na estimativa da severidade do MB com informações obtidas por dados hiperespectrais, é necessário que se tenha novas propostas e aplicações usando equipamentos comerciais multiespectrais acoplados a RPA mais acessíveis e técnicas que possam ser usadas mais facilmente na lavoura. Logo, o objetivo deste estudo foi desenvolver um modelo para detectar a incidência e severidade do MB, e da produtividade do feijoeiro em presença do MB, utilizando imagens multiespectrais obtidas a partir de RPA e técnicas de aprendizado de máquina.

2.2. MATERIAL E MÉTODOS

Área experimental

Este estudo foi realizado nas cidades de Viçosa e Oratórios, no estado de Minas Gerais, Brasil (Figura 2.1), totalizando três experimentos. Dois experimentos, um em 2021 e outro em 2022, foram conduzidos no campo experimental Diogo Alves de Melo da Universidade Federal de Viçosa (UFV), no município de Viçosa (Latitude: 20°46'7.42"S, Longitude: 42°52'11.98"W, Altitude: 660 m). O terceiro experimento foi conduzido em 2022 na Fazenda Experimental do Vale do Piranga, da Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), no município de Oratórios (Latitude: 20°24'23.10"S, Longitude: 42°49'7.05"W, Altitude: 476 m). O clima da região de Viçosa e Oratórios se enquadra na classificação Cwa de Köppen-Geiger (Alvares et al., 2013), com verões quentes e inverno seco.

Figura 2.1 – Localização dos locais dos dois experimentos nas cidades de Viçosa e Oratórios, no estado de Minas Gerais.



Fonte: elaborado pelo próprio autor.

As áreas experimentais foram preparadas com arado, seguido de duas grades niveladoras. Foi usado na adubação de plantio 350 kg ha⁻¹ da formulação NPK 08-28-16. Em cobertura, usaram-se 150 kg ha⁻¹ de ureia no estágio de desenvolvimento V4, quando as plantas apresentavam três folhas trifolioladas. Foram realizadas aplicações para tratamentos de prevenção de doenças fúngicas (exceto para o mofo-branco) e pragas. Foi usado o fungicida com princípio ativo à base de trifloxistrobina (10% m/v) e tebuconazol (20% m/v), e inseticida com princípio ativo à base de imidacloprido (70% m/m). Para o manejo das plantas daninhas foram usados herbicidas à base de fluazifop-p-butílico (25% m/v) e fomesafem (25% m/v). Como o cultivo foi realizado no outono-inverno, a área foi irrigada com lâmina de água de 10 mm a cada três dias até a emergência das plântulas e 40 mm semanalmente após a emergência.

Foi utilizada a linhagem CNFC 10720 para os três experimentos. As plantas dessa cultivar apresenta resistência parcial ao MB, hábito de crescimento tipo II, com porte ereto e caule pouco ramificado. Cada experimento foi conduzido em 32 parcelas, totalizando 96 parcelas. Entretanto, três parcelas foram perdidas, reduzindo o total de parcelas para 93. As parcelas experimentais foram instaladas seguindo o delineamento em blocos casualizados (DBC), em fatorial 4 (densidades) x 2 (espaçamento entre linhas) e quatro repetições. As densidades de plantio usadas foram 7, 10, 13 ou 16 plantas por metro, e os espaçamentos entre linhas de 0,25 ou 0,50 m. As linhas externas das parcelas não foram consideradas nas avaliações.

Determinação da severidade do mofo-branco

O índice de severidade do mofo-branco foi avaliado por meio da escala proposta por Hall e Phillips (1996) com notas de 0 a 4, em que a nota 0 é a ausência de sintomas, nota 1 corresponde a 25% da planta com sintomas de mofo-branco, nota 2 de 26 a 50% de plantas com sintomas, nota 3 de 51 a 75% da planta com sintomas, e nota 4 de 76 a 100% da planta com sintomas. Após a atribuição das notas de severidade no final do ciclo foi calculado o índice de severidade do mofo-branco, conforme a Equação 2.1 (KOLKMAN & KELLY, 2002). A incidência de mofo-branco (IMB) foi a porcentagem das plantas da

parcela com a presença do MB (Equação 2.2). A avaliação do MB aconteceu no estágio R9 do feijoeiro-comum (fase de maturação do desenvolvimento da planta).

$$\text{ISMB}(\%) = 100 \frac{\sum(\text{notas de todas as plantas})}{4 \times (\text{número total de plantas})} \quad (2.1)$$

$$\text{IMB}(\%) = 100 \frac{(\text{número de plantas da parcela com mofo-branco})}{(\text{número total de plantas da parcela})} \quad (2.2)$$

Produtividade de grãos

Para determinar a produtividade, colheram-se os grãos de cada parcela excluindo uma linha de cada borda e 0,50 m em cada extremidade da linha. Foi utilizado o método da estufa, segundo as Regras para Análise de Sementes (BRASIL, 1992), para a determinação do teor de água dos grãos. O teor de água dos grãos em base úmida (% b.u.) foi utilizado para converter os valores de produtividade para o padrão de umidade de 13% b.u. Por fim, a produtividade foi convertida para kg ha⁻¹.

Aquisição e processamento das imagens

As imagens foram adquiridas por meio de uma câmera multiespectral Micasense, modelo RedEdge MX (MicaSense, Seattle, WA, USA). Esta câmera possui cinco sensores CMOS (Complementary metal-oxide semiconductor), cujas informações, segundo o fabricante, estão apresentadas na Tabela 2.1. A câmera foi acoplada a uma aeronave remotamente pilotada da marca DJI, modelo Matrice 100 (DJI Innovations, Shenzhen, China).

Tabela 2.1. Informações da câmera multiespectral MicaSense RedEdge MX, segundo o fabricante (Micasense - *RedEdge-MX Integration Guide*, 2017).

Banda	Comprimentos de onda (nm)	Largura da banda (nm)
Azul	475	20
Verde	560	20
Vermelho	668	10
Borda do vermelho	717	10
NIR ¹	840	40

Legenda: ¹infravermelho próximo.

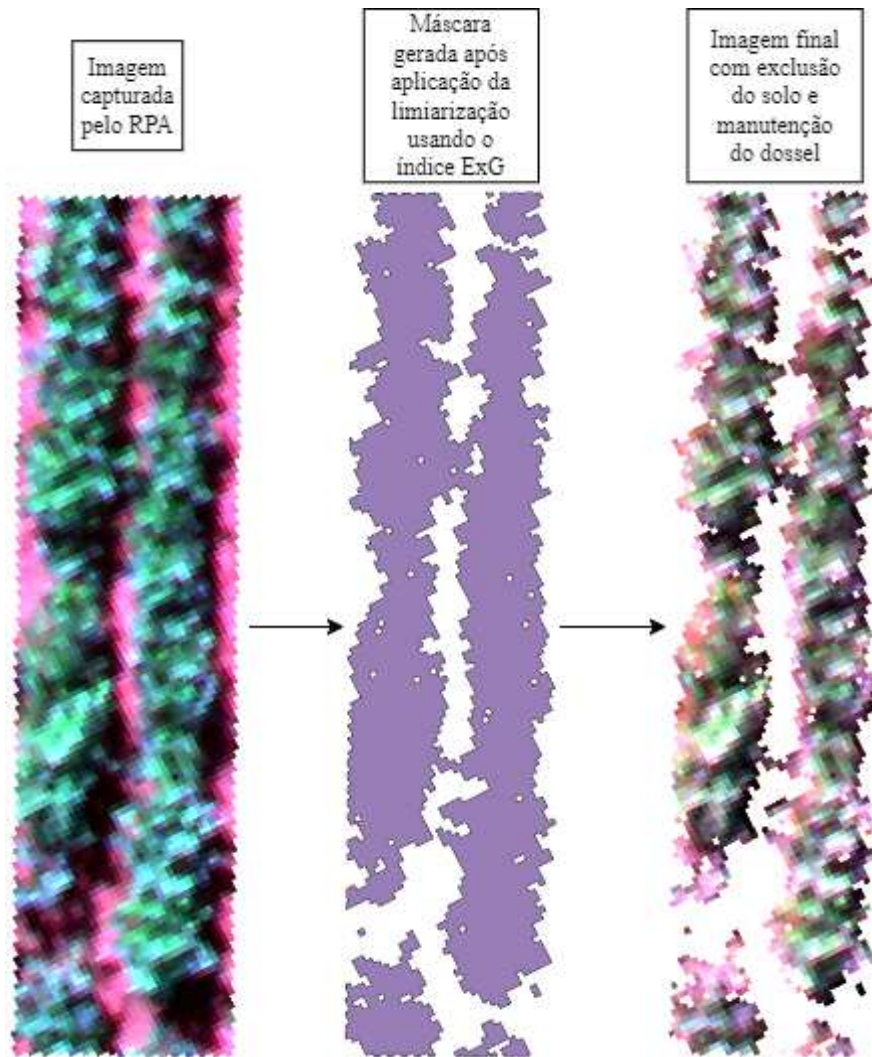
Os voos do RPA foram conduzidos semanalmente, das 11h às 13h do horário local, aos 38, 45, 52 e 59 dias após a emergência (DAE) das plantas. Um plano de voo predefinido, preparado pelo software DroneDeploy (DroneDeploy Inc., San Francisco, CA, EUA), foi utilizado para conduzir os voos. Os voos foram realizados a 40 metros de altura, resultando em um *Ground sampling distance* (GSD) de 0,03 m. A aquisição das imagens se deu com imagens sobrepostas frontalmente em 75% e lateralmente em 75%. Para a calibração radiométrica em pós-processamento, as imagens do alvo de calibração de refletância foram capturadas a uma altura de 1 m antes e após cada voo.

Após os voos, as imagens foram processadas para obtenção do ortomosaico no software Agisoft™ MetaShape software, version 1.5.3 (Agisoft LLC, St. Petersburg, Russia). O alinhamento das imagens e georreferenciamento foram feitos no software Agisoft™ MetaShape. Para o georreferenciamento preciso das imagens foram usados cinco pontos de controle (PCs) no solo. O equipamento usado para georreferenciar os PCs foi o receptor topográfico GNSS (*Global Navigation Satellite System*), modelo Trimble ProXT (Trimble Inc., Sunnyvale, CA, EUA). A correção diferencial dos pontos georreferenciados foi feita usando a base de dados da Rede de Monitoramento Contínuo (RBMC), do Instituto Brasileiro de Geografia e Estatística (IBGE), localizada no município de Viçosa-MG.

Extração de dados espectrais

Após gerar os ortomosaicos, foi usado o software QGIS, versão 3.22 (QGIS Development Team, 2016) para extrair os dados espectrais. Antes de extrair os dados espectrais das parcelas foi realizado o processamento das imagens para a exclusão do fundo da imagem (solo) por meio de limiarização, conforme apresentado na Figura 2.2. Dessa forma a análise ocorreu apenas para o dossel das plantas. Além das cinco bandas da câmera multiespectral, foram calculados 51 índices de vegetação (Tabela 2.3). Os índices foram selecionados com base em pesquisas científicas com o objetivo de detectar doenças em plantas. O cálculo dos índices foi realizado na área útil das parcelas, excluindo uma linha de cada borda, e excluindo os 0,5 m de cada extremidade das parcelas.

Figura 2.2 – Etapa de exclusão do solo nas imagens capturadas por meio da câmera acoplada à RPA aplicada às parcelas para posterior cálculo dos índices de vegetação apenas no dossel.



Fonte: elaborado pelo próprio autor.

Tabela 2.3 – Índices de vegetação (IVs) calculados para uso no modelo de predição de infestação por mofo-branco.

IVs	Equação	Referência
NDVI ¹	$\frac{NIR - R}{NIR + R}$	ROUSE et al. (1973)
NDRE ²	$\frac{NIR - RE}{NIR + RE}$	FITZGERALD et al. (2006)
GNDVI ³	$\frac{NIR - G}{NIR + G}$	GITELSON et al. (1996)
DVI ⁴	$IR - R$	TUCKER (1979)
MSAVI ⁵	$0,5 \left[2(IR + 1) - \sqrt{[(2IR + 1)^2 - 8(IR - R)]} \right]$	QI et al. (1996)
GDVI ⁶	$\frac{NIR - G}{NIR + G}$	TUCKER et al. (1979)
CCCI ⁷	$\frac{NDRE}{NDVI}$	EL-SHIKHA et al. (2008)
GRVI ⁸	$\frac{NIR}{G}$	SRIPADA et al. (2005)
RVI ⁹	$\frac{NIR}{R}$	PEARSON e MILLER (1972)
RDVI ¹⁰	$\frac{NIR - R}{\sqrt{(NIR + R)}}$	ZARCO-TEJADA et al. (2013)
GLI ¹¹	$\frac{2G - B - R}{2G + B + R}$	LOUHAICHI et al. (2001)
VARI-Green ¹²	$\frac{G - R}{G + R - B}$	GITELSON et al. (2002)
VARI-Red Edge ¹³	$\frac{RE - 1,7 \times R - 0,7 \times B}{RE + 2,3 \times R - 1,3 \times B}$	GITELSON et al. (2002)
ExG ¹⁴	$\frac{2G - R - B}{R + G + B}$	WOEBBECKE et al. (1995)
ExR ¹⁵	$\frac{1,4 R - G}{R + G + B}$	MEYER et al. (1998)
ExB ¹⁶	$\frac{1,4 B - G}{R + G + B}$	MAO e WANG (2003)

ExGR ¹⁷	$\frac{ExG - ExR}{R^2}$	MAO et al. (2003)
RI ¹⁸	$\frac{R^2}{B \times G \times 3}$	MADEIRA et al. (1997)
CI-RED ¹⁹	$\frac{NIR}{R} - 1$	GITELSON et al. (2005)
CVI ²⁰	$NIR \frac{R}{G^2}$	DATT et al. (2003)
EVI ²¹	$2,5 \frac{NIR - R}{NIR + 6 \times R - 7,5 \times B + 1}$	HUETE et al. (1997)
EVI-2 ²²	$2,5 \frac{NIR - R}{NIR + 2,4 \times R + 1}$	JIANG et al. (2008)
MEXG ²³	$1,262G - 0,884R - 0,311B$	BURGOS-ARTIZZU et al. (2011)
MNGDR ²⁴	$\frac{G^2 - R^2}{G^2 + R^2}$	BENDIG et al. (2015)
NGDR ²⁵	$\frac{G - R}{G + R}$	HAMUDA et al. (2016)
OSAVI ²⁶	$\frac{NIR - R}{NIR + R + 0,16}$	RONDEAUX et al. (1996)
RECI ²⁷	$\frac{NIR}{RE} - 1$	GITELSON et al. (2003)
SAVI ²⁸	$1,5 \frac{NIR - R}{NIR + R + 0,5}$	HUETE (1988)
TGI ²⁹	$G - 0,39 \times R - 0,61 \times B$	HUNT JR et al. (2011)
HI ³⁰	$\frac{2R - G - B}{G - B}$	ESCADAFAL (1994)
SI ³¹	$\frac{R - B}{R + B}$	ESCADAFAL (1994)
VVI ³²	$\left(1 - \frac{ R - 30 }{ R + 30 }\right) \left(1 - \frac{ G - 50 }{ G + 50 }\right) \left(1 - \frac{ B - 1 }{ B + 1 }\right)$	PONTI (2012)
GLAI ³³	$25 \frac{G - R}{G + R - B} + 1,25$	REUDENBACH et al. (2021)

CI ³⁴	$\frac{R - B}{R}$	BIRTH e MCVEY (1968)
HUE ³⁵	$atan \left[\left(2 \frac{R - G - B}{30,5} (G - B) \right) \right]$	HENRICH et al. (2011)
SHP ³⁶	$2 \frac{R - G - B}{G - B}$	BOOCHS et al. (1990)
CIVE ³⁷	$0,441R - 0,811G - 0,385B + 18,787$	KATAOKA et al. (2003)
NDWI ³⁸	$\frac{G - NIR}{G + NIR}$	MCFEETERS (1996)
Chl ³⁹	$723,04 \left[\left(\frac{NIR}{RE} \right) - 1 \right] - 10,76$	VIÑA et al. (2008)
IAF ⁴⁰	$exp^{1,426 + \left(\frac{-0,542}{NDVI} \right)}$	GALVÍNCIO et al. (2013)
RGBVI ⁴¹	$\frac{G^2 - R \times B}{G^2 + B \times R}$	BENDIG et al. (2015)
MCARI ⁴²	$\left[(RE - R) - 0,2(RE - G) \right] \frac{RE}{R}$	DAUGHTRY et al. (2000)
TCARI ⁴³	$3 \times \left[(RE - R) - 0,2(RE - G) \left(\frac{RE}{R} \right) \right]$	HABOUDANE et al. (2002)
WDRVI01 ⁴⁴	$\frac{0,1 \times NIR - R}{0,1 \times NIR + R}$	GITELSON (2004)
WDRVI02 ⁴⁵	$\frac{0,2 \times NIR - R}{0,2 \times NIR + R}$	GITELSON (2004)
MSR ⁴⁶	$\left(\frac{NIR}{R} - 1 \right) / \left[\left(\sqrt{\frac{NIR}{R}} \right) - 1 \right]$	CHEN (1996)
MTCI ⁴⁷	$\frac{NIR - RE}{RE - R}$	DASH e CURRAN (2004)
CI_GREEN ⁴⁸	$\frac{NIR}{G} - 1$	GITELSON et al. (2003)
WI ⁴⁹	$\frac{G - B}{R - G}$	WOEBBECKE et al. (1995)

VEG⁵⁰

$$G / (R^{0,667} \times B^{1-0,667})$$

MARCHANT e ONYANGO (2002)

COM⁵¹

$$0,25 \times ExG + 0,3 \times ExGR + 0,33 \times CIVE + 0,12 \times VEG$$

GUIJARRO et al. (2011)

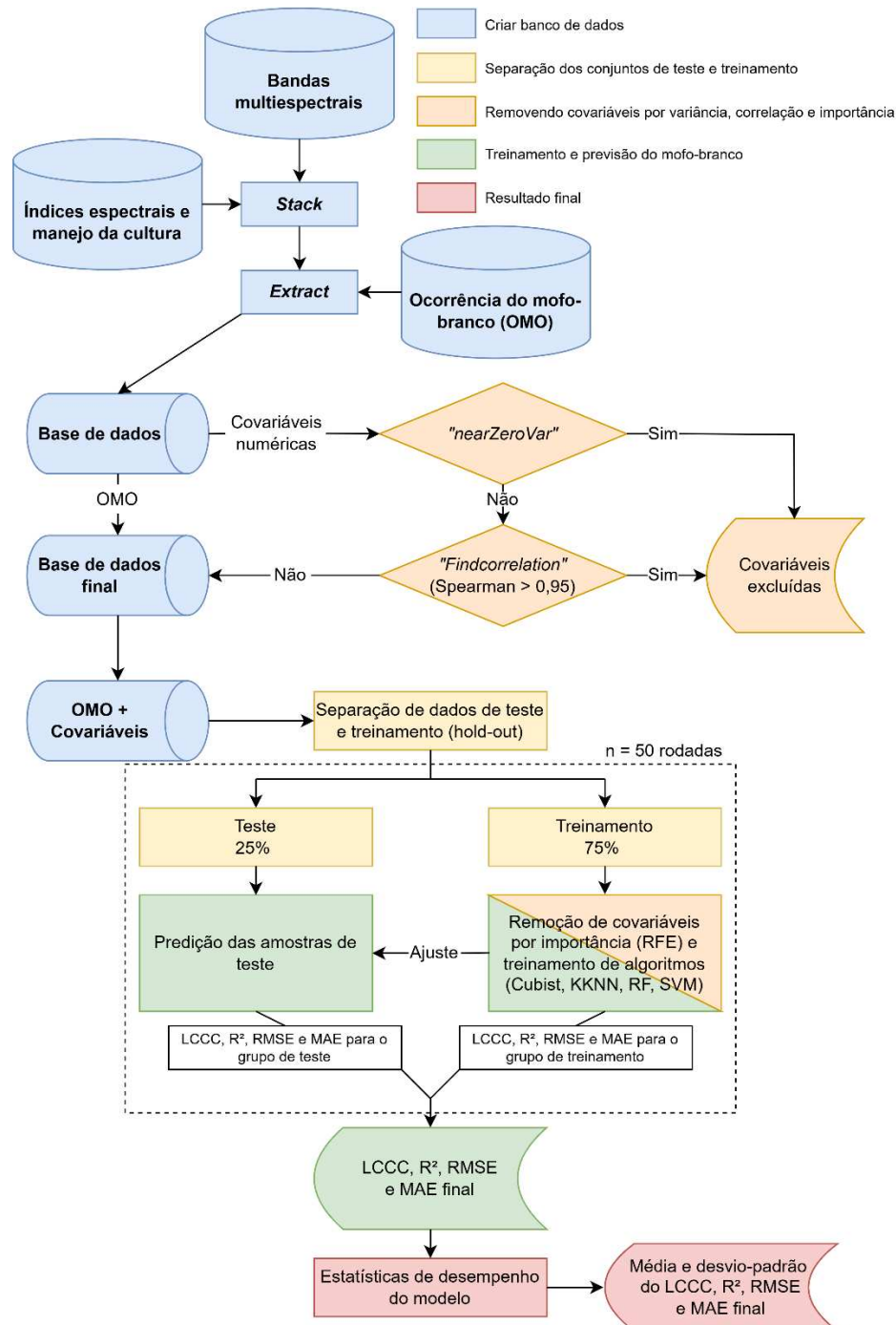
¹Normalized Difference Vegetation Index; ²Normalized difference red edge index; ³Green normalized difference vegetation index; ⁴Difference Vegetation Index; ⁵Modified Soil Adjusted Vegetation Index; ⁶Difference NIR/Green Green Difference Vegetation Index; ⁷Canopy Chlorophyll Content Index; ⁸Green Ratio Vegetation Index; ⁹Ratio Vegetation Index; ¹⁰Renormalized Difference Vegetation Index; ¹¹Green Leaf Index; ¹²Visible Atmospherically Resistant Index; ¹³Visible Atmospherically Resistant Index; ¹⁴Excess green vegetation index; ¹⁵Excess red vegetation index; ¹⁶Excess blue vegetation index; ¹⁷Excess green minus Excess red index; ¹⁸Redness index; ¹⁹Chlorophyll index red; ²⁰Chlorophyll vegetation index; ²¹Enhanced Vegetation Index; ²²Enhanced Vegetation Index 2 -2; ²³Modified excess green; ²⁴Modified normalized green red difference; ²⁵Normalized green red difference; ²⁶Optimized soil-adjusted vegetation index; ²⁷Red edge chlorophyll index; ²⁸Soil adjusted vegetation index; ²⁹Triangular greenness index; ³⁰Primary Colors Hue Index; ³¹Spectra Slope Saturation Index; ³²Visible Vegetation Index; ³³Greenish Leaf Area Index; ³⁴Coloration Index; ³⁵Hue Index; ³⁶Shape Index; ³⁷Color index of vegetation; ³⁸Normalized Difference Water Index; ³⁹Total chlorophyll content; ⁴⁰Leaf area index; ⁴¹Red green blue vegetation index; ⁴²Modified ¹Chlorophyll Absorption in Reflectance Index; ⁴³Transformed Chlorophyll absorption reflectance index; ⁴⁴Wide dynamic range vegetation index; ⁴⁵Wide dynamic range vegetation index; ⁴⁶Modified Simple Ratio NIR/RED; ⁴⁷Meris terrestrial chlorophyll index; ⁴⁸Green Chlorophyll Index; ⁴⁹Woebbecke index; ⁵⁰Vegetative index; ⁵¹Combination.

Fonte: elaborado pelo próprio autor.

Processo de modelagem para a detecção do mofo-branco

O modelo de risco de ocorrência do mofo-branco desenvolvido foi estimado com dados espectrais na área de experimento durante o ciclo do feijão-comum. As visitas à área experimental foram feitas para a coleta das informações espectrais utilizando a câmera multiespectral acoplada ao RPA. As covariáveis usadas no desenvolvimento do modelo de predição de risco do mofo-branco constam na Tabela 2.3, são os índices de vegetação calculados por meio das informações espectrais obtidas da câmera multiespectral nos voos. Sendo também utilizadas como covariáveis de entrada para o modelo de predição as refletâncias determinadas pelos cinco canais da câmera (bandas do azul, verde, vermelho, borda do vermelho e infravermelho próximo). O fluxograma do processo de modelagem está demonstrado na Figura 2.3.

Figura 2.3 – Fluxograma metodológico ilustrando a sequência de métodos aplicados na predição do mofo-branco. Os algoritmos usados foram o *Cubist*, *Random Forest* (RF), *Support Vector Machine* (SVMs) e *Kernel K Nearest Neighbors* (KNN) para modelar a previsão do mofo-branco.



Fonte: elaborado pelo próprio autor.

Para verificar potenciais interações e associações entre as variáveis de entrada, aumento da parcimônia do modelo final e redução do custo computacional no treinamento do modelo (De Castro Paes et al. 2023; Silva et al., 2023) foi conduzido um processo de seleção de covariáveis. Por esse processo, excluiu-se covariáveis com correlação superior a 95% pelo teste de *Spearman*. Para executar essa etapa foi utilizado o software R (R Core Team, 2023) por meio da aplicação dos pacotes “*stats*” (Hothorn, 2021) e “*caret*” (Kuhn et al., 2020) e suas respectivas funções, “*cor*” e “*findCorrelation*” (Kuhn; Johnson, 2013). As covariáveis que passaram por essa etapa de seleção foram consideradas como aptas a fazer parte do modelo de predição de infestação por mofo-branco.

Os dados de treinamento e teste foram separados usando o método *holdout test* – HOT (Burman, 1989). Nessa separação 75% dos dados foram destinados para treinamento e 25 % para teste. O HOT é indicado para situações em que o conjunto de dados é reduzido, conforme verificado por Eertink et al. (2022). Isso justifica o uso do HOT neste trabalho, uma vez que o conjunto de amostras foi pequeno, em número de 93 parcelas. Ainda, o HOT apresenta eficiência superior à tradicional, visto que reduz a variabilidade do estimador (Kim, 2009). Outra vantagem é que o HOT diminui a possibilidade de *overfitting* (Valier, 2020). O processo de HOT aconteceu 50 vezes no conjunto de dados, em que cada repetição teve uma divisão aleatória de dados, considerado um valor adequado visto o tamanho do conjunto amostral. O procedimento foi implementado da seguinte forma: o conjunto de dados foi submetido a 50 iterações independentes. Em cada iteração, os dados foram realocados aleatoriamente em uma proporção de 75% para treinamento e 25% para teste. O modelo foi então treinado e avaliado em cada uma dessas 50 divisões. O resultado final reportado consiste na média das métricas de desempenho obtidas nas 50 avaliações. O número de 50 repetições foi considerado adequado para garantir a estabilização da média e a robustez estatística, dado o tamanho amostral disponível (93 parcelas).

A remoção por importância foi realizada visando a otimização do número de covariáveis a qual o modelo apresentasse melhor desempenho. Este processo de remoção foi realizado utilizando a função *Recursive Feature Elimination* (RFE, Eliminação Recursiva de Feições). A função realiza uma seleção de covariáveis usando um

procedimento tipo *backward*, a qual é calculado primeiramente a importância do conjunto completo de covariáveis, sendo posteriormente obtido o desempenho para subconjuntos com número de covariáveis inferior ao conjunto total (*subsets*) definidas pela função RFE (Reis, et al., 2021; Dias, et al., 2021; Mell, et al., 2022; De Castro Paes, et al., 2022). O *conjunto* de covariáveis ideal foi o que apresentou melhor resultado do parâmetro de desempenho definido pela função. O melhor conjunto das covariáveis selecionadas foi avaliado pelo Coeficiente de Correlação de Concordância de Lin - LCCC (Lawrence; Lin, 1989) conforme a Equação 2.3. O LCCC é um parâmetro de desempenho mais preciso e exato quando comparado com o coeficiente de determinação (Khaledian; Miller, 2020).

$$LCCC = \frac{2r\sigma_O\sigma_F}{\sigma_O^2 + \sigma_F^2 + (\bar{O} - \bar{F})^2} \quad (2.3)$$

Em que, “*r*” é o coeficiente de determinação entre os valores observados e preditos, σ_O e σ_F são as variâncias dos valores observados e preditos, respectivamente, $\bar{O}$ e $\bar{F}$ as médias dos valores observados e preditos, respectivamente.

Para a modelagem, foram empregados quatro algoritmos de aprendizado de máquina: *Cubist*, *Kernel K-Nearest Neighbors* (KNN), *Random Forest* (RF) e *Support Vector Machine* com kernel radial (SVM). O treinamento de todos os algoritmos foi conduzido utilizando-se o subconjunto de covariáveis selecionadas pela função RFE do pacote “*caret*” no software R (Kuhn et al., 2020). A otimização dos hiperparâmetros de cada um dos modelos foi realizada mediante a técnica de Validação Cruzada Repetida (*Repeated Cross-Validation* - VCR). Este procedimento de VCR foi configurado com 10 folds e 5 repetições, totalizando 50 iterações de ajuste e validação para cada combinação de hiperparâmetros testada. Em resumo, o conjunto de treinamento é dividido aleatoriamente em 10 partes (folds) em 5 ocasiões distintas. Em cada uma das 5 repetições, o processo de validação cruzada de 10 folds é executado integralmente: 9 folds são usados para o treinamento e 1 fold para validação, rotacionando até que todos os folds tenham sido usados como conjunto de validação. O desempenho médio, calculado a partir de todas as 50 validações, é utilizado para avaliar a robustez da combinação de hiperparâmetros. Foram testados valores de hiperparâmetros em uma malha predefinida, variando o número de covariáveis utilizadas

nos subconjuntos de 5 até 20, além dos valores 25, 30 e o conjunto completo (42 covariáveis). O subconjunto de covariáveis que resultou no melhor desempenho médio na VCR para cada algoritmo foi, então, selecionado para ser utilizado em cada uma das rodadas do procedimento de HOT.

O algoritmo *Cubist* explica problemas não lineares por meio de uma série combinada de modelos lineares (Quinlan, 1992). Os hiperparâmetros ajustados do *Cubist* foram o número de comitês (“*committees*”) e o número de instâncias (“*neighbors*”). O *Kernel K Nearest Neighbors* – KNN (Yu; Ji; Zhang, 2002) é uma implementação mais avançada do algoritmo KNN, o qual apresenta um número superior de hiperparâmetros para a otimização. Os hiperparâmetros otimizados neste trabalho foram o número máximo de vizinhos (“*kmax*”), distância (“*distance*”) e o *Kernel*. O *Random Forest* (RF) é um algoritmo com base em árvores de decisão na resolução de problemas não lineares, podendo ser usado tanto para regressão quanto para classificação por meio de previsões das múltiplas árvores (Breiman, 2001). O hiperparâmetro ajustado no RF é o número de preditores selecionados aleatoriamente (“*mtry*”). E no algoritmo SVM (Cortes e Vapnik, 1995), usados na aplicação de problemas de regressão, as variáveis de entrada são transformadas para um hiperplano por meio de funções (*kernel*) de base radial (*Radial Basis Function* – RBF) para reproduzir um ambiente relacional não linear entre as variáveis. Para isso, os hiperparâmetros ajustados foram o *sigma* e o custo (“*cost*”). Sendo assim, a seleção de variáveis e os hiperparâmetros de cada algoritmo foram otimizados automaticamente pela função RFE do pacote “*caret*” (Kuhn et al., 2020).

O resultado obtido na seleção pela importância foi utilizado no treinamento final do algoritmo em cada rodada do HOT. O treinamento é feito para escolha dos valores hiperparâmetros ideais com o conjunto de covariáveis ideal. O treinamento foi otimizado usando a validação cruzada repetida (*k-fold*), com 10 *folds*, 5 repetições e 5 valores de hiperparâmetros internos de cada algoritmo. A importância das covariáveis foram calculadas pelo método padrão do pacote “*caret*”, o qual calcula o RMSE normalizado (NRMSE). Esse método substitui uma das covariáveis por valores aleatórios, e avalia o erro (RMSE) que isso causa com a amostra de validação, quanto maior o erro mais importante é a variável. Isso é feito com todas as covariáveis, e a importância é normalizada considerando o maior valor como 100% e o menor valor como zero. Os

resultados das 50 repetições do HOT foram agrupados estatisticamente usando a média dos parâmetros de performance. O melhor algoritmo foi escolhido a partir da análise estatística dos parâmetros LCCC utilizando o teste de Kruskal-Wallis com o teste de *post hoc* de Dunn, os dois com 5% de significância.

