POLLYANNA CAPOBIANGO DA FONSECA

MAIZE COMMON RUST RESISTANCE CLASSIFICATION WITH MACHINE LEARNING ANALYZES

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

Orientador: Aluízio Borém de Oliveira

Coorientador: Francelino Augusto R. Junior

VIÇOSA – MINAS GERAIS
2023

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

T
F676m
2023

Fonseca, Pollyanna Capobiango da, 1991-
Maize common rust resistance classification with machine learning analyzes / Pollyanna Capobiango da Fonseca. – Viçosa, MG, 2023.
1 tese eletrônica (54 f.): il. (algumas color.).

Orientador: Aluízio Borém de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2023.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.202
Modo de acesso: World Wide Web.

1. Milho - Doenças e pragas. 2. Ferrugem-comum. 3. Aprendizado do computador. 4. Mineração de dados (Computação). I. Oliveira, Aluízio Borém de, 1959-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 633.1594

Bibliotecário(a) responsável: Alice Regina Pinto Pires CRB-6/2523

POLLYANNA CAPOBIANGO DA FONSECA

MAIZE COMMON RUST RESISTANCE CLASSIFICATION WITH MACHINE LEARNING ANALYZES

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

APROVADA: 23 de fevereiro de 2023.

Assentimento:

Documento assinado digitalmente
gov.br
POLLYANNA CAPOBIANGO DA FONSECA
Data: 03/05/2023 14:37:32-0300
Verifique em https://validar.iti.gov.br

Pollyanna Capobiango da Fonseca
Autora

Documento assinado digitalmente
gov.br
ALUIZIO BOREM DE OLIVEIRA
Data: 03/05/2023 17:23:41-0300
Verifique em https://validar.iti.gov.br

Aluízio Borém de Oliveira
Orientador

AGRADECIMENTOS

Agradeço a Deus por ser guia e proteção ao longo dessa jornada.

À família e aos amigos pela assistência e ensinamentos.

Ao professor Aluízio Borém pela confiança, apoio, conselhos e tempo empenhado em compartilhar seu conhecimento. Ao pesquisador Francelino Rodrigues pelo constante apoio, incentivo e experiências compartilhadas no desenvolvimento desse trabalho. A ambos agradeço o exemplo pessoal e profissional.

À Universidade Federal de Viçosa, pela oportunidade de realizar a pós-graduação.

Ao Programa de Pós-graduação em Genética e Melhoramento, pela oportunidade de realizar o curso e me preparar para o mercado de trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

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

Quem é apenas bravo tem só um assomo,
quem é apenas valente tem só um temperamento,
quem é apenas corajoso tem só uma virtude:
o obstinado na verdade tem a grandeza.
Quase todo segredo dos grandes corações está nesta palavra: "perseverando".
A perseverança está para coragem como a roda para a alavanca:
é a renovação perpétua do ponto de apoio.
Os Trabalhadores do Mar - Victor Hugo

ABSTRACT

FONSECA, Pollyanna Capobiango da, D.Sc., Universidade Federal de Viçosa, February, 2023. **Maize common rust resistance classification with machine learning analyzes.** Adviser: Aluízio Borém de Oliveira. Co-adviser: Francelino Augusto Rodrigues Junior.

Maize (*Zea mays ssp.* Mays) is a widely cultivated crop, having one of the highest productivities among cereals, and it is of great importance in human consumption, both *in natura* and processed. In addition, it has applications in industry as a source of energy through corn ethanol and animal feed. Many diseases can affect maize yield such as the Maize Common Rust (MCR) (*Puccinia sorghi* Schwein), a leaf disease which causes the appearance of pustules. The aim of this study was to classify maize lines between resistant and susceptible, selecting 50% of them to be carried on the breeding pipeline. A dataset containing three time-point evaluations in two years using a visual score scale and two Unmanned Aerial Vehicle (UAV) - couple sensors (multispectral and thermal) data were analyzed with six machine learning algorithms in order to identify the training time set to deliver the best classification performance. The three time-point evaluations phenotypic data along with the genetic markers data were used to explore the performance of the Support Vector Machine (SVM) and the Artificial Neural Network (ANN) algorithms in a k-fold cross-validation analysis with nine datasets. Their learning curves and feature importance rank were analyzed using the SVM algorithm. Our results showed that the last evaluation training set delivered the highest accuracies, of approximately 80 per cent, with Logistic Regression and SVM outperforming the other algorithms. The results obtained with the analysis by year suggest that a homogenous distribution of scores is of great importance for an effective MCR resistance classification. Our results also demonstrated the advantageous use of the SVM algorithm, in which models had the capacity to generalize using a smaller number of features. Similar performance metrics were achieved with SVM when the third evaluation and the three time-point evaluations combined together were employed. The SVM learning curves indicate that the addition of more training samples would be beneficial for all datasets analyzed. The five most important features for each dataset were listed, resulting in a predominance of the Red wavelength in the first position of the rank. In addition, the protein-coding genes aligned with the markers' allele sequence ranked as important should be further explored in genomic-functional studies.

Keywords: Maize common rust. Machine learning. SVM. ANN. Data mining

RESUMO

FONSECA, Pollyanna Capobiango da, D.Sc., Universidade Federal de Viçosa, fevereiro de 2023. **Classificação da resistência à ferrugem comum do milho via análise por machine learning.** Orientador: Aluízio Borém de Oliveira. Coorientador: Francelino Augusto Rodrigues Junior.

O milho (*Zea mays ssp.* Mays) é uma espécie amplamente cultivada, com uma das maiores produtividades entre os cereais, e de grande importância na alimentação humana *in natura* ou processado. Além disso, é usado na indústria como fonte de energia, através do etanol de milho, e ração animal. Muitas doenças podem afetar a produtividade do milho, como a ferrugem comum do milho (MCR) (*Puccinia sorghi* Schwein), uma doença foliar que causa o aparecimento de pústulas. O objetivo deste estudo foi classificar linhagens de milho entre resistentes e suscetíveis, selecionando 50% delas para serem usadas em etapas subsequentes do programa de melhoramento. Três avaliações em tempos diferentes do ciclo de desenvolvimento do milho foram realizadas em dois anos usando escala de avaliação visual e dois sensores (multiespectral e termal) acoplados ao *Unmanned aerial Vehicle* (UAV) e seus respectivos dados foram analisados usando seis algoritmos de aprendizado de máquinas para identificar a avaliação que fornece a melhor performance de classificação da resistência. Os dados fenotípicos das três avaliações e os dados de marcadores genéticos foram usados para explorar o desempenho dos algoritmos *Support Vector Machine* (SVM) e *Artificial Neural Network* (ANN) em uma análise de validação cruzada k-fold com nove conjuntos de dados. As curvas de aprendizado e ranking de importância das variáveis preditoras foram analisados usando o algoritmo SVM. Os resultados mostraram que a terceira avaliação forneceu os maiores valores de acurácia, de aproximadamente 80%, com *Logistic Regression* e SVM superando os demais algoritmos. Os resultados obtidos com a análise por ano sugerem que uma distribuição homogênea de notas visuais é de grande importância para uma classificação eficaz da resistência à MCR. Os resultados também demonstraram que o uso do algoritmo SVM será vantajoso dependendo da capacidade do modelo de generalizar usando um número menor de variáveis preditoras. Métricas de desempenho semelhantes foram alcançadas usando SVM quando a terceira avaliação e as três avaliações combinadas foram empregadas. As curvas de aprendizado usando o algoritmo SVM indicam que a adição de mais amostras de treinamento seria benéfica para todos os conjuntos de dados analisados. As cinco variáveis preditoras mais importantes para cada conjunto de dados foram listadas, resultando em uma predominância do comprimento de onda vermelho na primeira posição do ranking. Além

disso, os genes codificadores de proteínas alinhados com a sequência alélica dos marcadores de importância ranqueados devem ser explorados posteriormente em estudos genômico-funcionais.

Palavras-chave: Ferrugem comum do milho. Aprendizado de máquinas. SVM. ANN. Mineração de dados

SUMMARY

1. GENERAL INTRODUCTION

1.1 UAV-based high-throughput phenotyping

Remote sensing technologies applied to access plant traits can reduce the time and cost of obtaining phenotypic data while increasing the accuracy, precision, and reproducibility of evaluations under greenhouse or field conditions.

Image-based phenotyping methods use spectral signatures to distinguish light reflected in different parts of the electromagnetic spectrum based on the spectral properties of the material. For example, vegetation, water, and soil have distinct spectral signatures due to their different reflectance. Changes in plant physiology, such as biotic stress, can alter tissue color, transpiration rate, and canopy morphology, altering their respective spectral signatures (West et al., 2010). High-throughput phenotyping (HTP) platforms coupled with sensors (*e.g.*, RGB, thermal and multispectral cameras) can capture images in different bands of the electromagnetic spectrum.

Images collected in the visible light wavelength range (400 to 700 nm) are greatly influenced by the leaf pigment content (Mahlein, 2016). The RGB images are commonly used to access morphological plant traits or organs, biomass, and plant growth (W. Yang et al., 2020). In two different years, for example, an Unmanned aerial vehicle (UAV) coupled with an RGB sensor was used to acquire images to access vegetative indices as well as to detect and quantify the severity of wheat leaf rust (Bhandari et al., 2020).

Multispectral sensors are widely used in the detection and monitoring of plant diseases. Multispectral sensors are equipped with lenses that can access specific reflected wavelengths, allowing the collection of important spectral bands for data analysis. Because distinct patterns of spectral responses can be found for resistant and susceptible plants, changes in the spectral response curves are an important factor to consider when studying the spectral response of plants to diseases (Zhang et al., 2019). Although the use of UAV-coupled sensors offers an opportunity to improve disease-scoring efficiency, additional investigation into the spectral signature of healthy and stressed plants at all plant growth stages may be necessary to obtain trustful measurements (McNish & Smith, 2021). Multispectral sensors coupled to UAVs are extensively used in the study of plant diseases, such as ramularia leaf blight in cotton (Xavier et al., 2019).

Thermal images from thermal sensors can be used to determine plant or leaf temperature (W. Yang et al., 2020). Its importance is due to the fact that the temperature of

plant tissues varies depending on the infection stage and the type of infection caused by the pathogen; additionally, the canopy temperature can be negatively correlated with the transpiration rate (Mutka & Bart, 2015). Because the transpiration rate is related to leaf water potential and stomatal conductance, thermal images could be used to monitor abiotic and biotic stresses (G. Yang et al., 2017). Some pathogens induce a reduction in the stomatal opening, generally causing a decrease in transpiration rate and an increase in leaf temperature. Because thermal images, depending on the disease, do not provide enough spectral information, combining different imaging technologies has the potential to provide greater efficiency in detecting and quantifying changes caused by pathogens (W. Yang et al., 2020). For example, multispectral and thermal sensors combined have been successfully used in the identification, monitoring, and prediction of ascochyta blight in chickpea (Zhang et al., 2019).

1.2 Maize genotyping and phenotyping

The maize reference genome was sequenced using as a source the B73 inbred line (Jiao et al., 2017). The sequencing of other maize lines with different genetic backgrounds, such as the HuangZaoSi maize line, is important due to their high genetic variability (Li et al., 2019). Sequencing a large number of lines aids in the identification of structural variants as well as in the interconnection between genetic and phenotypic variability (Liu et al., 2020). Furthermore, it is relevant in the time maximization of maize breeding and in genomic studies.

Maize genomic data is used to develop haplotype maps containing a wide genetic variability from maize lines, cultivars, and closely related species. The HapMap3 consortium, for example, was developed from the sequencing of 1,218 strains, and 83 million VNPs (variation nucleotide polymorphisms) were identified (Bukowski et al., 2018). Likewise, genome-wide association studies (GWAS) use genomic data, such as those provided by GBS (genotyping by sequencing), to identify linkage disequilibrium blocks, VNPs, and putative genes. For example, maize resistance to *Fusarium verticillioides* was studied with GWAS using 226,446 VNPs after data quality processing, including imputation, call rate, and minor allele frequency (Stagnati et al., 2019).

The high genetic variability of maize and its close relatives can be used in genomic studies. These studies often use techniques such as hybridization, genetic transformation, mapping of VNPs and QTLs, induction of haploidy, and CRISPR/Cas9 (Liu et al., 2020). The overexpression of the transcription factor *zmm28* and the interaction of the produced protein

with promoters of several genes altered maize transcription regulation, positively influencing photosynthetic capacity, nitrogen usage efficiency, and increased growth and yield (Wu et al., 2019). Genomic data platforms such as MaizeGDB, MaizeGO, and qTellerMaizeGDB play a key role in integrating the large volume of data generated in various studies. Thus, information retrieved from publicly available databases on maize genetics and functional variability can help us comprehend the genotype-phenotype relationship.

Maize phenotyping has been studied for several applications in maize breeding programs such as abiotic and biotic stresses. The spatial variation of nitrogen deficiency tolerance, for example, was accessed in an experimental field using a UAV-coupled multispectral sensor to build vegetative indices, resulting in a strong correlation with senescence and grain yield (Zaman-Allah et al., 2015). In addition, in the study of parameters related to maize tar spot complex resistance using UAV-coupled multispectral and thermal sensors, a high correlation was found between yield and area under the disease progress curve (Loladze et al., 2019).

1.3 Maize common rust

Maize common rust (MCR) (*Puccinia sorghi* Schwein) causes the appearance of brown to reddish-brown pustules (Ramirez-Cabral et al., 2017). Temperatures between 16 and 23°C and high humidity favor spore germination and leaf mesophyll colonization (Rochi et al., 2016). This pathogen has as alternative hosts plants of the genus *Oxalis*, which are sources of sexual genetic variability since the generation of new physiological variants occurs even if there is no maize being cultivated. This characteristic contributes to the adaptation of the fungus to different climatic conditions and locations, besides potentially reducing resistance durability. Protective measures known are the use of resistant cultivars, crop rotation, eradication of alternative hosts, chemical control, and the use of resistant cultivars.

Our decision to use high-throughput phenotyping and high-throughput genotyping data in this project for ML analyzes illustrates our belief that they are promising techniques to support strategic decisions which will impact, for example, a project's structure, backlog, lead time, personnel requirement, and financial return.

