

JOÃO AMARO FERREIRA VIEIRA NETTO

USO DE INTELIGÊNCIA COMPUTACIONAL NA FENOTIPAGEM DE SOJA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

Orientador: Leonardo Lopes Bhering

**VIÇOSA - MINAS GERAIS
2024**

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

T

V658u
2024
Vieira Netto, João Amaro Ferreira, 1995-
 Uso de inteligência computacional na fenotipagem de soja /
 João Amaro Ferreira Vieira Netto. – Viçosa, MG, 2024.
 1 dissertação eletrônica (82 f.): il. (algumas color.).

Orientador: Leonardo Lopes Bhering.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Biologia Geral, 2024.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2024.257>

Modo de acesso: World Wide Web.

1. *Glycine max* - Genética. 2. Fenótipo. 3. Aprendizado do
computador. 4. Vegetação - Índices. I. Bhering, Leonardo Lopes,
1980-. II. Universidade Federal de Viçosa. Departamento de
Biologia Geral. Programa de Pós-Graduação em Genética e
Melhoramento. III. Título.

CDD 22. ed. 633.342


JOÃO AMARO FERREIRA VIEIRA NETTO

USO DE INTELIGÊNCIA COMPUTACIONAL NA FENOTIPAGEM DE SOJA


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADO: 28 de Fevereiro de 2024.

Assentimento:

Documento assinado digitalmente
 **JOAO AMARO FERREIRA VIEIRA NETTO**
Data: 29/07/2024 11:07:23-0300
Verifique em <https://validar.iti.gov.br>

João Amaro Ferreira Vieira Netto
Autor

Documento assinado digitalmente
 **LEONARDO LOPES BHERING**
Data: 29/07/2024 11:21:28-0300
Verifique em <https://validar.iti.gov.br>

Leonardo Lopes Bhering
Orientador

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa de estudos.

À minha família, por todo apoio.

À Isadora, por todo amor.

À UFV, pela oportunidade.

Ao professor Leonardo Lopes Bhering, pelo exemplo.

Aos colegas de laboratório e da UFV, pela amizade.

A Deus, por tudo.

Muito obrigado!

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

RESUMO

NETTO, João Amaro Ferreira Vieira, M. Sc., Universidade Federal de Viçosa, março de 2024. **Uso de inteligência computacional na fenotipagem de soja.** Orientador: Leonardo Lopes Bhering.

A soja (*Glycine max* (L.)) é uma fonte valiosa de alimento humano, animal e como matéria prima industrial. Para atender às demandas crescentes, a soja enfrenta uma série de desafios complexos. Esses desafios estão intrinsecamente ligados à busca por variedades mais produtivas, resistentes a doenças e adaptadas a condições ambientais variáveis. A fenotipagem de alto rendimento emerge como uma ferramenta crucial nesse processo, acelerando o desenvolvimento eficiente de novas variedades mais produtivas, capazes de enfrentar os desafios ambientais e sociais em constante evolução. Esta oferece vantagens quando na comparação com a fenotipagem tradicional, uma vez que a partir do uso de imagens e sensores, auxiliados com softwares e algoritmos de inteligência computacional, otimizam o processo de medição, possibilita maior fenotipagem em escala e reduz a variabilidade da mensuração humana. A utilização de índices de vegetação está entre os principais meios utilizados na fenotipagem de alto rendimento. Com a utilização destes, é possível realizar uma variedade de estudos, incluindo a avaliação do teor de nitrogênio nas folhas, a determinação de características físicas como biomassa, altura da planta e área foliar, a análise da heterogeneidade das plantas no campo, a estimativa do teor de clorofila, a avaliação do teor de água nas plantas, a quantificação do teor de lignina, e a detecção de danos causados por pragas e doenças. Com o uso de técnicas de fenotipagem em larga escala, torna-se crescente volume e a complexidade dos dados obtidos, aumentando a exigência para abordagens inovadoras para análise dos dados com foco em seleção eficiente. Nesse contexto, a aplicação de inteligência computacional tem emergido como uma ferramenta essencial para transformar a forma como se aborda a fenotipagem e o melhoramento de plantas. Assim, este trabalho tem como objetivo avaliar o desempenho de algoritmos de Perceptron Multicamadas (PMC) e Random Forest (RF) na predição de características fenotípicas de soja a partir dos seguintes índices de vegetação: Índice de vegetação de Diferença

Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), Índice de vegetação de Diferença Normalizada Verde (GNVDI), Índice de Vegetação Aprimorado (EVI), Índice de Diferença Normalizada de Borda Vermelha (NDRE) e Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI). Além disso, busca-se investigar o impacto dos parâmetros das redes neurais no desempenho de um PMC e como esses parâmetros influenciam sua eficácia na predição de características fenotípicas de soja. Constatou-se que, para os modelos de PMC, conforme a quantidade de neurônios aumentava até 10, e o número de folds na validação cruzada aumentava até 15, os modelos apresentavam resultados progressivamente melhores. Por outro lado, o aumento do número de épocas resultava em um aumento nos valores de R^2 dos modelos, alcançando um limite de 30000, após o qual os resultados começavam a diminuir. Ademais, o acréscimo no número de camadas ocultas ocasionava uma redução nos valores dos coeficientes de determinação, indicando que o melhor número de camadas foi um. Além disso, houve destaque na predição das variáveis AIV e AP, especialmente em relação aos modelos de Random Forest avaliados.

Palavras-chave: Predição; Aprendizado de Máquina; Índices de Vegetação; *Glycine max (L)*.

ABSTRACT

NETTO, João Amaro Ferreira Vieira, M. Sc., Universidade Federal de Viçosa, March, 2024. **Use of computational intelligence in soybean phenotyping.** Adviser: Leonardo Lopes Bhering.

Soybean (*Glycine max (L.)*) is a valuable source of human, animal, and industrial food. To meet growing demands, soybeans face a series of complex challenges. These challenges are intrinsically linked to the search for more productive varieties, resistant to diseases, and adapted to variable environmental conditions. High-throughput phenotyping emerges as a crucial tool in this process, accelerating the efficient development of new, more productive varieties capable of addressing constantly evolving environmental and social challenges. It offers advantages compared to traditional phenotyping, as it uses images and sensors aided by software and computational intelligence algorithms, optimizing the measurement process, enabling greater phenotyping on scale, and reducing human variability. The use of vegetation indices is among the main methods used in high-throughput phenotyping. With their use, a variety of studies can be conducted, including the evaluation of nitrogen content in leaves, determination of physical characteristics such as biomass, plant height, and leaf area, analysis of plant heterogeneity in the field, estimation of chlorophyll content, evaluation of plant water content, quantification of lignin content, and detection of damage caused by pests and diseases. With the use of large-scale phenotyping techniques, the volume and complexity of the obtained data are increasing, increasing the demand for innovative approaches to data analysis focused on efficient selection. In this context, the application of computational intelligence has emerged as an essential tool for transforming the way plant phenotyping and improvement are approached. Thus, this work aims to evaluate the performance of Multilayer Perceptron (MLP) and Random Forest (RF) algorithms in predicting soybean phenotypic characteristics from the following vegetation indices: Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Modified Soil-Adjusted Vegetation Index (MSAVI), Green Normalized Difference Vegetation Index (GNVDI), Enhanced Vegetation Index (EVI), Normalized Difference Red Edge Index (NDRE),

and Simplified Canopy Chlorophyll Content Index (SCCCI). Additionally, it seeks to investigate the impact of neural network parameters on the performance of an MLP and how these parameters influence its effectiveness in predicting soybean phenotypic characteristics. It was found that, for MLP models, as the number of neurons increased up to 10, and the number of folds in cross-validation increased up to 15, the models showed progressively better results. On the other hand, an increase in the number of epochs resulted in an increase in the R^2 values of the models, reaching a limit of 30000, after which the results began to decrease. Furthermore, an increase in the number of hidden layers led to a reduction in the values of the determination coefficients, indicating that the optimal number of layers was one. Moreover, there was a highlight in the prediction of the variables AIV and AP, especially regarding the evaluated Random Forest models.

Keywords: Prediction; Machine Learning; Vegetation Indices; *Glycine max (L)*.

SUMÁRIO

RESUMO.....	4
ABSTRACT	6
INTRODUÇÃO GERAL.....	10
Fenotipagem de alto rendimento	11
Inteligência computacional.....	12
REFERÊNCIAS	15
CAPÍTULO 1 - TOPOLOGIA E PARÂMETROS DE MODELOS DE PERCEPTON MULTICAMADAS NA FENOTIPAGEM DE SOJA A PARTIR DE IMAGENS MULTIESPECTRAIS.....	21
RESUMO.....	21
CHAPTER 1 - TOPOLOGY AND PARAMETERS OF MULTILAYER PERCEPTRON MODELS IN SOYBEAN PHENOTYPING USING MULTISPECTRAL IMAGES	23
ABSTRACT	23
INTRODUÇÃO	25
MATERIAIS E MÉTODOS	27
Descrição do experimento	27
Inteligência Computacional	29
RESULTADOS	30
Efeito do número de folds em relação ao número de neurônios por camada oculta	30
Relação camadas ocultas x folds x neurônios nas variáveis avaliadas	36
Efeito do número de épocas em relação ao número de neurônios por camada oculta	42
Relação camadas ocultas x épocas x neurônios entre as variáveis avaliadas	48
Média geral na relação entre o número de camadas e número de neurônios por camada.....	54
Média geral entre o número de camadas e número de épocas	55
Média geral entre número de camadas e número de folds.....	55
DISCUSSÃO	56
CONCLUSÃO	57
REFERÊNCIAS	58
CAPÍTULO 2 – ALGORITMOS DE INTELIGÊNCIA COMPUTACIONAL NA FENOTIPAGEM DE SOJA ATRAVÉS DE IMAGENS MULTIESPECTRAIS	64
RESUMO.....	64

CHAPTER 2 – COMPUTATIONAL INTELLIGENCE ALGORITHMS IN PLANT PHENOTYPING USING MULTISPECTRAL IMAGES	66
ABSTRACT	66
INTRODUÇÃO	68
MATERIAIS E MÉTODOS	69
Descrição do experimento	69
Inteligência Computacional	70
RESULTADOS	72
Altura de inserção da vagem (AIV)	72
Altura de planta (AP)	72
Dias para maturação (DM)	72
Produtividade (PROD)	73
DISCUSSÃO	73
CONCLUSÃO	75
REFERÊNCIAS	75
CONCLUSÃO GERAL	82

INTRODUÇÃO GERAL

A soja (*Glycine max* (L.)) foi domesticada entre seis e nove mil anos atrás a partir da soja selvagem (*Glycine soja* Sieb. & Zucc.) no leste asiático (Sedivy et al., 2017) e tem o nordeste da China como região com maior diversidade genética (Li et al., 2010). Desde então, se tornou uma fonte valiosa de alimento humano, animal e como matéria prima industrial (Modgil et al., 2021). Atualmente é cultivada em todos os continentes, com destaque para Brasil, EUA e Argentina como os principais países produtores (USDA, 2023).

A soja, com sua riqueza genética, é destaque no melhoramento vegetal. Essa diversidade oferece oportunidades para o desenvolvimento de variedades adaptadas a uma ampla gama de condições climáticas e latitudes, apresentando resistência a pragas, doenças e outras características altamente desejáveis para a agricultura e indústria (Zhangxiong Liu et al., 2017; Xuenqin Liu et al., 2017; Smýkal et al., 2015).

O processo de melhoramento da soja, apesar de sua importância para atender às demandas crescentes por alimentos e recursos sustentáveis, enfrenta uma série de desafios complexos. Esses desafios estão intrinsecamente ligados à busca por variedades mais produtivas, resistentes a doenças e adaptadas a condições ambientais variáveis. Um dos principais desafios é a crescente pressão exercida pelas mudanças climáticas. A variação nos padrões climáticos, como aumento da temperatura e eventos climáticos extremos, afeta diretamente o cultivo da soja, tornando essencial o desenvolvimento de variedades capazes de enfrentar essas condições adversas. Silva et al. (2021) reporta que o melhoramento genético da soja deve priorizar variedades de ciclo longo, já que as mudanças climáticas podem reduzir o ciclo médio de todas as classes de maturidade avaliadas em seu trabalho, assim como modificar as cultivares de soja para possuírem folhas menores, o que ajudaria na dissipação do calor e minimizaria o estresse térmico nas plantações.

Battisti et al. (2017) ressaltam que, diante das projeções climáticas, estratégias de melhoramento da soja focadas em aprimorar a profundidade das raízes, restringir a transpiração conforme o déficit de pressão de vapor e minimizar a sensibilidade do período de enchimento de grãos à escassez de água, têm o potencial de ampliar a resposta positiva das plantas de soja no sul do Brasil diante das futuras condições climáticas previstas. Enquanto isso, na China, pode-se encontrar uma redução da

produtividade da soja entre 2021 e 2060 de até 24% (Guo et al., 2022). Além disso, mediante às mudanças climáticas, áreas que hoje são produtivas, deixarão de ser e áreas que hoje não se cultiva soja, poderão ser utilizadas futuramente, aumentando a pressão para o desenvolvimento de variedades de soja adaptadas para novas regiões (Fodor et al., 2017).

À medida que a demanda global por alimentos cresce (Van Dijk et al., 2021; Bodirsky et al., 2015), fatores climáticos e tecnológicos representam desafios para o aumento da produtividade vegetal necessário para atender a essa demanda futura (Guo et al., 2022; Iizumi et al., 2014). Uma estratégia fundamental para contornar esses desafios envolve o desenvolvimento de variedades de plantas adaptadas às condições específicas de cada região. A fenotipagem de alto rendimento emerge como uma ferramenta crucial nesse processo, acelerando o desenvolvimento eficiente de novas variedades mais produtivas, capazes de enfrentar os desafios ambientais e sociais em constante evolução (Prasanna et al., 2021; Yang et al., 2020; Shakoor et al., 2017;).

Fenotipagem de alto rendimento

A fenotipagem de alto rendimento, refere-se à adoção de plataformas terrestres e aéreas equipadas com múltiplos sensores para caracterizar plantas de interesse e seu entorno, sendo realizada através da aplicação de ferramentas como imagens terrestres, fenotipagem aérea e sensoriamento remoto (Gill et al., 2022). Esta apresenta aspectos relevantes em comparação com a fenotipagem tradicional, uma vez que a partir do uso de imagens e sensores, auxiliados com softwares e algoritmos de inteligência computacional, otimizam o processo de medição, possibilita maior fenotipagem em escala e reduz a variabilidade da mensuração humana (Yang et al., 2020; Xie & Yang, 2020).

A utilização de índices de vegetação está entre os principais meios utilizados na fenotipagem de alto rendimento (Xie & Yang, 2020). Utilizando Índices de Vegetação (IVs), é possível realizar uma variedade de estudos, incluindo a avaliação do teor de nitrogênio nas folhas, a determinação de características físicas como biomassa, altura da planta e área foliar, a análise da heterogeneidade das plantas no campo, a estimativa do teor de clorofila, a avaliação do teor de água nas plantas, a quantificação do teor de lignina, e a detecção de danos causados por pragas e

doenças (Tayade et al., 2022).

Os índices de vegetação são baseados em dados gerados pela leitura de diversas bandas de reflectância por sensores. O Índice de Vegetação de Diferença Normalizada (NDVI) é o mais amplamente utilizado apresentando uma resposta sensível à vegetação verde mesmo em áreas com cobertura vegetal baixa. No entanto, o NDVI é sensível aos efeitos do brilho do solo, cor do solo, atmosfera, nuvens e sombras de nuvens, além de sombras do dossel foliar, e requer calibração por sensoriamento remoto (Xue & Su, 2017).

Uma vez que o NDVI, sofre dessas limitações, o SAVI (Índice de Vegetação Ajustado ao Solo), foi desenvolvido por Liu & Huete (1995) para possibilitar uma análise abrangente da vegetação em condições que levam em consideração tanto as características do solo quanto da atmosfera. O SAVI modificado, ou Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), consegue reduzir com mais eficiência os efeitos de solo descoberto, sendo mais utilizado quando no início do plantio e em menores densidades de plantio (Xue & Su, 2017). Além disso, o Índice de vegetação de Diferença Normalizada Verde (GNDVI) foi desenvolvido para apresentar maior resistência aos efeitos da atmosfera (Gitelson et al., 1996). Enquanto que o Índice de Vegetação Aprimorado (EVI), apresenta melhores respostas à altas densidades de vegetação (Zhao et al., 2021). Já outros índices de vegetação como Índice de Diferença Normalizada de Borda Vermelha (NDRE) e Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI), apresentam melhores sensibilidades como indicadores de clorofila e N foliar (Sumner et al., 2021; Li et al., 2014).

Com o uso de técnicas de fenotipagem em larga escala, torna-se crescente volume e a complexidade dos dados obtidos, aumentando a exigência para abordagens inovadoras para análise dos dados com foco em seleção eficiente. Nesse contexto, a aplicação de inteligência computacional tem emergido como uma ferramenta essencial para transformar a forma como se aborda a fenotipagem e o melhoramento de plantas (Shakoor et al., 2017).

Inteligência computacional

De todos os algoritmos de inteligência computacional, o uso de redes neurais artificiais (RNAs) se destaca quando na etapa de fenotipagem, sendo este o principal algoritmo utilizado para este fim (Van Klompenburg et al., 2020). Quando aplicadas à

fenotipagem, RNAs podem ser utilizadas na predição de produção, conteúdo de componentes bioquímicos em sementes, tolerância à salinidade, e de características morfológicas como altura de planta e volume de madeira (Van Klompenburg et al., 2020; Niazian & Niedbała, 2020; Osco et al., 2020; Bhering et al., 2015).

Uma RNA é um modelo computacional inspirado no funcionamento do cérebro humano, projetado para realizar tarefas de aprendizado e reconhecimento de padrões. A RNA organiza-se em camadas, sendo a camada de entrada responsável por receber os dados, as camadas intermediárias processam informações e a camada de saída produz a resposta final (Figura 1). A conexão entre neurônios possui um peso associado, ajustado durante o treinamento com o objetivo de minimizar erros. O aprendizado ocorre por meio da apresentação repetida de exemplos ao modelo, permitindo que a RNA ajuste seus pesos para otimizar o desempenho na tarefa específica (Worden et al., 2023; Cruz & Nascimento, 2018).

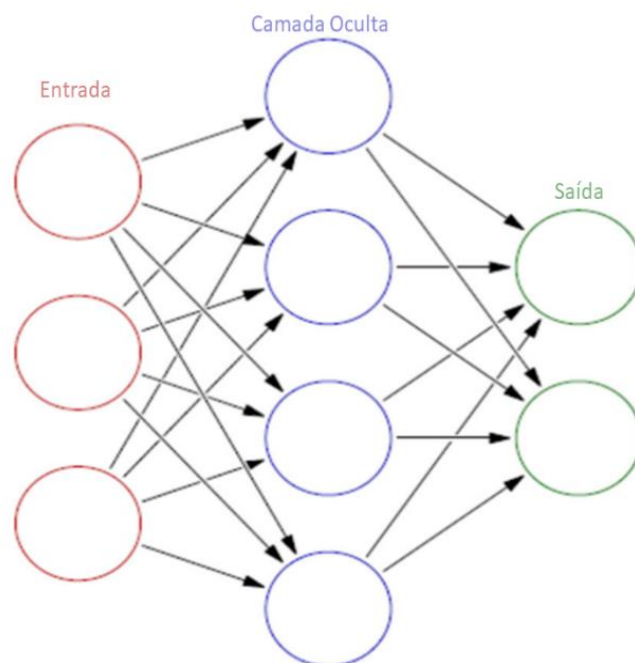


Figura 1: Rede Neural Artificial com uma camada de entrada (vermelho), uma oculta e uma de saída (verde). Adaptado de Wu & Feng (2018).

Segundo Wu & Feng (2018), as características e superioridade das Redes Neurais Artificiais são principalmente manifestadas em três aspectos: sua capacidade de autoaprendizado, eficiente armazenamento e rápida busca por soluções ótimas em

problemas complexos. O método de treinamento de uma RNA, frequentemente empregando o algoritmo de *backpropagation*, visa ajustar os pesos das conexões entre os neurônios para minimizar a diferença entre o valor estimado pelo modelo e o valor real dos dados de treinamento (Islam et al., 2019). O treinamento ocorre ao longo de épocas, o que define o número de vezes que o conjunto de treinamento é apresentado à rede (Rawat et al., 2013). A escolha de uma função de ativação, como tangente hiperbólica ou logística, influencia a capacidade da RNA em aprender relações complexas (Sharma & Athaiya, 2020). O número de neurônios em cada camada e o passo, que determina o tamanho do ajuste dos pesos, são parâmetros que impactam a capacidade da RNA em representar padrões (Worden et al., 2023; Kruse et al., 2022). Assim, a otimização desses elementos é essencial para maximizar o desempenho da rede em tarefas específicas.