Os parâmetros de performance utilizados neste trabalho foram o LCCC, o coeficiente de determinação (R^2), Raiz do Erro Quadrático Médio (RMSE) e Erro Médio Absoluto (MAE), representados pelas Equações 2.4, 2.5 e 2.6, respectivamente, em que “ $\hat{y}$ ” é o valor predito, “ y ” é o valor real e n o número de amostras. O LCCC fornece uma medida global de concordância considerando a exatidão e precisão entre os valores observados e preditos. O R^2 é uma métrica que caracteriza a qualidade do ajuste do modelo quantificando o quanto da variável dependente é explicada pelas variáveis independentes. O RMSE quantifica a eficácia do modelo por meio da padronização dos resultados, enquanto o MAE mensura qualidade do modelo por meio da avaliação da diferença entre os valores preditos e reais.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.4)$$

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.5)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.6)$$

Foram calculados também, usando a função “*nullModel*” do pacote “*caret*” (Kuhn et al., 2020), os modelos nulos (“*baseline model*”) das métricas RMSE e MAE, apresentados nas Equações 2.7 e 2.8, respectivamente. Estes modelos nulos foram calculados para fins de comparação com os valores de RMSE e MAE das modelagens realizadas pelos algoritmos. O modelo nulo serve como uma linha de base para avaliar se há benefício na construção do modelo proposto pelos algoritmos. Sendo assim, o modelo nulo faz apenas uma previsão baseada apenas na média dos dados de teste. Dessa forma, se os resultados obtidos utilizando um determinado algoritmo não for

melhor que o modelo proposto pelo modelo nulo, ou seja, menores RMSE e MAE, indica que o modelo gerado pelo algoritmo não apresentou aprendizado significativo, e que compensasse prever os dados com sua utilização.

$$\text{RMSE_nulo} = \sqrt{\left(\frac{1}{n} \sum_{i=0}^n (\overline{Q_{train_i}} - Q_{obs_i})^2\right)} \quad (2.7)$$

$$\text{MAE_nulo} = \frac{1}{n} \sum |\overline{Q_{train_i}} - Q_{obs_i}| \quad (2.8)$$

Em que, $\overline{Q_{train_i}}$ é a média das amostras de treinamento, Q_{obs_i} é a amostra de validação e, n é o número de amostras (loop).

2.3. RESULTADOS

Nas Tabelas 2.4 e 2.5 são apresentadas a caracterização climática dos experimentos realizados e as informações relacionadas ao desenvolvimento da lavoura, do mofo-branco e da produtividade, respectivamente. Os três experimentos apresentaram os estágios de desenvolvimento de início do estágio R5, estágio R5 pleno, Início do estágio R6 e Estádio R6 pleno, aos 38, 45, 52 e 59 dias após a emergência (DAE), respectivamente. Os Gráficos 2.1, 2.2 e 2.3 mostram a distribuição da incidência do mofo-branco (IMB) e índice de severidade do mofo-branco (ISMB) em 2021 e 2022 em Viçosa, e em 2022 em Oratórios, respectivamente.

Tabela 2.4 – Caracterização climática durante a realização dos experimentos em Viçosa, considerando a mínima, média e máxima temperaturas diárias e umidade relativa média.

Variável	Viçosa 2021	Viçosa 2022
Média das temperaturas máximas diárias (°C)	17,82	17,38
Média das temperaturas médias diárias (°C)	17,08	16,63
Média das temperaturas mínimas diárias (°C)	16,33	15,89
Umidade relativa média (%)	79,53	78,64
bPrecipitação (mm)	0	0

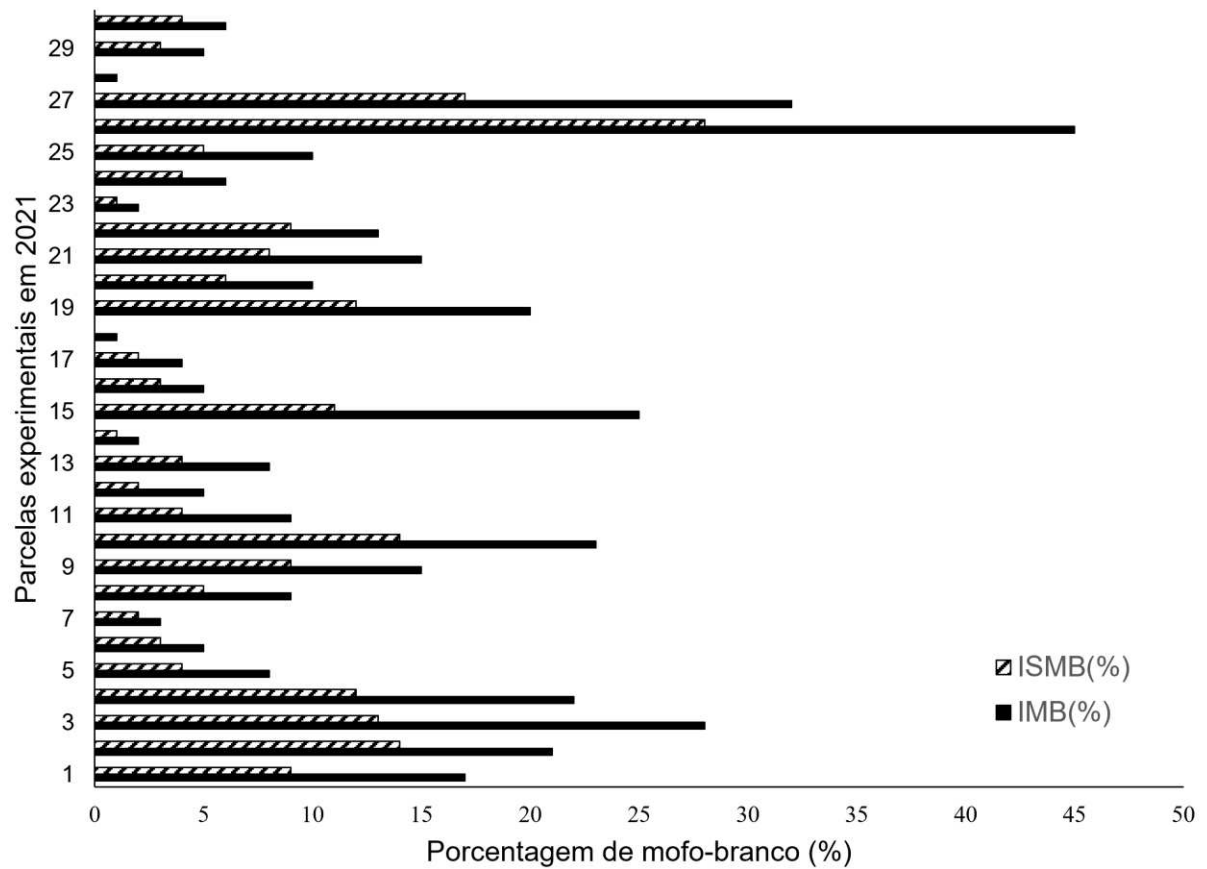
Fonte: elaborado pelo próprio autor.

Tabela 2.5 – Estatística descritiva do desenvolvimento do feijoeiro-comum, do mofo-branco e da produtividade dos experimentos.

Informações da lavoura e do mofo-branco	Oratórios 2022	Viçosa 2021	Viçosa 2022
Início do mofo-branco (DAE)	64	68	68
Avaliação mofo-branco / Colheita (DAE)	104	111	116
Incidência de mofo-branco mínima (%)	0,0	1,0	6,0
Incidência de mofo-branco média (%)	17,7	12,5	29,3
Incidência de mofo-branco máxima (%)	45,0	45,0	68,0
Índice de severidade do mofo-branco mínimo (%)	0,0	0,0	3,0
Índice de severidade do mofo-branco médio (%)	8,0	7,0	17,8
Índice de severidade do mofo-branco máximo (%)	24,0	28,0	45,0
Produtividade mínima (kg.ha ⁻¹)	3240	1813	1732
Produtividade média (kg.ha ⁻¹)	4247	2759	2678
Produtividade máxima (kg.ha ⁻¹)	5245	3388	4185

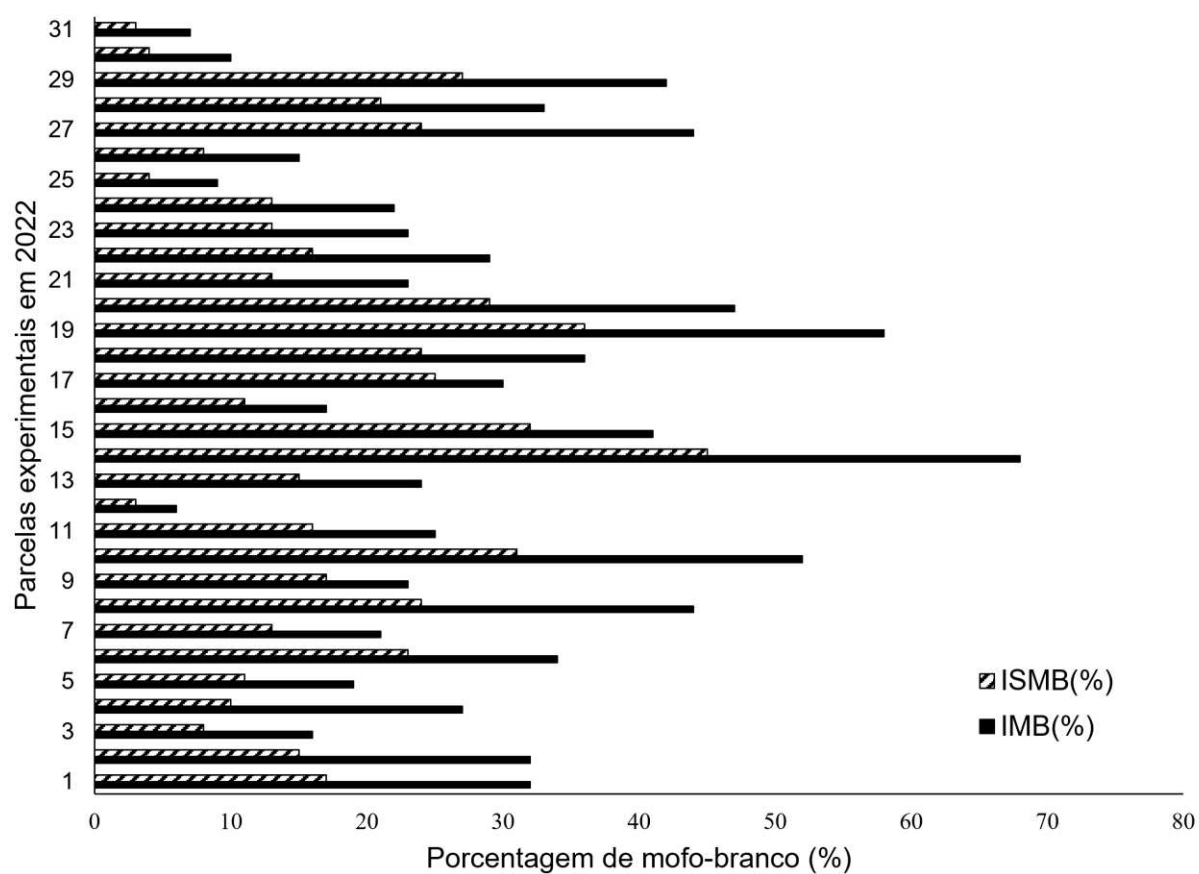
Fonte: elaborado pelo próprio autor.

Gráfico 2.1 – Distribuição da incidência do mofo-branco (IMB) e do índice de severidade do mofo-branco (ISMB) na colheita em 2021 em Viçosa.



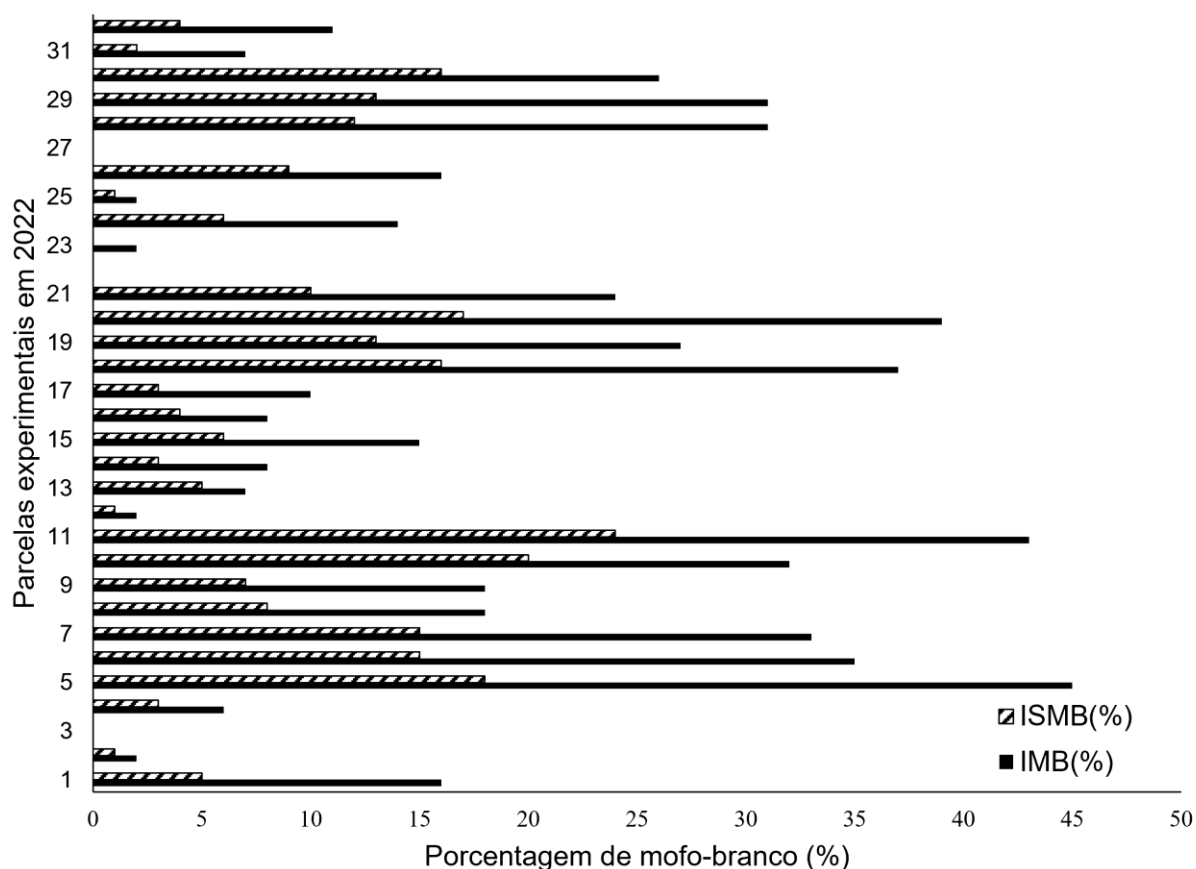
Fonte: elaborado pelo próprio autor.

Gráfico 2.2 – Distribuição da incidência do mofo-branco (IMB) e do índice de severidade do mofo-branco (ISMB) na colheita em 2022 em Viçosa.



Fonte: elaborado pelo próprio autor.

Gráfico 2.3 – Distribuição da incidência do mofo-branco (IMB) e do índice de severidade do mofo-branco (ISMB) na colheita em 2022 em Oratórios.



Fonte: elaborado pelo próprio autor.

Desempenho dos algoritmos de *machine learning* para predição da incidência de mofo-branco (IMB), índice de severidade do mofo-branco (ISMB) e da produtividade do feijoeiro-comum

Neste trabalho, foram usados quatro algoritmos de aprendizagem de máquina para prever a incidência do mofo-branco (IMB), o índice de severidade do mofo-branco (ISMB) e a produtividade do feijoeiro-comum. Entre os modelos usados estão o *Cubist*, *Kernel K Nearest Neighbors* (KNN), *Random Forest* (RF) e *Support Vector Machine Radial Sigma* (SVM). Cada modelo foi avaliado com base em métricas de desempenho, como o coeficiente de determinação (R^2), Erro Quadrático Médio (RMSE), Erro Médio Absoluto (MAE) e o Coeficiente de Correlação de Concordância de Lin (LCCC).

O modelo gerado para predição de produtividade apresentou melhor desempenho que os modelos para predição da incidência do mofo-branco (IMB) e do índice de

severidade do mofo-branco (ISMB) (Tabela 2.6). No entanto, mesmo as modelagens das variáveis de saída que apresentaram avaliações de desempenho mais baixas, apresentaram desempenhos promissores, visto que todos os modelos apresentaram valores de RMSE e MAE superiores ao modelo nulo.

Tabela 2.6 – Valores dos parâmetros de desempenho (R^2 , RMSE, MAE e LCCC) de teste dos modelos nas predições da incidência (IMB) e índice de severidade do mofo-branco (ISMB), e da produtividade (Prod) do feijoeiro-comum.

R^2				
	SVM	RF	Cubist	KKNN
IMB	0,194 a	0,245 ab	0,274 b	0,277 b
ISMB	0,241 a	0,278 a	0,301 a	0,269 a
Prod	0,651 a	0,694 a	0,626 a	0,689 a

RMSE					
	SVM	RF	Cubist	KKNN	MN
IMB	14,35 b	13,39 a	13,41 a	13,49 a	15,04
ISMB	8,45 a	8,08 a	8,07 a	8,22 a	9,22
Prod	545,98 ab	511,21 a	580,62 b	511,11 a	899,53

MAE					
	SVM	RF	Cubist	KKNN	MN
IMB	11,36 b	10,76 a	10,58 a	10,70 a	12,5
ISMB	6,65 a	6,54 a	6,40 a	6,61 a	7,41
Prod	435,01 ab	419,01 a	458,18 b	408,73 a	748,59

LCCC				
	SVM	RF	Cubist	KKNN
IMB	0,360 a	0,379 a	0,454 b	0,448 b
ISMB	0,408 a	0,399 a	0,463 a	0,433 a
Prod	0,776 ab	0,797 ab	0,755 a	0,806 b

*Letras minúsculas na mesma linha diferem entre si ao nível de 5% de significância pelo teste de Kruskal-Wallis.

Fonte: elaborado pelo próprio autor.

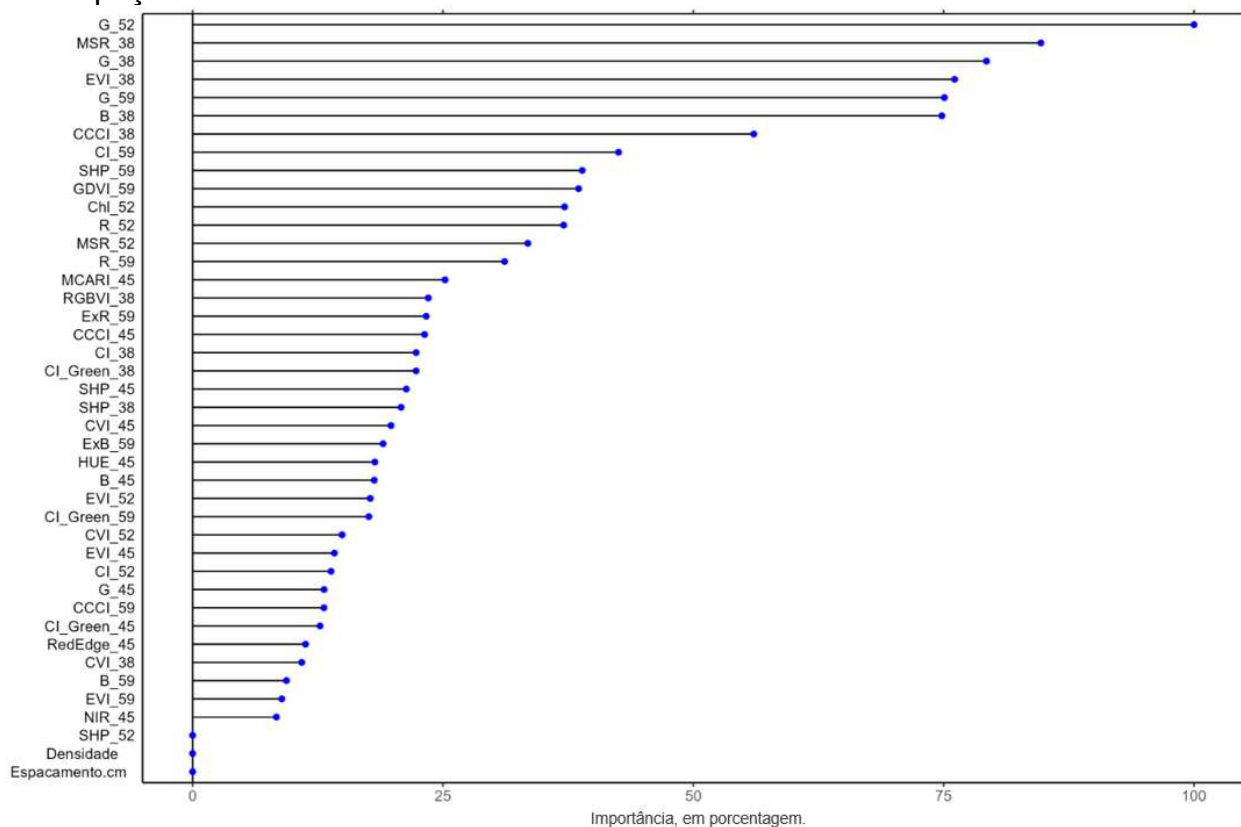
Os modelos dos algoritmos *Cubist*, SVM, RF e KKNN apresentaram valores de MAE e RMSE menores que os valores de MAE e RMSE dos modelos nulos (Tabela 2.6). Para a predição da IMB, o algoritmo que apresentou o melhor desempenho foi o *Cubist*, com os maiores valores de LCCC e menor MAE. Como base no teste Kruskal-Wallis, os

valores de R^2 do *Cubist* e KNN, e os valores do RMSE para o *Cubist*, KNN e RF não diferem significativamente. No entanto, o modelo do *Cubist* apresentou o melhor desempenho na predição da IMB. O pior desempenho foi o do modelo construído pelo SVM. O SVM apresentou os menores valores de R^2 e LCCC, não diferindo significativamente do RF pelo teste de Kruskal-Wallis a 5% de significância. O SVM ainda apresentou os maiores valores de RMSE e MAE, sendo estatisticamente diferente ao teste de Kruskal-Wallis 5% de significância que os demais algoritmos.

Para a predição do ISMB, o algoritmo *Cubist* apresentou o melhor desempenho, com os maiores valores de R^2 e LCCC, e menores valores de RMSE e MAE (Tabela 2.6). Mesmo o algoritmo *Cubist* apresentando maiores valores absolutos de R^2 e LCCC, os algoritmos SVM, KNN e RF não diferiram do *Cubist*, pelo teste de Kruskal-Wallis a 5 % de significância. Em valores absolutos, o algoritmo SVM teve o pior desempenho, mas ele não diferiu significativamente dos demais algoritmos. Quanto a produtividade, o desempenho do algoritmo KNN apresentou maior valor de LCCC e menores valores de RMSE e MAE, o que o caracteriza como melhor modelo em valores absolutos. No entanto, o RF apresentou maior valor de R^2 e valores de LCCC e MAE que não diferiram significativamente do KNN. O pior desempenho na modelagem da produtividade foi o do algoritmo *Cubist*, com menores valores de R^2 e LCCC e maiores valores de RMSE e MAE.

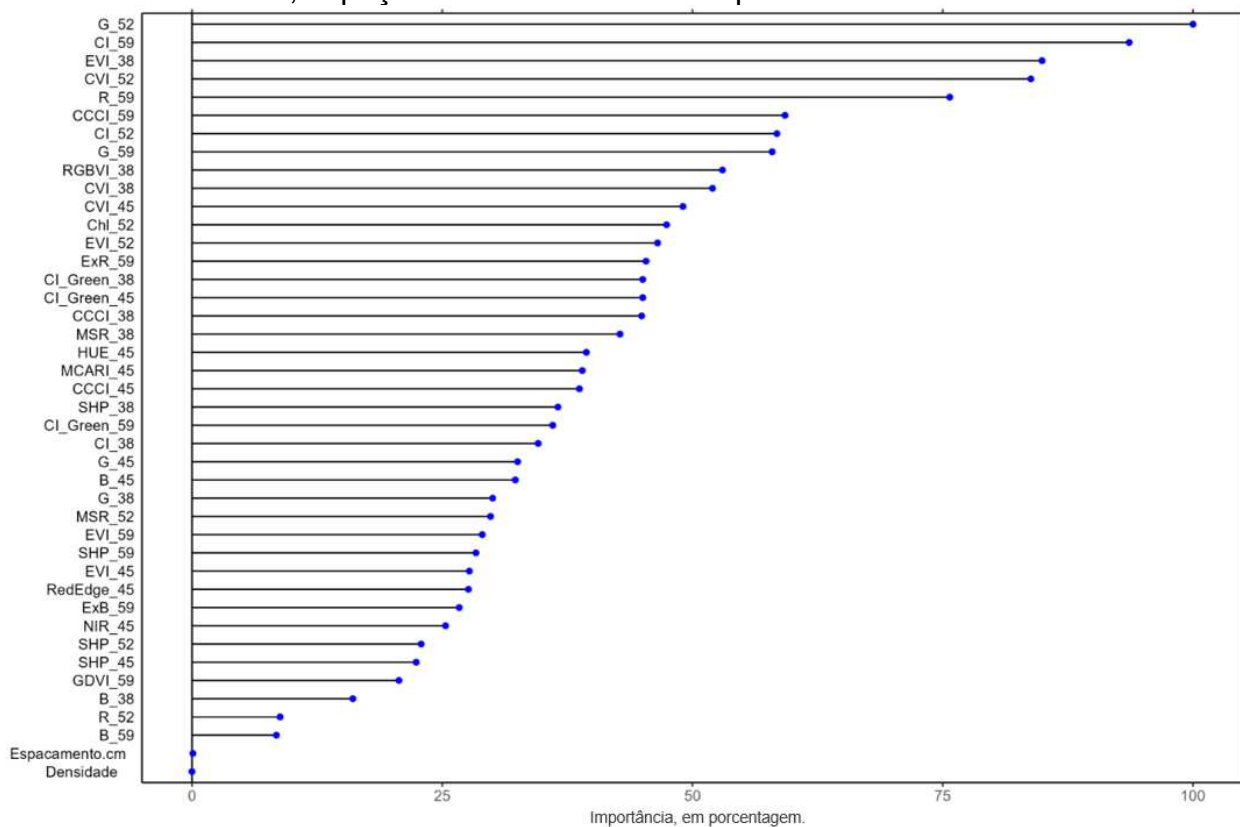
As covariáveis mais importantes selecionadas por meio da função RFE foram diferentes para cada algoritmo. Essa diferença no conjunto de covariáveis selecionadas se deve aos procedimentos adotados em cada algoritmo. Os gráficos de importância das covariáveis selecionadas na modelagem de cada algoritmo para o IMB, ISMB e produtividade estão apresentadas nas Figuras 2.4 a 2.7, nas Figuras 2.8 a 2.11, e nas Figuras 2.12 a 2.15, respectivamente.

Figura 2.4 – Seleção de covariáveis mais importantes para a modelagem da predição do IMB por meio do algoritmo *Cubist*. As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G aos 52 DAE (dias após a emergência), o MSR aos 38 DAE, banda G aos 38 DAE, EVI ao 38 DAE, banda G aos 59 DAE, banda B aos 38 DAE, CCCI aos 38 DAE, CI aos 59 DAE, SHP aos 59 DAE, GDVI aos 59 DAE, ChI aos 52 DAE, banda R aos 52 DAE, MSR aos 52 DAE, banda R aos 59 DAE, MCARI aos 45 DAE, RGBVI aos 38 DAE, ExR aos 59 DAE, CCCI aos 45 DAE, CI aos 38 DAE, CI_Green aos 38 DAE, SHP aos 45 DAE, SHP aos 38 DAE, CVI aos 45 DAE, ExB aos 59 DAE, HUE aos 45 DAE, banda B aos 45 DAE, EVI aos 52 DAE, CI_Green aos 59 DAE, CVI aos 52 DAE, EVI aos 45 DAE, CI aos 52 DAE, banda G aos 45 DAE, CCCI aos 59 DAE, CI_Green aos 45 DAE, banda RedEdge aos 45 DAE, CVI aos 38 DAE, banda B aos 59 DAE, EVI aos 59 DAE, banda NIR aos 45 DAE, SHP aos 52 DAE, densidade de plantio e o espaçamento.



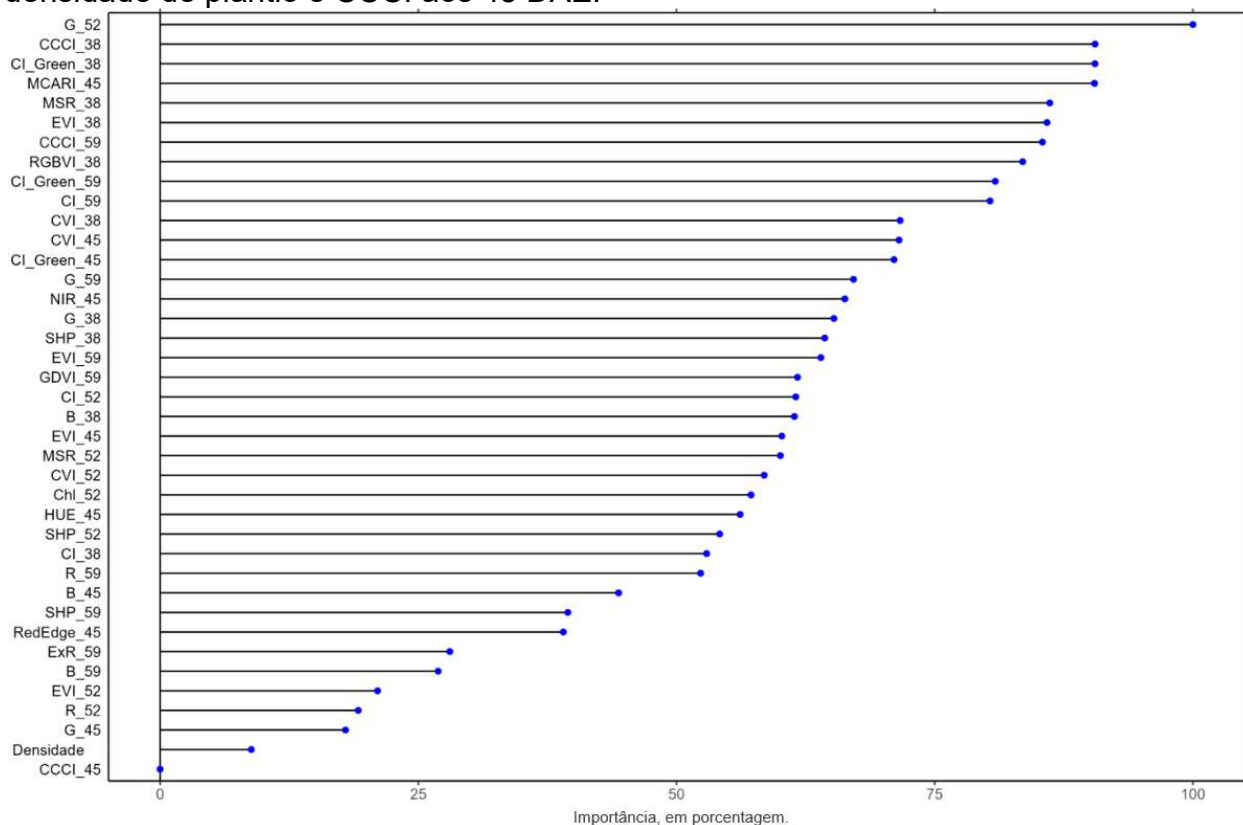
Fonte: elaborado pelo próprio autor.

Figura 2.5 – Seleção das covariáveis mais importantes para a modelagem da predição do IMB por meio do algoritmo *Kernel K Nearest Neighbors* (KKN). As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G ao 52 dias após a emergência (DAE), CI aos 59 DAE, EVI aos 38 DAE, CVI aos 52 DAE, banda R aos 59 DAE, CCCI aos 59 DAE, CI aos 52 DAE, banda G aos 59 DAE, RGBVI aos 38 DAE, CVI aos 38 DAE, CVI aos 45 DAE, Chl aos 52 DAE, EVI aos 52 DAE, ExR aos 59 DAE, CI_Green aos 38 DAE, CI_Green aos 45 DAE, CCCI aos 38 DAE, MSR aos 38 DAE, HUE aos 45 DAE, MCARI aos 45 DAE, CCCI aos 45 DAE, SHP aos 38 DAE, CI_Green aos 59 DAE, CI aos 38 DAE, banda G ao 45 DAE, banda B ao 45 DAE, banda G ao 38 DAE, MSR aos 52 DAE, EVI aos 59 DAE, SHP aos 59 DAE, EVI aos 45 DAE, banda RedEdge aos 45 DAE, ExB aos 59 DAE, banda NIR aos 45 DAE, SHP aos 52 DAE, SHP aos 45 DAE, GDVI aos 59 DAE, banda B ao 38 DAE, banda R ao 52 DAE, banda B ao 59 DAE, espaçamento e densidade de plantio.



