UNIVERSIDADE FEDERAL DE VIÇOSA

Determination of wheat seed quality in the presence of Fusarium graminearum using near-infrared spectroscopy combined with machine learning technique

José Maria da Silva
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2024

JOSÉ MARIA DA SILVA

Determination of wheat seed quality in the presence of Fusarium graminearum using near-infrared spectroscopy combined with machine learning technique

Thesis submitted to the Plant Production Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Laercio Junio da Silva

VIÇOSA - MINAS GERAIS
2024

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

T
S586d
2024

Silva, José Maria da, 1992-
Determination of wheat seed quality in the presence of *Fusarium graminearum* using near-infrared spectroscopy combined with machine learning technique / José Maria da Silva. – Viçosa, MG, 2024.
1 tese eletrônica (96 f.): il. (algumas color.).

Texto em inglês.
Orientador: Laércio Junio da Silva.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2024.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2025.037
Modo de acesso: World Wide Web.

1. Trigo - Doenças e pragas. 2. Sementes - Doenças e pragas. 3. Germinação. 4. Sementes - Qualidade - Modelos matemáticos. 5. Espectroscopia de infravermelho.
6. Aprendizado do computador. I. Silva, Laércio Junio da, 1984-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Fitotecnia.
III. Título.

CDD 22. ed. 633.11523

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

JOSÉ MARIA DA SILVA

Determination of wheat seed quality in the presence of Fusarium graminearum using near-infrared spectroscopy combined with machine learning technique

Thesis submitted to the Plant Production Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: October 29, 2024.

Assent:

_____

José Maria da Silva
Author

_____

Laercio Junio da Silva
Adviser

ACKNOWLEDGMENTS

A Deus pela minha vida e por estar presente em todos os momentos.

Aos meus pais, Helvécio e Efigênia, pela confiança e por acreditarem sempre na importância dos estudos e em seu poder de transformação.

À minha namorada e melhor amiga, Jucilane Almeida, por todo incentivo, carinho e paciência em todos esses anos.

Aos meus irmãos Regiane, Vanessa e Antônio, meus tios, primos e demais familiares por toda ajuda e apoio.

À Universidade Federal de Viçosa (UFV) pelo acolhimento e oportunidades concedidas desde a Graduação em Agronomia até o Doutorado em Fitotecnia. Foram anos incríveis de desenvolvimento pessoal e profissional.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pela concessão da bolsa de estudos.

Ao Programa de Pós-Graduação em Fitotecnia e ao Departamento de Agronomia pela oportunidade de realização do curso e por toda a minha formação acadêmica. Agradeço a todos os colegas de pós-graduação, graduação, professores, servidores e funcionários do Departamento que contribuíram para a realização deste trabalho.

Ao Professor Laércio Junio da Silva, pela orientação, confiança, amizade, ajuda, paciência e sobretudo por tornar possível a realização desse trabalho.

À Professora Denise Cunha Fernandes dos Santos Dias pelos ensinamentos, contribuições e apoio.

Ao Professor Franklin Jackson Machado, do Departamento de Fitopatologia da UFV, pelo acolhimento no laboratório de Manejo de Doenças de Plantas, pelas orientações e apoio nas análises de sanidade de sementes, além das

sugestões valiosas no Exame de Qualificação.

Ao Professor José Antônio dos Santos Dias pela disponibilização do laboratório de Agroenergia para realização das análises de espectroscopia do infravermelho próximo.

Aos amigos do Laboratório de Sementes pelo acolhimento e ajuda com a condução do trabalho, em especial ao Wander, Manuel, José Custódio, Wanderson, Giulia, Nayara, André, Guilherme, Joyce e Júlia.

Ao Gabriel e à Lara, do Departamento de Fitopatologia, pela ajuda na realização e condução do experimento.

Aos amigos Carla e Edimaldo por todo incentivo, companheirismo e apoio em todos os momentos.

Aos Professores Daniel Pinheiro e Martha Freire pelas sugestões durante o Exame de Qualificação que foram fundamentais para o aperfeiçoamento do trabalho.

Aos membros da banca de Defesa da Tese, Professor Laércio, Professora Denise, Professor Hugo Catão, Doutora Jussara Roque e Doutor Wander Pereira por todas sugestões e correções fundamentais para melhoria do trabalho.

Aos amigos de república/alojamento pela convivência ao longo desses anos, em especial ao Júnior, Felipe, Lusvânio, Elias, Evandro e Nathan.

Aos colegas e amigos da Empresa de Assistência Técnica e Extensão Rural do estado de Minas Gerais (EMATER-MG), pelo incentivo na conclusão deste trabalho.

Enfim, a todos que direta ou indiretamente contribuíram para a realização e conclusão desse trabalho.

Muito obrigado!

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

“*I can do all things through Christ who strengthens me.*”
(Philippians 4:13)

ABSTRACT

SILVA, José Maria da, D.Sc., Universidade Federal de Viçosa, October, 2024. **Determination of wheat seed quality in the presence of Fusarium graminearum using near-infrared spectroscopy combined with machine learning technique**. Adviser: Laercio Junio da Silva.

Wheat is a key global crop that significantly contributes to food security. In Brazil, approximately 50% of wheat consumption relies on imports, making the increase in domestic production a priority. Expanding wheat cultivation necessitates the use of high-quality seeds, with seed health being critical due to the presence of various pathogens. Among these, *Fusarium graminearum*, the causative agent of Fusarium head blight (FHB), is known to reduce seed germination and vigor, thereby affecting field establishment. Near-infrared (NIR) spectroscopy, combined with machine learning technique, offers a non-destructive method for assessing seed quality. However, the effectiveness of this approach depends on the appropriate data preprocessing methods. This study aimed to evaluate different preprocessing techniques to determine the quality of wheat seeds infected by *F. graminearum* and to assess the application of the second derivative of Savitzky-Golay as a preprocessing method across four datasets derived from final germination percentage, seedling growth, emergence, and blotter tests. Two wheat cultivars were grown in a greenhouse and inoculated with *F. graminearum* at three stages: anthesis, early milk, and early dough. Following the crop cycle, individual seeds were harvested and analyzed using NIR spectroscopy. The seeds then underwent quality tests to classify each spectrum. The datasets were preprocessed, and the Partial Least Squares Discriminant Analysis (PLS-DA) algorithm was used to classify the seeds. Each model was evaluated using multiple performance metrics to assess its classification capability, and key wavelength regions that contributed to class separation were identified. The findings indicated that binary classification models outperformed three-class models, regardless of the preprocessing method used. Models based on growth tests exhibited lower performance across all preprocessing methods. The spectral regions of 1000-1300 nm and 1800-2000 nm were consistently important for class separation across different seed quality tests. These regions could be utilized in seed quality control programs employing NIR spectroscopy for non-destructive analysis. Additionally, the study highlighted the importance of using diverse metrics to ensure robust results, particularly in cases of imbalanced datasets.

Keywords: preprocessing; accuracy ; wheat scab; pls-da

RESUMO

SILVA, José Maria da, D.Sc., Universidade Federal de Viçosa, outubro de 2024. **Determinação da qualidade de sementes de trigo na presença de Fusarium graminearum usando espectroscopia de infravermelho próximo combinada com técnica de aprendizado de máquina**. Orientador: Laercio Junio da Silva.

O trigo é uma cultura de extrema importância para a segurança alimentar global. No Brasil, aproximadamente 50% do consumo de trigo depende de importações, tornando o aumento da produção doméstica uma prioridade. A expansão da área cultivada com trigo exige o uso de sementes de alta qualidade, sendo a sanidade das sementes fundamental devido à presença de vários patógenos. Entre esses, *Fusarium graminearum*, o agente causador da giberela (FHB), é conhecido por reduzir a germinação e o vigor das sementes, afetando o estabelecimento no campo. A espectroscopia no infravermelho próximo (NIR), combinada com técnica de aprendizado de máquina, oferece um método não destrutivo para avaliar a qualidade das sementes. No entanto, a eficiência da técnica depende dos métodos apropriados de pré-processamento dos dados. Este estudo teve como objetivo avaliar diferentes técnicas de pré-processamento para determinar a qualidade das sementes de trigo infectadas por *F. graminearum* e avaliar a aplicação da segunda derivada de Savitzky-Golay como método de pré-processamento em quatro conjuntos de dados derivados dos testes de germinação, crescimento de plântulas, emergência e teste de *Blotter*. Duas cultivares de trigo foram cultivadas em casa de vegetação e inoculadas com *F. graminearum* em três estágios: antese, grão leitoso e grão pastoso. Após o ciclo da cultura, sementes individuais foram colhidas e analisadas por espectroscopia NIR. Em seguida, as mesmas sementes foram submetidas a análise de germinação e vigor para obtenção da respectiva classe para compor o arquivo de dados. Os conjuntos de dados foram pré-processados, e o algoritmo de Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) foi utilizado para classificar as sementes. Cada modelo foi avaliado usando múltiplas métricas de desempenho para verificar sua capacidade de classificação, e as principais regiões de comprimento de onda que contribuíram para a separação das classes foram identificadas. Os resultados indicaram que modelos de classificação binária superaram os modelos de três classes, independentemente do método de pré-processamento utilizado. Os modelos baseados em testes de crescimento apresentaram menor desempenho em todos os métodos de pré-processamento. As regiões espectrais de 1000-1300 nm e 1800-2000 nm foram consistentemente importantes para a separação das classes em

diferentes testes de qualidade de sementes. Essas regiões poderiam ser utilizadas em programas de controle de qualidade de sementes que empregam espectroscopia NIR para análise não destrutiva. Além disso, o estudo destacou a importância do uso de diferentes métricas para garantir resultados robustos, especialmente em casos de conjuntos de dados desbalanceados.

Palavras-chave: pré-processamento; acurácia; giberela; pls-da

LIST OF ILLUSTRATIONS

LIST OF TABLES

SUMMARY

GENERAL INTRODUCTION

Sustainable agricultural systems are the core of producing enough food for a growing population worldwide despite limited resources. According to projections of the Food and Agriculture Organization (FAO) of the United Nations, the world population is expected to increase to 9.1 billion people in 2050. This would require raising overall food production to about 60% to meet the increased food demand (FAO, 2009). The increase in food production has to be done in the scenario of significant climate changes which poses some challenges to the agricultural sector. Therefore, the agricultural systems must increase food production with limited resources and reduce environmental impacts. In this context, it is necessary to increase the efficiency of crop production systems, which are dependent on seed technology.

Seed technology is directly related to quality, which comprises the attributes of physical, physiological, genetic, and sanitary quality. Physical quality is related to the purity of the seed lot, meaning the absence of foreign materials such as stones, other seeds, etc. Physiological quality corresponds to the ability of a given seed to form a new plant. It is commonly measured through germination tests, which are required by law, as well as vigor tests. Genetic quality ensures that a specific seed lot is composed solely of seeds from the cultivar in question, thereby eliminating the occurrence of cross-breeding with other cultivars. Finally, the fourth pillar of seed quality aims to ensure the absence of pathogens and pests associated with the seeds (MCDONALD and, COPELAND, 1997).

The infection of wheat seeds by *Fusarium graminearum* negatively impacts the quality of seeds. It has been associated to cause a reduction in final germination percentage (GARCIA JUNIOR et al., 2007), increasing the number of abnormal seedlings (HASSANI et al. 2019), and the number of dead seeds of tested lots (ASRAN and AMAL, 2011). Also, the infected seeds are a source of contamination of new areas, which could impact other susceptible crops in the field. Therefore, determining the presence of this pathogen associated with the seeds could be part of a quality control program to guarantee the health of commercialized seeds.

The industry has increasingly adopted nondestructive seed analysis, considering certain advantages, such as the speed of delivering results, which can be determined in the field, and the ability to use the sample to monitor processing and storage strategies over time. Near-infrared spectroscopy (NIR) is a rapid, non-destructive technique that provides insights into the chemical composition of products like seeds, including key components such as lipids, proteins, and carbohydrates (XIA et al., 2019). NIR covers a spectral range from 780 to 2500 nm and is based on the measurement of vibrations of atoms in organic molecules, primarily involving the C-H, C-O, O-H, and N-H bonds (BURNS and CIURCZACK, 2007; ORINA et al., 2017).

Different authors have studied the utilization of NIR to determine seed quality (VENKATESAN et al., 2020). Fan et al. (2020) explored the application of NIR combined with machine learning techniques to assess the vigor of individual wheat kernels. All tested models achieved accuracies above 84%, indicating the effectiveness of NIR for classifying wheat seed vigor non-destructively. The study suggests that integrating NIR with advanced machine learning algorithms offers a promising technique for rapid and non-destructive seed vigor assessment, potentially aiding in plant breeding and quality control processes. Additionally, it could save time and reduce costs by enabling the early identification and rejection of inferior seed lots before the cleaning process. However, the non-destructive wheat seed quality determination using NIR in the context of infection by *Fusarium graminearum* is still uncleared. This study aimed to evaluate the physiological and sanitary quality of wheat seeds infected by *Fusarium graminearum* using NIR spectroscopy and different pre-processing algorithms.

REFERENCES

Asran, M. R., Amal, M. I. E. (2011). Aggressiveness of certain *Fusarium graminearum* isolates on wheat seedlings and relation with their trichothecene production. *Plant Pathology Journal*, *10*(1), 36–41. https://doi.org/10.3923/ppj.2011.36.41

Burns, D.A., & Ciurczak, E.W. (Eds.). (2007). Handbook of Near-Infrared Analysis (3rd ed.). CRC Press. https://doi.org/10.1201/9781420007374

Fan, Y., Ma, S., Wu, T. (2020). Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Physics and Technology*, *105*(1), 103213. https://doi.org/10.1016/j.infrared.2020.103213

Food and Agriculture Organization (FAO). 2009. How to Feed the World in 2050. Available online: https://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Garcia Júnior, D., Vechiato, M. H., Menten, J. O. M., Lima, M. I. P. M. (2007). Influência de *Fusarium graminearum* na germinação de genótipos de trigo (*Triticum aestivum* L.). *Arquivos do Instituto Biológico*, 74(2), 157–162.

Hassani, F. Leila Z., and Nima K. (2019). Evaluation of germination and vigor indices associated with *Fusarium*-infected seeds in pre-basic seeds wheat fields. *Journal of Plant Protection Research,* 59 (1): 69–85. https://doi.org/10.24425/jppr.2019.126037.

McDonald, M.B., Copeland, L.O. (1997). Seed Quality and Performance. In: Seed Production. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-4074-8_8

Orina, I., Manley, M., Williams, P. J. (2017). Non-destructive techniques for the detection of fungal infection in cereal grains. *Food Research International*, *100*, 74-86. https://doi.org/10.1016/j.foodres.2017.07.069

Venkatesan, S., Masilamani, P., Janaki, P., Eevera, T., Sundareswaran, S., Rajkumar, P. (2020). Role of near-infrared spectroscopy in seed quality evaluation: A review. *Agricultural Reviews*. 41(2): 106-115. https://doi.org/10.18805/ag.R-1960

Xia, Y., Xu, Y., Li, J., Zhang, C., Fan, S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, *1*, 35–47. https://doi.org/10.1016/j.aiia.2019.05.001

CHAPTER I

Nondestructive Evaluation of Wheat Seed Viability and Vigor Under *Fusarium graminearum* Infection Using Near-infrared Spectroscopy and Different Preprocessing Methods.

1.1. Abstract

Wheat scab, caused by *Fusarium graminearum*, has a serious impact on wheat production by infecting plants and compromising seed quality. This disease can result in substantial losses during anthesis or from late infections. Seed quality, including germination and vigor, is adversely affected by mycotoxins produced by the pathogen. Recent advances in near-infrared spectroscopy (NIR) offer a rapid, non-destructive alternative for seed quality evaluation. This study investigates various NIR data preprocessing methods to improve seed physiological and sanitary quality predictions in wheat cultivars infected with *Fusarium graminearum*. The wheat cultivars, BR18 and Guamirim, were grown in 3-liter pots containing commercial substrate. After reaching the reproductive phase, the plants were separated and inoculated with *Fusarium graminearum* at three stages (complete anthesis, early milk, and early dough stage) and a control (no inoculation). At the end of the cycle, the seeds were harvested, individualized, and scanned using a Fourier Transform Infrared (FT-NIR) spectrometer. After scanning, the seeds underwent germination, growth, emergence, and blotter tests to determine the class of each seed. The datasets were submitted to 12 different preprocessing methods and separated into train and test to classify the seeds. Optimal models for seed classification using NIR were identified for different quality tests. For germination, the first derivative of Savitzky-Golay with a second-order polynomial and a window size of 11 (SG_D1_W11) was more efficient in distinguishing the classes from BR18 and the standard normal variate combined with the second-order Savitzky-Golay (SNV_SG) method accurately distinguished normal and dead seeds, for Guamirim cultivar. In the growth test, misclassification was observed for all tested models. The classification of seeds from the emergence test benefited from applying the second-order Savitzky-Golay to the Standard Normal Variate

dataset (SNV_SG) for BR18 and the Savitzky-Golay combined with gap-segment derivative algorithm (SG_D1) for Guamirim. In contrast, the SNV for BR18 and D1 for Guamirim were highlighted for the Blotter test as the most effective for detecting infected seeds. The regression coefficients showed key spectral regions for class separation, particularly around 1000-1300 nm, 2000 nm, and 2300 nm. NIR successfully distinguished between viable and non-viable seeds but struggled with intermediate classes like abnormal or low-vigor seeds. The Savitzky-Golay preprocessing algorithm, especially when combined with standard normal variate, proved effective across various tests and cultivars, enhancing spectral peak resolution and classification accuracy.

Keywords: Fusarium head blight. Wheat scab. Machine learning. Savitzky-Golay filter.

1.2. Introduction

Wheat scab, a disease caused by the fungus *Fusarium graminearum*, is a global concern in wheat production. The fungus spreads through spores released from infected plant debris, carried by wind or water to infect healthy flowers. This disease can affect all parts of the wheat plant, from the stem to the ears, leading to significant seed production losses (DEL PONTE et al., 2004; GARCIA JUNIOR et al., 2007). Infections during anthesis can result in total production loss, as the pathogen hinders seed formation (ALISAAC et al., 2021). Conversely, late infections can lead to a surge in contaminated seeds that are challenging to remove through traditional post-harvest cleaning methods.

Companies conduct various commercialization and quality control tests to ensure that a specific lot meets legal and institutional standards for high-quality seeds. The germination test, a mandatory step for commercialization, requires the lot to demonstrate a minimum of 80% normal seedlings capable of establishing and forming an adequate stand under favorable conditions (MAPA, 2013). This test distinguishes normal seedlings, abnormal seedlings, and dead seeds. Institutions also employ vigor tests to assess the potential for rapid and uniform seedling emergence. Wheat scab directly impacts the quality of the seeds by compromising basic seed structures, which is evidenced by these tests. The

mycotoxins can inhibit cell division, leading to seedling deformities and reduced initial development, potentially causing stand failures (ASRAN & AMAL, 2011; GARCIA JUNIOR et al., 2007; HASSANI et al., 2019).

Standard laboratory tests evaluate the physiological and biochemical aspects of a seed lot, often involving destructive techniques that induce radicle protrusion or expose seeds to stress conditions. However, in recent years, rapid and non-destructive techniques like near-infrared spectroscopy have emerged as promising alternatives (XIA et al., 2019). NIR can identify the chemical composition of seeds, including lipids, proteins, and carbohydrates (FAN et al., 2020). It has also been used to detect fungi like *Fusarium* and mycotoxins in seeds, offering a more rapid and non-destructive method of seed quality assessment (PEIRIS et al., 2010).

NIR corresponds to the wavelength range between 780 and 2500 nm of the electromagnetic spectrum. This technique is based on absorbing infrared radiation by chemical bonds in molecules, causing specific vibrational transitions. Each type of chemical bond, such as Carbon-Hydrogen (C-H), Oxygen-Hydrogen (O-H), and Nitrogen-Hydrogen (N-H), absorbs infrared radiation at characteristic wavelengths, allowing the identification and quantification of various chemical components in a sample (BURNS et al., 2007). With the obtained spectrum, the data are used to generate a dataset and, through chemometrics, extract information about the sample under study.

For the development of classification models, the spectrum undergoes a preparation phase to remove intrinsic noise from the device and the sample. Various preprocessing methodologies and algorithms for NIR data are found in the literature (RINNAN et al., 2009). However, to integrate the technique into routine seed analysis, it is necessary to establish specific methodologies capable of analyzing a dataset to extract valuable information rapidly and efficiently. Thus, this study aimed to test different combinations of NIR data preprocessing methods from seeds inoculated with *Fusarium graminearum* to predict the physiological and sanitary quality of two wheat cultivars.

1.3. Material and methods

1.3.1. Seed sample preparation

Commercial wheat seeds of two cultivars known as susceptible (BR18) (SOUSA, 2002), and moderate resistant (Guamirim) (SCHEEREN, et al. 2007) to wheat scab were cultivated in a greenhouse for 210 days at the Universidade Federal de Viçosa-MG, Brasil. The selected plants were separated and sprayed with *Fusarium graminearum* inoculum with suspension adjusted to a final concentration of 1 x 10^4 macroconidia/mL at anthesis, early milk, and early dough. Additionally, there were plants without inoculation (control). The plants were kept inside the greenhouse until the seed development process was complete. Then, they were harvested and thrilled by hand to avoid damage to the kernels. Each inoculation period and the control constituted a seed lot to be evaluated by near-infrared spectroscopy.

1.3.2. Near Infrared spectral acquisition (NIR)

The spectral data were obtained with a Fourier Transform Infrared Spectrometer FT-NIR (Thermo Scientific Antaris II). Individual seeds from each lot were placed in a specific support to the size and shape of the wheat seeds. For each seed, 3111 points per spectrum were collected within the 1000 – 2500 nm wavelength range and a resolution of 8 cm^{-1}. The average spectrum of 30 successive scans for each seed was evaluated.

1.3.3. Seed class determination

After obtaining the spectra, the seeds were submitted to seed testing to determine the seed class. Four main germination and vigor tests were evaluated: Final germination percentage, growth, emergence, and blotter test. For the final germination test, 800 seeds from each cultivar were placed to germinate on towel paper for 8 days, according to the International Seed Testing Association (ISTA, 2018). After this period, the seeds were classified as normal, abnormal, or dead seeds.

The growth test was performed using 320 seeds from each cultivar. They were incubated for 72 hours at 20 ºC. After this period, images of seedlings were captured and measured with ImageJ software. Seedlings greater than 30 mm were classified as high vigor, and those with less than 30 mm were classified as

low vigor. The class of dead seeds were those without protrusion of the radicle after the test period and with intense deterioration by pathogens.

The emergence test was conducted with 800 seeds for each cultivar. After NIR readings, the seeds were sowed to a trail filled with sand and kept in a greenhouse for 15 days. Daily count of the seedlings greater than 5 mm above the surface was performed. After this period, the seeds were classified as viable or non-viable.

Finally, 400 disinfested seeds from each cultivar were submitted for the Blotter test, according to Brasil, 2009. The seeds were sown in acrylic boxes (40x40x40 cm) on three sheets of towel paper, moistened with saline solution (NaCl) to prevent seed germination. The whole process was done under sterile conditions inside the laminar flow cabinet. The boxes were placed in BOD regulated at 25 °C for seven days. After this period, an expert visually evaluated mycelium growth typical of *Fusarium graminearum* after incubation time, and the seeds were classified as infected or non-infected.

1.3.4. Preprocessing and classification methods

The spectral data were organized with singular entrances corresponding to the seed and the respective class from seed testing. Then, the dataset was submitted to 12 preprocessing methods (Table 1).

Table 1. Preprocessing methods applied to the datasets obtained with NIR spectroscopy.

Preprocessing method	Description	Reference
Raw spectra	NIR spectra data without preprocessing	-
SNV	The data was normalized using the Standard Normal Variate (SNV) method to remove multiplicative variations in spectroscopic data. For each variable (spectrum) it was calculated the mean (μ) and standard	MARTENS and STARK (1991)

	deviation (σ), then it was applied the formula: $Xsnv = \frac{x-\mu}{\sigma}$.	
SNV_1D	After the SNV normalization, the data was submitted to the first derivative to calculate how the intensity of each wavelength varies with respect to the adjacent wavelengths after the SNV normalization. This preprocessing step can enhance spectral features and reduce baseline variations, making identifying subtle differences between spectra easier.	SOREN and NORGAARD (1999)
SNV_2D	This method takes the derivative of the SNV_1D. It was applied to enhance the spectral features further and reduce baseline variations more effectively.	SOREN and NORGAARD (1999)
D1	The first derivative was taken from raw spectra data.	WORKMAN and WEYER (2007)
D2	The second derivative of raw spectra (D1)	WORKMAN and WEYER (2007)
SG	The Savitzky-Golay method was applied to reduce noise and eliminate baseline variations while preserving the spectral features. The method fitted a second-degree polynomial function to 11 adjacent data points and replaced the central data point with the value predicted by the polynomial fit along the whole spectra. This method applied smoothing without differentiation (m=0).	SAVITZKY and GOLAY (1964)

SNV_SG	The SG method was applied to the normalized data using SNV.	CHEN et al. (2009)
SG_D1	The Savitzky-Golay method was initially used to smooth the spectral data using an 11-window size. After smoothing, the gap-segment derivative algorithm was applied to calculate the first derivative, with each segment having 11 widths and a step size of 10 data points between segments.	ZHANG et al. (2020)
SG_D1_W5	The first derivative of Savitzky-Golay with a 5-window-size	SAVITZKY and GOLAY (1964)
SG_D1_W11	This method applied the first derivative of Savitzky-Golay with a second-order polynomial and a window size of 11 data points, performing smoothing and differentiation.	SAVITZKY and GOLAY (1964)
SG_D2_W5	The second derivative of Savitzky-Golay with a 5-window-size	SAVITZKY and GOLAY (1964)
SG_D2_W11	The second derivative of Savitzky-Golay with an 11-window-size	SAVITZKY and GOLAY (1964)

1.3.5. Model evaluation metrics

After preprocessing, 70% of the data was used for training and cross-validation (10 layers), and 30% was used for independent model validation using the Partial Least Squares – Discriminant Analysis (PLS-DA) algorithm. The metrics obtained from the classification results applied to evaluate the methods are described in Table 2.

Table 2. Evaluation metrics extracted from the classification models using PLS-DA.

Metric	Description	Equation
Accuracy	Accuracy measures the overall correctness of the classification model.	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
Kappa	Kappa statistic measures the agreement between observed and predicted classifications while accounting for the possibility of agreement occurring by chance.	$Kappa = \frac{2 * (TP * Tn - FN * FP)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)}$
Sensitivity	Sensitivity, also known as the true positive rate, measures the proportion of actual positives correctly identified by the model.	$Sensitivity = \frac{TP}{TP + FN}$
Specificity	Specificity, also known as the true negative rate, measures the proportion of actual negatives correctly identified by the model.	$Specificity = \frac{TN}{TN + FP}$
Balanced accuracy	Balanced accuracy provides a balanced measure of sensitivity and specificity, considering the imbalance in class distribution.	$Balanced\ accuracy = \frac{Sensitivity + Specificity}{2}$
Precision	Precision measures the proportion of positive predictions that are correct.	$Precision = \frac{TP}{TP + FP}$

Negative Predictive Value (NPV)	NPV measures the proportion of negative predictions that are correct.	$NPV = \frac{TN}{TN + FN}$
Prevalence	The proportion of positive instances in the dataset.	$Prevalence = \frac{TP + FN}{TP + TN + FP + FN}$
Detection prevalence	The proportion of instances classified as positive by the model.	$Detection\ prevalence = \frac{TP + FP}{TP + TN + FP + FN}$

Where *TP, TN, FP, and FN* correspond to a true positive, true negative, false positive, and false negative, respectively.