As Perceptrons Multicamadas (PMC) ou MLP, Multilayer Perceptrons, representam uma evolução das redes neurais simples, sendo compostas por camadas de neurônios interconectados, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio nas camadas ocultas e de saída utiliza funções de ativação não-lineares, permitindo que as PMC capturem relações complexas nos dados. Essa arquitetura mais complexa confere às PMC a capacidade de aprender representações hierárquicas e realizar tarefas mais sofisticadas, como reconhecimento de padrões e classificação em contextos diversos, tornando-as um componente essencial em aplicações de aprendizado de máquina (Kruse et al., 2022).

O Random Forest representa um algoritmo de aprendizado de máquina, geralmente utilizado para tarefas de classificação e predição. O funcionamento dessa técnica envolve a construção de múltiplas árvores de decisão (Costa & Pedreira, 2022) durante o treinamento. Durante a fase de predição, as árvores geram suas próprias predições, que são combinadas para chegar a um resultado final. Essa abordagem de conjunto permite que o algoritmo de Random Forest capture uma variedade mais ampla de padrões nos dados, melhore a robustez e reduza o risco de overfitting, tornando-a uma escolha eficaz para uma variedade de aplicações de aprendizado de máquina (Biau & Scornet, 2016).

Com base no exposto, este trabalho tem como objetivo avaliar o desempenho de algoritmos de Perceptron Multicamadas (PMC) e Random Forest (RF) na predição de características fenotípicas de soja utilizando índices de vegetação. Além disso,

busca-se investigar o impacto dos parâmetros das redes neurais no desempenho de um PMC e como esses parâmetros influenciam sua eficácia na predição de características fenotípicas de soja.

REFERÊNCIAS

BATTISTI, R.; SENTEHLAS, P. C.; BOOTE, K. J.; CÂMARA, G. M. de S.; FARIAS, J. R. B.; BASSO, C. J. Assessment of soybean yield with altered water-related genetic improvement traits under climate change in Southern Brazil. **Eur. J. Agron.**, v. 83, p. 1-14, 2017. ISSN: 1161-0301. doi: 10.1016/j.eja.2016.11.004.

BIAU, G.; SCORNET, E. A random forest guided tour. **TEST**, v. 25, p. 197–227, 2016. doi: 10.1007/s11749-016-0481-7.

BHERING, L. L.; CRUZ, C. D.; PEIXOTO, L. de A.; ROSADO, A. M.; LAVIOLA, B. G.; NASCIMENTO, M. Application of neural networks to predict volume in eucalyptus. **Crop. Breed. Appl. Biotechnol.**, v. 15, n. 3, p. 125–131, 2015. doi: 10.1590/1984-70332015v15n3a23.

BODIRSKY, B. L.; ROLINSKI, S.; BIEWALD, A.; WEINDL, I.; POPP, A.; LOTZE-CAMPEN, H. Global Food Demand Scenarios for the 21st Century. **PLoS ONE**, v. 10, n. 11, e0139201, 2015. doi: 10.1371/journal.pone.0139201.

COSTA, V. G.; PEDREIRA, C. E. Recent advances in decision trees: an updated survey. **Artif. Intell. Rev.**, v. 56, p. 4765–4800, 2023. doi: 10.1007/s10462-022-10275-5.

CRUZ, C.; NASCIMENTO, M. 2018. **Inteligência Computacional Aplicada ao Melhoramento Genético**. Editora UFV, Viçosa, 414p. ISBN: 9788572696067.

FODOR, N.; CHALLINOR, A.; DROUTSAS, I.; RAMIREZ-VILLEGAS, J.; ZABEL, F.; KOEHLER, A.-K.; FOYER, C. H. Integrating Plant Science and Crop Modeling: Assessment of the Impact of Climate Change on Soybean and Maize Production.

Plant Cell Physiol., v. 58, n. 11, p. 1833–1847, November 2017. doi: 10.1093/pcp/pcx141.

GILL, T.; GILL, S. K.; SAINI, D. K. et al. A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping. **Phenomix**, p. 156–183, 2022. doi: 10.1007/s43657-022-00048-z.

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sens. Environ.**, v. 58, n. 3, p. 289-298, 1996. doi: 10.1016/S0034-4257(96)00072-7.

GUO, S.; ZHANG, Z.; GUO, E.; FU, Z.; GONG, J.; YANG, X. Historical and projected impacts of climate change and technology on soybean yield in China. **Agric. Syst.**, v. 203, 2022, p. 103522. ISSN: 0308-521X. doi: 10.1016/j.agsy.2022.103522.

IIZUMI, T.; YOKOZAWA, M.; SAKURAI, G.; TRAVASSO, M. I.; ROMANENKOV, V.; OETTLI, P.; NEWBY, T.; ISHIGOOKA, Y.; FURUYA, J. Historical changes in global yields. **Glob. Ecol. Biogeogr.**, v. 23, p. 346-357, 2014. doi: 10.1111/geb.12120.

ISLAM, M.; CHEN, G.; JIN, S. An Overview of Neural Network. **Am. J. Neural Netw. Appl.**, v. 5, n. 1, p. 7-11, 2019. doi: 10.11648/j.ajinna.20190501.12.

KRUSE, R.; BORGELT, C.; BRAUNE, C.; MOSTAGHIM, S.; STEINBRECHER, M. **Computational intelligence**. Springer, 2022. doi: 10.1007/978-3-030-42227-1.

PRASANNA, B. M.; CAIRNS, J. E.; ZAIDI, P. H. Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. **Theor. Appl. Genet.**, v. 134, p. 1729–1752, 2021. doi: 10.1007/s00122-021-03773-7.

LI, F.; MIAO, Y.; FENG, G.; YUAN, F.; YUE, S.; GAO, X.; LIU, Y.; LIU, B.; USTIN, S. L.; CHEN, X. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. **Field Crops Res.**, v. 157, p. 111-123, 2014. ISSN 0378-4290. doi: 10.1016/j.fcr.2013.12.018.

LI, Y.-H.; LI, W.; ZHANG, C.; YANG, L.; CHANG, R.-Z.; GAUT, B.S.; QIU, L.-J. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. **New Phytol.**, v. 188, p. 242-253, 2010. <https://doi.org/10.1111/j.1469-8137.2010.03344.x>

LIU, H. Q.; HUETE, A. A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. **IEEE Trans. Geosci Remote Sens.**, v. 33, n. 2, p. 457-465, 1995. doi: 10.1109/TGRS.1995.8746027.

LIU, X.; WU, J. A.; REN, H.; QI, Y.; LI, C.; CAO, J.; ZHANG, X.; ZHANG, Z.; CAI, Z.; GAI, J. Genetic variation of world soybean maturity date and geographic distribution of maturity groups. **Breed Sci.** v. 67, n.3, p. 221-232, 2017. doi: 10.1270/jsbbs.16167.

LIU, Z.; LI, H.; WEN, Z.; FAN, X.; LI, Y.; GUAN, R.; GUO, Y.; WANG, S.; WANG, D.; QIU, L.; Comparison of genetic diversity between chinese and american soybean (*glycine max* (l.)) Accessions Revealed by High-Density SNPs. **Front. Plant Sci.**, v. 8, p. 2014, 2017. doi: 10.3389/fpls.2017.02014

MODGIL, R.; TANWAR, B.; GOYAL, A.; KUMAR, V. Soybean (*Glycine max*). In: TANWAR, B.; GOYAL, A. (eds). **Oilseeds: Health Attributes and Food Applications**. Springer, Singapore, 2021. doi: 10.1007/978-981-15-4194-0_1.

NIAZIAN, M.; NIEDBAŁA, G. Machine Learning for Plant Breeding and Biotechnology. **Agriculture**, v. 10, p. 436, 2020. doi: 10.3390/agriculture10100436.

OSCO, L. P.; JUNIOR, J. M.; RAMOS, A. P. M.; FURUYA, D. E. G.; SANTANA, D. C.; TEODORO, L. P. R.; GONÇALVES, W. N.; BAIO, F. H. R.; PISTORI, H.; JUNIOR, C. A. d. S. et al. Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. **Remote Sens.**, v. 12, p. 3237, 2020. doi: 10.3390/rs12193237.

RAWAT, R.; PATEL, J. K.; MANRY, M. T. **Minimizing validation error with respect to network size and number of training epochs**. In: The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 2013, pp. 1-7. doi:

10.1109/IJCNN.2013.6706919.

SEDIVY, E.J.; WU, F.; HANZAWA, Y. Soybean domestication: the origin, genetic architecture and molecular bases. **New Phytol**, v. 214, p. 539-553, 2017. <https://doi.org/10.1111/nph.14418>

SILVA, E. H. F. M.; ANTOLIN, L. A. S.; ZANON, A. J.; ANDRADE, A. S.; SOUZA, H. A.; CARVALHO, K. S.; VIEIRA, N. A.; MARIN, F. R. Impact assessment of soybean yield and water productivity in Brazil due to climate change. **Eur. J. Agron.**, v. 129, p. 126329, 2021. doi: 10.1016/j.eja.2021.126329.

SHAKOOR, N., LEE, S., MOCKLER, T. C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. **Curr. Opin. Plant Biol.**, v. 38, 2017, P. 184-192. ISSN: 1369-5266. doi: 10.1016/j.pbi.2017.05.006.

SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **Int. J. Eng. Appl. Sci. Technol.**, v. 4, n. 12, p. 310-316, 2020. doi: 10.33564/IJEAST.2020.v04i12.054.

SMÝKAL, P.; COYNE, C. J.; AMBROSE, M. J.; MAXTED, N.; SCHAEFER, H.; BLAIR, M. W.; BERGER, J.; GREENE, S. L.; NELSON, M. N.; BESHARAT, N.; VYMYSLICKÝ, T.; TOKER, C.; SAXENA, R. K.; ROORKIWAL, M.; PANDEY, M. K.; HU, J.; LI, Y. H.; WANG, L. X.; GUO, Y.; QIU, L. J. Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. **Crit. Rev. Plant Sci.**, v. 34, p. 43–104, 2014 <https://doi.org/10.1080/07352689.2014.897904>

SUMNER, Z.; VARCO, J. J; DHILLON, J. S; FOX, A. A. A; CZARNECKI, J; HENRY, W. B. Ground versus aerial canopy reflectance of corn: red-edge and non-red edge vegetation indices. **Agron. J.**, v. 113, p. 2773-2788, 2021. doi: 10.1002/agj2.20657.

TAYADE, R.; YOON, J.; LAY, L.; KHAN, A.L.; YOON, Y.; KIM, Y. Utilization of Spectral Indices for High-Throughput Phenotyping. **Plants**, v. 11, p. 1712, 2022. <https://doi.org/10.3390/plants11131712>.

UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. **World Agricultural**

Production – Circular Series, dez 2023. USDA. 2023.

VAN DIJK, M.; MORLEY, T.; RAU, M. L. A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. **Nat. Food**, v. 2, p. 494–501, 2021. doi: 10.1038/s43016-021-00322-9.

VAN KLOMPENBURG, T.; KASSAHUN, A.; CATAL, C. Crop yield prediction using machine learning: A systematic literature review. **Comput. Electron. Agric.**, v. 177, 2020, p. 105709. ISSN 0168-1699. doi: 10.1016/j.compag.2020.105709.

WORDEN, K.; TSIALIAMANIS, G.; CROSS, E. J.; ROGERS, T. J. Artificial Neural Networks. In: RABCZUK, T.; BATHE, KJ. **Machine Learning in Modeling and Simulation**. Computational Methods in Engineering & the Sciences. Springer, Cham, 2023. doi: 10.1007/978-3-031-36644-4_2.

WU, Y. C.; FENG, J. W. Development and Application of Artificial Neural Network. **Wirel. Pers. Commun.**, v. 102, p. 1645–1656, 2018. doi: 10.1007/s11277-017-5224-x.

XIE, C.; YANG, C. A review on plant high-throughput phenotyping traits using UAV-based sensors. **Comput. Electron. Agric.**, v. 178, p. 105731, 2020. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2020.105731>

XUE, J.; SU, B. Significant Remote Sensing Vegetation Indices: A review of Developments and Applications. **J. Sens.**, v. 2017, p.17, 2017. <https://doi.org/10.1155/2017/1353691>

YANG, W., FENG, H., ZHANG, X., ZHANG, J., DOONAN, J. H., BATCHELOR, W. D., XIONG, L. e YAN, J. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. **Mol. Plant**, v. 13, p. 187–214, 2020.

ZHAO, H., LI, Y., CHEN, X. et al. Monitoring monthly soil moisture conditions in China with temperature vegetation dryness indexes based on an enhanced vegetation index

and normalized difference vegetation index. **Theor Appl Climatol**, v. 143, p. 159–176, 2021. doi: 10.1007/s00704-020-03422-x.

CAPÍTULO 1 - TOPOLOGIA E PARÂMETROS DE MODELOS DE PERCEPTON MULTICAMADAS NA FENOTIPAGEM DE SOJA A PARTIR DE IMAGENS MULTIESPECTRAIS

RESUMO

A soja (*Glycine max*) é uma importante fonte de alimento para humanos e animais, além de ser utilizada como matéria-prima na indústria. Para atender às demandas crescentes, a soja enfrenta uma série de desafios complexos. Esses desafios estão intrinsecamente ligados à busca por variedades mais produtivas, resistentes a doenças e adaptadas a condições ambientais variáveis. A fenotipagem de alto rendimento emerge como uma ferramenta crucial nesse processo, acelerando o desenvolvimento eficiente de novas variedades mais produtivas, capazes de enfrentar os desafios ambientais e sociais em constante evolução. Esta oferece vantagens quando na comparação com a fenotipagem tradicional, uma vez que a partir do uso de imagens e sensores, auxiliados com softwares e algoritmos de inteligência computacional, otimizam o processo de medição, possibilita maior fenotipagem em escala e reduz a variabilidade da mensuração humana. A utilização de índices de vegetação está entre os principais meios utilizados na fenotipagem de alto rendimento. Com a utilização destes, é possível realizar uma variedade de estudos, incluindo a avaliação do teor de nitrogênio nas folhas, a determinação de características físicas como biomassa, altura da planta e área foliar, a análise da heterogeneidade das plantas no campo, a estimativa do teor de clorofila, a avaliação do teor de água nas plantas, a quantificação do teor de lignina, e a detecção de danos causados por pragas e doenças. Com o uso de técnicas de fenotipagem em larga escala, torna-se crescente volume e a complexidade dos dados obtidos, aumentando a exigência para abordagens inovadoras para análise dos dados com foco em seleção eficiente. Nesse contexto, a aplicação de inteligência computacional tem emergido como uma ferramenta essencial para transformar a forma como se aborda a fenotipagem e o melhoramento de plantas. Assim, este trabalho tem como objetivo avaliar o impacto dos parâmetros das redes neurais no desempenho de um Perceptron Multicamadas (PMC) e como esses parâmetros influenciam sua eficácia na predição de características fenotípicas de soja a partir dos seguintes índices de vegetação: Índice de vegetação de Diferença Normalizada (NDVI), Índice de Vegetação Ajustado

ao Solo (SAVI), Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), Índice de vegetação de Diferença Normalizada Verde (GNVDI), Índice de Vegetação Aprimorado (EVI), Índice de Diferença Normalizada de Borda Vermelha (NDRE) e Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI). Constatou-se que, para os modelos de PMC, conforme a quantidade de neurônios aumentava até 10, e o número de folds na validação cruzada aumentava até 15, os modelos apresentavam resultados progressivamente melhores. Por outro lado, o aumento do número de épocas resultava em um aumento nos valores de R^2 dos modelos, alcançando um limite de 30000, após o qual os resultados começavam a diminuir. Ademais, o acréscimo no número de camadas ocultas ocasionava uma redução nos valores dos coeficientes de determinação, indicando que o melhor número de camadas foi um.

Palavras-chave: Predição, Aprendizado de Máquina, Índices de Vegetação, *Glycine max (L)*.

CHAPTER 1 - TOPOLOGY AND PARAMETERS OF MULTILAYER PERCEPTRON MODELS IN SOYBEAN PHENOTYPING USING MULTISPECTRAL IMAGES

ABSTRACT

Soybean (*Glycine max*) is a valuable source of food for humans and animals, as well as a key raw material in industry. To meet growing demands, soybean faces a series of complex challenges. These challenges are intrinsically linked to the pursuit of more productive varieties, resistant to diseases, and adapted to variable environmental conditions. High-throughput phenotyping emerges as a crucial tool in this process, accelerating the efficient development of new, more productive varieties capable of addressing constantly evolving environmental and social challenges. It offers advantages compared to traditional phenotyping, as it optimizes the measurement process through the use of images and sensors, aided by software and computational intelligence algorithms, enabling greater scaling of phenotyping and reducing human variability. The use of vegetation indices is among the main methods used in high-throughput phenotyping. With their use, a variety of studies can be conducted, including the evaluation of nitrogen content in leaves, the determination of physical characteristics such as biomass, plant height, and leaf area, the analysis of plant heterogeneity in the field, the estimation of chlorophyll content, the assessment of water content in plants, lignin content quantification, and the detection of damage caused by pests and diseases. With the use of large-scale phenotyping techniques, the volume and complexity of the data obtained are increasing, increasing the demand for innovative approaches to data analysis focused on efficient selection. In this context, the application of computational intelligence has emerged as an essential tool to transform the way phenotyping and plant breeding are approached. Thus, this work aims to evaluate the impact of neural network parameters on the performance of a Multilayer Perceptron (MLP) and how these parameters influence its effectiveness in predicting soybean phenotypic characteristics using the following vegetation indices: Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Modified Soil-Adjusted Vegetation Index (MSAVI), Green Normalized Difference Vegetation Index (GNVDI), Enhanced Vegetation Index (EVI), Normalized Difference Red Edge Index (NDRE), and Simplified Canopy Chlorophyll Content Index (SCCCI). It was found that for the MLP models, as the number of neurons increased

up to 10, and the number of folds in cross-validation increased up to 15, the models showed progressively better results. On the other hand, an increase in the number of epochs resulted in an increase in the R^2 values of the models, reaching a limit of 30,000, after which the results began to decrease. In addition, an increase in the number of hidden layers resulted in a reduction in the values of the determination coefficients, indicating that the optimal number of layers was one.

Keywords: Prediction, Machine Learning, Vegetation Indices, *Glycine max (L)*.

INTRODUÇÃO

A soja foi domesticada entre seis e nove mil anos no leste asiático a partir da soja selvagem, com a China apresentando a maior diversidade genética. Hoje, é uma importante fonte de alimento humano, animal e matéria-prima industrial, sendo cultivada em todo o mundo, com Brasil, EUA e Argentina como os principais produtores (USDA, 2023; Modgil et al., 2021; Sedivy et al., 2017; Li et al., 2010). Apesar de ser crucial para atender às crescentes demandas por alimentos e recursos sustentáveis, o melhoramento da soja enfrenta uma série de desafios. Estes desafios estão intimamente relacionados à busca por variedades mais produtivas, resilientes a doenças e capazes de se adaptar a condições ambientais variáveis (Fodor et al., 2017; Battisti et al., 2017). Assim, é crescente a busca pelo rápido desenvolvimento de variedades adaptadas às condições adversas, utilizando técnicas avançadas, como a fenotipagem de alto rendimento, para aprimorar os programas de melhoramento (Prasanna et al., 2021).

Os Índices de Vegetação (IVs) são amplamente utilizados no contexto da fenotipagem de alto rendimento, uma técnica que envolve plataformas terrestres e/ou aéreas equipadas com múltiplos sensores para analisar as características das plantas e seu ambiente (Tayade et al., 2022). A fenotipagem de alto rendimento otimiza o processo de medição através do uso de imagens e sensores, juntamente com softwares e algoritmos de inteligência computacional, resultando em uma análise em larga escala mais eficiente e com redução da influência da subjetividade humana (Yang et al., 2020; Xie & Yang, 2020).

Dentre os índices de vegetação, o Índice de Vegetação de Diferença Normalizada (NDVI) é amplamente utilizado, porém sensível a vários efeitos ambientais. Em resposta a essas limitações, foram desenvolvidos o Índice de Vegetação Ajustado ao Solo (SAVI) e o Índice de Vegetação Ajustado ao Solo Modificado (MSAVI) para considerar características específicas do solo e da atmosfera, enquanto o Índice de Vegetação de Diferença Normalizada Verde (GNVDI) apresenta maior resistência aos efeitos atmosféricos (Xue & Su, 2017; Gitelson et al., 1996; Liu & Huete, 1995;). O Índice de Vegetação Aprimorado (EVI) é particularmente eficaz em altas densidades de vegetação (Zhao et al., 2021). Além disso, o Índice de Diferença Normalizada de Borda Vermelha (NDRE) e o Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI) são preferidos para avaliação de clorofila e

teor de nitrogênio foliar, respectivamente (Sumner et al., 2021; Li et al., 2014;).

Com o uso de técnicas de fenotipagem em larga escala, torna-se crescente volume e a complexidade dos dados obtidos, aumentando a exigência para abordagens inovadoras para análise dos dados com foco em seleção eficiente. Nesse contexto, a aplicação de inteligência computacional tem emergido como uma ferramenta essencial para transformar a forma como se aborda a fenotipagem e o melhoramento de plantas (Shakoor et al., 2017).