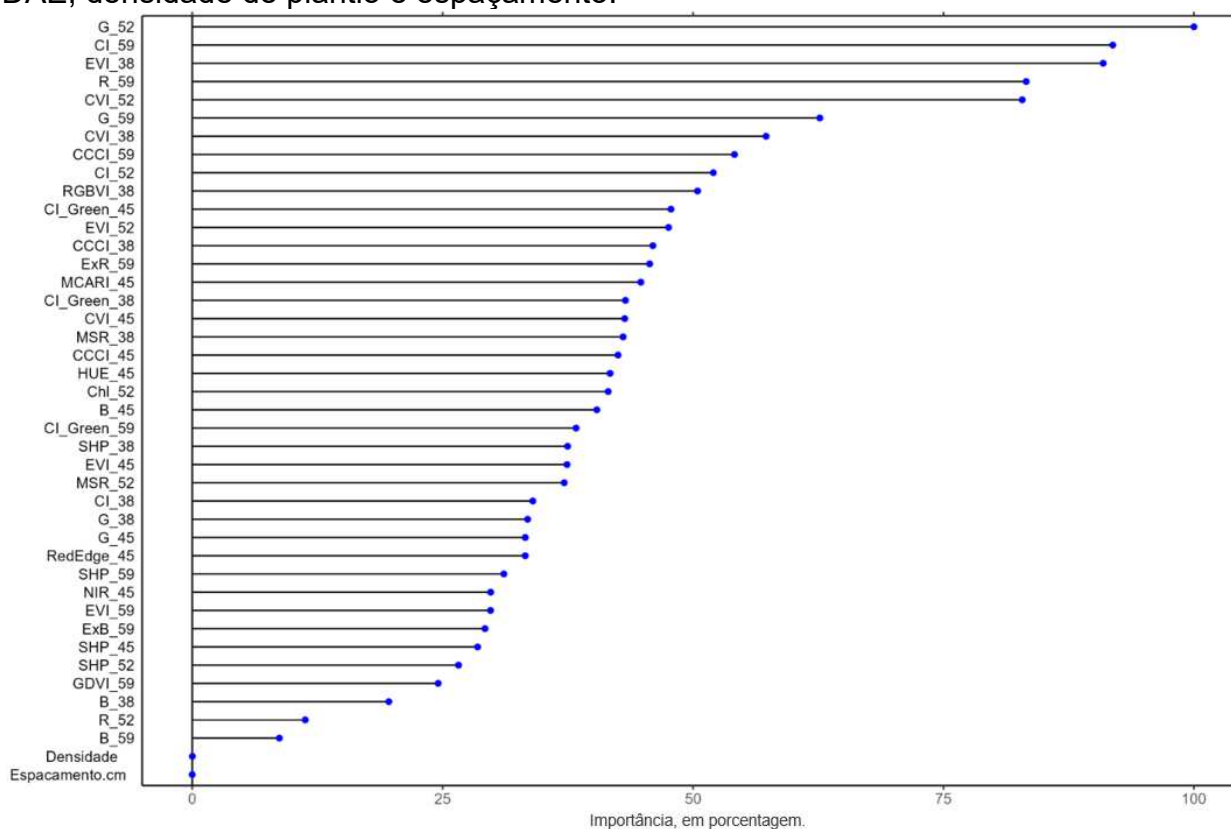
Fonte: elaborado pelo próprio autor.

Figura 2.6 – Seleção das covariáveis mais importantes para a modelagem da predição do IMB por meio do algoritmo *Random Forest* (RF). As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G aos 52 dias após a emergência (DAE,) CCCI aos 38 DAE, CI_Green aos 38 DAE, MCARI aos 45 DAE, MSR aos 38 DAE, EVI aos 38 DAE, CCCI aos 59 DAE, RGBVI aos 38 DAE, CI_Green aos 59 DAE, CI aos 59 DAE, CVI aos 38 DAE, CVI aos 45 DAE, CI_Green aos 45 DAE, G aos 59 DAE, NIR aos 45 DAE, G aos 38 DAE, SHP aos 38 DAE, EVI aos 59 DAE, GDVI aos 59 DAE, CI aos 52 DAE, B aos 38 DAE, EVI aos 45 DAE, MSR aos 52 DAE, CVI aos 52 DAE, Chl aos 52 DAE, HUE aos 45 DAE, SHP aos 52 DAE, CI aos 38 DAE, banda R aos 59 DAE, banda B aos 45 DAE, SHP aos 59 DAE, banda RedEdge aos 45 DAE, ExR aos 59 DAE, banda B aos 59 DAE, EVI aos 52 DAE, banda R aos 52 DAE, banda G aos 45 DAE, densidade de plantio e CCCI aos 45 DAE.



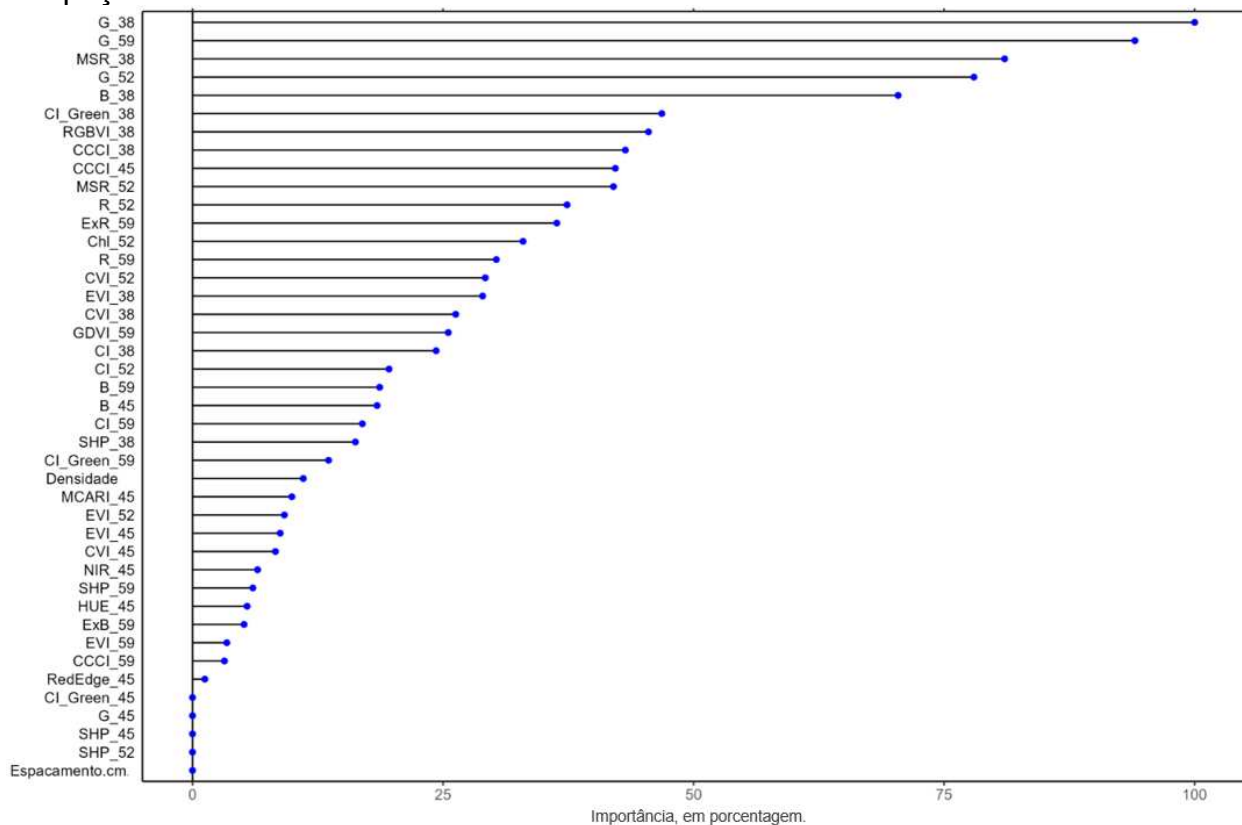
Fonte: elaborado pelo próprio autor.

Figura 2.7 – Seleção das covariáveis mais importantes para a modelagem da predição do IMB por meio do algoritmo *Support Vector Machine* (SVM). As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G aos 52 dias após a emergência (DAE), CI aos 59 DAE, EVI aos 38 DAE, R aos 59 DAE, CVI aos 52 DAE, banda G aos 59 DAE, CVI aos 38 DAE, CCCI aos 59 DAE, CI aos 52 DAE, RGBVI aos 38 DAE, CI_Green aos 45 DAE, EVI aos 52 DAE, CCCI aos 38 DAE, ExR aos 59 DAE, MCARI aos 45 DAE, CI_Green aos 38 DAE, CVI aos 45 DAE, MSR aos 38 DAE, CCCI aos 45 DAE, HUE aos 45 DAE, Chl aos 52 DAE, banda B aos 45 DAE, CI_Green aos 59 DAE, SHP aos 38 DAE, EVI aos 45 DAE, MSR aos 52 DAE, CI aos 38 DAE, banda G aos 38 DAE, banda G aos 45 DAE, banda RedEdge aos 45 DAE, SHP aos 59 DAE, banda NIR aos 45 DAE, EVI aos 59 DAE, ExB aos 59 DAE, SHP aos 45 DAE, SHP aos 52 DAE, GDVI aos 59 DAE, banda B aos 38 DAE, banda R aos 52 DAE, banda B aos 59 DAE, densidade de plantio e espaçamento.



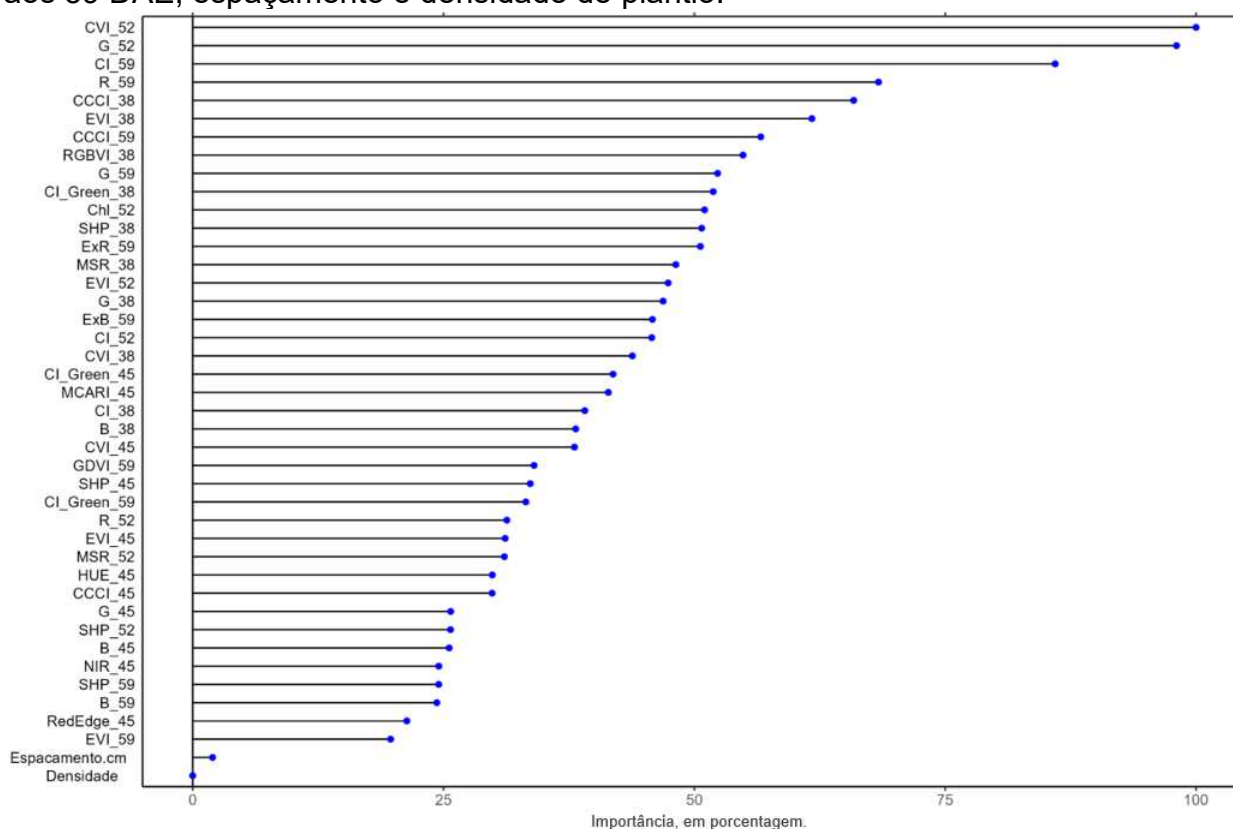
Fonte: elaborado pelo próprio autor.

Figura 2.8 – Seleção das covariáveis mais importantes para a modelagem da predição do ISMB por meio do algoritmo *Cubist*. As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G aos 38 dias após a emergência (DAE), banda G aos 59 DAE, MSR aos 38 DAE, banda G aos 52 DAE, banda B aos 38 DAE, CI_Green aos 38 DAE, RGBVI aos 38 DAE, CCCI aos 38 DAE, CCCI aos 45 DAE, MSR aos 45 DAE, banda R aos 52 DAE, ExR aos 59 DAE, Chl aos 52 DAE, banda R aos 59 DAE, CVI aos 52 DAE, EVI aos 38 DAE, CVI aos 38 DAE, GDVI aos 59 DAE, CI aos 38 DAE, CI aos 52 DAE, banda B aos 59 DAE, banda B aos 45 DAE, CI aos 59 DAE, SHP aos 38 DAE, CI_Green aos 59 DAE, densidade de plantio, MCARI aos 45 DAE, EVI aos 52 DAE, EVI aos 45 DAE, CVI aos 45 DAE, banda NIR aos 45 DAE, SHP aos 59 DAE, HUE aos 45 DAE, ExB aos 59 DAE, EVI aos 59 DAE, CCCI aos 59 DAE, banda RedEdge aos 45 DAE, CI_Green aos 45 DAE, banda G aos 45 DAE, SHP aos 45 DAE, SHP aos 52 DAE e espaçamento.



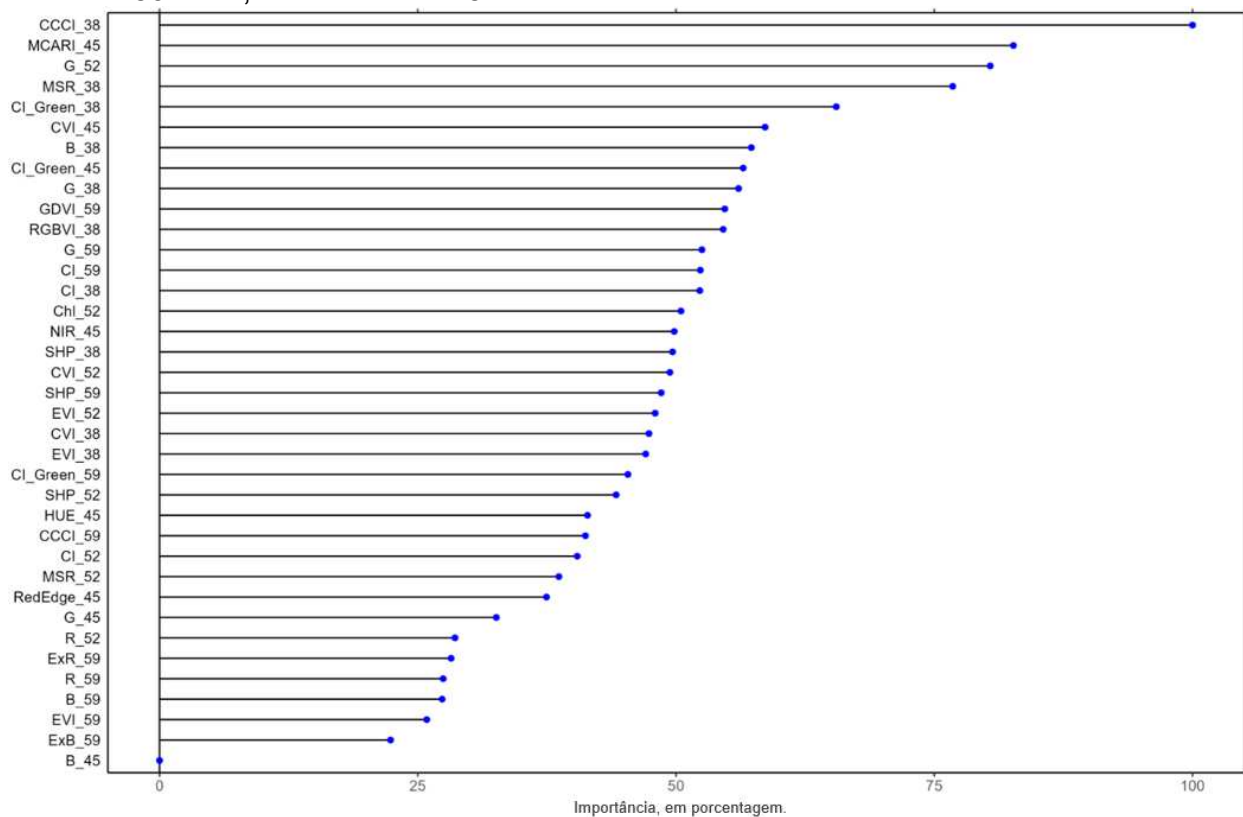
Fonte: elaborado pelo próprio autor.

Figura 2.9 – Seleção das covariáveis mais importantes para a modelagem da predição do ISMB por meio do algoritmo *Kernel K Nearest Neighbors* (KNN). As covariáveis selecionadas para a modelagem em ordem de importância foram: CVI aos 52 dias após a emergência (DAE), banda G aos 52 DAE, CI aos 59 DAE, banda R aos 59 DAE, CCCI aos 38 DAE, EVI aos 38 DAE, CCCI aos 59 DAE, RGBVI aos 38 DAE, banda G aos 59 DAE, CI_Green aos 38 DAE, Chl aos 52 DAE, SHP aos 38 DAE, ExR aos 59 DAE, MSR aos 38 DAE, EVI aos 52 DAE, banda G aos 38 DAE, ExB aos 59 DAE, CI aos 52 DAE, CVI aos 38 DAE, CI_Green aos 45 DAE, MCARI aos 45 DAE, CI aos 38 DAE, banda B aos 38 DAE, CVI aos 45 DAE, GDVI aos 59 DAE, SHP aos 45 DAE, CI_Green aos 59 DAE, banda R aos 52 DAE, EVI aos 45 DAE, MSR aos 52 DAE, HUE aos 45 DAE, CCCI aos 45 DAE, banda G aos 45 DAE, SHP aos 52 DAE, banda B aos 45 DAE, banda NIR aos 45 DAE, SHP aos 59 DAE, banda B aos 59 DAE, banda RedEdge aos 45 DAE, EVI aos 59 DAE, espaçamento e densidade de plantio.



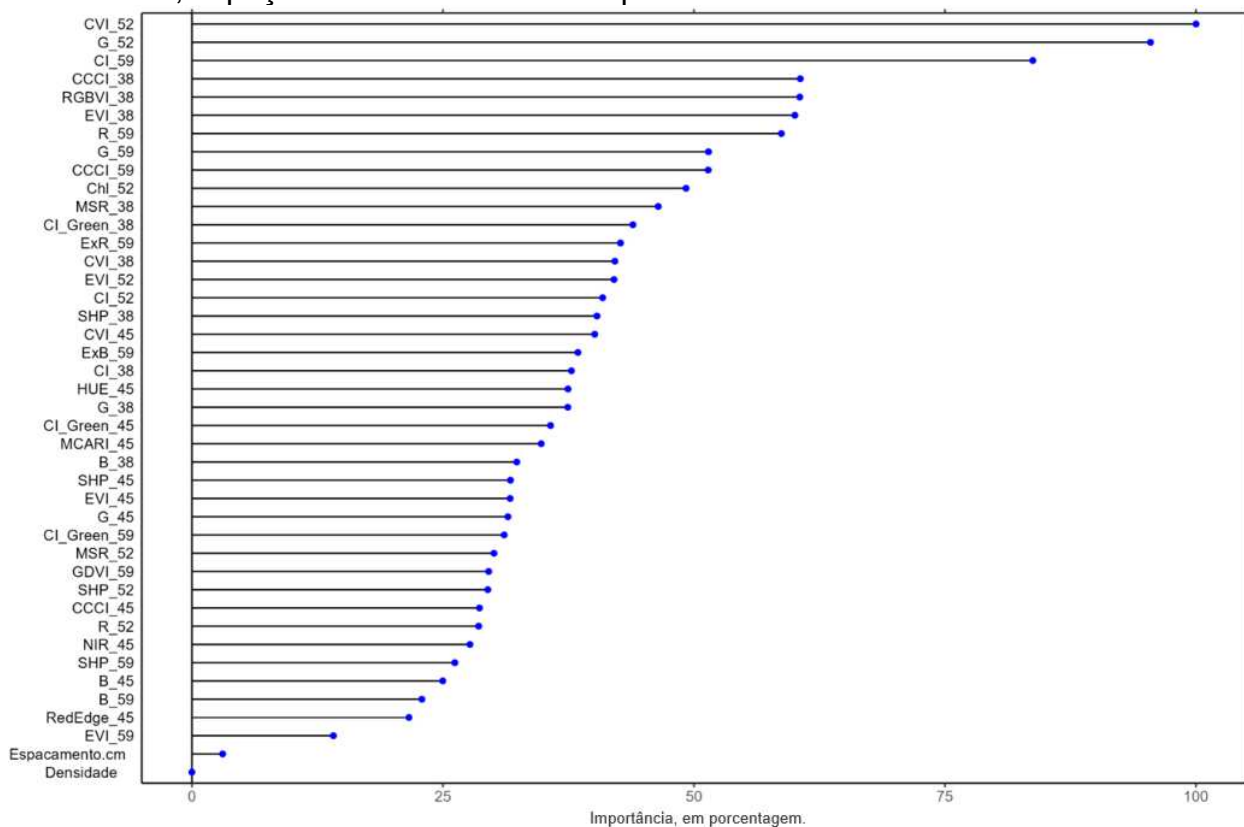
Fonte: elaborado pelo próprio autor.

Figura 2.10 – Seleção das covariáveis mais importantes para a modelagem da predição do ISMB por meio do algoritmo *Random Forest* (RF). As covariáveis selecionadas para a modelagem em ordem de importância foram: CCCl aos 38 dias após a emergência (DAE), MCARI aos 45 DAE, banda G aos 52 DAE, MSR aos 38 DAE, CI_Green aos 38 DAE, CVI aos 45 DAE, banda B aos 38 DAE, CI_Green aos 45 DAE, banda G aos 38 DAE, GDVI aos 59 DAE, RGBVI aos 38 DAE, banda G aos 59 DAE, CI aos 59 DAE, CI aos 38 DAE, Chl aos 52 DAE, banda NIR aos 45 DAE, SHP aos 38 DAE, CVI aos 52 DAE, SHP aos 59 DAE, EVI aos 52 DAE, CVI aos 38 DAE, EVI aos 38 DAE, CI_Green aos 59 DAE, SHP aos 52 DAE, HUE aos 45 DAE, CCCl aos 59 DAE, CI aos 52 DAE, MSR aos 52 DAE, banda RedEdge aos 45 DAE, banda G aos 45 DAE, banda R aos 52 DAE, ExR aos 59 DAE, banda R aos 59 DAE, banda B aos 59 DAE, EVI aos 59 DAE, ExB aos 59 DAE, banda B aos 45 DAE.



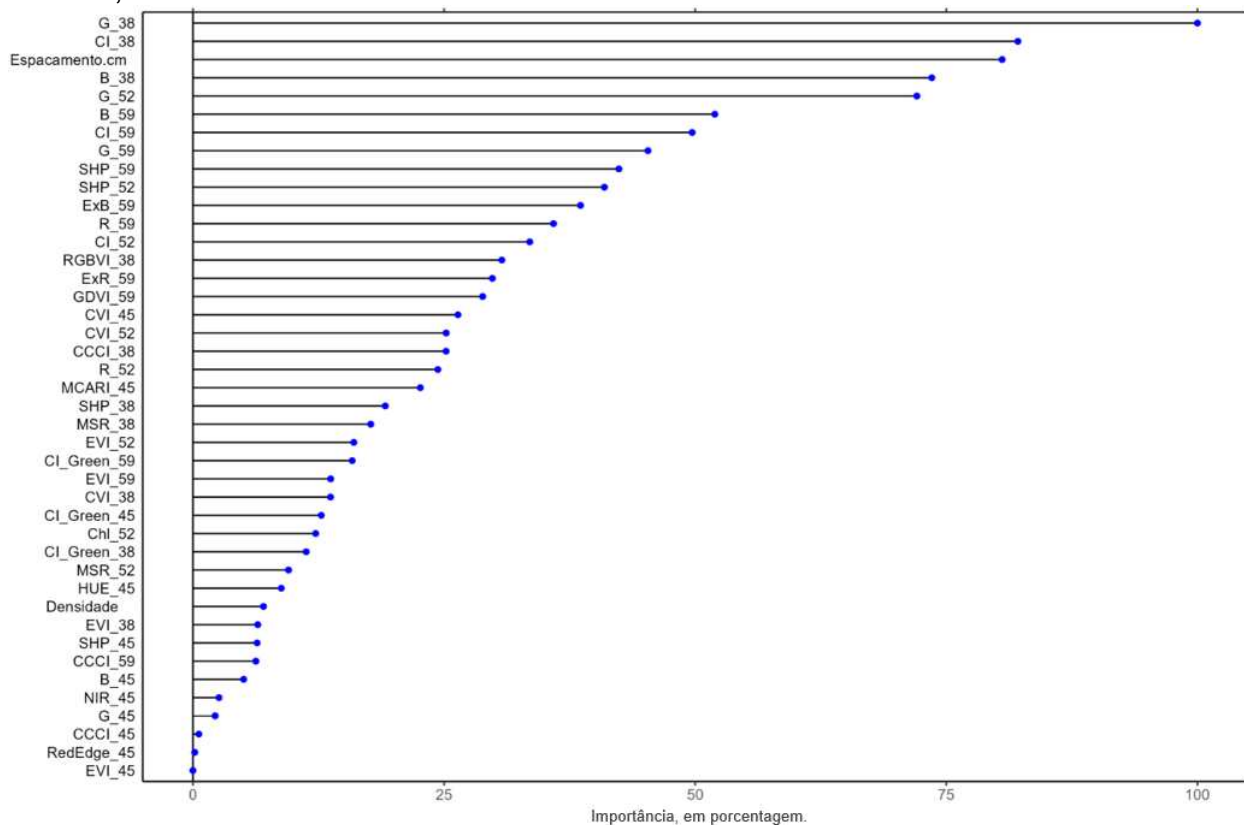
Fonte: elaborado pelo próprio autor.

Figura 2.11 – Seleção das covariáveis mais importantes para a modelagem da predição do ISMB por meio do algoritmo *Support Vector Machine* (SVM). As covariáveis selecionadas para a modelagem em ordem de importância foram: CVI aos 52 dias após a emergência (DAE), banda G aos 52 DAE, CI aos 59 DAE, CCCI aos 38 DAE, RGBVI aos 38 DAE, EVI aos 38 DAE, banda R aos 59 DAE, banda G aos 59 DAE, CCCI aos 59 DAE, Chl aos 52 DAE, MSR aos 38 DAE, CI_Green aos 38 DAE, ExR aos 59 DAE, CVI aos 38 DAE, EVI aos 52 DAE, CI aos 52 DAE, SHP aos 38 DAE, CVI aos 45 DAE, ExB aos 59 DAE, CI aos 38 DAE, HUE aos 45 DAE, banda G aos 38 DAE, CI_Green aos 45 DAE, MCARI aos 45 DAE, banda B aos 38 DAE, SHP aos 45 DAE, EVI aos 45 DAE, banda G aos 45 DAE, CI_Green aos 59 DAE, MSR aos 52 DAE, GDVI aos 59 DAE, SHP aos 52 DAE, CCCI aos 45 DAE, banda R aos 52 DAE, banda NIR aos 45 DAE, SHP aos 59 DAE, banda B aos 45 DAE, banda B aos 59 DAE, banda RedEdge aos 45 DAE, EVI aos 59 DAE, espaçamento e densidade de plantio.



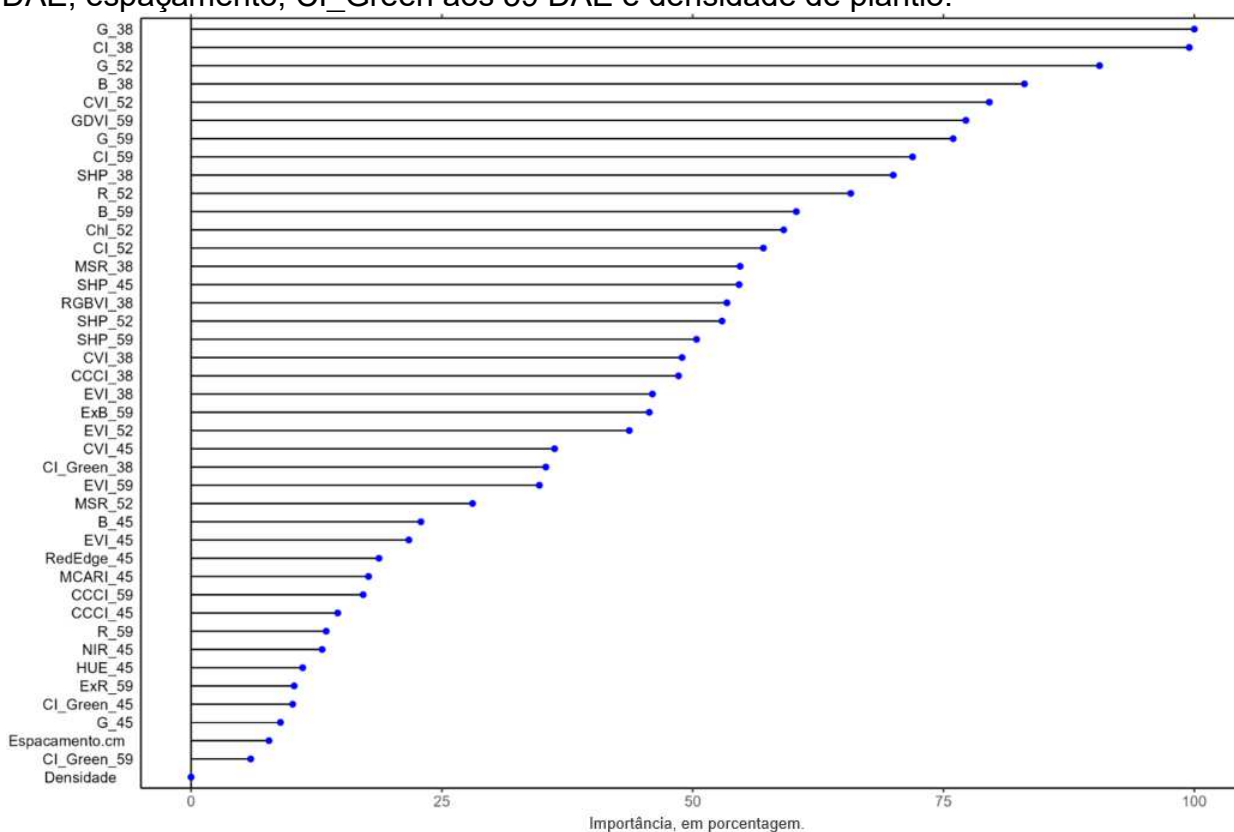
Fonte: elaborado pelo próprio autor.