1.3.6. Non-Information rate

To establish a baseline accuracy for the classification problem, we calculated the Non-Information Rate. It represents the accuracy achieved by always predicting the majority class in the dataset and it was calculated according to the formula:

$$Non - information\ Rate = \frac{Number\ of\ instances\ in\ majority\ class}{Total\ number\ of\ instances}$$

The majority class consisted of normal seedlings, high vigor, viable, and non-infected seeds for germination, growth, emergence, and blotter tests, respectively.

1.3.7. McNemar Test

The performance of the classification models was compared using the McNemar test. The McNemar test statistic was calculated using the formula:

$$X^2 = \frac{(b-c)^2}{b+c}$$

where b represents the number of discordant pairs, and c represents the number of concordant pairs. We set the significance level (α) at 0.05. Based on the calculated test statistic, we compared it to a chi-squared distribution with one degree of freedom to determine statistical significance of the model.

1.3.8. Regression coefficients plots

The regression coefficients from the most accurate models in each seed analysis test — evaluated through cross-validation and balanced accuracy on the test dataset while considering the significance of the non-information rate — were plotted. These plots were then used to visually identify the regions of interest that most significantly contributed to class separation.

1.3.9. Statistical analysis

Statistical analysis and model evaluation were conducted using packages such as *caret*, *pls*, and *e1071* of the R software (R Core Team, 2024) and Python.

1.4. Results

In the present study, the optimal models for discriminating between classes of seeds from different seed quality tests were identified based on near-infrared spectral data. The best-performing models for each class were determined through rigorous analysis of various preprocessing methods. Considering the imbalance of the datasets, the results were analyzed using balanced accuracy, which provides an average of specificity and sensitivity for the testing results.

1.4.1. Germination test

The models better-classified seeds based on final germination results for the two different classes (normal and dead seeds) (Figure 1). For BR18, the highest cross-validation accuracy and kappa were obtained with SNV model, reaching 89.3%, and 75%, respectively. Among testing results, the RAW, SG, and the SG_D1_W11 methods were the most effective for the normal class, achieving a balanced accuracy of 86%. For the abnormal and dead seeds classes, the models RAW and SG stood out, achieving a balanced accuracy of 68% and 85%, respectively (Table 3).

For Guamirim, the highest cross-validation accuracy and kappa were obtained with the SG_D1 model, reaching 86% and 67%, respectively. Among testing results, the SNV, SG, and SNV_SG methods were the most effective, achieving a balanced accuracy of 83%, and 87% for the normal, and dead seeds classes, respectively. The model SNV2D was the most effective for the abnormal class, reaching 51% of balanced accuracy (Table 4).

Figure 1 - Confusion matrix of BR18 wheat cultivar from different models based on final seed germination results.

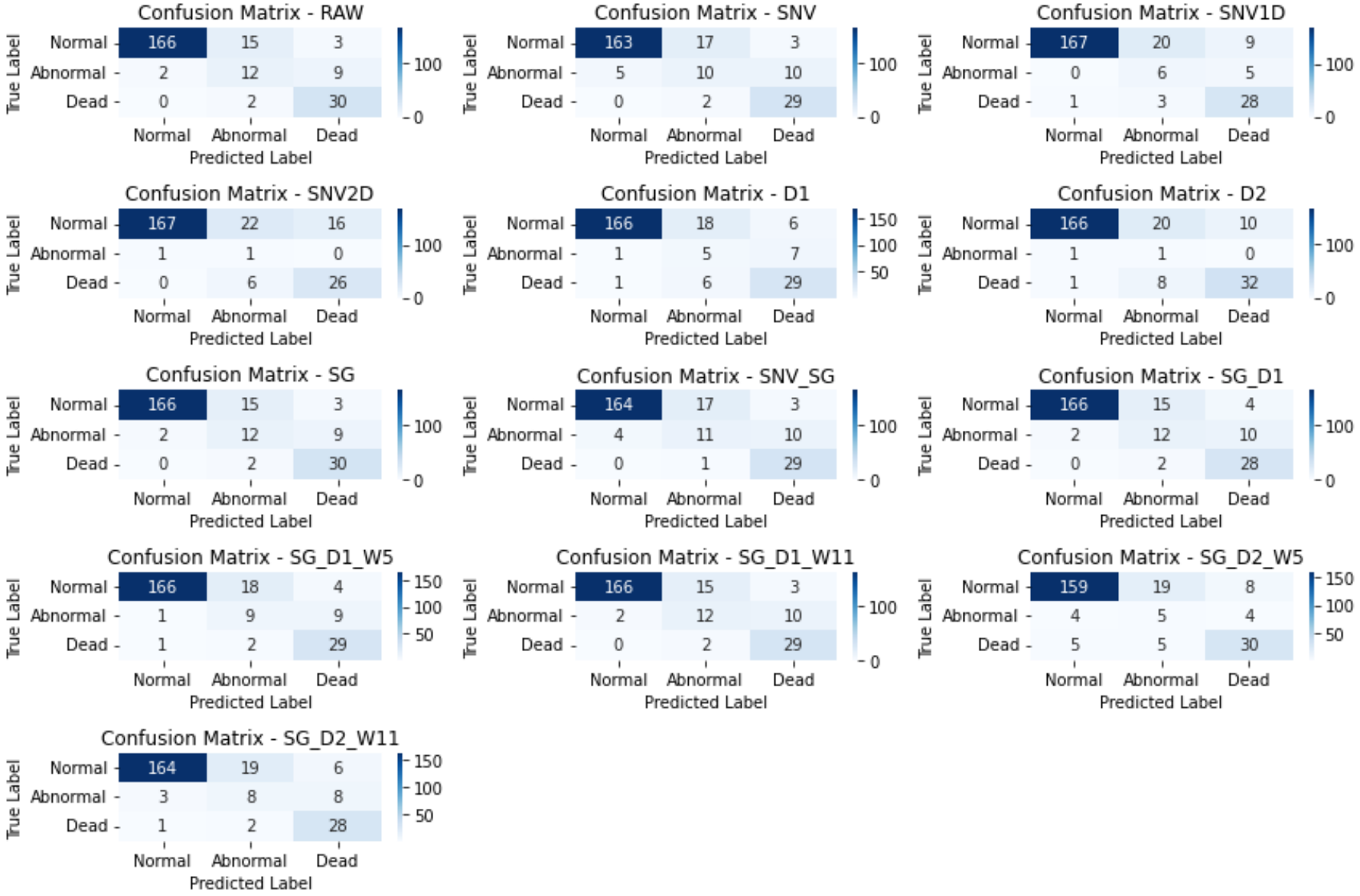


Table 3 - Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the germination test of the BR18 wheat seed cultivar.

Metrics		Preprocessing method												
		Raw	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy		0.8900	0.8930	0.8502	0.8230	0.8569	0.8081	0.8901	0.8907	0.8901	0.8752	0.8847	0.8129	0.8360
Cross-validated Kappa		0.7427	0.7513	0.6238	0.5390	0.6502	0.5169	0.7428	0.7441	0.7416	0.6974	0.7282	0.5474	0.5855
Overall Accuracy		0.8703	0.8452	0.8410	0.8699	0.8368	0.8326	0.8703	0.8536	0.8619	0.8536	0.8661	0.8117	0.8368
Overall Kappa		0.6938	0.6368	0.5969	0.6840	0.6002	0.5735	0.6938	0.6547	0.6725	0.6462	0.6841	0.5484	0.6036
Sensitivity	Normal	0.9881	0.9702	0.9940	0.9940	0.9881	0.9881	0.9881	0.9762	0.9881	0.9881	0.9881	0.9464	0.9762
	Abnormal	0.4138	0.3448	0.2069	0.0345	0.1724	0.0345	0.4138	0.3793	0.4138	0.3103	0.4138	0.1724	0.2759
	Dead	0.7143	0.6905	0.6667	0.6190	0.6905	0.7619	0.7143	0.6905	0.6667	0.6905	0.6905	0.7143	0.6667
Specificity	Normal	0.7465	0.7183	0.5915	0.4648	0.6620	0.5775	0.7465	0.7183	0.7324	0.6901	0.7465	0.6197	0.6479
	Abnormal	0.9476	0.9286	0.9762	0.9952	0.9619	0.9952	0.9476	0.9333	0.9429	0.9524	0.9429	0.9619	0.9476
	Dead	0.9898	0.9898	0.9797	0.9695	0.9645	0.9543	0.9898	0.9949	0.9898	0.9848	0.9898	0.9492	0.9848
Precision	Normal	0.9022	0.8907	0.8520	0.8146	0.8737	0.8469	0.9022	0.8913	0.8973	0.8830	0.9022	0.8548	0.8677
	Abnormal	0.5217	0.4000	0.5455	0.5000	0.3846	0.5000	0.5217	0.4400	0.5000	0.4737	0.5000	0.3846	0.4211
	Dead	0.9375	0.9355	0.8750	0.8125	0.8056	0.7805	0.9375	0.9667	0.9333	0.9062	0.9355	0.7500	0.9032
Neg Pred Value	Normal	0.9636	0.9107	0.9767	0.9706	0.9592	0.9535	0.9636	0.9273	0.9630	0.9608	0.9636	0.8302	0.9200
	Abnormal	0.9213	0.9112	0.8991	0.8819	0.8938	0.8819	0.9213	0.9159	0.9209	0.9091	0.9209	0.8938	0.9046
	Dead	0.9420	0.9375	0.9324	0.9227	0.9360	0.9495	0.9420	0.9378	0.9330	0.9372	0.9375	0.9397	0.9327
Prevalence	Normal	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029	0.7029
	Abnormal	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213	0.1213
	Dead	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757	0.1757
Detection Prevalence	Normal	0.7699	0.7657	0.8201	0.8577	0.7950	0.8201	0.7699	0.7699	0.7741	0.7866	0.7699	0.7782	0.7908
	Abnormal	0.0962	0.1046	0.0460	0.0084	0.0544	0.0084	0.0962	0.1046	0.1004	0.0795	0.1004	0.0544	0.0795
	Dead	0.1339	0.1297	0.1339	0.1339	0.1506	0.1715	0.1339	0.1255	0.1255	0.1339	0.1297	0.1674	0.1297
Balanced Accuracy	Normal	0.8673	0.8443	0.7928	0.7294	0.8250	0.7828	0.8673	0.8473	0.8602	0.8391	0.8673	0.7831	0.8120
	Abnormal	0.6807	0.6367	0.5915	0.5149	0.5672	0.5149	0.6807	0.6563	0.6783	0.6314	0.6783	0.5672	0.6117
	Dead	0.8521	0.8402	0.8232	0.7943	0.8275	0.8581	0.8521	0.8427	0.8283	0.8376	0.8402	0.8318	0.8257

Figure 2 - Confusion matrix of Guamirim wheat cultivar from different models based on final seed germination results

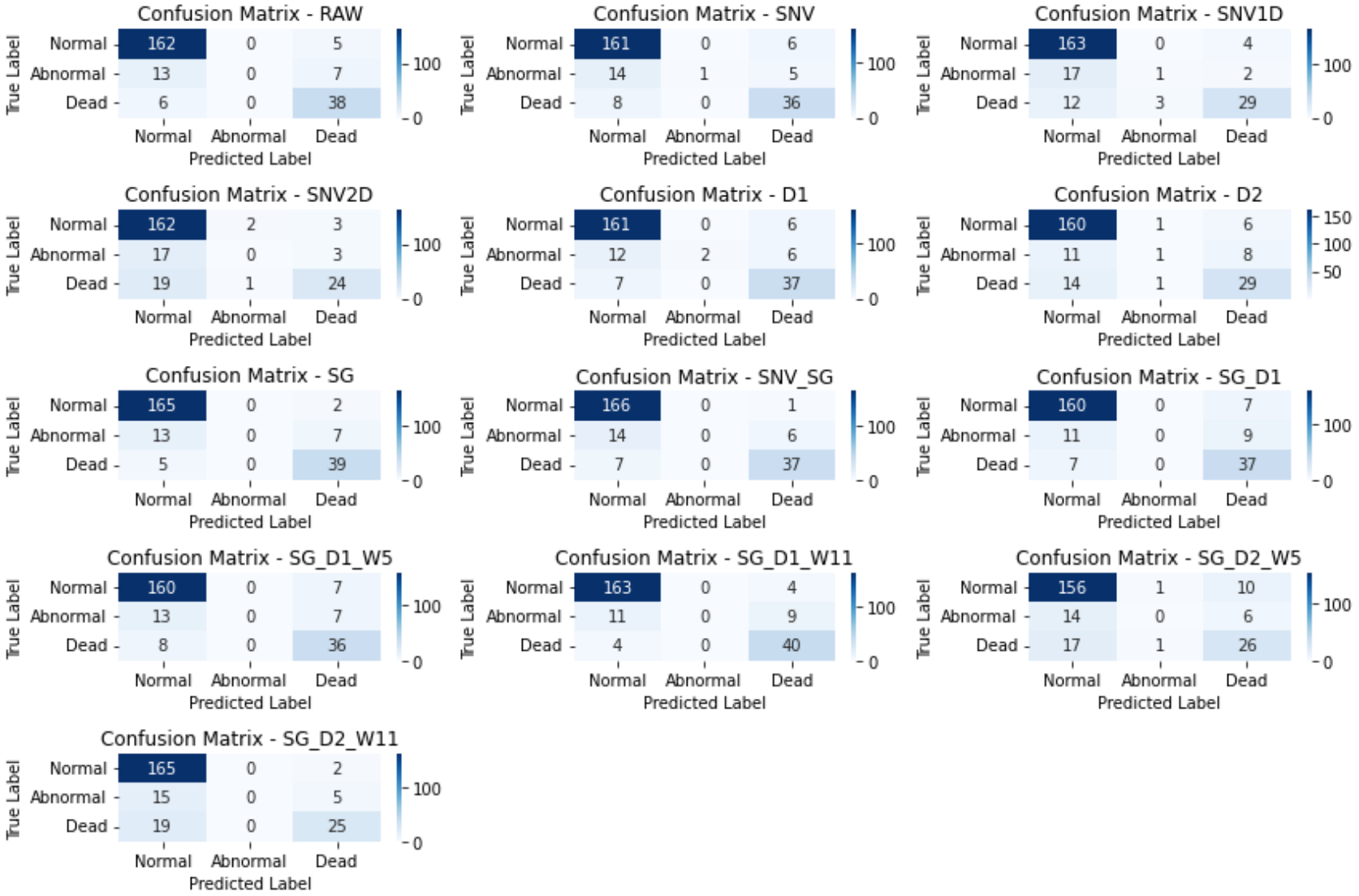
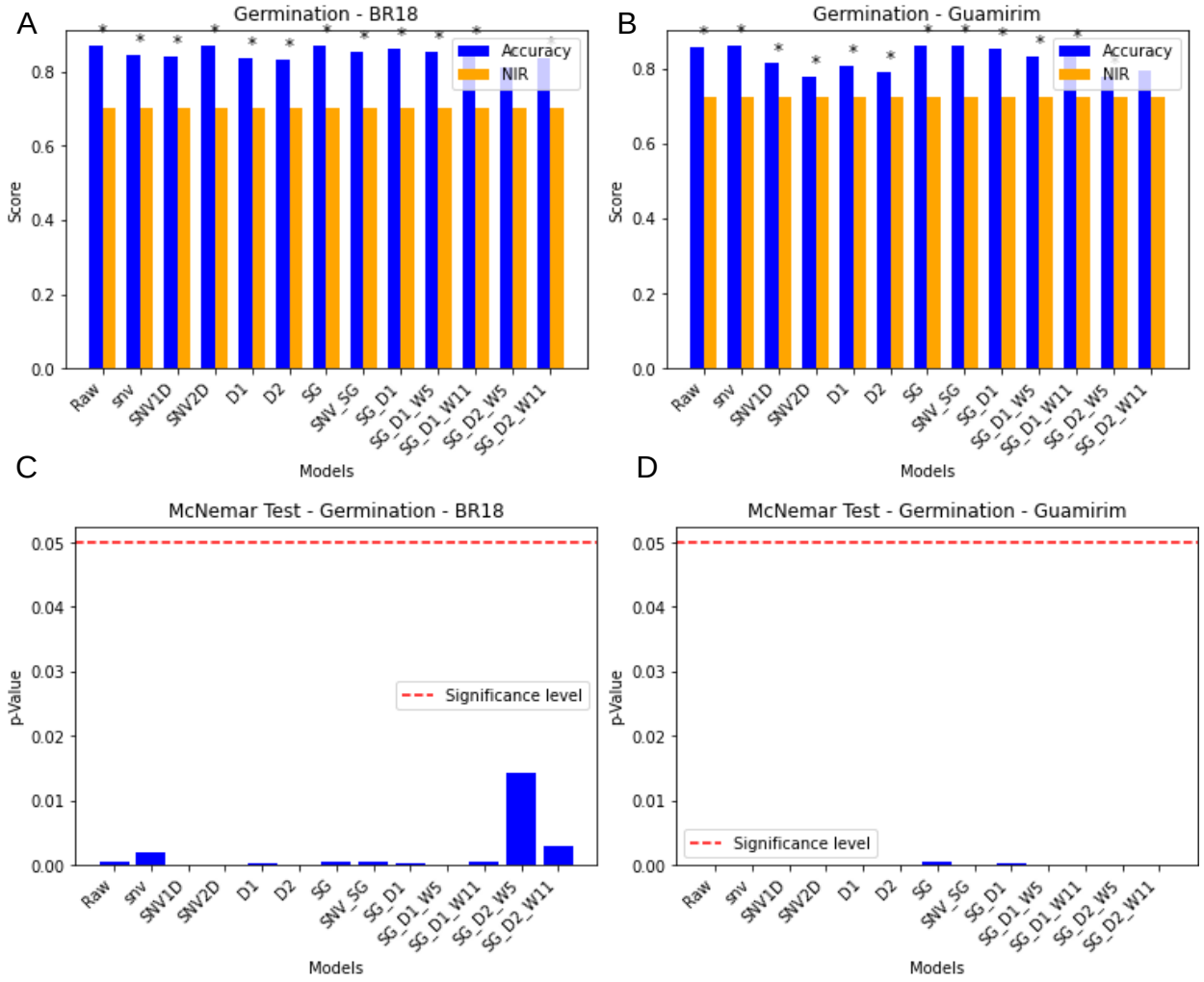


Table 4. Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the germination test of the Guamirim wheat seed cultivar.

Metrics		Preprocessing method												
		Raw	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy		0.8627	0.8598	0.8425	0.8309	0.8474	0.8241	0.8628	0.8585	0.8646	0.8609	0.8640	0.8174	0.8345
Cross-validated Kappa		0.6514	0.6407	0.5877	0.5542	0.5945	0.5392	0.6516	0.6377	0.6590	0.6524	0.6569	0.4982	0.5482
Overall Accuracy		0.8571	0.8615	0.8139	0.7792	0.8095	0.7922	0.8615	0.8615	0.8528	0.8312	0.8485	0.7792	0.7965
Overall Kappa		0.6271	0.6406	0.4933	0.3820	0.4641	0.4191	0.6406	0.6406	0.6111	0.5462	0.5997	0.3537	0.4084
Sensitivity	Normal	0.9820	0.9820	0.9760	0.9641	0.9820	0.9701	0.9820	0.9820	0.9820	0.9820	0.9820	0.9760	0.9880
	Abnormal	0.0000	0.0000	0.0000	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Dead	0.7727	0.7955	0.5682	0.4091	0.5227	0.4773	0.7955	0.7955	0.7500	0.6364	0.7273	0.3864	0.4318
Specificity	Normal	0.6719	0.6875	0.5469	0.4375	0.4844	0.4688	0.6875	0.6875	0.6406	0.5938	0.6406	0.3750	0.4219
	Abnormal	1.0000	1.0000	0.9953	0.9858	1.0000	1.0000	1.0000	1.0000	1.0000	0.9953	1.0000	1.0000	1.0000
	Dead	0.9358	0.9358	0.9305	0.9358	0.9412	0.9251	0.9358	0.9358	0.9412	0.9358	0.9358	0.9412	0.9465
Precision	Normal	0.8865	0.8913	0.8490	0.8173	0.8325	0.8265	0.8913	0.8913	0.8770	0.8632	0.8770	0.8030	0.8168
	Abnormal	NaN	NaN	0.0000	0.2500	NaN	NaN	NaN	NaN	NaN	0.0000	NaN	NaN	NaN
	Dead	0.7391	0.7447	0.6579	0.6000	0.6765	0.6000	0.7447	0.7447	0.7500	0.7000	0.7273	0.6071	0.6552
Neg Pred Value	Normal	0.9348	0.9362	0.8974	0.8235	0.9118	0.8571	0.9362	0.9362	0.9318	0.9268	0.9318	0.8571	0.9310
	Abnormal	0.9134	0.9134	0.9130	0.9163	0.9134	0.9134	0.9134	0.9134	0.9134	0.9130	0.9134	0.9134	0.9134
	Dead	0.9459	0.9511	0.9016	0.8707	0.8934	0.8827	0.9511	0.9511	0.9412	0.9162	0.9358	0.8670	0.8762
Prevalence	Normal	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229	0.7229
	Abnormal	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866
	Dead	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905	0.1905
Detection Prevalence	Normal	0.8009	0.7965	0.8312	0.8528	0.8528	0.8485	0.7965	0.7965	0.8095	0.8225	0.8095	0.8788	0.8745
	Abnormal	0.0000	0.0000	0.0043	0.0173	0.0000	0.0000	0.0000	0.0000	0.0000	0.0043	0.0000	0.0000	0.0000
	Dead	0.1991	0.2035	0.1645	0.1299	0.1472	0.1515	0.2035	0.2035	0.1905	0.1732	0.1905	0.1212	0.1255
Balanced Accuracy	Normal	0.8270	0.8348	0.7615	0.7008	0.7332	0.7194	0.8348	0.8348	0.8113	0.7879	0.8113	0.6755	0.7049
	Abnormal	0.5000	0.5000	0.4976	0.5179	0.5000	0.5000	0.5000	0.5000	0.5000	0.4976	0.5000	0.5000	0.5000
	Dead	0.8543	0.8656	0.7493	0.6725	0.7320	0.7012	0.8656	0.8656	0.8456	0.7861	0.8316	0.6638	0.6892

The results of the non-information rate, which represents the accuracy that could be achieved by simply predicting the majority class without using any predictive spectra feature, from the tested models are shown in Figure 3. In this case, the majority class is represented by the normal class. For both cultivars, all the models had a significantly different accuracy from the test (Figures 3A and 3B). This indicates that the models capture some patterns in the minority class, such as abnormal and dead seeds. Figures 3C and 3D show the MCNemar test for germination results of the BR18 cultivar and Guamirim, respectively. The results showed that all models exhibited statistically significant differences ($p < 0.05$) compared to the ground truth, as indicated by p-values below the significance threshold (red dotted line). These findings highlight systematic differences between model predictions and actual classes, suggesting the need for further refinement to improve classification accuracy.

Figure 3 - Statistical Results Comparison for Cultivars BR18 and Guamirim based on germination test results. A) non-information rate for BR18; B) non-information rate for Guamirim; C) McNemar test for BR18; D) Mcnemar test for Guamirim.



(*) means that the model outperforms the non-information rate.

1.4.2. Growth test

Figure 4 and Table 5 for BR18 and Figure 5 and Table 6 for the Guamirim cultivar show the results of 12 preprocessing methods to classify seeds according to the near-infrared data and growth test results. The models did not perform well using these approaches, with a high error rate between high-vigor and low-vigor seeds. The results

of the models using different cultivars show variations in classification performance across different preprocessing methods and seed classes.

For BR18, after cross-validation, the best performance was obtained with the models SNV, SNV_1D, D1, and SNV_SG, achieving 68% accuracy and the model D1 achieving 45% kappa. The performance of the models on the testing dataset varied across different preprocessing methods and seed classes. The SNV, SNV_2D, SNV_SG, and SG_D1 methods for high-vigor seeds showcased the highest balanced accuracy of 75%. The low-vigor seeds were most accurately classified using the SG_D1 preprocessing method, yielding a balanced accuracy of 61% The RAW and the SG preprocessing method demonstrated the highest balanced accuracy of 80% for dead seeds (Table 5).

For Guamirim, the methods of RAW, SG, and SG_D1 reached high cross-validated accuracy with 66%, and the method SG_D1 reached a cross-validated kappa of 40%. For test results, the SNV and SNV_SG models reached the highest balanced accuracy of 76% for the high-vigor class, and the model SNV_2D stood out with 58% for the low-vigor class. The RAW and SG achieved the highest balanced accuracy of 95% for the dead seeds class, indicating robust classification performance. (Table 6).

Figure 4 - Confusion matrix of BR18 wheat cultivar from different models based on growth test results.