Os algoritmos de inteligência computacional, o uso de redes neurais artificiais (RNAs) se destaca quando na etapa de fenotipagem, sendo este o principal algoritmo utilizado para este fim (Van Klompenburg et al., 2020). Quando aplicadas à fenotipagem, RNAs podem ser utilizadas na predição de produção, conteúdo de componentes bioquímicos em sementes, tolerância à salinidade, e de características morfológicas como altura de planta e volume de madeira (Van Klompenburg et al., 2020; Niazián & Niedbała, 2020; Osco et al., 2020; Bhering et al., 2015).

Os Perceptrons Multicamadas (PMC) ou MLP, abreviação de Multilayer Perceptrons, são uma progressão das redes neurais simples, constituídas por várias camadas de neurônios conectados, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (Cruz & Nascimento, 2018). Cada neurônio nessas camadas utiliza funções de ativação não-lineares, possibilitando que os PMC capturem relações complexas nos dados. Essa estrutura mais elaborada confere aos PMC a capacidade de aprender representações hierárquicas e executar tarefas mais avançadas, como reconhecimento de padrões e classificação em diversos contextos, tornando-os componentes fundamentais em aplicações de aprendizado de máquina (Kruse et al., 2022).

A correta determinação da topologia e dos parâmetros de um Perceptron Multicamadas (PMC) é essencial para o desempenho deste. A escolha adequada do número de camadas ocultas, bem como o número de neurônios em cada camada, tem um impacto significativo na capacidade da rede em aprender e representar padrões complexos nos dados (Worden et al., 2023; Kruse et al., 2022). Uma topologia inadequada pode resultar em problemas como *overfitting*, onde a rede se ajusta muito aos dados de treinamento, mas não generaliza para novos dados, ou *underfitting*, onde a rede não é capaz de capturar com precisão a complexidade dos dados (Vujicic et al., 2016). Além disso, a seleção das funções de ativação e a configuração dos parâmetros de treinamento, como a taxa de aprendizado e o número de épocas,

também desempenham um papel crucial no desempenho geral do PMC (Sharma & Athaiya, 2020; Rawat et al., 2013;). Considerações sobre o número de folds em uma validação cruzada também são importantes, pois influenciam na robustez da avaliação do desempenho do modelo (Kärkkäinen, 2014). Portanto, uma análise cuidadosa e uma abordagem sistemática na determinação da topologia e dos parâmetros são fundamentais para garantir um desempenho eficaz e uma boa capacidade de generalização da rede em uma variedade de problemas de aprendizado de máquina.

No entanto, persiste uma lacuna substancial no entendimento do impacto dos parâmetros, incluindo o número de neurônios e camadas ocultas, em Perceptrons Multicamadas (PMC) quando aplicados a conjuntos de dados reais obtidos para fins de fenotipagem e integrados com índices de vegetação. Em grande parte, os estudos nessa área se concentram na capacidade dos modelos em previsão ou classificação, comparando modelos com diferenças significativas. Porém, há uma escassez de pesquisas que examinem modelos com pequenas variações entre si, visando avaliar o impacto dessas pequenas diferenças nos modelos, a fim de analisar os efeitos das mudanças nos parâmetros, uma vez que conhecer o impacto de cada parâmetro de um PMC é essencial na construção do melhor modelo (Uzair & Jamil, 2020; Ibnu et al., 2019; Andersen & Martinez, 1999).

Portanto, os objetivos deste estudo consistem em avaliar o impacto do número de neurônios, camadas ocultas, épocas de treinamento e folds da cross-validação, bem como a interação entre esses fatores, sobre o desempenho de um Perceptron Multicamadas (PMC) com finalidade de prever valores fenotípicos de soja a partir de dados de índices de vegetação.

MATERIAIS E MÉTODOS

Todas as análises foram realizadas no software Rbio® juntamente com o R (R core Team, 2024; Bhering, 2017).

Descrição do experimento

Os experimentos com a cultura da soja foram conduzidos na área experimental da Universidade Federal de Mato Grosso do Sul, Campus de Chapadão do Sul

(localizado em 18°46'26"S, 52°37'28"W, com uma altitude média de 810 m), o método de preparo do solo adotado foi convencional.

Foram analisados 32 genótipos, EM quatro blocos, submetidos a avaliação por meio de um delineamento de blocos casualizados nas safras de 19/20, 20/21 e 21/22.

As características agrícolas avaliadas incluem: dias para maturação (DM), altura da inserção da primeira vagem (AIV, cm), altura de plantas (AP, cm) e produção de grãos (PROD, kg ha⁻¹). O DM foi determinado a partir da contagem dos dias entre a emergência e a maturidade de mais da metade das plantas em cada unidade experimental. AIV e AP foram medidas em cinco plantas por parcela, utilizando uma fita métrica. A produção de grãos foi determinada pela colheita dos dois metros centrais de cada parcela e ajuste para umidade em 13%.

Os índices de vegetação utilizados (Tabela 1) foram obtidos no florescimento pleno da cultura (Ramos et al., 2020; Osco et al., 2020).

Tabela 1: Índices de vegetação utilizados e seus métodos de obtenção

Índice de vegetação	Método de obtenção
NDVI	Xue & Su, 2017
SAVI	Liu & Huete, 1995
MSAVI	Xue & Su, 2017
GNVDI	Gitelson et al., 1996
EVI	Zhao et al., 2021
NDRE	Li et al., 2014
SCCCI	Sumner et al., 2021

Índice de vegetação de Diferença Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), Índice de vegetação de Diferença Normalizada Verde (GNVDI), Índice de Vegetação Aprimorado (EVI), Índice de Diferença Normalizada de Borda Vermelha (NDRE), Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI).

Foi utilizada uma aeronave remotamente pilotada (RPA) de asa fixa, a Sensefly eBee RTK, com capacidade de decolagem, voos e aterrissagens autônomas. O eBee foi equipado com o sensor multiespectral Sensefly Sequoia, e realizou a captura da reflectância dos comprimentos de onda previamente mencionados. Além disso, esse sensor possui uma tecnologia de medição de luminosidade para a calibração dos dados adquiridos. O levantamento aéreo foi feito com a tecnologia RTK, garantindo precisão de posicionamento da câmera no momento da coleta da imagem em até 2,5 cm. As imagens foram processadas com o uso do software Pix4Dmapper, criando

mosaicos e correções ortográficas. A precisão posicional foi avaliada através de pontos de controle em campo, medidos com o RTK.

Inteligência Computacional

A predição das variáveis fenotípicas foi baseada em PMC conforme tabela 2, com combinações de um, três, cinco e dez neurônios com uma, duas ou três camadas ocultas, cinco, 10 ou 15 folds utilizados na validação cruzada e 10000, 30000 e 50000 épocas, com *Backpropagation* combinado com a função de ativação do tipo logística através pacote neuralnet (Fritsch et al., 2019).

Tabela 2: Configuração gerais dos Modelos de Perceptron Multicamadas (PMC) utilizados

Número de camadas	1, 2 e 3
Número de Neurônios por camada	1, 3, 5 e 10
Função de ativação	Logística
Função de treinamento	<i>Backpropagation</i>
Método de Avaliação	R ²
Número de épocas de treinamento	10000, 30000 e 50000
Número de dados	384 linhas
Tipo de treinamento	Validação Cruzada
Número de folds/repetições	5, 10 e 15 / 2 Repetições

As variáveis de entrada (Variáveis X) foram as leituras de comprimento de ondas e/ou os índices de vegetação, enquanto que a variável de saída (Variável Y) foi a variável fenotípica predita.

O sucesso da rede neural foi testado comparando-se o coeficiente de determinação (R²) de cada modelo avaliado.

Uma vez captados, os dados relativos às variáveis Y foram normalizados pela técnica de Min-Max normalization (Cabello-Solorzano et al., 2023).

RESULTADOS

Efeito do número de folds em relação ao número de neurônios por camada oculta

Uma camada oculta

De maneira geral, a resposta do modelo mediante as alterações no número de folds se manteve constante para as diferentes variáveis (Tabela 3 e Figura 2).

Tabela 3: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos com uma camada oculta.

Variável	Folds	Neurônios			
		1	3	5	10
AIV	5	53.38	61.24	61.06	68.57
	10	33.99	60.89	69.53	70.68
	15	44.76	69.94	69.18	70.15
AP	5	60.48	80.04	79.82	79.57
	10	66.61	81.71	79.69	80.25
	15	60.20	78.67	80.96	80.89
DM	5	6.17	9.67	12.80	11.74
	10	8.59	12.48	10.60	14.78
	15	9.60	13.07	15.97	11.56
PROD	5	14.77	24.29	23.17	26.14
	10	13.69	24.12	24.77	24.11
	15	20.96	28.65	24.95	27.42

AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

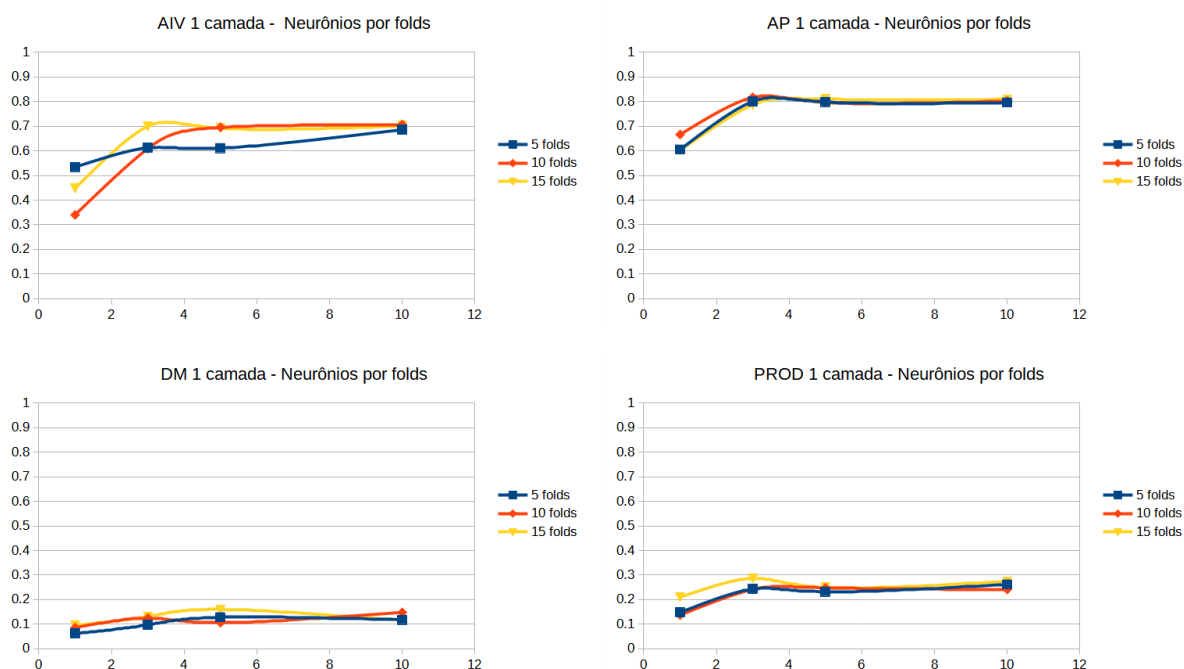


Figura 2: Estimativa do R^2 obtidos pelos modelos nas diferentes configurações para uma camada oculta, com diferentes números de neurônios e números de folds. AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

A variável altura de inserção de vagem apresentou maior variação dentro da relação entre folds e neurônios. Para o primeiro neurônio, o menor coeficiente de determinação foi observado com 10 folds, atingindo 33.99%, seguido de 15 folds com 44.76%, e 53.38% para cinco folds. No caso de três neurônios, os resultados foram de 61.24% para cinco folds e 60.89% para 10 folds. O ponto alto foi registrado com 15 folds, alcançando 69.94%, representando o máximo do coeficiente de determinação dentro da relação entre neurônios e folds. Ao considerar cinco neurônios, os modelos com 10 e 15 folds alcançaram resultados semelhantes, em torno de 69%, enquanto o modelo com cinco folds apresentou um desempenho inferior com 61.06%. Para 10 neurônios, todos os números de folds estudados apresentaram resultados próximos, variando entre 68.57% para cinco folds e 70.85% para 10 folds.

A variável que apresentou maior predictibilidade foi altura de planta, iniciando com coeficiente de determinação de aproximadamente 60 % para um neurônio na camada oculta, tendo seu pico com três neurônios, com um R^2 de 81,71 para 10 folds, e se mantendo relativamente constante entre 79.82% e 80.89% até 10 neurônios entre cinco e 15 folds. Para esta variável, única etapa onde houve mais de 5% de diferença entre os folds foi no primeiro neurônio, onde para cinco folds o modelo apresentou

60.48% e para 10 folds apresentou 66.61% de coeficiente de determinação.

Para dias de maturação, os R^2 se mantiveram semelhantes, com a maior variação sendo para três neurônios, onde para cinco folds foi apresentado R^2 de 9.67%, 10 folds 12.48% e 15 folds 13.07%. Para cinco neurônios, o menor coeficiente de determinação foi para 10 folds, com 10.6% e o maior, sendo este o máximo dentre a relação neurônios x folds para uma camada oculta, 15.97%.

Para produtividade, modelos com 10 e 15 folds se mantiverem semelhantes tendo o mesmo comportamento ao longo dos acréscimos de neurônios, tendo a maior variação para os 10 neurônios, sendo 26.14% de R^2 para cinco folds e 24.11% para 10 folds. Para esta variável os modelos com 15 folds se destacaram dos demais, com 20.96% de R^2 para um neurônio, apresentando um máximo de 28.65% para três neurônios, e em seguida se mantendo semelhante dos demais.

Duas camadas ocultas

Para duas camadas ocultas, a variação do coeficiente de determinação à medida em que se altera o número de folds por neurônios para cada variável se apresentou de maneira semelhante ao que foi apresentado para uma camada oculta.

A variável que apresentou maior alteração ao longo do número de neurônios por folds foi AIV. Os modelos com cinco folds registraram os menores R^2 em todas as etapas, enquanto os modelos com 10 e 15 folds demonstraram resultados semelhantes a partir de três neurônios. O maior coeficiente de determinação obtido ocorreu com 15 folds e cinco neurônios, atingindo 72.12%, o que não diferiu muito dos valores de 70.52 e 70.40 apresentados por 10 folds com cinco e 10 neurônios, respectivamente (Tabela 4 e Figura 3).

Tabela 4: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos com duas camadas ocultas.

Variável	Folds	Neurônios			
		1	3	5	10
AIV	5	31.86	56.85	62.38	64.69
	10	33.69	60.77	70.52	70.40
	15	43.44	62.80	71.12	67.95
AP	5	66.73	82.40	81.10	81.77
	10	57.28	82.14	82.99	82.31
	15	67.59	81.77	81.52	81.35
DM	5	5.98	7.73	11.89	9.88
	10	7.09	9.56	12.70	11.74
	15	8.19	12.08	14.82	13.06
PROD	5	14.87	23.74	24.48	24.51
	10	15.27	25.51	24.09	22.84
	15	17.29	24.93	24.87	26.45

AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

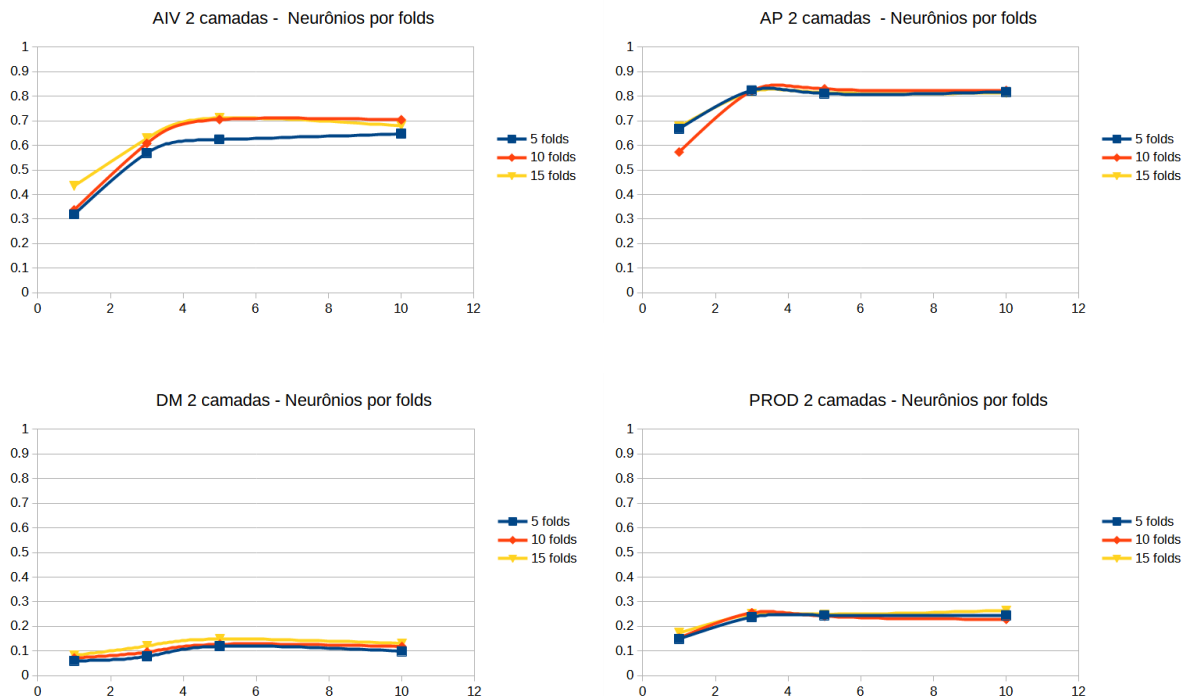


Figura 3: Estimativa do R^2 obtidos pelos modelos nas diferentes configurações para duas camadas ocultas, com diferentes números de neurônios e números de folds. AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

A variável altura de planta apresentou resultados que se diferiram menos de 1% a partir de três neurônios, apresentando a maior diferença entre os R^2 apresentados para um neurônio, com 57.28% para 10 folds, e 67.59% para 10 folds. Para esta variável o pico foi de 82.99%, sendo este para 10 folds e cinco neurônios.

Em relação aos dias para maturação a variação do coeficiente de determinação ao longo da relação neurônios x folds se manteve constante. Seu pico foi de 14.82%, ocorrendo para cinco neurônios e 15 folds. Assim como para uma camada, duas camadas ocultas apresentaram menor R^2 para um neurônio.

A variável produtividade também manteve padrão de variação semelhante para as alterações na relação folds por neurônios, se diferenciando pouco para cada neurônio. Seu pico ocorreu com 26.45% para 15 folds e 10 neurônios, mas sem se diferir mais de 2% para todas as variações dentro de cinco e 10 neurônios.

Três camadas ocultas

Com três camadas ocultas, houve um pouco mais de variação dentro da relação folds por neurônios, mas ainda não se diferenciando muito dos resultados obtidos para uma e duas camadas (Tabela 5 e Figura 4).

Tabela 5: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos com três camadas ocultas.