Figura 2.12 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade por meio do algoritmo *Cubist*. As covariáveis selecionadas para a modelagem em ordem de importância foram: G aos 38 dias após a emergência (DAE), CI aos 38 DAE, espaçamento, banda B aos 38 DAE, banda G aos 52 DAE, banda B aos 59 DAE, CI aos 59 DAE, banda G aos 59 DAE, SHP aos 59 DAE, SHP aos 52 DAE, ExB aos 59 DAE, banda R aos 59 DAE, CI aos 52 DAE, RGBVI aos 38 DAE, ExR aos 59 DAE, GDVI aos 59 DAE, CVI aos 45 DAE, CVI aos 52 DAE, CCCI aos 38 DAE, banda R aos 52 DAE, MCARI aos 45 DAE, SHP aos 38 DAE, MSR aos 38 DAE, EVI aos 52 DAE, CI_Green aos 59 DAE, EVI aos 59 DAE, CVI aos 38 DAE, CI_Green aos 45 DAE, Chl aos 52 DAE, CI_Green aos 38 DAE, MSR aos 52 DAE, HUE aos 45 DAE, densidade de plantio, EVI aos 38 DAE, SHP aos 45 DAE, CCCI aos 59 DAE, banda B aos 45 DAE, banda NIR aos 45 DAE, banda G aos 45 DAE, CCCI aos 45 DAE, banda RedEdge aos 45 DAE, EVI aos 45 DAE.



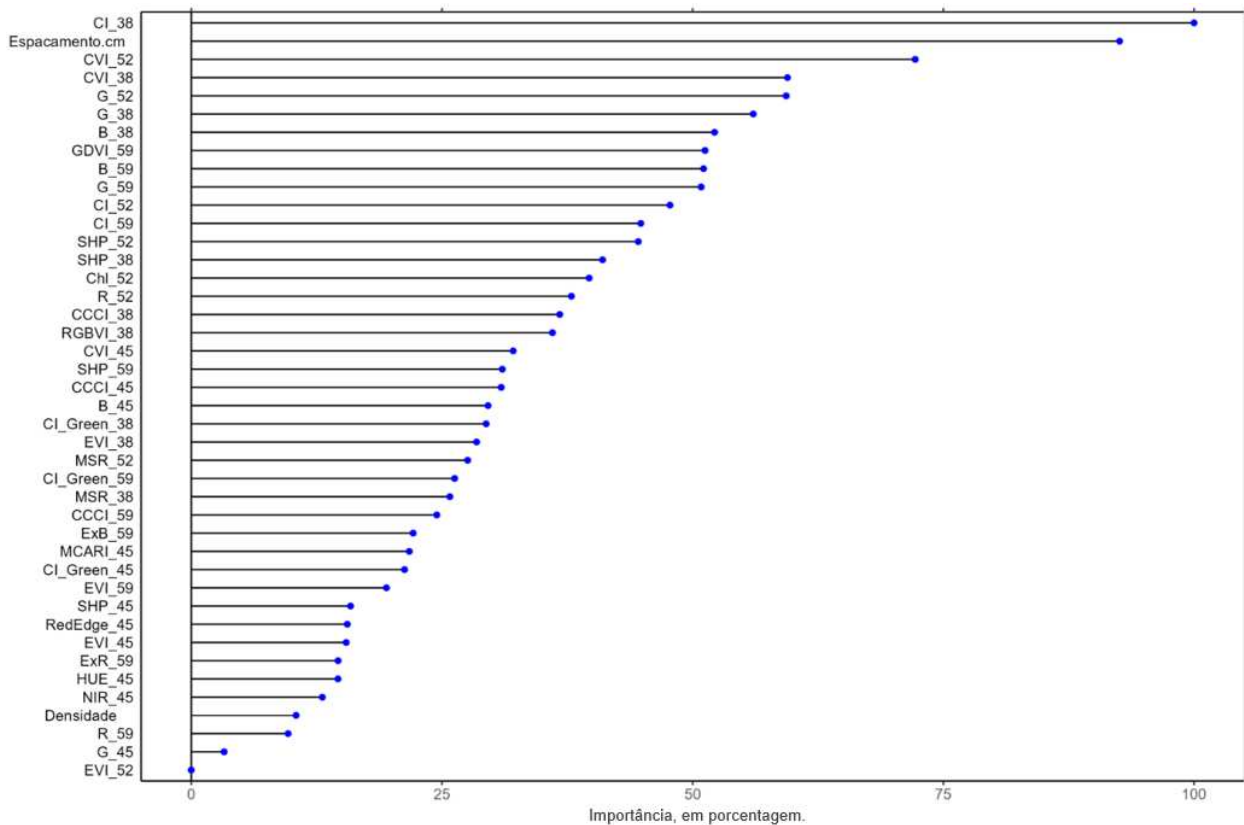
Fonte: elaborado pelo próprio autor.

Figura 2.13 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade por meio do algoritmo *Kernel K Nearest Neighbors* (KNN). As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G aos 38 dias após a emergência (DAE), CI aos 38 DAE, banda G aos 52 DAE, banda B aos 38 DAE, CVI aos 52 DAE, GDVI aos 59 DAE, banda G aos 59 DAE, CI aos 59 DAE, SHP aos 38 DAE, banda R aos 52 DAE, banda B aos 59 DAE, Chl aos 52 DAE, CI aos 52 DAE, MSR aos 38 DAE, SHP aos 45 DAE, RGBVI aos 38 DAE, SHP aos 52 DAE, SHP aos 59 DAE, CVI aos 38 DAE, CCCI aos 38 DAE, EVI aos 38 DAE, ExB aos 59 DAE, EVI aos 52 DAE, CVI aos 45 DAE, CI_Green aos 38 DAE, EVI aos 59 DAE, MSR aos 52 DAE, banda B aos 45 DAE, EVI aos 45 DAE, banda RedEdge aos 45 DAE, MCARI aos 45 DAE, CCCI aos 59 DAE, CCCI aos 45 DAE, banda R aos 59 DAE, banda NIR aos 45 DAE, HUE aos 45 DAE, ExR aos 59 DAE, CI_Green aos 45 DAE, banda G aos 45 DAE, espaçamento, CI_Green aos 59 DAE e densidade de plantio.



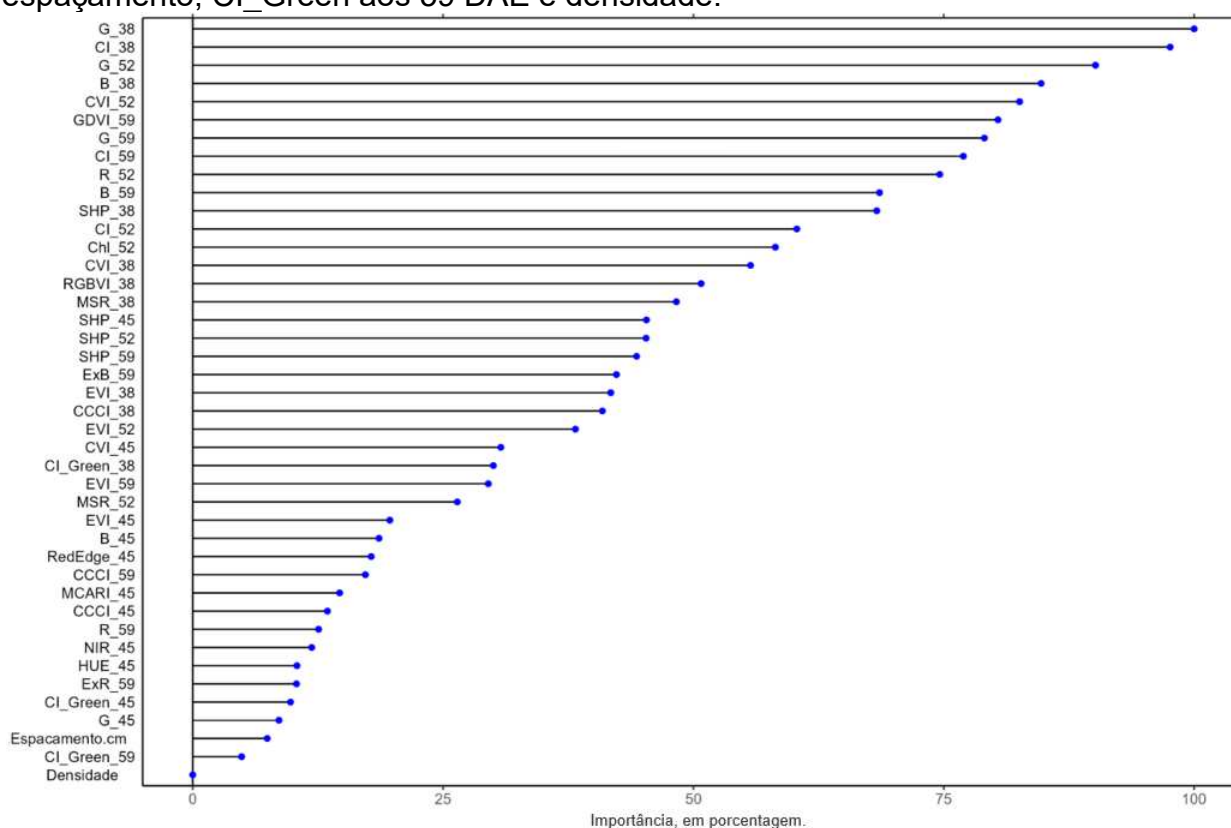
Fonte: elaborado pelo próprio autor.

Figura 2.14 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade por meio do algoritmo *Random Forest* (RF). As covariáveis selecionadas para a modelagem em ordem de importância foram: CI aos 38 dias após a emergência (DAE), espaçamento, CVI aos 52 DAE, CVI aos 38 DAE, banda G aos 52 DAE, banda G aos 38 DAE, banda B aos 38 DAE, GDVI aos 59 DAE, banda B aos 59 DAE, banda G aos 59 DAE, CI aos 52 DAE, CI aos 59 DAE, SHP aos 52 DAE, SHP aos 38 DAE, Chl aos 52 DAE, banda R aos 52 DAE, CCCI aos 38 DAE, RGBVI aos 38 DAE, CVI aos 45 DAE, SHP aos 59 DAE, CCCI aos 45 DAE, banda B aos 45 DAE, CI_Green aos 38 DAE, EVI aos 38 DAE, MSR aos 52 DAE, CI_Green aos 59 DAE, MSR aos 38 DAE, CCCI aos 59 DAE, ExB aos 59 DAE, MCARI aos 45 DAE, CI_Green aos 45 DAE, EVI aos 59 DAE, SHP aos 45 DAE, banda RedEdge aos 45 DAE, EVI aos 45 DAE, ExR aos 59 DAE, HUE aos 45 DAE, NIR aos 45 DAE, densidade de plantio, banda R aos 59 DAE, banda G aos 45 DAE e EVI aos 52 DAE.



Fonte: elaborado pelo próprio autor.

Figura 2.15 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade por meio do algoritmo *Support Vector Machine*. As covariáveis selecionadas para a modelagem em ordem de importância foram: banda G aos 38 dias após a emergência (DAE), CI aos 38 DAE, banda G aos 52 DAE, banda B aos 38 DAE, CVI aos 52 DAE, GDVI aos 59 DAE, banda G aos 59 DAE, CI aos 59 DAE, banda R aos 52 DAE, banda B aos 59 DAE, SHP aos 38 DAE, CI aos 52 DAE, Chl aos 52 DAE, CVI aos 38 DAE, RGBVI aos 38 DAE, MSR aos 38 DAE, SHP aos 45 DAE, SHP aos 52 DAE, SHP aos 59 DAE, ExB aos 59 DAE, EVI aos 38 DAE, CCCI aos 38 DAE, EVI aos 52 DAE, CVI aos 45 DAE, CI_Green aos 38 DAE, EVI aos 59 DAE, MSR aos 52 DAE, EVI aos 45 DAE, banda B aos 45 DAE, banda RedEdge aos 45 DAE, CCCI aos 59 DAE, MCARI aos 45 DAE, CCCI aos 45 DAE, banda R aos 59 DAE, banda NIR aos 59 DAE, HUE aos 45 DAE, ExR aos 59 DAE, CI_Green aos 45 DAE, banda G aos 45 DAE, espaçamento, CI_Green aos 59 DAE e densidade.

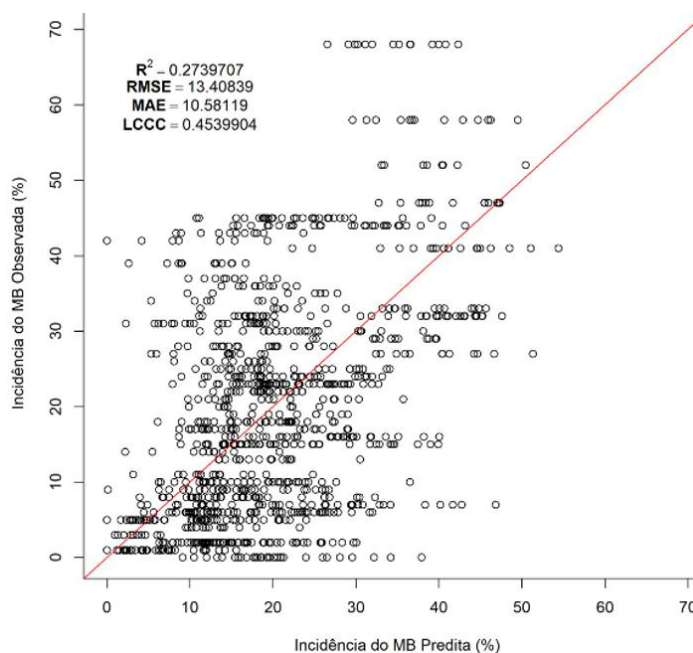


Fonte: elaborado pelo próprio autor.

Os algoritmos aqui usados, Cubist, KNN, RF e SVM, selecionaram diferentes conjuntos de covariáveis para a predição das três variáveis de saída, IMB, ISMB e produtividade. Como exposto nas Figuras 2.4 a 2.15, as covariáveis mais importantes foram selecionadas para os modelos gerados para a predição do IMB, ISMB e produtividade pelos algoritmos, respectivamente. Nas Figuras 2.16, 2.17 e 2.18 são apresentados os gráficos de dispersão entre os valores observados e os preditos para o

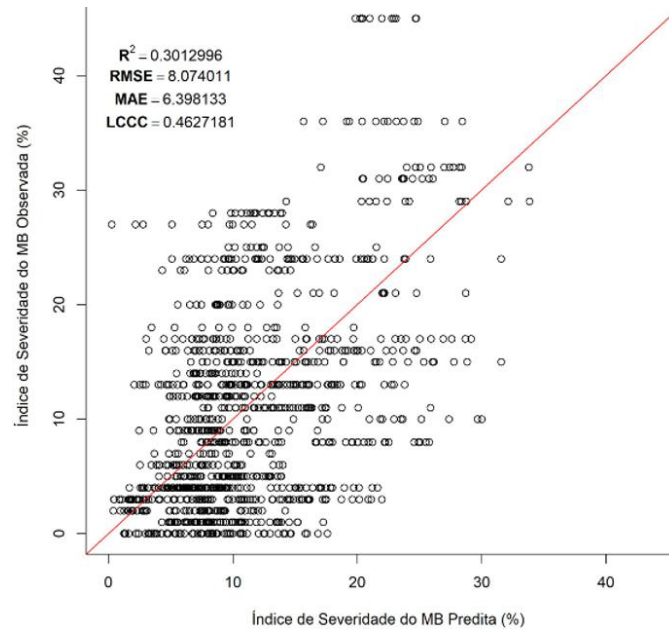
conjunto covariáveis e para o algoritmo que apresentou o melhor desempenho nas previsões do IMB, ISMB e produtividade do feijoeiro-comum, respectivamente. Nas Figuras 2.19 e 2.20 são apresentados o desempenho temporal das bandas e índices das melhores modelagens pelos algoritmos, sendo o Cubist para previsões do IMB (Figura 2.4) e ISMB (Figura 2.8), e o KKNn para a previsão da produtividade (Figura 2.13). Na Figura 2.19 os valores médios de produtividade para as condições de IMB = 0%, $0\% < \text{IMB} \leq 25\%$, $25\% < \text{IMB} \leq 50\%$ e $50\% < \text{IMB} \leq 75\%$ foram de 4.043, 3.214, 3.349 e 2.100 kg.ha^{-1} , respectivamente. As médias de produtividade de ISMB = 0%, $0\% < \text{ISMB} \leq 25\%$ e $25\% < \text{ISMB} \leq 50\%$ foram 3.428, 3.326 e 2.151 kg.ha^{-1} (Fig. 2.20).

Figura 2.16 – Gráfico de dispersão dos resultados de teste entre os valores observados e preditos da modelagem do algoritmo *Cubist* na previsão da incidência do mofo-branco (MB).



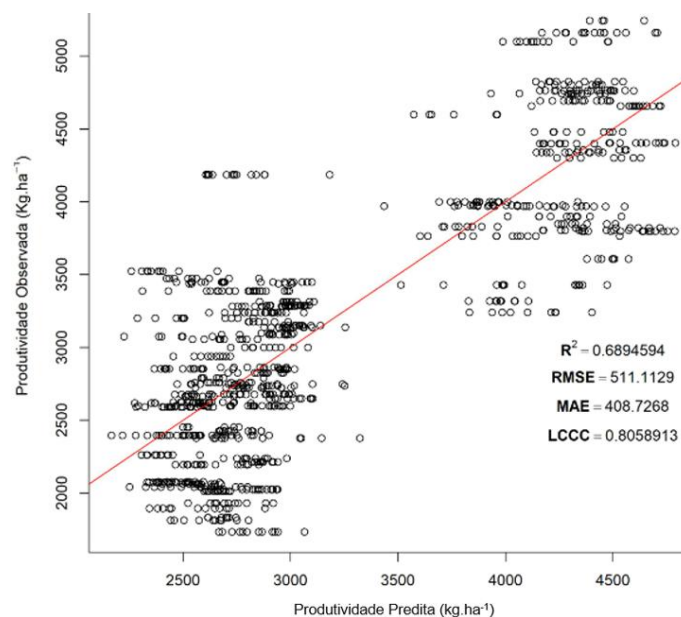
Fonte: elaborado pelo próprio autor.

Figura 2.17 – Gráfico de dispersão dos resultados de teste entre os valores observados e preditos da modelagem do algoritmo *Cubist* na predição do índice de severidade mofo-branco (ISMB).



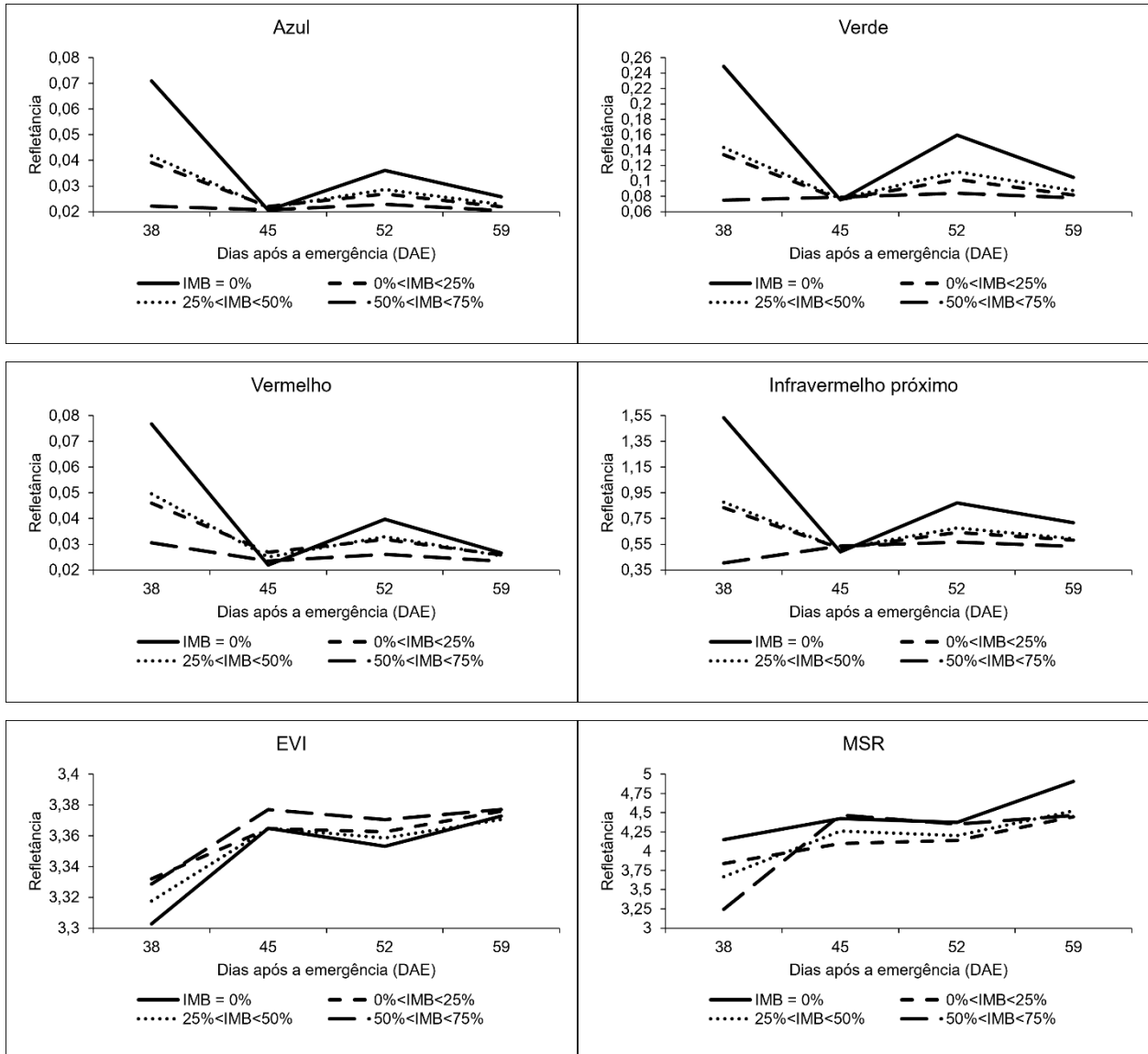
Fonte: elaborado pelo próprio autor.

Figura 2.18 – Gráfico de dispersão dos resultados de teste entre os valores observados e preditos da modelagem do algoritmo *Kernel K Nearest Neighbors* (KNN) na predição da produtividade do feijoeiro-comum.



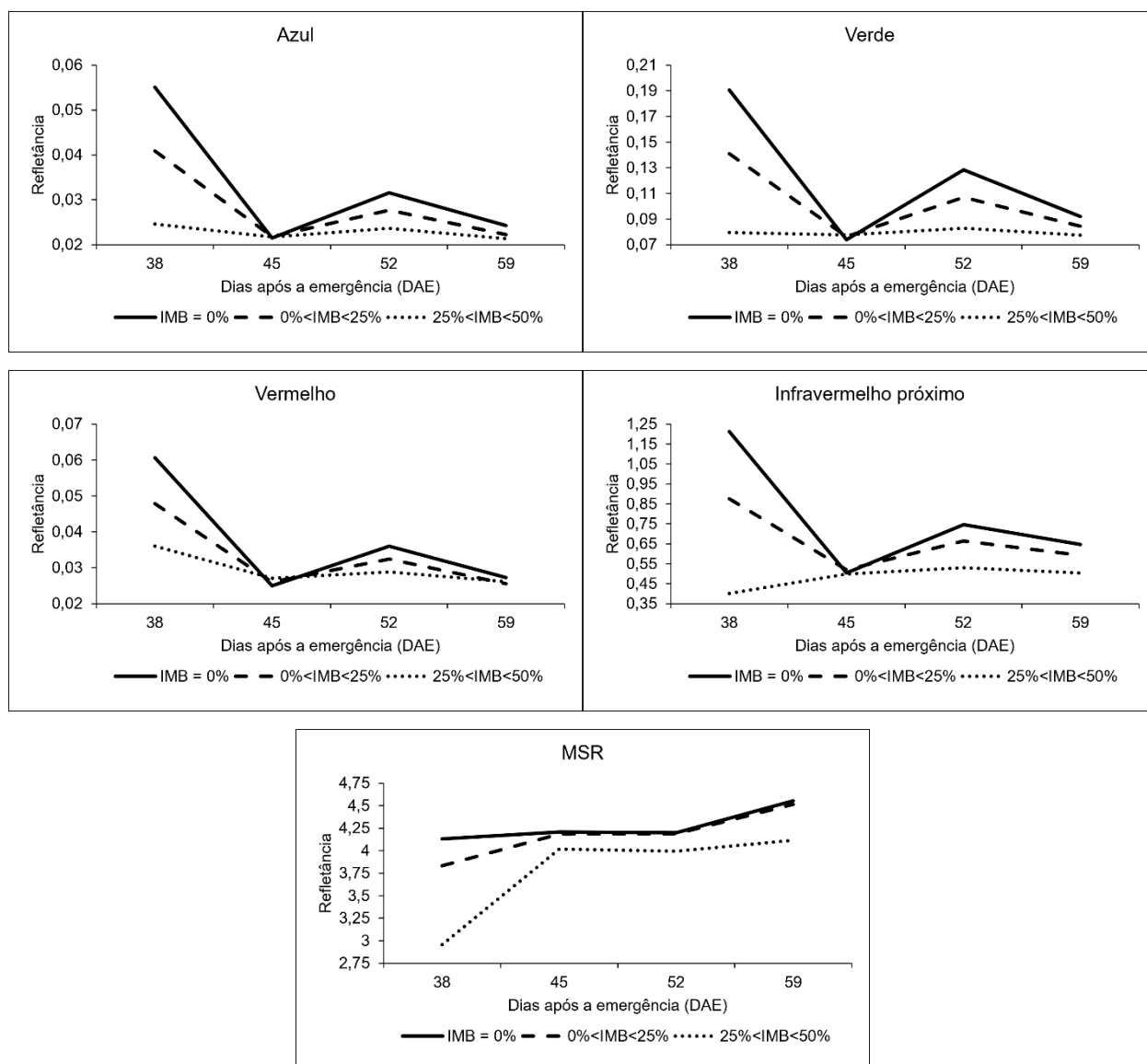
Fonte: elaborado pelo próprio autor.

Figura 2.19 – Médias, dos três experimentos, da variação temporal das refletâncias do feijoeiro-comum das bandas azul, verde, vermelho e infravermelho próximo e dos índices de vegetação EVI e MSR em diferentes incidências do mofo-branco (IMB).



Fonte: elaborado pelo próprio autor.

Figura 2.20 – Médias, dos três experimentos, da variação temporal das refletâncias do feijoeiro-comum das bandas azul, verde, vermelho, infravermelho próximo e do índice de vegetação MSR, sob diferentes condições de índice de severidade do mofo-branco (ISMB).



Fonte: elaborado pelo próprio autor.

2.5. DISCUSSÃO

Este trabalho trouxe uma nova abordagem para a avaliação de características que se relacionam com a ocorrência do mofo-branco no feijoeiro-comum por meio de dados multiespectrais adquiridos por uma RPA em nível de análise da área de produção,

juntamente com dados de espaçamento e densidade de plantio. Outros trabalhos, como o de Boechat et al. (2014) e Machado et al. (2015a) usaram informações hiperespectrais a nível de parcela, no entanto, usando espectrorradiômetro. A vantagem no uso de RPAs em relação ao espectrorradiômetro é sua maior capacidade operacional, facilidade na operação e menor custo.

As informações extraídas dos dados, expressas pelas bandas e índices de vegetação foram aplicadas nas modelagens por meio dos algoritmos. Os resultados indicaram que os algoritmos selecionaram as melhores covariáveis entre as usadas como covariáveis de entrada, revelando uma melhor data e índice de vegetação para cada resposta específica. Logo, para cada resposta específica, como incidência e índice de severidade do mofo-branco, e produtividade do feijoeiro, as modelagens forneceu as covariáveis mais importantes na previsão de ocorrência do mofo-branco e da produtividade.

Pesquisando na base do Google Scholar pelos indexadores "*white mold*", "*Sclerotinia sclerotiorum*", "*Phaseolus vulgaris*" e "*remote sensing*", encontramos 68 resultados, porém um número pouco expressivo se relaciona com a aplicação de sensoriamento remoto no manejo do mofo-branco do feijoeiro-comum. Nenhum desses cinco estudos usaram câmera multiespectral e diferentes índices de vegetação aplicados em aprendizagem de máquinas na investigação da interação do mofo-branco no feijoeiro. Para reforçar essa informação, pesquisamos estudos na base do Google Scholar usando os indexadores "*modeling*", "*phaseolus vulgaris*", "*Sclerotinia sclerotiorum*" e "spectral data". Entre os resultados, apareceu o trabalho de Pozza, Alves e Sanches (2022). Estes autores usaram índices de vegetação em sementes de feijão na identificação de patógenos, e não trabalharam com MB. Logo, há demanda por mais pesquisas usando SR por meio de imagens adquiridas por RPA no manejo da detecção do MB.

Diante da carência na aplicação de métodos que usaram imagens obtidas por câmeras multiespectrais acopladas à RPA combinadas com técnicas de aprendizagem de máquinas, buscou-se, neste trabalho, analisar diferentes índices de vegetação e seu potencial na predição das características do feijão atacado por MB. Uma razão para aplicar dados espectrais advindos de sensores multiespectrais em vários índices de vegetação é o elevado custo dos sensores hiperespectrais (Pandal et al., 2020). Por isso,

o objetivo no presente estudo foi buscar soluções de baixo custo envolvendo a aplicação do sensoriamento remoto para auxiliar os produtores na detecção do MB no feijoeiro-comum. Este trabalho procurou verificar a eficácia e potencial dos modelos aqui aplicados. Tão importante quanto propor novos modelos, foi nossa intenção mostrar as limitações para que se possa construir abordagens mais efetivas e aplicáveis ao campo.

Quanto ao processo de modelagem realizado utilizando algoritmos de aprendizagem de máquinas, uma análise importante que é trazida por este estudo é o uso do método da validação cruzada repetida (VCR). Kim (2009) conclui que a eficiência da VCR é superior a validação cruzada tradicional, visto que reduz a variabilidade do estimador. Como avaliamos 93 parcelas, o estudo pode ser caracterizado como um conjunto de dados reduzidos. De acordo com Eertink et al. (2022), a VCR é a forma de análise mais indicada para essas situações.

Para a previsão da IMB, o modelo que apresentou melhor desempenho foi o *Cubist*. Anderegg et al. (2019) também verificaram que o *Cubist* foi o melhor modelo quando aplicaram técnicas de sensoriamento remoto junto com aprendizagem de máquinas para prever a severidade de *Septoria tritici* em trigo. A modelagem da incidência do MB pelo algoritmo *Cubist* apresentou o maior valor de LCCC (LCCC igual a 0,4540), e o menor MAE, de 10,58%. Apesar de esses resultados serem modestos, é importante entender como o modelo identificou a incidência do MB. Essa importância está relacionada ao fato de a IMB ser uma medida envolvida no cálculo do índice de severidade do MB, conforme a Equação 2.1. Consequentemente, a incidência está associada também, além da severidade do MB, à densidade de plantio e à produtividade (Vieira et al., 2012). Dessa forma, verifica-se que houve uma relação causal de densidade e severidade de MB.

O *Cubist* teve o melhor desempenho na predição da IMB, enquanto *Random Forest* (RF), *Support Vector Machine* e o *Kernel-K-Nearest Neighbors* (KNN) tiveram desempenho inferior. O teste de Kruskal-Wallis sugeriu que o R^2 para o KNN e o *Cubist* foram equivalentes, assim como para o RMSE do *Cubist*, KNN e RF. O melhor desempenho do algoritmo *Cubist* possivelmente deveu-se ao fato de esse algoritmo possuir boa capacidade de encontrar padrões em relações não lineares complexas. Um fator que pode ter contribuído para o baixo desempenho do modelo, ao prever a

incidência do MB usando dados espectrais, foi a baixa incidência da doença observada nos três experimentos, com média de 19,88% de IMB no conjunto de dados. Harikrishnan e Rio (2008) propuseram um modelo de previsão de risco do MB em feijoeiro. Eles observaram que o modelo preditivo apresentava maior precisão quando a média de IMB era maior que 20%.

O algoritmo Cubist também foi o que apresentou o melhor desempenho na predição do ISMB, com valores mais elevados de R^2 e LCCC, e os menores valores de RMSE e MAE. No entanto, os testes de Kruskal-Wallis mostraram que, para a predição do ISMB, não há diferenças significativas entre os algoritmos. Logo, pode-se usar outro algoritmo, uma vez que o Cubist apresenta o maior custo entre os algoritmos aqui avaliados, conforme verificado por Liu et al. (2022). Como o algoritmo KNN é uma extensão do KNN convencional, ele apresenta alto custo computacional, pois seu desempenho depende do número de dimensões (Liu et al., 2021). Resque et al. (2019) verificaram que o custo computacional do SVM é mais alto ao do RF. Por isso, este último algoritmo é mais adequado por exigir menor custo computacional e ter apresentado parâmetros de desempenho equivalentes ao algoritmo KNN.