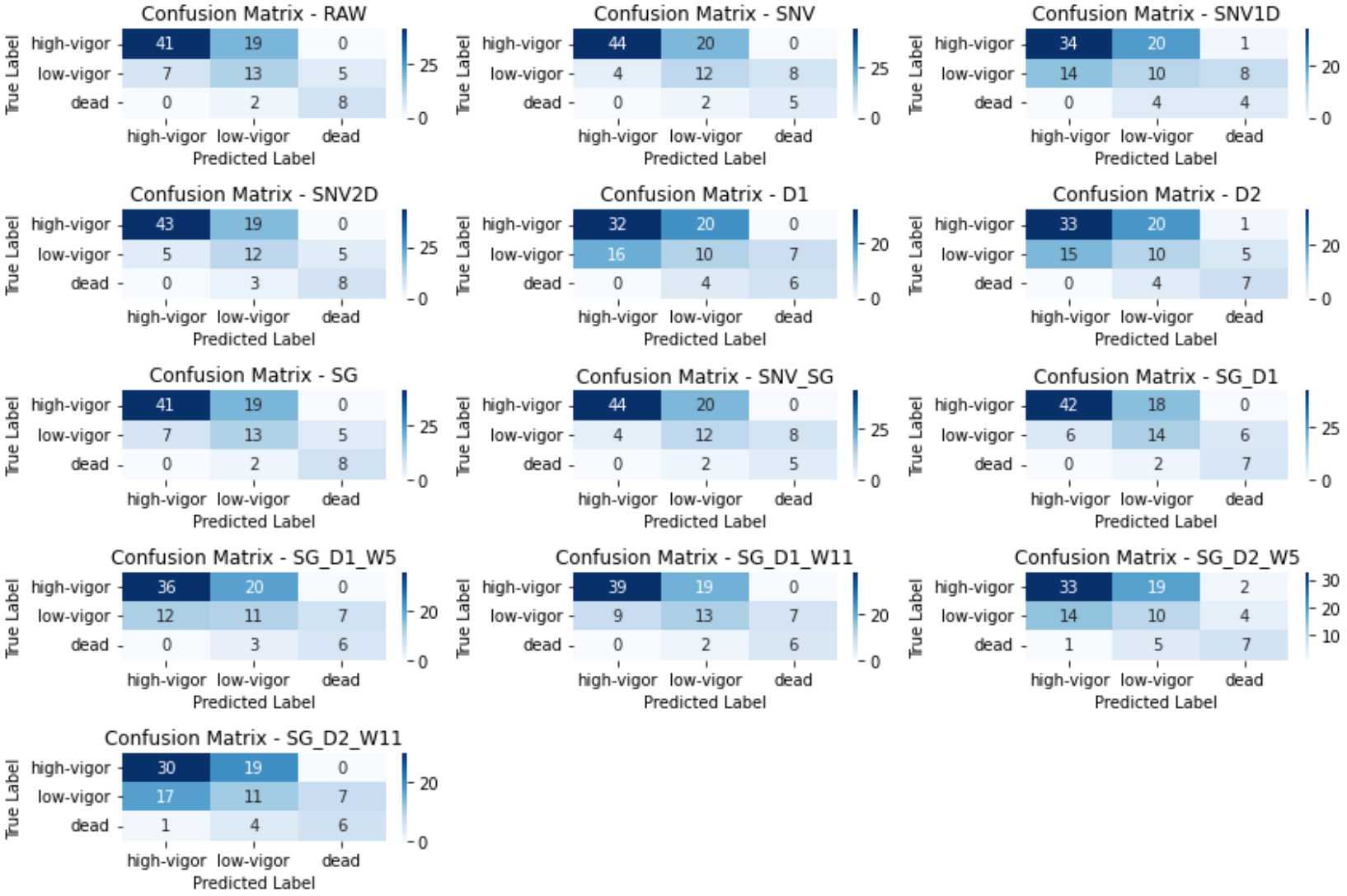


Table 5. Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the growth test of the BR18 wheat seed cultivar.

Metrics		Preprocessing method												
		Raw	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy		0.6613	0.6790	0.6767	0.6428	0.6788	0.5984	0.6611	0.6759	0.6399	0.6541	0.6360	0.6144	0.6404
Cross-validated Kappa		0.4109	0.4327	0.4409	0.3656	0.4504	0.3076	0.4093	0.4271	0.3785	0.4039	0.3726	0.3461	0.3920
Overall Accuracy		0.6526	0.6421	0.5053	0.6632	0.5053	0.5263	0.6526	0.6421	0.6632	0.5579	0.6105	0.5263	0.4947
Overall Kappa		0.3930	0.3599	0.1402	0.4106	0.1539	0.1888	0.3930	0.3599	0.4090	0.2327	0.3176	0.1952	0.1461
Sensitivity	High-vigor	0.8542	0.9167	0.7083	0.8958	0.6667	0.6875	0.8542	0.9167	0.8750	0.7500	0.8125	0.6875	0.6250
	Low-vigor	0.3824	0.3529	0.2941	0.3529	0.2941	0.2941	0.3824	0.3529	0.4118	0.3235	0.3824	0.2941	0.3235
	Dead seeds	0.6154	0.3846	0.3077	0.6154	0.4615	0.5385	0.6154	0.3846	0.5385	0.4615	0.4615	0.5385	0.4615
Specificity	High-vigor	0.5957	0.5745	0.5532	0.5957	0.5745	0.5532	0.5957	0.5745	0.6170	0.5745	0.5957	0.5532	0.5957
	Low-vigor	0.8033	0.8033	0.6393	0.8361	0.6230	0.6721	0.8033	0.8033	0.8033	0.6885	0.7377	0.7049	0.6066
	Dead seeds	0.9756	0.9756	0.9512	0.9634	0.9512	0.9512	0.9756	0.9756	0.9756	0.9634	0.9756	0.9268	0.9390
Precision	High-vigor	0.6833	0.6875	0.6182	0.6935	0.6154	0.6111	0.6833	0.6875	0.7000	0.6429	0.6724	0.6111	0.6122
	Low-vigor	0.5200	0.5000	0.3125	0.5455	0.3030	0.3333	0.5200	0.5000	0.5385	0.3667	0.4483	0.3571	0.3143
	Dead seeds	0.8000	0.7143	0.5000	0.7273	0.6000	0.6364	0.8000	0.7143	0.7778	0.6667	0.7500	0.5385	0.5455
Neg Pred Value	High-vigor	0.8000	0.8710	0.6500	0.8485	0.6279	0.6341	0.8000	0.8710	0.8286	0.6923	0.7568	0.6341	0.6087
	Low-vigor	0.7000	0.6901	0.6190	0.6986	0.6129	0.6308	0.7000	0.6901	0.7101	0.6462	0.6818	0.6418	0.6167
	Dead seeds	0.9412	0.9091	0.8966	0.9405	0.9177	0.9286	0.9412	0.9091	0.9302	0.9186	0.9195	0.9268	0.9167
Prevalence	High-vigor	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053
	Low-vigor	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579	0.3579
	Dead seeds	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368	0.1368
Detection Prevalence	High-vigor	0.6316	0.6737	0.5789	0.6526	0.5474	0.5684	0.6316	0.6737	0.6316	0.5895	0.6105	0.5684	0.5158
	Low-vigor	0.2632	0.2526	0.3368	0.2316	0.3474	0.3158	0.2632	0.2526	0.2737	0.3158	0.3053	0.2947	0.3684
	Dead seeds	0.1053	0.0737	0.0842	0.1158	0.1053	0.1158	0.1053	0.0737	0.0947	0.0947	0.0842	0.1368	0.1158
Balanced Accuracy	High-vigor	0.7250	0.7456	0.6308	0.7458	0.6206	0.6203	0.7250	0.7456	0.7460	0.6622	0.7041	0.6203	0.6104
	Low-vigor	0.5928	0.5781	0.4667	0.5945	0.4585	0.4831	0.5928	0.5781	0.6075	0.5060	0.5600	0.4995	0.4650
	Dead seeds	0.7955	0.6801	0.6295	0.7894	0.7064	0.7448	0.7955	0.6801	0.7570	0.7125	0.7186	0.7327	0.7003

Figure 5 - Confusion matrix of Guamirim wheat cultivar from different models based on growth test results.

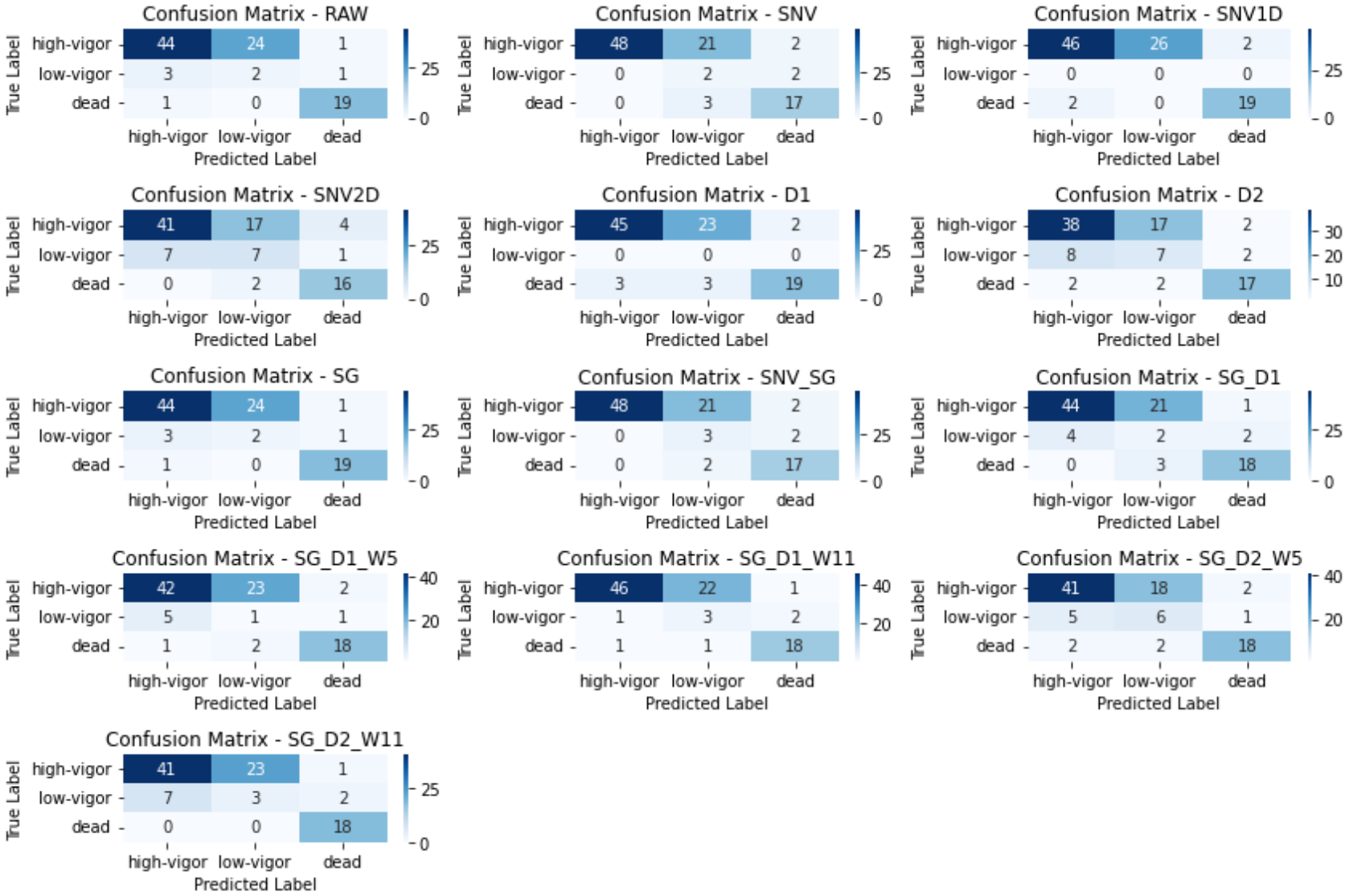
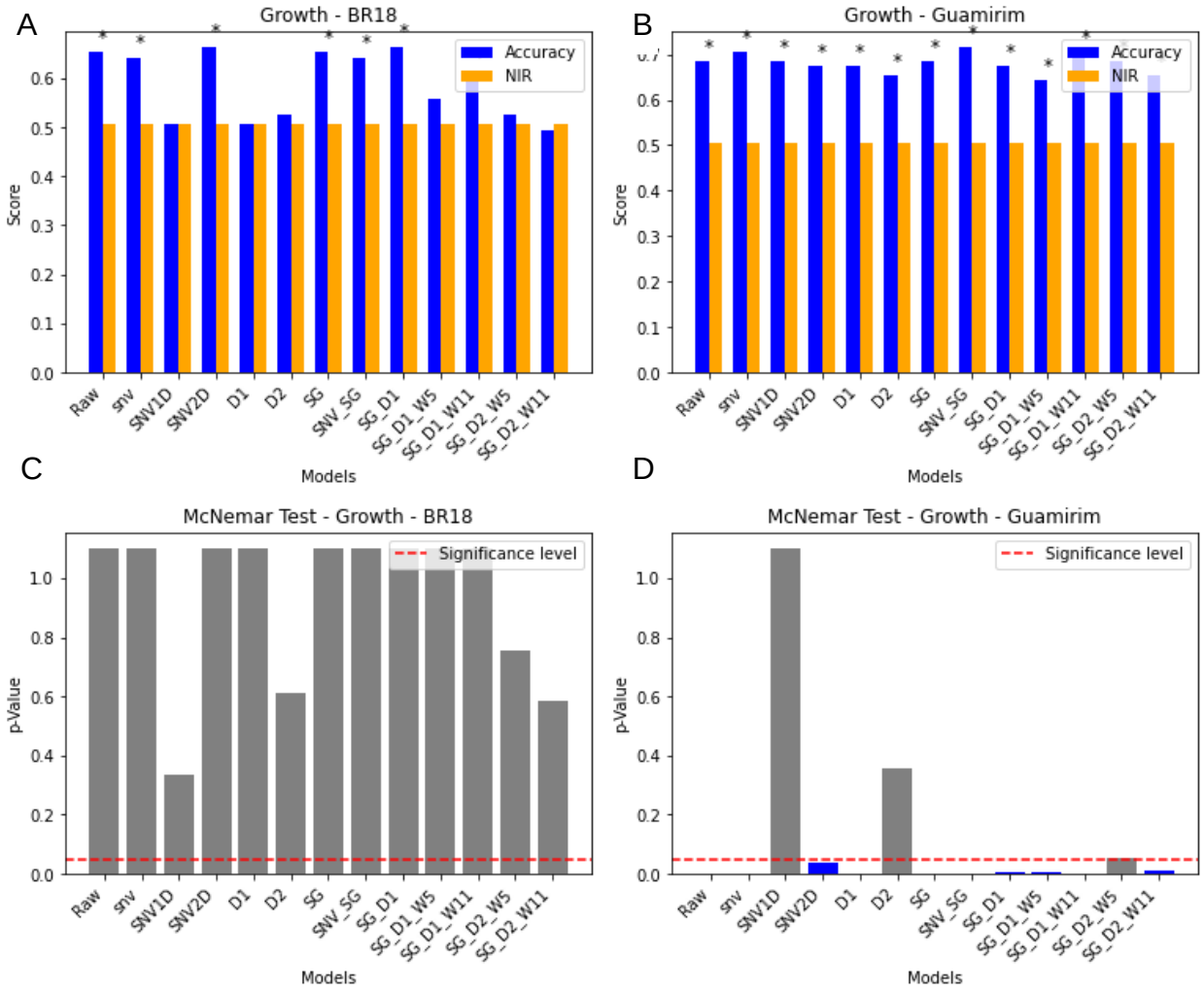


Table 6. Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the growth test of the Guamirim wheat seed cultivar.

Metrics		Preprocessing method												
		Raw	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy		0.6568	0.6536	0.6239	0.6255	0.6270	0.6192	0.6568	0.6535	0.6622	0.6251	0.6427	0.6043	0.5927
Cross-validated Kappa		0.3914	0.3826	0.3123	0.3595	0.3288	0.3697	0.3910	0.3834	0.4044	0.3401	0.3775	0.3375	0.3101
Overall Accuracy		0.6842	0.7053	0.6842	0.6737	0.6737	0.6526	0.6842	0.7158	0.6737	0.6421	0.7053	0.6842	0.6526
Overall Kappa		0.4452	0.4777	0.4336	0.4423	0.4270	0.4201	0.4452	0.4959	0.4345	0.3772	0.4822	0.4646	0.3988
Sensitivity	High-vigor	0.9167	1.0000	0.9583	0.8542	0.9375	0.7917	0.9167	1.0000	0.9167	0.8750	0.9583	0.8542	0.8542
	Low-vigor	0.0769	0.0769	0.0000	0.2692	0.0000	0.2692	0.0769	0.1154	0.0769	0.0385	0.1154	0.2308	0.1154
	Dead seeds	0.9048	0.8095	0.9048	0.7619	0.9048	0.8095	0.9048	0.8095	0.8571	0.8571	0.8571	0.8571	0.8571
Specificity	High-vigor	0.4681	0.5106	0.4043	0.5532	0.4681	0.5957	0.4681	0.5106	0.5319	0.4681	0.5106	0.5745	0.4894
	Low-vigor	0.9420	0.9710	1.0000	0.8841	1.0000	0.8551	0.9420	0.9710	0.9130	0.9130	0.9565	0.9130	0.8696
	Dead seeds	0.9865	0.9595	0.9730	0.9730	0.9189	0.9459	0.9865	0.9730	0.9595	0.9595	0.9730	0.9459	1.0000
Precision	High-vigor	0.6377	0.6761	0.6216	0.6613	0.6429	0.6667	0.6377	0.6761	0.6667	0.6269	0.6667	0.6721	0.6308
	Low-vigor	0.3333	0.5000	NaN	0.4667	NaN	0.4118	0.3333	0.6000	0.2500	0.1429	0.5000	0.5000	0.2500
	Dead seeds	0.9500	0.8500	0.9048	0.8889	0.7600	0.8095	0.9500	0.8947	0.8571	0.8571	0.9000	0.8182	1.0000
Neg Pred Value	High-vigor	0.8462	1.0000	0.9048	0.7879	0.8800	0.7368	0.8462	1.0000	0.8621	0.7857	0.9231	0.7941	0.7667
	Low-vigor	0.7303	0.7363	0.7263	0.7625	0.7263	0.7564	0.7303	0.7444	0.7241	0.7159	0.7416	0.7590	0.7229
	Dead seeds	0.9733	0.9467	0.9730	0.9351	0.9714	0.9459	0.9733	0.9474	0.9595	0.9595	0.9600	0.9589	0.9610
Prevalence	High-vigor	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053	0.5053
	Low-vigor	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737	0.2737
	Dead seeds	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211	0.2211
Detection Prevalence	High-vigor	0.7263	0.7474	0.7789	0.6526	0.7368	0.6000	0.7263	0.7474	0.6947	0.7053	0.7263	0.6421	0.6842
	Low-vigor	0.0632	0.0421	0.0000	0.1579	0.0000	0.1790	0.0632	0.0526	0.0842	0.0737	0.0632	0.1263	0.1263
	Dead seeds	0.2105	0.2105	0.2211	0.1895	0.2632	0.2211	0.2105	0.2000	0.2211	0.2211	0.2105	0.2316	0.1895
Balanced Accuracy	High-vigor	0.6924	0.7553	0.6813	0.7037	0.7028	0.6937	0.6924	0.7553	0.7243	0.6715	0.7345	0.7143	0.6718
	Low-vigor	0.5095	0.5240	0.5000	0.5766	0.5000	0.5622	0.5095	0.5432	0.4950	0.4758	0.5360	0.5719	0.4925
	Dead seeds	0.9456	0.8845	0.9389	0.8674	0.9118	0.8777	0.9456	0.8912	0.9083	0.9083	0.9151	0.9015	0.9286

The results of the non-information rate from the tested models are shown in Figure 6. The models RAW, SNV, SNV_2D, SG, SNV_SG, SG_D1, and SG_D1_W11 for the BR18 cultivar (Figure 6A) and all the models for Guamirim (Figure 6B) were statistically significant, meaning these models outperformed the guests of only the majority class represented by high-vigor seeds. Figures 6 C and D show the McNemar test for growth results of the BR18 cultivar and Guamirim. The models SNV_1D, D2, SG_D2_W5, and SG_D2_W11 for the BR18 cultivar and the D2 and SG_D2_W5 for Guamirim were not significant, indicating that there is no significant difference in the classification errors for the three classes under the conditions tested. Several models returned N/A p-values (represented as exceeding a p-value of 1), which may indicate an inability to detect differences due to insufficient variability in predictions or perfect agreement with the ground truth in certain cases.

Figure 6 - Statistical results comparison for cultivars BR18 and Guamirim based on growth test results. A) non-information rate for BR18; B) non-information rate for Guamirim; C) McNemar test for BR18; D) Mcnemar test for Guamirim.



(*) means that the model outperforms the non-information rate.

1.4.3. Emergence test

The results of the binary classification model using different preprocessing methods reveal variations in performance metrics across the methods employed. From the confusion matrix, the models were more efficient at identifying viable seeds than non-viable ones on either BR18 (Figure 7) or Guamirim (Figure 8) cultivars. For BR18, RAW, SNV, SG, SNV_SG, SG_D1, and SG_D1_W11 reached 90% of cross-validated accuracy and the SNV method stood out for cross-validated kappa reaching 74%. The balanced accuracy, which considers both sensitivity and specificity, reaches the highest value (83%) using the RAW, the SNV, SG, SNV_SG, SG_D1, SG_D1_W5, and SG_D2_W5 preprocessed methods (Table 7). For Guamirim, the RAW, SNV, SG, SNV_SG, and SG_D1 methods reached the highest accuracy (91%) and Kappa (74%) from the cross-validation technique. However, the best model for the test dataset was obtained with SG_D1, reaching a balanced accuracy of 85% (Table 8).

Figure 7 - Confusion matrix of BR18 wheat cultivar from different models based on emergence test results.

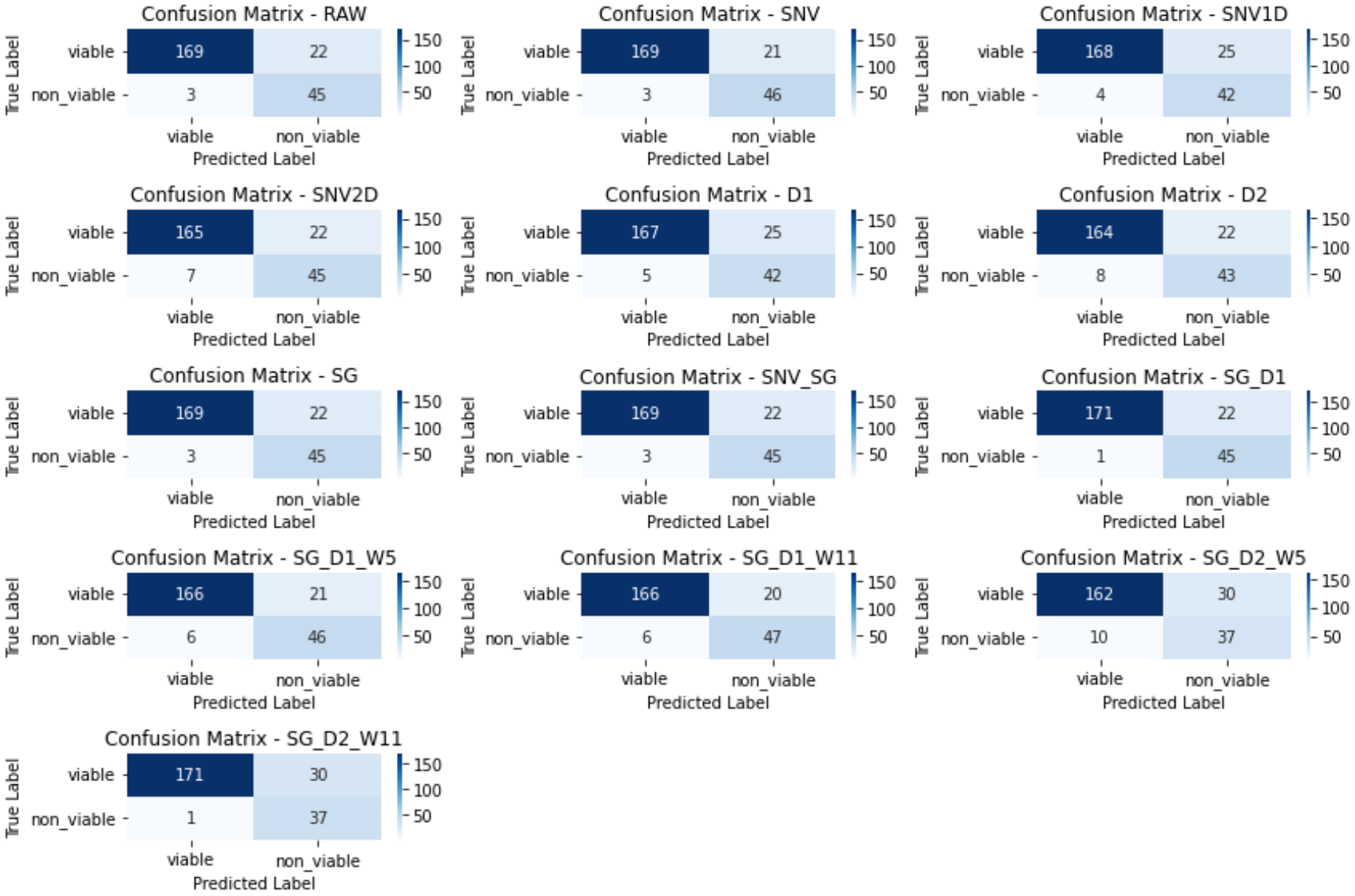


Table 7. Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the Emergence test of the BR18 wheat seed cultivar.

Metrics	**Preprocessing method**												
Non-viable (+) Viable (-)	**Raw**	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy	0.9009	0.9032	0.8740	0.8646	0.8835	0.8628	0.9008	0.9003	0.8979	0.8877	0.8954	0.8439	0.8694
Cross-validated Kappa	0.7349	0.7426	0.6484	0.6253	0.6853	0.6401	0.7347	0.7299	0.7250	0.7015	0.7247	0.5891	0.6348
Accuracy	0.8954	0.8996	0.8787	0.8787	0.8745	0.8661	0.8954	0.8954	0.9038	0.8870	0.8912	0.8326	0.8703
Kappa	0.7162	0.7289	0.6675	0.6772	0.6577	0.6421	0.7162	0.7162	0.7363	0.6995	0.7120	0.5436	0.6296
Sensitivity	0.6716	0.6866	0.6269	0.6716	0.6269	0.6418	0.6716	0.6716	0.6716	0.6866	0.7015	0.5522	0.5522
Specificity	0.9826	0.9826	0.9767	0.9593	0.9709	0.9535	0.9826	0.9826	0.9942	0.9651	0.9651	0.9419	0.9942
Pos Pred Value	0.9375	0.9388	0.9130	0.8654	0.8936	0.8431	0.9375	0.9375	0.9783	0.8846	0.8868	0.7872	0.9737
Neg Pred Value	0.8848	0.8895	0.8705	0.8824	0.8698	0.8723	0.8848	0.8848	0.8860	0.8877	0.8925	0.8438	0.8507
Prevalence	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803
Detection Prevalence	0.2008	0.2050	0.1925	0.2176	0.1967	0.2134	0.2008	0.2008	0.1925	0.2176	0.2218	0.1967	0.1590
Balanced Accuracy	0.8271	0.8346	0.8018	0.8155	0.7989	0.7976	0.8271	0.8271	0.8329	0.8258	0.8333	0.7470	0.7732

Figure 8 - Confusion matrix of Guamirim wheat cultivar from different models based on emergence test results.

