



OPEN ACCESS

EDITED BY

Pengchao Chen,
South China Agricultural University, China

REVIEWED BY

Yasin Kaya,
Adana Alparslan Turkes Science and
Technology University, Türkiye
Moisés Roberto Vallejo-Pérez,
Autonomous University of San Luis Potosí,
Mexico

*CORRESPONDENCE

Maurício Rossato
✉ mauricio.rossato@unb.br

RECEIVED 17 September 2025

REVISED 26 December 2025

ACCEPTED 29 December 2025

PUBLISHED 26 January 2026

CITATION

Carvalho ICB, Ferreira LdC, Neves ARdM,
Carvalho AMS, Lima HPR and Rossato M
(2026) A hyperspectral imaging and machine
learning approach for rapid and non-invasive
diagnosis of cassava bacterial blight.
Front. Plant Sci. 16:1707646.
doi: 10.3389/fpls.2025.1707646

COPYRIGHT

© 2026 Carvalho, Ferreira, Neves, Carvalho,
Lima and Rossato. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A hyperspectral imaging and machine learning approach for rapid and non-invasive diagnosis of cassava bacterial blight

Ian Carlos Bispo Carvalho¹, Luciellen da Costa Ferreira¹,
Ana Régia de Mendonça Neves², Alice Maria Silva Carvalho¹,
Henrique Póvoa Rodrigues Lima¹ and Maurício Rossato^{1*}

¹Universidade de Brasília, Brasília, Distrito Federal, Brazil, ²Instituto Federal de Brasília (IFB), Eixo de
Informação e Comunicação, Distrito Federal, Brasília, Brazil

This study explores the use of hyperspectral imaging (HSI) combined with machine learning to detect physiological alterations in cassava leaves caused by *Xanthomonas phaseoli* pv. *manihotis* (Xpm), a bacterial plant disease that causes significant yield losses worldwide. Therefore, the use of hyperspectral images associated with machine learning can provide information rapidly and accurately, aiming to support decision-making. HSI captures spectral data that reflects biochemical changes in infected plant tissues. An image set of cassava healthy and symptomatic leaves (402 and 450, respectively) were imaged using a hyperspectral camera across wavelengths from 400 to 1000 nm, with image calibration and spectral normalization to improve data quality. Spectral parameters, such as mean reflectance and spectral differences (healthy vs. infected), were analyzed. Six machine learning models were tested for classification: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP). SVM performed best, achieving the highest accuracy (91.41%), followed by MLP (87.89%), XGBoost (79.69%), and RF (77.34%). DT and KNN had the lowest accuracy (71.88% and 70.31%, respectively). The results suggest that HSI, particularly when combined with SVM, offers a rapid and accurate method for diagnosing cassava bacterial blight, with potential for large-scale field applications.

KEYWORDS

bacterial blight, HSI, plant disease, plant physiology, *Xanthomonas*

1 Introduction

Cassava bacterial blight, caused by *Xanthomonas phaseoli* pv. *manihotis* (Xpm), is a major phytosanitary challenge that threatens cassava production globally. Occurring in nearly 50 countries (Taylor et al., 2017; Zárate-Chaves et al., 2021), this disease causes yield losses ranging from 30% to 90%. Without proper management, complete crop failure can occur within two to three growing cycles (Mansfield et al., 2012; Zárate-Chaves et al., 2021). Symptoms include translucent, water-soaked leaf spots that progress to necrosis and may cause wilting in cases of systemic infection.

Successful disease management depends on the grower's ability to detect outbreaks in the field, thereby anticipating and mitigating the deleterious effects of the phytopathogenic agents. In recent decades, considerable progress has been made in developing non-invasive techniques for plant disease diagnosis, such as fluorescence spectroscopy, VNIR spectroscopy, fluorescence imaging, and hyperspectral imaging (Sankaran et al., 2010; Golhani et al., 2018). Among these, hyperspectral sensors have gained prominence for their efficiency in extracting diverse types of information from plant tissues (Ortenberg, 2018). These techniques have found widespread application in agriculture, including seed quality analysis (Feng et al., 2019; Ferreira et al., 2024) and soil assessment (Demattè et al., 2010).