1.4 References

Bhandari, M., Ibrahim, A. M. H., Xue, Q., Jung, J., Chang, A., Rudd, J. C., Maeda, M., Rajan,

N., Neely, H., & Landivar, J. (2020). Assessing winter wheat foliage disease severity using aerial imagery acquired from small Unmanned Aerial Vehicle (UAV). *Computers and Electronics in Agriculture*, *176*, 105665. https://doi.org/https://doi.org/10.1016/j.compag.2020.105665

Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., Wang, B., Xu, D., Yang, B., Xie, C., Fan, L., Gao, S., Xu, X., Zhang, G., Li, Y., Jiao, Y., Doebley, J. F., Ross-Ibarra, J., Lorant, A., … Xu, Y. (2018). Construction of the third-generation Zea mays haplotype map. *GigaScience*, *7*(4), 1–12. https://doi.org/10.1093/gigascience/gix134

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C. S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Li, C., Song, W., Luo, Y., Gao, S., Zhang, R., Shi, Z., Wang, X., Wang, R., Wang, F., Wang, J., Zhao, Y., Su, A., Wang, S., Li, X., Luo, M., Wang, S., Zhang, Y., Ge, J., Tan, X., … Zhao, J. (2019). The HuangZaoSi Maize Genome Provides Insights into Genomic Variation and Improvement History of Maize. *Molecular Plant*, *12*(3), 402–409. https://doi.org/10.1016/j.molp.2019.02.009

Liu, J., Fernie, A. R., & Yan, J. (2020). The Past, Present, and Future of Maize Improvement: Domestication, Genomics, and Functional Genomic Routes toward Crop Enhancement. *Plant Communications*, *1*(1), 100010. https://doi.org/10.1016/j.xplc.2019.100010

Loladze, A., Rodrigues, F. A., Toledo, F., San Vicente, F., Gérard, B., & Boddupalli, M. P. (2019). Application of Remote Sensing for Phenotyping Tar Spot Complex Resistance in Maize. In *Frontiers in Plant Science* (Vol. 10, p. 552). https://www.frontiersin.org/article/10.3389/fpls.2019.00552

Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease*, *100*(2), 1–11. https://doi.org/10.1007/s13398-014-0173-7.2

McNish, I. G., & Smith, K. P. (2021). Oat Crown Rust Disease Severity Estimated at Many Time Points Using Multispectral Aerial Photos. *Phytopathology®*, *112*(3), 682–690. https://doi.org/10.1094/PHYTO-09-20-0442-R

Mutka, A. M., & Bart, R. S. (2015). Image-based phenotyping of plant disease symptoms. *Frontiers in Plant Science*, *5*, 1–8. https://doi.org/10.3389/fpls.2014.00734

Ramirez-Cabral, N. Y. Z., Kumar, L., & Shabani, F. (2017). Global risk levels for corn rusts (Puccinia sorghi and Puccinia polysora) under climate change projections. *Journal of Phytopathology*, *165*(9), 563–574. https://doi.org/10.1111/jph.12593

Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R., Cuyeu, A. R., Turjanski, A., Kreff, E. D., & Sacco, F. (2016). Characterization and comparative analysis of the genome of Puccinia sorghi Schwein, the causal agent of maize common rust. *Fungal Genetics and Biology*, *112*(16), 31–39.

https://doi.org/10.1016/j.fgb.2016.10.001

Stagnati, L., Lanubile, A., Samayoa, L. F., Bragalanti, M., Giorni, P., Busconi, M., Holland, J. B., & Marocco, A. (2019). A genome wide association study reveals markers and genes associated with resistance to fusarium verticillioides infection of seedlings in a maize diversity panel. *G3: Genes, Genomes, Genetics*, *9*(2), 571–579. https://doi.org/10.1534/g3.118.200916

West, J. S., Bravo, C., Oberti, R., Moshou, D., Ramon, H., & McCartney, H. A. (2010). Detection of Fungal Diseases Optically and Pathogen Inoculum by Air Sampling. In E.-C. Oerke, R. Gerhards, G. Menz, & R. A. Sikora (Eds.), *Precision Crop Protection - the Challenge and Use of Heterogeneity* (pp. 135–149). Springer Netherlands. https://doi.org/10.1007/978-90-481-9277-9_9

Wu, J., Lawit, S. J., Weers, B., Sun, J., Mongar, N., van Hemert, J., Melo, R., Meng, X., Rupe, M., Clapp, J., Collet, K. H., Trecker, L., Roesler, K., Peddicord, L., Thomas, J., Hunt, J., Zhou, W., Hou, Z., Wimmer, M., … Habben, J. E. (2019). Overexpression of zmm28 increases maize grain yield in the field. *Plant Biology*, *116*(47), 23850–23858. https://doi.org/10.1073/pnas.1902593116

Xavier, T. W. F., Souto, R. N. V, Statella, T., Galbieri, R., Santos, E. S., S. Suli, G., & Zeilhofer, P. (2019). Identification of Ramularia Leaf Blight Cotton Disease Infection Levels by Multispectral, Multiscale UAV Imagery. In *Drones* (Vol. 3, Issue 2). https://doi.org/10.3390/drones3020033

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., & Yang, H. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Frontiers in Plant Science*, *8*(June). https://doi.org/10.3389/fpls.2017.01111

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular Plant*, *13*(2), 187–214. https://doi.org/10.1016/j.molp.2020.01.008

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., Hornero, A., Albà, A. H., Das, B., Craufurd, P., Olsen, M., Prasanna, B. M., & Cairns, J. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods*, *11*(1), 1–10. https://doi.org/10.1186/s13007-015-0078-2

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., & Huang, W. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, *165*, 104943. https://doi.org/https://doi.org/10.1016/j.compag.2019.104943

2. ARTICLE 1: MAIZE COMMON RUST RESISTANCE CLASSIFICATION WITH UAV-BASED SENSORS DATA USING TIME-SERIES MACHINE LEARNING ANALYZES

2.1 Abstract

This study is aimed at investigating maize common rust (MCR) disease resistance or susceptibility classification by identifying the best training time of three time-point evaluations using Machine Learning (ML) algorithms. The training datasets were used to develop ML models to be tested with the last time-evaluation set, aiming to reduce the number of evaluations to less than the three implemented in our experiments and to evaluate MCR later in the crop cycle only through imaging. The classification performance of the three time-point evaluations over two years was accessed using the UAV (unmanned aerial vehicle)-derived spectral reflectance and surface temperature data as the predictor variables and the resistant/susceptible annotated visual score as the predicted variable. Based on the time-point evaluation, four training sets were constructed. These datasets were analyzed using six different supervised classification algorithms. Following the identification of the time-point evaluation that delivered the highest accuracy across models, four further sets were constructed, splitting its data by year. In addition, a cross-validation analysis was implemented in that set with the purpose of comparing the ML algorithms relative standard deviation. The prediction analysis of the time training sets showed that no other set performance was closer to the last evaluation of approximately 80 per cent accuracy across most algorithms. The analysis of the last evaluation by year resulted in a visual score distribution discrepancy among years, which might have played a role in its distinct performance results and indicates the importance of a multi-year approach for MCR resistance classification. Logistic Regression and Support Vector Machine algorithms stood out in the cross-validation analysis due to their consistent accuracy and lowest relative standard deviation compared with other algorithms and its use should be considered in future studies.

Keywords: Maize leaf disease, Machine learning, Classification, Resistance prediction, Temporal data mining

2.2 Introduction

Maize common rust (MCR) (*Puccinia sorghi* Schwein) is a foliar disease-causing damage to maize at any stage of its vegetative or reproductive development. This disease is usually found coexisting with maize, but epidemics have occurred and are expected to become more serious according to climate change risk projections (Ramirez-Cabral et al., 2017). One of the most recommended methods for MCR control is the use of resistant cultivars.

Plant breeding programs use multidisciplinary approaches to address the growing challenges of evaluating and selecting the most promising genotypes within plant breeding trials. For example, modern high-throughput phenotyping (HTP) techniques for large-scale disease screening can maximize its time-consuming nature. Traditional phenotyping is considered laborious, time-consuming, and usually destructive, while HTP emerges as an approach to manage such limitations. Remote HTP platforms are used, for example, to estimate phenotypic data of a single plant over time or to assess crop canopy more quickly and accurately (Mutka et al., 2016). For example, the nitrogen deficiency tolerance in maize was accessed in an experimental field using a multispectral sensor coupled to a UAV (Unmanned aerial vehicle), and indices were built, resulting in a strong correlation with senescence and grain yield (Zaman-Allah et al., 2015). The HTP technology is also used to explore biotic stresses. In a study to access parameters related to resistance of maize to the tar spot complex using multispectral and thermal sensors, for instance, a high correlation was found between grain yield, a vegetative index, and canopy temperature (Loladze et al., 2019).

Prediction is a critical task in large-scale breeding programs. In predicting categorical values or class labels, we make use of prediction algorithms for classification tasks. It can take advantage of datasets containing historical data encompassing intrinsic patterns that ML (Machine Learning) prediction algorithms might be able to withdraw to deliver a more appropriate comprehension of the phenomenon studied, depending on the dataset structure and the algorithm's assumptions (Parmezan et al., 2019). Multi-year or multi-season evaluations aid in better understanding the variability of phenotypic assessments due to distinct environment and field conditions over time. Multi-time evaluation in the same year may help to comprehend, for instance, plant growth or disease development and to select a target time to perform evaluations. The multi-time prediction approach has been applied to yield and disease studies with large databases. For example, Cheng et al. (2022) investigated

maize yield prediction with the purpose of identifying a time earlier in the growth cycle when maturity accuracy performance is sustained.

Most recent studies complement the ML classification analyzes with comparisons between algorithms. Neelakantan (2021) used a complex image dataset of five tomato diseases to confront six different supervised ML algorithms. In a metanalysis exploration, Uddin et al. (2019) compared the frequency of employment and accuracy results of seven ML algorithms used in human diseases prediction. Therefore, historical HTP data is feasible to be analyzed with a multimodal ML prediction approach to aid in phenotype classification and to speed up breeding time.

This temporal data mining study proposes an MCR resistant/susceptible classification by comparing three time-point evaluations from two years using six ML algorithms. Our aims were to: (1) identify the best training time-evaluation set for MCR resistance prediction through the use of supervised multimodal ML analyzes; (2) investigate the year effect in the identified set by breaking it down to compare prediction results; and (3) determine which of the ML algorithms employed delivered a higher cross-validation performance using the set identified in (1). By identifying the best training time-evaluation and ML algorithm employed, this approach also aims to optimize the number and timing of evaluations and to maximize computer use, pre-processing time, and the final analysis efficiency of MCR resistance classification.

2.3 Materials and methods

2.3.1 Data origin

The visual score and remotely sensed spectral reflectance and surface temperature data were obtained from CIMMYT's (International Maize and Wheat Improvement Center) trials of maize lines screened for MCR resistance. Details about the plant material origin, experimental design, inoculation, disease score evaluation, image acquisition, and processing can be found meticulously described by Loladze et al. (2023). Briefly, three biparental populations of doubled haploid maize lines with two replicates were tested in 2019 and 2020 evaluation trials. Crosses CHWTI23/DTMA-85, CHWTI59/DTMA-85, and DTMA-17/DTMA-85 and their parental lines were evaluated in a total of 900 entries at three different time-points, which are approximately the reproductive stages R1, R2, and R3 (Ritchie et al., 1986). Each replicate was visually evaluated with scores varying from 1 (very resistant) to 9 (very susceptible), and the entry's average score was calculated for further processing. Images

were gathered using a UAV platform with multispectral and thermal infrared sensors. Within-plot average spectral reflectance of Green (550nm ± 40nm), Red (660nm ± 40nm), Red-edge (735nm ± 10nm), and NIR (790nm ± 40nm), and average surface temperature of Thermal (7.5–13.5μm) bandwidth were retrieved from the orthomosaics constructed for the three time-point score evaluations each year so that the five bandwidths could be used as predictor variables in the current analysis.

2.3.2 Visual scores and UAV-derived data analysis

The MCR visual scores average data of two replicates for each maize line were analyzed for evaluation per year through descriptive statistics so that we could understand the score frequency distribution and interpret the ML models' results. The original visual scores were also analyzed with Pearson's correlation in order to comprehend the statistical relationship between the three score evaluations by year and to make sense of the approach of selecting a time-point evaluation to be used as training set. Pearson's correlation of spectral reflectance and surface temperature data were analyzed for each time-point evaluation using the data from both years. Besides, using the two-year third evaluation data, the one hundred maize lines with the highest reflectance data for NIR and Red-edge wavelengths, and the lowest values for the Red wavelength were ranked to retrieve the number of maize lines annotated as resistant and susceptible phenotypes.

The visual scores and UAV-derived data analysis were accomplished using Python 3.8 and the software libraries Numpy 1.19.2, Pandas 1.4.3, and Scipy 1.6.2. The software Matplotlib 3.3.3, Seaborn 0.11.1, and Python's API module Pylab were used for data visualization.

2.3.3 Visual score and UAV-derived data annotation and pre-processing

In each of the three time-point evaluations, we expected to prevail a certain range of scores because of the different phases of disease development accessed. Therefore, maize lines data annotation of each evaluation set was implemented according to its distribution. We used each dataset second quartile (Q2) as the threshold to categorize maize lines between resistant and susceptible, aiming to select 50% of the lines to be carried out in the breeding program. For each training and test set, the threshold used was: (1) if the average visual score > Q2, the maize line was categorized as susceptible; (2) if average visual score ≤ Q2, the line

was categorized as resistant. The same rules were applied for all sets, regardless of whether the dataset contained one- or two-years' data. The chosen threshold might change in a breeding program depending, for example, on the number of maize lines available in the dataset, the disease severity in the season, and the breeder selection target.

The four spectral reflectance and the temperature surface data pre-processing was done with the software Scikit-learn 1.1.1 min-max scaler, which standardizes the ML predictor variable data by scaling it into a chosen range. The training set range of values we chosen was between 0 and 1, mathematically:

$$X_{scaled} = {(X - X_{min})}/{(X_{max} - X_{min})}$$

where for each maize line X_{scaled} is the standardized value of a predictor variable, X is the original value of the variable, and X_{max} and X_{min} are the maximum and minimum values of the variable for the entire training set.

Because it is not a model requirement, sensors' data used with the classification algorithms Decision Tree and Random Forest were not normalized. Parameter tuning was made based on the three evaluations in both years data.

2.3.4 ML algorithms for classification analysis

The K Nearest Neighbor (kNN) ML algorithm uses a measure of distance, such as Euclidian distance, to find the lowest distance between the new sample and k others existent on the set (Panigrahi et al., 2020). The kNN algorithm requires scaling, has a high computational cost, and is slow to perform predictions with a large training dataset. This algorithm does not allow the construction of a model. Instead, it calculates the distance and updates the classification set with the new sample class label. Python's software ML library Scikit-learn 1.1.1 was used to implement the algorithms used for this model and the next four. The KNeighborsClassifier algorithm was used with all default parameters but 20 as the number of neighbors.