O algoritmo que se destacou na predição da produtividade foi o KNN, visto que apresentou o maior valor de LCCC e menores valores de MAE e RMSE. Ao contrário do que foi observado nas predições do IMB e do ISMB, o algoritmo Cubist apresentou o pior desempenho na predição da produtividade. Os maiores valores das métricas na predição da produtividade, comparados às predições da IMB e do ISMB, pode ter ocorrido pelo fato de os índices de vegetação estarem relacionados com o vigor das plantas e a quantidade de biomassa da cultura. Assim, de forma indireta, a produtividade pode estar mais correlacionada com esses índices do que a presença do mofo-branco. Trabalhos como os de Silva et al. (2020) e Shammi et al. (2024) em soja, Ramos et al. (2020) em milho e Badagliacca et al. (2023) em trigo, também mostraram bom desempenho na predição da produtividade usando câmera multiespectral acopladas a RPA.

As covariáveis selecionadas pelos algoritmos para predição IMB, ISMB e produtividade foram diferentes. Esses resultados indicam que a resposta dos índices de vegetação à lavoura varia de acordo com a variável de saída. A seleção de variáveis é etapa importante na aplicação dos algoritmos de aprendizagem de máquinas, visto que

considera apenas as covariáveis mais importantes para a modelagem, e exclui as que possam prejudicar o desempenho dos algoritmos. Logo, entre as vantagens da aplicação da seleção de covariáveis estão a redução do *overfitting*, redução do tempo de treinamento e melhoria da precisão do modelo (Herrera et al., 2015). No estudo desses autores, 42 covariáveis foram selecionadas para todas as variáveis de saída. Entretanto, cada algoritmo atribuiu importância de forma peculiar à contribuição de cada covariável em seu modelo.

Na predição da IMB, o algoritmo Cubist selecionou 42 covariáveis importantes para a modelagem, destas, duas covariáveis foram o espaçamento e a densidade de plantio, entretanto elas apresentaram baixa significância para a modelagem. Das 40 covariáveis restantes têm-se os índices espectrais, em que das 224 informações espectrais de entrada (bandas e índices), foram selecionadas apenas 19 informações espectrais (14 índices e 5 bandas da câmera multiespectral). Dos 14 índices de vegetação, apenas oito possuem alguma banda NIR ou RE, além da RGB. Isso evidencia o potencial que câmeras RGB possuem na estimativa da incidência do MB, além de ter bons resultados, é de menor custo comparado ao de uma câmera multiespectral. Ratificando o potencial das câmeras RGB em estimar a IMB, a Figura 2.4 mostra que as seis primeiras covariáveis selecionadas foram a banda G aos 52 DAE, MSR aos 38 DAE, banda G aos 38, EVI aos 38 DAE, banda G aos 59 DAE e banda B aos 38 DAE, com contribuição de 100%, 84,71%, 79,23%, 76,09%, 75,07% e 74,82%, respectivamente. Outro ponto a ser notado é que dentre essas seis covariáveis mais importantes, quatro são informações espectrais do 38 DAE, demonstrando que o modelo de predição tem potencial em prever a ocorrência da IMB.

Resultados encontrando a banda G como um indicativo de infecção do MB é muito importante de modo que possam ser usados sensores mais simples e baratos no manejo da doença. Além disso, contribui ainda para que novas alternativas aos fungicidas convencionais possam ser mais usadas, como a aplicação de fosfitos de Zn e Cu na inibição micelial do mofo-branco no feijoeiro-comum (Fagundes-Nacarath et al., 2018a). Uma observação importante é que a banda G aos 52 DAE foi mais importante (100%), no entanto, na prática, aos 52 DAE a lavoura já estava na segunda semana após a floração, já podendo apresentar a infecção pelo MB. Além disso, duas semanas após

iniciada a fase de florescimento, um técnico capacitado já consegue identificar o murchamento na lavoura em consequência da infecção do MB. A banda G aos 38 DAE é uma covariável com maior potencial de previsibilidade, visto que sua informação está na iminência e entrando na fase de florescimento. Dessa forma, possui maior viabilidade prática em seu uso para prever a ocorrência da infecção pelo MB.

O uso do Cubist na modelagem do ISMB teve comportamento semelhante à modelagem na IMB, selecionando o mesmo número de covariáveis importantes, e o mesmo número e índices de vegetação que não usam apenas o espectro RGB em suas equações (CI_Green, CCCI, EVI, CVI, MCARI, GDVI, Chl e MSR). As covariáveis que mais contribuíram para a modelagem do ISMB foram a banda G aos 38 DAE (100%), banda G aos 59 DAE (94,04%), MSR aos 38 DAE (81,04%), banda G aos 52 DAE (77,98%) e banda B aos 38 DAE (70,40%). Assim como a predição do IMB, a forte influência do espectro RGB sendo representado por quatro das cinco mais importantes na modelagem. É sabido que a IMB influencia na ISMB, dessa forma, a seleção da banda G por ambas as modelagens comprova o quão forte é esse comportamento.

As seis covariáveis mais importantes na predição da produtividade foram a banda G aos 38 DAE, CI aos 38 DAE, banda G aos 52 DAE, banda B aos 38 DAE, CVI aos 52 DAE, GDVI aos 59 DAE e banda G aos 59 DAE, com 100%, 99,50%, 90,55%, 83,07%, 79,57%, 77,23% e 75,96%, respectivamente, de contribuição para o modelo. Predomina as informações espectrais na paleta do RGB, representando cinco das sete mais importantes, considerando as covariáveis com importância maior que 75%. A banda G se destaca com a maior importância a partir da primeira data de voo, aos 38 DAE. Isso traz grandes benefícios, considerando que onde há tendência de maior produtividade, pode ocorrer menor presença de IMB e ISMB.

O comportamento temporal da assinatura espectral das bandas azul (B) e verde (G), mostradas nas Figuras 2.19 e 2.20, apresentaram-se semelhantes. Observa-se que a refletância é maior em condições ausentes da doença, diminuindo com o aumento da IMB e ISMB. As datas de 38, 45 e 59 DAE na Figura 2.19 evidenciam a importância das bandas B e G na predição do IMB, visto suas notáveis diferenças entre as pressões de IMB. A diferença dos valores de refletância e do índice MSR entre as condições de 0% e 25%, e 25% e 50%, foram maiores para o ISMB. Esse comportamento das bandas B e

G, e do MSR na predição da ISMB pode explicar a linearidade da diminuição da produtividade com o aumento da severidade. Isso evidencia que o ISMB compromete mais a produtividade do que a IMB. Nas Figuras 2.19 e 2.20, o ponto mínimo no valor das bandas e dos índices aos 45 DAE foi devido a coleta de dados ter sido feito em condições nubladas, resultando em menores valores de refletância.

No que tange a observações sobre possíveis limitações deste estudo, por mais que diversos estudos tenham mostrado bons resultados na predição da ocorrência do MB no feijoeiro, é importante se levar em conta que muitos destes trabalhos foram realizados em laboratório. Claramente os trabalhos em laboratório obtiveram sucesso, no entanto, a dinâmica da doença muda dentro de uma mesma lavoura. Dessa forma, a interação entre o MB, o feijoeiro e o ambiente em campo ditarão quão severo será o ataque do MB na lavoura, ou seja, seu comportamento *in loco*.

No que se refere ao vértice ambiental do triângulo da doença, o ambiente não se resume apenas na variação do clima. O ambiente é um fator complexo que representa, além da atuação do clima, a ação do solo e a dinâmica dos fenômenos interligados, como a dinâmica da água. Como exemplo disso, há o trabalho de Clarkson et al. (2004) com a cultura da alface, em que é necessário o potencial de água no solo máximo de -100 kPa de uma a duas semanas para que ocorra a germinação dos apotécios. Ainda, segundo Clarkson et al. (2004), juntamente a essa condição de potencial hídrico do solo, é necessário que o clima apresente temperaturas de 5 a 25 °C.

Dessa forma, por mais que os resultados aqui pareçam modestos, este trabalho, assim como os demais trabalhos já realizados, é de elevada importância para que se possa continuar a buscar como identificar com mais eficácia, em campo, a dinâmica dos fatores que envolvem o ataque do MB no feijoeiro. Os resultados deste trabalho na modelagem para a predição da IMB e do ISMB revelam que ainda é pouco eficaz apenas o uso de informações espectrais. Esse argumento é ratificado pelos resultados dos trabalhos apresentados por Clarkson et al. (2004) e Shah, Dillard e Pethybridge (2019).

Os modestos resultados dos modelos de predição da IMB e ISMB provavelmente se devem à resolução temporal usada na coleta das imagens multiespectrais. Como mencionado, nesse estudo os voos para a aquisição das informações espectrais foram realizados semanalmente, o que pode ter afetado o desempenho do modelo.

Corroborando com essa hipótese, Watts et al. (2011) verificaram que aumentar o número de coletas espectrais pode melhorar o desempenho do modelo. Outro fato que corrobora com a possibilidade de os modelos terem apresentados modestos resultados está na própria dinâmica de ataque do MB. Por exemplo, uma vez que o MB infecta uma flor, o fungo pode infectar os órgãos adjacentes, como vagens e caule, dentre 2 a 3 dias (Steadman, 1983). Dessa forma, uma estratégia a ser pensada é a coleta diária da coleta espectral, inicialmente, entre o período proposto neste trabalho, ou seja, uma semana antes da entrada no estágio vegetativo da floração e duas semanas após o estágio da floração. Fagundes-Nacarath et al. (2018b) verificaram sinais de dessecação no feijoeiro-comum após 60 horas de ocorrer a infecção do MB. Dessa forma, intervalos de voo menores que 60 horas são fundamentais para observar uma possível variação espectral como resposta ao início da infecção do MB. Ainda sobre o baixo desempenho dos modelos ao predizer a IMB e o ISMB, isso pode ter ocorrido com o método do ISMB usado, em que a escala de 0 a 4 pode levar a um grande erro na estimativa da severidade em campo.

Sob as condições e resultados deste trabalho, algumas observações trouxeram algumas propostas para futuros trabalhos relacionados à infecção do MB. Entre as propostas está a criação de um modelo de escoamento e molhamento superficial de água na folha e correlacioná-lo com a informação espectral com os diferentes níveis de escoamento e molhamento em campo. Dessa maneira, ao se coletar informações espectrais ter-se-á valores correlacionados para poder estimar com maior eficácia condições potenciais para o desenvolvimento e início de infecção do MB por meio dos ascósporos.

Os algoritmos implementados para predição da produtividade apresentaram melhor desempenho do que os utilizados para predizer o IMB e o ISMB, o que é importante para auxiliar os produtores nas estratégias diante ao mercado. Uma perspectiva importante que há para que melhorem a previsibilidade da ocorrência do MB, é a redução da poluição ambiental e menor exposição do trabalhador rural aos produtos fitossanitários. Além disso, o custo com tratamento fitossanitário poderia ser menor, promovendo menores custos com consequente aumento da lucratividade da produção.

2.6. CONCLUSÃO

Neste trabalho foram explorados quatro modelos de aprendizagem de máquina na previsão do comportamento de três variáveis da cultura do feijoeiro-comum, sendo a incidência do mofo-branco (IMB), índice de severidade do mofo-branco (ISMB) e a produtividade. Na previsão do comportamento dessas três variáveis observou-se que os algoritmos mais adequados para a modelagem das respectivas previsões foram o *Cubist* e o KKNn.

Para a previsão da IMB pelo algoritmo *Cubist*, as covariáveis mais importantes em sua predição foram banda G aos 52 DAE, MSR aos 38 DAE, banda G aos 38 DAE, EVI aos 38 DAE, banda G aos 59 DAE e banda B aos 38 DAE. Para o ISMB as covariáveis mais importantes foram banda G aos 38 DAE, banda G aos 59 DAE, MSR aos 38 DAE, banda G aos 52 DAE e banda B aos 38 DAE, também pela modelagem do algoritmo *Cubist*. Por fim, a modelagem da previsão da produtividade pelo algoritmo KKNn teve como covariáveis mais importantes a banda G aos 38 DAE, CI aos 38 DAE, banda G aos 52 DAE, banda B aos 38 DAE, CVI aos 52 DAE, GDVI aos 59 DAE e banda G aos 59 DAE. O algoritmo RF foi semelhante ao KKNn na predição da produtividade. Portanto, o RF é a alternativa mais viável do ponto da perspectiva da economia computacional.

Considerando as condições que este trabalho foi realizado, a variável que apresentou melhor capacidade de predição em ambiente não controlado para o MB foi a produtividade. Para a produtividade os modelos apresentaram melhores desempenhos e maior potencial de previsibilidade em relação às outras variáveis de saída. Dessa forma, há a possibilidade de aplicação em campo na estimativa de produtividade pelos produtores de feijão-comum.

2.7. AGRADECIMENTOS

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

2.8. REFERÊNCIAS

ABDULRIDHA, J.; AMPATZIDIS, Y.; KAKARLA, S. C.; ROBERTS, P. Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. **Precision Agriculture**, v. 21, n. 5, p. 955-978, 2020.

ABDULRIDHA, J.; BATUMAN, O.; AMPATZIDIS, Y. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. **Remote Sensing**, v. 11, n. 11, p. 1373, 2019.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M. Köppen's climate classification map for Brazil. **Meteorologische zeitschrift**, v. 22, n. 6, p. 711-728, 2013.

ANDEREGG, J.; HUND, A.; KARISTO, P.; MIKABERIDZE, A. In-field detection and quantification of *Septoria tritici* blotch in diverse wheat germplasm using spectral-temporal features. **Frontiers in plant science**, v. 10, p. 1355, 2019.

ASHOURLOO, D.; MOBASHERI, M. R.; HUETE, A. Evaluating the effect of different wheat rust disease symptoms on vegetation indices using hyperspectral measurements. **Remote Sensing**, v. 6, n. 6, p. 5107-5123, 2014.

BADAGLIACCA, G.; MESSINA, G.; PRATICÒ, S.; LO PRESTI, E.; PREITI, G.; MONTI, M.; MODICA, G. Multispectral vegetation indices and machine learning approaches for durum wheat (*Triticum durum* Desf.) yield prediction across different varieties. **AgriEngineering**, v. 5, n. 4, p. 2032-2048, 2023.

BENDIG, J.; YU, K.; AASEN, H.; BOLTEN, A.; BENNERTZ, S.; BROSCHEIT, J.; GNYP, M. L.; BARETH, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. **International Journal of Applied Earth Observation and Geoinformation**, v. 39, p. 79-87, 2015.

BIRTH, G. S.; MCVEY, G. R. Measuring the color of growing turf with a reflectance spectrophotometer 1. **Agronomy Journal**, v. 60, n. 6, p. 640-643, 1968.

BOECHAT, L. T.; PINTO, F. A. C.; PAULA JÚNIOR, T. J.; QUEIROZ, D. M.; TEIXEIRA, H. Detecção do mofo-branco no feijoeiro, utilizando características espectrais. **Revista Ceres**, v. 61, n. 6, p. 907-915, 2014.

BOLAND, G. J.; HALL, R. Index of plants of hostes *Sclerotinia sclerotiorum*. **Canadian Journal of Plant Pathology**, v. 16, n.1, p. 93-108, 1994.

BOOCHS, F.; KUPFER, G.; DOCKTER, K.; KÜHBAUCH, W. Shape of the red edge as vitality indicator for plants. **Remote sensing**, v. 11, n. 10, p. 1741-1753, 1990.

BREIMAN, L. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.

BURGOS-ARTIZZU, X. P.; RIBEITO, A.; GUIJARRO, M.; PAJARES, G. Real-time image processing for crop/weed discrimination in maize fields. **Computers and Electronics in Agriculture**, v. 75, n. 2, p. 337-346, 2011.

CHEN, J. M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. **Canadian Journal of Remote Sensing**, v. 22, n. 3, p. 229-242, 1996.

CLARKSON, J. P.; PHELPS, K.; WHIPPS, J. M.; YOUNG, C. S.; SMITH, J. A.; WATLING, M. Forecasting Sclerotinia disease on lettuce: toward developing a prediction model for carpogenic germination of sclerotia. **Phytopathology**, v. 94, n. 3, p. 268-279, 2004.

CORTES, C; VAPNIK, V. Support-vector networks. **Machine learning**, v. 20, p. 273-297, 1995.

DASH, J.; CURRAN, P. J. The MERIS terrestrial chlorophyll index. **International Journal of Remote Sensing**, v. 25, p. 5403–5413, 2004.

DATT, B.; MCVICAR, T.R; VAN NIEL, T. G.; JUPP, D. L.; PEARLMAN, J.S. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, n. 6, p. 1246-1259, 2003.

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

EERTINK, J. J.; HEYMANS, M. W.; ZWEZERIJNEN, G. J.; ZIJLSTRA, J. M.; DE VET, H. C.; BOELLAARD, R. et al. External validation: a simulation study to compare cross-validation versus holdout or external testing to assess the performance of clinical prediction models using PET data from DLBCL patients. **EJNMMI research**, v. 12, n. 1, p. 58, 2022.

EL-SHIKHA, D. M.; BARNES, E. M.; CLARKE, T. R.; HUNSAKER, D. J.; HABERLAND, J. A.; PINTER, P. J.; WALLER, P. M.; THOMPSON, T. L. Remote sensing of cotton nitrogen status using the canopy chlorophyll content index (CCCI). **Transactions of the ASABE**, v. 51, n. 1, p. 73-82, 2008.

ESCADAFAL, R. Soil spectral properties and their relationships with environmental parameters-examples from arid regions. **Imaging spectrometry—A tool for environmental observations**, p. 71-87, 1994.

FAGUNDES-NACARATH, I. R. F.; DEBONA, D.; BRÁS, V. V.; SILVEIRA, P. R.; RODRIGUES, F. A. Phosphites attenuate Sclerotinia sclerotiorum-induced physiological impairments in common bean. **Acta Physiologiae Plantarum**, v. 40, n. 11, p. 198, 2018a.

FAGUNDES-NACARATH, I. R. F.; DEBONA, D.; RODRIGUES, F. A. Oxalic acid-mediated biochemical and physiological changes in the common bean-*Sclerotinia sclerotiorum* interaction. **Plant Physiology and Biochemistry**, v. 129, p. 109-121, 2018b.

FERREIRA, R. G.; SILVA, D. D.; ELESBON, A. A. A.; FERNANDES-FILHO, E. I.; VELOSO, G. V.; FRAGA, M. S.; FERREIRA, L. B. Machine learning models for streamflow regionalization in a tropical watershed. **Journal of Environmental Management**, v. 280, p. 111713, 2021.

FITZGERALD, G. J.; RODRIGUEZ, D.; CHRISTENSEN, L. K.; BELFORD, R.; SADRAS, V. O.; CLARKE, T. R. Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. **Precision Agriculture**, v. 7, p. 233-248. 2006.

GALVÍNCIO, J. D.; MOURA, M. S. B.; SILVA, T. G. F.; SILVA, B. B.; NAUE, C. R. LAI improved to dry forest in semiarid of the Brazil. **International Journal of Remote Sensing Applications**, v. 3, n. 4, p. 193-202, 2013.

GITELSON, A. A.; VIÑA, A.; CIGANDA, V.; RUNDQUIST, D. C.; ARKEBAUER, T. J. Remote estimation of canopy chlorophyll content in crops. **Geophysical research letters**, v. 32, n. 8, 2005.

GITELSON, A.; KAUFMAN, Y.; MERZLYAK, M. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sensing of Environment**, v.58, p.289-298, 1996.

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

GITELSON, A. A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. **Journal of plant physiology**, v. 161, n. 2, p. 165-173, 2004.

GÖRGEN, C.A.; HIKISHIMA, M.; NETO, A.N.S.; CARNEIRO, L.C.; JUNIOR, M.L. Mofo-branco (*Sclerotinia sclerotiorum*). In: ALMEIDA, A.M.R.; SEIXAS, C.D.S. Soja: Doenças Radiculares e de Hastes e Inter-relações com o Manejo do Solo e da Cultura. 1 ed. Londrina: **Embrapa Soja**, p.73-104, 2010.

GÖRGEN, C. A.; SILVEIRA NETO, A. N.; CARNEIRO, L. C.; RAGAGNIN, V.; LOBO JÚNIOR, M. Controle do mofo-branco com palhada e *Trichoderma harzianum* 1306 em soja. **Pesquisa Agropecuária Brasileira**, v. 44, p. 1583-1590, 2009.

GUIJARRO, M.; PAJARES, G.; RIOMOROS, I.; HERRERA, P. J.; BURGOS-ARTIZZU, X. P.; RIBEIRO, A. Automatic segmentation of relevant textures in agricultural images. **Computers and Electronics in Agriculture**, v. 75, n. 1, p. 75-83, 2011.

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

HARIKRISHNAN, R.; DEL RÍO, L. E. A logistic regression model for predicting risk of white mold incidence on dry bean in North Dakota. **Plant Disease**, v. 92, n. 1, p. 42-46, 2008.

HERRERA, L. J.; LAFUENTE, V.; GHINEA, R.; PEREZ, M. M.; NEGUERUELA, I.; POMARES, H.; ROJAS, I.; GUILLHÉN, A. Mutual information-based feature selection in spectrometric data for agriculture applications. In: **Proceedings of the International MultiConference of Engineers and Computer Scientists**. 2015.

HIND, T. L.; ASH, G. J.; MURRAY, G. M. Prevalence of sclerotinia stem rot of canola in New South Wales. **Australian Journal of Experimental Agriculture**, v. 43, n. 2, p. 163-168, 2003.

HOTHORN, T.; FRICK, H.; KOOK, L. CRAN task view: Machine learning & statistical learning. 2025.

HUETE, A. R.; LIU, H.; BATCHILY, K.; VAN LEEUWEN, W. J. D. A. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote sensing of environment**, v. 59, n. 3, p. 440-451, 1997.

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

KATAOKA, T.; KANEKO, T.; OKAMOTO, H.; HATA, S. I. Crop growth estimation system using machine vision. In: **Proceedings 2003 IEEE/ASME international conference on advanced intelligent mechatronics (AIM 2003)**. IEEE, 2003. p. b1079-b1083 vol. 2.

KIM, J. H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. **Computational statistics & data analysis**, v. 53, n. 11, p. 3735-3745, 2009.

KUHN, M.; WESTON, S.; KEEFER, C.; COULTER, N. Cubist models for regression. **R package Vignette R package version 0.0**, v. 18, p. 480, 2012.

KUHN, M.; WING, J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; TEAM, R. C.; BENESTY, M.; LESCARBEAU, R.; ZIEM, A.; SCRUCICA, L.; TANG, Y.; CANDAN, C.; HUNT, T. Package 'caret'. **The R Journal**, v. 223, n. 7, p. 48, 2020.

KUHN, M.; JOHNSON, K. **Applied predictive modeling**. New York: Springer, 2013.

LEITE, R. M. V. B. C. Ocorrência de doenças causadas por *Sclerotinia sclerotiorum* em girassol e soja. **Embrapa Soja. Comunicado Técnico 76**, 2005.

LAWRENCE, I.; LIN, K. A concordance correlation coefficient to evaluate reproducibility. **Biometrics**, p. 255-268, 1989.

LIU, B.; VU-BAC, N.; ZHUANG, X.; FU, X.; RABCZUK, T. Stochastic integrated machine learning based multiscale approach for the prediction of the thermal conductivity in carbon nanotube reinforced polymeric composites. **Composites Science and Technology**, v. 224, p. 109425, 2022.

LIU, C.; CICHON, A.; KRÓLCZYK, G.; LI, Z. Technology development and commercial applications of industrial fault diagnosis system: a review. **The International Journal of Advanced Manufacturing Technology**, v. 118, n. 11, p. 3497-3529, 2022.

LOUHAICHI, M.; BORMAN, M. M.; JOHNSON, D. E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, v. 16, n. 1, p. 65-70, 2001.

MACHADO, M. L.; PINTO, F. A.; QUEIROZ, D. M.; PAULA JÚNIOR, T. J.; VIEIRA, R. F. Estimativa de severidade do mofo-branco em lavouras de feijão utilizando-se sensores hiper e multiespectral. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 5, p. 426-432, 2015a.

MACHADO, M. L.; PINTO, F. A. C.; PAULA, T. J.; QUEIROZ, D. M.; CERQUEIRA, O. A. T. White mold detection in common beans through leaf reflectance spectroscopy. **Engenharia Agrícola**, v. 35, n. 6, p. 1117-1126, 2015.

MADEIRA, J.; BEDIDI, A.; CERVELLE, B.; POUGET, M.; FLAY, N. Visible spectrometric indices of hematite (Hm) and goethite (Gt) content in lateritic soils: the application of a Thematic Mapper (TM) image for soil-mapping in Brasília, Brazil. **International Journal of Remote Sensing**, v. 18, n. 13, p. 2835-2852, 1997.

MAO, W.; WANG, Y.; WANG, Y. Real-time detection of between-row weeds using machine vision. In: **2003 ASAE Annual Meeting**. American Society of Agricultural and Biological Engineers, 2003. p. 1.

MARCHANT, J. A.; ONYANGO, C. M. Shadow-invariant classification for scenes illuminated by daylight. **Journal of the Optical Society of America A**, v. 17, n. 11, p. 1952-1961, 2000.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International journal of remote sensing**, v. 17, n. 7, p. 1425-1432, 1996.

MCKINNON, T.; HOFF, P. Comparing RGB-based vegetation indices with NDVI for drone based agricultural sensing. **Agribotix. com**, v. 21, n. 17, p. 1-8, 2017.

MELLO, D. C.; VELOSO, G. V.; LANA, M. G. D.; MELLO, F. A. D. O.; POPPIEL, R. R.; CABRERO, D. R. O.; DI RAIMO, L. A. D. L.; SCHAEFER, C. E. G. R.; FERNANDES FILHO, E. I.; LEITE, E. P.; DEMATTÊ, J. A. M. A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. **Geoscientific Model Development**, v. 15, n. 3, p. 1219-1246, 2022.

MEYER, G. E.; HINDMAN, T. W.; LAKSMI, K. Machine vision detection parameters for plant species identification. In: **Precision agriculture and biological quality**. SPIE, 1999. p. 327-335.

NASSER, L. C. B.; SUTTON, J. C.; BOLAND, G. J.; JAMES, T. W. Influence of crop residues and soil moisture on *Sclerotinia sclerotiorum* from the Cerrados region in Brazil. **Canadian Journal of Plant Pathology**, v. 17, p. 353-365, 1995.

NEGRISOLI, M. M.; NEGRISOLI, R.; SILVA, F.; LOPES, L. S.; SOUZA JÚNIOR, F. S.; VELINI, E. D.; CARBONARI, C. A.; RODRIGUES, S. A.; RAETANO, C. G. Soybean rust detection and disease severity classification by remote sensing. **Agronomy Journal**, v. 114, n. 6, p. 3246-3262, 2022.

PALLATHADKA, H.; RAVIPATI, P.; SAJJA, G. S.; PHASINAM, KASSANUK, T.; SANCHEZ, D. T.; PRABHU, P. Application of machine learning techniques in rice leaf disease detection. **Materials today: proceedings**, v. 51, p. 2277-2280, 2022.

PANDAY, U. S.; PRATIHAST, A. K.; ARYAL, J.; KAYASTHA, R. B. A review on drone-based data solutions for cereal crops. **Drones**, v. 4, n. 3, p. 41, 2020.

PARVANDEH, S.; MCKINNEY, B. A. EpistasisRank and EpistasisKatz: interaction network centrality methods that integrate prior knowledge networks. **Bioinformatics**, v. 35, n. 13, p. 2329-2331, 2019.

PAULA JÚNIOR, T. J.; MORANDI, M. A. B.; LOBO JUNIOR, M.; VIEIRA, R. F.; ZAMBOLIM, L. Controle alternativo do mofo-branco do feijoeiro. **Embrapa Meio Ambiente-Capítulo em livro científico (ALICE)**, 2006.

PEARSON, R. L.; MILLER, L. D. Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. **Remote sensing of environment**, VIII, p. 1355, 1972.

PONTI, M. P. Segmentation of low-cost remote sensing images combining vegetation indices and mean shift. **IEEE Geoscience and Remote Sensing Letters**, v. 10, n. 1, p. 67-70, 2012.

POZZA, E. A.; ALVES, M. C.; SANCHES, L. Using computer vision to identify seed-borne fungi and other targets associated with common bean seeds based on red-green-blue spectral data. **Tropical Plant Pathology**, p. 1-18, 2022.

QI, J.; CHEHBOUNI, A.; HUETE, A. R.; KERR, Y. H.; SOROOSHIAN, S. A modified soil adjusted vegetation index. **Remote sensing of environment**, v. 48, n. 2, p. 119-126, 1994.

QUINLAN, J. R. Learning with continuous classes. In: **5th Australian joint conference on artificial intelligence**. 1992. p. 343-348.

RAMOS, A. P. M.; OSCO, L. P.; FURUYA, D. E. G.; GONÇALVES, W. N.; SANTANA, D. C.; TEODORO, L. P. R.; SILVA JÚNIOR, C. A.; CAPRISTO-SILVA, G. F.; LI, J.; BAIO, F. H. R.; MARCATO JÚNIOR, J.; TEODORO, P. E.; PISTORI, H. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. **Computers and Electronics in Agriculture**, v. 178, p. 105791, 2020.

RedEdge-MX Integration Guide. Micasense, 2017. Disponível em: <<https://support.micasense.com/hc/en-us/articles/360011389334-RedEdge-MX-Integration-Guide>>. Acesso em: 05 de janeiro de 2024.

RESQUE, P.; BARROS, A.; ROSÁRIO, D.; CERQUEIRA, E. An investigation of different machine learning approaches for epileptic seizure detection. In: **2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)**. IEEE, 2019. p. 301-306.

REUDENBACH, C.; MEYER, H.; DETSCH, F.; MÖLLER, F.; NAUSS, T.; OPGENOORTH, L.; MARBURG, E. I. uavRst: Unmanned Aerial Vehicle R Tools. 2021.

ROUSE, W. Monitoring vegetation system in the great plain with ERTS. In: **3rd ERTS symposium, NASA Washington DC, 1973**. 1973. p. 309-317.

SHAHOTEISI, F.; MANESH, M. R.; MENDOZA, L. E. D. R. Modeling risk of Sclerotinia sclerotiorum-induced disease development on canola and dry bean using machine learning algorithms. **Scientific Reports**, v. 12, n. 1, p. 864, 2022.

SHAMMI, S. A.; HUANG, Y.; FENG, G.; TEWOLDE, H.; ZHANG, X.; JENKINS, J.; SHANKLE, M. Application of UAV multispectral imaging to monitor soybean growth with yield prediction through machine learning. **Agronomy**, v. 14, n. 4, p. 672, 2024.

SILVA, E. E.; BAIO, F. H. R.; TEODORO, L. P. R.; SILVA JÚNIOR, C. A.; 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**, v. 18, p. 100318, 2020.

SINGH, S. P.; SCHWARTZ, H. F. Breeding common bean for resistance to diseases: a review. **Crop Science**, v. 50, n. 6, p. 2199-2223, 2010.

SRIPADA, R. P.; HEINIGER, R. W.; WHITE, J. G.; WEISZ, R. Aerial color infrared photography for determining late-season nitrogen requirements in corn. **Agronomy Journal**, v. 97, n. 5, p. 1443-1451, 2005.

STEADMAN, J. R. White mold--a serious yield-limiting disease of bean. **Plant Disease**, v. 67, n. 4, p. 346-350, 1983.

TSOUROS, D. C.; BIBI, S.; SARIGIANNIDIS, P. G. A review on UAV-based applications for precision agriculture. **Information**, v. 10, n. 11, p. 349, 2019.

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

TUCKER, C. J.; ELGIN, J. H.; MCMURTREY III, J. E.; FAN, C. J. Monitoring corn and soybean crop development with hand-held radiometer spectral data. **Remote Sensing of Environment**, v. 8, n. 3, p. 237-248, 1979.

VARMA, S.; SIMON, R. Bias in error estimation when using cross-validation for model selection. **BMC bioinformatics**, v. 7, n. 1, p. 91, 2006.

VIEIRA, R. F.; PAULA JÚNIOR, T. J.; CARNEIRO, J. E. S.; TEIXEIRA, H.; QUEIROZ, T. F. N. Management of white mold in type III common bean with plant spacing and fungicide. **Tropical Plant Pathology**, v. 37, p. 91-101, 2012.

VIÑA, A.; GITELSON, A. A.; RUNDQUIST, D. C.; KEYDAN, G.; LEAVITT, B.; SCHEPERS, J. Monitoring maize (*Zea mays* L.) phenology with remote sensing. **Agronomy Journal**, v. 96, n. 4, p. 1139-1147, 2004.

WATTS, J. D.; POWELL, S. L.; LAWRENCE, R. L.; HILKER, T. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. **Remote Sensing of Environment**, v. 115, n. 1, p. 66-75, 2011.

