

TEORIA DA SELEÇÃO FUNDAMENTADA EM APRENDIZADO DE MÁQUINA NO MELHORAMENTO DE FEIJÃO: OTIMIZAÇÃO DA ESTRATÉGIA SELETIVA POR MEIO DE CARACTERES ADAPTATIVOS

Mauro Bitencourt de Souza, Henrique de Sá Albino, Lucas Daniel Chaves, João Antonio Dalmagro, Paulo Henrique Cerutti, Luan Tiago dos Santos Carbonari, Carlos Zacarias Joaquim Júnior, Marissa Prá de Souza, Jefferson Luís Meirelles Coimbra

INTRODUÇÃO

A seleção de plantas foi e continua sendo essencial para o desenvolvimento de cultivares superiores. Essa prática, fundamental para o melhoramento de plantas, é realizada por duas vias, com a escolha de indivíduos superiores ou a eliminação dos inferiores (Falconer & Mackay, 1996). Porém, estas estratégias enfrentam limitações quando aplicadas a uma grande quantidade de genótipos, em virtude dos inúmeros dados gerados e aos reflexos acentuados do ambiente no fenótipo. Assim a utilização do aprendizado de máquina surge como alternativa promissora (Farooq et al., 2025), entretanto, o uso desta técnica em populações segregantes ainda é escasso. Deste modo, o objetivo do trabalho foi a verificação da eficiência do aprendizado de máquina na predição do desempenho de genótipos de feijão, por meio de caracteres adaptativos.

DESENVOLVIMENTO

Um ensaio foi realizado na área experimental do Instituto de Melhoramento e Genética Molecular (IMEGEM/CAV), no município de Lages. Foram utilizados genótipos (Genitores e populações segregantes nas gerações F₄ a F₇) oriundos de trabalhos anteriores do grupo de pesquisa (IMEGEM), com o intuito de formar um conjunto de dados para aplicação do aprendizado de máquina. Um total de 3.924 plantas deste ensaio foram mensuradas para os seguintes caracteres adaptativos, estatura da planta (cm), diâmetro do caule (mm), inserção do primeiro legume (cm), além do caráter massa de grãos por planta (g), totalizando 15,696 unidades informativas. A massa de grãos por planta foi particionada em quatro classes com base nos valores dos quartis obtidos com análise descritiva. Assim, as classes foram definidas do seguinte modo: *i*) Inferior (<4,99g); *ii*) Moderado (entre 4,99 e 8,10g); *iii*) Bom (entre 8,10 e 12,85g); e *iv*) Superior (>12,85g). O algoritmo Random Forest Classifier, descrito em Breiman (2001), foi empregado para prever o desempenho dos genótipos ($y = f(x_n)$) quanto a massa de grãos por planta (y), com base nos caracteres adaptativos (x_n), assim, os dados foram divididos em 70% para treino, 30% para teste do modelo (Géron, 2021; Fernandes et al., 2024; Sun et al., 2024). Para a avaliação do modelo, foram utilizadas as seguintes métricas: acurácia, precisão e sensibilidade. As análises foram realizadas na linguagem Python (versão 3.12.6), empregando as bibliotecas Pandas (Biblioteca para tabulação dos dados), Numpy (Biblioteca para cálculos), Scikit-learn (Biblioteca principal de aprendizado de máquina), Matplotlib e Seaborn (Bibliotecas para visualização dos dados).

RESULTADOS

A partir dos resultados obtidos com a avaliação do modelo de aprendizado de máquina, se verificou um valor para a acurácia de 0,52 (Tabela 1). Essa métrica é amplamente utilizada em trabalhos envolvendo aprendizado de máquina, revelando o desempenho geral do modelo de classificação (Sun et al., 2020). As estatísticas precisão e sensibilidade foram utilizadas também para a avaliação do desempenho do modelo localizada para cada classe. As classes intermediárias apresentaram valores menores de precisão (0,38 e 0,42) e sensibilidade (0,39 e