The Logistic Regression (LR) algorithm is most used for linear classification problems. It uses a sigmoid function-based model to predict which class the sample should be labelled in (Ibrahim & Abdulazeez, 2021). The model's input variables should not have

complex relationships to deliver good predictions (Uddin et al., 2019). This algorithm has a low computational cost, and feature scaling is required. The LogisticRegression algorithm was implemented with all default parameters.

The Support Vector Machine (SVM) method "learns" an optimal hyperplane that maximizes the margin among classes and is used for complex tasks with linear or non-linear data for classification and regression problems (Iniyan et al., 2020). It is possible to use a range of kernel types, such as linear radial basis, gaussian, polynomial, and tangent hyperbolic. The SVM algorithm is considered slow to perform predictions and is a black box method, not intuitive and not easy to visualize its results graphically. The SVC algorithm was used with all default parameters but a linear kernel and C value of 3.0.

The Decision Tree (DT) algorithm uses rules learned from the training data to build a classification or regression model. A rank of attributes is created based on its entropy and gain and used in constructing the supervised model. Because larger trees are challenging to draw conclusions from, it is usual to leave less important attributes out of the analysis. The DecisionTreeClassifier algorithm was implemented with all default parameters but the criterion entropy.

Random Forest (RF) is an ensemble model derived from DT. The RF model uses k features to construct n trees and the classification result is the one most recurrent in the n trees. Although RF is a complex and time-consuming algorithm, it manages to overcome training overfitting compared with the DT algorithm (Uddin et al., 2019). The RandomForestClassifier algorithm was used with all default parameters but the criterion entropy and the number of estimators of 5.

The Artificial Neural Network (ANN) is an ML algorithm inspired by the concepts of the human brain function and structure and applied mathematically in prediction analyzes. One of the Artificial Neural Network's most important parameters is the layers' weights, which are updated when the model is trained. It uses the gradient descent method to train the network and minimize the cost function, and backpropagation to update the weights. Overfitting is a recurrent problem in ANN if appropriate precautions are not observed, such as regularization, increased training data, and early stopping. TensorFlow library 2.5.0 with Keras API was used to run the ANN. The ANN structure used was dense (fully-connected), sequential, with two hidden layers, Rectified linear unit (ReLU) as the activation function, random uniform as kernel initializer, and its respective dropout layers. The output layer used had sigmoid as the activation function. For each dataset, the number of neurons used in each hidden layer was half of the number of predictors plus predicted variables. The ANN model

was compiled with Adam optimizer, binary cross-entropy as loss function, and binary accuracy as the model performance metric. The ANN training phase consisted of 200 epochs with a batch size of 60 samples.

2.3.5 Time-point evaluations sub-datasets analysis

Each sampled maize line data consisted of reflectance data, surface temperature, and the annotated disease score per evaluation (E) per year (Y). In our multimodal disease resistance supervised classification analysis, we used the binary annotated visual score data as the predicted variable and the five UAV-extracted within-plot average spectral reflectance and surface temperature standardized data as predictor variables.

The three time-point evaluations data from two years combined were divided into four smaller sets for training purposes, sub-datasets 1 to 4 (Table 1), with the aim to compare the MCR resistance classification models' accuracy and F1-score. While accuracy is the correct prediction rate, the F1-score represents the harmonic mean between precision and recall. The accuracy and F1-score are mathematically defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1\text{-}score = \frac{2 * TP}{(2 * TP) + FP + FN}$$

where true negative (TN) represents the number of susceptible samples predicted correctly, true positive (TP) is the number of resistant samples predicted correctly, false negative (FN) indicates the number of resistant samples incorrectly predicted as susceptible, false positive (FP) is the number of susceptible samples incorrectly predicted as resistant.

In sub-datasets 1, 2, and 3, we used E1, E2, individually, and E1 and E2 data combined, respectively, to train the classification models to be tested with E3. In sub-dataset four, E3 data was randomly divided with a stratified split of 70/30 (training/test) using the software library Scikit-learn version 1.1.1 based on class labels (resistant and susceptible) and year and used to run the six ML models.

Table 1 – Experiment sub-datasets training and test data year (Y) and evaluation (E) source, training Q2 score used as threshold for visual score resistance/susceptible annotation, number (#) of resistant (R) phenotypes in the training set, percentage (%) of resistant (R) phenotypes in the training set, test set Q2 score used as threshold for visual score resistance/susceptible annotation, number (#) of resistant (R) phenotypes in the test set, and percentage (%) of resistant (R) phenotypes in the training set.

sub-dataset	**training set**	**training Q2 score (threshold)**	**# of R phenotypes training set**	**% of R phenotypes training set**	**test set**	**test Q2 score (threshold)**	**# of R phenotypes test set**	**% of R phenotypes test set**
1	Y1Y2-E1	3.0	1142/1800	0.63	Y1Y2-E3	5.0	964/1800	0.53
2	Y1Y2-E2	4.0	1019/1800	0.57	Y1Y2-E3	5.0	964/1800	0.53
3	Y1Y2-E1E2	3.0/4.0	2161/3600	0.60	Y1Y2-E3	5.0	964/1800	0.53
4	Y1Y2-E3	5.0	672/1260	0.53	Y1Y2-E3	5.0	292/540	0.54
5	Y1-E3	5.5	325/630	0.52	Y1-E3	5.5	139/270	0.51
6	Y2-E3	4.5	350/630	0.55	Y2-E3	4.5	150/270	0.55
7	Y1-E3	5.5	464/900	0.51	Y2-E3	4.5	500/900	0.55
8	Y2-E3	4.5	500/900	0.55	Y1-E3	5.5	464/900	0.51

Apart from identifying the best training time-point evaluation, we aimed to investigate the impact of each year on the classification accuracy. From the sub-datasets 1 to 4, the training set that resulted in the highest accuracies across models, time-point evaluation three, was split by year and analyzed using the same year data and crossing years in the training and testing phases with sub-datasets 5 to 8 (Table 1). As with sub-dataset 4, the same year data was used to train and test the models in sub-datasets 5 and 6, where each set data was randomly divided with a stratified split of 70/30 (training/test) based on class labels (resistant and susceptible) and used to run the six ML models.

We also cross-validated the best-performing training set from the sub-datasets 1 to 4 to ensure the accuracies' consistency. In this cross-validation analysis of evaluation three, the six ML algorithms were executed 30 times, each time using ten randomly split folds (k-folds), and the algorithms' average accuracy was calculated.

2.4 Results

2.4.1 Visual score data analysis

Each evaluation-year (E-Y) original score data is composed of the average of 2 replicates, each varying from 1 to 9, and its frequency distribution (score x frequency) can be observed in Figure 1A. In Table 2 is presented a general description of the scores. The visual scores mean of the three evaluations, rounded, were 4-4-5 in the first year and 2-3-4 in the second year. Scores of 4 or above were already well represented in E1-Y1, while in E1-Y2 the same score value was the highest visual score of the set. The last evaluation presented symptoms corresponding to the highest score in both years and resulted in a more unbalanced distribution of scores in Y1 compared with Y2 (Figure 1B). Besides, a more unbalanced distribution was found in E1 compared with E2 and E3 when the two-year data combined were analyzed (Figure 1C). The biggest differences in CV's (%) between years were 17.6 and 9.6 in E3 and E2, respectively, with Y2 showing superior values in both cases. Regarding the Pearson's correlation between evaluation-year data, E1-Y1 x E2-Y1, E1-Y2 x E2-Y2, and E2-Y2 x E3-Y2, resulted in the highest values, respectively, of 0.97, 0.92, and 0.95 (Figure 2).

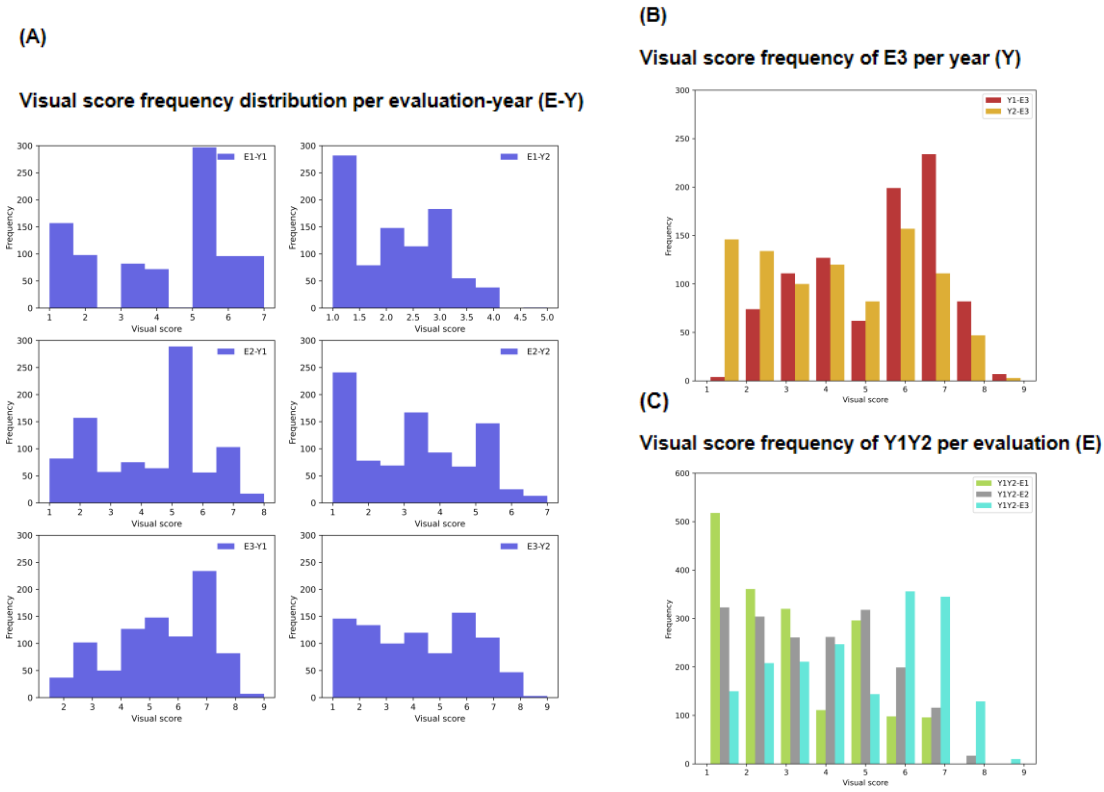

Figure 1 – Visual score frequency distribution histogram per evaluation-year (E-Y) (A); The third evaluation (E3) visual score frequency distribution in year 1 (Y1) and 2 (Y2), separately (B); and the three time-point evaluations visual score frequency distribution in years 1 and 2 combined (Y1Y2) (C).

Table 2 – Evaluation-year (E-Y) visual score data descriptive statistics.

Evaluation-Year	mean	min	max	CV(%)
E1-Y1	4.03	1	7	48.48
E1-Y2	2.09	1	5	44.88
E2-Y1	4.26	1	8	41.30
E2-Y2	3.13	1	7	50.90
E3-Y1	5.35	1.5	9	31.32
E3-Y2	4.19	1	9	48.92

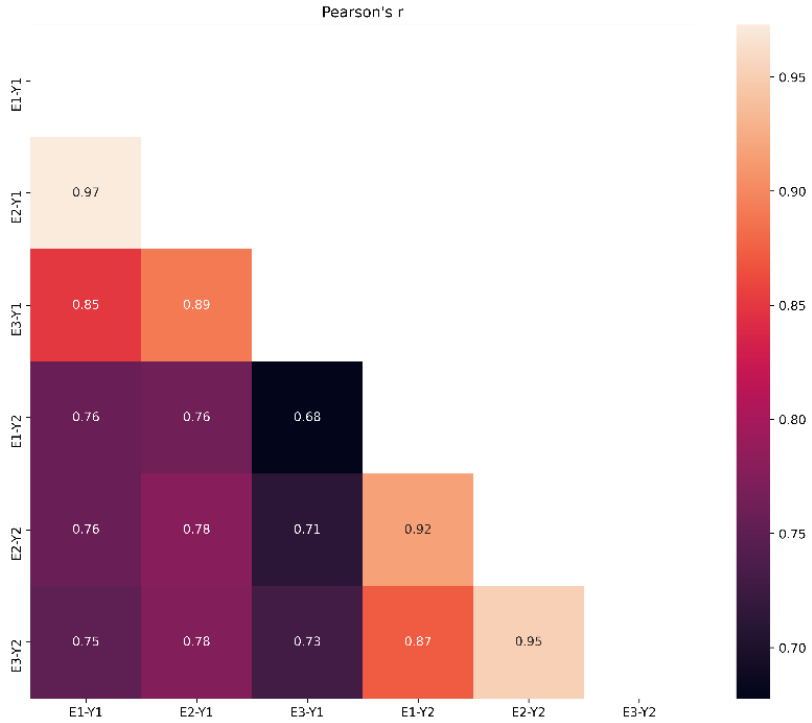

Figure 2 – Pearson's r for evaluation-year (E-Y) visual score data.

2.4.2 Visual score, spectral reflectance, and surface temperature data analysis

The average original visual score, spectral reflectance, and temperature surface data were retrieved from each genotype in both years for the three time-point evaluations. Correlation coefficients from the visual scores and the UAV-derived data are displayed in Figure 3. High correlation values were obtained between visual scores of different evaluations, varying from 0.79 to 0.92, and the correlations between the time-point visual score evaluations and the UAV-derived data were not superior to them. The highest absolute correlations found with the visual scores of evaluations 3 were of NIR and Red wavelengths third evaluation data. The correlation between NIR's third evaluation and the Red wavelength in evaluations 1, 2, and 3 was almost constant, of -0.8, -0.81, and -0.82, respectively. The correlation between NIR and the Red-edge wavelength in evaluation 3 had as result 0.97. Temperature surfaces data also presented high correlations between evaluations. Temperature surface correlations between evaluations 1 and 3 had a positive value of 0.84, while for evaluations 2 and 3 of 0.97, and for evaluations 1 and 2 of -0.91.

Simply ranking the one hundred maize lines with the highest values of NIR and Red-edge wavelengths using the two-year third evaluation data showed that only three and eleven maize lines, respectively, were annotated as susceptible to MCR. When the one hundred

maize lines with the lowest values of the Red wavelength were ranked using the same set, all of these maize lines were annotated as resistant to MCR.