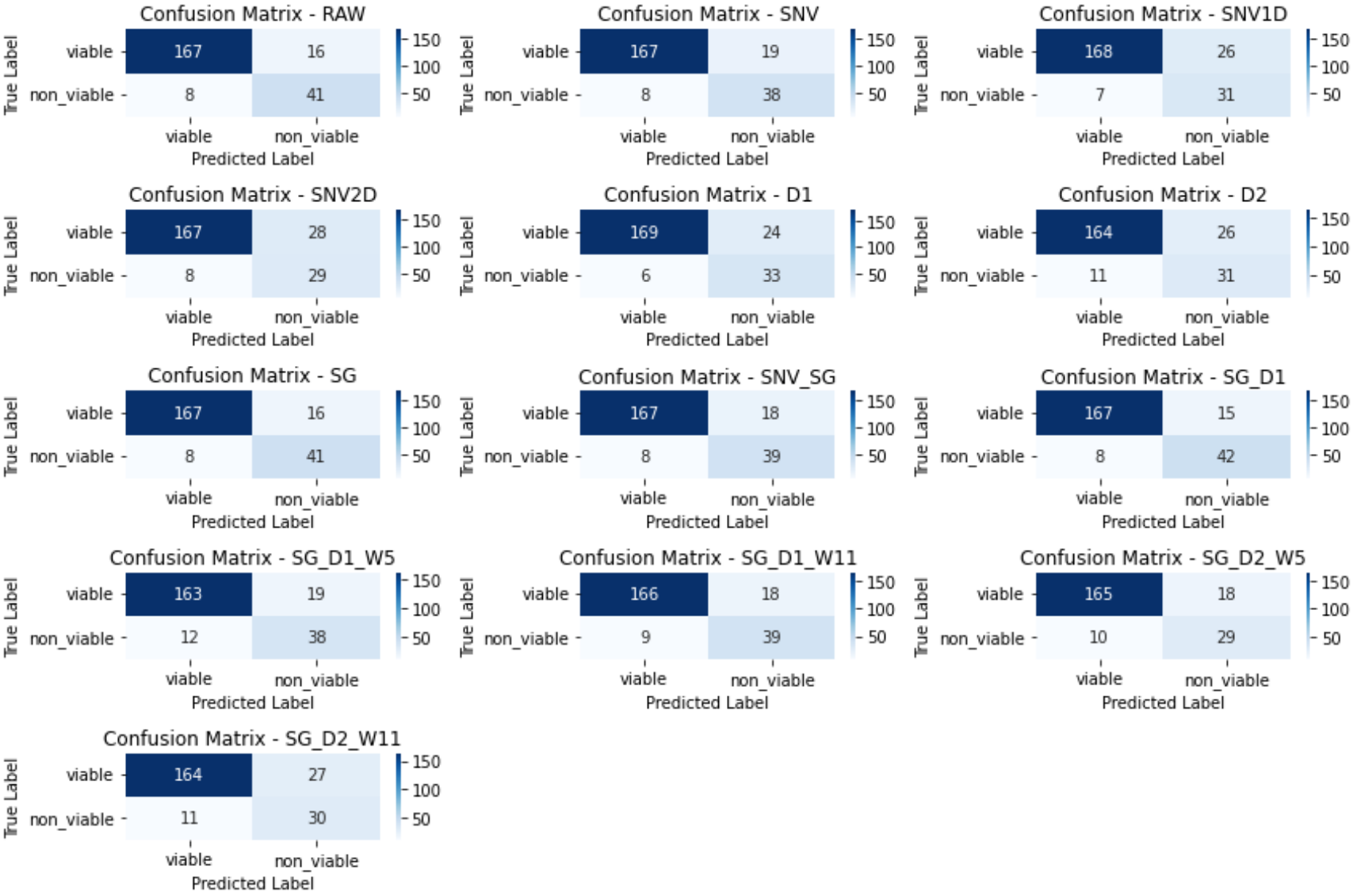
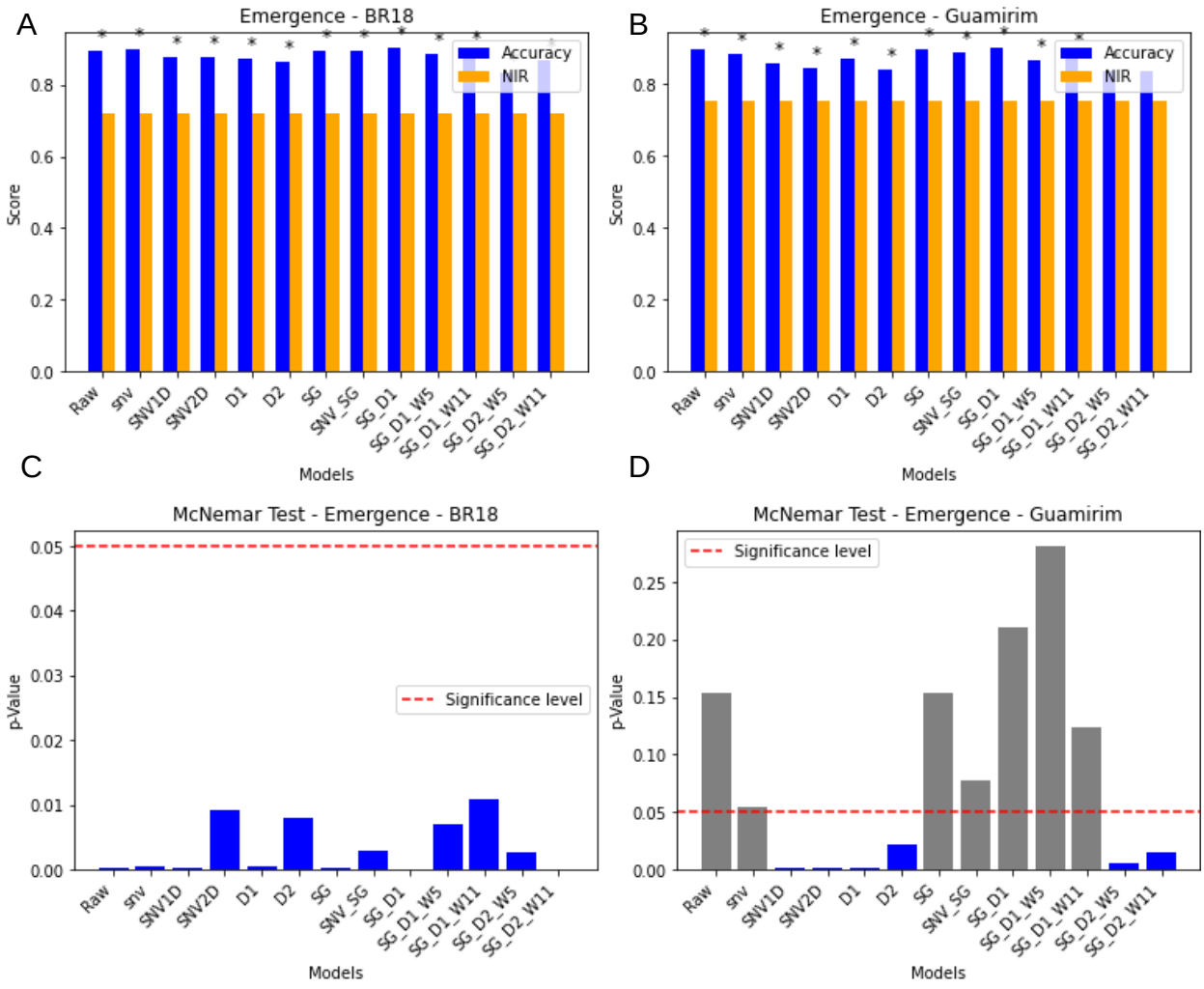


Table 8 - Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the emergence test of the Guamirim wheat seed cultivar.

Metrics	**Preprocessing method**												
Non-viable (+) Viable (-)	**Raw**	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy	0.9055	0.9050	0.8859	0.8601	0.8932	0.8399	0.9061	0.9050	0.9068	0.9019	0.9007	0.8349	0.8576
Cross-validated Kappa	0.7352	0.7350	0.6662	0.5658	0.6895	0.5300	0.7366	0.7353	0.7387	0.7247	0.7229	0.5231	0.5655
Accuracy	0.8966	0.8836	0.8578	0.8448	0.8707	0.8405	0.8966	0.8879	0.9009	0.8664	0.8836	0.8362	0.8362
Kappa	0.7070	0.6642	0.5677	0.5252	0.6096	0.5278	0.7070	0.6786	0.7210	0.6239	0.6684	0.5054	0.5119
Sensitivity	0.7193	0.6667	0.5439	0.5088	0.5789	0.5439	0.7193	0.6842	0.7368	0.6670	0.6842	0.5088	0.5263
Specificity	0.9543	0.9543	0.9600	0.9543	0.9657	0.9371	0.9543	0.9543	0.9543	0.9314	0.9486	0.9429	0.9371
Pos Pred Value	0.8367	0.8261	0.8158	0.7838	0.8462	0.7381	0.8367	0.8298	0.8400	0.7600	0.8125	0.7436	0.7317
Neg Pred Value	0.9126	0.8978	0.8660	0.8564	0.8756	0.8632	0.9126	0.9027	0.9176	0.8956	0.9022	0.8549	0.8586
Prevalence	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457	0.2457
Detection Prevalence	0.2112	0.1983	0.1638	0.1595	0.1681	0.1810	0.2112	0.2026	0.2155	0.2155	0.2069	0.1681	0.1767
Balanced Accuracy	0.8368	0.8105	0.7519	0.7315	0.7723	0.7405	0.8368	0.8192	0.8456	0.7990	0.8164	0.7258	0.7317

The results of the non-information rate from the tested models are shown in Figure 9. The viable class represents the majority class. For the BR18 cultivar (Figure 9A and Guamirim (Figure 9B), all the models of the study had significantly different accuracy (Figure 9a). Figures 9 C and D show the McNemar test for the emergence results of the BR18 cultivar and Guamirim. All the models for the BR18 cultivar and SNV_1D, SNV_2D, D1, D2, SG_D2_W5, and SG_D2_W11, for the Guamirim cultivar were significant, indicating that these models misclassify one class more than another. On the other hand, the models RAW, SNV, SG, SNV_SG, SG_D1, SG_D1_W5, SG_D1_W11 were nonsignificant for the Guamirim cultivar (Figure 9D),

Figure 9 - Statistical results comparison for cultivars BR18 and Guamirim based on emergence test results. A) non-information rate for BR18; B) non-information rate for Guamirim; C) McNemar test for BR18; D) Mcnemar test for Guamirim.



(*) means that the model outperforms the non-information rate.

1.4.4. Blotter test

Figures 10 and 11 and Tables 9 and 10 show the performance of a binary classification model trained on pre-processed datasets from Blotter test results evaluated for two cultivars. This approach considers the challenge of imbalanced data, where the negative class (non-infected seeds) was twice as prevalent as the positive class (infected seeds) (Figures 10 and 11). For BR18, the cross-validated accuracy and Kappa reached 90% and 76%, respectively, using the models of RAW, SNV, SG, SNV_SG, SG_D1, and SG_D1_W11. In the training results, the balanced accuracy reached 90% for SG_D1 (Table 9). For Guamirim, the cross-validated accuracy reached 91% using the model D1. The cross-validated Kappa reached 65%, using the D2 model. The model SG_D1_W5 stood out on testing performance, reaching 70% of balanced accuracy (Table 10).

Figure 10 - Confusion matrix of BR18 wheat cultivar from different models based on Blotter test results.

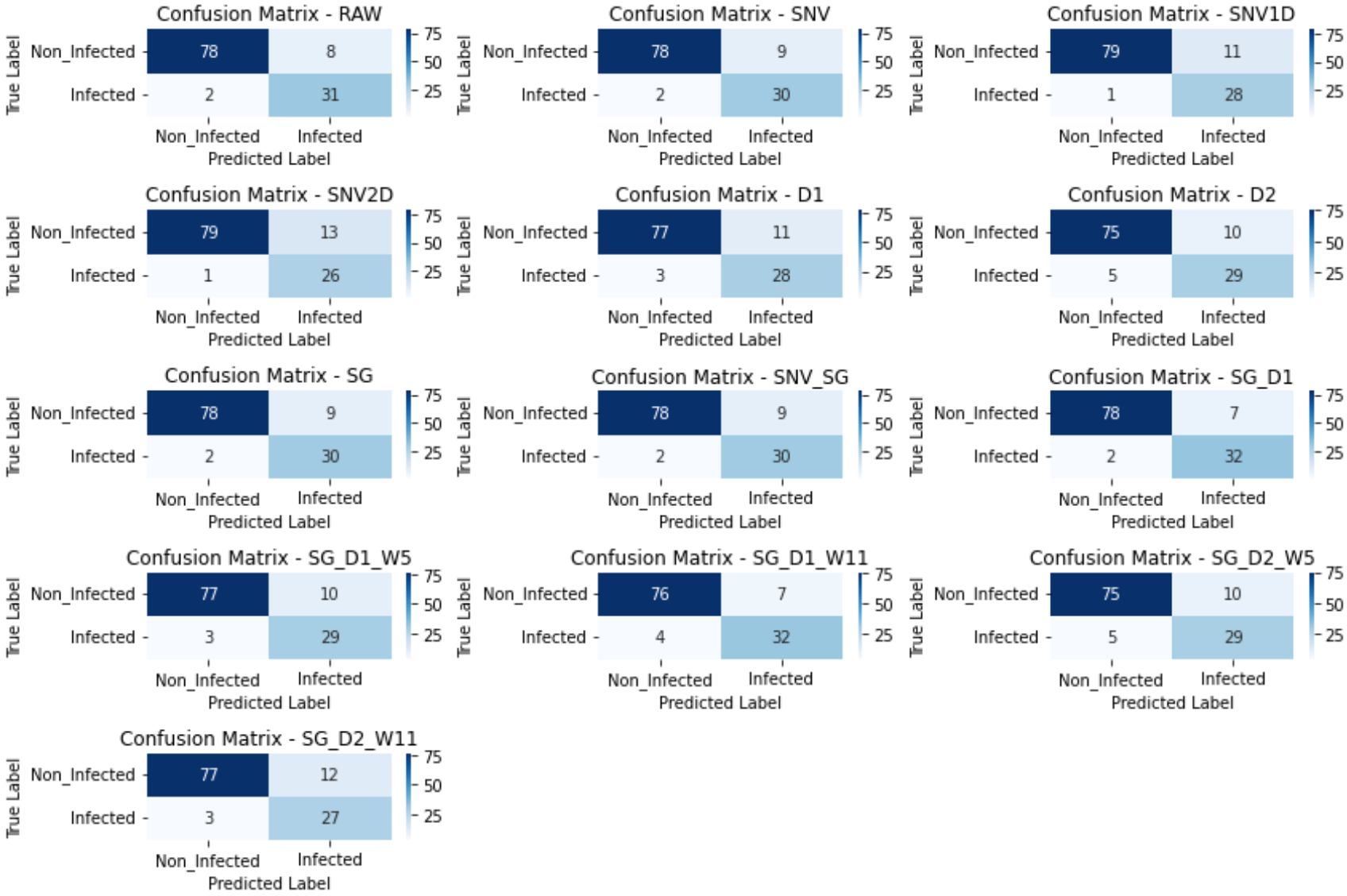


Table 9. Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the blotter test of the BR18 wheat seed cultivar.

Metrics	**Preprocessing method**												
Infected (+) Non-infected (-)	**Raw**	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy	0.9042	0.9042	0.8827	0.8662	0.8709	0.8507	0.9030	0.9030	0.9031	0.8886	0.9031	0.8350	0.8651
Cross-validated Kappa	0.7633	0.7624	0.7089	0.6597	0.6875	0.6432	0.7634	0.7599	0.7637	0.7249	0.7652	0.6000	0.6597
Accuracy	0.9160	0.9076	0.8992	0.8824	0.8824	0.8739	0.9076	0.9076	0.9244	0.8908	0.9076	0.8739	0.8739
Kappa	0.8015	0.7801	0.7551	0.7102	0.7182	0.7042	0.7801	0.7801	0.8225	0.7401	0.7860	0.7042	0.6960
Sensitivity	0.7949	0.7692	0.7179	0.6667	0.7179	0.7436	0.7692	0.7692	0.8205	0.7436	0.8205	0.7436	0.6923
Specificity	0.9750	0.9750	0.9875	0.9875	0.9625	0.9375	0.9750	0.9750	0.9750	0.9625	0.9500	0.9375	0.9625
Pos Pred Value	0.9394	0.9375	0.9655	0.9630	0.9032	0.8529	0.9375	0.9375	0.9412	0.9062	0.8889	0.8529	0.9000
Neg Pred Value	0.9070	0.8966	0.8778	0.8587	0.8750	0.8824	0.8966	0.8966	0.9176	0.8851	0.9157	0.8824	0.8652
Prevalence	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277	0.3277
Detection Prevalence	0.2773	0.2689	0.2437	0.2269	0.2605	0.2857	0.2689	0.2689	0.2857	0.2689	0.3025	0.2857	0.2521
Balanced Accuracy	0.8849	0.8721	0.8527	0.8271	0.8402	0.8405	0.8721	0.8721	0.8978	0.8530	0.8853	0.8405	0.8274

Figure 11 - Confusion matrix of Guamirim wheat cultivar from different models based on Blotter test results.

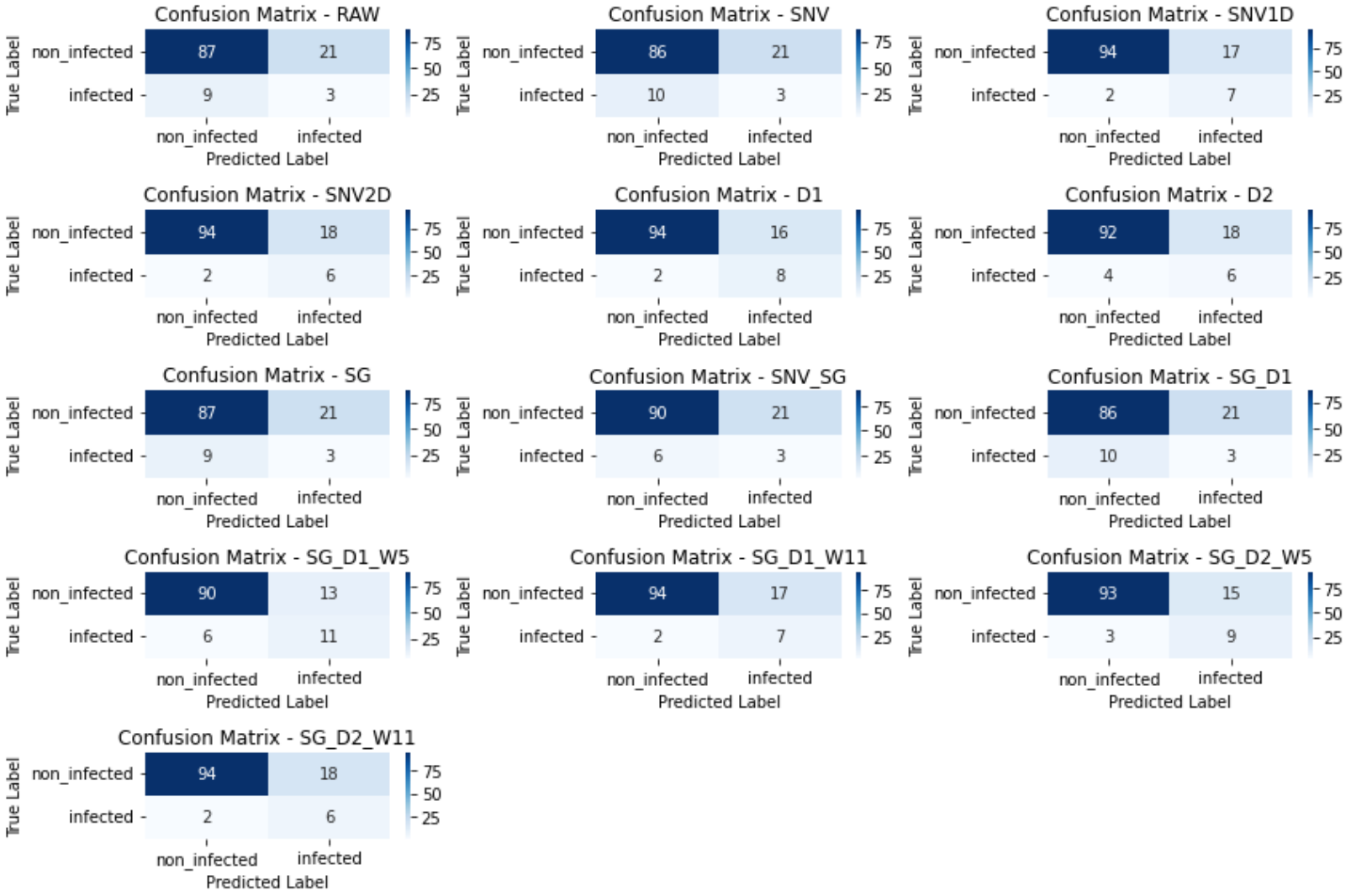
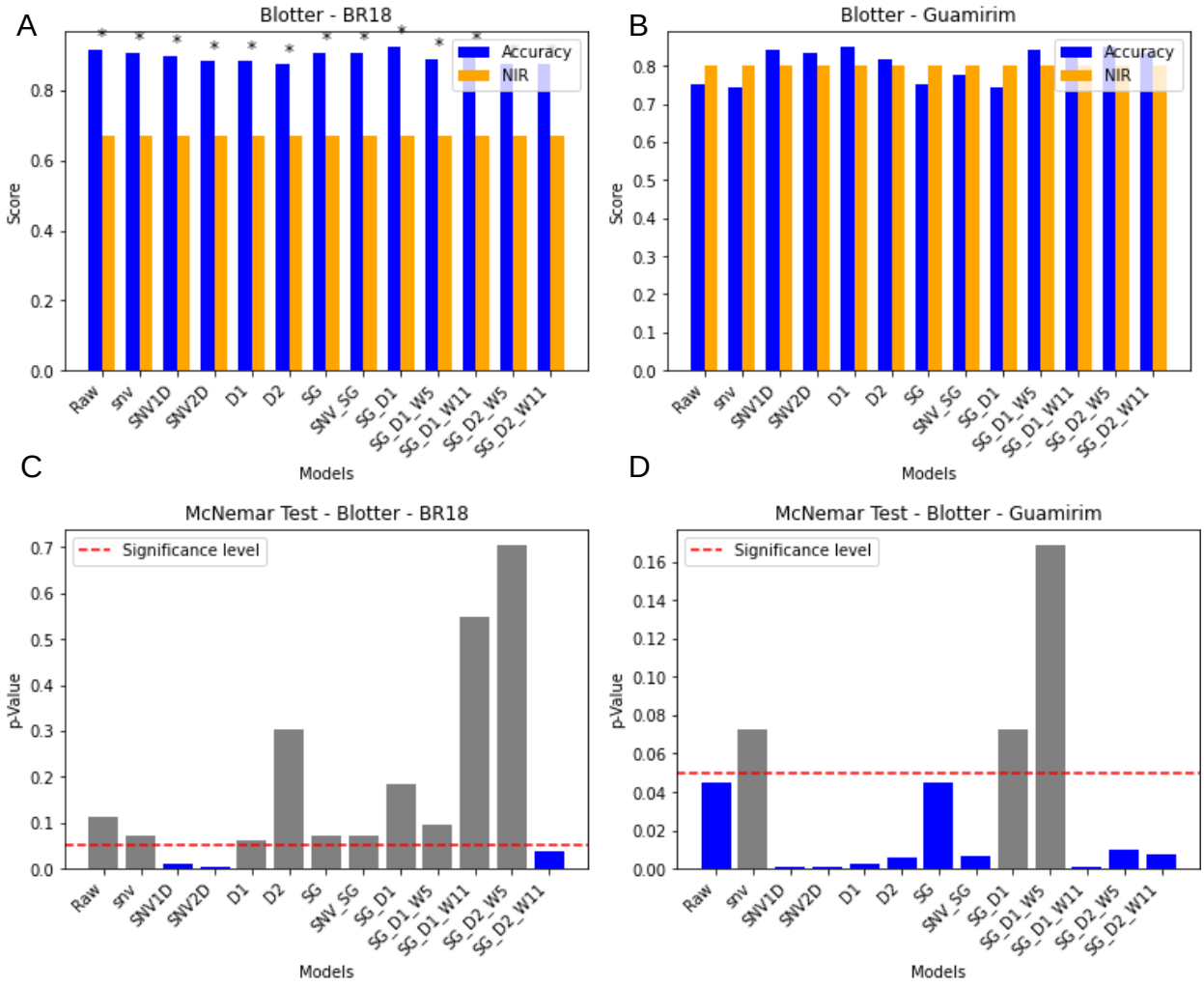


Table 10 - Metrics from machine learning cross-validated and test results using different preprocessing methods to classify seeds according to the blotter test of the Guamirim wheat seed cultivar.

Metrics	**Preprocessing method**												
Infected (+) Non-infected (-)	**Raw**	**SNV**	**SNV1D**	**SNV2D**	**D1**	**D2**	**SG**	**SNV_SG**	**SG_D1**	**SG_D1 W5**	**SG_D1 W11**	**SG_D2 W5**	**SG_D2 W11**
Cross-validated Accuracy	0.8123	0.8084	0.8967	0.9040	0.9051	0.9039	0.8124	0.8051	0.8037	0.8931	0.8727	0.8944	0.8826
Cross-validated Kappa	0.2654	0.2753	0.6000	0.6264	0.6441	0.6503	0.2655	0.2514	0.2926	0.6113	0.5346	0.5910	0.5556
Accuracy	0.7500	0.7417	0.8417	0.8333	0.8500	0.8167	0.7500	0.7750	0.7417	0.8417	0.8417	0.8500	0.8333
Kappa	0.0385	0.0252	0.3537	0.3056	0.4000	0.2667	0.0385	0.0816	0.0252	0.4444	0.3537	0.4231	0.3056
Sensitivity	0.1250	0.1250	0.2917	0.2500	0.3333	0.2500	0.1250	0.1250	0.1250	0.4583	0.2917	0.3750	0.2500
Specificity	0.9062	0.8958	0.9792	0.9792	0.9792	0.9583	0.9062	0.9375	0.8958	0.9375	0.9792	0.9688	0.9792
Pos Pred Value	0.2500	0.2308	0.7778	0.7500	0.8000	0.6000	0.2500	0.3333	0.2308	0.6471	0.7779	0.7500	0.7500
Neg Pred Value	0.8056	0.8037	0.8469	0.8393	0.8546	0.8364	0.8056	0.8108	0.8037	0.8738	0.8469	0.8611	0.8393
Prevalence	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000
Detection Prevalence	0.1000	0.1083	0.0750	0.0668	0.0833	0.0833	0.1000	0.0750	0.1083	0.1416	0.0750	0.1000	0.0667
Balanced Accuracy	0.5156	0.5104	0.6354	0.6146	0.6563	0.6042	0.5156	0.5312	0.5104	0.6979	0.6354	0.6719	0.6146

Figures 12a and 12b show the non-information rate results from the tested models based on Blotter test. The non-infected class represents the majority class. In all the models obtained from the BR18 cultivar, the accuracies were significantly different (Figure 12a), but none of the models were significant from the Guamirim cultivar (Figure 12b). Figures 12c and 12d show the McNemar test for the blotter results of the BR18 cultivar and Guamirim. The models raw, SNV, D1, D2, SG, SNV_SG, SG_D1, SG_D1_W5, SG_D1_W11, and SG_D2_W5 for the BR18 cultivar (Figure 12c) and SNV, SG_D1, SG_D1_W5, for Guamirim (Figure 12d) were not significant, which suggests the misclassifications of these models are distributed between the two classes (infected and non-infected).

Figure 12 - Statistical results comparison for cultivars BR18 and Guamirim based on Blotter test results. A) non-information rate for BR18; B) non-information rate for Guamirim; C) McNemar test for BR18; D) Mcnemar test for Guamirim.