Variável	Folds	Neurônios			
		1	3	5	10
AIV	5	34.59	52.70	53.53	51.47
	10	30.41	55.34	59.67	55.88
	15	36.32	55.51	55.96	56.94
AP	5	49.14	76.46	81.26	80.65
	10	55.09	81.25	81.35	80.77
	15	47.94	78.75	78.96	79.75
DM	5	3.19	7.57	6.83	6.32
	10	6.91	10.12	11.64	12.88
	15	6.99	9.89	10.43	14.52
PROD	5	13.46	23.06	21.98	20.12
	10	10.87	21.62	21.14	20.54
	15	13.39	24.82	26.39	22.11

AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

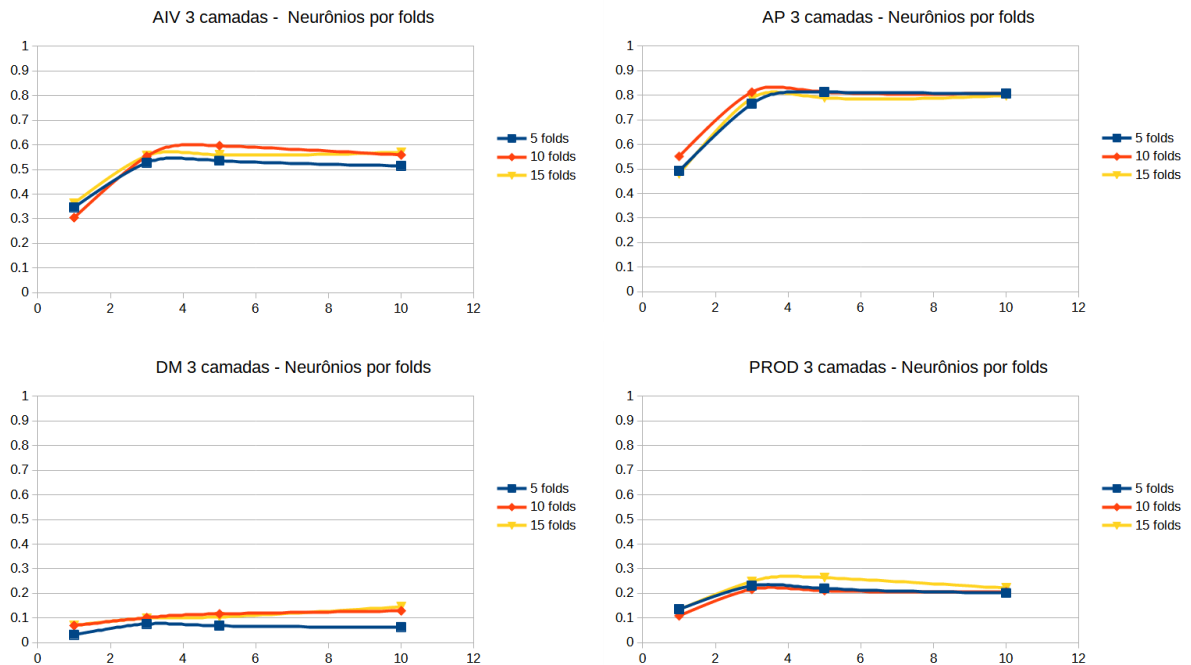


Figura 4: Estimativa do R^2 obtidos pelos modelos nas diferentes configurações para três camadas ocultas, com diferentes números de neurônios e números de folds. AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

A variável altura de inserção de vagem iniciou-se com R^2 entre 30.41% e 36.32%, para 10 e 15 folds respectivamente. Com três neurônios todos apresentaram aumento, para 52.70, 55.35 e 55.51, para cinco, 10 e 10 folds respectivamente. O pico ocorreu com cinco neurônios e 10 folds, apresentando R^2 de 59.67%. Com 10 neurônios, não houve diferença maior que 5% entre os folds estudados para esta variável.

No que se refere à altura de planta, houve baixa variação ao longo de toda a relação. Sendo esta, mais variável para um neurônio, onde a configuração de 15 folds apresentou um R^2 de 47.94%, cinco folds 49.14% e 10 folds 55.09%. O pico foi obtido com cinco neurônios e 10 folds, com coeficiente de determinação de 81.35%, seguido de 81.26% para cinco neurônios e cinco folds e 81.25% para 10 folds e três neurônios. A partir de cinco neurônios, não houve diferença maior que 3% entre os folds.

Dias para maturação teve cinco folds com o pior desempenho para todos os neurônios, com 10 e 15 folds apresentando resultados que não diferiram mais que 1% para cada quantidade de neurônios estudados. O pico ocorreu com 15 folds e 10 neurônios, com um aumento do R^2 apresentado de 0.0699 para um neurônio para 14.52 para 10 neurônios com 15 folds.

Produtividade apresentou baixa variação para cada número de folds estudados para a mesma quantidade de neurônios. Porém, 15 folds liderou os coeficientes de determinação apresentados ao longo de toda a variação do número de neurônios, iniciando com 13.39% para um neurônio, passando pelo pico de 26.39 com cinco neurônios e finalizando com 22.11% com 10 neurônios. Modelos com cinco e 10 folds se alternaram em segundo lugar na medida em que houve variação dos neurônios, mas sem se diferirem mais que 3%.

Relação camadas ocultas x folds x neurônios nas variáveis avaliadas

Altura de inserção de vagem (AIV)

Para altura de inserção de vagem (Tabela 6 e Figura 5), na relação entre o número de camadas ocultas e número de neurônios, o coeficiente de determinação apresentou-se menor para o modelo com três camadas ocultas, enquanto que uma e duas camadas apresentaram resultados semelhantes de maneira geral. Os

coeficientes apresentados pelos modelos com uma e duas camadas alternaram-se quanto ao maior valor, a depender do número de neurônios e número de folds. Dentro todas as configurações, o modelo com duas camadas, cinco neurônios e 15 folds apresentou maior coeficiente, seguido por duas camadas com 10 folds com cinco e 10 neurônios e uma camada com 10 e 15 folds e 10 neurônios, não diferindo mais que 2% dentre todos estes. Para três camadas ocultas, o maior valor encontrado foi de 59.67%.

Tabela 6: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável altura de inserção de vagem.

Camadas	Folds	Neurônios			
		1	3	5	10
1	5	53.38	61.24	61.06	68.57
	10	33.99	60.89	69.53	70.68
	15	44.76	69.94	69.18	70.15
2	5	31.86	56.85	62.38	64.69
	10	33.69	60.77	70.52	70.40
	15	43.44	62.80	71.12	67.95
3	5	34.59	52.70	53.53	51.47
	10	30.41	55.34	59.67	55.88
	15	36.32	55.51	55.96	56.94

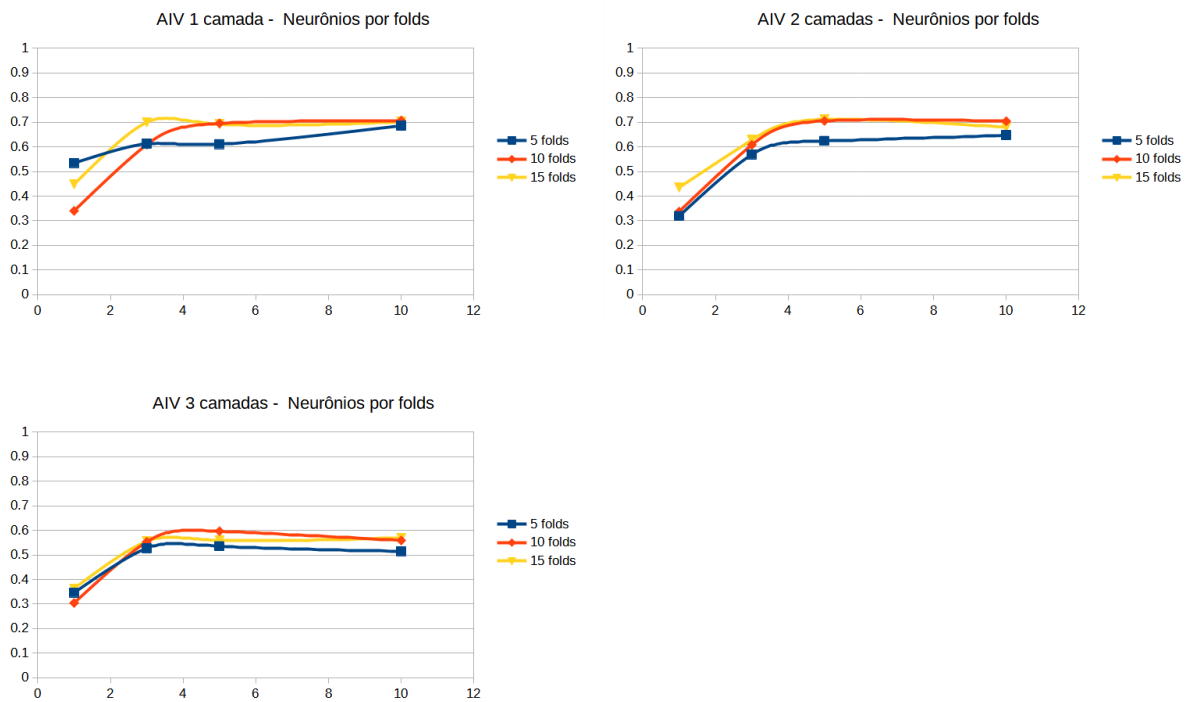


Figura 5: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e números de folds para a variável altura de inserção de vagem.

Altura de Planta (AP)

Já para altura de planta (Tabela 7 e Figura 6), os valores de coeficientes de determinação apresentado para uma, duas ou três camadas ocultas variou de menor forma de maneira geral. Todos apresentaram comportamentos semelhantes na média em que houve acréscimo do número de neurônios por camadas. Com exceção de um neurônio por camada, onde houve maior variação, a partir de três neurônios, os resultados se diferiram pouco, não havendo mais que 5% de diferença entre os R^2 apresentados para todas as configurações para cada quantidade de neurônio. O maior valor apresentado foi, 82.99%, resultado apresentado com duas camadas, cinco neurônios e 10 folds, mas não diferindo mais de 1% dos resultados encontrados com duas camadas e 10 folds a partir de três neurônios.

Tabela 7: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável altura de planta.

Camadas	Folds	Neurônios			
		1	3	5	10
1	5	60.48	80.04	79.82	79.57
	10	66.61	81.71	79.69	80.25
	15	60.20	78.67	80.96	80.89
2	5	66.73	82.40	81.10	81.77
	10	57.28	82.14	82.99	82.31
	15	67.59	81.77	81.52	81.35
3	5	49.14	76.46	81.26	80.65
	10	55.09	81.25	81.35	80.77
	15	47.94	78.75	78.96	79.75

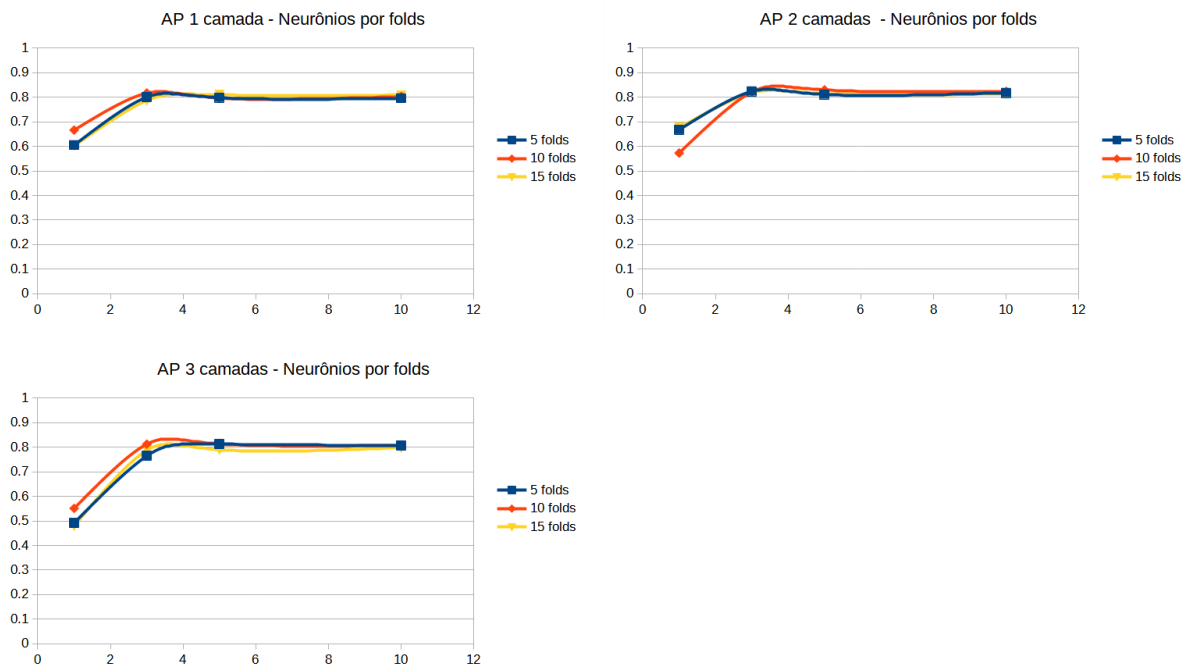


Figura 6: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e números de folds para a variável altura de planta.

Dias para maturação (DM)

Dias para maturação apresentou resultados semelhantes para os modelos com uma, duas e três camadas (Tabela 8 e Figura 7), sendo que para cinco e 15 folds, três camadas apresentaram um resultado ligeiramente menor que uma e duas camadas. O pico do coeficiente de determinação ocorreu com uma camada, 15 folds e cinco

neurônios, com um R^2 de 15.97%, seguido de duas camadas, 15 folds e cinco neurônios, apresentando 14.82%. O menor coeficiente de determinação por quantidade de neurônios ocorreu com apenas um neurônio por camada, estabilizando a partir de três neurônios.

Tabela 8: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável dias para maturação.

Camadas	Folds	Neurônios			
		1	3	5	10
1	5	6.17	9.67	12.80	11.74
	10	8.59	12.48	10.60	14.78
	15	9.60	13.07	15.97	11.56
2	5	5.98	7.73	11.89	9.88
	10	7.09	9.56	12.70	11.74
	15	8.19	12.08	14.82	13.06
3	5	3.19	7.57	6.83	6.32
	10	6.91	10.12	11.64	12.88
	15	6.99	9.89	10.43	14.52

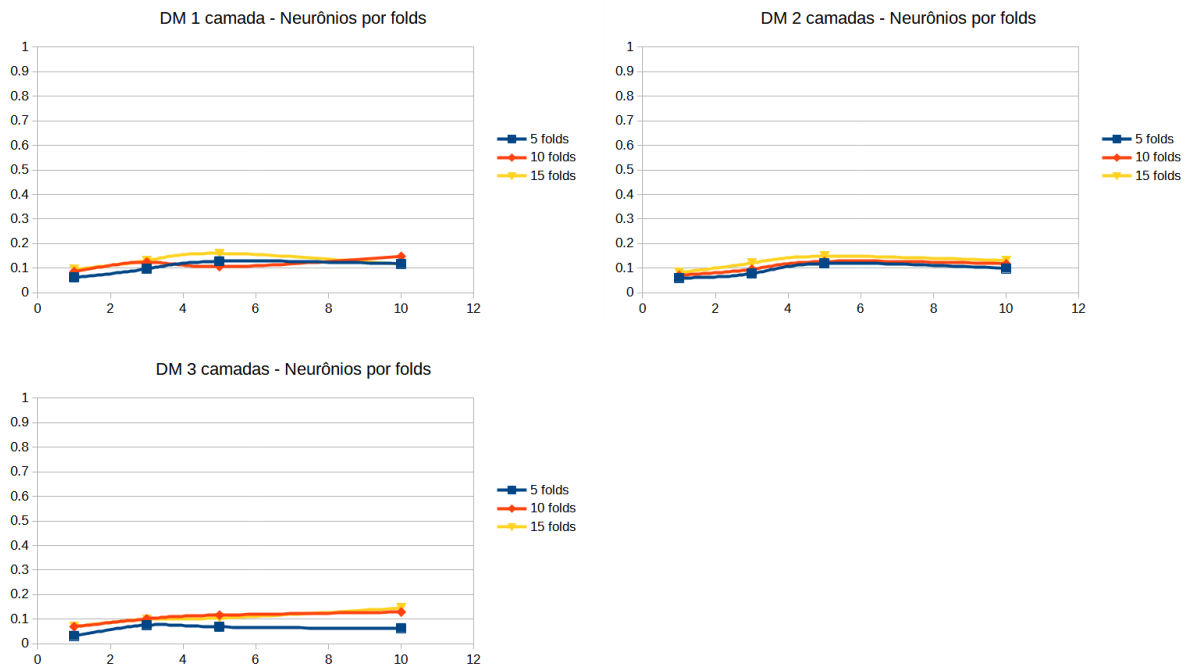


Figura 7: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e números de folds para a variável dias para maturação.

Produtividade (PROD)

À semelhança do ocorrido para as demais variáveis estudadas, o modelo com três camadas apresentou resultados inferiores para a variável produtividade quando comparado com os modelos com uma e duas camadas ocultas (Tabela 9 e Figura 8). O pico ocorreu com uma camada, 15 folds e três neurônios, sendo de 28.65%, mas não diferiu mais de 4% para estes mesmos parâmetros com cinco e 10 neurônios. Houve estabilização a partir de três neurônios, não havendo variações maiores que 4% quando na comparação entre modelos com o mesmo número de camadas ocultas para cada quantidade de neurônio.

Tabela 9: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável produtividade.

Camadas	Folds	Neurônios			
		1	3	5	10
1	5	14.77	24.29	23.17	26.14
	10	13.69	24.12	24.77	24.11
	15	20.96	28.65	24.95	27.42
2	5	14.87	23.74	24.48	24.51
	10	15.27	25.51	24.09	22.84
	15	17.29	24.93	24.87	26.45
3	5	13.46	23.06	21.98	20.12
	10	10.87	21.62	21.14	20.54
	15	13.39	24.82	26.39	22.11

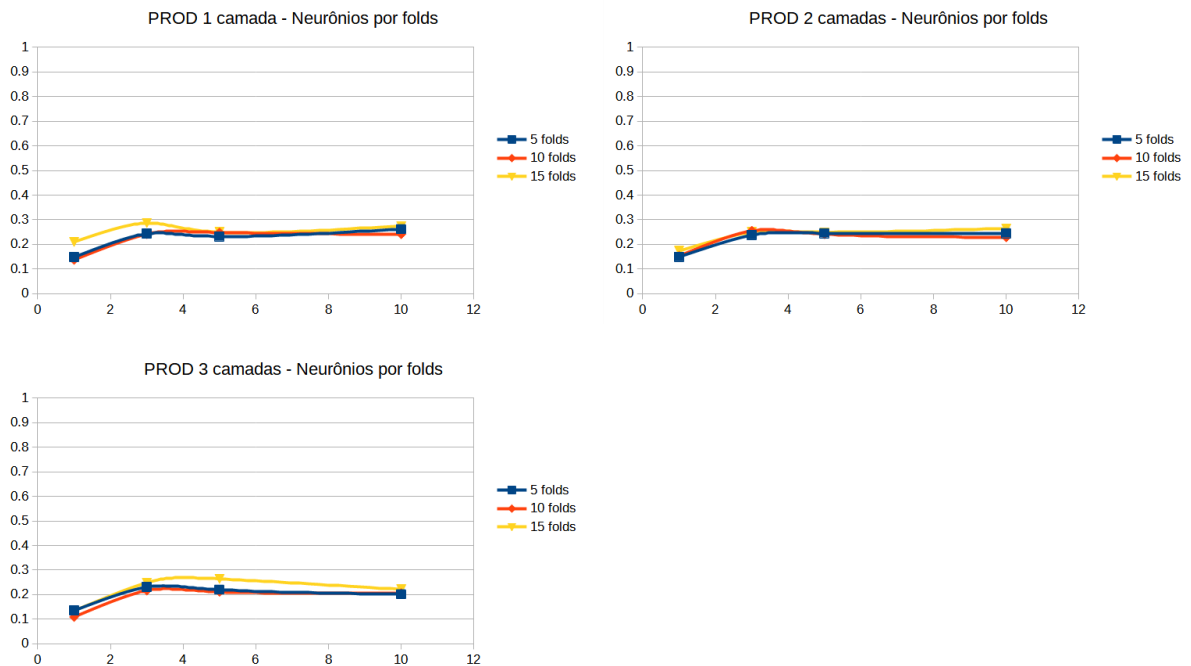


Figura 8: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e números de folds para a variável produtividade.

Efeito do número de épocas em relação ao número de neurônios por camada oculta

Uma camada oculta

Para uma camada oculta (Tabela 10 e Figura 9), a variável altura de inserção de vagem apresentou variações no coeficiente de determinação para as três quantidades de épocas avaliadas. Com um neurônio por camada, 30000 e 50000 épocas apresentaram um R^2 semelhante, de 44.76% e 42.83% respectivamente, com 10000 épocas apresentando R^2 superior de 53.38%. Porém, a partir de três neurônios, 30000 épocas passa a apresentar resultados superiores que as demais épocas, sendo o pico do coeficiente de determinação obtido 70.15%, com 10 neurônios por camada oculta, mas não diferindo mais de 1% em relação à esta mesma quantidade de épocas com três e cinco neurônios. 10000 e 50000 épocas apresentaram R^2 inferiores para três e cinco neurônios, mas com 10 neurônios, 10000 épocas apresentou um R^2 de 68.57%. Já 50000 épocas apresentou um R^2 de 65.49%.

Tabela 10: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos com uma camada oculta.