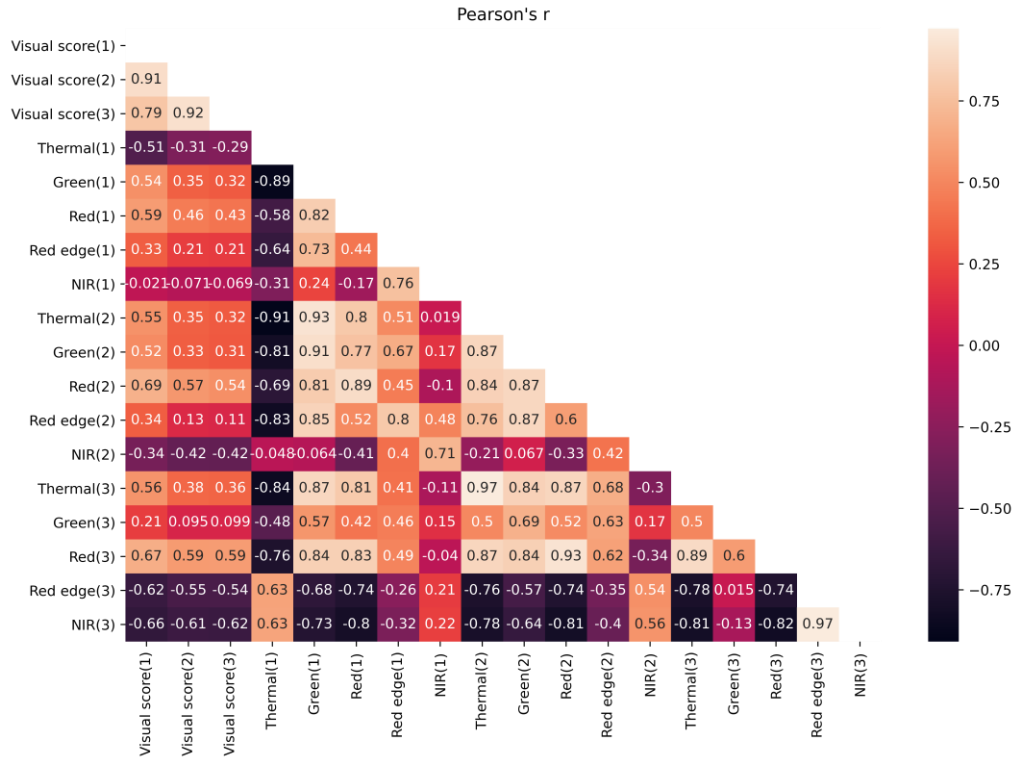

Figure 3 – Pearson's r for (1), (2), and (3) time-point evaluations visual scores, and UAV-derived temperature surface (Thermal) and spectral reflectance (Green, Red, Red-edge, and NIR).

2.4.3 Datasets analyzes results

The four different sets selected were used to build six ML models. The highest accuracies across models in sub-datasets 1, 2, and 3 presented similar values (Table 3). The algorithm SVM resulted in a good performance when sub-datasets 1, 2, and 4 were analyzed, while LR was prominent when sub-datasets 1 and 3 were analyzed. Sub-dataset 4 accuracies across models remarkably surpassed the others, resulting in accuracy values of 81 per cent across the algorithms kNN, SVM, and ANN. Considering the F1-scores, generally, sub-dataset 1 presented higher values than 2 and 3 but smaller values than 4 (Table 4).

Table 3 – Machine Learning Algorithms' accuracies of sub-datasets 1 to 4.

sub-dataset	**sub-dataset (training/test)**	**kNN**	**LR**	**SVM**	**DT**	**RF**	**ANN**

1	Y1Y2-E1/Y1Y2-E3	0.66	0.75	0.76	0.67	0.66	0.53
2	Y1Y2-E2/Y1Y2-E3	0.69	0.71	0.76	0.67	0.71	0.7
3	Y1Y2-E1E2/Y1Y2-E3	0.68	0.72	0.68	0.67	0.67	0.7
4	Y1Y2-E3/Y1Y2-E3	0.81	0.79	0.81	0.76	0.77	0.81

Table 4 – Machine Learning Algorithms' F1-scores of sub-datasets 1 to 4.

sub-dataset	**sub-dataset (training/test)**	**kNN**	**LR**	**SVM**	**DT**	**RF**	**ANN**
1	Y1Y2-E1/Y1Y2-E3	0.73	0.79	0.78	0.75	0.72	0.7
2	Y1Y2-E2/Y1Y2-E3	0.68	0.71	0.75	0.62	0.69	0.69
3	Y1Y2-E1E2/Y1Y2-E3	0.69	0.76	0.76	0.64	0.67	0.63
4	Y1Y2-E3/Y1Y2-E3	0.82	0.79	0.81	0.77	0.78	0.82

To better comprehend sub-dataset 4, we separated it by year, training and testing with each year individually to investigate the influence of the year on our results. To do so, we organized sub-datasets 5 to 8, applied the same ML models, and retrieved the accuracy and F1-scores (Tables 5 and 6). Sub-dataset 6 presented higher accuracy results than 5, with the ANN algorithm showing the highest accuracy of 84 per cent. Following the same path, F1-scores presented higher values in sub-dataset 6. Sub-dataset 8 even resulted in zero or negligible F1-scores for all algorithms but SVM and ANN.

Table 5 – Machine Learning Algorithms' accuracies of sub-datasets 5 to 8.

sub-dataset	**sub-dataset (training/test)**	**kNN**	**LR**	**SVM**	**DT**	**RF**	**ANN**
5	Y1-E3/Y1-E3	0.77	0.77	0.77	0.67	0.72	0.78
6	Y2-E3/Y2-E3	0.83	0.83	0.83	0.8	0.83	0.84
7	Y1-E3/Y2-E3	0.59	0.61	0.62	0.64	0.64	0.75
8	Y2-E3/Y1-E3	0.49	0.49	0.51	0.49	0.49	0.63

Table 6 – Machine Learning Algorithms' F1-scores of sub-datasets 5 to 8.

sub-dataset	**sub-dataset (training/test)**	**kNN**	**LR**	**SVM**	**DT**	**RF**	**ANN**
5	Y1-E3/Y1-E3	0.79	0.79	0.76	0.69	0.73	0.79
6	Y2-E3/Y2-E3	0.86	0.86	0.86	0.83	0.86	0.86
7	Y1-E3/Y2-E3	0.73	0.74	0.75	0.75	0.75	0.81
8	Y2-E3/Y1-E3	0	0	0.67	0.02	0	0.48

The cross-validation results of the third evaluation in both years resulted in mean accuracies equal or higher than 80 per cent when the ML models kNN, LR, and SVM were adopted (Table 7). Algorithms LR and SVM also presented the lowest CVs among all of

them. The ANN was the model with the highest CV, and the accuracy value discrepancy between the cross-validation and the previous test (sub-dataset 4) was considerable.

Table 7 – The ML algorithms' previous accuracy results for sub-dataset 4, the cross-validation results for the same set, and the cross-validation coefficient of variation (%).

ML algorithm	**sub-dataset 4 accuracy score**	**E3 cross-validation accuracy mean**	**E3 cross-validation CV (%)**
kNN	0.81	0.81	0.340902
LR	0.79	0.80	0.211206
SVM	0.81	0.82	0.202229
DT	0.76	0.76	1.158948
RF	0.77	0.79	0.927386
ANN	0.81	0.70	8.985242

2.5 Discussion

2.5.1 Visual scores

The within-plot average visual score data for each evaluation-year were descriptively analyzed. As we can observe in Figure 1 and in the relative standard deviations, the distributions differed in each evaluation-year set. In E3-Y2, we detected a more homogenous distribution through the 1-9 range of scores, which is of great importance for the ML prediction analyzes. Homogeneously distributed data are easier to make accurate predictions due to their higher trainability. In the study by Lo et al. (2021), datasets with higher heterogeneity had lower trainability and influenced the models' performance when used as test sets. That might be one of the reasons for the overall algorithms' higher accuracy values and F1-scores of sub-datasets 6 over 5, and 7 over 8 shown in Tables 5 and 6. Although the predicted variable data used to build the ML models were binary annotated, having a wide range of visual scores leading to it is a manner to capture and incorporate data variability to the analyzes, which is unprecedent to achieve trustful classification performances.

In our MCR visual score analysis, we achieved high correlation results between years and evaluations, as shown in Figures 2 and 3. In Figure 2, we observed that the correlation values between the two years for the same time-point evaluation were of 0.73, 0.78, and 0.76 for evaluations 1, 2, and 3, respectively. Therefore, using both years data was a suitable approach for this analysis. Also, the more balanced visual score distribution with the two-year data shown in Figure 1C is evidence of the advantage of using data from multiple years to perform MCR resistance classification in this scenario. We also obtained high correlation

values between different time-point evaluations using both years data, as shown in Figure 3. The correlations between evaluations 1 and 2, 1 and 3, and 2 and 3 were of 0.91, 0.79, and 0.92, respectively. Hence, using a time-point evaluation as training set and other as test set turned out to be a plausible approach.

2.5.2 Spectral reflectance

High correlations between temperature surface data across evaluations were expected based on the nature of the data. Overall, the third evaluation displayed highest correlation scores (positive or negative) between Thermal, Red, Red-edge, and NIR wavelengths. A reason for it might be that in the third evaluation, the plants are mature enough to express these wavelengths' reflectance correlation patterns more intensely. The highest correlation values with the visual scores of time-point evaluation three were obtained using NIR and Red correspondent evaluation, which highlights the importance of these wavelengths in our current analysis. An interesting result was the practically constant correlation values between the NIR third evaluation and the three time-point Red wavelength evaluations. That result could be further studied in the future to optimize the use of this wavelength in earlier time-point evaluations.

The expected plant spectral signature pattern tells that in healthy vegetation, Red and Red-edge wavelengths have smaller reflectance than stressed plants, while the opposite is true with the NIR wavelength (Ibarrola-Ulzurrun et al., 2019). In the analysis of the third evaluation, NIR and Red had a negative correlation (-0.82). The NIR and Red-edge wavelengths had a positive correlation (0.97), the opposite of the spectral pattern expected according with these authors. Nonetheless, as shown in the one hundred maize lines ranking, higher values of NIR and low values of Red wavelength were mainly represented by the annotated resistant phenotype, which is consistent with previous studies. Zahir et al. (2022), for example, state that low NIR reflectance is an indicator of cellular structural damage, with its pattern depending on the disease's severity and stage of development.

2.5.3 Datasets

The sub-dataset 4 was superior to 1, 2, and 3 at delivering higher accuracies and F1-scores across algorithms. These higher metric prediction performances in sub-dataset 4 probably happened because we did not cross different time-point evaluations when training

and testing the ML models. In addition, a wider range of visual scores was identified in the last evaluation, which did not occur with earlier training time evaluations. That turned out to be extremely important to perform MCR resistance prediction in our target classification scenario. Similar results were encountered by Lo et al. (2021) for human Type 1 Diabetic Retinopathy, where the authors concluded that training data should be similar to the target data to guarantee robust prediction. Another analogous result was found with oat crown rust severity prediction based on 27 time-point evaluations using visual scores and multispectral aerial photos. The authors discovered that the accuracy results were higher shortly before the plants began to senesce and that the accuracies were usually low outside this timeframe (McNish & Smith, 2021).

Sub-datasets 1, 2, and 3 presented similar accuracies results. Even with the addition of two time-point evaluations on set 3, its models were not capable of making predictions more accurately. That may have occurred because sub-datasets 1 and 2 models were not good at capturing important data patterns for the disease classification, or, even more likely, the earlier training evaluations were not enough to adequately study the disease with the approach used in this work. In the first and second evaluations, for example, the disease did not develop a large range of visual scores, nor did its symptoms develop to the highest potential compared with the last evaluation. Our results indicate that the third evaluation training set is the most suitable to be used in prediction analyzes in this scenario.

The use of earlier time-point evaluations data as training set, however, can potentially be suitable in other scenarios. In our classification results, mainly using the SVM model, an accuracy value of 0.76 was achieved using evaluations 1 and 2 as training set. A value not so distant from the 0.81 resulted from the use of evaluation 3. In future studies, alternative strategies might be applicable to perhaps elevate its potential use in earlier time classification analysis, such as increase the number of maize lines studied, re-design the visual score evaluation scale for use in earlier timings, and adjust the models's parameter using only the earlier evaluations.

Aiming to study in detail sub-dataset 4, the set in which the models were most successful in predicting MCR resistance, sub-datasets 5 to 8 were created and analyzed. With sub-datasets 5 and 6, we wanted to investigate the effect of years 1 and 2 separately in the last evaluation. Sub-dataset 6 resulted in higher accuracies and F1-scores than 5 and 4, likely due to the more homogenous distribution of scores in year 2 compared with year 1. Because visual scores, mainly in year 1, had instances much more underrepresented than others, samples associated with these scores are, comparatively, more challenging to make predictions. With

sub-datasets 7 and 8, we aimed to investigate the capacity of a model trained with a year's data to predict resistance with another year effectively and have an indication of whether model updates through years should be mandatory. In sub-dataset 8, trained with year 2 and tested with year 1, all six models' accuracies were smaller than those of sub-dataset 7. This result probably happened because of the more heterogenous distribution of the test set. Besides, the models did poorly at correctly identifying the resistant phenotype, which led to the zero or near zero values of F1-scores in sub-dataset 8. Consequently, it would be wise to gather data from more years, build and update a more robust model to be used in a multi-year MCR classification task.

The cross-validation technique was used to compare the ML algorithms employed in our models using the two-year third time-point evaluation data. The LR and SVM classifiers' performances surpassed the others with their high accuracy values and by exhibiting the lowest CVs, making them more trustworthy with this dataset. The algorithms compared in our study are known to perform well in prediction studies. For example, the SVM classifier delivered the most consistent performances of prediction error and hit rate and was one of the best ML methods identified for data modelling and prediction in a robust time-series study (Parmezan et al., 2019). In the study conducted by Neelakantan (2021), the SVM algorithm delivered 83 percent average accuracy in predicting health plants and four diseases accessed with a phone camera. This was the second highest value when compared with RF, KNN, DT, and Naïve Bayes (NB) classifiers, that resulted in accuracies of 89, 82, 77, and 81. In a comparative analysis of human disease prediction studies employing ANN, DT, kNN, LR, RF, SVM, and NB algorithms, RF and SVM were superior in, respectively, 53% and 41% of the studies selected by the authors (Uddin et al., 2019). Ibrahim & Abdulazeez (2021) compared nineteen papers on disease prediction and found that SVM, kNN, RF, and DT had good accuracies. They also discovered that accuracies might oscillate among datasets depending on the size of the dataset, the number of features, and feature selection.

The ANN algorithm presented the worst cross-validation result compared with the analysis of dataset 4 and resulted in the highest CV among all the algorithms. In our previous analysis, the random training/test split was made just once, which strongly influenced its accuracy value. In each of the 30 cross-validation executions, we had different training and test samples, minimizing the error by withdrawing its average accuracy. The ANN is a method that depends on a considerable amount of data because of the backpropagation step, and more samples and evaluation-years are likely necessary to reduce the CV. It is known that the scarcity of publicly available datasets is a major bottleneck in disease prediction analysis,

which limits models' performance (Thakur et al., 2022). In small sets, as is our case, traditional ML algorithms might perform as well as or even outperform an ANN, as concluded by Montesinos-López et al. (2021) and Azodi et al. (2019). Consequently, the use of LR and SVM algorithms should be highlighted for MCR resistance classification studies.