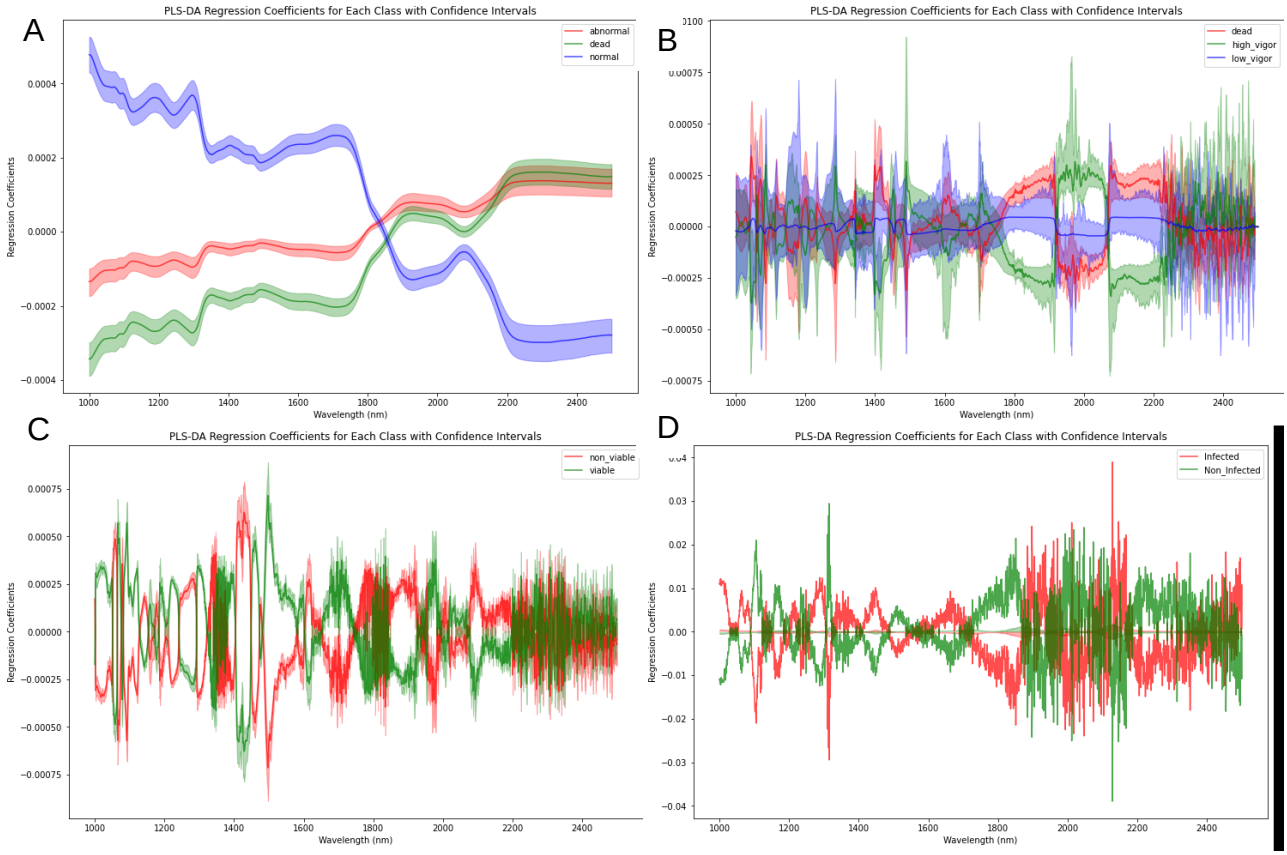
(*) means that the model outperforms the non-information rate.

1.4.5. Regression coefficients

Figures 13 and 14 show the regression coefficient plots for BR18 and Guamirim cultivars, respectively. They represent the models that reach the best scores for overall accuracy and were significant for the NIR test. For BR18, the pre-processing methods SG_D1_W11, SG_D1, SNV_SG, and SNV have better results for the germination, growth, emergence, and blotter tests, respectively. The results for the Guamirim cultivar reveal the following preprocessing methods: SNV_SG for the germination and growth tests, SG_D1 for the emergence test, and D1 for the blotter test.

The class separation based on NIR of the BR18 cultivar for the germination test shows significant regions of the spectra that could be used to separate the normal seedlings from the abnormal and dead. The normal seedlings, represented by the blue line, show higher regression coefficients than the other classes, especially from 1000 nm to 1300 nm. In the region of 1400-1800, it was observed that the confidence intervals are relatively wide, indicating a significant model coefficient variability among the three classes. The region of 2200-2500 did not separate the classes of abnormal and dead seeds (Figure 13a). The growth test plot (Figure 13b) shows a high overlap among the regression coefficients across the entire wavelength range. The regions around 1000-1200 nm, 1300 nm, and 1700 nm show a slight divergence that could be used to separate the three classes (high vigor, low vigor, and dead seeds). In the 1800-1900 nm region, the dead seeds show positive coefficients that diverge from the other classes. This also could be a region of interest for separating dead seeds from high and low-vigor seeds. For the emergence test, it was observed a significant divergence between viable and non-viable seeds in the range of 1100-1250 nm, 1400-1600 nm, 1600-1800, and 2000-2200 nm, with minimal overlap in confidence intervals (Figure 13c). The plot for the blotter test evidenced the regions around 1100 nm, 1300 nm, 1500 nm, and 2100 nm as critical for distinguishing between infected and non-infected seeds (Figure 13d).

Figure 13 - Regression coefficient plots of the BR18 cultivar for each class with confidence intervals.

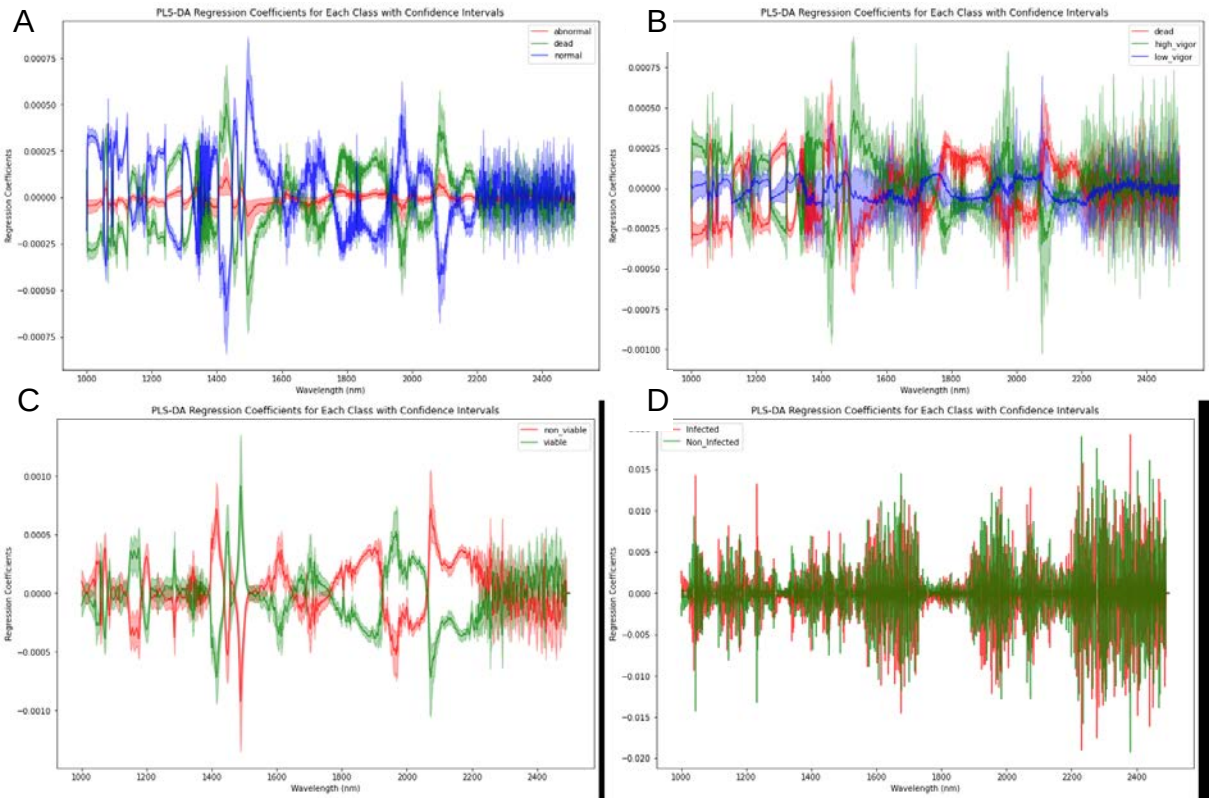


(a) Coefficients of the model obtained after SG_D1_W11 preprocessing method for the germination test. (b) Coefficients of the model obtained after SG_D1 preprocessing method for the growth test. (c) Coefficients of the model obtained after the SNV_SG preprocessing method for the emergence test, and (d) Coefficients of the model obtained after the SNV preprocessing method for the blotter test.

For the Guamirim cultivar, applying the SNV_SG pre-processing method to the germination test shows a wide significant region of the spectra that can distinguish between normal seedlings, abnormal seedlings, and dead seeds. There are several regions where the lines of each class did not overlap in the range between 1000 to 2200 nm, indicating significant variability in model coefficients among the three classes. The normal and dead seeds did not overlap in important regions of the spectra, especially from 1000 nm to 2200 nm. In this range, it was observed positive peaks in the regions of 1000 to 1100 nm, 1220 to 1300 nm, 1500 to 1600 nm, and 1900 to 2050 nm. The abnormal class appears relatively stable across the spectrum with minimal fluctuations, indicating fewer wavelengths significantly contributing to its distinction (Figure 14a). The growth test plot reveals a strong overlap in regression coefficients for all variables in the range of 1000 to 2500 nm. However, the spectral regions between 1400 to 1600 nm and 1800 to 2000 nm exhibit the most significant variation in regression coefficients,

suggesting these wavelengths may be critical for distinguishing between high vigor, low vigor, and dead seeds, despite the overlap observed in the confidence intervals (Figure 14b). Significant divergence between viable and non-viable seeds is observed for the emergence test in the same regions of BR18. Some peaks could contribute to separate classes in the 1100-1200, 1400-1600 nm, 1900 nm, and 2000-2200 nm ranges, with minimal overlap in confidence intervals. The viable class showed positive coefficients at 1150 nm, 1500 nm, and 2000 nm (Figure 14c). The plot for the blotter test was similar to the growth test, and we could not visually identify regions of the spectra that most contributed to separating the infected from non-infected using this pre-processing method (Figure 14d).

Figure 14 - Regression coefficient plots of the Guamirim cultivar for each class with confidence intervals.



(a) Coefficients of the model obtained after SNV_SG preprocessing method for the germination test. (b) Coefficients of the model obtained after SNV_SG preprocessing method for the growth test. (c) Coefficients of the model obtained after the SG_D1 preprocessing method for the emergence test, and (d) Coefficients of the model obtained after the D1 preprocessing method for the blotter test.

1.5. Discussion

Adopting rapid and non-destructive seed quality determination methods represents a significant advancement in agricultural technology. The benefits of these methods include the obtention of quick results that allow timely decisions in the agricultural process and the possibility to perform the same test several times, which is particularly important to seed quality control for seed companies. This study presents the utilization of Near-Infrared Spectroscopy to determine wheat seed quality in the context of contamination by *Fusarium graminearum*. This pathogen has been associated with the reduction of seed germination and vigor, mainly due to the consumption of seed reserves and the resumption of fungal growth and invasion of seed tissue during seed germination and emergence in the field (GARCIA JÚNIOR et al., 2007).

Near-infrared spectroscopy (NIR) involves exposing seeds to near-infrared light, which interacts with the seeds and is absorbed, transmitted, or reflected to varying degrees depending on the chemical bonds present (e.g., water, proteins, carbohydrates). This interaction generates a spectrum collected by a detector, carrying valuable information about the molecular vibrations and overtones within the seeds (DU et al., 2022). Although viable and dead seeds may appear similar before germination, their compositions differ significantly. Viable seeds contain more essential compounds like lipids, proteins, and antioxidants. These chemical differences, detectable by the NIR, are crucial for supporting the initial stages of seed growth. In this study, the NIR could separate viable and non-viable seeds but did not perform well in the intermediate class, represented by abnormal and low-vigor seeds from germination and growth tests, respectively. Fan et al. (2020) also reported difficulty distinguishing the intermediate class, known as moderate vigor, in a study to assess the vigor of wheat seeds based on NIR associated with machine learning methodologies. According to them, a robust classification model would be necessary to accomplish this task.

Research has shown that several factors, such as the seed coat and intrinsic equipment, impact the NIR. These noises directly affect the performance of the technique but are overcome by some pre-processing of the data with great performance using a machine learning approach (ZHU et al., 2015). In this study, the 12 methods were tested for each seed quality test results. As stated before, the methods did perform differently for 3-class and 2-class evaluations. The models for germination and growth tests classified seeds into three different classes. The models were observed to perform better

for germination tests than growth tests. This discrepancy could be due to the methodology used to categorize the dataset classes. The germination test classifies seeds into normal seedlings, abnormal seedlings, or dead (not germinated) based on a standard evaluation of seedling development. This test directly assesses seeds by examining the presence of fundamental structures essential for producing a normal seedling in the field (BRASIL, 2009). In contrast, the growth test classes were determined by the size of the seedlings after a 3-day period of test. Seedling sizes were then categorized into high and low vigor. Since the NIR instrument measures the intensity of light absorbed by the seed and the test is conducted over three days, the subtle differences between classes might not be as easily detectable. This can result in a significant misclassification by the model.

The regression plots from the best models applied to each dataset reveal the differences between germination and growth tests. In this last test, there was a strong overlapping of the classes, indicating that the model does not find strongly distinguishing features in the provided spectral data, which leads to misclassification. Further refinement of the spectral analysis would be necessary to improve the model, which integrates other data types, such as imaging, and specifies the target class better. Also, it could incorporate the automatic analysis of the image to extract the seedling length and incorporate it as a target class, especially using regression-based models.

Both 2-class models achieved above 90% balanced accuracy for some preprocessing methods, indicating that the NIR could more effectively distinguish between viable, non-infected seeds and non-viable, infected seeds. These correspond to the classes for the emergence and blotter tests, respectively. However, the non-information rate test reveals that none of the models were significant considering the blotter test for the Guamirim cultivar. It suggests that considering the accuracy obtained with the models, they were not learning meaningful patterns from the data and were not adding much value beyond simply predicting the majority class. In this context, the regression coefficients for the blotter test in this cultivar due to the infection were not so distinctive. The difficulty in separating the classes may be attributed to the moderate resistance of the Guamirim cultivar to *Fusarium graminearum* (SCHEEREN et al., 2007). Nevertheless, this technique could be utilized to facilitate the differentiation between susceptible and resistant cultivars during the breeding process.

The McNemar test provides valuable insights into the performance of models, particularly in contexts where researchers aim to analyze changes or differences in paired categorical data, often involving matched or dependent samples (McNemar, 1947). In

machine learning, the test is used to assess the ability of the model to classify target classes correctly compared to a reference class. Here, the null hypothesis posits that there is no evidence of the model being biased towards misclassifying one class over another. Generally, it was observed that the models were more likely to fail to reject the null hypothesis for the BR18 cultivar compared to the Guamirim cultivar. This outcome may be related to the lower representation of the Guamirim class, which could affect the error ratio for this cultivar.

The models maximize the differences between the classes by using similar spectra regions. For the germination test, the most discrepant classes, represented by dead seeds and normal seedlings, were in the range of 1400-1800 nm, which is related to overtones of C-H, O-H, and N-H (SHENK; WORKMAN JR.; WESTERHAUS, 2007). These are chemical bonds that compose organic chemicals present in high amounts for seeds that can generate normal seedlings. The healthy seeds also show absorption bands around 2000 nm that are associated with proteins (XU et al., 2019) and 2300 related to lipids (CAZZANIGA et al., 2023), both components crucial for seed metabolism and germination. Interestingly, these regions were also considered to separate viable and non-viable, for both cultivars from emergence test and for separating infected and non-infected seeds of the BR 18 from blotter test. This could be associated with the impact of the pathogen on the chemical components of the seeds, which would affect various quality seed tests.

The Savitzky-Golay algorithm was revealed to be the most applied method among the best results for classifying wheat seeds, considering different vigor and viability tests. Various authors have applied this algorithm to classify seeds of different crops based on NIRS (DELWICHE; HARELAND, 2004; KUSUMANINGRUM et al., 2018; MEDEIROS et al., 2022; OLESEN et al., 2011; QIU et al., 2018; SEO et al., 2016). This pre-processing method stood out for the germination test for both cultivars and could be applied to classify seeds as normal, abnormal, or dead. Also, this algorithm was applied to the growth test of the BR18 cultivar and the emergence test of BR18 and Guamirim, confirming the importance of this method. According to Shenk et al. (2007), using the derivatives helps visualize the spectra peaks since apparent band resolution enhancement occurs. The same authors pointed out that applying a second derivative to agricultural products such as seeds has the advantage of generating few if any, false peaks.

1.6. Conclusion

The findings reveal that the distribution of each category in the dataset significantly influences the performance of the models. It is crucial to consider this to prevent misinterpretation of results, especially when dealing with imbalanced datasets, which are common in seed analysis. In assessing seed viability, better results were obtained using the SG_D1_W11 pre-treatment for BR18 and the SNV_SG for Guamirim. When evaluating the vigor of the BR18 cultivar, better performance was observed with the SG_D1, SNV_SG, and SNV pre-treatments for growth, emergence, and blotter tests, respectively. For the Guamirim cultivar, the SNV_SG, SG_D1, and D1 pre-treatments were particularly effective for growth, emergence, and blotter tests, respectively.

1.7. References

Alisaac, E., Rathgeb, A., Karlovsky, P., Mahlein, A. K. (2021). Fusarium head blight: Effect of infection timing on spread of *Fusarium graminearum* and spatial distribution of deoxynivalenol within wheat spikes. *Microorganisms*, 9(1), 1–12. https://doi.org/10.3390/microorganisms9010079

Asran, M. R., Amal, M. I. E. (2011). Aggressiveness of certain *Fusarium graminearum* isolates on wheat seedlings and relation with their trichothecene production. *Plant Pathology Journal*, *10*(1), 36–41. https://doi.org/10.3923/ppj.2011.36.41

BRASIL (2009). Ministério da Agricultura. Manual de análise sanitária de sementes. Brasília, LANARV/SNDA/MA. 200 p.

Burns, D.A., & Ciurczak, E.W. (Eds.). (2007). Handbook of Near-Infrared Analysis (3rd ed.). CRC Press. https://doi.org/10.1201/9781420007374

Cazzaniga, E. et al.(2023) Lipids in a nutshell: quick determination of lipid content in hazelnuts with NIR spectroscopy. *Foods*, v. 12, n. 1, p. 1–11.

Chen, Y. R., Baek, S. J., Cho, B. K. (2009) Improvement of PLS modeling for the non-destructive determination of soluble solids content in Intact 'Fuji' apples by Vis/NIR spectroscopy using a new SNV-Detrend pre-processing method," *Postharvest Biology and Technology*, vol. 51, no. 2, p. 210-216.

Del Ponte, E. M., Fernandes, J. M. C., Pierobom, C. R., Bergstrom, G. C. (2004). Giberela do trigo: aspectos epidemiológicos e modelos de previsão. *Fitopatologia Brasileira*, 29(6), 587–605. https://doi.org/10.1590/s0100-41582004000600001

Delwiche, S. R.; Hareland, G. A.(2004) Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. *Cereal Chemistry*, 81: 643-649. https://doi.org/10.1094/CCHEM.2004.81.5.643

Du, Z. et al. (2022). Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*, v. 21, n. 3, p. 2956–3009. https://ift.onlinelibrary.wiley.com/doi/10.1111/1541-4337.12958.

Fan, Y., Ma, S., Wu, T. (2020). Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Physics and Technology*, *105*(1), 103213. https://doi.org/10.1016/j.infrared.2020.103213

Food and Agriculture Organization (FAO). 2009. How to Feed the World in 2050. Available online: https://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Garcia Júnior, D., Vechiato, M. H., Menten, J. O. M., Lima, M. I. P. M. (2007). Influência de *Fusarium graminearum* na germinação de genótipos de trigo (*Triticum*

aestivum L.). *Arquivos do Instituto Biológico*, 74(2), 157–162.

Hassani, F. Leila Z., and Nima K. (2019). Evaluation of germination and vigor indices associated with *Fusarium*-infected seeds in pre-basic seeds wheat fields. *Journal of Plant Protection Research,* 59 (1): 69–85. https://doi.org/10.24425/jppr.2019.126037.

INTERNATIONAL SEED TESTING ASSOCIATION. *Handbook on Seedling Evaluation*. 4. ed. Bassersdorf: International Seed Testing Association, 2018.

Kusumaningrum, D.; Lee, H.; Lohumi, S.; Mo, C.; Kim, M.S.; Cho, B.-K. (2018) Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy. *Journal of the Science of Food and Agriculture*, 98: 1734–1742. https://doi.org/10.1002/jsfa.8646.

MAPA – MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO -. Padrões para a produção e a comercialização de sementes de trigo (*Triticum aestivum* L.). Available in: https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/publicacoes-sementes-e-mudas/copy_of_INN45de17desetembrode2013.pdf . Acessed on April 15, 2024.

Martens, H.; Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near-infrared spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 9, no. 8, pp. 625-635, 1991.

McDonald, M.B., Copeland, L.O. (1997). Seed Quality and Performance. In: Seed Production. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-4074-8_8

Medeiros, M. L. S. et al. (2022). Assessment oil composition and species discrimination of brassicas seeds based on hyperspectral imaging and portable near infrared (NIR) spectroscopy tools and chemometrics. *Journal of Food Composition and Analysis*, [*s. l.*], v. 107.

Olesen, M. H., Nisha, S., Rene, G., Birte B. (2011). “Classification of viable and non-viable spinach (*Spinacia oleracea* l.) seeds by single seed near infrared spectroscopy and extended canonical variates analysis. *Journal of Near Infrared Spectroscopy* 19 (3): 171–80. https://doi.org/10.1255/jnirs.928.

Peiris, K., Pumphrey, M., Dong, Y., Maghirang, E., Berzonsky, W., Dowell, F. (2010). Near-infrared spectroscopic method for identification of fusarium head blight damage and prediction of deoxynivalenol in single wheat kernels. *Cereal Chemistry*, *87*(6), 511–517.

Qiu, G., Enli, L., Huazhong, L., Sai, X., Fanguo, Z., Qin, S. (2018). Single-kernel FT-NIR spectroscopy for detecting supersweet corn (*Zea may*s l. Saccharata Sturt) seed viability with multivariate data analysis.” *Sensors (Switzerland)* 18 (4). https://doi.org/10.3390/s18041010.

R CORE TEAM. R Development Core Team. R: A Language and Environment for Statistical Computing, 2023. http://dx.doi.org/ http://www.R-project.org

Rinnan, A., Vandenberg, F., Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10), 1201-1222. DOI: 10.1016/j.trac.2009.07.007

Savitzky, A., Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry*, vol. 36, no. 8, pp. 1627-1639, 1964.

Scheeren, P. L., Caierão, E., Só e Silva, M., Del Duca, L.J. A., Nascimento Junior, A. Linhares, A., Eichelberger, L. (2007). BRS Guamirim: cultivar de trigo da classe pão, precoce e de baixa estatura." *Pesquisa Agropecuária Brasileira* 42 (2): 293–96. https://doi.org/10.1590/s0100-204x2007000200020.

Seo, Y., Ahn, C., Lee, H. Park, E., Mo, C., Cho, B. (2016). non-destructive sorting techniques for viable pepper (*Capsicum annuum* l.) seeds using Fourier transform near-infrared and Raman spectroscopy." *Journal of Biosystems Engineering* 41 (1): 51–59. https://doi.org/10.5307/JBE.2016.41.1.051.

Shenk, J. S., Workman Jr, J.J., Westerhaus, M. O. (2007). Application of NIR spectroscopy to agricultural products." In *Handbook of Near-Infrared Analysis*, edited by Burns, D. A. and Ciurczak, E. W. 3rd ed. CRC Press. https://doi.org/https://doi.org/10.1201/9781420007374.

Soren Balling Engelsen, S., Norgaard, L., Bro, R. (1999). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, vol. 53, no. 9, pp. 1104-1109.

Sousa, P. G.. (2002). BR 18-Terena: cultivar de trigo para o Brasil. Pesquisa Agropecuária Brasileira, 37(7), 1039–1043. https://doi.org/10.1590/S0100-204X2002000700019

Workman Jr., J., Weyer, L. Practical guide to interpretive near-infrared spectroscopy. 1st ed. Boca Raton: CRC Press, 2007. 344 p. https://doi.org/10.1201/9781420018318

Xia, Y., Xu, Y., Li, J., Zhang, C., Fan, S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, *1*, 35–47. https://doi.org/10.1016/j.aiia.2019.05.001

Xu, J., Chinedu, C. N., Nadil Shah, Y. Z., Chunyu, Z. (2019). Identification of genetic variation in *Brassica napus* seeds for tocopherol content and composition using near-infrared spectroscopy technique. *Plant Breeding* 138 (5): 624–34. https://doi.org/10.1111/pbr.12708.

Zhang, C., Wang, J., Zhang, W., et al. (2020). Application of the gap-segment derivative method for quantitative analysis of NIR spectra. *Applied Spectroscopy*, vol. 74, no. 3, pp. 303-308, 2020.

Zhu, L., Wen-guang, M., Jin Hu, Y. Z., Yi-xin, T., Ya-jing, G., Wei-min, H. (2015). Advances of NIR spectroscopy technology applied in seed quality detection. *Journal of Infrared and Millimeter Waves*, 35: 346–49. https://doi.org/10.3964/j.issn.1000-0593(2015)02-0346-04.

CHAPTER II

Utilization of Near-Infrared Spectroscopy (NIR) to Evaluate the Physiological and Sanitary Quality of Wheat Seeds.

2.1. Abstract

Fusarium head blight (FHB) is a major disease that affects wheat, reducing both physiological and sanitary quality of the seeds. The infection can occur at any stage of plant development, leading to the possibility of infected seeds without visible symptoms. The combination of NIR spectroscopy and the PLS-DA algorithm has been explored for detecting Fusarium Damaged Kernels (FDK), as it is a fast and non-destructive method. However, its application in the seed sector still needs improvement. This study aimed to investigate the potential of NIR spectroscopy combined with the PLS-DA algorithm to classify individual seeds according to their physiological and sanitary quality. The study obtained 4640 seeds of two wheat cultivars from plants inoculated with *Fusarium graminearum* at different stages of seed development. All seeds were individually submitted to near-infrared spectroscopy in the 1000 to 2500 nm spectral range. The classes of each seed were obtained through the tests of germination, seedling growth, emergence, and sanitary (Blotter). The spectral data associated with each class were pre-processed using the Savitzky-Golay second derivative methodology and analyzed using the PLS-DA algorithm. The models obtained resulted in an accuracy above 87% for training data and above 83% in validation for the correct classification of each seed regarding their germination, emergence, and Blotter tests for both cultivars. However, an overfitting of the model for separating the classes from the seedling growth test was verified. The spectral bands corresponding to 1000-1200, 1400-1600, 1900-2200, and 2400 nm were more important for separating the classes. Hence, it is concluded that NIR spectroscopy combined with the PLS-DA algorithm can efficiently classify seeds according to the physiological and sanitary quality of wheat seeds.