0,31). Por outro lado, as classes “Inferior” e “Superior” apresentaram maiores valores de precisão (0,57 e 0,66) e sensibilidade (0,62 e 0,75), respectivamente, indicando um desempenho superior para as classes localizadas nos extremos da distribuição de massa de grãos. Estes desempenhos contrastantes ocasionam um valor intermediário de acurácia, refletindo a média entre os desempenhos do modelo durante a classificação. Considerando que, em estágios mais avançados dos programas de melhoramento, a tendência é a intensificação da seleção com foco na eliminação de genótipos inferiores, o erro mais crítico é classificar indivíduos inferiores como promissores. Isso, porque a condução dessas populações segregantes acarreta custos desnecessários aos programas de melhoramento (Allard, 1971). Assim, a precisão indica a proporção de indivíduos verdadeiramente superiores, demonstrando que entre as plantas superiores previstas pelo modelo, 34% dos genótipos classificados à classe superior foram erroneamente classificados. O erro em classificar plantas superiores como inferiores pode ser crítico devido a possibilidade de constituições genotípicas superiores serem erroneamente descartadas. Diante disso, a sensibilidade pode ser utilizada na obtenção da magnitude deste erro. Pois ela compreende a razão entre os indivíduos verdadeiramente superiores que foram identificados pelo modelo sobre o total de indivíduos pertencentes a classe superior indicando um valor de 25% de genótipos superiores descartados. Com base nessa premissa, esse tipo de equívoco é menos prejudicial nas gerações em que os genótipos estão presentes no estudo (F₄ a F₇), em comparação a manutenção de genótipos pouco promissores. Visto que a condução de genótipos inferiores até fases avançadas implica em diversos custos elevados, como tempo, área experimental e recursos humanos. Sendo assim, o emprego de técnicas de aprendizado de máquina pode mitigar esse problema, contribuindo com a previsão de genótipos superiores com maior eficiência, seja em gerações iniciais ou avançadas no processo de melhoramento.

CONSIDERAÇÕES FINAIS

O aprendizado de máquina pode auxiliar o melhorista no processo de seleção utilizando caracteres adaptativos para a previsão de genótipos superiores e inferiores. Adicionalmente, a abordagem recomendada para os genótipos deste estudo é priorizar a seleção dos melhores indivíduos evitando o acúmulo de genótipos não promissores.

Palavras-chave: ciência de dados; inteligência artificial; *Phaseolus vulgaris* L.; melhoramento de plantas.

ILUSTRAÇÕES

Tabela 1. Resultados do modelo de classificação construído com o algoritmo de Random Forest com base nos caracteres adaptativos. CAV/UDESC. Lages, 2025.

Classe	Precisão	Sensibilidade
Inferior	0,57	0,62
Moderado	0,38	0,39
Bom	0,42	0,31
Superior	0,66	0,75
Acurácia	0,52	

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLARD, R. W. **Princípios do melhoramento genético das plantas**. 2. ed. São Paulo: Edgard Blücher, 1971.
- BREIMAN, L. Random forests. **Machine Learning**, Dordrecht, v. 45, n. 1, p. 5-32, 2001
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. Harlow: Longman Scientific & Technical; New York: Wiley, 1996.
- FAROOQ, M. A.; GAO, S.; HASSAN, M. A.; HUANG, Z.; RASHEED, A.; HEARNE, S.; PRASANNA, B.; LI, X.; LI, H. Artificial intelligence in plant breeding. **Trends in Genetics**, v. 40, n. 10, p. 891-908, out. 2024.
- FERNANDES, I. K.; VIEIRA, C. C.; DIAS, K. O. G.; FERNANDES, S. B. Using machine learning to combine genetic and environmental data for maize grain yield predictions across multi-environment trials. **Theoretical and Applied Genetics**. v. 137, n. 189, 2024.
- GÉRON, A. **Mãos à obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow: conceitos, ferramentas e técnicas para a construção de sistemas inteligentes**. 2. ed. São Paulo: Alta Books, 2021.
- SUN, K.; LIU, C.; HAN, J.; ZHANG, J.; QI, Y. Phenotypic detection of flax plants based on improved Flax-YOLOv5. **Frontiers in Plant Science**, v. 15, p. 1-14, 2024
- SUN, S.; WANG, C.; DING, H.; ZOU, Q. Machine learning and its applications in plant molecular studies. **Briefings in Functional Genomics**, v. 19, n. 1, p. 40-48, 2020.

DADOS CADASTRAIS

BOLSISTA: Mauro Bitencourt de Souza

MODALIDADE DE BOLSA: PROBIC/UDESC

VIGÊNCIA: 09/2024 a 08/2025 – Total: 12 meses

ORIENTADOR(A): Jefferson Luís Meirelles Coimbra

CENTRO DE ENSINO: CAV

DEPARTAMENTO: Agronomia

ÁREAS DE CONHECIMENTO: Ciências Agrárias / Agronomia

TÍTULO DO PROJETO DE PESQUISA: Reflexos do ambiente versus métodos de fenotipagem no sistema radicular em feijão.

Nº PROTOCOLO DO PROJETO DE PESQUISA: NPP3747-2021