2.6 Conclusion

Our study investigated the use of time series UAV image-based data in MCR resistance prediction using supervised ML classification algorithms. The combined use of multispectral and thermal sensors in the identification of MCR-resistant maize lines offered the possibility to obtain a more trustworthy data source and maximize evaluating efforts. The ML analyses results indicated that effectively setting the training time-point evaluation is of great importance to best capture data trends into the models. No other training set delivered prediction performance approximately as good as the last time-point evaluation, which was used as training and test set. In addition, the accuracy discrepancies between the two years accessed might be due to their distinct visual score distribution. The influence of these distribution patterns on the classification models' trainability was highlighted, which indicates the importance of a multi-year analysis approach. The use of a method to select the training set samples to contemplate equally the entire range of visual scores should be considered in future studies. Furthermore, the ML algorithms LR and SVM delivered outstanding prediction performances and consistent results in the cross-validation analysis, suggesting that their preferential employment may be advantageous.

2.7 Acknowledgments

The study was partially supported by the CGIAR Research Program on Maize (MAIZE). This study was also supported by CNPq (National Council for Scientific and Technological Development) and UFV (Universidade Federal de Viçosa). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. The authors thank CIMMYT's El Batan and Agua Fria experimental station professionals involved in the regional trials. The contents and opinions expressed herein are those of the authors and do not necessarily reflect the views of the associated and supporting institutions.

2.8 References

Azodi, C. B., Bolger, E., McCarren, A., Roantree, M., de los Campos, G., & Shiu, S.-H. (2019). Benchmarking Parametric and Machine Learning Models for Genomic Prediction of Complex Traits. *G3 Genes|Genomes|Genetics*, *9*(11), 3691–3702. https://doi.org/10.1534/g3.119.400498

Cheng, M., Penuelas, J., McCabe, M. F., Atzberger, C., Jiao, X., Wu, W., & Jin, X. (2022). Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China. *Agricultural and Forest Meteorology*, *323*, 109057. https://doi.org/https://doi.org/10.1016/j.agrformet.2022.109057

Ibarrola-Ulzurrun, E., Marcello, J., Gonzalo-Martín, C., & Martín-Esquivel, J. L. (2019). Temporal dynamic analysis of a mountain ecosystem based on multi-source and multi-scale remote sensing data. *Ecosphere*, *10*(6), 1–17. https://doi.org/https://doi.org/10.1002/ecs2.2708

Ibrahim, I. M., & Abdulazeez, A. M. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, *2*(01), 10–19. https://doi.org/10.38094/jastt20179

Iniyan, S., Jebakumar, R., Mangalraj, P., Mohit, M., & Nanda, A. (2020). Plant Disease Identification and Detection Using Support Vector Machines and Artificial Neural Networks. In S. S. Dash, C. Lakshmi, S. Das, & B. K. Panigrahi (Eds.), *Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing, vol 1056* (pp. 15–27). Springer Singapore. https://doi.org/10.1007/978-981-15-0199-9_2

Lo, J.-E., Kang, E. Y.-C., Chen, Y.-N., Hsieh, Y.-T., Wang, N.-K., Chen, T.-C., Chen, K.-J., Wu, W.-C., Hwang, Y.-S., Lo, F.-S., & Lai, C.-C. (2021). Data Homogeneity Effect in Deep Learning-Based Prediction of Type 1 Diabetic Retinopathy. *Journal of Diabetes Research*, *2021*, 2751695. https://doi.org/10.1155/2021/2751695

Loladze, A., Rodrigues, F. A., Toledo, F., San Vicente, F., Gérard, B., & Boddupalli, M. P. (2019). Application of Remote Sensing for Phenotyping Tar Spot Complex Resistance in Maize. In *Frontiers in Plant Science* (Vol. 10, p. 552). https://www.frontiersin.org/article/10.3389/fpls.2019.00552

Loladze, A., Rodrigues Jr, F. A., Petroli, C. D., Munoz, C., Naranjo, S. M., Vicente, F. S., Gerard, B., Montesinos-Lopez, O. A., Crossa, J., Martini, J. W. R. (2023). Use of Remote Sensing for Genome-Wide Association Studies and Genomic Prediction [Unpublished manuscript]. International Maize and Wheat Improvement Center (CIMMYT El Batan, Mexico).

McNish, I. G., & Smith, K. P. (2021). Oat Crown Rust Disease Severity Estimated at Many Time Points Using Multispectral Aerial Photos. *Phytopathology®*, *112*(3), 682–690. https://doi.org/10.1094/PHYTO-09-20-0442-R

Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W. R., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., &

Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Genomics*, *22*(1), 19. https://doi.org/10.1186/s12864-020-07319-x

Mutka, A. M., Fentress, S. J., Sher, J. W., Berry, J. C., Pretz, C., Nusinow, D. A., & Bart, R. (2016). Quantitative, image-based phenotyping methods provide insight into spatial and temporal dimensions of plant disease. *Plant Physiology*, *172*(2), 650–660. https://doi.org/10.1104/pp.16.00984

Neelakantan, P. (2021). Analyzing the best machine learning algorithm for plant disease classification. *Materials Today: Proceedings*. https://doi.org/https://doi.org/10.1016/j.matpr.2021.07.358

Panigrahi, K. P., Das, H., Sahoo, A. K., & Moharana, S. C. (2020). Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. In H. Das, P. K. Pattnaik, S. S. Rautaray, & K.-C. Li (Eds.), *Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing, vol. 1119* (pp. 659–669). Springer. https://doi.org/10.1007/978-981-15-2414-1

Parmezan, A. R. S., Souza, V. M. A., & Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences*, *484*, 302–337. https://doi.org/10.1016/j.ins.2019.01.076

Ramirez-Cabral, N. Y. Z., Kumar, L., & Shabani, F. (2017). Global risk levels for corn rusts (Puccinia sorghi and Puccinia polysora) under climate change projections. *Journal of Phytopathology*, *165*(9), 563–574. https://doi.org/10.1111/jph.12593

Ritchie, S. W., Hanway, J. J., & Benson, G. O. (1986). *How a Corn Plant Develops. Spec. Rep. 48*.

Thakur, P. S., Khanna, P., Sheorey, T., & Ojha, A. (2022). Trends in vision-based machine learning techniques for plant disease identification: A systematic review. *Expert Systems with Applications*, *208*, 118117. https://doi.org/https://doi.org/10.1016/j.eswa.2022.118117

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, *19*(1), 281. https://doi.org/10.1186/s12911-019-1004-8

Zahir, S. A. D. M., Omar, A. F., Jamlos, M. F., Azmi, M. A. M., & Muncan, J. (2022). A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Sensors and Actuators A: Physical*, *338*, 113468. https://doi.org/https://doi.org/10.1016/j.sna.2022.113468

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., Hornero, A., Albà, A. H., Das, B., Craufurd, P., Olsen, M., Prasanna, B. M., & Cairns, J. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods*, *11*(1), 1–10. https://doi.org/10.1186/s13007-015-0078-2

3. ARTICLE 2: EXPLORING THE USE OF SVM AND ANN FOR MAIZE COMMON RUST RESISTANCE CLASSIFICATION WITH GENETIC MARKERS AND UAV-COUPLED SENSORS DATA

3.1 Abstract

Recent advances in high-throughput phenotyping (HTP) and high-throughput genotyping (HTG) technologies enabled the generation of a considerable amount of data. These high-throughput data can be analyzed using Machine learning (ML) approaches in order to improve prediction performance. The identification of Maize common rust (MCR) resistant lines in large-scale field trials is a challenge that could be solved by combining the use of these technologies. The HTP data was obtained from images collected with multispectral and thermal sensors in three time-point evaluations, while the HTG data was extracted from 912 SNP markers selected after quality control filtering. To explore MCR resistance classification performance of two classification algorithms, Support Vector Machine (SVM) and Artificial Neural Network (ANN), a k-fold cross-validation analysis was performed with nine HTP/HTG/HTP+HTG datasets. We also compared learning curves and ranked the five most important features in each dataset using the SVM algorithm. Our results indicated that the datasets performed differently at classifying MCR resistance depending on the algorithm used. The ANN outperformed SVM when marker data was employed. The greatest algorithm x dataset performance, on the other hand, was achieved when the SVM algorithm was used with the third and the three time-point HTP evaluations combined. The SVM's learning curves indicated that more training samples are required to reduce the models' variability. In general, the feature importance rank indicated that the Red wavelength is the most valuable for successfully classifying maize lines into resistant or susceptible to MCR. In the rank of importance of markers, it was highlighted five SNPs aligned with the transcription factors, and protein-coding genes mostly related with stress tolerance according with recent studies. The disparities in computational power and amount of training data to effectively compare the HTP and HTG sets call attention to the importance of carefully examining whether using a more complex model, such as ANN, is advantageous in a particular scenario.

Keywords: Maize common rust, high-throughput phenotyping, Machine learning, Support Vector Machine, Artificial Neural Network

3.2 Introduction

Maize common rust (MCR) is a disease caused by *Puccinia sorghi* Schwein, which produces brown to reddish-brown pustules on the leaves (Ramirez-Cabral et al., 2017). Crop rotation, eradication of alternative hosts, chemical control, and the use of resistant cultivars are commonly used protective measures against this pathogen. Chemical control is frequently used at the beginning of the crop cycle when the disease has the greatest potential to negatively affect productivity. However, due to its high cost, limited effectiveness, and environmental effects, the most recommended protective method is the use of resistant cultivars.

High-throughput phenotyping (HTP) data are widely employed in breeding programs to identify and predict the most outstanding genotypes. Spectral vegetation reflectance measured in various spectral regions, such as near-infrared (NIR) and thermal infrared radiation, can be utilized to explain phenotypic traits (Omari et al., 2020). For example, plant reflectance in the thermal infrared region can be used to estimate canopy temperature (C. Zhao et al., 2019). Different spectral regions were used in the development of vegetation indices applied to traits such as biomass (Quirós Vargas et al., 2019), tolerance or resistance to biotic or abiotic stresses (Kim et al., 2020), growth and development (Tirado et al., 2020), and productivity (Bombrun et al., 2020).

Maize origin, domestication, and genetic variability have all been extensively investigated using high-throughput genotyping (HTG) technologies, such as with the sequencing of the B73 inbred line reference genome (Jiao et al., 2017). The B73 reference genome was used in MCR resistance studies, revealing more than twenty major effect genes and almost 40 quantitative trait loci (QTLs) (Olukolu et al., 2016). Other 5 QTLs and 9 candidate genes were identified in a subsequent study (Zheng et al., 2018).

The Machine learning (ML) classification algorithms Support Vector Machine (SVM) and Artificial Neural Network (ANN) are frequently used in image-based plant disease prediction analyzes. To name a few applications, the SVM algorithm was used to predict thrips incidence in banana (Manrique-Silupu et al., 2021), in the early detection of soybean mosaic virus (Gui et al., 2021), and in wheat powdery mildew severity estimation (J. Zhao et al., 2020). The selection of the kernel function is an essential step in SVM analysis because this algorithm optimizes classification performance through the surface separation by a hyperplane using the kernel function as a parameter. On the other hand, to successfully train a supervised ANN classifier, it’s important to pay attention to the number of nodes and hidden

layers, and in its ability to iteratively adjust the weights in the backpropagation step. For instance, the ANN algorithm was used to estimate the severity of three wheat diseases through UAV-coupled multispectral sensor data (Bebronne et al., 2020). The ANN and SVM algorithms have been applied, among others, to identify and quantify field trials of yellow sigatoka in banana using RGB sensor data (Calou et al., 2020).

ML analyzes are applied in genotyping data analyzes for a variety of reasons. UAV images, for example, were used in a GWAS study conducted by Chen et al. (2019), in which the authors identified seven QTLs for winter survival in wheat. In a post-GWAS analysis, Nicholls et al. (2020) investigated the capacity of ML models to predict the most important loci to be prioritized in functional analysis. Another study used HTP and HTG data for SNP allele prediction using deep convolutional neural networks in pre-training and SVM or ANN models for the classification (Z. Zhang et al., 2022).

In k-fold cross-validation analyzes, samples are randomly split into k parts, one of which is used to iteratively validate the testing model and predict the samples' class, and each of the remaining k-1 parts are used to train the model (Rohani et al., 2018). Iterations are performed so that each k fold can be used to validate the testing model. After that, the average prediction performance is calculated to generate a single estimation value. The k-fold cross-validation approach is used to evaluate prediction performance and optimize the model's parameters (Dong et al., 2021).

The use of data retrieved from HTP and HTG technologies in supervised disease resistance prediction appears to have the potential to achieve high classification performance using ML approaches. However, it is necessary to elucidate the magnitude and stability of the prediction performance for a given input data, which can be done with the cross-validation analysis using the ML algorithms chosen.

This study is aimed at exploring the performance of SVM and ANN algorithms in MCR resistance or susceptibility cross-validation classification using nine datasets derived from three time-point evaluations (HTP) and 912 SNP markers (HTG). In this work, we considered as cross-validation optimization targets the model's cross-validation prediction performance output and its relative standard deviation. Furthermore, we also aimed at analyzing the learning curves and ranking the five most important features for each dataset using the SVM algorithm. In comparing the three time-point disease evaluations, we focused on the maximization of temporal information assessment, storage, and processing time.

3.3 Materials and methods

3.3.1 Visual score, spectral reflectance, and genotyping data acquisition

A CIMMYT (International Maize and Wheat Improvement Center) trial of 899 maize lines with two replicates was screened for MCR resistance using traditional and HTP techniques. Visual scores and HTP data were obtained with a disease score scale ranging from 1 (very resistant) to 9 (very susceptible) and multispectral and thermal UAV-coupled sensors, respectively. The spectral reflectance and surface temperature data were accessed in the bandwidths of Green (550nm ± 40nm), Red (660nm ± 40nm), Red-edge (735nm ± 10nm), NIR (790nm ± 40nm), and Thermal (7.5–13.5μm). The lines were evaluated in two seasons, 2019 and 2020, at three stages of maize development, R1 (silking), R2 (blister), and R3 (milk) (Ritchie et al., 1986). The within-plot average visual score, spectral reflectance, and temperature surface of each maize line per year were retrieved and used in the following analyzes. The maize lines' SNPs allele calls were made using HTG DArT-seq data, which focuses on protein-coding regions. After removing markers with more than 15% missing values and minor allele frequency smaller than 0.05, a total of 912 SNP markers were selected. Imputation of the remaining missing data was made using the marker's mode. More details on the experimental design, plant inoculation, disease score evaluation, image acquisition and processing, and the genotypic data can be found on Loladze et al. (2023).