Keywords: Fusarium head blight; Machine learning; PLS-DA; Seed quality

2.2. Introduction

Wheat belongs to the Poaceae family, to the genus *Triticum*, and the main cultivated species are *Triticum durum*, *Triticum aestivum*, and *Triticum monococcum*. It is one of the most cultivated crops in the world, along with maize and rice, and thus plays a fundamental role in the global economy and food security. Brazil is in the 15th position in the ranking of the largest wheat producers in the world, but national production is still not enough to meet the domestic market, estimated at 12.6 million tons (CONAB, 2023). Investments in technologies and practices that improve the national wheat sector are essential to meet domestic demand. This investment must go through the production of quality seeds.

Seed quality comprises the physical quality, seeds that present a high degree of purity, free of inert materials, impurities, and mechanical damage; physiological quality, seeds that have high germination capacity and vigor, essential for a strong and uniform stand; genetic quality, seeds that present varietal purity; and sanitary quality, disease-free seeds. In wheat, research on pathogens that can harm humans through consuming the cereal has been extensively studied, as it goes beyond seed quality in terms of health concerns. The presence of *Fusarium graminearum* in grains is associated with deoxynivalenol (DON) synthesis. This mycotoxin causes health problems in humans and animals, such as abdominal discomfort, diarrhea, vomiting, and anorexia (Khaneghah et al., 2018). Besides the impacts on food, studies have shown a significant reduction in seed production and quality (Asran and Amal, 2011; Hassani et al., 2019; Garcia Júnior et al., 2007).

Fusarium head blight, caused by *Fusarium graminearum*, is one of the most relevant diseases that affect the wheat crop. It is a monocyclic disease in which the fungus survives as a saprophyte in crop residues forming perithecia (sexual spores) and macroconidia (asexual spores). Under conditions of high humidity and temperatures between 20 and 30 °C, ascospores may be released from inside the perithecia and transported over long distances by wind, while rain splashes release macroconidia over short distances (Del Ponte et al., 2004; Garcia Junior et al., 2007). At anthesis, upon reaching the ear, the fungus can destroy the flower, preventing seed formation, or act slowly, causing the formation of empty and deformed seeds (Del Ponte et al., 2004). On the other hand, when the infection occurs at later stages, there will be the formation of filled seeds that may be contaminated (Del Ponte et al., 2007;

Alisaac et al., 2021; Siou et al., 2014). These seeds are difficult to eliminate in the cleaning process, directly affecting the sanitary and physiological quality of the seed lot.

In a study conducted by Hassani et al. (2019), it was observed that wheat seeds infected with *Fusarium* spp. had a higher percentage of abnormal seedlings and reduced growth. This is due to the ability of the *Fusarium* to colonize the pericarp and embryonic axis of the seeds. Gilbert et al. (1997) also observed abundant growth of *F. graminearum* in the scutellum and embryonic axis of wheat seeds during imbibition (Phase I), even without any visible symptoms of fungal contamination.

Garcia Junior et al. (2007) found a positive correlation between the presence of the pathogen and dead seeds, which can impact the final germination percentage. Also, the presence of the pathogen in the tissues can cause damping off and death of seedlings, compromising the formation and establishment of the culture in the field (Asran and Amal, 2011). Additionally, the respiration of the pathogen can increase the temperature and humidity during storage, leading to the loss of dry matter (Hassani et al., 2019). This significantly reduces the viability and vigor.

The presence of *Fusarium graminearum* in seeds is usually determined using the Blotter or filter paper tests. This test consists of identifying the growth of the pathogen in seeds sowed on a layer of filter paper inside containers, such as gerboxes or Petri dishes, for 7 days for *Fusarium graminearum*. Alternatively, for the breeding process, some companies are using the visual method, which consists of attributing a score for a sample by an expert worker. These are laborious and time-consuming procedures that requires practice to ensure test accuracy. Thus, different alternative methodologies capable of determining fungal infection quickly and non-destructively have emerged, such as near-infrared spectroscopy (NIR) analysis.

Near-infrared spectroscopy is a fast, non-destructive technique that allows obtaining information on the chemical composition, such as lipids, proteins, and carbohydrates, of products such as seeds (Xia et al., 2019). Obtaining the spectral signature through NIR and analyzing the spectra through machine learning algorithms has proven to be highly efficient for classifying materials with high DON indices (Alisaac et al., 2019; Liang et al., 2020), evaluation of *Fusarium* infection (Shahin, Symons, 2011; Singh et al., 2012) and physiological quality (Fan et al., 2020; Zhang et al., 2018) of wheat seeds.

Classifying wheat seeds based on *Fusarium* infection is crucial to advance Brazilian wheat production. Furthermore, the seed technology industry requires a fast and non-destructive method to assess seed quality. Inserting NIR into the wheat seed production chain to eliminate Fusarium-contaminated seeds can help ensure proper growth in the field and prevent the disease from spreading to new areas. Therefore, the objective of this study was to investigate the potential of NIR spectroscopy combined with PLS-DA algorithm to classify individual seeds according to their physiological and sanitary quality.

2.3. Material and methods

2.3.1. Plant material

Seeds from two wheat cultivars, BR18, known as susceptible (SOUSA, 2002), and Guamirim, showing moderate resistance to Fusarium Head Blight (FHB) (SCHEEREN, et al. 2007), were sown in 3-liter pots filled with substrate and cultivated in a greenhouse with temperatures between 25 °C and 30 °C. Twelve plants per pot were maintained, and secondary tillers were eliminated throughout the experiment. After reaching the reproductive phase, the plants were sprayed with *Fusarium graminearum* inoculum with suspension adjusted to a final concentration of 1×10^4 macroconidia mL^{-1}, at different growth stages, according to the Zadoks scale (Zadoks et al., 1974).

Single-time spray inoculation was performed at anthesis (ZS 65), early milk (ZS 73), and early dough (ZS 83). The control consisted of non-inoculated plants. After completing the cultivation cycle, the seeds were harvested and scanned using a Fourier transform infrared spectrometer. Then, the same seeds were submitted to physiological (germination and vigor) and sanitary (Blotter) tests. The analyses were conducted in the Agroenergy, Seed Research, and Epidemiology laboratories at the Federal University of Viçosa.

2.3.2. Acquisition of spectra and obtaining classes

2.3.2.1. NIR spectroscopy analysis

The spectral data were obtained with a Fourier transform infrared spectrometer FT-NIR (Thermo Scientific Antaris II). Individual seeds from each lot were placed in a specific support to the size and shape of the wheat seeds. For each seed, 3111 points per spectrum were collected within the 1000 – 2500 nm wavelength range and a resolution of 8 cm^{-1}. The average spectrum of 30 successive scans for each seed was evaluated. It was evaluated 800 seeds of each cultivar, for the germination test; 800 seeds per cultivar, for the emergence test; 320 seeds per cultivar, for the seedling growth test; and 400 seeds per cultivar for the Blotter test.

2.3.2.2. Germination

After obtaining the spectra, the corresponding seeds were placed to germinate on towel paper moistened with 2.5 times their dry weight at a temperature of 20 °C for 8 days, according to the International Seed Testing Association (ISTA, 2018). The seeds were separated into three classes based on seedling growth: normal, abnormal, or dead seed, which were used in the supervised machine learning model.

2.3.2.3. Emergence

For constructing the machine learning model to predict the emergence capacity of wheat seeds, after obtaining the NIR spectra data of the individual seeds, they were sowed in expanded polystyrene trays filled with sand and kept in a greenhouse with temperatures ranging between 25 and 30 °C. The seedlings 5 mm above the surface of the substrate were counted. Daily counts were performed for 15 consecutive days. The emergence test results were used to classify the seeds as viable and non-viable, which were used as the target class to build the machine learning model.

2.3.2.4. Seedling Growth

NIR spectroscopy-based model to differentiate high and low vigor classes and dead seeds was trained from the seedling growth test result. After obtaining the spectra, the tracked seeds were placed to germinate on towel paper, moistened with 2.5 times the weight of dry paper at 20 °C in a Mangelsdorf germination chamber for

72 hours. After this period, images of seedlings were captured using a scanner (HP, Scanjet 200) fixed upside-down inside the metal box at a scanning resolution of 300 dpi. The seedling length was later measured with ImageJ software. The classes were defined according to the size of each seedling. Seedlings greater than 30 mm were classified as high vigor, and those with less than 30 mm were classified as low vigor. The class corresponding to dead seeds were those without protrusion of the radicle after the test period and with intense deterioration by pathogens. These classes were used to create the supervised ML model. Seedling length was also utilized to assess the indices of vigor, growth, and uniformity of the seeds, using the statistical software SeedCalc (Silva, Medeiros, and Oliveira, 2019).

2.3.2.5. Blotter

After obtaining the spectra, the seeds were disinfested to eliminate possible microorganisms in the seed coat or associated with them through the air. At this stage, the seeds were packed in perforated plates and then dipped in alcohol 70% for 60 seconds, followed by immersion in sodium hypochlorite solution (1%) for 2 minutes. After that, the seeds were rinsed twice in distilled sterile water. After washing, the plates were left on paper towels to remove excess water. Then the seeds were sown in acrylic boxes (40x40x40 cm) on three sheets of towel paper, moistened with saline solution (NaCl) to prevent seed germination. The whole process was done under sterile conditions inside the laminar flow cabinet. The boxes were placed in BOD regulated at 25 °C for seven days. After this period, the boxes were opened, and each seed was visually evaluated for signs of *Fusarium graminearum* growth. If observed growth of mycelium, the seed was classified as infected, and in the absence, the seed was classified as non-infected. These classes were attributed to the previously obtained spectra for the training and testing of the model.

2.3.3. Statistical analysis

2.3.3.1. Box plot

Boxplots were created to visualize the distribution of physiological test results across different inoculation periods (ZS 65, ZS 73, ZS 83, and the control). Each boxplot displays the median, interquartile range (IQR), and potential outliers, with whiskers

extending to the most extreme data points within 1.5 times the IQR. These visualizations were generated to enable a comparative assessment of seed quality and physiological parameters, highlighting differences across inoculation periods.

2.3.3.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was employed as an exploratory multivariate technique. For each seed test, the dataset was structured into matrices, and groups of potentially correlated variables were transformed into linearly uncorrelated variables. The first two principal components for each seed test were then plotted to visualize the data.

2.3.3.3. Linear Discriminant Analysis (LDA)

The results of the NIR were organized and modeled using Principal Component Analysis (PCA) to extract the principal component (PC) scores from the spectra. These scores were then used as input for Linear Discriminant Analysis (LDA). The software primarily generates graphical representations, such as decision boundaries and class separation, to visualize the performance of the model and highlight feature contributions.

2.3.3.4. Obtaining Machine Learning Model

After obtaining the NIR spectra and subsequent determination of the classes of each seed, the data were organized into matrices to obtain the supervised models referred to each seed quality test. The data were preprocessed using the methodology of the second derivative of Savitzky-Golay to eliminate noise intrinsic to the equipment and the surface of the seeds. After pre-processing, 70% of the data were used for training and cross-validation (10 layers), and 30% for independent model validation using the PLS-DA algorithm. Based on the classification results, the accuracy, sensitivity, specificity, precision, F1-score, and Kappa index were determined according to the following equations:

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

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 - Score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$$

$$Kappa = \frac{(TP * TN) - (FP * FN)}{(TP * TN) + (TP * FP) + (2 * TP * TN) + FN^2 + (FN * TN) + FP^2 + (FP * TN)}$$

Where *TP, TN, FP, and FN* correspond to a true positive, true negative, false positive, and false negative, respectively.

2.3.3.5. Importance of variables

The importance of the variables was assessed using the varImp function from the caret package in R 4.1.2 (R Core Team, 2024). This method is based on the analysis of absolute regression coefficients, weighted by the reduction in the residual sum of squares across the components of the Partial Least Squares Regression (PLSR) model. The importance of each variable was calculated separately for each class. The variable weights were obtained by considering their contribution to explaining the variability in the model, with more relevant variables having larger absolute coefficients. The visualization of variable importance was performed using a scatter plot.

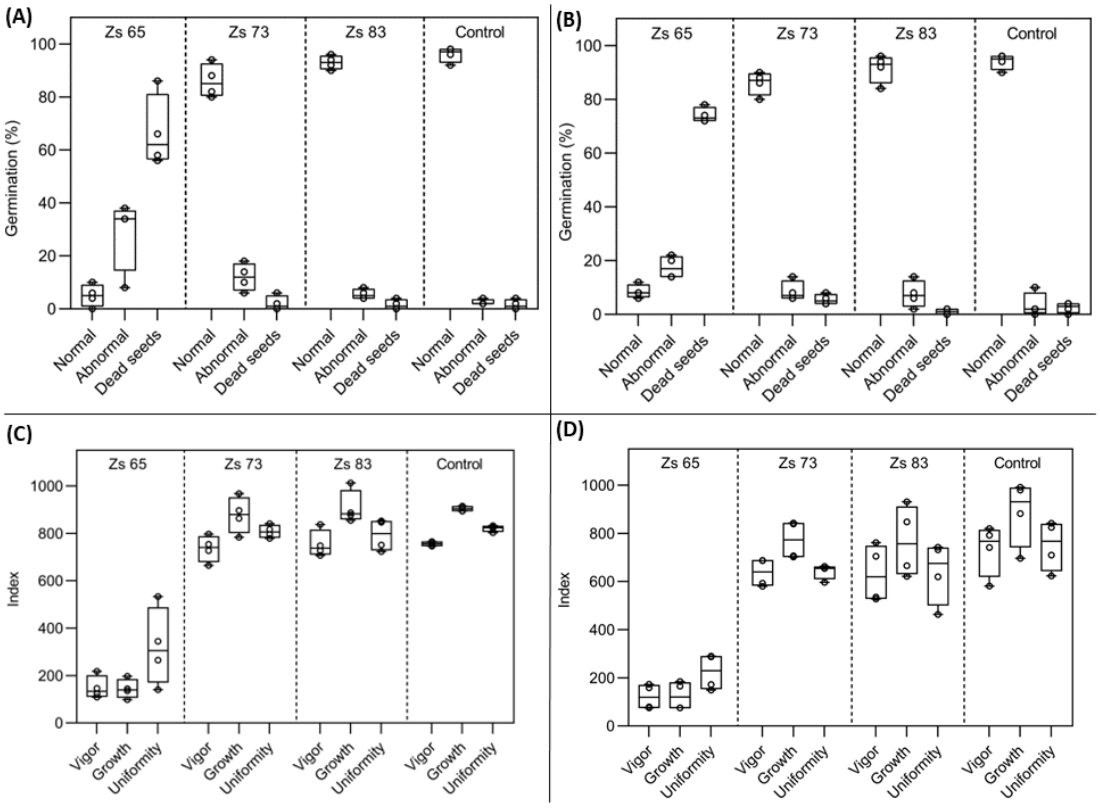
2.4. Results

2.4.1. Physiological data

The inoculation of the wheat cultivars at anthesis (ZS 65) significantly reduced the seed quality (Figure 15). At this stage, for both cultivars, the lowest final germination percentage was observed, quantified by the ability of the seed to establish a normal seedling, and a significant reduction in the vigor, growth, and uniformity indices, evaluated by the growth test. At ZS 73 and ZS 83, corresponding to early milk and early dough, respectively, the test result of final germination reached the minimal

percentage (80%) to be commercialized as a seed (MAPA, 2013). However, it was inferior to the control, which showed the highest percentage of normal seedlings for both cultivars (Figure 15 A and B). The vigor evaluated through the seedling growth test did not differ at these stages from the control for both cultivars (Figure 15 C and D).

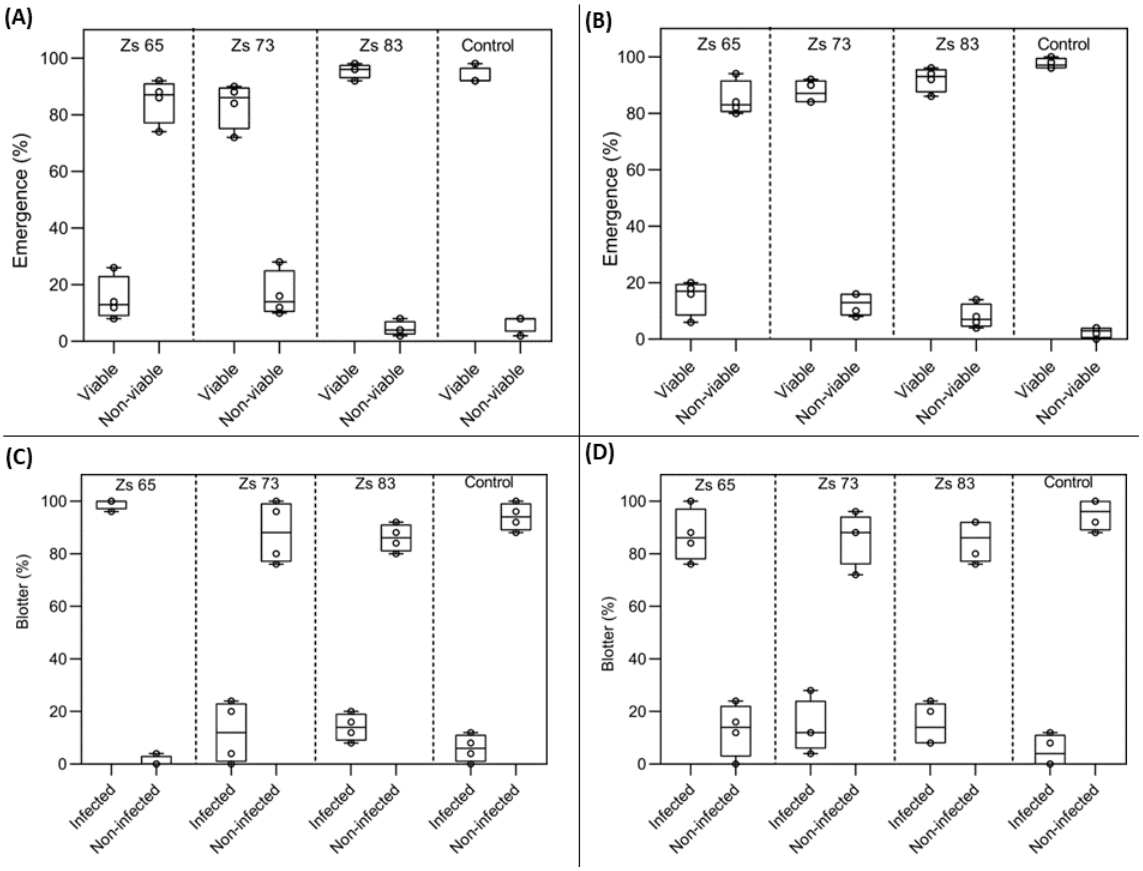
Figure 15 - Box-plot of the wheat seed quality test results*. Final germination percentage of BR18 (A) and Guamirim (B) cultivars. Vigor, Growth and Uniformity indices from BR18 (C) and Guamirim (D) cultivars.



*The data represent average and mean standard errors for individual seeds harvested from plants inoculated at different growth stages. Zs 65: anthesis, Zs 73: early milk, ZS 83: early dough, Control: non-inoculated plants.

The percentage of seedlings emergence was significantly reduced with the inoculation at the anthesis stage (≈20%) and increased with the subsequent stages. However, there was no difference between the inoculation at ZS 83 and the control for BR 18 cultivar. For Guamirim, the control was superior to the inoculated stages (Figure 16 A and B). The blotter results confirmed the trend observed with the physiological quality tests. They showed that approximately 100% of the seeds were infected at ZS 65, and the infection was significantly reduced in later stages and with the control (Figure 16 C and D).

Figure 16 - Box-plot of the wheat seed quality test results*: Seedling emergence of BR18 (A) and Guamirim (B) cultivars. Blotter results from BR18 (C) and Guamirim (D) cultivars.



*The data represent average and mean standard errors for individual seeds harvested from plants inoculated at different growth stages. Zs 65: anthesis, Zs 73: early milk, ZS 83: early dough, Control: non-inoculated plants.

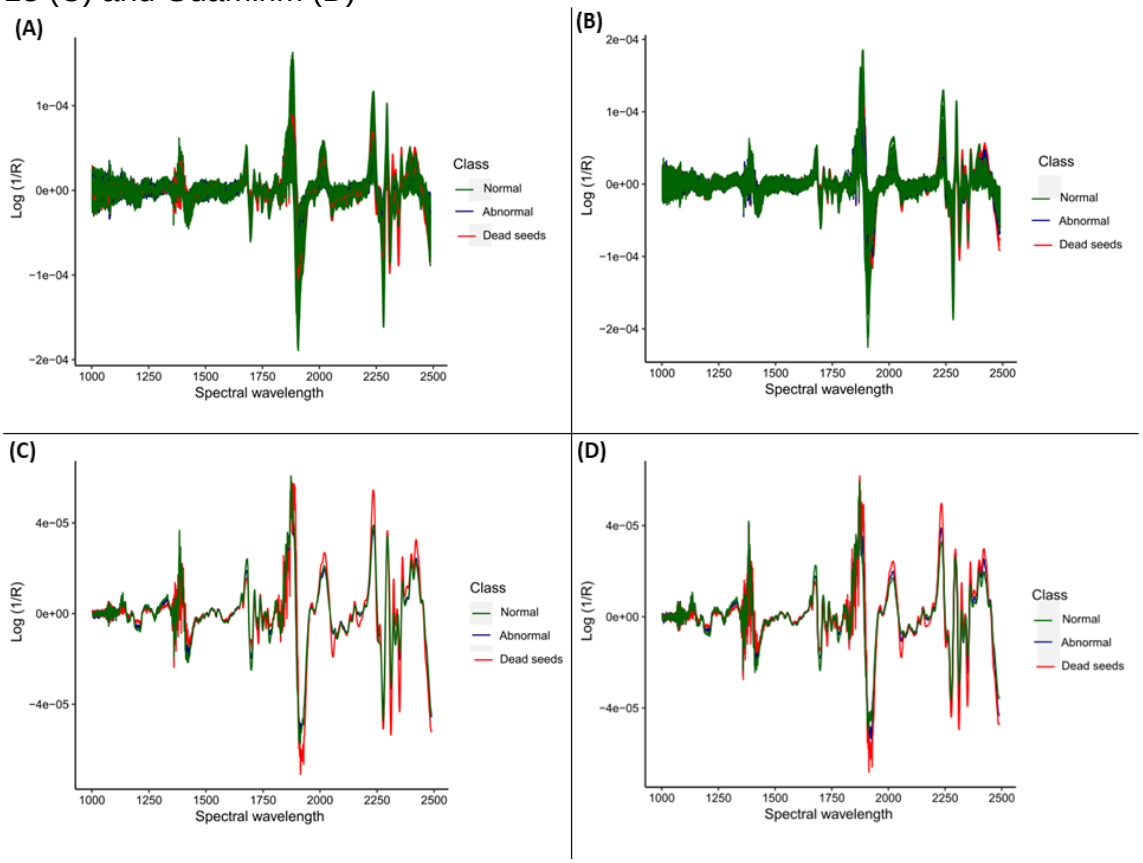
2.4.2. Models

2.4.2.1. Germination

The spectra and the averaged spectra, processed by second derivative of Savitzky-Golay method, of normal, abnormal, and dead seeds obtained with the germination test results of BR 18 and Guamirim cultivars are shown in Figure 17. The data was preprocessed to minimize the noises associated with the equipment and associated with the samples. We observed a similar pattern with the spectra for all

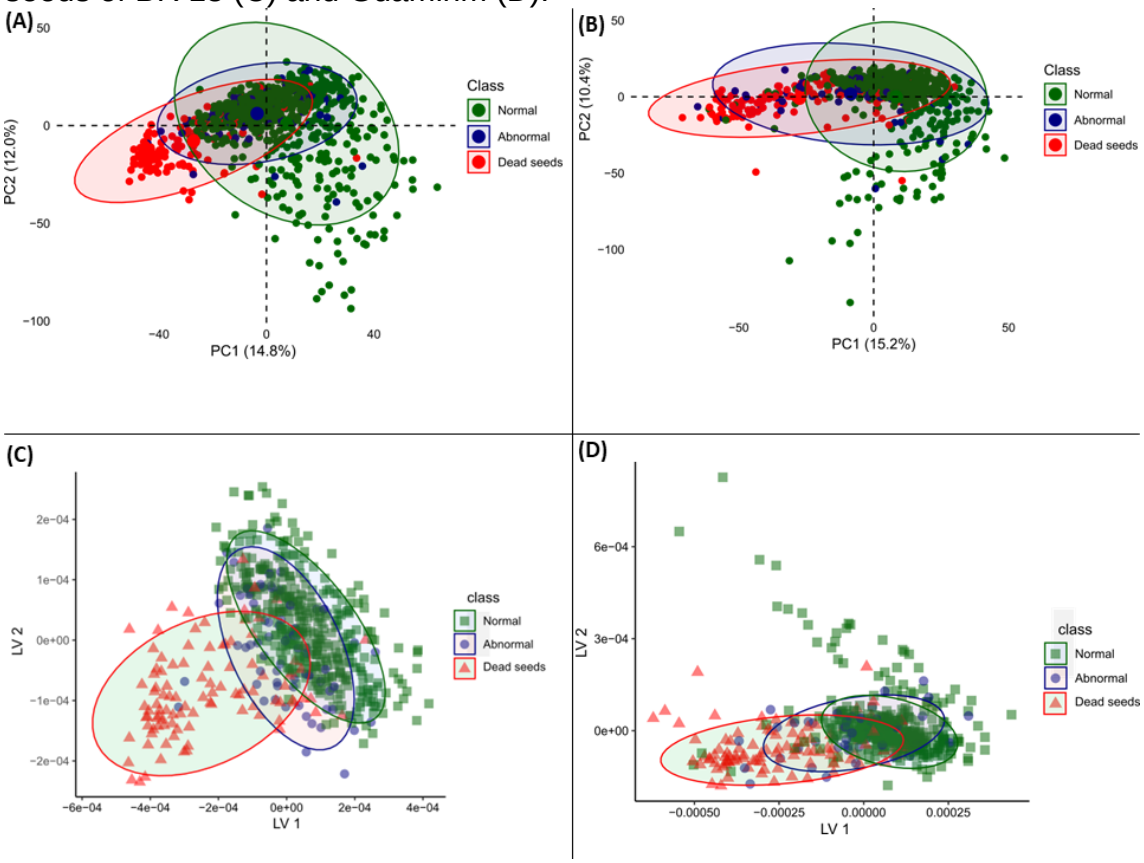
classes evaluated, indicating that observing peaks and valleys would not be sufficient to identify differences between samples.

Figure 17 - FT-NIR absorption spectra of wheat seeds classified according to germination results. Spectra of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay; Average spectra of BR 18 (C) and Guamirim (D)



The principal component analysis showed that the first two components explained 26.8% of the variance for BR18 and 25.6% for the Guamirim cultivar. Therefore, the PCA could not separate between classes by using this transformation and only two components since there was a significant overlapping between individuals (Figure 18). Nevertheless, the spectra were evaluated by the chemometrics approach.