Variável	Épocas	Neurônios			
		1	3	5	10
AIV	10000	53.38	61.24	61.06	68.57
	30000	44.76	69.94	69.18	70.15
	50000	42.83	62.87	64.18	65.49
AP	10000	60.48	80.04	79.82	79.57
	30000	77.43	79.65	80.07	81.01
	50000	72.31	79.98	80.39	80.71
DM	10000	6.17	9.67	12.80	11.74
	30000	5.76	10.40	10.96	11.30
	50000	5.76	11.09	9.15	11.80
PROD	10000	14.77	24.29	23.17	26.14
	30000	18.42	23.95	24.61	26.54
	50000	13.80	23.47	22.06	24.14

AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

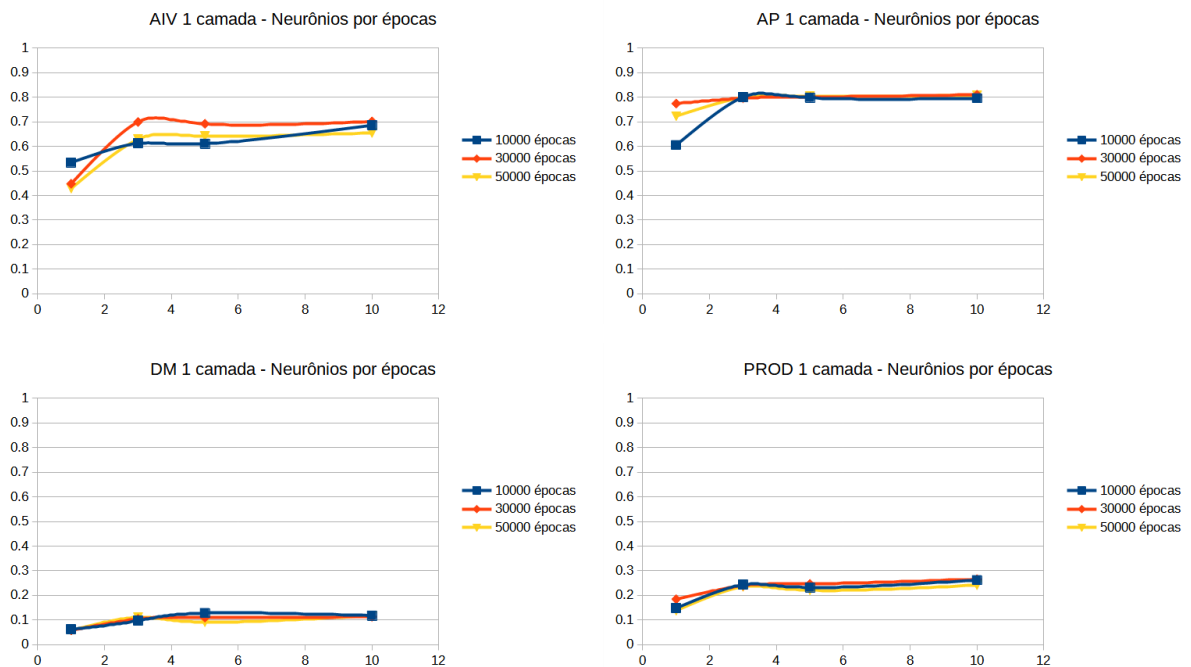


Figura 9: Estimativa do R^2 obtidos pelos modelos com uma camada oculta, com diferentes números de neurônios e número de épocas. AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

A variável altura de planta apresentou maior estabilidade a partir de três neurônios, com todos os modelos com R^2 de aproximadamente 80% não diferindo mais de 1%, tendo seu pico com o modelo de 30000 épocas, com coeficiente de determinação de 81.01%. Para um neurônio por camada, 10000 épocas apresentou resultado no coeficiente de determinação entre 12% a 17% inferior em relação à 50000 e 30000 épocas respectivamente. O modelo com 30000 épocas iniciou com 77.43% de R^2 para um neurônio e chegou a 81.01% com 10 neurônios, apresentando aumento gradual do coeficiente de determinação.

Para dias para maturação, houve baixa diferença entre os R^2 apresentados, sendo a maior diferença com cinco neurônios onde 50000 épocas obteve um coeficiente e determinação de 9.15%, enquanto que 30000 e 10000 épocas apresentaram um R^2 de 10.96% e 12.80% respectivamente, sendo o último o maior R^2 obtido dentre todas as configurações.

Os modelos com 50000 épocas apresentaram menor R^2 para produtividade para toda quantidade de neurônios por camada testados, iniciando com 13.80% para um neurônio e se mantendo entre 22.06% e 24.14% a partir de três neurônios. Já os modelos com 10000 e 30000 épocas se alternaram com maior R^2 a depender da quantidade de neurônios, mas mantendo um padrão semelhante de comportamento na medida em que houve acréscimo da quantidade de neurônios, sendo o pico do coeficiente de determinação com 10 neurônios, 26.54% para 30000 épocas e 26.14% para 10000 épocas.

Duas camadas ocultas

Diferentemente de uma camada oculta, para a variável AIV os modelos não se diferiram muito em relação ao número de épocas testadas com duas camadas ocultas, não havendo diferença maior que 4% quando na mesma quantidade de neurônios por camada (Tabela 11 e Figura 10). O maior R^2 ocorreu no modelo com 30000 épocas e 10 neurônios, sendo este de 65.48%, seguido de 64.69% do modelo com 10000 épocas e 10 neurônios. Houve grande aumento para todos os números de épocas testados na comparação entre os coeficientes de determinação testados com um e três neurônios, havendo um acréscimo de até 25%. E apresentou-se ligeiro aumento do valor de R^2 para todas as épocas a partir de três neurônios.

Tabela 11: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos com duas camadas ocultas.

Variável	Épocas	Neurônios			
		1	3	5	10
AIV	10000	31.86	56.85	62.38	64.69
	30000	35.57	57.11	61.79	65.48
	50000	37.74	58.78	61.32	61.45
AP	10000	66.73	82.40	81.10	81.77
	30000	66.84	82.58	82.43	81.56
	50000	64.01	79.84	81.55	80.47
DM	10000	5.98	7.73	11.89	9.88
	30000	8.24	11.69	12.81	11.87
	50000	7.10	10.37	12.81	11.05
PROD	10000	14.87	23.74	24.48	24.51
	30000	16.00	24.79	26.63	25.46
	50000	15.61	23.78	21.78	23.41

AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

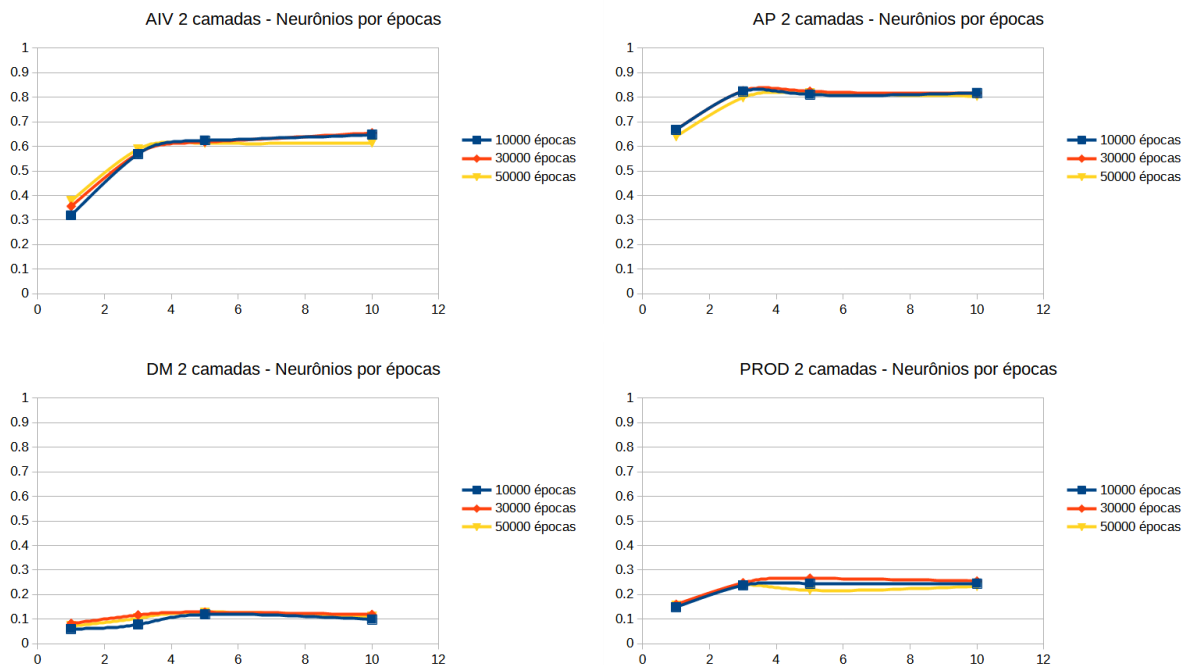


Figura 10: Estimativa do R^2 obtidos pelos modelos com duas camadas ocultas, com diferentes números de neurônios e número de épocas. AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

A variável altura de planta apresentou aumento do valor do R^2 para todas as épocas quando na comparação entre um e três neurônios, com acréscimos de até 25%, à semelhança da variável AIV. Porém, o pico de R^2 ocorreu com três neurônios, tanto para 10000 quanto para 30000 épocas, atingindo valores de 82.40% e 82.58% respectivamente. A partir de três neurônios, houve queda menor que 1% de maneira geral, com os coeficientes de determinação apresentando-se estáveis até 10 neurônios.

Dias para maturação apresentou maior R^2 com cinco neurônios por camada oculta, sendo este de 12.81%, obtido tanto por 30000 épocas, quanto por 50000 épocas. Considerando 10000 épocas, os resultados foram inferiores para todas as quantidades de neurônios avaliadas, quando na comparação com as outras épocas testadas. 30000 e 50000 épocas apresentaram estabilização do valor de R^2 a partir de três neurônios.

Produtividade apresentou estabilidade nos coeficientes de determinação obtidos a partir de três neurônios, com o pico ocorrendo para cinco neurônios e 30000 épocas. Entre um e três neurônios houve variação de até 8% e de três a 10 neurônios não houve variação maior que 3%. As épocas de 10000 e 30000 não diferiram mais que 2% quando na mesma quantidade de neurônios, enquanto que 50000 apresentou R^2 menor em até 5% quando na comparação com as outras épocas avaliadas.

Três camadas

Para a variável AIV, 50000 épocas apresentou menores R^2 em comparação com as outras épocas testadas para todas as quantidades de neurônios avaliadas. 10000 e 30000 apresentaram resultados semelhantes para um e três neurônios (Tabela 12 e Figura 11). Porém a partir de cinco neurônios 30000 épocas se mantém com aumento no valor do coeficiente de determinação, chegando em até 59.33%, ocorrendo em 10 neurônios, valor este o maior dentre todos os R^2 obtidos. Já 10000 épocas decaiu ligeiramente entre três e 10 neurônios.

Tabela 12: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos com três camadas ocultas.

Variável	Épocas	Neurônios			
		1	3	5	10
AIV	10000	34.59	52.70	53.53	51.47
	30000	32.39	53.57	57.59	59.33
	50000	24.53	45.46	51.12	49.56
AP	10000	49.14	76.46	81.26	80.65
	30000	40.67	80.77	82.52	79.09
	50000	61.43	79.98	81.72	80.47
DM	10000	3.19	7.57	6.83	6.32
	30000	3.89	7.90	7.56	8.04
	50000	3.89	9.74	9.47	9.78
PROD	10000	13.46	23.06	21.98	20.12
	30000	7.46	20.18	22.61	21.84
	50000	8.99	20.42	22.53	23.82

AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

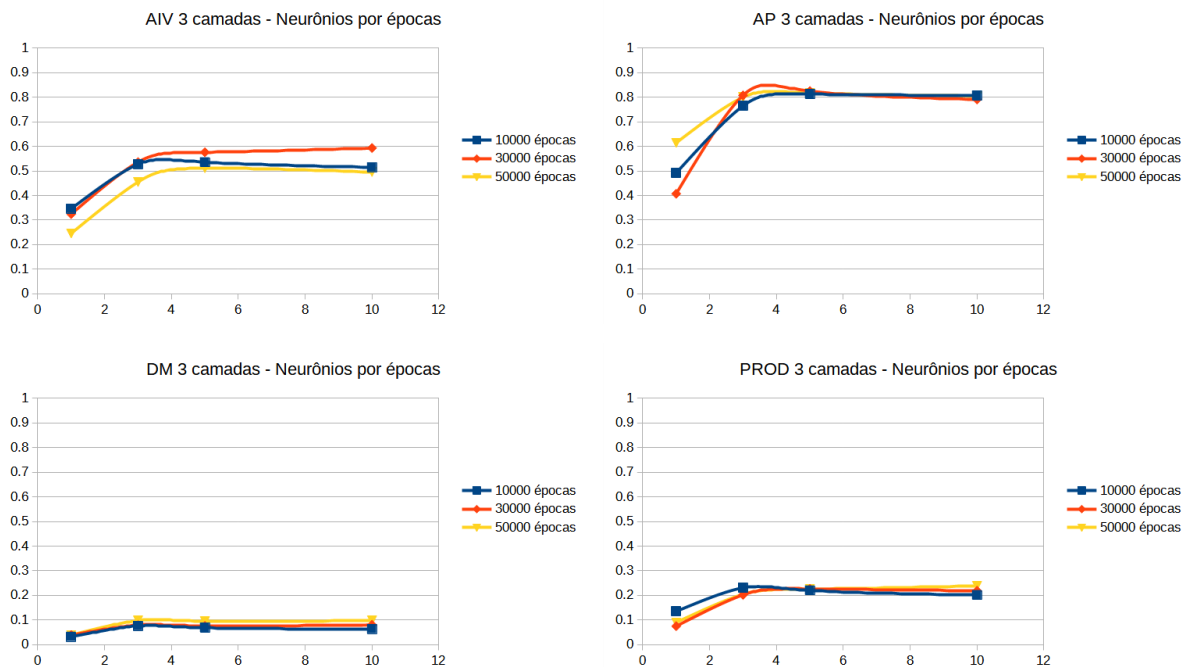


Figura 11: Estimativa do R^2 obtidos pelos modelos com três camadas ocultas, com diferentes números de neurônios e número de épocas. AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade.

Os maiores coeficientes de determinação para altura de planta foram obtidos

com cinco neurônios para todas as épocas, sendo 82.52% o maior e obtido com 30000 épocas, seguido de 81.72% por 50000 épocas e 81.26% com 10000 épocas. Houve diferença de até 40% do valor do R^2 entre um e três neurônios estabilizando com cinco e 10 neurônios, não variando mais que 2% entre as épocas.

Dias para maturação apresentou-se abaixo de 10% para todas as configurações avaliadas, sendo 9.78% o maior R^2 , sendo este obtido com 50000 épocas e 10 neurônios. A diferença entre o menor e o maior valores dos coeficientes de determinação encontrados foi de 5.89%, relativos à 50000 épocas com um e 10 neurônios respectivamente.

Com relação à produtividade, não houve diferença maior que 2% para todas as configurações obtidas a partir de cinco neurônios. 23.82% foi o maior coeficiente de determinação encontrado, obtido a partir de 50000 épocas e 10 neurônios. Os R^2 obtidos para todas as configurações com um neurônio, aumentaram em até 15% quando na comparação com configurações com outros neurônios.

Relação camadas ocultas x épocas x neurônios entre as variáveis avaliadas

Altura da inserção da vagem (AIV)

Para a relação entre o número de camadas ocultas, número de épocas e número de neurônios para a variável AIV, os coeficientes de determinação encontrados utilizando dos modelos com uma camada foram maiores que duas e três camadas ocultas para todas as épocas avaliadas (Tabela 13 e Figura 12). O maior R^2 encontrado ocorreu com uma camada, 30000 épocas e 10 neurônios. Observou-se que com três camadas, o desempenho dos modelos foi inferior em todas as configurações.

Tabela 13: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável altura da inserção da vagem.

Camadas	Épocas	Neurônios			
		1	3	5	10
1	10000	53.38	61.24	61.06	68.57%
	30000	44.76	69.94	69.18	70.15
	50000	42.83	62.87	64.18	65.49
2	10000	31.86	56.85	62.38	64.69
	30000	35.57	57.11	61.79	65.48
	50000	37.74	58.78	61.32	61.45
3	10000	34.59	52.70	53.53	51.47
	30000	32.39	53.57	57.59	59.33
	50000	24.53	45.46	51.12	49.56

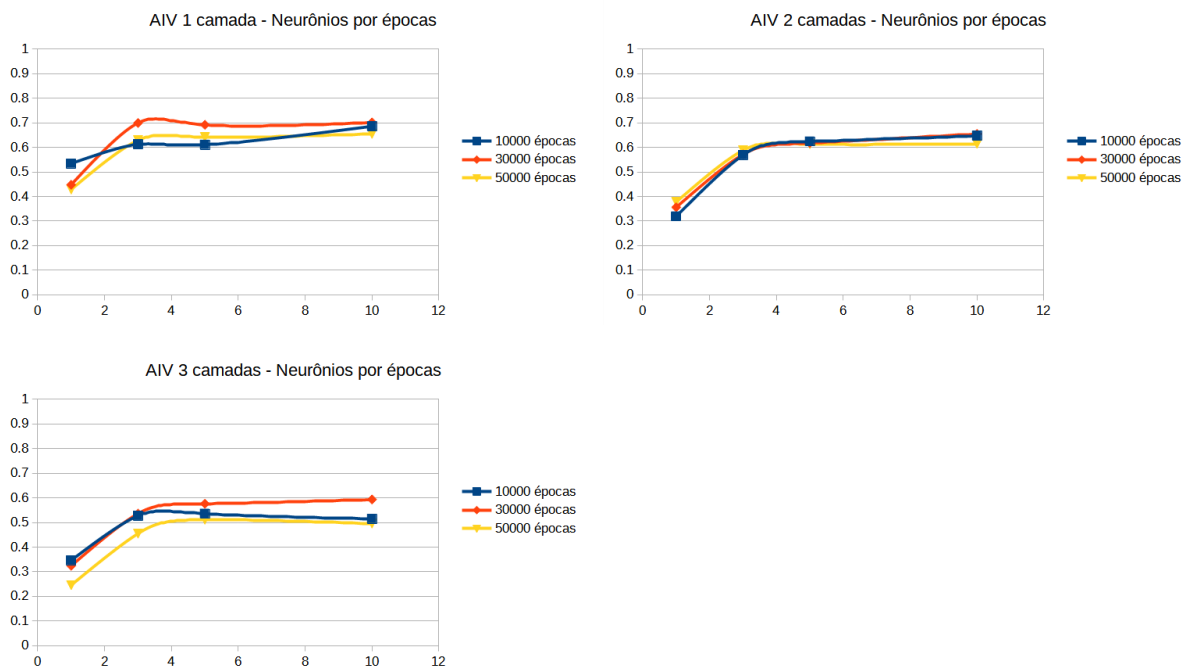


Figura 12: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e número de épocas para a variável altura da inserção da vagem.

Altura de planta (AP)

Contrastando com altura de inserção de vagem, a altura de planta apresentou coeficientes de determinação semelhantes a partir de três neurônios para todas as configurações, variando em maior grandeza apenas com um neurônio. Os modelos

com duas camadas, com 10000 e 30000 épocas e três neurônios se destacaram quanto aos maiores R^2 , com 82.40% e 82.58% respectivamente, mas não divergindo mais que 2% em relação a essas mesmas configurações com cinco e 10 neurônios. Já os modelos com três camadas foram, de maneira geral, inferiores em todas as alternativas (Tabela 14 e Figura 13). Porém, os valores obtidos com três, cinco e 10 neurônios, apesar de inferiores não se diferiram mais que 5% dentro da mesma quantidade de neurônios em relação às outras quantidades de camadas e nas diversas épocas estudadas.

Tabela 14: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável altura de planta.

Camadas	Épocas	Neurônios			
		1	3	5	10
1	10000	60.48	80.04	79.82	79.57
	30000	77.43	79.65	80.07	81.01
	50000	72.31	79.98	80.39	80.71
2	10000	66.73	82.40	81.10	81.77
	30000	66.84	82.58	82.43	81.56
	50000	64.01	79.84	81.55	80.47
3	10000	49.14	76.46	81.26	80.65
	30000	40.67	80.77	82.52	79.09
	50000	61.43	79.98	81.72	80.47

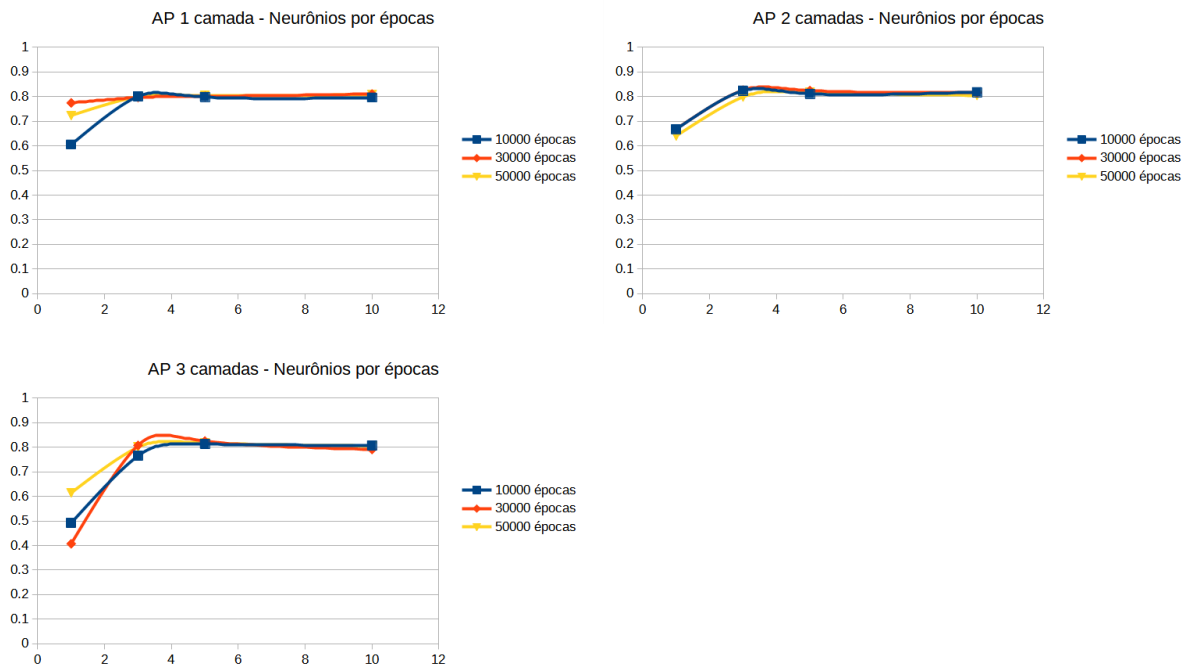


Figura 23: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e número de épocas para a variável altura de planta.