3.3.2 ML models' features

The ML models were developed to categorize maize lines between resistant and susceptible to MCR using HTP and HTG-derived data, which were analyzed with the ML algorithms SVM and ANN aiming to reduce the uncertainties that arise from visually accessing a large number of plots for MCR resistance, as well as the challenges of quickly processing it while delivering an adequate classification performance.

With the aim to select 50% of the maize lines, the third visual score evaluation was annotated using its second quartile as the threshold to distinguish resistant from susceptible genotypes and used as predicted variable in all our models. Nine datasets containing five to 927 predictor variables were created (Table 1). Each marker was used as a model feature in the sets encompassing the HTG data of 912 SNP markers (datasets 5 to 9). The datasets comprising HTP data from each time-point evaluation included the corresponding five wavelengths as features (datasets 1, 2, and 3), and when combined with marker data, each set

had 917 features (datasets 6, 7, and 8). The remaining datasets were made of the three time-point HTP evaluation data (dataset 4), and the same set plus the HTG data (dataset 9) contained 15 and 927 features, respectively.

Table 1 – Datasets given number, training dataset, and total number of features per dataset. HTP (high throughput phenotyping)-E (evaluation); HTG (high throughput genotyping).

dataset	**training set**	**# of features**
1	HTP-E1	5
2	HTP-E2	5
3	HTP-E3	5
4	HTP-E1E2E3	15
5	HTG	912
6	HTP-E1+HTG	917
7	HTP-E2+HTG	917
8	HTP-E3+HTG	917
9	HTP-E1E2E3+HTG	927

3.3.3 Models' deployment

The datasets were cross-validated using SVM and ANN ML models with a binary supervised classification approach. The software libraries Pandas 1.4.3, Numpy 1.19.2, Scipy 1.6.2, and Matplotlib 3.3.3 were used in data processing. Spectral reflectance and temperature surface data were normalized with the min-max scaler from Scikit-learn 1.1.1, which was also used to run the SVM classification algorithm, with C of 3.0, linear kernel, and the remaining default parameters. We utilized Yellowbrick 1.5 to visualize the SVM learning curves.

The ANNs were built with the TensorFlow 2.5.0 software library with Keras API. Their structure consisted of a sequential fully-connected network of two hidden layers and the output layer, using as the activation functions ReLU (Rectified linear unit) and sigmoid, respectively. A random uniform method was employed for the kernel initializer. Each model's compilation was made using Adam optimizer, binary cross entropy to access the loss function, and binary accuracy to access model performance. In each dataset, the number of hidden layer's neurons used was half of the number of predictors plus predicted variables. The ANNs training phase had 30 epochs and a batch size of 64 samples. Parameter tuning was done based on the corresponding training set analyzed. To better capture the variability of the training datasets in this cross-validation analysis, the algorithms were executed in ten

shuffled iterations, each of them with five randomly split k-folds (10x5). The average accuracy and coefficient of variation (CV) were computed using the cross-validation results.

3.3.4 Feature importance rank and genes description

Python's Scikit-learn API Yellowbrick 1.5 was used to rank the five most important features in each of the nine datasets using the SVM algorithm. The allele sequences containing the SNP markers found within these ranked features were aligned with the maize reference genome version 5 (Jiao et al., 2017) in MaizeGDB (Maize Genetics and Genomics Database) website (https://maizegdb.org/) using blastn, a BLAST program (Altschul et al., 1990). Their respective protein-coding gene information was retrieved using MaizeGDB jbrowse (https://jbrowse.maizegdb.org/). Additional gene information was obtained from relevant literature.

3.4 Results

3.4.1 SVM and ANN cross-validation prediction performance

The accuracy results of the nine cross-validated datasets using SVM and ANN algorithms are shown in Figure 1. The SVM models were able to more accurately discriminate maize lines between resistant and susceptible when HTP datasets were used. Instead, with datasets containing HTG or HTP+HTG data, ANN models slightly outperformed SVM prediction metrics.

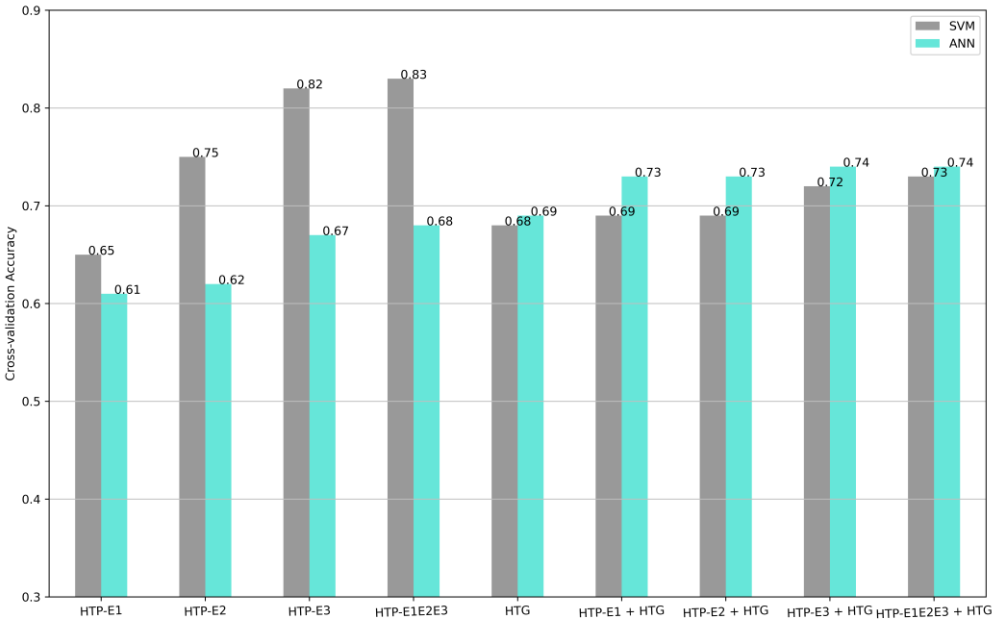

Figure 1 - Cross-validation accuracy results from the comparative analysis of Support Vector Machine (SVM) and Artificial Neural Network (ANN) models using 10x5 k-fold cross-validation of nine datasets. HTP (high throughput phenotyping)-E (evaluation); HTG (high throughput genotyping).

Although the addition of HTG data improved the prediction performances of the ANN models, the SVM algorithm used with HTP data alone still delivered higher accuracies than the ANN for evaluations two, three and the three evaluations combined. Only with time-point evaluation one the addition of the marker data was effective in increasing the ANN accuracy above both algorithms used with the HTP data. The HTP-E3 and HTP-E1E2E3 data predicted with the SVM algorithm yielded the highest accuracy values for the analyzes, 82 and 83 per cent, respectively.

The CVs' results can be visualized in Figure 2. When HTP data were analyzed using the ANN algorithm, the CV values results were remarkably high. However, when these datasets were examined using the SVM algorithm, the estimations were more precise than any other set analyzed using both algorithms. The sets containing HTP+HTG data showed similar CVs across both algorithms. The addition of marker data in the HTP datasets increased the SVM models' CVs while decreasing the CVs of ANN models to a great extent. The HTG set

resulted in similar accuracies and standard deviations compared to HTP+HTG sets using SVM and ANN algorithms.

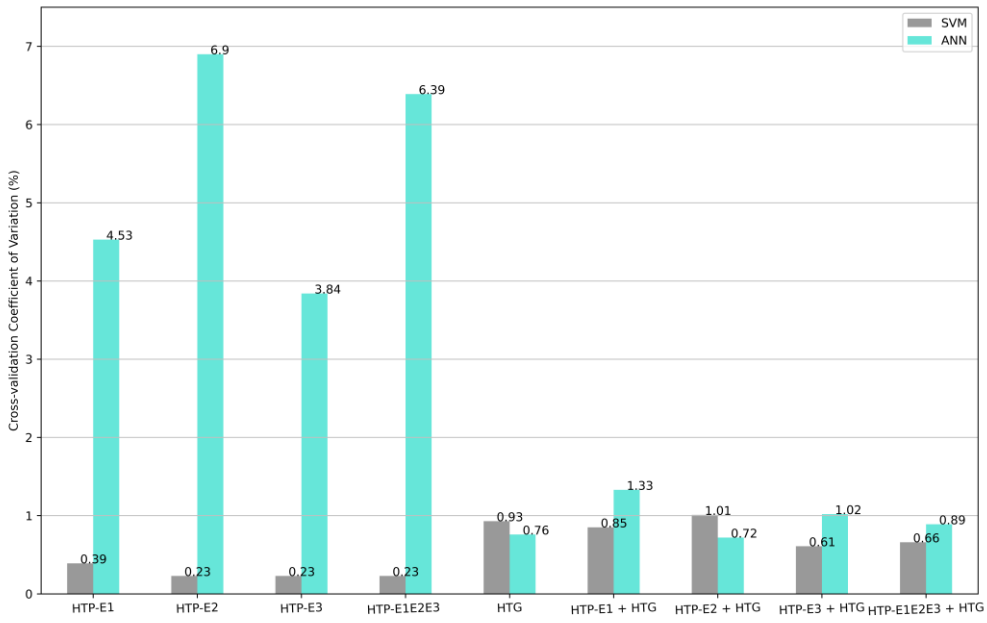

Figure 2 - Cross-validation Coefficient of Variation (CV) per cent results from the comparative analysis of Support Vector Machine (SVM) and Artificial Neural Network (ANN) models using 10x5 k-fold cross-validation of nine datasets. HTP (high throughput phenotyping)-E (evaluation); HTG (high throughput genotyping).

3.4.2 SVM learning curves

The SVM model validation curves assisted us in the interpretation of the cross-validation F1 mean score compared with the F1-score from a predictive model developed with the dataset training/test split for the same number of sampled instances. In models built with HTP data only, the scores of both training and cross-validation curves converged as more data was included in the prediction analysis (Figure 3). The HTP-E1 data required more training samples for the models to converge than the HTP data of other time-point evaluations. The HTP-E2 SVM and ANN models converged after 700 training instances, but F1-scores did not stabilize until 1300. Although the final scores of HTP-E3 and HTP-E1E2E3

models were similar, only the latter achieved a score above 82 per cent using up to 900 training instances on both curves.

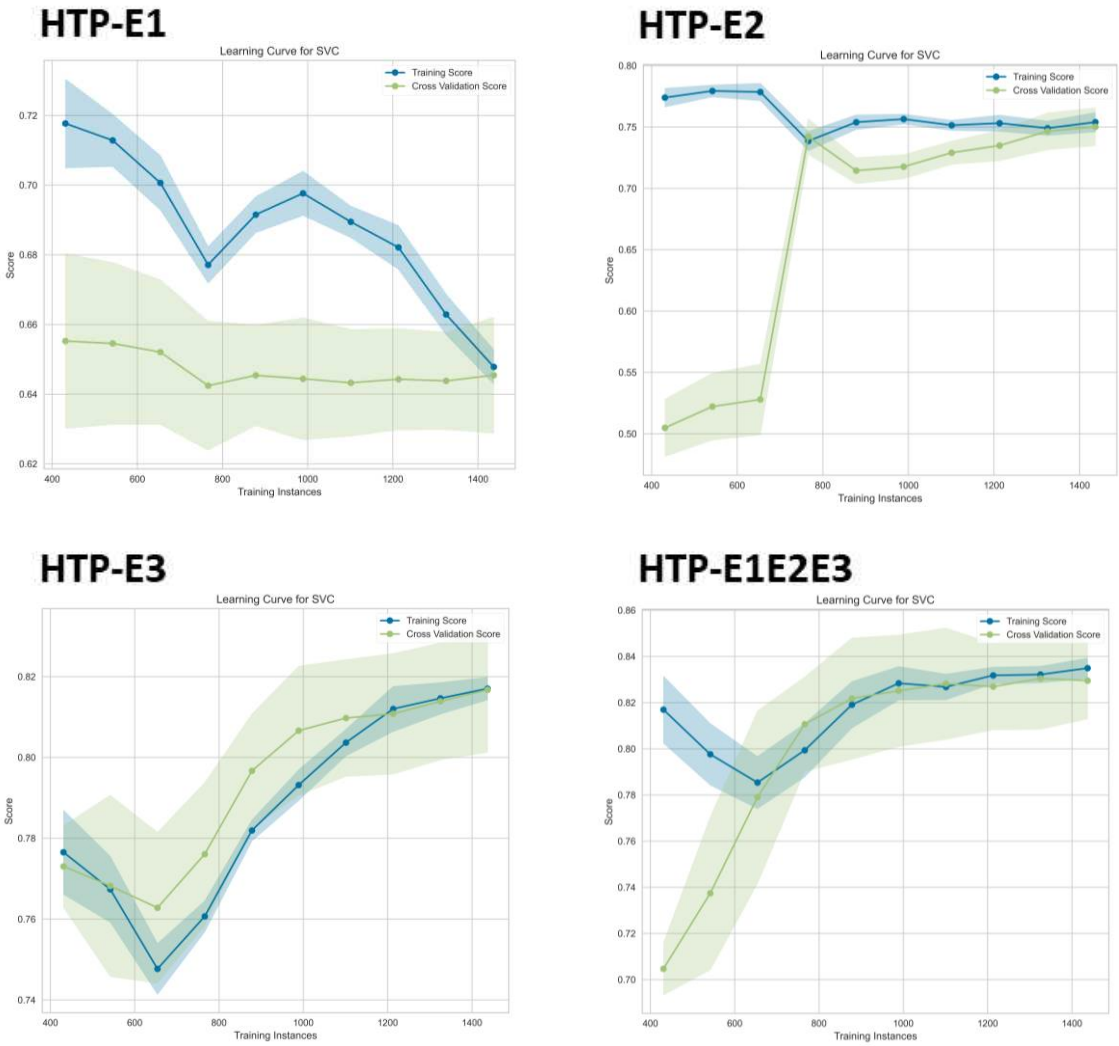

Figure 3 –HTP datasets learning curves showing the relationship between train/test and cross-validation SVM models as the number of training instances increases. HTP (high throughput phenotyping)-E (evaluation); HTG (high throughput genotyping).

The learning curves of HTG and HTP+HTG datasets revealed a similar pattern (Figure 4). The train/test models delivered high scores, and the across-validation scores improved as the quantity of training samples increased. Their respective training and cross-validation scores did not converge.