Figure 18 - Principal component analysis using FT-NIR data of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay and germination test results. Latent variables from the model used to classify seeds of BR 18 (C) and Guamirim (D).



The results from the PLS-DA model applied to the train and test datasets are shown in Table 1. Evidently, there was a significant imbalance of the datasets for both cultivars. The most representative class was the normal seedlings, followed by the class of dead seeds and the abnormal seedlings. The model reached an accuracy and sensitivity greater than 83% for both cultivars and subsets of the data (train and test), indicating efficiency in classifying the different classes from the germination test.

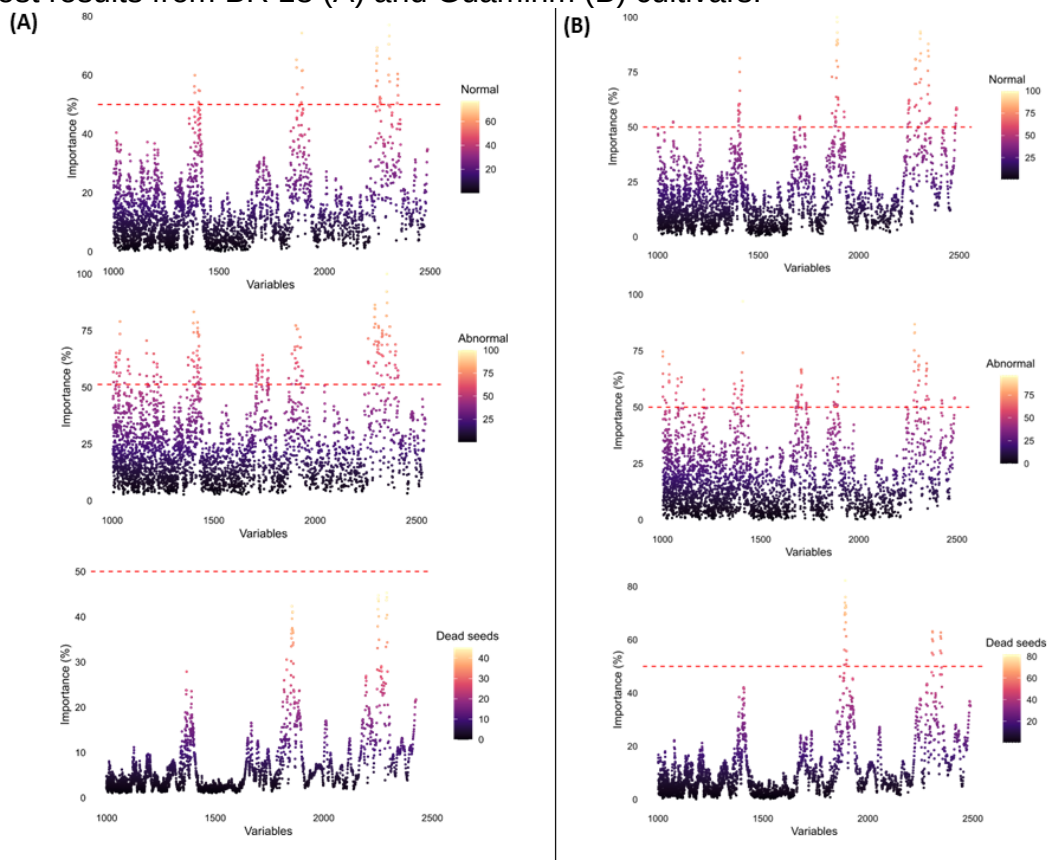
Table 11 - Hit number per class and metrics obtained with the PLS-DA classification model, according to germination results, using the FT-NIR resources of wheat seeds from two different cultivars. The number in parenthesis means the occurrence of the class in the dataset.

Class	BR 18		Guamirim	
	Train	Test	Train	Test
	Hits (Total)			
Normal	390 (392)	165 (168)	382 (391)	165 (167)

Abnormal	35 (68)	7 (29)	0 (49)	0 (20)
Dead seeds	76 (101)	28 (42)	93 (104)	28 (44)
Accuracy	0.893	0.837	0.873	0.835
Sensitivity	0.754	0.630	0.624	0.541
Specificity	0.902	0.845	0.889	0.836
Precision	0.862	0.763	-	-
F1-Score	0.797	0.670	-	-
Kappa	0.748	0.589	0.682	0.549

The spectral range near to 1400, 1900, and 2300 nm were more important for the model to separate the classes for both cultivars (Figure 19).

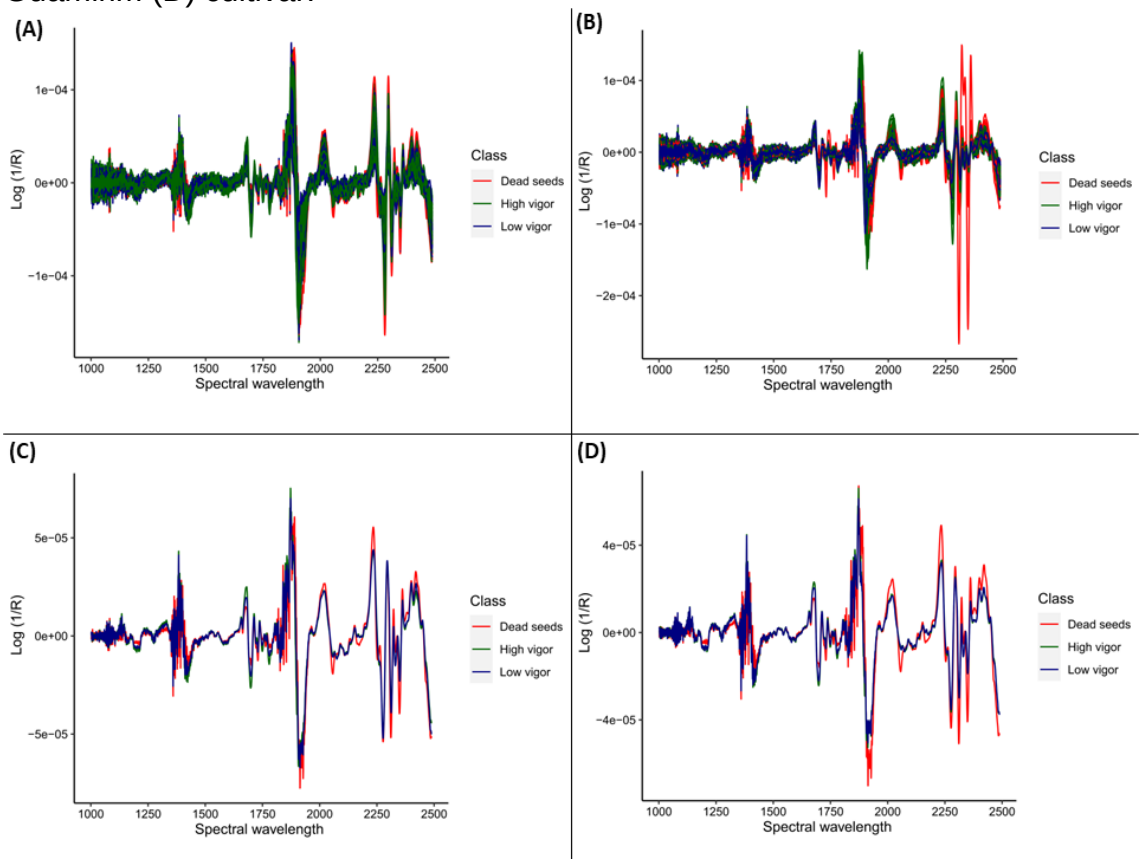
Figure 19 - Importance of variables used to classify seeds according to the germination test results from BR 18 (A) and Guamirim (B) cultivars.



2.4.2.2. Growth

From the growth test, it was obtained 3 classes: High vigor, low vigor, and dead seeds. There was a pattern between both spectra and averaged spectra of the cultivar graphs, indicating a separation between dead seeds and viable ones. However, it could not differentiate the high-vigor seeds from the seeds of low vigor (Figure 20).

Figure 20 - FT-NIR absorption spectra of wheat seeds classified according to the Growth results Spectra of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay; Average spectra of BR 18 (C) and Guamirim (D) cultivar.



The principal component analysis showed that the first two components explained 28.2% of the variance for BR18 and 23.8% for the Guamirim cultivar. The graphs showed more concentration of individuals that resulted in dead seeds on the negative scores of PC1 and viable seeds kept on the positive scores of PC1 for the Guamirim cultivar. The low-vigor seeds were more distributed along dead and high-vigor seeds for the BR 18 cultivar than the low-vigor seeds from the Guamirim cultivar. The latent variable graphs showed the same pattern, which could not separate between 3 classes satisfactorily (Figure 21).

Figure 21 - Principal component analysis using FT-NIR data BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay and growth test results. Latent variables from the model used to classify seeds of BR 18 (C) and Guamirim (D) cultivars according to growth test results.

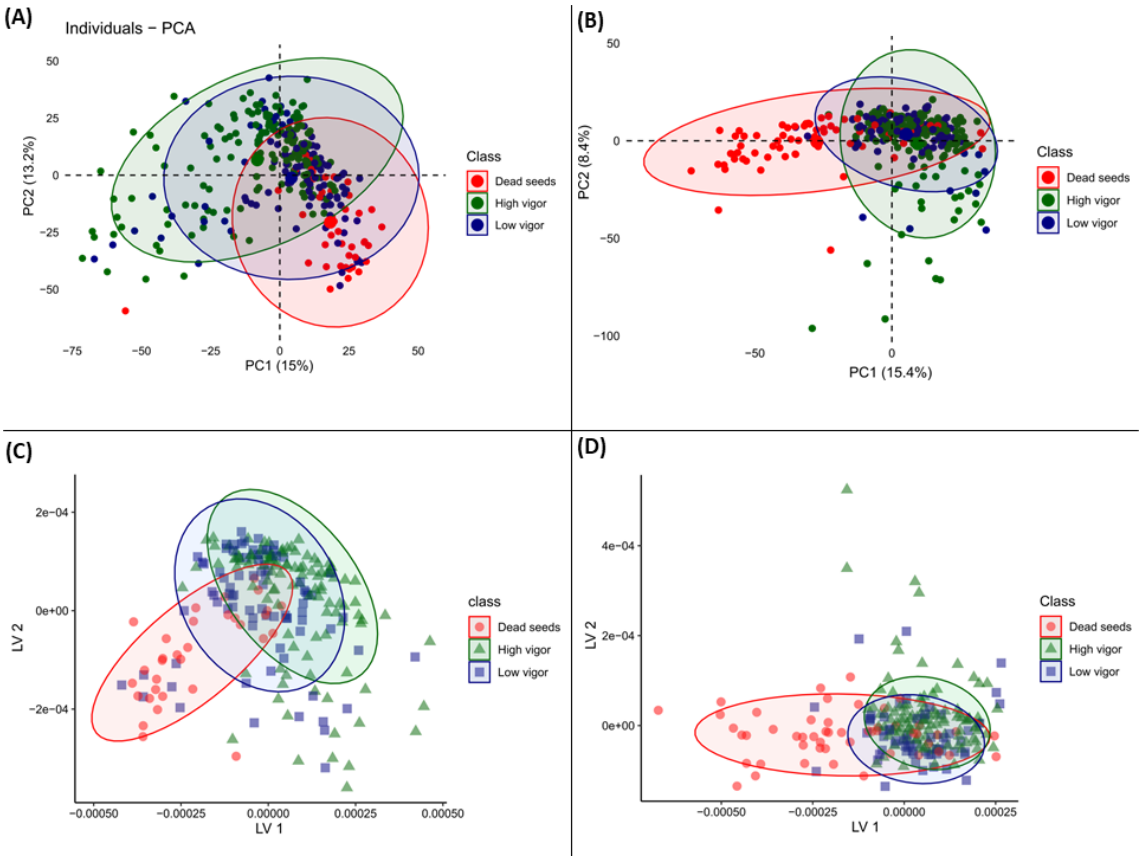


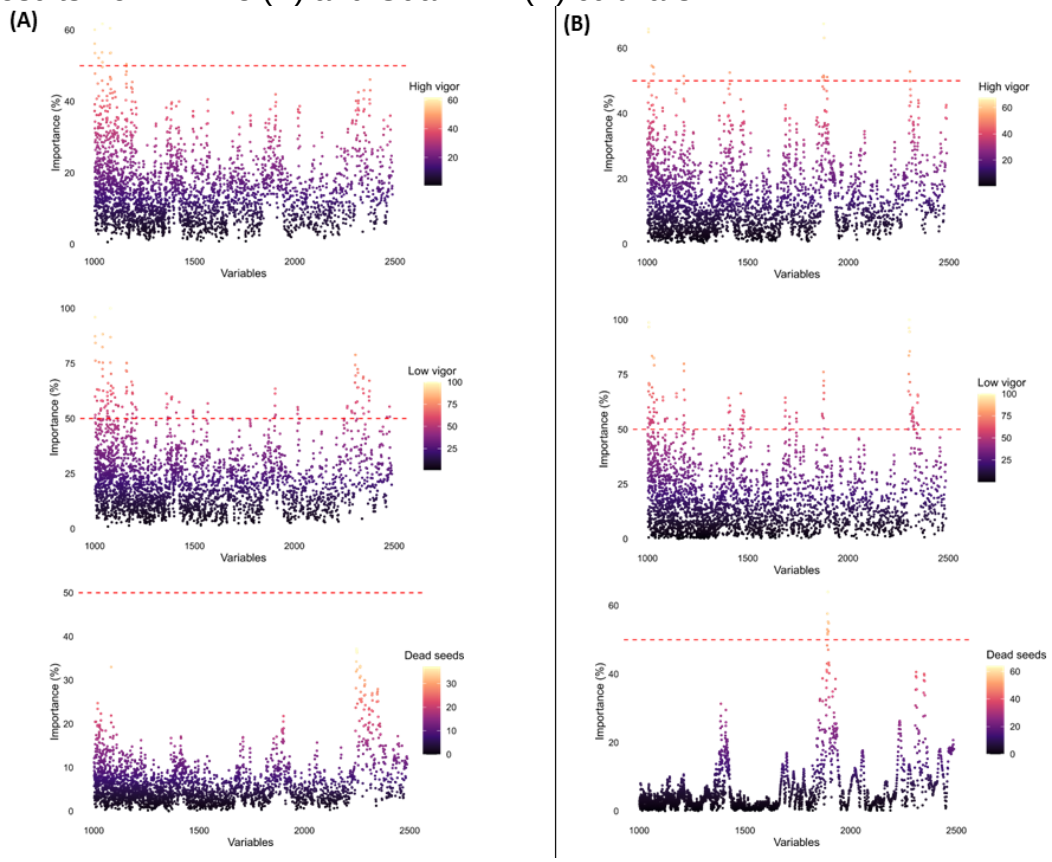
Table 12 shows the chemometrics results from the machine learning model tested to classify the seeds according to the classes from seedling growth test. This model was evaluated considering the accuracy, sensitivity, specificity, precision, mean F1, and kappa index. All these indices indicated a strong overfitting of the model since it performed very well with the training dataset but drastically decreased the performance with new data. This result can be associated with the worst performance for identifying the low vigor class in the testing dataset, especially in the BR18 cultivar. However, the specificity, which measures the proportion of true negatives that are correctly identified by the model, was better than the other metrics (Table 12).

Table 12 - Hit number per class and metrics obtained with the PLS-DA classification model, according to growth test results, using the FT-NIR resources of wheat seeds from two different cultivars. The number in parenthesis means the occurrence of the class in the dataset.

Class	BR 18		Guamirim	
	Train	Test	Train	Test
	Hits (Total)			
High vigor	112 (112)	31 (48)	106 (112)	42 (48)
Low vigor	78 (81)	9 (34)	34 (45)	3 (26)
Dead seeds	29 (32)	8 (13)	30 (51)	18 (21)
Accuracy	0.973	0.505	0.756	0.663
Sensitivity	0.956	0.509	0.694	0.616
Specificity	0.990	0.712	0.851	0.794
Precision	0.959	0.534	0.795	0.615
F1-Score	0.957	0.519	0.723	0.596
Kappa	0.956	0.155	0.583	0.418

The importance of variables graphs indicates several regions that reach more than 50% importance to discriminate between classes. The spectral regions of 1000, 1400, 1900, and 2300 nm were more important to separate the classes for both cultivars. Interestingly, the low vigor class showed more frequency (greater than 50%) throughout the whole spectra, and this could have bothered the separation from other classes since it had peaked in all regions (Figure 22).

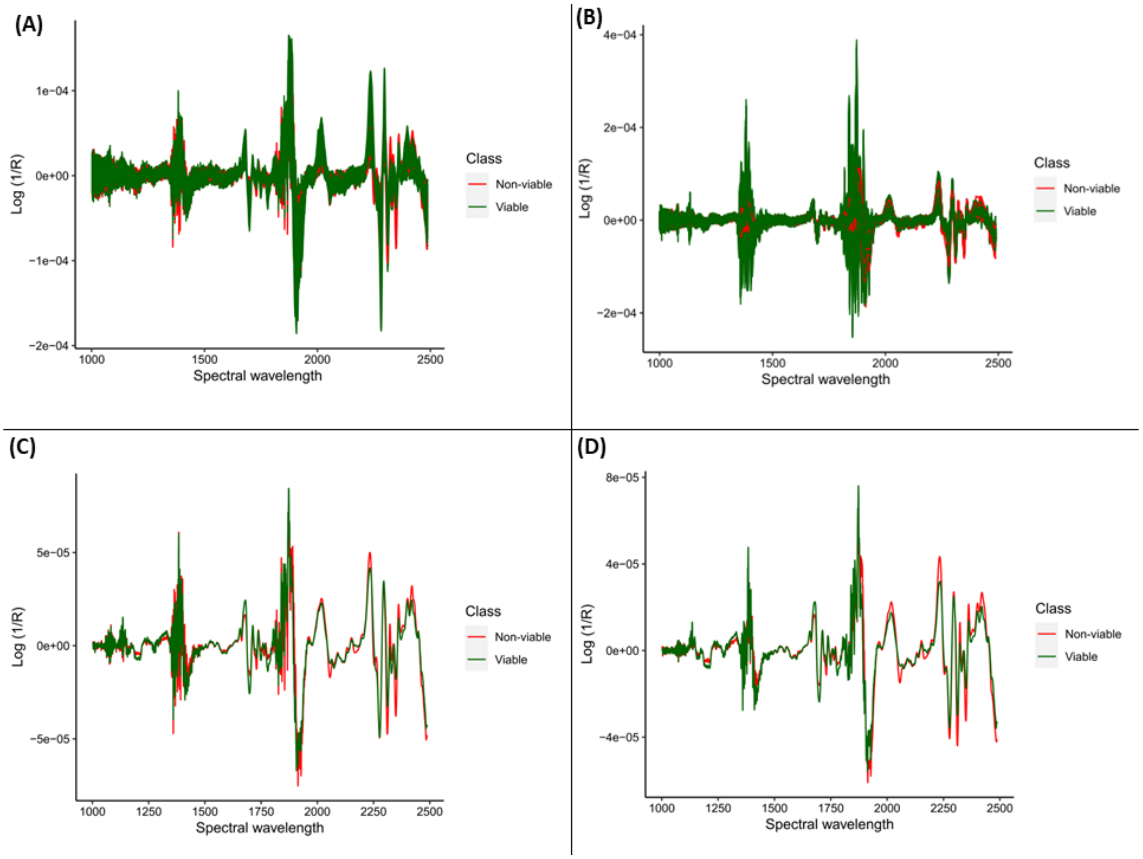
Figure 22 - Importance of variables used to classify seeds according to the growth test results from BR 18 (A) and Guamirim (B) cultivars



2.4.2.3. Emergence

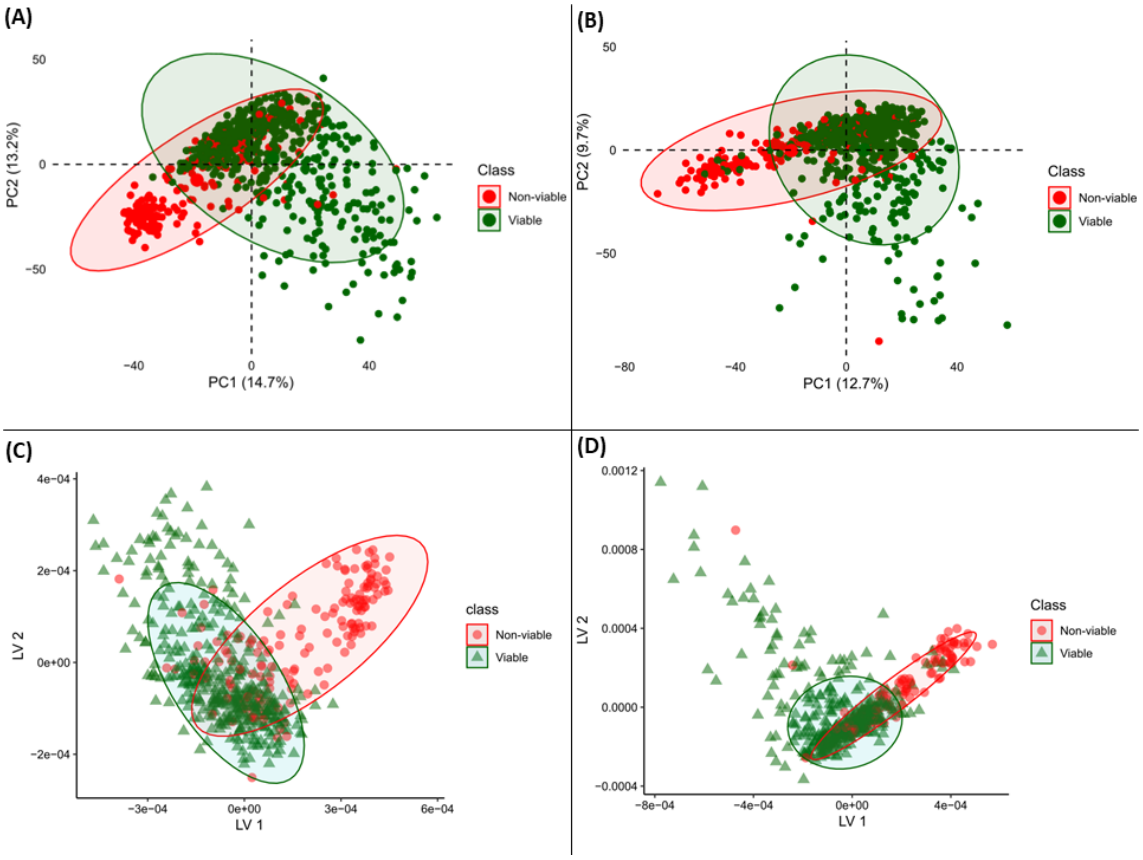
The spectra and averaged spectra of BR18 and Guamirim cultivars are shown in Figure 23. Through the emergence test, the seeds were classified as viable or non-viable after 15 days of running the test under greenhouse conditions. We observed in the spectra that the signal intensity was stronger for non-viable seeds than for the viable ones, especially with the averaged spectra graphs. By visual observation, we noted peaks in the range of 2300-2450 that may be used to separate the classes. These may be related to the absorption of energy by seed tissue.

Figure 23 - FT-NIR absorption spectra of wheat seeds classified according to the emergence test results. Spectra of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay; Average spectra of BR 18 (C) and Guamirim (D) cultivar seeds transformed by the second derivative of the Savitzky–Golay.



The PCA for BR 18 and Guamirim cultivars explained 27.9 and 22.4 of the variation, respectively, using only two principal components. The two components are represented in Figure 24 A and B. Unlikely for the other physiological tests, we observed a more apparent separation between the classes even using only two components. Similarly, the LDA graphs showed more distinction between classes represented by the green and red circles (Figure 24 C and D).

Figure 24 - Principal component analysis using FT-NIR data of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay and emergence test results. Latent variables from the model used to classify seeds of BR 18 (C) and Guamirim (D) cultivars according to emergence test results.



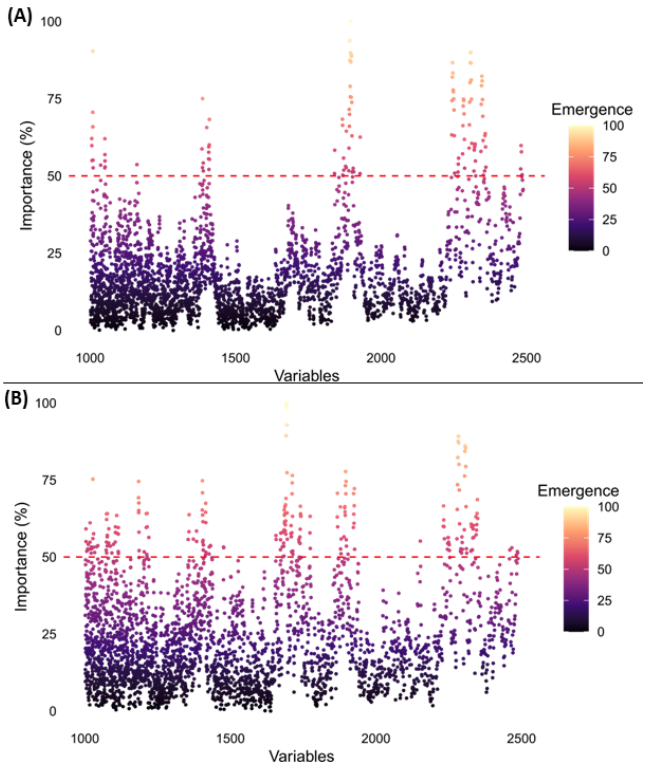
The results from the PLS-DA classification are shown in Table 13. As with the other physiological test results, these data sets are unbalanced, which requires careful evaluation of the metrics. Using the data from the emergence test was not evident an overfitting of the model, suggesting that it could be used to separate viable from non-viable seeds. The viable seeds were nearly 2.5 times the non-viable ones. The accuracy and specificity of the model were greater than 87% in both train and test datasets for BR 18 and Guamirim cultivars. However, there was a decrease in the sensitivity of the model for both cultivars, reaching 82% as the highest level. The Kappa coefficients were near 80% for training and 70% for the test.

Table 13 - Hit number per class and metrics obtained with the PLS-DA classification model, according to emergence test results, using the FT-NIR resources of wheat seeds from two different cultivars. The number in parenthesis means the occurrence of the class in the dataset.

Class	BR 18		Guamirim	
	Train	Test	Train	Test
	Hits (Total)			
Viable	396 (403)	167 (172)	397 (409)	166 (175)
Non-viable	124 (158)	42 (67)	111 (134)	42 (57)
Accuracy	0.930	0.874	0.936	0.900
Sensitivity	0.785	0.627	0.828	0.738
Specificity	0.982	0.971	0.971	0.949
Precision	0.947	0.894	0.902	0.824
F1-Score	0.856	0.737	0.864	0.777
Kappa	0.809	0.658	0.822	0.711

The most important variables used to separate between classes were close to 1000, 1400, 1900, and 2400 nm for both cultivars. The Guamirim also showed a significant peak (above 50%) at 1700 nm (Figure 11).