Dias para maturação (DM)

Dias para maturação não apresentou R^2 maior que 12.81% para todas as configurações avaliadas, sendo este obtido com duas camadas, neurônios e tanto 30000 quanto 50000 épocas. Mas este valor máximo não divergiu mais que 3% quando na comparação com o mesmo número de neurônios para todas as épocas estudadas dentro de uma e duas camadas. Os modelos com três camadas apresentaram, de maneira geral, resultados inferiores (Tabela 15 e Figura 14).

Tabela 15: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável dias para maturação.

Camadas	Épocas	Neurônios			
		1	3	5	10
1	10000	6.17	9.67	12.80	11.74
	30000	5.76	10.40	10.96	11.30
	50000	5.76	11.09	9.15	11.80
2	10000	5.98	7.73	11.89	9.88
	30000	8.24	11.69	12.81	11.87
	50000	7.10	10.37	12.81	11.05
3	10000	3.19	7.57	6.83	6.32
	30000	3.89	7.90	7.56	8.04
	50000	3.89	9.74	9.47	9.78

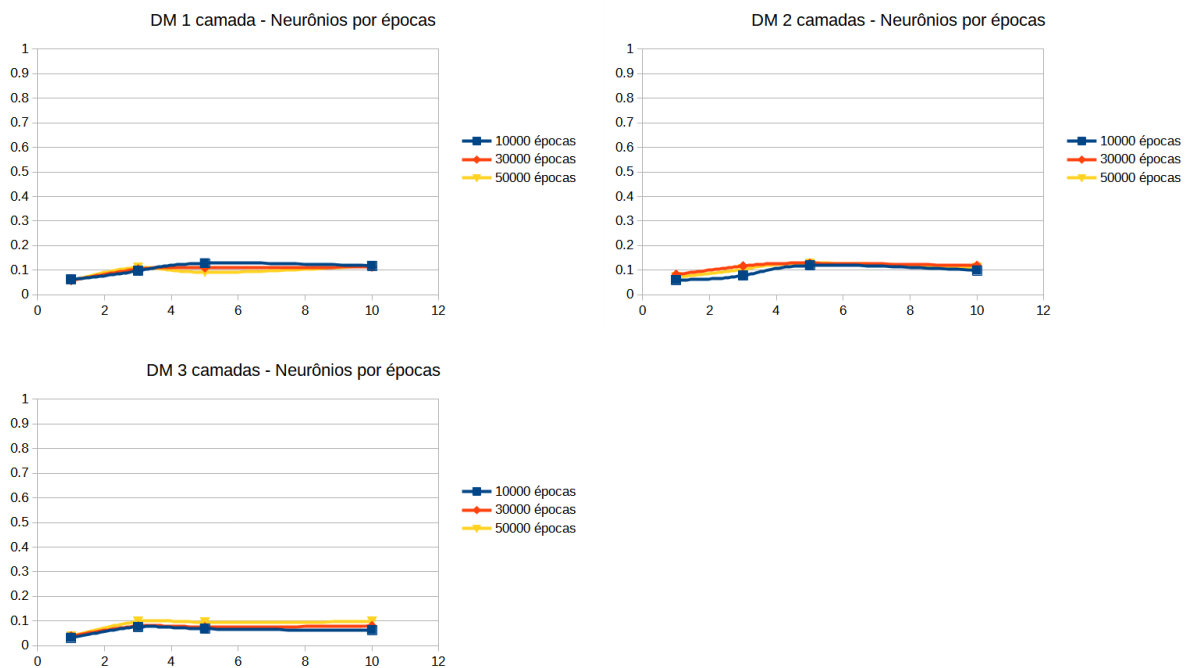


Figura 34: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e número de épocas para a variável dias para maturação.

Produtividade (PROD)

Os modelos com três camadas para a variável produtividade, assim como para todas as outras, apresentaram resultados inferiores em sua maioria. Já os modelos com uma e duas camadas obtiveram resultados semelhantes, com ora uma camada

apresentando maiores R^2 , ora duas camadas apresentando R^2 superiores, mas estes não se diferindo mais que 2% dentro da mesma quantidade de neurônios para todas as épocas estudadas (Tabela 16 e Figura 15). O maior R^2 foi obtido através de modelo com duas camadas, 30000 épocas e cinco neurônios por camada oculta. Seguido de 26.54% e 26.16% obtidos pelos modelos com uma camada, 10 neurônios e 30000 e 10000 épocas respectivamente.

Tabela 16: Porcentagem da estimativa do R^2 (%) obtidos pelos modelos para a variável produtividade.

Camadas	Épocas	Neurônios			
		1	3	5	10
1	10000	14.77	24.29	23.17	26.14
	30000	18.42	23.95	24.61	26.54
	50000	13.80	23.47	22.06	24.14
2	10000	14.87	23.74	24.48	24.51
	30000	16.00	24.79	26.63	25.46
	50000	15.61	23.78	21.78	23.41
3	10000	13.46	23.06	21.98	20.12
	30000	7.46	20.18	22.61	21.84
	50000	8.99	20.42	22.53	23.82

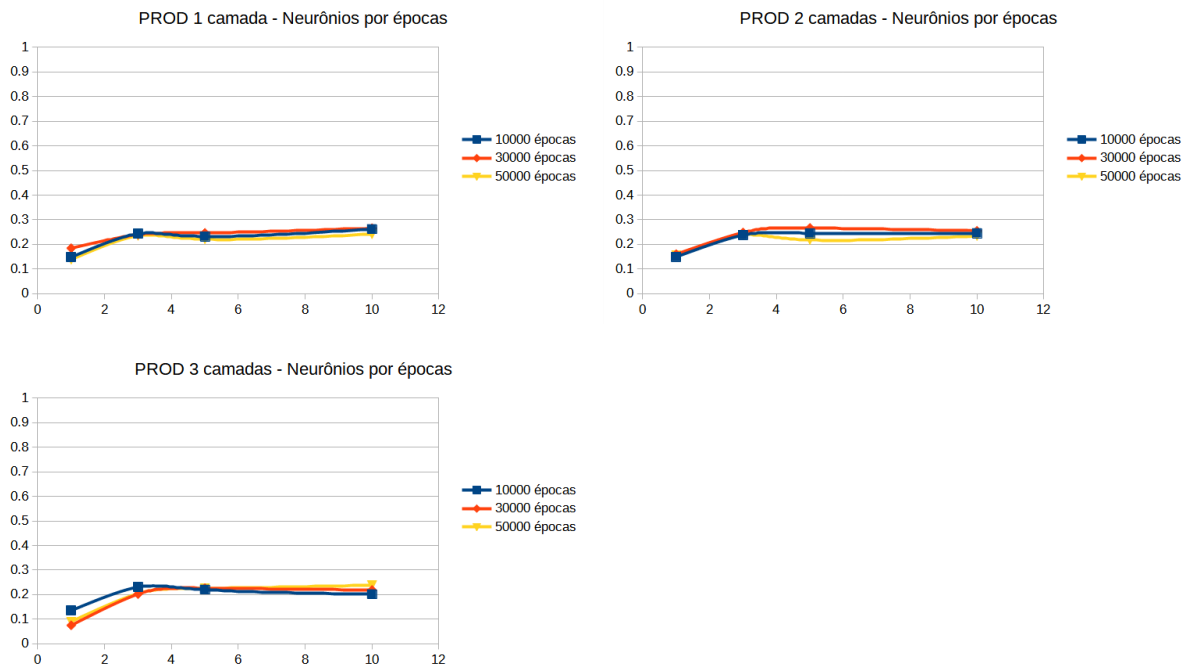


Figura 45: Estimativa do R^2 obtidos pelos modelos com diferentes números de neurônios e número de épocas para a variável produtividade.

Média geral na relação entre o número de camadas e número de neurônios por camada

Quando na média entre todas as variáveis e todas as quantidades de folds avaliados, observa-se que houve aumento de até 15% para as médias entre todas as variáveis na comparação entre um e três neurônios e aumento gradual das médias dos coeficientes de determinação apresentados a partir de três neurônios sendo que, em média, os modelos com 10 neurônios apresentaram maiores R^2 , sendo os modelos com cinco neurônios não apresentando valores menores que 1% quando na comparação com 10 neurônios (Tabela 17). Entre todas as médias avaliadas, os modelos com uma camada oculta e 10 neurônios apresentaram maiores médias, mas não diferindo mais que 2% com relação à 10 neurônios e duas camadas e cinco neurônios com uma ou duas camadas.

Tabela 17: Média entre todas as variáveis, número de folds e épocas por quantidade de neurônios por camada oculta.

Camadas	1 Neurônio	3 Neurônios	5 Neurônios	10 Neurônios
1	32.77	45.40	46.04	47.16
2	30.77	44.19	46.87	46.41
3	25.69	41.42	42.43	41.83

Coefficientes de determinação apresentados em porcentagem (%).

Média geral entre o número de camadas e número de épocas

Quando na relação entre a média do número de épocas com o número de camadas, o número intermediário de épocas, 30000, apresentou maior média de coeficiente de determinação, sendo uma camada apresentando maior média em todos os valores de média avaliados. Considerando duas camadas ocultas, as médias não diferiram mais de 2% na comparação com uma camada para todas as épocas avaliadas (Tabela 18). Para três camadas, apresentando a maior diferença em relação aos outros modelos avaliados, sendo esta diferença de até 5%.

Tabela 18: Relação entre a média geral do número de épocas com o número de camadas.

Camadas	10000 Épocas	30000 Épocas	50000 Épocas
1	44.84	46.48	44.61
2	44.29	45.35	43.88
3	40.16	41.75	40.34

Coefficientes de determinação apresentados em porcentagem (%).

Média geral entre número de camadas e número de folds

Na média geral entre o número de camadas e número de folds, houve acréscimo dos coeficientes de determinação médios para todos os modelos na medida em que se houve o aumento do número de folds, sendo que, portanto, 15 folds apresentou o os melhores R^2 médios para todos os modelos testados, tendo os modelos com uma camada e 15 folds apresentado maior média a respeito do R^2 , seguido dos modelos com duas camadas e por final três camadas (Tabela 19).

Tabela 19: Relação entre a média geral do número de folds com número de camadas.

Camadas	5 Folds	10 Folds	15 Folds
1	44.84	46.13	47.62
2	44.29	46.30	46.89
3	40.16	42.68	42.84

Coeficientes de determinação apresentados em porcentagem (%).

DISCUSSÃO

Em geral, embora haja estabilização em alguns modelos específicos, o incremento no número de neurônios resultou em elevação dos coeficientes de determinação. Uzair (2020), destacou que o aumento no número de neurônios por camada está associado à redução das taxas de erro nas predições. Este estudo evidencia a possibilidade de redução das taxas de erro em até 40 neurônios por camada. Outros estudos também atestam para que haja priorização da quantidade de neurônios em relação à quantidade de camadas no momento da determinação da topologia (Shafi et al., 2007).

Em relação ao número de camadas, os modelos com uma camada apresentaram em média maiores R^2 na comparação com modelos com duas e três camadas com um, três e 10 neurônios por camadas, e os modelos com duas camadas apresentaram maiores R^2 médios com cinco neurônios por camadas, mas não diferindo mais que 1% em relação à média dos modelos com uma camada. Ou seja, modelos com uma camada se sobressaíram nas predições, não diferindo muito dos modelos com duas camadas ocultas, com os modelos com três camadas ocultas apresentando resultados inferiores em relação aos modelos com uma e duas camadas. Ibnu et al. (2019) sugerem que modelos treinados com conjuntos de dados limitados podem ser mais eficazes ao serem configurados com uma ou duas camadas ocultas. O aumento do número de camadas em redes de perceptrons multicamadas (PMC) seria necessário apenas em conjuntos de dados mais complexos e extensos, como também indicado em (Uzair & Jamil, 2020), o que não se aplica ao cenário deste estudo. A diminuição dos valores de R^2 ao adicionar uma terceira camada oculta, conforme observado por Raut & Dani (2020), reflete a ideia de que a precisão do conjunto de teste pode diminuir se o número de camadas ocultas utilizadas na rede

for significativamente maior do que o necessário para o conjunto de dados fornecido. Outros estudos que empregaram tanto dados simulados quanto dados reais também sugerem que, para conjuntos de dados de tamanho reduzido, a utilização de uma ou duas camadas ocultas é frequentemente suficiente para alcançar o máximo desempenho das redes neurais (Vujicic, 2016; Panchal & Panchal, 2014; Huang, 2013; Panchal et al., 2011; Shafi et al., 2007).

Em relação ao número de folds utilizados na validação cruzada, observou-se que o aumento de cinco para 10 folds resultou, em média, em um aumento nos coeficientes de determinação. Da mesma forma, o aumento de 10 para 15 folds também resultou em um aumento nos coeficientes de determinação. Berrar (2018), Witten et al. (2011) e Kohavi (2001) concluíram em seus estudos que 10 é o número de folds ideal, no entanto, este valor foi indicado buscando o equilíbrio entre a precisão e o viés dos modelos. Uma vez que a qualidade dos dados influencia na cross-validação (Kärkkäinen, 2014) e os valores dos coeficientes de determinação deste estudo não atingiram valores altos para acontecer sobreajuste, o número de 15 folds, ou mais se houver ampliação este estudo, pode ser o mais indicado para os dados em questão.

O número de épocas demonstrou ter impacto na precisão dos modelos, corroborando com as descobertas de Khan & Narvekar (2020). Diferentemente do número de neurônios e do número de folds, o aumento no número de épocas resultou em aumento da precisão ao passar de 10000 para 30000 épocas, porém, reduziu a precisão dos modelos ao passar de 30000 para 50000 épocas. Resultados semelhantes foram encontrados por Shabani et al. (2017), onde as taxas de erro dos modelos diminuíram com o aumento do número de épocas até atingir um valor mínimo, após o qual começaram a aumentar com o contínuo aumento do número de épocas.

CONCLUSÃO

A análise dos diferentes fatores revelou um impacto significativo no desempenho dos modelos avaliados. Constatou-se que, conforme a quantidade de neurônios aumentava até 10, e o número de folds na validação cruzada aumentava

até 15, os modelos apresentavam resultados progressivamente melhores. Por outro lado, o aumento do número de épocas resultava em um aumento nas estimativas de R^2 dos modelos, alcançando um limite de 30000, após o qual os resultados começavam a diminuir. Ademais, o acréscimo no número de camadas ocultas ocasionava uma redução nos valores dos coeficientes de determinação, indicando que o melhor número de camadas foi um.

REFERÊNCIAS

ANDERSEN, T.; MARTINEZ, T. **Cross validation and MLP architecture selection**. In: IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339), Washington, DC, USA, 1999, vol. 3, pp. 1614-1619. 1999. doi: 10.1109/IJCNN.1999.832613.

BERRAR, D. (2018). **Cross-validation**. In: Encyclopedia of Bioinformatics and Computational Biology, Volume 1, Elsevier, pp. 542-545. doi: 10.1016/B978-0-12-809633-8.20349-X.

BHERING, L. L.; CRUZ, C. D.; PEIXOTO, L. de A.; ROSADO, A. M.; LAVIOLA, B. G.; NASCIMENTO, M. Application of neural networks to predict volume in eucalyptus. **Crop. Breed. Appl. Biotechnol.**, v. 15, n. 3, p. 125–131, 2015. doi: 10.1590/1984-70332015v15n3a23.

BHERING, L. L. Rbio: A Tool For Biometric And Statistical Analysis Using The R Platform. **Crop. Breed. Appl. Biotechnol.**, v.17, p. 187-190, 2017.

CABELLO-SOLORZANO, K., ORTIGOSA DE ARAUJO, I., PEÑA, M., CORREIA, L., TALLÓN-BALLESTEROS, A. J. (2023). **The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis**. In: GARCÍA BRINGAS, P., et al. 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023). SOCO 2023. Lecture Notes in Networks and Systems, vol. 750. Springer, Cham. doi: 10.1007/978-3-031-42536-3_33.

FODOR, N.; CHALLINOR, A.; DROUTSAS, I.; RAMIREZ-VILLEGAS, J.; ZABEL, F.; KOEHLER, A.-K.; FOYER, C. H. Integrating Plant Science and Crop Modeling: Assessment of the Impact of Climate Change on Soybean and Maize Production. **Plant Cell Physiol.**, v. 58, n. 11, p. 1833–1847, November 2017. doi: 10.1093/pcp/pcx141.

FRITSCH, S.; GUENTHER, F.; WRIGHT, M. N. **Neuralnet: Training of Neural Networks. R package version 1.44.6**, 2019.

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sens. Environ.**, v. 58, n. 3, p. 289-298, 1996. doi: 10.1016/S0034-4257(96)00072-7.

HUANG, G. Learning Capability and Storage Capacity of Two-Hidden-Layer Feed Forward Networks. **IEEE**, v. 14, n. 2, 2013.

IBNU, C. R. M.; SANTOSO, J.; SURENDRO, K. **Determining the Neural Network Topology: A Review**. In: Proceedings of the 2019 8th International Conference on Software and Computer Applications (ICSCA '19). New York, NY, USA: Association for Computing Machinery, 2019. p. 357–362. 2019. doi: 10.1145/3316615.3316697.

KÄRKKÄINEN, T. On Cross-Validation for MLP Model Evaluation. In: FRÄNTI, P.; BROWN, G.; LOOG, M.; ESCOLANO, F.; PELILLO, M. (eds.) **Structural, Syntactic, and Statistical Pattern Recognition**. S+SSPR 2014. Lecture Notes in Computer Science, vol. 8621. Springer, Berlin, Heidelberg, 2014. p. 30. 2014. doi: 10.1007/978-3-662-44415-3_30.

KHAN, S., NARVEKAR, M. (2020). **Disorder Detection in Tomato Plant Using Deep Learning**. In: VASUDEVAN, H., MICHALAS, A., SHEKOKAR, N., NARVEKAR, M. (eds) **Advanced Computing Technologies and Applications. Algorithms for Intelligent Systems**. Springer, Singapore. https://doi.org/10.1007/978-981-15-3242-9_19

KOHAVI, R. (2001). **A Study of Cross-Validation and Bootstrap for Accuracy**

Estimation and Model Selection. p. 14.

KRUSE, R.; BORGELT, C.; BRAUNE, C.; MOSTAGHIM, S.; STEINBRECHER, M. **Comput. intell.** Springer, 2022. doi: 10.1007/978-3-030-42227-1.

LI, F.; MIAO, Y.; FENG, G.; YUAN, F.; YUE, S.; GAO, X.; LIU, Y.; LIU, B.; USTIN, S. L.; CHEN, X. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. **Field Crops Res.**, v. 157, p. 111-123, 2014. ISSN 0378-4290. doi: 10.1016/j.fcr.2013.12.018.

LI, Y.-H.; LI, W.; ZHANG, C.; YANG, L.; CHANG, R.-Z.; GAUT, B.S.; QIU, L.-J. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. **New Phytol.**, v. 188, p. 242-253, 2010. <https://doi.org/10.1111/j.1469-8137.2010.03344.x>

LIU, H. Q.; HUETE, A. A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. **IEEE Trans. Geosci Remote Sens.**, v. 33, n. 2, p. 457-465, 1995. doi: 10.1109/TGRS.1995.8746027.

MODGIL, R.; TANWAR, B.; GOYAL, A.; KUMAR, V. Soybean (*Glycine max*). In: TANWAR, B.; GOYAL, A. (eds). **Oilseeds: Health Attributes and Food Applications.** Springer, Singapore, 2021. doi: 10.1007/978-981-15-4194-0_1.

NIAZIAN, M.; NIEDBAŁA, G. Machine Learning for Plant Breeding and Biotechnology. **Agriculture**, v. 10, p. 436, 2020. doi: 10.3390/agriculture10100436.

OSCO, L. P.; JUNIOR, J. M.; RAMOS, A. P. M.; FURUYA, D. E. G.; SANTANA, D. C.; TEODORO, L. P. R.; GONÇALVES, W. N.; BAIO, F. H. R.; PISTORI, H.; JUNIOR, C. A. d. S. et al. Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. **Remote Sens.**, v. 12, p. 3237, 2020. doi: 10.3390/rs12193237.