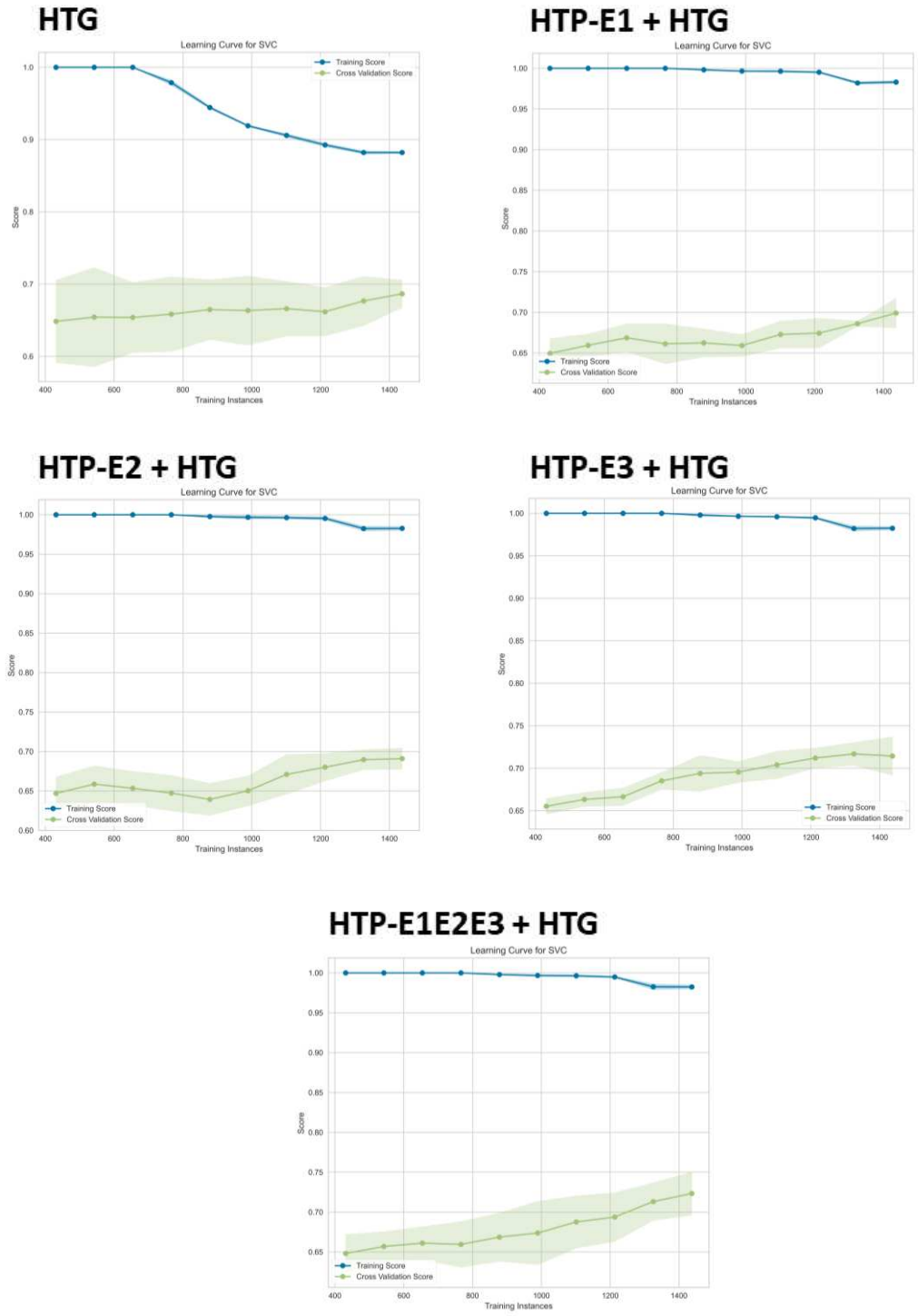

Figure 4 – HTG and HTP+HTG datasets learning curves showing the relationship between train/test and cross-validation SVM models as the number of training instances increases. HTP (high throughput phenotyping)-E (evaluation); HTG (high throughput genotyping).

3.4.3 SVM feature importance

The five most important features in each dataset are shown in Table 2. In six out of eight models trained with HTP or HTP+HTG sets, the Red wavelength was the most relevant attribute. Despite the fact that datasets 4 and 9 comprised HTP data from the three time-point evaluations, the red wavelength from evaluation three was ranked in the highest position. The importance of Green and Red-edge wavelengths in the HTP datasets decreased from earlier evaluations to the third evaluation. In the opposite direction, NIR and Thermal wavelengths ranked higher in the third evaluation of HTP sets. In three out of four HTP+HTG datasets, the NIR wavelength was ranked as the second most important feature. The markers 100401010|F|0NA10:C>GNA10:C>G and 100625723|F|0NA8:G>ANA8:G>A were found in three and two out of the four HTP+HTG datasets, respectively. They were also ranked in the first two positions of importance using the HTG dataset. The allele sequences containing the five SNP markers listed in the importance rank were aligned with the protein-coding genes described in Table 3.

Table 2 – Top five SVM feature importance rank for the nine datasets. Wavelength (correspondent time-point evaluation). Wavelengths: Green, Red, Red-edge, NIR, and Thermal. We named the ranked markers 100401010|F|0NA10:C>GNA10:C>G as marker 1, 100625723|F|0NA8:G>ANA8:G>A as marker 2, 2419277|F|0NA8:C>TNA8:C>T as marker 3, 100608684|F|0NA41:T>CNA41:T>C as marker 4, and 2504229|F|0NA16:C>TNA16:C>T as marker 5.

SVM feature importance rank						
dataset	**training set**	**1º**	**2º**	**3º**	**4º**	**5º**
1	HTP-E1	Red(1)	Green(1)	Red-edge(1)	Thermal(1)	NIR(1)
2	HTP-E2	Red(2)	Green(2)	Red-edge(2)	NIR(2)	Thermal(2)
3	HTP-E3	Red(3)	Thermal(3)	NIR(3)	Green(3)	Red-edge(3)
4	HTP-E1E2E3	Red(3)	Thermal(2)	NIR(3)	Red(2)	Red(1)
5	HTG	marker 1	marker 2	marker 3	marker 4	marker 5
6	HTP-E1+HTG	Thermal(1)	NIR(1)	marker 1	Green(1)	marker 2
7	HTP-E2+HTG	Red-edge(2)	Thermal(2)	marker 1	NIR(2)	marker 2
8	HTP-E3+HTG	Red(3)	NIR(3)	Red-edge(3)	Thermal(3)	marker 1
9	HTP-E1E2E3 + HTG	Red(3)	NIR(3)	Thermal(1)	Thermal(2)	Red-edge(3)

Table 3 – SNP markers' allele sequence aligned gene information. Gene name, description, position, and GO annotation of the ranked markers. The listed markers were named with the same numbers of Table 2. Position: chr (chromosome). GO: gene ontology.

Marker	Gene name	Gene description	Position	GO annotations
1	Zm00001eb406000	100383957, Putative WRKY transcription factor 38	chr10: 3537595..3539838 (+ strand)	GO:0005634: nucleus GO:0003700: DNA-binding transcription factor activity GO:0043565: sequence-specific DNA binding
2	Zm00001eb406020	RHC2A, Putative E3 ubiquitin-protein ligase RHC2A; 103642259, Putative RING zinc finger domain superfamily protein	chr10: 3560820..3561735 (+ strand)	GO:0061630: ubiquitin protein ligase activity GO:0006511: ubiquitin-dependent protein catabolic process
3	Zm00001eb407810	100273197, GATA transcription factor 14; Uncharacterized protein	chr10: 9232979..9235197 (+ strand)	GO:0005634: nucleus GO:0043565: sequence-specific DNA binding GO:0008270: zinc ion binding GO:0030154: cell differentiation GO:0006355: regulation of transcription, DNA-templated
4	Zm00001eb184110	100273793, B-box zinc finger protein 24 (Orphans transcription factor) (Salt tolerance-like protein)	chr4: 139057053..139058564 (- strand)	GO:0005634: nucleus GO:0008270: zinc ion binding GO:0009640: photomorphogenesis GO:0006355: regulation of transcription, DNA-templated
5	Zm00001eb088630	Unknown	chr2: 110999024..111000271 (- strand)	Unknown

3.5 Discussion

3.5.1 Comparing ANN and SVM algorithms

The SVM models were more effective at correctly classifying the resistant and susceptible samples and showed lower standard deviations than ANN models when only the HTP data was used. The HTP-E3 and HTP-E1E2E3 data analysis using the SVM algorithm resulted in the highest accuracy values of 0.82 and 0.83, respectively, and similar CVs. Because the third evaluation is the closest to the disease infection and development full potential, these similar prediction results are consistent with the data analyzed. It is interesting to notice that the prediction accuracies for HTP-E3 and HTP-E1E2E3 using SVM or ANN were fairly similar. That may have happened because evaluations one and two did not add new information/patterns that the models could use to improve their performance.

In contrast, when HTG and HTP+HTG data were employed, the ANN models generalized better and produced similar CVs compared with SVM models. However, the cross-validation prediction metrics of most HTP+HTG datasets using ANN did not suppress the accuracies found with the respective HTP data analyzed with SVM. That may have occurred because the SVM models performed better with smaller datasets, whereas ANN models require a higher number of features to update the weights more effectively and the addition of marker data was not able to aid in the model generalization. The exception was HTP-E1 and HTP-E1+HTG analysis results, where the addition of marker data was beneficial for the model generalization when both algorithms were employed, which possibly happened because the use of phenotypic data alone was insufficient to deliver higher accuracies.

A cross-validation study with liver drug toxicity showed that fewer molecular descriptors delivered a higher SVM classification accuracy (Jaganathan et al., 2021). In another work, Iniyan et al. (2020) compared SVM and ANN models for plant disease identification. The authors concluded that fewer samples and features were required for the SVM model to identify the disease-affected plants using a less complex model, time, and disk space. Therefore, using a k-fold cross validation approach, we concluded that the third evaluation HTP data, which has a small number of features, analyzed with the SVM algorithm yielded a superior resistance classification performance using the data of only one evaluation and should be prioritized in further analysis. In accordance with our findings, J. Zhang et al. (2019) stated that the lack of publicly available image datasets is a limitation in remote sensing, making the SVM algorithm commonly used for disease and pest monitoring due to its ability to generalize effectively with a restricted number of training samples.

The SVM validation curves show a shaded area above and below the F1 mean score for the increasing number of training instances, which represents the standard deviation of the cross-validation scores. For the training curves, the width of the shaded area is related to error due to bias. As a result, the HTP sets appear to suffer from error owing to variance in the cross-validation rather than bias during the training. The variability in the cross-validation curves was higher in the HTP-E1, HTP-E3, and HTP-E1E2E3 sets than HTP-E2, although the HTP-E2 model did not generalize better than the HTP-E3 and HTP-E1E2E3 data-based models. The HTP-E1 training model scores decreased as the number of training samples increased. The addition of more instances may have influenced this model to avoid overfitting and the scores to converge at the end of this training curve. The HTP-E2 and HTP-E1E2E3 datasets followed a pattern of increasing scores in the cross-validation curve and more stable scores in the training curve, culminating with the curves converging. The four HTP curves converging F1-scores indicate that we have added enough data for those HTP data-based models to perform adequately.

The learning curves built with HTG and HTP+HTG data revealed that SVM models suffered from cross-validation variance but not from error due to bias. In addition, cross-validation scores values did not increase much as the number of training instances increased. As a result, their curves did not converge, indicating that more training data are required for the models to generalize better. This learning pattern also suggests that the SVM models were inefficient at preventing overfitting, which could be attributed to the restricted number of maize lines used to train the models and to the inefficiency of marker data to provide useful information for the SVM models' generalization. Consequently, alternative approaches should be addressed to analyze the HTG/HTP+HTG datasets. The ANN algorithm outperformed SVM accuracies when marker data was employed in this analysis, for example. However, a large number of samples are still required for the weights to be properly updated, in addition to other ANN model improvements such as modifying model architecture and increase the number of input features.

3.5.2 SVM feature importance

The Red wavelength reflectance was, generally, the most important feature of our SVM models for MCR resistance classification. The Red transmittance channel is related to photosynthesis, and absorption of leaf, chlorophyll A, and chlorophyll B in plants (X. Zhang et al., 2019). In a study conducted by Chivasa et al. (2020), among eight UAV-derived

multispectral data, the Red band delivered the highest correlation with the visual scores for maize streak virus. J. Zhang et al. (2019) discovered that the diseased-healthy/healthy spectral ratio for the Red wavelength was the highest when compared to Green and three vegetation indices' spectral responses of powdery mildew in wheat, both at canopy and satellite levels. The NIR and Thermal wavelengths were ranked higher in later evaluations, possibly due to the development of more disease symptoms compared to early evaluations. The NIR reflectance is reported to depend on leaf structure and leaf water absorption (Mahlein, 2016). The temperature surface delivered from thermal images is usually associated with plant transpiration rates and stresses-related responses (Mutka & Bart, 2015). In a study using Thermal, hyperspectral, and RGB sensors to monitor powdery mildew in wheat, the parameters derived from the Thermal sensor were the second in order of suitability to predict this disease (Feng et al., 2022).

The SNP markers to feature in the top five importance rank were aligned to the maize reference genome, and their correspondent protein coding-gene information was retrieved. Markers 1 and 3 are located inside the genomic regions of the transcription factors (TF) putative WRKY 38 and GATA 14, respectively. The TF WRKY 38 is a DNA-binding protein that modulates transcription levels in response to stress (Phukan et al., 2016). The GATA TF family is involved in general plant development, photosynthesis, and chlorophyll biosynthesis (Schwechheimer et al., 2022). The OsGATA16, a GATA-type zinc finger TF, conferred cold tolerance to rice at the seedling stage when overexpressed (H. Zhang et al., 2021). In previous studies, markers 2 and 4 were found within protein-coding genes related to stress tolerance. The marker 2 gene is similar to a putative E3 ubiquitin-protein ligase and a RING zinc finger domain superfamily protein, which literally acts as an E3 ubiquitin ligase. Ubiquitin protein ligases are linked with stress-related responses through the post-transcriptional regulation of the leucine-rich repeat resistance proteins, which triggers cell death via hypersensitive response (Han et al., 2022). This gene has been linked to tolerance to a variety of abiotic and biotic stresses, including blast disease resistance in rice (K. Wang et al., 2023) and the regulation of maize hypersensitive response (Karre et al., 2021). The marker 4 gene encodes a B-box zinc finger type protein involved in plant growth and development as well as biotic and abiotic stress responses in maize (Xu et al., 2023). A study in Arabidopsis found that the overexpression of a B-box zinc finger protein enhanced tolerance to salt and drought stress via modulating abscisic acid pathways (Liu et al., 2019). The function of the marker 5 gene is unknown. Although a ML method was used to rank the SNP markers in order of importance for MCR classification and no genetic technique was added, this information can be of

contribution in future studies of resistance to MCR. However, the subsequent validation processes of the five genes identified should be systematically applied to confirm their importance in MCR resistance response.