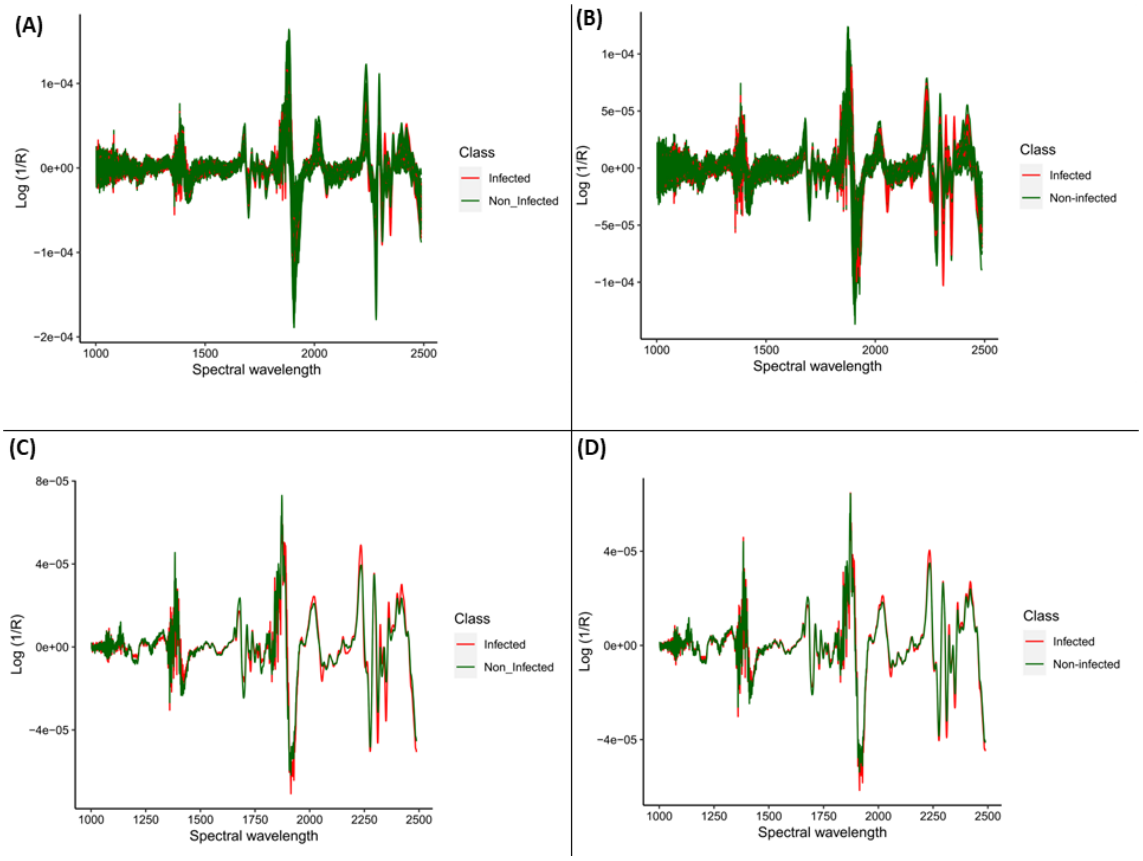
Figure 25 - Importance of variables used to classify seeds according to the emergence test results from BR 18 (A) and Guamirim (B) cultivars



2.4.2.4. Blotter test

The blotter test results were used to classify the seeds as infected or non-infected. Similarly to the emergence results, we could identify some regions with different patterns between the two classes. The mean data graphs of the two cultivars were very similar, except for the stronger signal of infected seeds in the range of 2350-2400 nm for the BR18 cultivar (Figure 26).

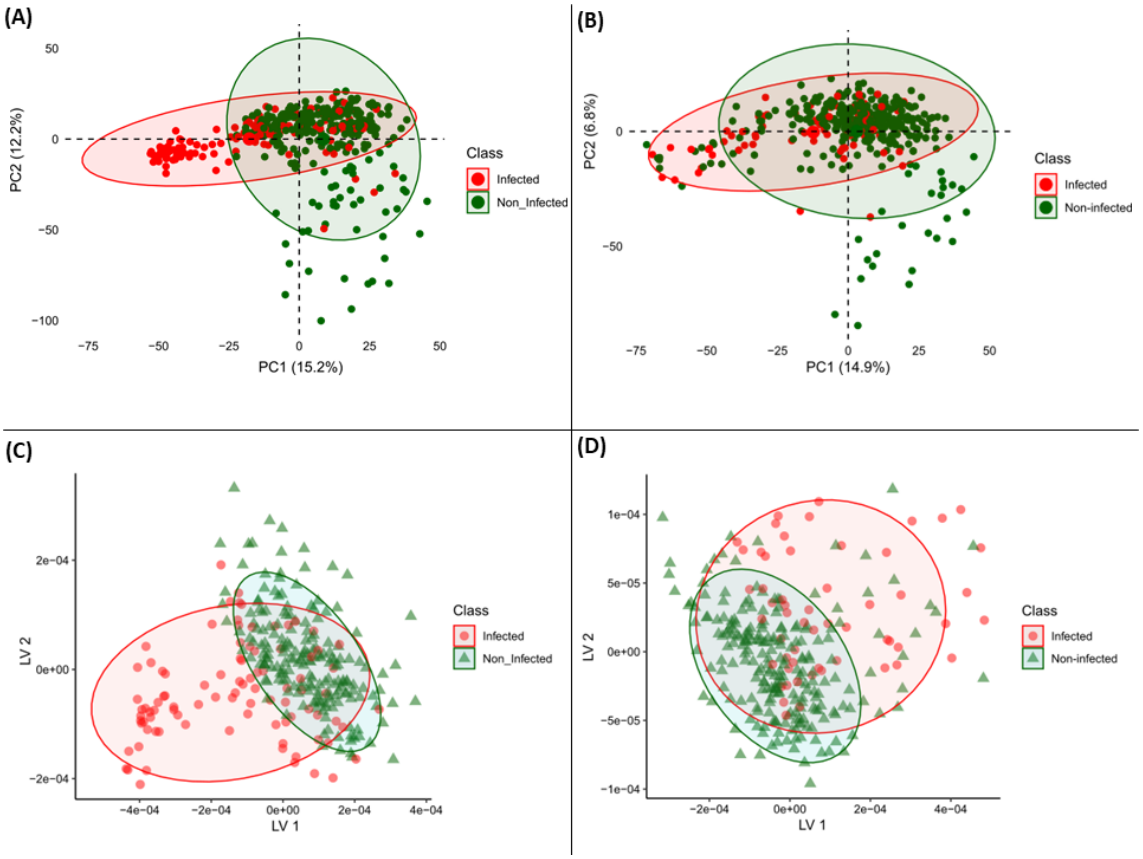
Figure 26 - FT-NIR absorption spectra of wheat seeds classified according to the Blotter test results. Spectra of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay; Average spectra of BR 18 (C) and Guamirim (D) cultivar seeds transformed by the second derivative of the Savitzky–Golay.



The two principal components represented 27.4% and 21.7% of the variation in data from BR18 and Guamirim cultivars, respectively. We observed a more apparent separation between classes from BR18 than Guamirim (Figure 27 A and B). However,

as stated before, the two components were not significant in showing class separation. These results suggest that the detection by the NIR instrument of infected seeds could vary from cultivar. The LDA graphs followed the same pattern of PCA (Figure 27 C and D).

Figure 27 - Principal component analysis using FT-NIR data of BR 18 (A) and Guamirim (B) cultivar seeds transformed by the second derivative of the Savitzky–Golay and Blotter test results. Latent variables from the model used to classify seeds of BR 18 (C) and Guamirim (D) cultivars according to Blotter test results.



As stated for the physiological test results, the blotter dataset was also strongly unbalanced. The non-infected class was twice more represented than the infected for BR 18 and about four times for the Guamirim cultivar. The model more efficiently detected non-infected seeds in both cultivars for training or test datasets. The evaluated metrics showed above 87% for accuracy and specificity. The sensitivity reached 85% for the train and test datasets in BR18 and 88% for training, and 58% for the test in Guamirim. Similar results were observed either for Mean F1 and precision, which reached above 90% for training in both cultivars and 86% for test in BR18 and 64% for Guamirim. This result suggests an overfitting of the model for the Guamirim cultivar. The Kappa showed a better performance of the model for BR18, even though

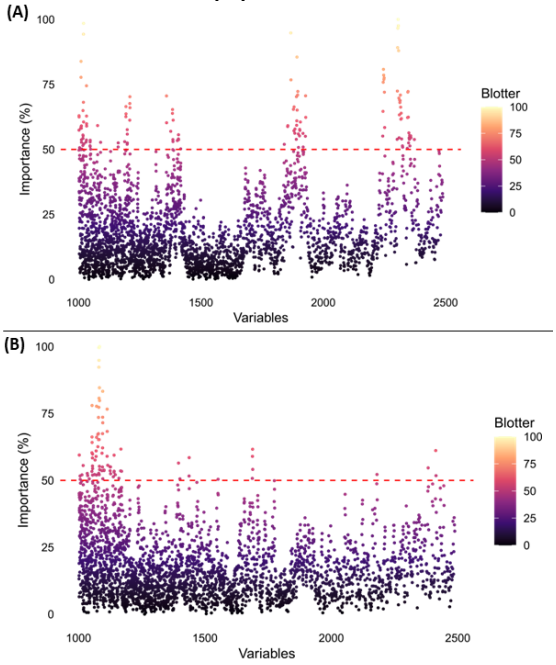
it was near 86% for training and 79% for the test, than the Guamirim cultivar, which showed above 90% for training but a poor result in the test dataset (56%) (Table 14).

Table 14 - Hit number per class and metrics obtained with the PLS-DA classification model, according to Blotter test results, using the FT-NIR resources of wheat seeds from two different cultivars. The number in parenthesis means the occurrence of the class in the dataset.

Class	BR 18		Guamirim	
	Train	Test	Train	Test
	Hits (Total)			
Infected	78 (92)	33 (39)	49 (56)	14 (24)
Non-infected	186 (189)	75 (80)	223 (224)	90 (96)
Accuracy	0.940	0.908	0.971	0.867
Sensitivity	0.848	0.846	0.875	0.583
Specificity	0.984	0.938	0.996	0.938
Precision	0.963	0.868	0.980	0.700
F1-Score	0.902	0.857	0.925	0.636
Kappa	0.858	0.789	0.907	0.556

The graphs showing the importance of variables used by the model to differentiate between classes are shown in Figure 28. The BR18 cultivar showed more clear regions above 50% of importance than Guamirim cultivar, which agrees with the other results from the Blotter test. Based on this graph, the regions of 1000, 1250, 1400, 1900, and 2400nm were identified as more important to differentiate seeds infected from non-infected for BR18. In these wavelengths, the variables reached 75% of importance (Figure 28 A). On the other hand, the Guamirim showed fewer points above 50% of importance, in the regions of 1000, 1400, 2250, and 2400 nm (Figure 28 B).

Figure 28 - Importance of variables used to classify seeds according to the Blotter test results from BR 18 (A) and Guamirim (B) cultivars.



2.5. Discussion

In this work, it was studied the ability of NIR spectroscopy associated with the PLS-DA algorithm to assess the quality of wheat seeds harvested from artificially infected plants by *Fusarium graminearum* in different growth stages. As reported in this work and by other authors (Alisaac and Mahlein, 2023), when the infection occurs at anthesis, it drastically reduces the percentage of viable seeds. This experiment found the highest rate of dead seeds and the most abnormal seedlings for the infection at the anthesis. In fact, at this stage, most of the wheat would be discarded during the harvest and cleaning process. However, as stated by Siou et al. (2014), separating infected from non-infected seeds is difficult at late infection since there were no differences in their density or shape.

Similar results were observed by Del Ponte et al. (2007), showing that the incidence of contaminated grains reduced from the first application time (complete anthesis) to the early dough stage. It did not affect the weight of a thousand grains when compared to the control (without inoculation). Also, a different trend was observed at late stages (ZS 73 – ZS 83), where the frequencies of the seed quality increase with late infection. At ZS 73, corresponding to the early milk stage, the seeds were more susceptible to infection than ZS 83 (early dough) since the pathogen could

easily penetrate the pericarp and infect the embryonic axis without affecting the seed formation.

The fungus uses enzymes that degrade the seed cell wall, allowing the host to penetrate, grow and infect the plant tissue (Hassani et al., 2019). This compromises the efficiency of the cleaning process to separate the infected seeds. Therefore, the infection of seeds can impact the quality of the lot by decreasing the seed germination and vigor (Figures 15 and 16). In field production, these may be related to damping off and death of seedlings (Asran and Amal, 2011). The Blotter results confirmed that there were more infected seeds at ZS 73, in comparison to the infection at ZS 83 and with the control. Also, it was observed that the pathogen penetrated in the seed tissue since the test was conducted with surface-sterilized seeds. In this study, it was observed that seeds without visible symptoms of Fusarium Head Blight (FHB) could still be infected. While identifying asymptomatic infected seeds was not the primary objective, this observation highlights the limitations of visual separation methods. Therefore, traditional cleaning machines would not be effective and NIR spectroscopy emerged as a potentially valuable tool for this type of separation.

The NIR spectroscopy data associated with machine learning algorithms to identify FHB in seed lots have been studied (Dowell et al., 1999; Peiris et al., 2010). However, the main objectives were to use this technique in line with the grinding process in the mills to reduce the waiting time to get reliable results about the quality of the grain, to be acceptable or not for the industry. Dowell et al. (1999) studied the feasibility of NIR as an objective method for determining wheat seeds with FHB symptoms, DON, and ergosterol levels. The authors visually separated contaminated from non-contaminated seed lots and proceeded to read individual seeds in the wavelength range between 400 and 1700 nm. The spectra were subjected to PLS regression analysis. They showed that NIR was more efficient than human classification in separating infected from non-infected seeds and estimating DON and Ergosterol levels in individual seeds.

This study focused on seed quality parameters used in the commercialization. Therefore, the classes of datasets were the standard seed quality tests used by specialized laboratories for seed quality determination. The results showed that the infection might be associated with a negative impact on germination and vigor of the seeds, and it was detected by the NIR combined with the machine learning approach. The impacts on seed germination and vigor due to increased Fusarium-infected seeds

are reported by numerous studies. Hassani et al. (2019) observed a significant increase in abnormal seedlings and a decrease in shoot and root length in samples that contained more infected seeds. The pathogen was also associated with the worst indices of mean seed germination by Browne and Cooke (2005) and had a significant negative relationship with seed weight and germination (Browne, 2007).

The graphs representing the two main principal components did not clearly separate classes (Figs. 18, 21, 24, and 27). The two PCs applied to the SG transformed data were used to try to identify patterns in the dataset. However, these were limited by the 2D representation and the difficulty of representing many components in several dimensions. Instead, the PCA could be used as a dimensional reduction method to evaluate NIR spectra before submitting the data to ML models, as stated by Fan et al. (2020), with some improvements in the accuracy of the models. This study showed no evident improvement in the metrics using PCA (data not shown). Therefore, we used the whole spectra data to train and test the models.

A strongly imbalanced data between classes for all test results was evident, which is commonly observed in nature. The intermediate classes (abnormal seedlings from the germination test and low vigor from the growth test) were associated with a substantial rate of errors that may be related to the cutoff point that separates the classes. This could explain the lowest hit rates in these classes (Tables 11 and 12). Fan et al. (2020) also observed a misclassification portion of the intermediate class, suggesting the need for a robust classification model. However, the model detected most normal and vigorous seedlings correctly, which is required in practical applications.

The most important wavelengths for the model to separate between classes were in the range of 1100 - 1200, 1300 - 1400, 1700 - 1900, and 2200 - 2300. The region of 1100 - 1200 corresponds to the second overtone of carbohydrates, and it is considered a stable region that could be used to separate infected and non-infected seeds (Delwiche et al. 2004). The authors found that *Fusarium* infection strongly impacted the spectral range of 1130-1190 nm, with a peak at 1158 nm. Peiris et al. (2010) indicated that the region of 1300 - 1400 is associated with interactions between the water molecule and organic molecules. In fact, this region corresponds to the first overtone of -OH and -CH bonds and other functional groups (–CH, –CH2, –CH3, ArOH, and ROH). Delwiche et al. (2004) also found that this spectral range was affected by *Fusarium* infection with a remarkable peak at 1400 nm.

The pathogen may cause significant alterations in the pool of carbohydrates, lipids, and proteins that affect the vigor and germination of the seeds. The spectral region of 1700 - 1900 has been reported to be the first overtone -NH and -OH vibrations, which is related to chemical bonds of water and proteins (Fan et al. 2020). Mishra et al. (2017) established an important association between the peaks at 2200 - 2300 nm with the reduction of carbohydrates content, which could be associated with the reduction of seed weight.

The use of NIR and machine learning techniques for evaluating wheat seeds can provide fast and accurate results, especially in breeding cultivars resistant to FHB. The method can accurately separate contaminated from non-contaminated seeds, preventing the transmission of pathogens to new areas. It also efficiently determines seed viability, allowing for the commercialization of higher-quality lots. However, the influence of intermediate classes in the germination and growth test requires further research using more robust datasets and methodologies such as the PLS regression.

2.6. Conclusions

This study investigated the use of NIR spectroscopy to predict the physiological and sanitary quality of wheat seeds. The proposed models had a global accuracy of 84%, 87%, 87%, and 50% for germination, emergence, blotter test, and seedling growth, respectively. It was observed an overfitting of the model regarding the seedling growth test. The presence of the pathogen significantly impacted the wheat seeds, as revealed by the physiological and NIR spectral data analyzed with the PLS-DA algorithm. The spectral regions of 1000-1200, 1400-1600, 1900-2200, and 2400 were more important in distinguishing between classes.

2.7. References

Alisaac, E., Behmann, J., Rathgeb, A., Karlovsky, P., Dehne, H. W., Mahlein, A. K. (2019). Assessment of *Fusarium* infection and mycotoxin contamination of wheat kernels and flour using hyperspectral imaging. *Toxins*, *11*(10), 1–18. https://doi.org/10.3390/toxins11100556

Alisaac, E., Rathgeb, A., Karlovsky, P., Mahlein, A. K. (2021). Fusarium head blight: Effect of infection timing on spread of *Fusarium graminearum* and spatial distribution of deoxynivalenol within wheat spikes. *Microorganisms*, 9(1), 1–12. https://doi.org/10.3390/microorganisms9010079

Alisaac, E., Mahlein, A. K. (2023). *Fusarium* Head Blight on Wheat: Biology, modern detection and diagnosis and integrated disease management. *Toxins*, *15*(3), 192. https://doi.org/10.3390/toxins15030192

Amer, H., Sahi, S. T., Javed, N., Ahmad, S. (2011). Prevalence of seed-borne fungi on wheat during storage and its impact on seed germination. *Pakistan Journal of Phytopathology*, *23*(1), 42–47.

Asran, M. R., Amal, M. I. E. (2011). Aggressiveness of Certain *Fusarium graminearum* isolates on wheat seedlings and relation with their trichothecene production. *Plant Pathology Journal*, *10*(1), 36–41. https://doi.org/10.3923/ppj.2011.36.41

BRASIL (2009). Ministério da Agricultura. Manual de análise sanitária de sementes. Brasília, LANARV/SNDA/MA. 200 p.

Browne, R. A., Cooke, B. M. (2005). A comparative assessment of potential components of partial disease resistance to Fusarium head blight using a detached leaf assay of wheat, barley and oats. *European Journal of Plant Pathology*, 112(3), 247–258. https://doi.org/10.1007/s10658-005-2077-z

Browne, R. A. (2007). Components of resistance to fusarium head blight (FHB) in wheat detected in a seed-germination assay with *Microdochium majus* and the relationship to FHB disease development and mycotoxin accumulation from *Fusarium graminearum* infection*. Plant Pathology*, 56(1), 65–72. https://doi.org/10.1111/j.1365-3059.2006.01509.x

CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. Trigo: Análise Mensal - Junho 2024 – safra 2023/2024. Brasília: *Companhia Nacional de Abastecimento*. 2022. Available in: www.conab.gov.br. Accessed on August 5, 2024.

Del Ponte, Emerson M., Fernandes, J. M. C., Pierobom, C. R., & Bergstrom, G. C. (2004). Giberela do trigo: aspectos epidemiológicos e modelos de previsão. *Fitopatologia Brasileira*, 29(6), 587–605. https://doi.org/10.1590/s0100-41582004000600001

Del Ponte, E. M., Fernandes, J. M. C., & Bergstrom, G. C. (2007). Influence of growth stage on fusarium head blight and deoxynivalenol production in wheat. *Journal of Phytopathology*, *155*(10), 577–581. https://doi.org/10.1111/j.1439-0434.2007.01281.x

Delwiche, S. R.; Hareland, G. A.(2004) Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. *Cereal Chemistry*, 81: 643-649. https://doi.org/10.1094/CCHEM.2004.81.5.643

Dowell, F.E., Ram, M.S. & Seitz, L.M. (1999), Predicting scab, vomitoxin, and ergosterol in single wheat kernels using near-infrared spectroscopy. *Cereal Chemistry*, 76: 573-576. https://doi.org/10.1094/CCHEM.1999.76.4.573

Fan, Y., Ma, S., & Wu, T. (2020). Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Physics and Technology*, *105*(1), 103213. https://doi.org/10.1016/j.infrared.2020.103213

Garcia Júnior, D., Vechiato, M. H., Menten, J. O. M., & Lima, M. I. P. M. (2007). Influência de *Fusarium graminearum* na germinação de genótipos de trigo (*Triticum aestivum* L.). *Arquivos Do Instituto Biológico*, 74(2), 157–162.

Gilbert, J., Tekauz, A., & Woods, S. M. (1997). Effect of storage on viability of fusarium head blight-affected spring wheat seed*. Plant Disease*, 81(2), 159–162. https://doi.org/10.1094/PDIS.1997.81.2.159

Hassani, F., Zare, L., Khaledi, N. (2019). Evaluation of germination and vigor indices associated with *Fusarium*-infected seeds in pre-basic seeds wheat fields. *Journal of Plant Protection Research*, *59*(1), 69–85. https://doi.org/10.24425/jppr.2019.126037

INTERNATIONAL SEED TESTING ASSOCIATION. *Handbook on Seedling Evaluation*. 4. ed. Bassersdorf: International Seed Testing Association, 2018.

Khaneghah, A. M., Martins, L. M., von Hertwig, A. M., Bertoldo, R., Sant'Ana, A. S. (2018). Deoxynivalenol and its masked forms: Characteristics, incidence, control and fate during wheat and wheat based products processing - A review. *Trends in Food Science & Technology*, *71*, 13–24. https://doi.org/https://doi.org/10.1016/j.tifs.2017.10.012

Liang, K., Huang, J., He, R., Wang, Q., Chai, Y., Shen, M. (2020). Comparison of Vis-NIR and SWIR hyperspectral imaging for the non-destructive detection of DON levels in fusarium head blight wheat kernels and wheat flour. *Infrared Physics and Technology*, *106*(40), 103281. https://doi.org/10.1016/j.infrared.2020.103281

MAPA – MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO -. Padrões para a produção e a comercialização de sementes de trigo (*Triticum aestivum* L.). Available in: https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/publicacoes-sementes-e-mudas/copy_of_INN45de17desetembrode2013.pdf . Acessed on April 15, 2024.

Mishra, G., Srivastava, S., Panda, B. K., Mishra, H. N. (2018). Rapid assessment of quality change and insect infestation in stored wheat grain using FT-NIR spectroscopy and chemometrics. *Food analytical methods*, 11, 1189-1198.

Peiris, K., Pumphrey, M., Dong, Y., Maghirang, E., Berzonsky, W., Dowell, F. (2010). Near-infrared spectroscopic method for identification of fusarium head blight damage and prediction of deoxynivalenol in single wheat kernels. *Cereal Chemistry*, *87*(6), 511–517.

Scheeren, P. L., Caierão, E., Só e Silva, M., Del Duca, L.J. A., Nascimento Junior, A. Linhares, A., Eichelberger, L. (2007). BRS Guamirim: cultivar de trigo da classe pão, precoce e de baixa estatura." Pesquisa Agropecuária Brasileira 42 (2): 293–96. https://doi.org/10.1590/s0100-204x2007000200020.

Silva, L. J. D., Medeiros, A. D. D., Oliveira, A. M. S. (2019). SeedCalc, a new automated R software tool for germination and seedling length data processing. *Journal of Seed Science*, *41*, 250-257. https://doi.org/10.1590/2317-1545v42n2217267.

Shahin, M. A., Symons, S. J. (2011). Detection of *Fusarium* damaged kernels in Canada western red spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. *Computers and Electronics in Agriculture*, *75*(1), 107–112. https://doi.org/10.1016/j.compag.2010.10.004

Singh, C. B., Jayas, D. S., Paliwal, J., White, N. D. G. (2012). Fungal damage detection in wheat using short-wave near-infrared hyperspectral and digital colour imaging. *International Journal of Food Properties*, *15*(1), 11–24. https://doi.org/10.1080/10942911003687223

Siou, D., Gélisse, S., Laval, V., Repinçay, C., Canalès, R., Suffert, F., Lannou, C. (2014). Effect of wheat spike infection timing on fusarium head blight development and mycotoxin accumulation. *Plant Pathology*, *63*(2), 390–399. https://doi.org/10.1111/ppa.12106

Sousa, P. G.. (2002). BR 18-Terena: cultivar de trigo para o Brasil. Pesquisa Agropecuária Brasileira, 37(7), 1039–1043. https://doi.org/10.1590/S0100-204X2002000700019

Xia, Y., Xu, Y., Li, J., Zhang, C., Fan, S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, *1*, 35–47. https://doi.org/10.1016/j.aiia.2019.05.001

Zadoks, J. C., Chang, T. T., Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, *14*(6), 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

Zhang, T., Wei, W., Zhao, B., Wang, R., Li, M., Yang, L., Wang, J., Sun, Q. (2018). A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. *Sensors (Switzerland)*, *18*(3). https://doi.org/10.3390/s18030813

3. GENERAL CONCLUSIONS

It was concluded that the representation of each category in the dataset significantly influences the performance of the models and it is crucial to consider this to prevent misinterpretation of results, especially when dealing with imbalanced datasets common in seed analysis. In assessing seed viability, better results were obtained using the SG_D1_W11 pre-treatment for BR18 and the SNV_SG for Guamirim. When evaluating the vigor of the BR18 cultivar, better performance was observed with the SG_D1, SNV_SG, and SNV pre-treatments for growth, emergence, and blotter tests, respectively. For the Guamirim cultivar, the SNV_SG, SG_D1, and D1 pre-treatments were particularly effective for growth, emergence, and blotter tests, respectively. Using the second derivative of Savitzky-Golay to pre-process the data had a global accuracy of 84%, 87%, 87%, and 50% for germination, emergence, blotter, and seedling growth, respectively. The model performance did not differ significantly for each cultivar, except for the BR18 cultivar in the seedling growth test, which showed an overfitting. The presence of the pathogen significantly impacted the wheat seeds, as revealed by NIR spectral data analyzed with the PLS-DA algorithm. The spectral regions of 1000-1200, 1400-1600, 1900-2200, and 2400 were more important in distinguishing between classes. The NIR constitutes an efficient tool for seed quality control that requires rapid and precise results.