PANCHAL, F., PANCHAL, M. Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. **IJCSMC**, v. 3, n. 11, pp. 455-465, 2014.

PANCHAL, G., GANATRA, A., KOSTA, Y., PANCHAL, D. Behavior Analysis of Multilayer Perceptron with Multiple Hidden Neurons and Hidden Layers. **Int. J. Comput. Theor. Eng.** , v. 3, n. 2, 2011.

PRASANNA, B. M.; CAIRNS, J. E.; ZAIDI, P. H. Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. **Theor. Appl. Genet.**, v. 134, p. 1729–1752, 2021. doi: 10.1007/s00122-021-03773-7.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

RAMOS, F. T.; FERREIRA, L. S.; PIVETTA, F.; MAIA, J. C. S. Leaf blade area of different plants estimated by linear and dry matter measures, calibrated with the ImageJ software. **Interciência**, v. 40, n. 8, p. 570-575, 2015.

RAUT, P., DANI, A. (2020). **Correlation Between Number of Hidden Layers and Accuracy of Artificial Neural Network**. In: VASUDEVAN, H., MICHALAS, A., SHEKOKAR, N., NARVEKAR, M. (eds) *Advanced Computing Technologies and Applications. Algorithms for Intelligent Systems*. Springer, Singapore. doi: 10.1007/978-981-15-3242-9_49.

RAWAT, R.; PATEL, J. K.; MANRY, M. T. **Minimizing validation error with respect to network size and number of training epochs**. In: *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, 2013, pp. 1-7. doi: 10.1109/IJCNN.2013.6706919.

SEDIVY, E.J.; WU, F.; HANZAWA, Y. Soybean domestication: the origin, genetic architecture and molecular bases. **New Phytol**, v. 214, p. 539-553, 2017. <https://doi.org/10.1111/nph.14418>

SHABANI, A., GHAFARY, K. A., SEPASKHAH, A. R., KAMGAR-HAGHIGHI, A. A. Using the Artificial Neural Network to Estimate Leaf Area. **Sci. Hortic.**, v. 216, pp. 103-110, 2017. ISSN 0304-4238. doi: 10.1016/j.scienta.2016.12.032.

SHAKOOR, N., LEE, S., MOCKLER, T. C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. **Curr. Opin. Plant Biol.**, v. 38, 2017, P. 184-192. ISSN: 1369-5266. doi: 10.1016/j.pbi.2017.05.006.

SHAFI, I. SHAH, A. JAMEEL, F. KSHIF. Impact of Varying Neurons and Hidden Layers in Neural Network Architecture for a Time Frequency Application. **IEEE**, 2007.

SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **Int. J. Eng. Appl. Sci. Technol.**, v. 4, n. 12, p. 310-316, 2020. doi: 10.33564/IJEAST.2020.v04i12.054.

SUMNER, Z.; VARCO, J. J; DHILLON, J. S; FOX, A. A. A; CZARNECKI, J; HENRY, W. B. Ground versus aerial canopy reflectance of corn: red-edge and non-red edge vegetation indices. **Agron. J.**, v. 113, p. 2773-2788, 2021. doi: 10.1002/agj2.20657.

TAYADE, R.; YOON, J.; LAY, L.; KHAN, A.L.; YOON, Y.; KIM, Y. Utilization of Spectral Indices for High-Throughput Phenotyping. **Plants**, v. 11, p. 1712, 2022. <https://doi.org/10.3390/plants11131712>.

UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. **World Agricultural Production – Circular Series**, dez 2023. USDA. 2023.

UZAIR, M.; JAMIL, N. **Effects of Hidden Layers on the Efficiency of Neural networks**. In: 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-6. p. 1-6. 2020. doi: 10.1109/INMIC50486.2020.9318195.

VAN KLOMPENBURG, T.; KASSAHUN, A.; CATAL, C. Crop yield prediction using machine learning: A systematic literature review. **Comput. Electron. Agric.**, v. 177, 2020, p. 105709. ISSN 0168-1699. doi: 10.1016/j.compag.2020.105709.

VUJICIC, T.; MATIJEVIC, T.; LJUCOVIC, J.; BALOTA, A.; SEVARAC, Z. **Comparative Analysis of Methods for Determining Number of Hidden Neurons in Artificial Neural Network**. In: Central European Conference on Information and Intelligent Systems. Varazdin: Faculty of Organization and Informatics Varazdin, 2016. p. 219-

223.

WITTEN, I.H.; FRANK, E.; HALL, M.A. **Data Mining: Practical Machine Learning Tools and Techniques**. 3. ed. Burlington: Morgan Kaufmann, 2011.

WORDEN, K.; TSIALIAMANIS, G.; CROSS, E. J.; ROGERS, T. J. Artificial Neural Networks. In: RABCZUK, T.; BATHE, KJ. **Machine Learning in Modeling and Simulation**. Computational Methods in Engineering & the Sciences. Springer, Cham, 2023. doi: 10.1007/978-3-031-36644-4_2.

XIE, C.; YANG, C. A review on plant high-throughput phenotyping traits using UAV-based sensors. **Comput. Electron. Agric.**, v. 178, p. 105731, 2020. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2020.105731>

XUE, J.; SU, B. Significant Remote Sensing Vegetation Indices: A review of Developments and Applications. **J. Sens.**, v. 2017, p.17, 2017. <https://doi.org/10.1155/2017/1353691>

YANG, W., FENG, H., ZHANG, X., ZHANG, J., DOONAN, J. H., BATCHELOR, W. D., XIONG, L. e YAN, J. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. **Mol. Plant**, v. 13, p. 187–214, 2020.

ZHAO, H., LI, Y., CHEN, X. et al. Monitoring monthly soil moisture conditions in China with temperature vegetation dryness indexes based on an enhanced vegetation index and normalized difference vegetation index. **Theor. Appl. Climatol.**, v. 143, p. 159–176, 2021. doi: 10.1007/s00704-020-03422-x.

CAPÍTULO 2 – ALGORITMOS DE INTELIGÊNCIA COMPUTACIONAL NA FENOTIPAGEM DE SOJA ATRAVÉS DE IMAGENS MULTIESPECTRAIS

RESUMO

A soja (*Glycine max*) é uma importante fonte de alimento para humanos e animais, além de ser utilizada como matéria-prima na indústria. Para atender às demandas crescentes, a soja enfrenta uma série de desafios. Esses desafios estão intrinsecamente ligados à busca por variedades altamente produtivas, resistentes a doenças e adaptadas a condições ambientais variáveis. A fenotipagem de alto rendimento emerge como uma ferramenta crucial, acelerando o desenvolvimento eficiente de novas variedades. Esta oferece vantagens quando na comparação com a fenotipagem tradicional, uma vez que a partir do uso de imagens e sensores, auxiliados com softwares e algoritmos de inteligência computacional, otimizam o processo de medição, possibilita maior fenotipagem em escala e reduz a variabilidade da mensuração humana. A utilização de índices de vegetação está entre os principais meios utilizados na fenotipagem de alto rendimento. Com a utilização destes, é possível realizar uma variedade de estudos, incluindo a avaliação do teor de nitrogênio nas folhas, a determinação de características físicas como biomassa, altura da planta e área foliar, a análise da heterogeneidade das plantas no campo, a estimativa do teor de clorofila, a avaliação do teor de água nas plantas, a quantificação do teor de lignina, e a detecção de danos causados por pragas e doenças. Com o uso de técnicas de fenotipagem em larga escala, torna-se crescente volume e a complexidade dos dados obtidos, aumentando a exigência para abordagens inovadoras para análise dos dados com foco em seleção eficiente. Nesse contexto, a aplicação de inteligência computacional tem emergido como uma ferramenta essencial para transformar a forma como se aborda a fenotipagem e o melhoramento de plantas. Assim, este trabalho tem como objetivo avaliar o desempenho de algoritmos de Perceptron Multicamadas (PMC) e Random Forest (RF) na predição de características da soja como Altura de Inserção de Vagem (AIV), Altura de Planta (AP), Dias para Maturação (DM) e Produtividade (PROD), a partir dos seguintes índices de vegetação: Índice de vegetação de Diferença Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), Índice de vegetação de Diferença Normalizada Verde (GNVDI), Índice de Vegetação

Aprimorado (EVI), Índice de Diferença Normalizada de Borda Vermelha (NDRE) e Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI). Houve destaque na predição das variáveis AIV e AP, especialmente em relação aos modelos de Random Forest avaliados. No entanto, nenhum dos modelos foi capaz de prever com êxito as variáveis DM e PROD.

Palavras-chave: Predição, Aprendizado de Máquina, Índices de Vegetação, *Glycine max* (L).

CHAPTER 2 – COMPUTATIONAL INTELLIGENCE ALGORITHMS IN PLANT PHENOTYPING USING MULTISPECTRAL IMAGES

ABSTRACT

Soybean (*Glycine max*) is a valuable source of food for humans and animals, as well as a key raw material in industry. To meet increasing demands, soybean faces a series of complex challenges. These challenges are intrinsically linked to the pursuit of more productive varieties, resistant to diseases, and adapted to variable environmental conditions. High-throughput phenotyping emerges as a crucial tool in this process, accelerating the efficient development of new, more productive varieties capable of addressing evolving environmental and social challenges. It offers advantages compared to traditional phenotyping, as it optimizes the measurement process through the use of images and sensors, aided by software and computational intelligence algorithms, enabling greater scaling of phenotyping and reducing human variability. The use of vegetation indices is among the main methods used in high-throughput phenotyping. With their use, a variety of studies can be conducted, including the evaluation of nitrogen content in leaves, the determination of physical characteristics such as biomass, plant height, and leaf area, the analysis of plant heterogeneity in the field, the estimation of chlorophyll content, the assessment of water content in plants, lignin content quantification, and the detection of damage caused by pests and diseases. With the use of large-scale phenotyping techniques, the volume and complexity of the data obtained are increasing, increasing the demand for innovative approaches to data analysis focused on efficient selection. In this context, the application of computational intelligence has emerged as an essential tool to transform the way phenotyping and plant breeding are approached. Thus, this work aims to evaluate the performance of Multilayer Perceptron (MLP) and Random Forest (RF) algorithms in predicting soybean characteristics such as Pod Insertion Height, Plant Height, Days to Maturity, and Yield, using the following vegetation indices: Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Modified Soil-Adjusted Vegetation Index (MSAVI), Green Normalized Difference Vegetation Index (GNVDI), Enhanced Vegetation Index (EVI), Normalized Difference Red Edge Index (NDRE), and Simplified Canopy Chlorophyll Content Index (SCCCI). There was emphasis on the prediction of the Pod Insertion Height and Plant Height variables,

especially concerning the evaluated Random Forest models. However, none of the models were able to successfully predict the Days to Maturity and Yield variables.

Keywords: Prediction, Machine Learning, Vegetation Indices, *Glycine max (L)*.

INTRODUÇÃO

A soja foi domesticada entre seis e nove mil anos atrás no leste asiático, originando-se da soja selvagem, com a China sendo o centro da maior diversidade genética (USDA, 2023; Sedivy et al., 2017; Modgil et al., 2021; Li et al., 2010). Hoje, Brasil, EUA e Argentina são os principais produtores mundiais desse cultivo, que se tornou essencial para atender às crescentes demandas por alimentos e recursos sustentáveis. No entanto, os desafios no processo de melhoramento da soja são complexos, tais como (Fodor et al., 2017; Battisti et al., 2017), impulsionando a busca por soluções inovadoras, como técnicas avançadas de fenotipagem de alto rendimento, para aprimorar os programas de melhoramento (Gill et al., 2022; Prasanna et al., 2021).

Os Índices de Vegetação (IVs) são amplamente empregados na fenotipagem de alto rendimento, uma técnica que utiliza plataformas terrestres e/ou aéreas com múltiplos sensores para analisar plantas e ambientes (Tayade et al., 2022). Essa abordagem, otimizada pelo uso de imagens, sensores e algoritmos de inteligência computacional, resulta em uma análise eficiente e menos influenciada pela subjetividade humana (Yang et al., 2020; Xie & Yang, 2020). Os principais índices de vegetação, amplamente utilizados em estudos e práticas relacionadas, são o Índice de Vegetação de Diferença Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), Índice de Vegetação de Diferença Normalizada Verde (GNVDI), Índice de Vegetação Aprimorado (EVI), Índice de Diferença Normalizada de Borda Vermelha (NDRE) e Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI), conforme estudado por Sumner et al. (2021), Zhao et al. (2021), Xue & Su (2017), Li et al. (2014), Gitelson et al. (1996) e Liu & Huete (1995). Esses índices desempenham papéis distintos na avaliação de características vegetais, com cada um oferecendo vantagens específicas em diferentes contextos e condições ambientais.

O aumento do volume e complexidade dos dados na fenotipagem em larga escala requer abordagens inovadoras para análise eficiente (Gill et al., 2022; Shakoore et al., 2017). A utilização de métodos de aprendizado de máquina surge como uma alternativa promissora e demonstram bom desempenho em diversas aplicações na agricultura como estimativa de biomassa, classificação de tipos de culturas e estimativa de produtividade (Eugenio et al., 2020; Maimaitijiang et al., 2020; Wu et al.,

2013).

As redes neurais artificiais (RNAs) se destacam nesse contexto, sendo uma das principais vantagens de seu uso a capacidade de resolver problemas não lineares sendo amplamente utilizadas para prever características como produção, conteúdo bioquímico de sementes e tolerância à salinidade, entre outros (Van Klompenburg et al., 2020; Niazian & Niedbała, 2020; Osco et al., 2020; Bhering et al., 2015). Os Perceptrons Multicamadas (PMC) representam uma evolução das RNAs simples, permitindo a captura de relações complexas nos dados e realização de tarefas sofisticadas (Kruse et al., 2022). Outro algoritmo eficaz é o Random Forest, utilizado para classificação e regressão, que combina múltiplas árvores de decisão para melhorar a robustez e a capacidade de predição dos modelos (Costa & Pedreira, 2022; Biau & Scornet, 2016).

Portanto, dada a importância da soja no contexto econômico atual e a constante pressão por produção de novas variedades adaptadas aos atuais e futuros cenários da agricultura, os objetivos deste trabalho são avaliar dois modelos de Perceptron Multicamadas e dois modelos de Random Forest quanto à sua capacidade de predição de quatro características fenotípicas da soja a partir de índices de vegetação, de modo a contribuir no contexto atual da fenotipagem de alto rendimento.

MATERIAIS E MÉTODOS

Todas as análises foram realizadas no software Rbio® (Bhering, 2017).

Descrição do experimento

Os experimentos com a cultura da soja foram conduzidos na área experimental da Universidade Federal de Mato Grosso do Sul, Campus de Chapadão do Sul (localizado em 18°46'26"S, 52°37'28"W, com uma altitude média de 810 m), o método de preparo do solo a ser adotado foi o convencional.

Foram analisados 32 genótipos, em quatro blocos, submetidos a avaliação por meio de um delineamento de blocos casualizados nas safras de 19/20, 20/21 e 21/22.

As características a serem avaliadas incluem: dias para maturação (DM), altura da inserção da primeira vagem (AIV, cm), altura de plantas (AP, cm) e produção de

grãos (PROD, kg ha⁻¹). O DM foi determinado a partir da contagem dos dias entre a emergência e a maturidade de mais da metade das plantas em cada unidade experimental. As AIV e AP foram medidos em cinco plantas por parcela, utilizando uma fita métrica. A produção de grãos foi determinada pela colheita dos dois metros centrais de cada parcela e ajuste para umidade em 13%.

Os índices de vegetação (Tabela 1) utilizados foram obtidos no florescimento pleno da cultura (Ramos et al., 2020; Osco et al., 2020).

Tabela 1: Índices de vegetação utilizados e seus métodos de obtenção

Índice de vegetação	Método de obtenção
NDVI	Xue & Su, 2017
SAVI	Liu & Huete, 1995
MSAVI	Xue & Su, 2017
GNVDI	Gitelson et al., 1996
EVI	Zhao et al., 2021
NDRE	Li et al., 2014
SCCCI	Sumner et al., 2021

Índice de vegetação de Diferença Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Vegetação Ajustado ao Solo Modificado (MSAVI), Índice de vegetação de Diferença Normalizada Verde (GNVDI), Índice de Vegetação Aprimorado (EVI), Índice de Diferença Normalizada de Borda Vermelha (NDRE), Índice Simplificado de Conteúdo de Clorofila do Dossel (SCCCI).

Foi utilizada uma aeronave remotamente pilotada (RPA) de asa fixa, a Sensefly eBee RTK, com capacidade de decolagem, voos e aterrissagens autônomas. O eBee foi equipado com o sensor multiespectral Sensefly Sequoia, e realizou a captura da reflectância dos comprimentos de onda previamente mencionados. Além disso, esse sensor possui uma tecnologia de medição de luminosidade para a calibração dos dados adquiridos. O levantamento aéreo foi feito com a tecnologia RTK, garantindo precisão de posicionamento da câmera no momento da coleta da imagem em até 2,5 cm. As imagens foram processadas com o uso do software Pix4Dmapper, criando mosaicos e correções ortográficas. A precisão posicional foi avaliada através de pontos de controle em campo, medidos com o RTK.

Inteligência Computacional

As predições das variáveis fenotípicas foram realizadas pelos seguintes algoritmos: Perceptron Multicamadas 1 (PMC1), Perceptron Multicamadas 2 (PMC2),

Random Forest 1 (RF1) e Random Forest 2 (RF2), conforme Tabela 2 e Tabela 3:

Tabela 2: Características gerais dos modelos Perceptron Multicamadas 1 e 2.

Parâmetros	PMC1	PMC2
Número de camadas	1	2
Número de Neurônios por camada	10	10
Função de ativação	Logística	Logística
Função de treinamento	<i>Backpropagation</i>	<i>Backpropagation</i>
Método de Avaliação	R ²	R ²
Número de épocas de treinamento	30000	30000
Número de dados	384 linhas	384 linhas
Tipo de treinamento	Validação Cruzada	Validação Cruzada
Número de folds/repetições	15 / 2 Repetições	15 / 2 Repetições

Tabela 3: Características Gerais dos modelos de Random Forest 1 e 2.

Parâmetros	RF1	RF2
Número de árvores	500	1000
Número de variáveis na RF	7	7
Número de dados	384 linhas	384 linhas
Tipo de treinamento	Validação Cruzada	Validação Cruzada
Número de folds/repetições	15 / 2 Repetições	15 / 2 Repetições

As análises foram realizadas através pacote neuralnet (Fritsch et al., 2019), da linguagem de programação R (R core Team, 2024), implementado no software Rbio[®] (Bhering, 2017)

Foi realizada uma regressão múltipla com finalidade de comparação entre uma abordagem de predição tradicional e as abordagens de predição via inteligência computacional.

As variáveis de entrada (Variáveis X) foram as leituras de comprimento de ondas e/ou os índices de vegetação, enquanto que a variável de saída (Variável Y) foi a variável fenotípica predita.

O sucesso da rede neural foi testado comparando-se o coeficiente de determinação (R²) de cada modelo avaliado e com a regressão múltipla.

Uma vez captados, os dados relativos às variáveis Y foram normalizados pela técnica de Min-Max normalization (Cabello-Solorzano et al., 2023).

RESULTADOS

Altura de inserção da vagem (AIV)

Para a variável AIV, não houve diferença maior que 3% do valor de R^2 dentre os algoritmos de inteligência computacional abordados neste estudo, havendo maior diferença, de até 33%, quando na comparação entre a regressão múltipla e o modelo RF 2, e tendo este apresentado até 1.08% a mais no valor de R^2 em relação à RF1. PMC1 e PMC2 não diferiram quanto ao valor do coeficiente de determinação para esta variável (Tabela 4).

Tabela 4: Estimativa do R^2 (%) obtido pelos modelos avaliados na validação.

Variável	Modelos				
	PMC 1	PMC 2	RF 1	RF 2	Reg
AIV	74.80	74.73	77.15	78.07	54.86
AP	80.03	81.57	82.96	82.31	72.70
DM	14.20	13.40	9.74	9.22	13.12
PROD	27.52	25.48	25.89	27.09	22.14
MÉDIA	49.14	48.80	48.94	49.17	40.71

Coefficientes de determinação apresentados em porcentagem (%). AIV: Altura de Inserção de Vagem; AP: Altura de Planta; DM: Dias para maturação; PROD: Produtividade. Perceptron Multicamadas 1 (PMC1), Perceptron Multicamadas 2 (PMC2), Random Forest 1 (RF1) e Random Forest 2 (RF2), Regressão Múltipla (Reg).

Altura de planta (AP)

RF1 foi o modelo que apresentou maior R^2 , mas não diferindo mais que 1% de RF2 e PMC2. Regressão múltipla apresentou um R^2 de 72.70%, 7.30% inferior ao modelo de IC que obteve pior desempenho, PMC1, com 80.03% (Tabela 4).