3.6 Conclusion

This study explored resistant or susceptible cross-validation classification of maize lines to MCR with Support Vector Machine and Artificial Neural Network with ML algorithms using HTP and HTG data. Our results indicate that, when the HTG data is combined with HTP data, the increasing number of features contributed for the superiority of ANN over SVM classification accuracy. However, the highest prediction performance found in our analyzes was achieved by using SVM trained with HTP data from the last evaluation and the three time-point evaluations combined. In addition to achieving the highest overall accuracy and lowest standard deviation, the SVM algorithm used significantly less computational power and a fewer number of samples to deliver a higher classification performance. Therefore, due to its less complex model architecture, SVM should be considered for maximized use with the third time-point evaluation dataset, and as new features are introduced, its prediction performance should be systematically monitored to achieve optimal results. The SVM models' learning curves suggest that improvements such as adding more training samples might be effective in reducing variance in the nine sets. The SVM feature rank of importance revealed that the Red band was the most crucial for MCR resistance/susceptible classification. The markers ranked are located within protein-coding genes, which can be further investigated in genomic-functional studies, such as in the validation of genes, transcription factors, and promoters.

3.7 Acknowledgments

The study was supported by the CGIAR Research Program on Maize (MAIZE). The authors thank CIMMYT's El Batan and Agua Fria experimental station scientists involved in the regional trials. This study was also supported by CNPq (National Council for Scientific and Technological Development) and UFV (Universidade Federal de Viçosa). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. The contents and opinions expressed herein are those of

the authors and do not necessarily reflect the views of the associated and supporting institutions.

3.8 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bebronne, R., Carlier, A., Meurs, R., Leemans, V., Vermeulen, P., Dumont, B., & Mercatoris, B. (2020). In-field proximal sensing of septoria tritici blotch, stripe rust and brown rust in winter wheat by means of reflectance and textural features from multispectral imagery. *Biosystems Engineering*, *197*, 257–269. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2020.06.011

Bombrun, M., Dash, J. P., Pont, D., Watt, M. S., Pearse, G. D., & Dungey, H. S. (2020). Forest-Scale Phenotyping: Productivity Characterisation Through Machine Learning. *Frontiers in Plant Science*, *11*(99), 1–14. https://doi.org/10.3389/fpls.2020.00099

Calou, V. B. C., Teixeira, A. dos S., Moreira, L. C. J., Lima, C. S., de Oliveira, J. B., & de Oliveira, M. R. R. (2020). The use of UAVs in monitoring yellow sigatoka in banana. *Biosystems Engineering*, *193*, 115–125. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2020.02.016

Chen, Y., Sidhu, H. S., Kaviani, M., McElroy, M. S., Pozniak, C. J., & Navabi, A. (2019). Application of image-based phenotyping tools to identify QTL for in-field winter survival of winter wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, *132*(9), 2591–2604. https://doi.org/10.1007/s00122-019-03373-6

Chivasa, W., Mutanga, O., & Biradar, C. (2020). UAV-Based Multispectral Phenotyping for Disease Resistance to Accelerate Crop Improvement under Changing Climate Conditions. In *Remote Sensing* (Vol. 12, Issue 15). https://doi.org/10.3390/rs12152445

Dong, X.-J., Shen, J.-N., He, G.-X., Ma, Z.-F., & He, Y.-J. (2021). A general radial basis function neural network assisted hybrid modeling method for photovoltaic cell operating temperature prediction. *Energy*, *234*, 121212. https://doi.org/https://doi.org/10.1016/j.energy.2021.121212

Feng, Z., Song, L., Duan, J., He, L., Zhang, Y., Wei, Y., & Feng, W. (2022). Monitoring Wheat Powdery Mildew Based on Hyperspectral, Thermal Infrared, and RGB Image Data Fusion. In *Sensors* (Vol. 22, Issue 1). https://doi.org/10.3390/s22010031

Gui, J., Fei, J., Wu, Z., Fu, X., & Diakite, A. (2021). Grading method of soybean mosaic disease based on hyperspectral imaging technology. *Information Processing in Agriculture*, *8*(3), 380–385. https://doi.org/https://doi.org/10.1016/j.inpa.2020.10.006

Han, G., Qiao, Z., Li, Y., Yang, Z., Wang, C., Zhang, Y., Liu, L., & Wang, B. (2022). RING Zinc Finger Proteins in Plant Abiotic Stress Tolerance. In *Frontiers in Plant Science* (Vol.

13). https://www.frontiersin.org/articles/10.3389/fpls.2022.877011

Iniyan, S., Jebakumar, R., Mangalraj, P., Mohit, M., & Nanda, A. (2020). Plant Disease Identification and Detection Using Support Vector Machines and Artificial Neural Networks. In S. S. Dash, C. Lakshmi, S. Das, & B. K. Panigrahi (Eds.), *Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing, vol 1056* (pp. 15–27). Springer Singapore. https://doi.org/10.1007/978-981-15-0199-9_2

Jaganathan, K., Tayara, H., & Chong, K. T. (2021). Prediction of Drug-Induced Liver Toxicity Using SVM and Optimal Descriptor Sets. In *International Journal of Molecular Sciences* (Vol. 22, Issue 15). https://doi.org/10.3390/ijms22158073

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C. S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., …Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Karre, S., Kim, S.-B., Samira, R., & Balint-Kurti, P. (2021). The maize ZmMIEL1 E3 ligase and ZmMYB83 transcription factor proteins interact and regulate the hypersensitive defence response. *Molecular Plant Pathology*, *22*(6), 694–709. https://doi.org/https://doi.org/10.1111/mpp.13057

Kim, J., Kim, K.-S., Kim, Y., & Chung, Y. S. (2020). A short review: Comparisons of high-throughput phenotyping methods for detecting drought tolerance. In *Scientia Agricola* (Vol. 78). scielo.

Liu, X., Li, R., Dai, Y., Yuan, L., Sun, Q., Zhang, S., & Wang, X. (2019). A B-box zinc finger protein, MdBBX10, enhanced salt and drought stresses tolerance in Arabidopsis. *Plant Molecular Biology*, *99*(4), 437–447. https://doi.org/10.1007/s11103-019-00828-8

Loladze, A., Rodrigues Jr, F. A., Petroli, C. D., Munoz, C., Naranjo, S. M., Vicente, F. S., Gerard, B., Montesinos-Lopez, O. A., Crossa, J., Martini, J. W. R. (2023). Use of Remote Sensing for Genome-Wide Association Studies and Genomic Prediction [Unpublished manuscript]. International Maize and Wheat Improvement Center (CIMMYT El Batan, Mexico).

Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease*, *100*(2), 1–11. https://doi.org/10.1007/s13398-014-0173-7.2

Manrique-Silupu, J., Campos, J. C., Paiva, E., & Ipanaqué, W. (2021). Thrips incidence prediction in organic banana crop with Machine learning. *Heliyon*, *7*(12). https://doi.org/10.1016/j.heliyon.2021.e08575

Mutka, A. M., & Bart, R. S. (2015). Image-based phenotyping of plant disease symptoms. *Frontiers in Plant Science*, *5*, 1–8. https://doi.org/10.3389/fpls.2014.00734

Nicholls, H. L., John, C. R., Watson, D. S., Munroe, P. B., Barnes, M. R., & Cabrera, C. P.

(2020). Reaching the End-Game for GWAS: Machine Learning Approaches for the Prioritization of Complex Disease Loci. In *Frontiers in Genetics* (Vol. 11). https://www.frontiersin.org/articles/10.3389/fgene.2020.00350

Olukolu, B. A., Tracy, W. F., Wisser, R., De Vries, B., & Balint-Kurti, P. J. (2016). A genome-wide association study for partial resistance to maize common rust. *Phytopathology*, *106*(7), 745–751. https://doi.org/10.1094/PHYTO-11-15-0305-R

Omari, M. K., Lee, J., Faqeerzada, Mohammad Akbar Joshi, R., Park, E., & Cho, B.-K. (2020). Digital image-based plant phenotyping: a review. *Korean Journal of Agricultural Science*, *20*(4), 5–13. https://doi.org/10.7744/kjoas.20200004

Phukan, U. J., Jeena, G. S., & Shukla, R. K. (2016). WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. *Frontiers in Plant Science*, *7*, 1–14. https://doi.org/10.3389/fpls.2016.00760

Quirós Vargas, J. J., Zhang, C., Smitchger, J. A., McGee, R. J., & Sankaran, S. (2019). Phenotyping of Plant Biomass and Performance Traits Using Remote Sensing Techniques in Pea (Pisum sativum, L.). *Sensors (Basel, Switzerland)*, *19*(9), 2031. https://doi.org/10.3390/s19092031

Ramirez-Cabral, N. Y. Z., Kumar, L., & Shabani, F. (2017). Global risk levels for corn rusts (Puccinia sorghi and Puccinia polysora) under climate change projections. *Journal of Phytopathology*, *165*(9), 563–574. https://doi.org/10.1111/jph.12593

Ritchie, S. W., Hanway, J. J., & Benson, G. O. (1986). *How a Corn Plant Develops. Spec. Rep. 48*.

Rohani, A., Taki, M., & Abdollahpour, M. (2018). A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renewable Energy*, *115*, 411–422. https://doi.org/https://doi.org/10.1016/j.renene.2017.08.061

Schwechheimer, C., Schröder, P. M., & Blaby-Haas, C. E. (2022). Plant GATA Factors: Their Biology, Phylogeny, and Phylogenomics. *Annual Review of Plant Biology*, *73*(1), 123–148. https://doi.org/10.1146/annurev-arplant-072221-092913

Tirado, S. B., Hirsch, C. N., & Springer, N. M. (2020). UAV-based imaging platform for monitoring maize growth throughout development. *Plant Direct*, *4*(6), 1–11. https://doi.org/10.1002/pld3.230

Wang, K., Li, S., Chen, L., Tian, H., Chen, C., Fu, Y., Du, H., Hu, Z., Li, R., Du, Y., Li, J., Zhao, Q., & Du, C. (2023). E3 ubiquitin ligase OsPIE3 destabilises the B-lectin receptor-like kinase PID2 to control blast disease resistance in rice. *New Phytologist*, *237*(5), 1826–1842. https://doi.org/https://doi.org/10.1111/nph.18637

Xu, X., Li, W., Yang, S., Zhu, X., Sun, H., Li, F., Lu, X., & Cui, J. (2023). Identification, evolution, expression and protein interaction analysis of genes encoding B-box zinc-finger proteins in maize. *Journal of Integrative Agriculture*, *22*(2), 371–388. https://doi.org/https://doi.org/10.1016/j.jia.2022.08.091

Zhang, H., Wu, T., Li, Z., Huang, K., Kim, N.-E., Ma, Z., Kwon, S.-W., Jiang, W., & Du, X. (2021). OsGATA16, a GATA Transcription Factor, Confers Cold Tolerance by Repressing OsWRKY45–1 at the Seedling Stage in Rice. *Rice*, *14*(1), 42. https://doi.org/10.1186/s12284-021-00485-w

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., & Huang, W. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, *165*, 104943. https://doi.org/https://doi.org/10.1016/j.compag.2019.104943

Zhang, X., Zhang, F., Li, M., Zhang, Z., & Liu, W. (2019). Design and applications of multichannel photosynthetic photon flux density sensor for artificial lighting. *Optical Engineering*, *58*(2), 27105. https://doi.org/10.1117/1.OE.58.2.027105

Zhang, Z., Pope, M., Shakoor, N., Pless, R., Mockler, T. C., & Stylianou, A. (2022). Comparing Deep Learning Approaches for Understanding Genotype × Phenotype Interactions in Biomass Sorghum. In *Frontiers in Artificial Intelligence* (Vol. 5). https://www.frontiersin.org/articles/10.3389/frai.2022.872858

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., Wang, J., & Fan, J. (2019). Crop phenomics: Current status and perspectives. *Frontiers in Plant Science*, *10*(June). https://doi.org/10.3389/fpls.2019.00714

Zhao, J., Fang, Y., Chu, G., Yan, H., Hu, L., & Huang, L. (2020). Identification of Leaf-Scale Wheat Powdery Mildew (Blumeria graminis f. sp. Tritici) Combining Hyperspectral Imaging and an SVM Classifier. In *Plants* (Vol. 9, Issue 8). https://doi.org/10.3390/plants9080936

Zheng, H., Chen, J., Mu, C., Makumbi, D., Xu, Y., & Mahuku, G. (2018). Combined linkage and association mapping reveal QTL for host plant resistance to common rust (Puccinia sorghi) in tropical maize. *BMC Plant Biology*, *18*(1), 1–14. https://doi.org/10.1186/s12870-018-1520-1

4. CONCLUSIONS

The earlier MCR evaluations were unable to provide accurate predictions as good as the late evaluation. Those evaluations did not represent the highest disease incidence potential, only accessing a limited range of visual scores. The last evaluation delivered a higher overall performance in this resistant/susceptible supervised classification analysis, although it was necessary to wait longer to evaluate the samples, resulting in greater resource deployment throughout the season. Algorithms LR and SVM performed well in the late evaluation set and had the lowest CV. So, it should be carefully considered in future analyzes.

An uneven distribution of classes in ML analyzes can lead to a deficiency to predict a certain class, which may have occurred with the late evaluation data for year one. Aiming to overcome this problem, future studies may choose a number of genotypes to represent each score, which can be used on a training set to build the models. With this set, it would be possible to select parameters, cross-validate the set, and build the ML models more effectively. With this selection, some variability that could be integrated into the models would be lost, but this extra step would ideally boost trainability, precision, and recall by assisting the models to better capture patterns of the data to be added to future predictions.

The comparison of SVM and ANN algorithms using k-fold cross-validation of HTP/HTG/HTP+HTG data resulted in different MCR resistance/susceptible classification performances depending on the data analyzed. Datasets with marker data performed better using ANN, but the overall accuracy was superior when SVM was utilized in the HTP sets. Ultimately, besides traditional performance metrics, computer power, processing time, and model architecture can play a decisive role in choosing the most appropriate algorithm for a specific task. With HTP and HTG data, SVM learning curves showed that the addition of more training samples would be beneficial to minimize variance. The Red wavelength was the most important feature in most ranks of importance using the SVM algorithm. The TF and genes ranked are related to stress response according to recently published literature and should be considered for validation procedures.