WOEBBECKE, D. M.; MEYER, G. E.; VON BARGEN, K.; MORTENSEN, D. A. Shape features for identifying young weeds using image analysis. Transactions of the ASAE (American Society of Agricultural Engineers), v. 38, n. 1, p. 271-281, 1995.

WOEBBECKE, D. M.; MEYER, G. E.; VON BARGEN, K.; MORTENSEN, D. A. Color indices for weed identification under various soil, residue, and lighting conditions. **Transactions of the ASAE**, v. 38, n. 1, p. 259-269, 1995.

YU, K.; JI, L.; ZHANG, X.. Kernel nearest-neighbor algorithm. **Neural Processing Letters**, v. 15, p. 147-156, 2002.

ZARCO-TEJADA, P. J.; GONZÁLEZ-DUGO, V.; WILLIAMS, L. E.; SUÁREZ, L.; BERNI, J. A.; GOLDHAMER, D.; FERRERES, E. A PRI-based water stress index combining structural and chlorophyll effects: Assessment using diurnal narrow-band airborne imagery and the CWSI thermal index. **Remote sensing of environment**, v. 138, p. 38-50, 2013.

ZHANG, C.; LANE, B.; FERNÁNDEZ-CAMPOS, M.; CRUZ-SANCAN, A.; LEE, D.; GONGORA-CANUL, C.; ROSS, T. J.; SILVA, C. R.; TELENKO, D. E. P. GOODWIN, S. B.; SCOFIELD, S. R.; OH, S.; JUNG, J.; CRUZ, C. D. Monitoring tar spot disease

in corn at different canopy and temporal levels using aerial multispectral imaging and machine learning. **Frontiers in plant science**, v. 13, p. 1077403, 2023.

ZHANG, N.; YANG, G.; ZHAO, C.; ZHANG, J.; YANG, X.; PAN, Y.; HUANG, W.; XU, B.; LI, M.; ZHU, X.; LI, Z. Progress and prospects of hyperspectral remote sensing technology for crop diseases and pests. **National remote sensing bulletin**, v. 25, n. 1, p. 403-422, 2021.

CAPÍTULO 3 - DESENVOLVIMENTO DE MODELO DE PREDIÇÃO DA SEVERIDADE DO MOFO-BRANCO EM FEIJOEIRO-COMUM COM BASE EM DADOS CLIMÁTICOS UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINAS

Resumo

O feijão pode ser cultivado o ano todo, sendo dividido em três safras: primeira, segunda e terceira. A terceira safra, conhecida como feijão de outono-inverno ou feijão irrigado, requer irrigação, e devido à alta umidade do solo e temperaturas amenas, favorece doenças fúngicas como o mofo-branco. Identificar e quantificar os fatores que promovem o mofo-branco é de elevada importância para controlá-lo. Sendo assim, o objetivo deste trabalho é desenvolver um modelo de *machine learning* para prever o risco da doença no feijoeiro-comum, utilizando dados climáticos de campo e produtividade. Para este trabalho foram usados dados coletados em cinco anos, sendo-os 2000, 2001, 2014, 2015 e 2017. Foram usados nos ensaios diferentes densidades de plantio e três cultivares. Foram usadas como variáveis temperaturas mínima, média e máxima, precipitação, índice de severidade do mofo-branco (ISMB) mínimo, médio e máximo. Na seleção de variáveis, aquelas que se mostraram mais importantes nas previsões foram ISMB, precipitação, cultivar e tratamento fungicida. Os algoritmos que apresentaram resultados viáveis à previsão do ISMB para cinco e três anos de experimentos foram o *Support Vector Machine* e *Linear Model*, respectivamente, e para a produtividade considerando três anos de experimentos o *Random Forest*. Sugere-se que os resultados modestos na generalização das modelagens das previsões podem ter ocorrido pelo tamanho restrito do banco de dados.

3.1. INTRODUÇÃO

O feijão (*Phaseolus vulgaris*) é um importante alimento, apresentando elevado consumo no Brasil e em outros países. Ratificando esse cenário, o Brasil é um dos maiores produtores mundiais desse produto. Dentre os estados brasileiros, Minas Gerais é o segundo maior produtor de feijão com 18,21% da produção, com produtividade média de 1.707 kg/ha na safra 2022/2023 (Conab, 2023).

O feijão é uma cultura que pode ser cultivada durante todo o ano. O cultivo do feijão é dividido em três safras, geralmente denominadas de primeira, segunda e terceira safra. Para o feijão de terceira safra, também denominado feijão de outono-inverno, a irrigação é indispensável (Leite, 2005), levando-o a ser conhecido também como feijão irrigado.

Os níveis mais elevados de teor de água do solo da terceira safra, advindos da irrigação, e as temperaturas amenas, entre 15 e 25°C (Paula Júnior et al., 2006) do outono-inverno, favorecem o desenvolvimento de doenças fúngicas, como o mofo-branco, causado pelo fungo *Sclerotinia sclerotiorum* (Lib.) de Bary. Possuindo um amplo espectro, os hospedeiros do fungo *S. sclerotiorum* apresentam-se em 75 famílias, 278 gêneros e 408 espécies (Boland; Hall, 1994). Geralmente, a infecção pelo *S. sclerotiorum* se inicia com o fechamento do dossel e a época de floração da lavoura (Canteri et al., 1999).

Os sintomas na planta, em consequência da infecção, apresentam-se na forma de lesões encharcadas nos órgãos atingidos (folhas, hastes e vagens), com coloração parda e consistência tenra e micélio branco de aspecto cotonoso, que dá o nome da doença de mofo-branco (Leite, 2005; Paula Júnior et al., 2006). A infestação pelo fungo leva à podridão das hastes e do caule comprometendo o desenvolvimento das plantas. Todos esses fatores levam à redução de produtividade e de qualidade.

A identificação dos fatores que favorecem o desenvolvimento do mofo-branco é de elevada importância, uma vez que podem facilitar o combate à doença por meio da identificação de quando esta pode surgir. A forma convencional de combate ao mofo-branco é realizar pelo menos duas aplicações preventivas de fungicida. A primeira aplicação ocorre no início da floração da lavoura, e a segunda, 10 dias após a primeira. Essa forma de produção pode se tornar mais sustentável se as condições que favorecem

o desenvolvimento do mofo-branco puderem ser identificadas de tal forma a possibilitar avaliar se é realmente necessário ou não a aplicação de fungicidas.

Uma das técnicas que vem sendo utilizadas para predição de doenças em cultura é a partir de modelos desenvolvidos com base em algoritmos de *machine learning*. Entre os algoritmos de *machine learning* utilizados estão os *Linear Model*, *Random Forest*, *Cubist*, *k-Nearest-Neighbours* e *Support Vector Machine*. Esses algoritmos foram utilizados nas culturas do dendê (Santoso; Tani; Wang, 2017), milho (Kusumo et al., 2018), soja (Morgan; Blank; Seetan, 2021) e mandioca (Mwebaze; Owomugisha, 2016) para previsão de doenças.

Avaliar o risco da ocorrência do mofo-branco por meio de um modelo de *machine learning*, desenvolvido com dados de campo permite identificar as condições da lavoura que favorecem a ocorrência da doença. Consequentemente, pode-se ter uma produção mais sustentável, visto que aplicações de fungicidas que são realizadas de forma preventiva na floração, para reduzir o mofo-branco, podem ser reduzidas ou evitadas.

Existe potencial para criar um modelo de *machine learning* que calcule o risco de mofo-branco, usando dados como densidade e espaçamento de plantio, condições climáticas, porcentagem de fechamento da cultura e altura das plantas. Shahoteisi et al. (2022), utilizaram a duração de molhamento, período seco e tempo acumulado na modelagem que realizaram. No entanto, eles observaram a incidência média do mofo-branco em experimentos de laboratório e não em condições de campo. Assim, Mota, Miquelluti e Ozaki (2020) sugerem que modelos com base em dados de campo têm maior eficácia na previsão da doença, sendo o intuito do presente estudo.

Diante do exposto, fica evidente a necessidade de desenvolvimento de modelos predição de risco de ocorrência de mofo-branco devido ao seu impacto na produtividade do feijão. É evidente também o potencial que técnicas de *machine learning* têm para o desenvolvimento desses modelos e que modelos desenvolvidos com base em dados coletados em campo devem prever com melhor exatidão os fenômenos associados à ocorrência de doenças. Dessa maneira, este trabalho teve como objetivo desenvolver um modelo de *machine learning* para prever o risco de ocorrência do mofo-branco no feijoeiro-comum com base em dados de ocorrência da doença obtidos em campo e sua produtividade.

3.2. MATERIAL E MÉTODOS

Caracterização do conjunto de dados

Os dados utilizados neste trabalho foram obtidos por parceria com a Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig). Para desenvolver o modelo de predição de risco da ocorrência do mofo-branco (MB) no feijoeiro-comum, o trabalho foi dividido basicamente em duas etapas. A primeira etapa foi projetar e organizar o banco de dados coletados em campo. Na segunda etapa foram testados diferentes algoritmos de *machine learning* para a predição de MB, isso foi feito tomando por base as covariáveis que mais afetam a severidade da doença.

O banco de dados continha dados de experimentos realizados em 2000, 2001 e 2017 em Viçosa (Latitude: 20°46'7,42"S, Longitude: 42°52'11,98"W, Altitude: 660 m), e 2014 e 2015 em Oratórios (Latitude: 20°24'23,10"S, Longitude: 42°49'7,05"W, Altitude: 476 m), totalizando cinco experimentos. Os solos de Viçosa e Oratórios são classificados como Argissolo Vermelho-Amarelo, segundo ao Sistema Brasileiro de Classificação de Solos (2018). Quanto a proporção granulométrica dos solos, até a profundidade de 20 cm, as áreas experimentais de Viçosa e Oratórios apresentam, respectivamente, 230 g kg⁻¹, 240 g kg⁻¹ e 530 g kg⁻¹ e 250 g kg⁻¹, 440 g kg⁻¹ e 310 g kg⁻¹, de areia, silte e argila.

Dentre os dados utilizados constam a densidade de plantio (número de plantas por metro), cultivar, presença e ausência de aplicação de fungicida, índice de severidade do mofo-branco (ISMB), precipitação (em milímetros) e temperaturas (em graus Celsius) mínima, média e máxima durante o ciclo de vida, e produtividade (kg ha⁻¹). Em 2000 e 2001, 72 e 48 amostras foram disponibilizadas, respectivamente; em 2014, 2015 e 2017, 64 amostras cada ano, totalizando 312 condições amostradas. Um resumo dos dados usados está na Tabela 3.1. Os dados dos anos 2000 e 2001 podem ser encontrados no artigo "White mold management in common bean by increasing within-row distance between plants" (Vieira et al., 2010). Os dados dos anos 2014, 2015 e 2017 podem ser encontrados no artigo "Managing white mold on common bean with type III growth habit by integrating partial resistance, plant density, and fungicide" (Vieira et al., 2022).

Tabela 3.1 – Caracterização dos locais, densidades de plantio, cultivares de feijão-comum usadas e número de amostras dos experimentos.

Ano	2000	2001	2014	2015	2017
Local	Viçosa	Viçosa	Oratórios	Oratórios	Viçosa
Densidade (plantas/m)	15	16	13	13	13
	7,5	12	10	10	10
	5	8	7	7	7
		4	4	4	4
Cultivar	Pérola	Pérola	Madrepérola VC17	Madrepérola VC17	Madrepérola VC17
Número de amostras	72	48	64	64	64

Nota: O espaçamento entre linhas usada nos cinco experimentos foi de 0,5 m. Todos os experimentos foram irrigados com lâmina de aproximadamente 50 mm por semana, divididas em duas irrigações de 25 mm.

Fonte: elaborado pelo próprio autor.

A severidade do MB foi avaliada por meio da escala proposta por Hall e Phillips (1996) e o índice de severidade do mofo-branco (ISMB) foi calculado por meio da equação proposta por Kolkman e Kelly (2002), conforme a Equação (3.1). Para a precipitação foi considerado a somatório da precipitação durante os ciclos de vida dos feijoeiros. Para as temperaturas mínima, média e máxima, foi calculado a média para as respectivas temperaturas no decorrer dos ciclos. Das cultivares mostradas na Tabela 3.1, suas características são apresentadas na Tabela 3.2.

$$\text{ISMB}(\%) = 100 \frac{\sum(\text{notas de todas as plantas})}{4 \times (\text{número total de plantas})} \quad (3.1)$$

Tabela 3.2 – Características das cultivares usadas no estudo.

Cultivar	Porte	Hábito de crescimento	Resistência ao MB
Pérola	Semiprostrado	III - Indeterminado	Moderadamente resistente
Madrepérola	Prostrado	III - Indeterminado	Suscetível
VC17	Semiprostrado	III - Indeterminado	Tolerante

Fonte: elaborado pelo próprio autor.

Desenvolvimento do modelo de *machine learning* para predição da severidade do mofo-branco e da produtividade do feijoeiro-comum utilizando dados climáticos

O modelo para predição da ocorrência do risco do MB e produtividade do feijoeiro-comum por meio de dados de campo foi desenvolvido por meio de técnicas de *machine learning*. As variáveis usadas para o desenvolvimento dos modelos de predição foram densidade de plantio, tipo de cultivar, presença e ausência de aplicação de fungicida, índice de severidade do MB (ISMB), precipitação, e temperaturas mínima (Tmin), média (Tmed) e máxima (Tmax), e produtividade.

Neste trabalho foram realizadas três propostas de geração de predição por meio das técnicas de *machine learning*: a predição da severidade do MB para os dados de cinco anos (2000, 2001, 2014, 2015 e 2017) em que se teve experimento (será identificada por 5-sev) ; a predição da severidade para dados de três anos (2014, 2015 e 2017) em que se teve experimento (será identificada por 3-sev); e a predição da produtividade do feijoeiro-comum para os três anos (2014, 2015 e 2017) em que se teve experimento (será identificada por 3-prod). As covariáveis de entrada para a construção dos modelos de predição da severidade do MB para os dados 5-sev e 3-sev, foram a densidade de plantio, tipo de cultivar, presença e ausência de aplicação de fungicida, precipitação, e temperaturas mínima, média e máxima (Tabela 3.3). As covariáveis de entrada na construção dos modelos para a previsão da produtividade do feijoeiro-comum foram densidade de plantio, cultivar, presença e ausência de aplicação de fungicida, índice de severidade do MB (ISMB), precipitação, e temperaturas mínima, média e máxima (Tabela 3.3). Na Tabela 3.4 são apresentadas as variáveis usadas nas modelagens.

Tabela 3.3 – Variáveis de entrada para a disposição dos modelos propostos para os respectivos anos.

Variáveis de entrada	Modelos propostos por experimentos		
	5-sev	3-sev	3-prod
Densidade de plantio	x	x	x
Cultivar	x	x	x
Fungicida	x	x	x
Precipitação	x	x	x
Temperatura mínima	x	x	x
Temperatura média	x	x	x
Temperatura máxima	x	x	x
Índice de severidade do mofo-branco	-	-	x

Legenda: “x” = presença da covariável no modelo; “-” = ausência da covariável no modelo.
Fonte: elaborado pelo próprio autor.

Tabela 3.4 – Estatística descritiva (mínima, média e máxima) das variáveis usadas nos modelos propostos para os respectivos anos.

	Ano				
	2000	2001	2014	2015	2017
	Viçosa	Viçosa	Oratórios	Oratórios	Viçosa
Temperatura mínima (°C)	11,68	13,03	12,20	14,50	11,97
Temperatura média (°C)	18,14	19,44	18,57	19,75	18,40
Temperatura máxima (°C)	24,60	25,85	24,93	25,00	24,83
Precipitação (mm)	18,14	68,10	12,10	37,90	15,60
Índice de severidade do mofo-branco (ISMB) mínimo (%)	28,25	0,01	0,00	0,86	1,00
ISMB médio (%)	69,80	40,19	9,25	37,84	17,00
ISMB máximo (%)	99,25	85,75	80,34	79,04	51,00
Produtividade (kg.ha ⁻¹)	-	-	3044,88	3029,44	2406,58

Fonte: elaborado pelo próprio autor.

A primeira etapa de desenvolvimento do modelo para predição do risco de severidade do mofo-branco e da produtividade envolveu a análise do banco de dados. Inicialmente, selecionaram-se covariáveis para examinar possíveis interações e associações entre as covariáveis de entrada, para simplificar o modelo final e reduzir os custos computacionais durante o treinamento. Nesse processo, foram excluídas as covariáveis com correlação superior a 95% pelo teste de Spearman. Foi utilizado o

software R (R Core Team, 2023), juntamente com os pacotes *stats* (Hothorn, 2021) e *caret* (Kuhn et al., 2020), e suas funções específicas, *cor* e *find correlation* (Kuhn e Johnson, 2013). As covariáveis que passaram por essa etapa foram submetidas a uma segunda fase de remoção, considerando sua importância no algoritmo avaliado.

Os dados destinados ao treinamento e teste foram segregados utilizando a metodologia de validação cruzada *holdout* estratificada - VCHE (Siqueira et al., 2022), sendo separados 75% para treinamento e 25% para teste. A VCHE separou os dados em subgrupos referentes ao ano do experimento, de modo a eliminar tendências que pudessem comprometer ou trazer resultados tendenciosos no treinamento do modelo. Em situações que há diferente distribuição de amostras entre os dados, a VCHE é indicada para garantir uma boa distribuição amostral (Precody; Phiri; Ashish, 2023).

O treinamento foi realizado para otimização dos hiperparâmetros internos dos algoritmos usando método de validação cruzada repetida por *k-fold*. Foram utilizados 10 folds, cinco repetições e cinco valores de hiperparâmetros internos de cada algoritmo. O parâmetro de desempenho usado para a otimização do treinamento foi o Coeficiente de Correlação de Concordância de Lin – LCCC (Lin, 1989), conforme a Equação 3.2. Os hiperparâmetros são características internas intrínsecas em cada algoritmo que podem ser modificados para otimizá-los. Cada algoritmo apresenta um número específico dessas características, e os valores desses números foram escolhidos automaticamente pela função *train* do pacote *caret*. Os hiperparâmetros que são modificados automaticamente em cada algoritmo podem ser identificados no manual do pacote “*caret*” (Kuhn et al., 2020). Os resultados foram agregados estatisticamente, calculando-se a média dos parâmetros de desempenho. A escolha do melhor algoritmo baseou-se na análise estatística do parâmetro LCCC.

$$LCCC = \frac{2r\sigma_O\sigma_F}{\sigma_O^2 + \sigma_F^2 + (\bar{O} - \bar{F})^2} \quad (3.2)$$

Em que, r é o coeficiente de correlação entre os valores observados e preditos, σ_O e σ_F são as variâncias dos valores observados e preditos, respectivamente, $\bar{O}$ e $\bar{F}$ as médias dos valores observados e preditos, respectivamente.

Os algoritmos usados neste trabalho foram o *Support Vector Machine* Radial Sigma (SVM-RS), *Random Forest* (RF), *Kernel K Nearest Neighbors* (KNN) e *Linear Model* (LM). O algoritmo *Support Vector Machine* Radial Sigma (Cortes; Vapnik, 1995) é

um algoritmo proposto para problemas de regressão usando um *kernel Radial Basis Function* (RBF). Os parâmetros principais ajustados no SVM foram o sigma e o custo. Esses dois parâmetros configuram o *kernel* RBF, em que o sigma atua na largura do *kernel* (controlando a suavidade da superfície de decisão), e o custo na penalização dos erros. Com isso, minimizamos o erro de treinamento e maximizamos a margem do hiperplano para reduzir o *overfitting*.

O algoritmo RF (Breiman, 2001) realiza a construção do modelo do *machine learning* por meio de florestas de árvores de decisão na fase de treinamento do modelo. O treinamento se dá por meio da predição do modelo com uma amostra aleatória do conjunto de dados, em que o resultado da predição da regressão é a média das amostras usadas. Uma vantagem do RF é que ele possui maior capacidade de generalização e evita *overfitting* (Pal, 2005; Bridewell, 2005). Neste trabalho o hiperparâmetro ajustado no modelo foi o número de preditores selecionados aleatoriamente (*mtry*).

O *Kernel K Nearest Neighbors* (KNN), desenvolvido por Yu, Ji e Zhang (2002), representa uma evolução do algoritmo KNN (Fix; Hodges, 1951), oferecendo um conjunto expandido de hiperparâmetros para otimização. No presente estudo, os hiperparâmetros otimizados incluíram o número máximo de vizinhos (*kmax*), a medida de distância (*distance*) e o tipo de *kernel* utilizado.

O algoritmo *Linear Model* (LM) utilizado foi o *Multiple Linear Regression*, o qual é uma técnica que permite modelar a relação entre uma variável dependente e duas ou mais independentes. Assumindo uma extensão da regressão linear simples, a MLR se apresenta como a Equação 3.3.

$$Y = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k + \epsilon, \quad (3.3)$$

em que, Y é a variável dependente, b_0 é o coeficiente de interceptação, $b_1, b_2, \dots, b_k$, são os coeficientes das variáveis independentes $X_1, X_2, \dots, X_k$, respectivamente, ϵ é o erro.

Uma vez que o presente trabalho é uma predição de dados numéricos, advindos de um processo de regressão, a avaliação do desempenho dos algoritmos se deu pelos parâmetros LCCC, coeficiente de determinação (R^2), Raiz do Erro Quadrático Médio (RMSE) e Erro Médio Absoluto (MAE), representados pelas Equações 3.4, 3.5 e 3.6,

respectivamente, em que $\hat{y}$ é o valor predito, y é o valor observado e n o número de amostras. O LCCC avalia a concordância geral, levando em conta a precisão e exatidão entre os valores previstos e observados. O R^2 é uma métrica que caracteriza a qualidade do ajuste do modelo quantificando o quanto da variável dependente é explicada pelas variáveis independentes. O RMSE quantifica a eficácia do modelo por meio da padronização dos resultados, enquanto o MAE mensura qualidade do modelo por meio da avaliação da diferença entre os valores preditos e reais.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.4)$$

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.5)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.6)$$

Também foram calculados os modelos nulos das métricas RMSE e MAE, utilizando a função *nullModel* do pacote *caret* (Kuhn et al., 2020), conforme descrito nas Equações 3.7 e 3.8, respectivamente. Esses modelos nulos foram gerados com o propósito de comparação com os valores de RMSE e MAE obtidos pelos modelos produzidos pelos algoritmos. O modelo nulo é utilizado como referência para determinar se há vantagem na construção dos modelos propostos pelos algoritmos. O modelo nulo realiza a predição apenas com base na média dos dados de teste. Assim, se os modelos dos algoritmos não superarem o modelo nulo, ou seja, se não resultarem em valores menores de RMSE e MAE, isso sugere que os modelos gerados pelos algoritmos não conseguiram aprender de forma significativa, tornando questionável a sua utilização para prever os dados.

$$RMSE_nulo = \sqrt{\left(\frac{1}{n} \sum_{i=0}^n (\overline{Q_{train_i}} - Q_{obs_i})^2\right)} \quad (3.7)$$

$$MAE_nulo = \frac{1}{n} \sum |\overline{Q_{train_i}} - Q_{obs_i}| \quad (3.8)$$

Em que, $\overline{Q_{train_i}}$ é a média das amostras de treinamento, Q_{obs_i} é a amostra de validação e, n é o número de amostras (loop).

3.3. RESULTADOS

Para a avaliação do desempenho dos algoritmos na modelagem da proposta de prever a severidade do MB nos conjuntos de dados referentes nos cinco experimentos (5-sev), o algoritmo que teve o melhor desempenho foi o SVM, sendo o único a superar os valores de MAE e RMSE modelo nulo, conforme apresentada na Tabela 3.5. Os valores de MAE e RMSE do LM foram, respectivamente, 47,3% e 51,5%, e do RMSE do RF de 37,4%, enquanto os valores de MAE e RMSE do modelo nulo foram 33,5% e 37,0%, respectivamente. O LM apresentou maiores valores de MAE e RMSE comparado ao modelo nulo, enquanto o RF obteve apenas o valor de RMSE maior que o modelo nulo. O modelo que apresentou o pior desempenho foi o LM, seguido do algoritmo RF.

Tabela 3.5 – Valores dos parâmetros de desempenho (R^2 , RMSE, MAE e LCCC) de teste dos modelos nas predições do índice de severidade do mofo-branco para o experimento de cinco anos (5-sev) e três anos (3-sev), e da produtividade para três anos (3-prod) do feijoeiro-comum.

R^2				
	SVM ¹	RF ²	KKNN ³	LM ⁴
5-sev	0,296	0,239	0,183	0,327
3-prod	0,337	0,333	0,336	0,309
3-sev	0,19	0,37	0,341	0,384

RMSE					
	SVM	RF	KKNN	LM	MN ⁵
5-sev	32,1	37,4	34,87	51,49	36,99
3-prod	842,31	750,81	752,65	2077,2	764,08
3-sev	21,6	19,86	19,49	17,9	22,73

MAE					
	SVM	RF	KKNN	LM	MN
5-sev	28,53	33,3	29,92	47,3	33,51
3-prod	705,56	648,29	664,91	1990	655,04
3-sev	17,47	16,7	16,19	14,64	19,79

LCCC				
	SVM	RF	KKNN	LM
5-sev	0,005	0,134	0,168	0,305
3-prod	0,105	0,268	0,248	0,121
3-sev	0,152	0,327	0,346	0,42

¹Support Vector Machine; ²Random Forest; ³Kernel K Nearest Neighbors; ⁴Linear Model; ⁵Modelo Nulo.

Fonte: elaborado pelo próprio autor.

Para a predição da produtividade, os algoritmos SVM, KKNN e LM apresentaram os piores resultados na predição da 3-prod, visto que apresentaram valores de MAE e RMSE superiores ao do modelo nulo (Tabela 3.5). Enquanto o SVM e o LM apresentaram valores de MAE e RMSE de 706 e 842, e 1990 e 2077 kg.ha⁻¹, respectivamente, o modelo nulo apresentou valores de MAE e RMSE de 655 e 764 kg.ha⁻¹. O algoritmo KKNN teve o valor de MAE (665 kg.ha⁻¹) maior que o modelo nulo. Para a predição da produtividade,

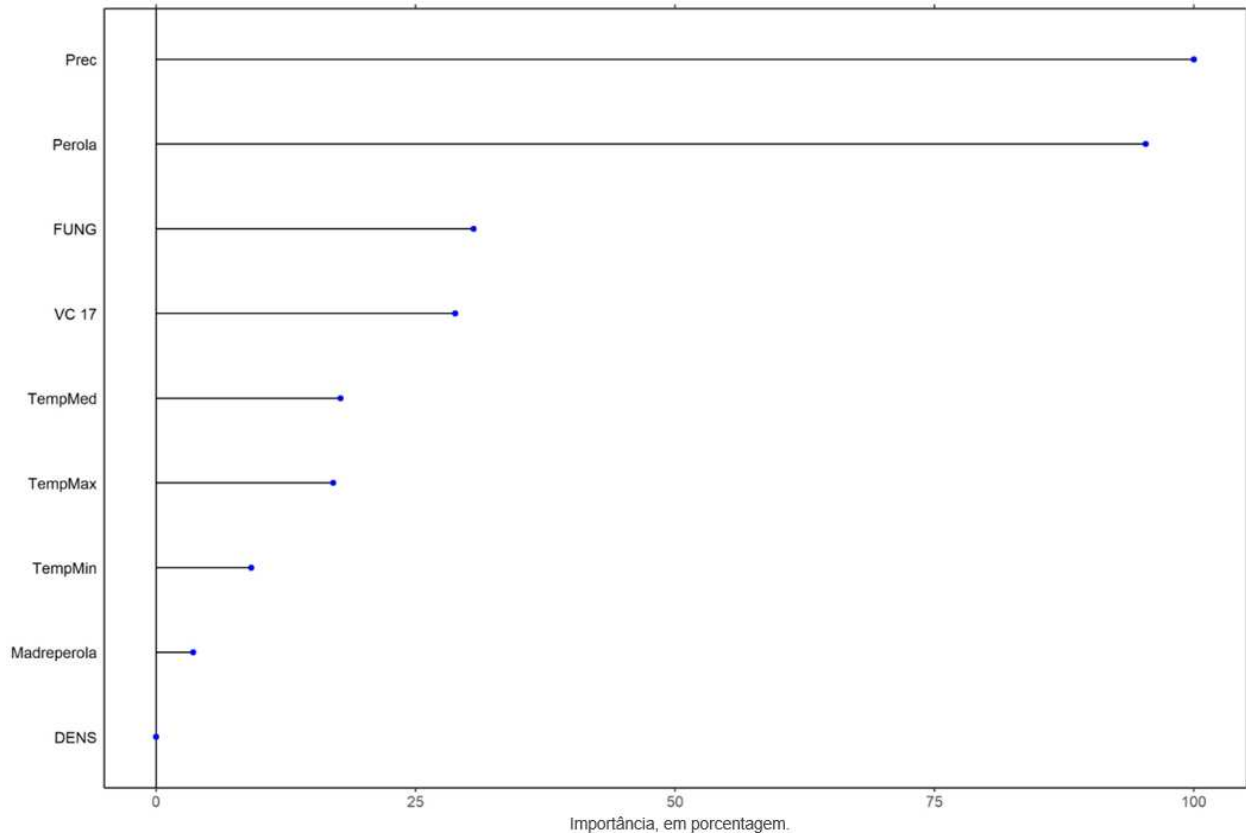
o algoritmo RF obteve desempenho superior ao modelo nulo, com valores de MAE e RMSE de 649 e 751 kg.ha⁻¹, respectivamente.

A predição da severidade utilizando os três anos (3-sev) proporcionou os melhores resultados quanto ao desempenho dos algoritmos, conforme mostra a Tabela 3.5. A modelagem dos algoritmos obtida na predição do 3-sev pode ser considerada que apresentou bons resultados, uma vez que apresentaram melhores valores de MAE e RMSE que o modelo nulo.

Seleção das variáveis

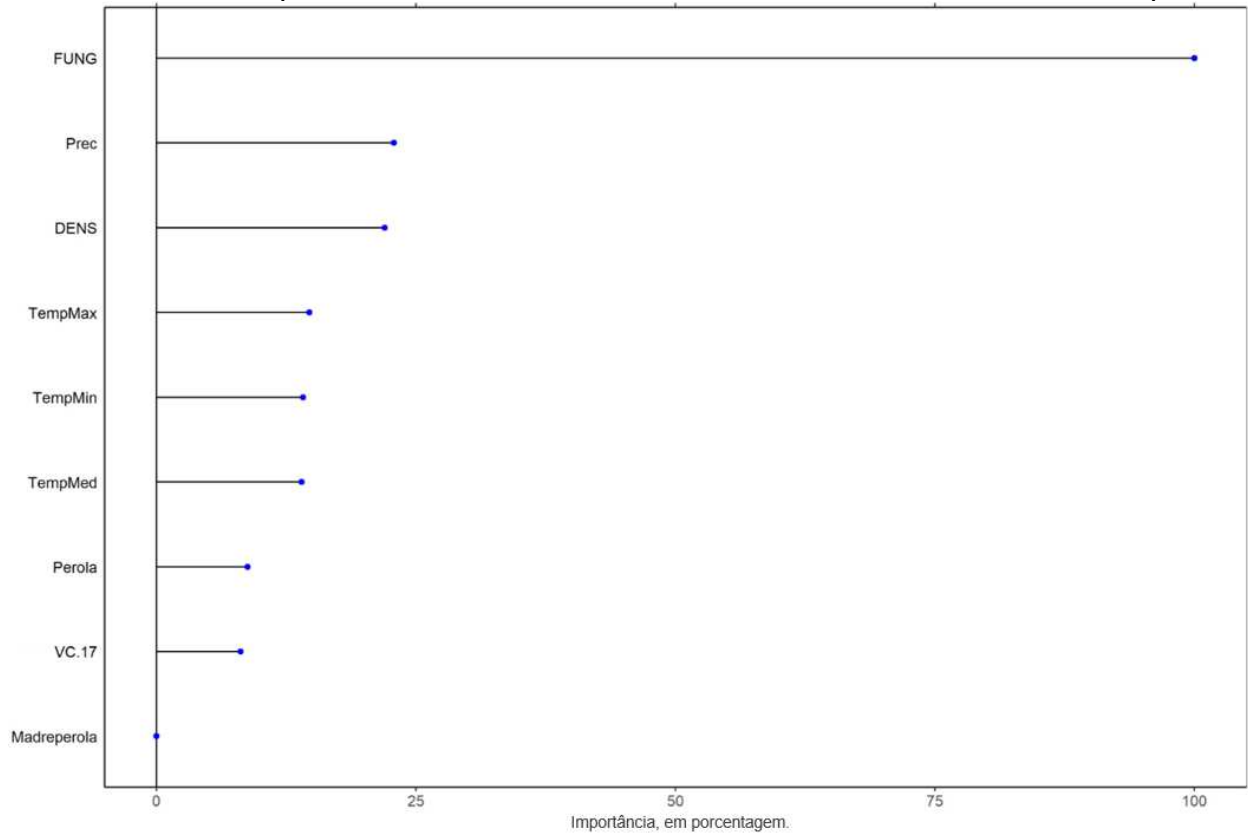
Neste trabalho, foram utilizados quatro algoritmos de machine learning: LM, RF, SVM e KKN. É possível que cada algoritmo selecione variáveis diferentes ou iguais para suas respectivas modelagens, com variação ou não em sua importância para os modelos. Mesmo quando as variáveis selecionadas são as mesmas, elas podem contribuir de forma diferente em cada modelo. Para a modelagem da severidade do MB nos cinco experimentos, todas as variáveis foram selecionadas pelos algoritmos KKN, RF e SVM, enquanto seis variáveis foram selecionadas pelo algoritmo LM, conforme demonstrado nas Figuras 3.1, 3.2, 3.3 e 3.4, respectivamente.