Dias para maturação (DM)

Ao contrário das variáveis até aqui apresentadas, os modelos de RF apresentaram resultados inferiores quanto ao valor do coeficiente de determinação dos modelos estudados, com valores inferiores até mesmo que a regressão múltipla. Os melhores modelos para esta variável foram PC1 e PMC2, com valores de R^2 de

14.20% e 13.40% respectivamente. A regressão linear apresentou um coeficiente de determinação de 13.12%. Os modelos RF1 e RF2 apresentaram, respectivamente, R^2 de 9.74% e 9.22% (Tabela 4).

Produtividade (PROD)

O modelo de regressão múltipla apresentou o menor R^2 dentre os modelos avaliados, sendo este de 22.14%. Os modelos de IC apresentaram resultados semelhantes, com PMC2 obtendo coeficiente de determinação de 25.48%, RF1 com R^2 de 25.89%, RF2, 27.09% e PMC1 28.52% (Tabela 4).

DISCUSSÃO

De maneira geral, os modelos de regressão múltipla apresentaram resultados inferiores em relação aos gerados via inteligência computacional, sejam eles modelos de redes neurais ou modelos de random forest. Em média, a regressão linear obteve R^2 de 40.71%, enquanto que a menor média obtida pelos algoritmos de IC foi de 48.80%, apresentada por PMC2, e a maior média foi de RF2, cujo valor foi de 49.17%, valores estes muito semelhantes. O desempenho inferior da regressão múltipla se dá uma vez que este tipo de modelo estatístico assume uma relação linear entre as variáveis, enquanto que, se maneira geral, este relacionamento é não-linear (Niedbala et al., 2019; Khairunniza-Bejo et al., 2014; Drummond et al., 2003).

O menor coeficiente de determinação obtido por todos os modelos na predição da produtividade, pode ser explicado, uma vez que a predição da produtividade é uma tarefa complexa que envolve diferentes fatores inerentes às plantas e ao ambiente. Como este estudo buscou prever a produtividade de soja a partir de dados de três safras, a grande variabilidade climática ocorrida nestas três safras pode ter interferido no sucesso dos modelos (Hara et al., 2021; Kross et al., 2020). Trabalhos que obtiveram sucesso na predição de produtividade aumentaram a complexidade de seus modelos de ML na tentativa de contornar a multiplicidade de fatores envolvidos neste processo, utilizando-se de análises multitemporais, modelos de deep learning, métodos de configuração de inputs, transformação de dados e outras séries de

melhorias que buscam extrair dos dados elementos essenciais para a melhor predição (Adak et al., 2021; Teodoro et al., 2021; Panda et al., 2010).

A altura de planta, apesar de também sofrer de interferências ambientais, apresentou resultados relativamente fatoráveis para todos os modelos testados, mas especialmente para os modelos de inteligência computacional. Diversos estudos relataram sucesso na predição de altura de plantas utilizando-se de imagens espectrais, recorrendo a modelos mais simples, até mesmo modelos de inteligência computacional mais complexos como deep learning (Kaur et al., 2021; Sandhu et al., 2021; Osco et al., 2020; Yue et al., 2017).

Altura de inserção de vagem foi a segunda variável que apresentou maior predictibilidade, sendo 78.07% o maior valor de R^2 , obtido pelo modelo RF2. Não foram encontrados trabalhos que utilizaram de algoritmos de inteligência computacional para prever esta característica a partir de índices de vegetação. Porém, AIV pode ser obtida de forma aproximada a partir dessa metodologia, uma vez que esta característica pode ser correlacionada com AP para certas populações de soja (Kuzbakova et al., 2022). Esta é uma variável importante para o melhoramento da soja, uma vez que a perda de grãos na colheita pode ser minimizada se a altura de inserção da primeira vagem for maior (Kang et al., 2017). Portanto esta tem sido uma variável a ser selecionada nos programas de melhoramento.

O melhor modelo para a predição de dias para maturidade foi PMC1, apresentando um coeficiente de determinação de 14.20%. Portanto os modelos avaliados não apresentaram desempenho satisfatório na predição desta característica. Modelos que obtiveram sucesso na predição, utilizaram de deep learning, métodos de inserção de inputs, outras variáveis ambientais e modelo aditivo generalizado para conseguir prever esta característica (Marcillo et al., 2021; Teodoro et al., 2021; Elizondo et al., 1994). Dessa forma, esta variável aparenta apresentar complexidade não compatível com os modelos aqui testados.

Um dos principais desafios dos programas de melhoramento atuais é a busca por realizar a fenotipagem de grandes populações, o que é agravado pela necessidade de amostrar em múltiplos ambientes com ensaios replicados (Gill et al., 2022). Lane & Murray (2021), em um estudo utilizando dados simulados de altura de planta, encontrou que, na comparação entre precisão e agilidade da fenotipagem, a fenotipagem de alto rendimento pode proporcionar melhores informações para o processo decisório de um programa de melhoramento. Isso se deve ao fato de que a

fenotipagem de alto rendimento possibilita avaliar uma ampla gama de genótipos e ambientes em um menor espaço de tempo, em comparação com a fenotipagem tradicional, que é mais precisa nas medições, mas avalia um menor número de genótipos e ambientes. Dessa forma, para as variáveis AIV e AP, os modelos de IC aqui avaliados podem ser utilizados na fenotipagem a partir de índices de vegetação, com destaque para RF1 e RF2.

Estes modelos podem ser avaliados quanto à capacidade de predição de características em estados iniciais de desenvolvimento de soja, com objetivo de antecipar o momento de seleção das plantas, reduzindo o tempo e o custo dos processos de avaliação nos programas de melhoramento (Araus et al., 2018; Tardieu et al., 2017; Rutkoski et al., 2016)

Apesar de, pelo menos para AIV e AP, os modelos poderem ser utilizados, pode-se avaliar metodologias para otimizar seus desempenhos, tais como métodos de seleção de variáveis dependentes e inputs (Yoosefzadeh-Najafabadi et al., 2021; Ramos et al., 2020; Hennessy et al., 2020; Zheng et al., 2020), otimização dos modelos via hiperparâmetros (Probst et al., 2019) e ampliação da rede de espectros (Maimaitijiang et al., 2020).

CONCLUSÃO

Todos os modelos apresentaram capacidade em prever as variáveis AIV e AP, com destaque para os modelos RF1 e RF2. No entanto, nenhum dos modelos foi capaz de prever com êxito as variáveis DM e PROD. Houve desempenho superior de todos os modelos em comparação com a regressão múltipla.

REFERÊNCIAS

ADAK, S.; BANDYOPADHYAY, K.K.; SAHOO, R.N.; MRIDHA, N.; SHRIVASTAVA, M.; PURAKAYASTHA, T.J. Prediction of Wheat Yield Using Spectral Reflectance Indices Under Different Tillage, Residue and Nitrogen Management Practices. **Curr. Sci.**, p. 1-12, 2021.

ARAUS, J. L.; KEFAUVER, S. C.; ZAMAN-ALLAH, M.; OLSEN, M. S.; CAIRNS, J. E. Translating high-throughput phenotyping into genetic gain. **Trend. Plant Sci.**, v. 23, p. 451–466, 2018. doi: 10.1016/j.tplants.2018.02.001.

BHERING, L. L.; CRUZ, C. D.; PEIXOTO, L. de A.; ROSADO, A. M.; LAVIOLA, B. G.; NASCIMENTO, M. Application of neural networks to predict volume in eucalyptus. **Crop. Breed. Appl. Biotechnol.**, v. 15, n. 3, p. 125–131, 2015. doi: 10.1590/1984-70332015v15n3a23.

BHERING, L. L. Rbio: A Tool For Biometric And Statistical Analysis Using The R Platform. **Crop. Breed. Appl. Biotechnol.**, v.17, p. 187-190, 2017.

BIAU, G.; SCORNET, E. A random forest guided tour. **TEST**, v. 25, p. 197–227, 2016. doi: 10.1007/s11749-016-0481-7.

CABELLO-SOLORZANO, K., ORTIGOSA DE ARAUJO, I., PEÑA, M., CORREIA, L., TALLÓN-BALLESTEROS, A. J. (2023). **The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis**. In: GARCÍA BRINGAS, P., et al. 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023). SOCO 2023. Lecture Notes in Networks and Systems, vol. 750. Springer, Cham. doi: 10.1007/978-3-031-42536-3_33.

COSTA, V. G.; PEDREIRA, C. E. Recent advances in decision trees: an updated survey. **Artif. Intell. Rev.**, v. 56, p. 4765–4800, 2023. doi: 10.1007/s10462-022-10275-5.

ELIZONDO, A.; MCCLENDON, R. W.; HOOGENBOOM, G. Neural Network Models for Predicting Flowering and Physiological Maturity of Soybean. **Trans. ASAE**, v. 37, n. 3, p. 981–988, 1994. DOI: 10.13031/2013.28168.

EUGENIO, F. C.; GROHS, M.; VENANCIO, L. P.; SCHUH, M.; BOTTEGA, E. L.; RUOSO, R.; SCHONS, C.; MALMANN, C. L.; BADIN, T. L.; FERNANDES, P.

Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. **Remote Sens. Appl.: Soc. Environ.** , v. 20, p. 100397, 2020. ISSN 2352-9385. doi: 10.1016/j.rsase.2020.100397.

FODOR, N.; CHALLINOR, A.; DROUTSAS, I.; RAMIREZ-VILLEGAS, J.; ZABEL, F.; KOEHLER, A.-K.; FOYER, C. H. Integrating Plant Science and Crop Modeling: Assessment of the Impact of Climate Change on Soybean and Maize Production. **Plant Cell Physiol.**, v. 58, n. 11, p. 1833–1847, November 2017. doi: 10.1093/pcp/pcx141.

FRITSCH, S.; GUENTHER, F; WRIGHT, M. N. **Neuralnet: Training of Neural Networks. R package version 1.44.6**, 2019.

GILL, T.; GILL, S. K.; SAINI, D. K. et al. A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping. **Phenomics**, p. 156–183, 2022. doi: 10.1007/s43657-022-00048-z.

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sens. Environ.**, v. 58, n. 3, p. 289-298, 1996. doi: 10.1016/S0034-4257(96)00072-7.

GUO, S.; ZHANG, Z.; GUO, E.; FU, Z.; GONG, J.; YANG, X. Historical and projected impacts of climate change and technology on soybean yield in China. **Agric. Syst.**, v. 203, 2022, p. 103522. ISSN: 0308-521X. doi: 10.1016/j.agsy.2022.103522.

HARA, P.; PIEKUTOWSKA, M.; NIEDBAŁA, G. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. **Land**, v. 10, p. 609, 2021. doi: 10.3390/land10060609.

HENNESSY, A.; CLARKE, K.; LEWIS, M. Hyperspectral classification of plants: a review of waveband selection generalisability. **Remote Sens.**, v. 12, p. 113, 2020. doi: 10.3390/rs12010113.

KAUR, B.; SANDHU, K.S.; KAMAL, R. et al. Omics for the improvement of abiotic,

biotic, and agronomic traits in major cereal crops: applications, challenges, and prospects. **Plants**, v. 10, p. 1989, 2021. doi: 10.3390/plants10101989.

KRUSE, R.; BORGELT, C.; BRAUNE, C.; MOSTAGHIM, S.; STEINBRECHER, M. **Computational intelligence**. Springer, 2022. doi: 10.1007/978-3-030-42227-1.

KUZBAKOVA, M.; KHASSANOVA, G.; OSHERGINA, I.; TEN, E.; JATAYEV, S.; YERZHEBAYEVA, R.; BULATOVA, K.; KHALBAYEVA, S.; SCHRAMM, C.; ANDERSON, P.; SWEETMAN, C.; JENKINS, C. L. D.; SOOLE, K. L.; SHAVRUKOV, Y. Height to first pod: A review of genetic and breeding approaches to improve combine harvesting in legume crops. **Front. Plant Sci.**, v. 13, 2022. doi: 10.3389/fpls.2022.948099. ISSN: 1664-462X.D.

LANE, H. M.; MURRAY, S. C. High throughput can produce better decisions than high accuracy when phenotyping plant populations. **Crop Science**, v. 61, p. 3301–3313, 2021. doi: 10.1002/csc2.20514.

LI, F.; MIAO, Y.; FENG, G.; YUAN, F.; YUE, S.; GAO, X.; LIU, Y.; LIU, B.; USTIN, S. L.; CHEN, X. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. **Field Crops Res.**, v. 157, p. 111-123, 2014. ISSN 0378-4290. doi: 10.1016/j.fcr.2013.12.018.

LI, Y.-H.; LI, W.; ZHANG, C.; YANG, L.; CHANG, R.-Z.; GAUT, B.S.; QIU, L.-J. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. **New Phytol.**, v. 188, p. 242-253, 2010. <https://doi.org/10.1111/j.1469-8137.2010.03344.x>

LIU, H. Q.; HUETE, A. A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. **IEEE Trans. Geosci Remote Sens.**, v. 33, n. 2, p. 457-465, 1995. doi: 10.1109/TGRS.1995.8746027.

MAIMAITIJIANG, M., SAGAN, V., SIDIKE, P., HARTLING, S., ESPOSITO, F., FRITSCHI, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote Sens. Environ.**, v. 237, p. 111599, 2020. doi: 10.1016/j.rse.2019.111599.

MARCILLO, G.S.; MARTIN, N.F.; DIERS, B.W.; DA FONSECA SANTOS, M.; LELES, E.P.; CHIGEZA, G.; FRANCISCHINI, J.H. Implementation of a Generalized Additive Model (GAM) for Soybean Maturity Prediction in African Environments. **Agronomy**, v. 11, p. 1043, 2021. doi: 10.3390/agronomy11061043.

MODGIL, R.; TANWAR, B.; GOYAL, A.; KUMAR, V. Soybean (*Glycine max*). In: TANWAR, B.; GOYAL, A. (eds). **Oilseeds: Health Attributes and Food Applications**. Springer, Singapore, 2021. doi: 10.1007/978-981-15-4194-0_1.

NIAZIAN, M.; NIEDBAŁA, G. Machine Learning for Plant Breeding and Biotechnology. **Agriculture**, v. 10, p. 436, 2020. doi: 10.3390/agriculture10100436.

OSCO, L. P.; JUNIOR, J. M.; RAMOS, A. P. M.; FURUYA, D. E. G.; SANTANA, D. C.; TEODORO, L. P. R.; GONÇALVES, W. N.; BAIO, F. H. R.; PISTORI, H.; JUNIOR, C. A. d. S. et al. Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. **Remote Sens.**, v. 12, p. 3237, 2020. doi: 10.3390/rs12193237.

PANDA, S.S.; AMES, D.P.; PANIGRAHI, S. Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. **Remote Sens.**, v. 2, n. 3, p. 673-696, 2010. DOI: 10.3390/rs2030673.

PRASANNA, B. M.; CAIRNS, J. E.; ZAIDI, P. H. Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. **Theor. Appl. Genet.**, v. 134, p. 1729–1752, 2021. doi: 10.1007/s00122-021-03773-7.

PROBST, P.; WRIGHT, M. N.; BOULESTEIX, A.-L. Hyperparameters and tuning strategies for random forest. **Rev.: Data Min. Know. Discov.**, 2019. doi: 10.1002/widm.1301.

RAMOS, F. T.; FERREIRA, L. S.; PIVETTA, F.; MAIA, J. C. S. Leaf blade area of different plants estimated by linear and dry matter measures, calibrated with the ImageJ software. **Interciência**, v. 40, n. 8, p. 570-575, 2015.

RAWAT, R.; PATEL, J. K.; MANRY, M. T. **Minimizing validation error with respect to network size and number of training epochs**. In: The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 2013, pp. 1-7. doi: 10.1109/IJCNN.2013.6706919.

RUTKOSKI, J.; POLAND, J.; MONDAL, S.; AUTRIQUE, E.; PÁREZ, L. G.; CROSSA, J. et al. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. **G3 (Bethesda)**, v. 6, p. 2799–2808, 2016. doi: 10.1534/g3.116.032888.

SANDHU, K.S.; AOUN, M.; MORRIS, C.F.; CARTER, A.H. Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. **Biol**, v. 10, p. 689, 2021. doi: 10.3390/biology10070689.

SEDIVY, E.J.; WU, F.; HANZAWA, Y. Soybean domestication: the origin, genetic architecture and molecular bases. **New Phytol**, v. 214, p. 539-553, 2017. <https://doi.org/10.1111/nph.14418>

SHAKOOR, N., LEE, S., MOCKLER, T. C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. **Curr. Opin. Plant Biol.**, v. 38, 2017, P. 184-192. ISSN: 1369-5266. doi: 10.1016/j.pbi.2017.05.006.

SUMNER, Z.; VARCO, J. J; DHILLON, J. S; FOX, A. A. A; CZARNECKI, J; HENRY, W. B. Ground versus aerial canopy reflectance of corn: red-edge and non-red edge vegetation indices. **Agron. J.**, v. 113, p. 2773-2788, 2021. doi: 10.1002/agj2.20657.

TAYADE, R.; YOON, J.; LAY, L.; KHAN, A.L.; YOON, Y.; KIM, Y. Utilization of Spectral Indices for High-Throughput Phenotyping. **Plants**, v. 11, p. 1712, 2022. <https://doi.org/10.3390/plants11131712>.

TEODORO, P.E.; TEODORO, L.P.R.; BAIO, F.H.R.; da SILVA JUNIOR, C.A.; dos SANTOS, R.G.; RAMOS, A.P.M.; PINHEIRO, M.M.F.; OSCO, L.P.; GONÇALVES,

W.N.; CARNEIRO, A.M.; et al. Predicting Days to Maturity, Plant Height, and Grain Yield in Soybean: A Machine and Deep Learning Approach Using Multispectral Data. **Remote Sens.**, v. 13, p. 4632, 2021. doi: 10.3390/rs13224632.

UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. **World Agricultural Production – Circular Series**, dez 2023. USDA. 2023.

VAN KLOMPENBURG, T.; KASSAHUN, A.; CATAL, C. Crop yield prediction using machine learning: A systematic literature review. **Comput. Electron. Agric.**, v. 177, 2020, p. 105709. ISSN 0168-1699. doi: 10.1016/j.compag.2020.105709.

WU, Q., QI, B., ZHAO, T.-J., YAO, X.-F., ZHU, Y., GAI, J.-Y. A tentative study on utilization of canopy hyperspectral reflectance to estimate canopy growth and seed yield in soybean. **Acta Agron. Sin.**, v. 39, p. 309, 2013. doi: 10.3724/SP.J.1006.2013.00309.

XIE, C.; YANG, C. A review on plant high-throughput phenotyping traits using UAV-based sensors. **Comput. Electron. Agric.**, v. 178, p. 105731, 2020. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2020.105731>

XUE, J.; SU, B. Significant Remote Sensing Vegetation Indices: A review of Developments and Applications. **J. Sens.**, v. 2017, p.17, 2017. <https://doi.org/10.1155/2017/1353691>

YANG, W., FENG, H., ZHANG, X., ZHANG, J., DOONAN, J. H., BATCHELOR, W. D., XIONG, L. e YAN, J. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. **Mol. Plant**, v. 13, p. 187–214, 2020.

YOOSEFZADEH-NAJAFABADI, M.; EARL, H. J.; TULPAN, D.; SULIK, J.; ESKANDARI, M. Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean. **Front. Plant Sci.**, v. 11, p. 624273, 2021. doi: 10.3389/fpls.2020.624273.

YUE, J.; YANG, G.; LI, C.; LI, Z.; WANG, Y.; FENG, H.; XU, B. Estimation of Winter Wheat Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Snapshot Hyperspectral Sensor and Crop Height Improved Models. **Remote Sens.**, v. 9, p. 708, 2017. doi: 10.3390/rs9070708.

ZHAO, H., LI, Y., CHEN, X. et al. Monitoring monthly soil moisture conditions in China with temperature vegetation dryness indexes based on an enhanced vegetation index and normalized difference vegetation index. **Theor. Appl. Climatol.**, v. 143, p. 159–176, 2021. doi: 10.1007/s00704-020-03422-x.

ZHENG, W.; ZHU, X.; WEN, G.; ZHU, Y.; YU, H.; GAN, J. Unsupervised feature selection by self-paced learning regularization. **Pattern Recognit. Lett.**, v. 132, p. 4–11, 2020. doi: 10.1016/j.patrec.2018.06.02.

CONCLUSÃO GERAL

Todos os parâmetros (número de épocas e número de folds) e elementos de topologia (número de neurônios e de camadas ocultas) de PMC avaliados revelaram-se influentes na capacidade de predição dos modelos. À medida que o número de folds e neurônios aumentou, até 15 e 10 respectivamente, observou-se incremento no desempenho dos modelos, enquanto o aumento do número de camadas ocultas resultou em uma diminuição do desempenho. Quanto ao número de épocas, houve aumento no desempenho até 30000 épocas, seguido de uma diminuição ao passar de 30000 para 50000 épocas. Os modelos PMC1, PMC2, RF1 e RF2 apresentaram êxito na predição das variáveis AIV e AP, havendo destaque para os modelos desenvolvidos com o algoritmo de Random Forest. Não houve modelo avaliado capaz de prever as variáveis DM e PROD.