Figura 3.1 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em cinco anos (5-sev), por meio do algoritmo KKN. As covariáveis selecionadas para a modelagem em ordem de importância foram: precipitação, cultivar Pérola, fungicida, cultivar VC 17, temperatura média, temperatura máxima, temperatura mínima, cultivar Madrepérola e densidade de plantio.



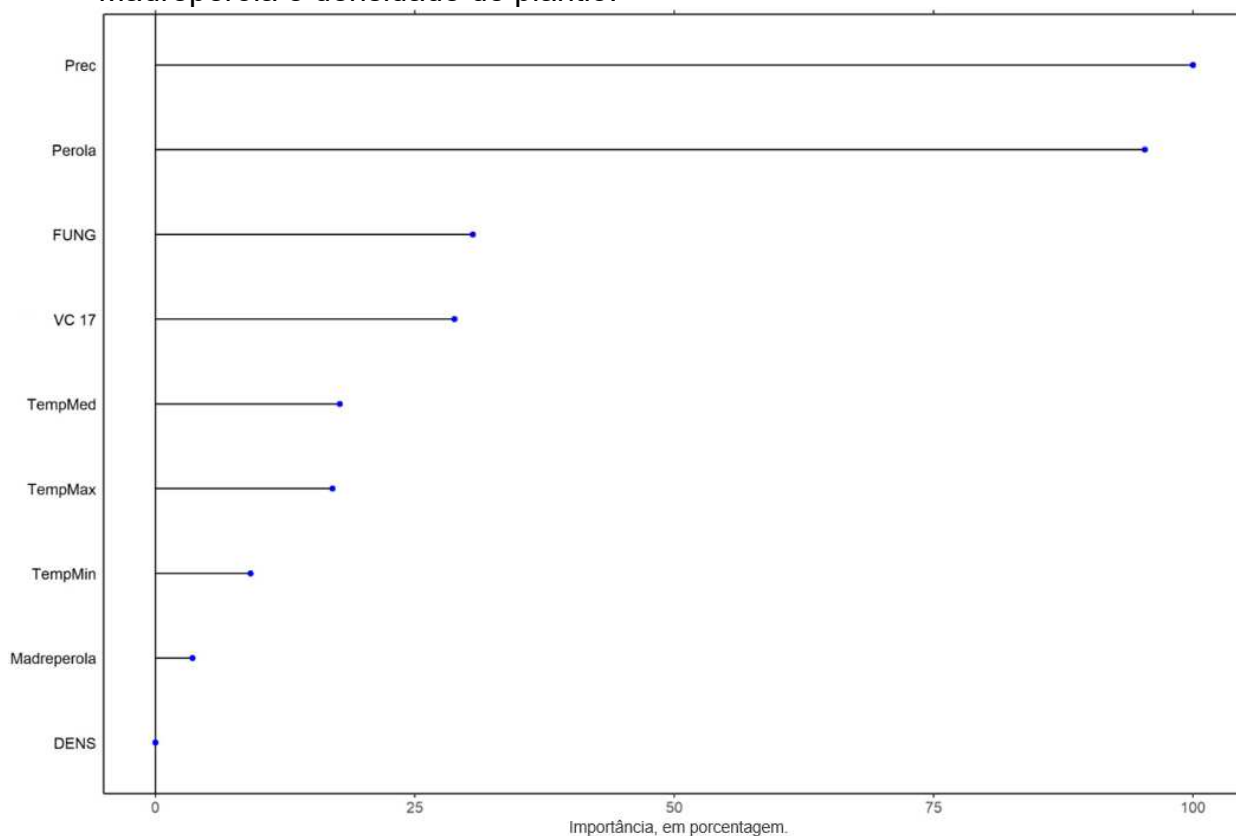
Fonte: elaborado pelo próprio autor.

Figura 3.2 – Seleção das variáveis mais importantes para a modelagem da predição da severidade do MB em cinco anos (5-sev), por meio do algoritmo RF. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, precipitação, densidade de plantio, temperatura máxima, temperatura mínima, temperatura média, cultivar Pérola, cultivar VC 17, cultivar Madrepérola.



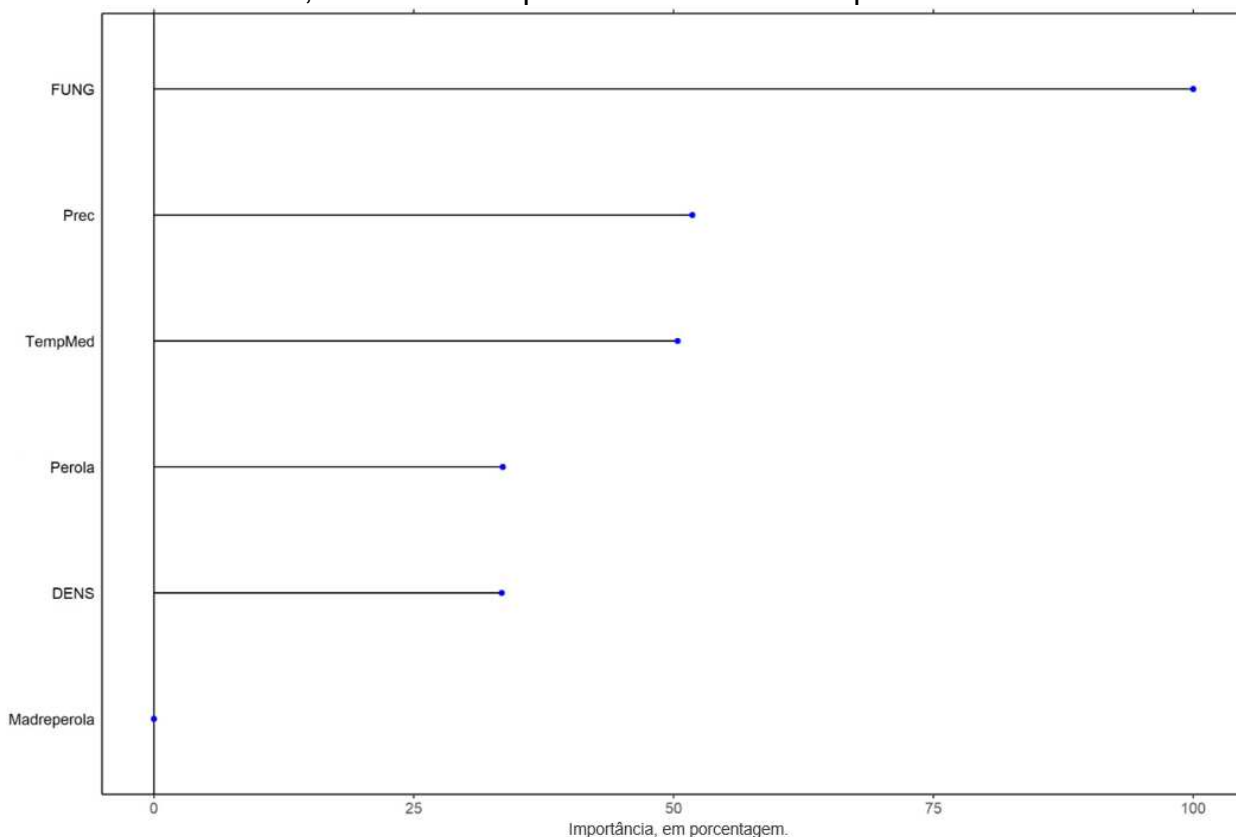
Fonte: elaborado pelo próprio autor.

Figura 3.3 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em cinco anos (5-sev), por meio do algoritmo *Support Vector Machine* (SVM). As covariáveis selecionadas para a modelagem em ordem de importância foram: precipitação, cultivar Pérola, fungicida, cultivar VC 17, temperatura média, temperatura máxima, temperatura mínima, cultivar Madrepérola e densidade de plantio.



Fonte: elaborado pelo próprio autor.

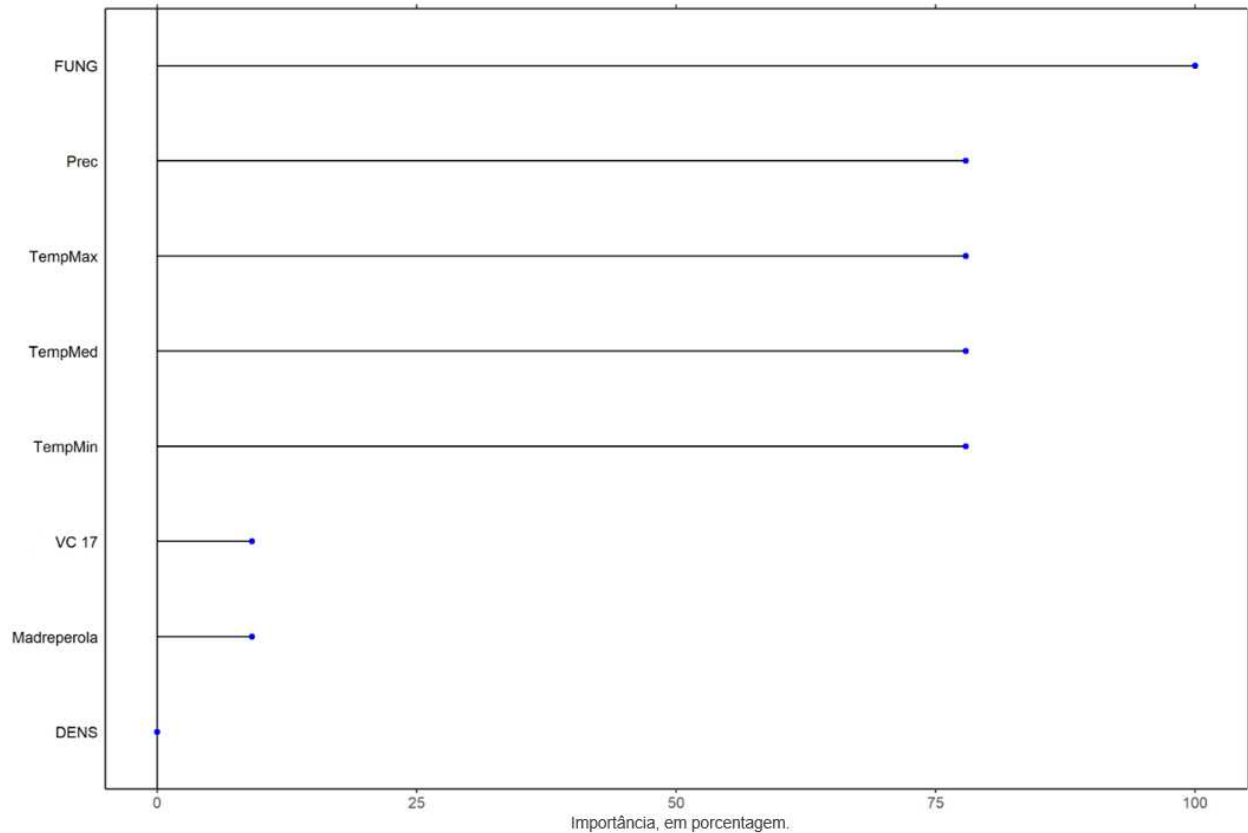
Figura 3.4 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em cinco anos (5-sev), por meio do algoritmo *Multiple Linear Regression* (LM). As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, precipitação, temperatura média, cultivar Pérola, densidade de plantio e cultivar Madrepérola.



Fonte: elaborado pelo próprio autor.

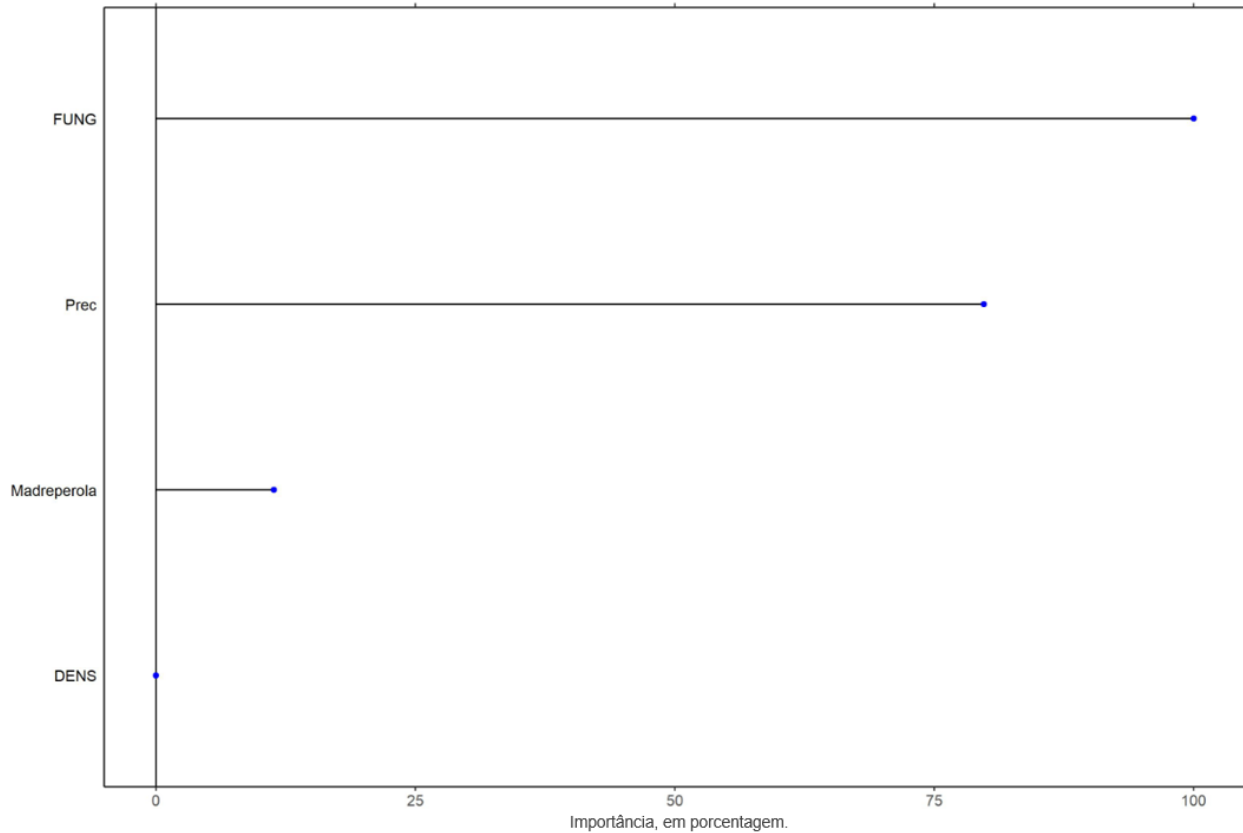
Assim como para a modelagem para cinco anos (5-sev), os algoritmos selecionaram todas as variáveis para a predição da severidade do MB em três anos (3-sev), conforme apresentados nas Figuras 3.5 a 3.8. Uma exceção ocorreu com o algoritmo LM, que selecionou quatro covariáveis (fungicida, precipitação, cultivar Madrepérola e densidade de plantio) das oito iniciais.

Figura 3.5 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em três anos (3-sev), por meio do algoritmo KKN. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, precipitação, temperatura máxima, temperatura média, temperatura mínima, cultivar VC 17, cultivar Madrepérola e densidade de plantio.



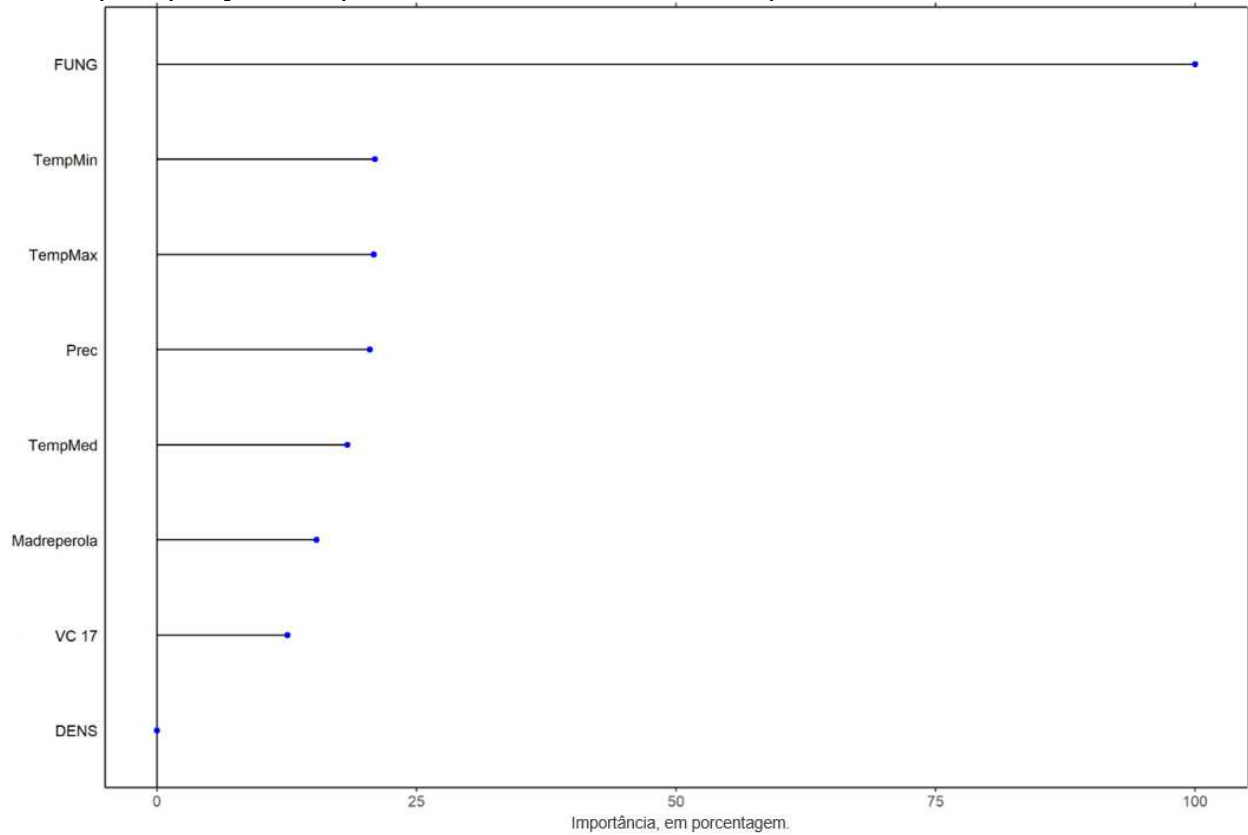
Fonte: elaborado pelo próprio autor.

Figura 3.6 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em três anos (3-sev), por meio do algoritmo LM. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, precipitação, cultivar Madrepérola e densidade de plantio.



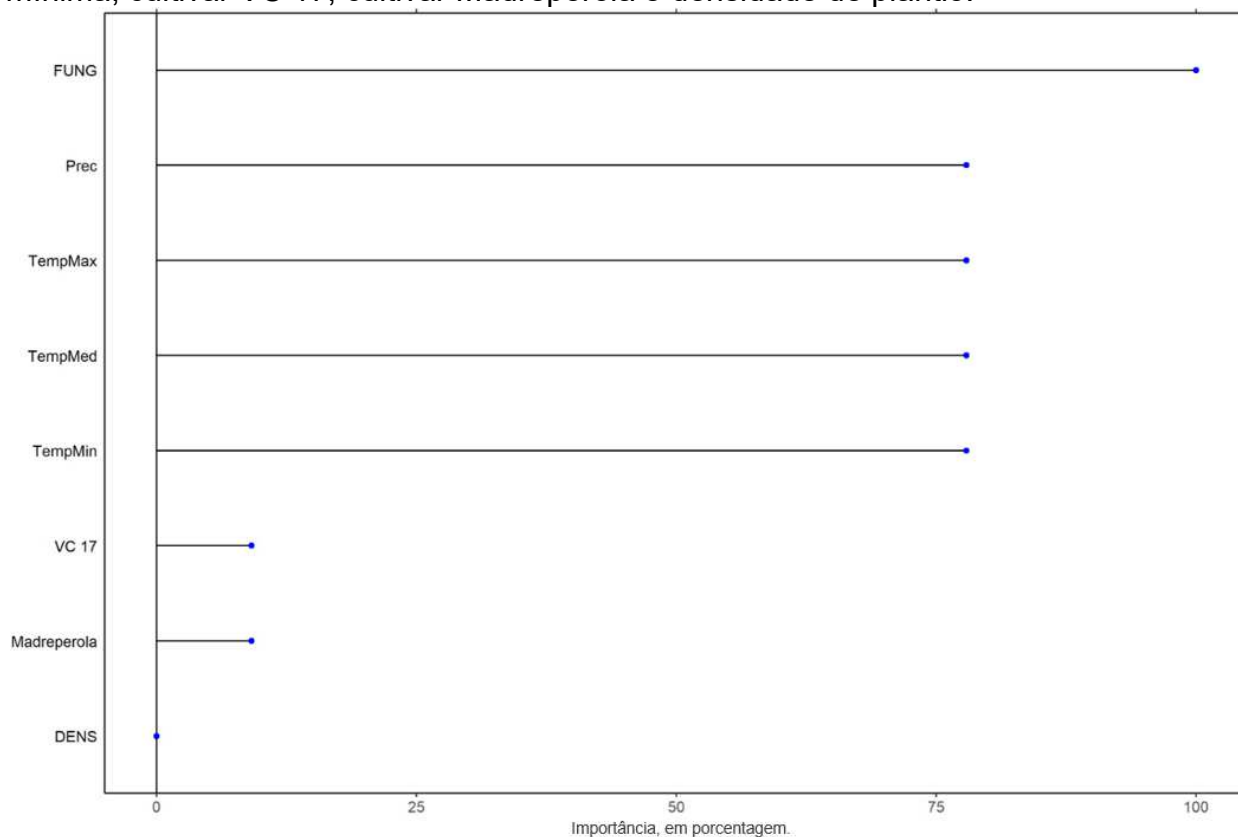
Fonte: elaborado pelo próprio autor.

Figura 3.7 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em três anos (3-sev), por meio do algoritmo *Random Forest*. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, temperatura mínima, temperatura máxima, precipitação, temperatura média, cultivar Madrepérola, cultivar VC 17 e densidade.



Fonte: elaborado pelo próprio autor.

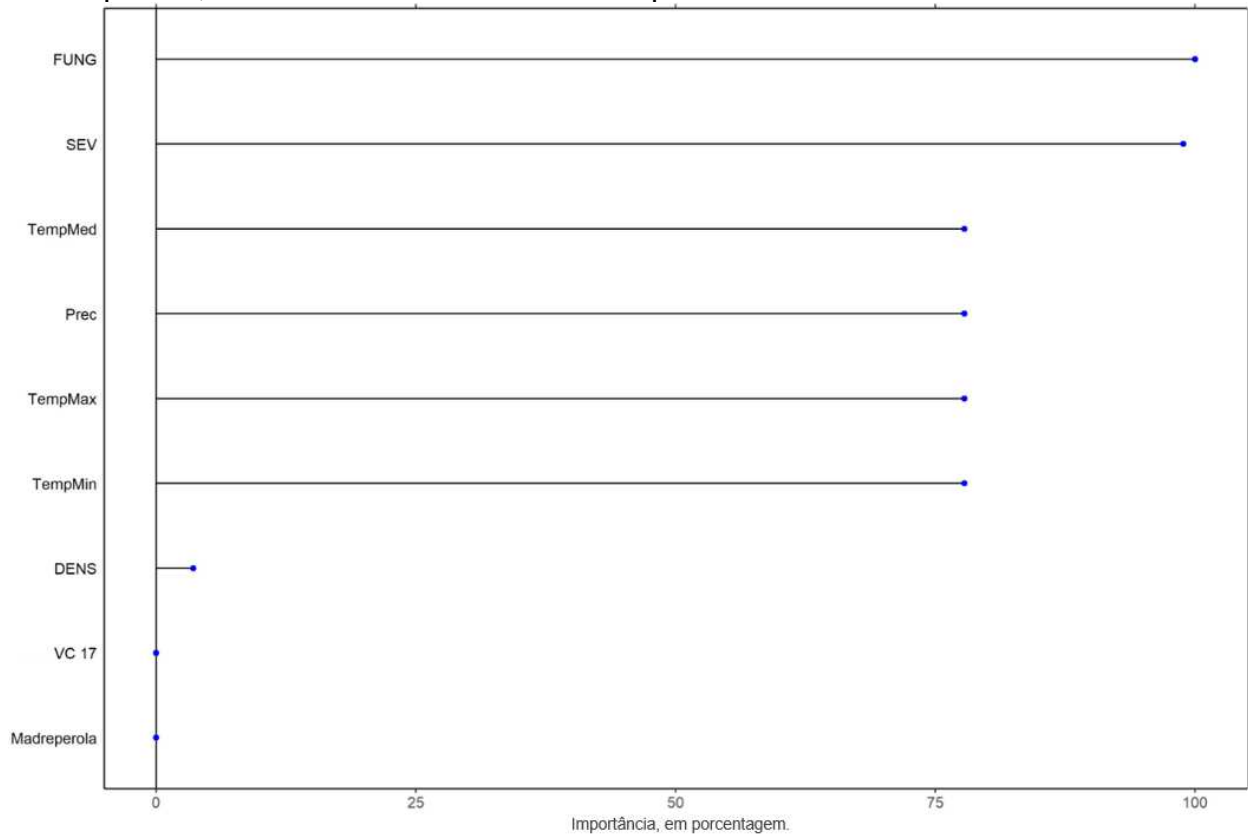
Figura 3.8 – Seleção das covariáveis mais importantes para a modelagem da predição da severidade do MB em três anos (3-sev), por meio do algoritmo *Support Vector Machine*. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, precipitação, temperatura máxima, temperatura média, temperatura mínima, cultivar VC 17, cultivar Madrepérola e densidade de plantio.



Fonte: elaborado pelo próprio autor.

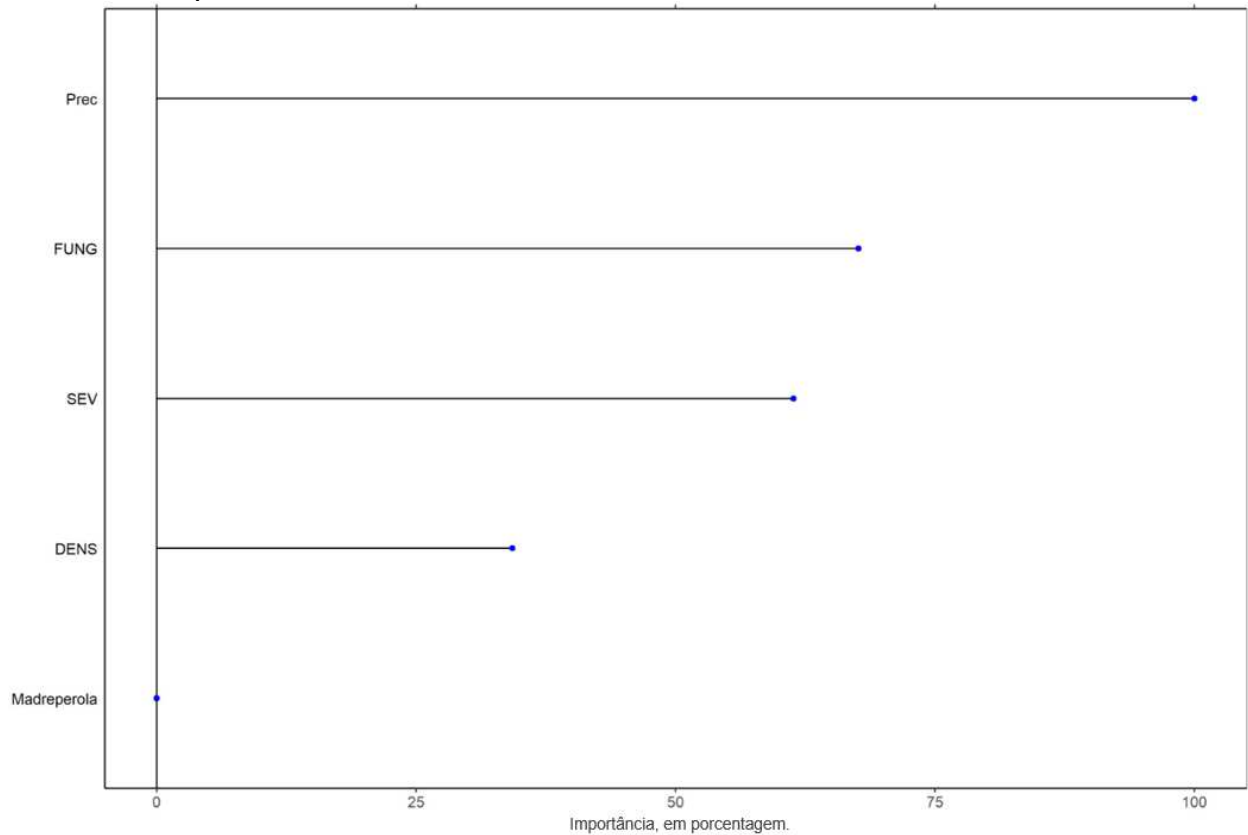
Os algoritmos para predição da produtividade do feijoeiro-comum apresentaram desempenho semelhantes às predições das severidades do mofo branco para três e cinco anos de experimentos. O KNN, RF e SVM selecionaram todas as variáveis como importantes e o LM selecionou cinco das variáveis iniciais como mais importantes para a modelagem da predição da produtividade (Figuras 3.9 a 3.12). As Figuras 3.13, 3.14 e 3.15 exibem os gráficos representando os dados observados e preditos dos melhores desempenhos obtidos pelas modelagens dos algoritmos nas previsões do ISMB para dados de cinco anos, ISMB para três anos e produtividade de três anos do feijoeiro-comum, respectivamente.

Figura 3.9 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade do feijoeiro-comum em três anos (3-prod), por meio do algoritmo *Kernel K Nearest Neighbors*. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, severidade, temperatura média, precipitação, temperatura máxima, temperatura mínima, densidade de plantio, cultivar VC 17 e cultivar Madrepérola.



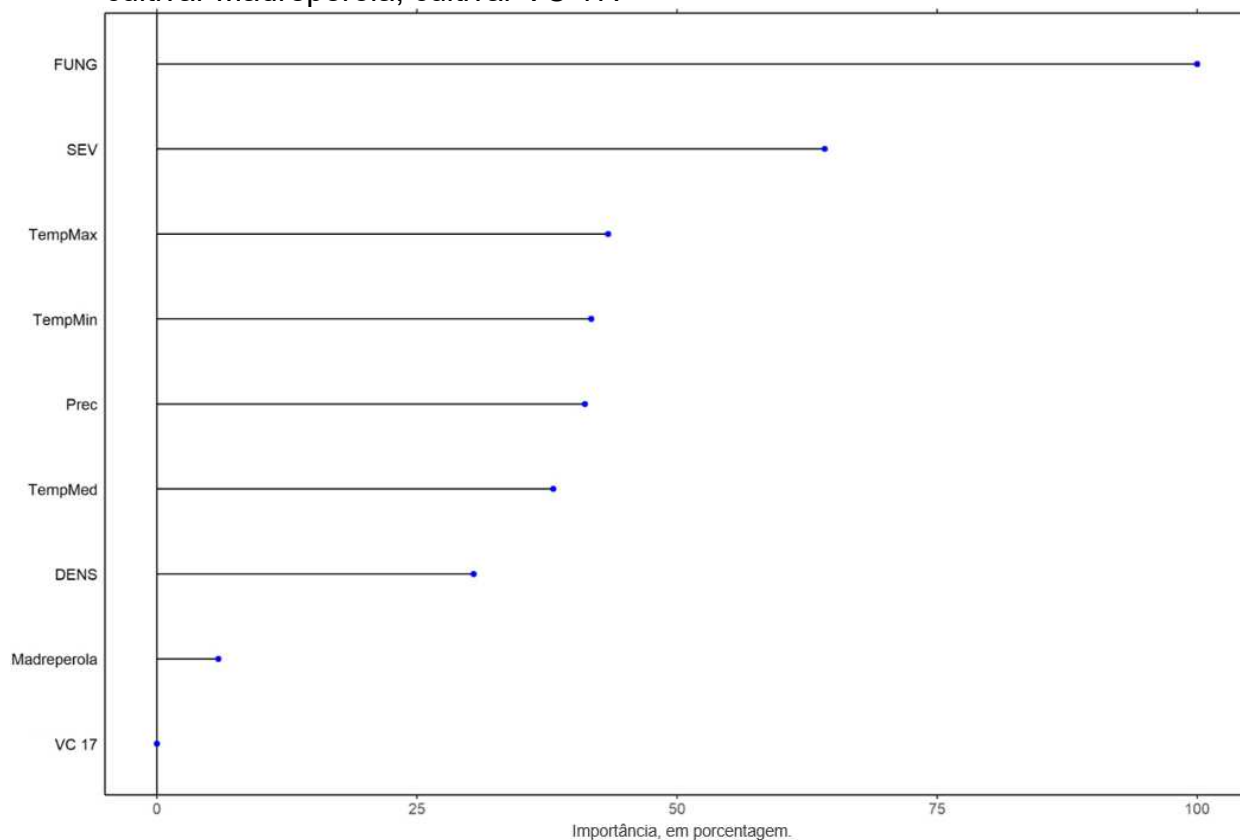
Fonte: elaborado pelo próprio autor.

Figura 3.10 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade do feijoeiro-comum em três anos (3-prod), por meio do algoritmo *Linear Model*. As covariáveis selecionadas para a modelagem em ordem de importância foram: precipitação, fungicida, severidade, densidade e cultivar Madreperola.



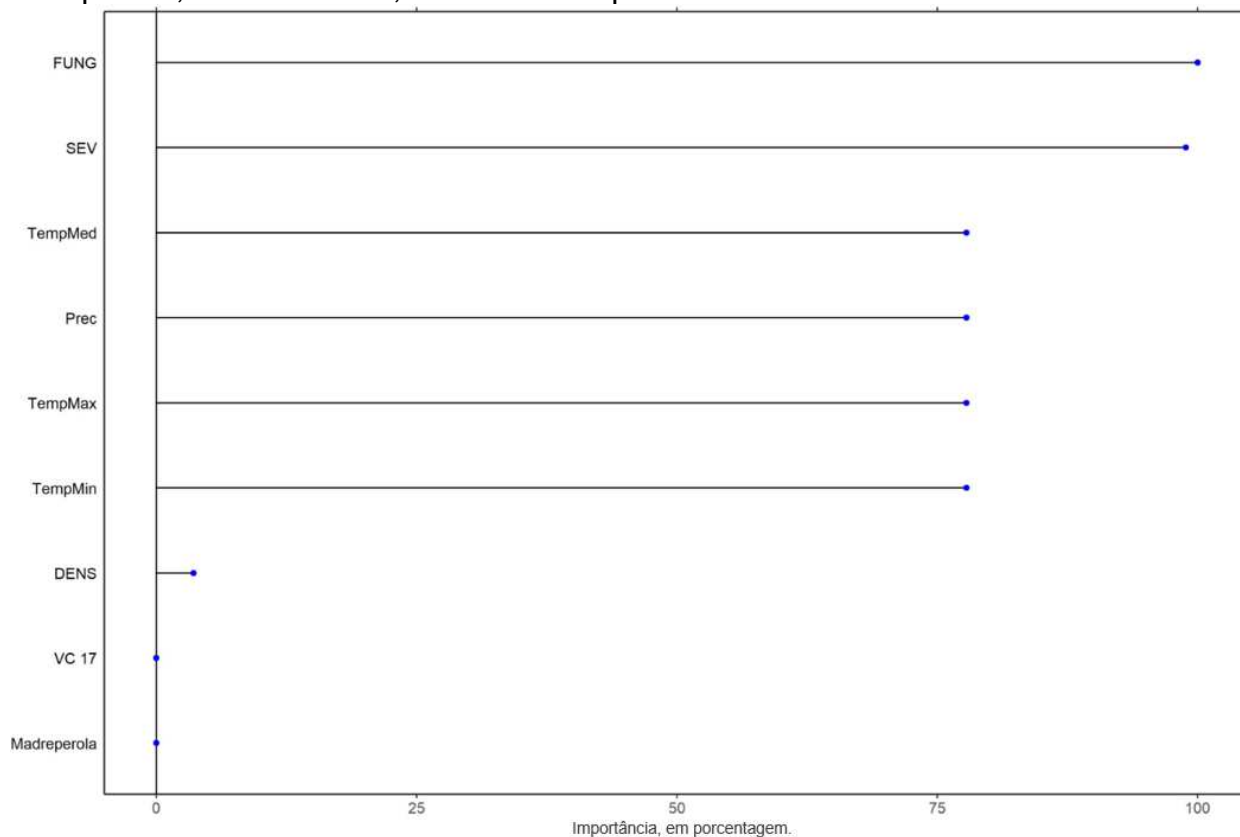
Fonte: elaborado pelo próprio autor.

Figura 3.11 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade do feijoeiro-comum em três anos (3-prod), por meio do algoritmo *Random Forest*. As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, severidade, temperatura máxima, temperatura mínima, precipitação, temperatura média, densidade de plantio, cultivar Madrepérula, cultivar VC 17.



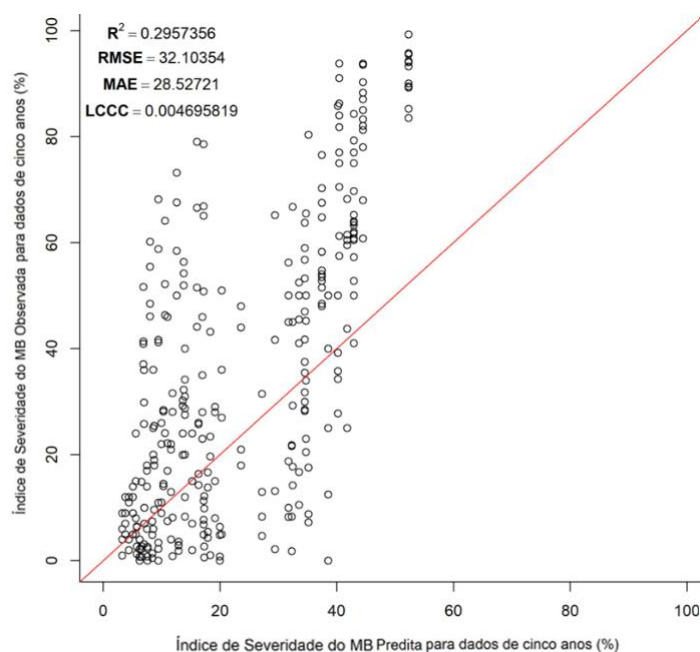
Fonte: elaborado pelo próprio autor.

Figura 3.12 – Seleção das covariáveis mais importantes para a modelagem da predição da produtividade do feijoeiro-comum em três anos (3-prod), por meio do algoritmo *Support Vector Machine* (SVM). As covariáveis selecionadas para a modelagem em ordem de importância foram: fungicida, severidade, temperatura média, precipitação, temperatura máxima, temperatura mínima, densidade de plantio, cultivar VC 17, cultivar Madrepérola.



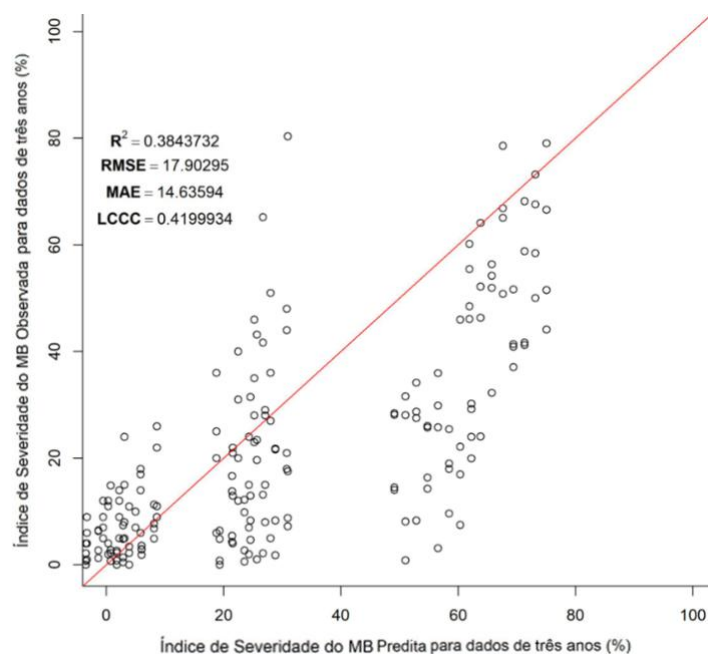
Fonte: elaborado pelo próprio autor.

Figura 3.13 – Gráfico de dispersão dos resultados de teste entre os valores observados e preditos da modelagem do algoritmo *Support Vector Machine* (SVM) na predição do índice de severidade mofo-branco (ISMB) para cinco anos de experimentos.



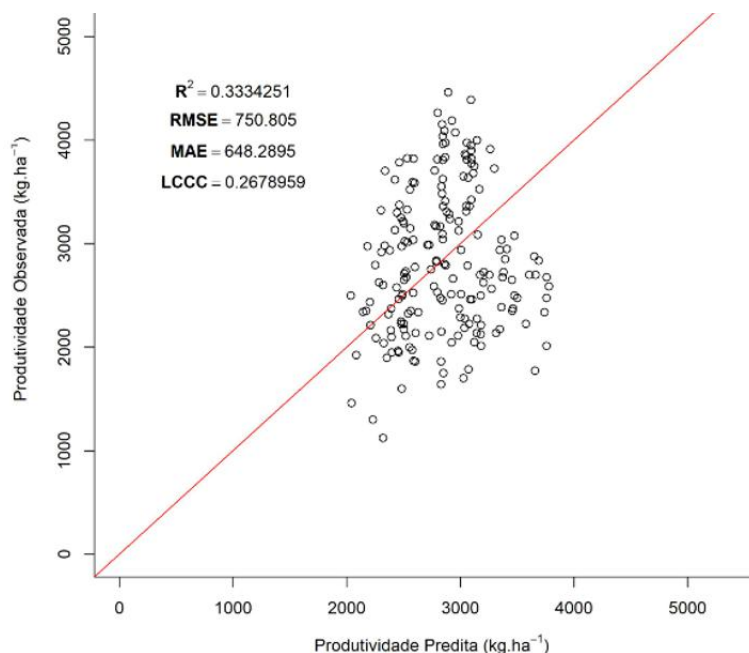
Fonte: elaborado pelo próprio autor.

Figura 3.14 – Gráfico de dispersão dos resultados de teste entre os valores observados e preditos da modelagem do algoritmo LM (*Multiple Linear Regression*) na predição do índice de severidade mofo-branco (ISMB) para três anos de experimentos.



Fonte: elaborado pelo próprio autor.

Figura 3.15 – Gráfico de dispersão dos resultados de teste entre os valores observados e preditos da modelagem do algoritmo *Random Forest* (RF) na predição da produtividade do feijoeiro-comum para três anos de experimentos.



Fonte: elaborado pelo próprio autor.

3.4. DISCUSSÃO

Neste trabalho foram fornecidos aos algoritmos de *machine learning* um conjunto de dados de experimentos realizados em campo, totalizando 312 amostras. Dos cinco experimentos realizados, três (em 2014, 2015 e 2017) apresentam o mesmo número de amostras, e em 2000 e 2001, números de amostras variadas entre si e entre os demais anos (Tabela 3.1). Este trabalho seguiu a proposta de trabalhar com os dados desses cinco experimentos para avaliar o quanto os modelos conseguem generalizar as previsões com um banco de dados restrito e com número de amostras diferentes entre experimentos. Dessa forma, foi usado, neste trabalho, a validação cruzada *holdout* estratificada (VCE), de forma que reduzisse a influência dos experimentos que tenham maior número de amostras na modelagem. Neetoo et al. (2021) e Sujatha et al. (2021) também utilizaram a VCE em seus trabalhos para avaliar doenças em culturas para atender reduzido número de amostras.

De maneira geral, os modelos usados neste trabalho conseguiram obter razoáveis desempenhos nas modelagens para os experimentos das severidades de três e cinco

anos, e para a produtividade de três anos. A razoabilidade mencionada ocorreu visto que a modelagem dos algoritmos superou o desempenho do modelo nulo na maioria dos casos, e os experimentos ocorreram em condições de campo em ambiente não controlado. Logo, sugere-se que os algoritmos possuem capacidade de detectar padrões no conjunto de dados nas predições de severidade do MB e produtividade do feijoeiro. Ao observar o desempenho dos modelos, nota-se que as métricas mostram que o conjunto de dados usado em cada predição influenciou a qualidade dos desempenhos. O modelo nulo fornece um resultado que os algoritmos precisam superar para trazer algum benefício em seu uso. Em outras palavras, o modelo nulo realiza a predição dos dados retornando valores da tendência central das métricas MAE e RMSE. Logo, se a modelagens dos algoritmos não apresentarem valores mais baixos, o modelo não consegue entender os padrões do conjunto de dados e fazer predições mais precisas do que um modelo básico fornecido pelo modelo nulo. Os valores das métricas, no geral, foram baixos, como provável consequência do reduzido tamanho do conjunto de dados, o que afetou o desempenho dos modelos, problema semelhante ao observado por Neetoo et al. (2021).

O algoritmo LM foi o único usado como modelo linear neste trabalho. O uso do LM se mostrou inviável na predição da severidade do MB para cinco anos e para a produtividade de três anos (5-sev e 3-prod, respectivamente), visto que apresentou valores de MAE e RMSE maiores que o modelo nulo, conforme visto na Tabela 3.5. Modelos lineares tendem a se limitar a relações lineares entre as variáveis. Portanto, em um contexto de em que as variáveis se comportam de maneira não linear, o modelo linear não responde bem a relações mais complexas.

Para a predição da severidade do MB usando o conjunto de dados de cinco anos (5-sev), o SVM foi o que apresentou o melhor desempenho, com os menores valores de MAE e RMSE. Shahoteisi, Manesh e Mendoza (2022) encontram resultados semelhantes quanto ao potencial uso do SVM na predição do MB no feijoeiro-comum em condições de laboratório. Sarangdhar e Pawar (2017) também concluíram que o SVM é um bom algoritmo para detectar doenças na folha do algodoeiro. Por mais que os algoritmos KNN e RF apresentassem bons resultados no treinamento, não superaram o SVM nos dados de teste.

Todos os algoritmos mostraram-se viáveis na modelagem da predição do 3-sev. No entanto, o algoritmo LM foi o algoritmo que apresentou o melhor desempenho (Tabela 3.5). Ele apresentou desempenho diferente ao apresentando na predição do 5-sev e 3-prod. Por isso, o LM teve os maiores valores das métricas R^2 e LCCC, e menores valores de MAE, RMSE. O segundo melhor algoritmo foi o KNN, com os maiores valores de R^2 e LCCC e menor valor de MAE após o algoritmo LM. O pior desempenho ficou para o algoritmo SVM com os valores mais baixos das métricas R^2 e LCCC e maiores de MAE e RMSE. Embora não seja comum melhores desempenhos de modelos lineares em comparação a modelos não lineares, isso pode ocorrer, como neste trabalho.

O algoritmo destaque para a modelagem da 3-prod foi o algoritmo RF. O desempenho superior do RF se deu por esse algoritmo apresentar menores valores de MAE e RMSE e maior valor do LCCC (Tabela 3.5). Por mais que o algoritmo KNN tenha apresentado bons valores das métricas, a métrica MAE o equiparou ao modelo nulo, indicando que seu uso não trouxe muitos benefícios para o conjunto de dados deste trabalho. Os demais algoritmos (SVM e LM) tiveram resultados mais baixos que o modelo nulo, mostrando serem inviáveis para a predição da produtividade com o conjunto de dados deste estudo.

Os resultados das modelagens mostraram que a seleção de variáveis dos algoritmos selecionou todas as variáveis do conjunto de dados como importantes para as predições, com exceção do algoritmo LM. O LM selecionou seis covariáveis (fungicida, precipitação, temperatura média, cultivar Pérola, densidade de plantio e cultivar Madrepérola) na modelagem da 5-sev, quatro na modelagem da 3-sev (fungicida, precipitação, cultivar Madrepérola e densidade de plantio), e cinco covariáveis na 3-prod (precipitação, fungicida, severidade, densidade de plantio e cultivar Madrepérola). Conforme apresentado na Figura 3.6, quatro foram as covariáveis mais importantes para a modelo. A covariável fungicida teve 100% de importância, seguida da precipitação (79,76%), cultivar Madrepérola (11,35%). A densidade de plantio teve menos peso entre as covariáveis selecionadas.

A seleção da covariável fungicida na 3-sev pelo algoritmo LM foi fortemente significativa com a produtividade, visto que é o método mais usado no controle do MB (Silva et al., 2024). A recomendação do tratamento com fungicida é que sejam feitas pelo

menos duas aplicações, a primeira no início da floração, e a segunda de sete a dez dias após. O fungicida usado nos experimentos que deu origem ao conjunto de dados deste trabalho foi o fluazinam. No entanto, fazer rotação de fungicidas é recomendável para manter a eficácia do controle do MB e evitar pressão de seleção do fungo na área aos fungicidas (Silva et al., 2024). A maior importância dada ao fungicida pelo algoritmo LM neste trabalho corrobora com os resultados obtidos por Vieira et al. (2012), que encontraram resultados em que o fungicida foi mais eficaz no controle do MB do que o espaçamento e a densidade de plantio. Neste trabalho, Vieira et al. (2012) também usaram a cultivar Pérola, utilizada no presente estudo.

A precipitação na predição do 3-sev apresentou 79,8% de contribuição na modelagem. A importância da precipitação está no tempo de molhamento foliar que causa na lavoura. Outros trabalhos, como o de Shahoteisi, Manesh e Mendoza (2022) e Phillips (1994), mostraram que o molhamento foliar é um importante fator para a ocorrência do MB em feijoeiro-comum. A temperatura, além da umidade relativa do ar, é um fator de elevada relevância no desenvolvimento do MB. Observa-se que nenhuma temperatura (T_{min} , T_{med} ou T_{max}) entrou como covariável importante no modelo LM na predição da 3-sev. Esse resultado traz ponderações importantes, visto que diversos autores citam a importância da temperatura como fator importante junto com a umidade no desenvolvimento do MB (Paula Júnior et al., 2006; Abawi; Grogan, 1975, 1979; Clarkson et al., 2004; Phillips, 1987). Uma hipótese para o modelo não considerar as temperaturas como covariáveis de entrada está na faixa de variação ser pequena entre os experimentos, conforme mostrado na Tabela 3.4., além de os ensaios terem sido conduzidos no inverno.

Entre as covariáveis que contribuíram mais na modelagem da 5-sev pelo algoritmo SVM estão a precipitação (Prec) e a cultivar Pérola, ambas com 100% e 95,3% de contribuição na, respectivamente. A seleção da covariável Prec se apresentar como a de maior contribuição reforça o resultado encontrado na modelagem do 3-sev pelo algoritmo LM. A Pérola é uma cultivar moderadamente resistente ao mofo-branco, o que a coloca como um importante fator na modelagem. Esse resultado corrobora com o trabalho de Teixeira et al. (2019), sugerindo que a cultivar Pérola permite um melhor controle do MB em condições de baixa pressão da doença na área, além de contribuir com a menor

necessidade de aplicação fungicida. Dessa forma, a cultivar Pérola mostra-se como uma boa preditora por apresentar um melhor controle do MB, resultando em menor severidade ao MB advindo da sua resistência parcial, podendo resultar em maior rendimento produtivo da lavoura. Ainda, Venegas e Saad (2010) concluíram que a cultivar Pérola respondeu de maneira eficaz ao tratamento fungicida para o controle do MB, corroborando com os resultados encontrados neste trabalho.

O algoritmo RF teve melhor desempenho na predição na 3-prod e foi o único a superar o modelo nulo no MAE e RMSE (Tabela 3.5). Observou-se que o RF selecionou como principais fatores que contribuiu com a modelagem as covariáveis fungicida e severidade, contribuindo com 100% e 65,1%, respectivamente. Mais uma vez o fungicida se mostrou com a mais importante variável no controle do MB. Esse resultado embasa as usuais aplicações de fungicida, preventivamente, no manejo do MB e evitar consequentes quebras de produtividade. No entanto, apesar do uso preventivo do fungicida ser comum, ele pode ser desnecessário em condições desfavoráveis ao desenvolvimento do MB (McCreary et al., 2016).

3.5. CONCLUSÕES

Este estudo teve como objetivo desenvolver um modelo de machine learning para a predição da severidade do mofo-branco do feijoeiro-comum e da sua produtividade com um conjunto de dados de cinco experimentos. Os resultados deste trabalho sugerem que os algoritmos testados tiveram um bom desempenho, pelo fato de terem sido superiores à modelagem do modelo nulo. Os algoritmos que sobressaíram na predição do 5-sev, 3-sev e 3-prod foram o SVM, LM e RF, respectivamente. Os baixos valores dos algoritmos nos testes podem ter ocorrido em razão do tamanho restrito do conjunto de dados, indicando que os algoritmos não apresentaram boa capacidade de generalização com base no tamanho do conjunto de dados usados. No entanto, o estudo trouxe percepções importantes do desempenho do manejo da doença quanto as covariáveis usadas, e realçou a importância de alguns fatores que contribuem para a severidade do MB.

3.6. AGRADECIMENTOS

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

3.7. REFERÊNCIAS

- ABAWI, G. S.; GROGAN, R. G. Epidemiology of diseases caused by *Sclerotinia* species. **Phytopathology**, v. 69, n. 8, p. 899-904, 1979.
- ABAWI, G.; GROGAN, R. Source of primary inoculum and effects of temperature. **Phytopathology**, v. 65, p. 300-309, 1975.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **The Journal Of Machine Learning Research**, Montréal, Qc, Canada, v. 13, n. 1, p.281-305, jan. 2012.
- BERUSKI, G. C.; PEREIRA, A. B.; JACCOUD FILHO, D. S.; SARTORI, F. F.; SENTELHAS, P. C. Effects of climate and agricultural practices on temporal progress of *Sclerotinia sclerotiorum* for soybean in Brazil. **Agricultural Sciences**, v. 7, n. 07, p. 425, 2016.
- BOLAND, G. J.; HALL, R. Index of plants of hostes *Sclerotinia sclerotiorum*. **Canadian Journal of Plant Pathology**, v. 16, n.1, p. 93–108, 1994.
- BRASIL. Ministério da Agricultura e Reforma Agrária. **Regras para análise de sementes. Brasília**, 1992. 365p.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.
- BRIDEWELL, W.; ASADI, N. B.; LANGLEY, P.; TODOROVSKI, L. Reducing overfitting in process model induction. In: **Proceedings of the 22nd international conference on Machine learning**. 2005. p. 81-88.
- CLARKSON, J. P.; PHELPS, K.; WHIPPS, J. M.; YOUNG, C. S.; SMITH, J. A.; WATLING, M. Forecasting *Sclerotinia* disease on lettuce: toward developing a prediction model for carpogenic germination of sclerotia. **Phytopathology**, v. 94 (3), 268-279, 2004.
- CONAB - Companhia Nacional de Abastecimento. Safrá Brasileira de Grãos – Feijão, Boletim Mensal - Abril, 2022. Disponível também em: <https://www.conab.gov.br/component/k2/item/download/41684_5d815afac2734e1a46eedcb0c15f6a27>.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, v. 20, p. 273-297, 1995.

FIX, E.; HODGES, J. L. Discriminatory analysis, nonparametric discrimination: Consistency properties. **Technical Report 4**, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

HALL, R.; PHILLIPS, L. G. Evaluation of parameters to assess resistance of white bean to white mold. **Annual Report of the Bean Improvement Cooperative**, East Lansing, v.39, p.306-307, 1996.

KANE, M. V.; GRABAU, L. J. Early planted, early maturing soybean cropping system: Growth, development, and yield. **Agronomy journal**, v. 84, n. 5, p. 769-773, 1992.

KUSUMO, B. S.; HERVANA, A.; MAHENDRA, O.; PARDELE, H. F. Machine learning-based for automatic detection of corn-plant diseases using image processing. In: **2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA)**. IEEE, 2018. p. 93-97.

LEITE, R. M. V. B. C. Ocorrência de doenças causadas por *Sclerotinia sclerotiorum* em girassol e soja. **Embrapa Soja. Comunicado Técnico**, 2005.

LIMA, R. C.; TEIXEIRA, P. H.; VIEIRA, R. F.; PAULA JÚNIOR, T. J. Densidade de plantas como componente do manejo integrado do mofo-branco do feijoeiro. **Controle alternativo de pragas e doenças: opção ou necessidade?**, p. 134.

MCCREARY, C. M.; DEPUYDT, D.; VYN, R. J.; GILLARD, C. L. Fungicide efficacy of dry bean white mold [*Sclerotinia sclerotiorum* (Lib.) de Bary, causal organism] and economic analysis at moderate to high disease pressure. **Crop protection**, v. 82, p. 75-81, 2016.

MORGAN, M.; BLANK, C.; SEETAN, R. Plant disease prediction using classification algorithms. **IAES International Journal of Artificial Intelligence**, v. 10, n. 1, p. 257, 2021.

MOTA, A. L.; MIQUELLUTI, D. L.; OZAKI, V. A. Predição de sinistros agrícolas: uma abordagem comparativa utilizando aprendizagem de máquina. **Economia Aplicada**, v. 24, n. 4, p. 533-554, 2020.

MWEBAZE, E.; OWOMUGISHA, G. Machine learning for plant disease incidence and severity measurements from leaf images. In: **2016 15th IEEE international conference on machine learning and applications (ICMLA)**. IEEE, 2016. p. 158-163.

NEETOO, H.; CHUTTUR, Y.; NAZURALLY, A.; TAKOOREE, S.; ALLY, N. M. Crop disease prediction using multiple linear regression modelling. In: **International Conference on Soft Computing and its Engineering Applications**. Cham: Springer International Publishing, 2021. p. 312-326.

PAL, M. Random forest classifier for remote sensing classification. **International journal of remote sensing**, v. 26, n. 1, p. 217-222, 2005.

PAULA JÚNIOR, T. J.; MORANDI, M. A. B.; LOBO JUNIOR, M.; VIEIRA, R. F.; ZAMBOLIM, L. Controle alternativo do mofo-branco do feijoeiro. **Embrapa Meio Ambiente-Capítulo em livro científico (ALICE)**, 2006.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, v. 12, p. 2825-2830, 2011.

PHILLIPS, A. J. Carpogenic germination of sclerotia of *Sclerotinia sclerotiorum*: a review. **Phytophylactica**, 19(3), 279-284, 1987.

PHILLIPS, A. J. L. Influence of fluctuating temperatures and interrupted periods of plant surface wetness on infection of bean leaves by ascospores of *Sclerotinia sclerotiorum*. **Annals of Applied Biology**, v. 124, n. 3, p. 413-427, 1994.

PRECODY, K. G.; PHIRI, K. F.; CHAKRAVERTI, A. K. A Transformer on Tabular Data Comparative Analysis with Linear and Tree Base Machine Learning Algorithm on Diabetic Dataset. **International Journal of Innovative Science and Research Technology**, 2023.

QUINLAN, J. R. Learning with continuous classes. In: **5th Australian joint conference on artificial intelligence**. 1992. p. 343-348.

SANTOSO, H.; TANI, H.; WANG, X. Random Forest classification model of basal stem rot disease caused by *Ganoderma boninense* in oil palm plantations. **International Journal of Remote Sensing**, v. 38, n. 16, p. 4683-4699, 2017.

SHAHOTEISI, F.; RIAHI MANESH, M.; DEL RÍO MENDOZA, L. E. Modeling risk of *Sclerotinia sclerotiorum*-induced disease development on canola and dry bean using machine learning algorithms. **Scientific reports**, v. 12, n. 1, p. 1-10, 2022.

SIQUEIRA, R. G.; VELOSO, G. V.; FERNANDES FILHO, E. I.; FRANCELINO, M. R.; SCHAEFER, C. E. G.; CORRÊA, G. R. Evaluation of machine learning algorithms to classify and map landforms in Antarctica. **Earth Surface Processes and Landforms**, v. 47, n. 2, p. 367-382, 2022.

SUJATHA, R.; CHATTERJEE, J. M.; JHANJHI, N. Z.; BROHI, S. N. Performance of deep learning vs machine learning in plant leaf disease detection. **Microprocessors and Microsystems**, v. 80, p. 103615, 2021.

TEIXEIRA, P. H.; LIMA, R. C.; BONICONTRO, B. F.; MENDES, O. L.; SOARES, B. A.; CARNEIRO, J. E. S.; PAULA JÚNIOR, T. J.; VIEIRA, R. F. Management of white mold in common bean using partial resistance and fungicide applications. **Crop Protection**, v. 124, p. 104867, 2019.

VENEGAS, F.; SAAD, J. C. C. Fungigação no controle do mofo branco e produtividade do feijoeiro em condições de cerrado brasileiro. **Irriga**, v. 15, n. 2, p. 159-172, 2010.

VIEIRA, R. F.; LIMA, R. C.; TEIXEIRA, P. H.; PAULA JÚNIOR, T. J.; CARNEIRO, J. E. E.; POSSAMAI, F.; BRAUN, E.; LIMA, M. S. Managing white mold on common bean with type III growth habit by integrating partial resistance, plant density, and fungicide. **Plant disease**, v. 106, n. 10, p. 2672-2677, 2022.

VIEIRA, R. F.; PAULA JÚNIOR, T. J.; TEIXEIRA, H. White mold management in common bean by increasing within-row distance between plants. **Plant disease**, v. 94, n. 3, p. 361-367, 2010.

YU, K.; JI, L.; ZHANG, X. Kernel nearest-neighbor algorithm. **Neural Processing Letters**, v. 15, p. 147-156, 2002.

CAPÍTULO 4 - CONCLUSÕES GERAIS

Foi proposto neste trabalho a aplicação de algoritmos de aprendizagem de máquina para o desenvolvimento de modelo para estimar a incidência do mofo-branco (IMB), o índice de severidade do mofo-branco (ISMB) e produtividade do feijoeiro-comum. Os trabalhos foram desenvolvidos pela coleta de dados espectrais adquiridos por meio de RPA, por dados do mofo-branco (MB) em campo e dados climáticos.

Na primeira proposta foi usado informações espectrais por meio de aprendizagem de máquinas por meio das modelagens realizadas pelos algoritmos na estimativa da IMB, ISMB e produtividade. Os algoritmos que apresentaram os melhores desempenhos das modelagens foram o Cubist e o KKNn. Enquanto o Cubist apresentou os melhores desempenhos na predição do IMB e do ISMB, o KKNn obteve o melhor desempenho na predição da produtividade. As covariáveis que apresentaram maior contribuição na predição do IMB foram a banda G aos 52 DAE, MSR aos 38 DAE, banda G aos 38 DAE, EVI aos 38 DAE, banda G aos 59 DAE e banda B aos 38 DAE. Para o ISMB as covariáveis mais importantes foram a banda G aos 38 DAE, banda G aos 59 DAE, MSR aos 38 DAE, banda G aos 52 DAE e banda B aos 38 DAE. A predição da produtividade apresentou como as covariáveis mais importantes o a banda G aos 38 DAE, CI aos 38 DAE, banda G aos 52 DAE, banda B aos 38 DAE, CVI aos 52 DAE, GDVI aos 59 DAE e banda G aos 59 DAE.

Para esta primeira proposta, o melhor desempenho foi para a predição da produtividade, por meio do algoritmo KKNn, mostrando potencial para aplicabilidade em campo sob condições não controláveis, como ocorreu neste trabalho.

A segunda proposta constou de os algoritmos de aprendizagem de máquinas predizerem o ISMB e a produtividade para conjunto de dados de três (3-sev e 3-prod, respectivamente) e cinco anos (5-sev). Os desempenhos dos algoritmos nas modelagens foram modestos, no entanto, superaram o modelo nulo, o que ainda justificaria o uso dos algoritmos para as respectivas predições. Os algoritmos que apresentaram melhores desempenhos nas predições do ISMB para cinco anos, ISMB para três anos e produtividade para três anos foram o SVM, LM e RF respectivamente.