Despite the promise of hyperspectral imaging, challenges remain in selecting optimal wavelengths and scaling this technology for practical use in large agricultural settings. Addressing these challenges often requires the application of machine learning algorithms to extract meaningful insights from the spectral data and develop effective disease management strategies (Feng et al., 2021; Zhang et al., 2020).

Recent studies have highlighted the potential of spectral data and machine learning to understand the behavior of plants under pathogen infection across different pathosystems, such as tomato bacterial blight (Abdulridha et al., 2020), rice bacterial blight (Zhang et al., 2025), bacterial blight disease in red kidney beans (Qiao et al., 2025) and cassava brown streak disease (Peng et al., 2022). Machine learning, a key area within artificial intelligence, employs computational algorithms that learn from input data to perform various tasks, such as classification or clustering. This approach is particularly well-suited for identifying parameters and trends in hyperspectral data (Dhakal et al., 2023).

Machine learning techniques can be broadly categorized into supervised and unsupervised learning. In supervised learning, predictive models are trained using labeled data, where each data point is associated with a known "ground truth", either assigned by experts or verified experimentally. In contrast, unsupervised learning identifies patterns in unlabeled data, without predefined labels (Greener et al., 2022; Asnicar et al., 2024). Additionally, semi-supervised learning, which combines both labeled and unlabeled data, can optimize the process, especially when data labeling is costly (Greener et al., 2022). Various machine learning algorithms, such as Support Vector Machines (SVM), Random Forest (RF), have been widely used to build plant disease prediction models,

each offering different performances depending on the dataset and the application (Ahmad et al., 2023; Omaye et al., 2024).

Despite advances in this field, a gap remains in the application of these techniques for the detection of Xpm in cassava plants. Therefore, this study aims to employ an innovative approach that combines hyperspectral imaging with machine learning to identify cassava plants infected with Xpm, thereby improving the diagnosis and management of cassava bacterial blight.

2 Materials and methods

2.1 Xpm inoculation on cassava plants

The bacterial inoculum (UnB 17 isolate) was prepared by streaking the bacteria into 523 culture medium (Kado and Heskett, 1970) and incubated at 28°C for 48 hours in a growth chamber. Typical colonies were transferred to new plates for another incubation period. A bacterial suspension was subsequently prepared in distilled water, and its concentration was measured and adjusted using a Shimadzu UV-1203 spectrophotometer (wavelength of 550 nm and absorbance of 0.350) to obtain a bacterial concentration of 10^8 CFU/mL.

Variety BGMC 962 of cassava, commonly used in Brazil and considered susceptible to Xpm, was chosen for the experiment and propagated by cuttings. Eight-week-old cassava plants were inoculated by spraying the aerial parts with the bacterial suspension using a handheld sprayer until runoff. The inoculated plants were maintained in a moist chamber for three days and then transferred to a greenhouse with controlled temperature (26°C), where they were monitored for disease development over the study period. Mock inoculated plants were kept with the same conditions. Twenty days after inoculation, both healthy and symptomatic leaves from cassava plants, exhibiting varying degrees of disease severity, were harvested for hyperspectral imaging. The leaf harvesting started 20 days after inoculation and extended over a 30-day period. By the end of this period, a total of 852 leaves were collected, comprising 402 healthy leaves and 450 symptomatic leaves.

2.2 Image capture

The collection of hyperspectral imaging of cassava leaves was performed using an FX10e camera (Specim, Finland), capable of measuring reflectance within the 400–1000 nm spectral range, attached to a LabScanner (Specim, Finland) with six halogen lamps (Figure 1). Images were captured with Software Breeze v. 2024.1 (Prediktera, Sweden) also used for storage and initial processing of images. Each group of symptomatic leaves was distributed along the LabScanner tray for sample movement and proper hyperspectral image capture. An RGB image was also captured to serve as a reference for the normal appearance of the samples.

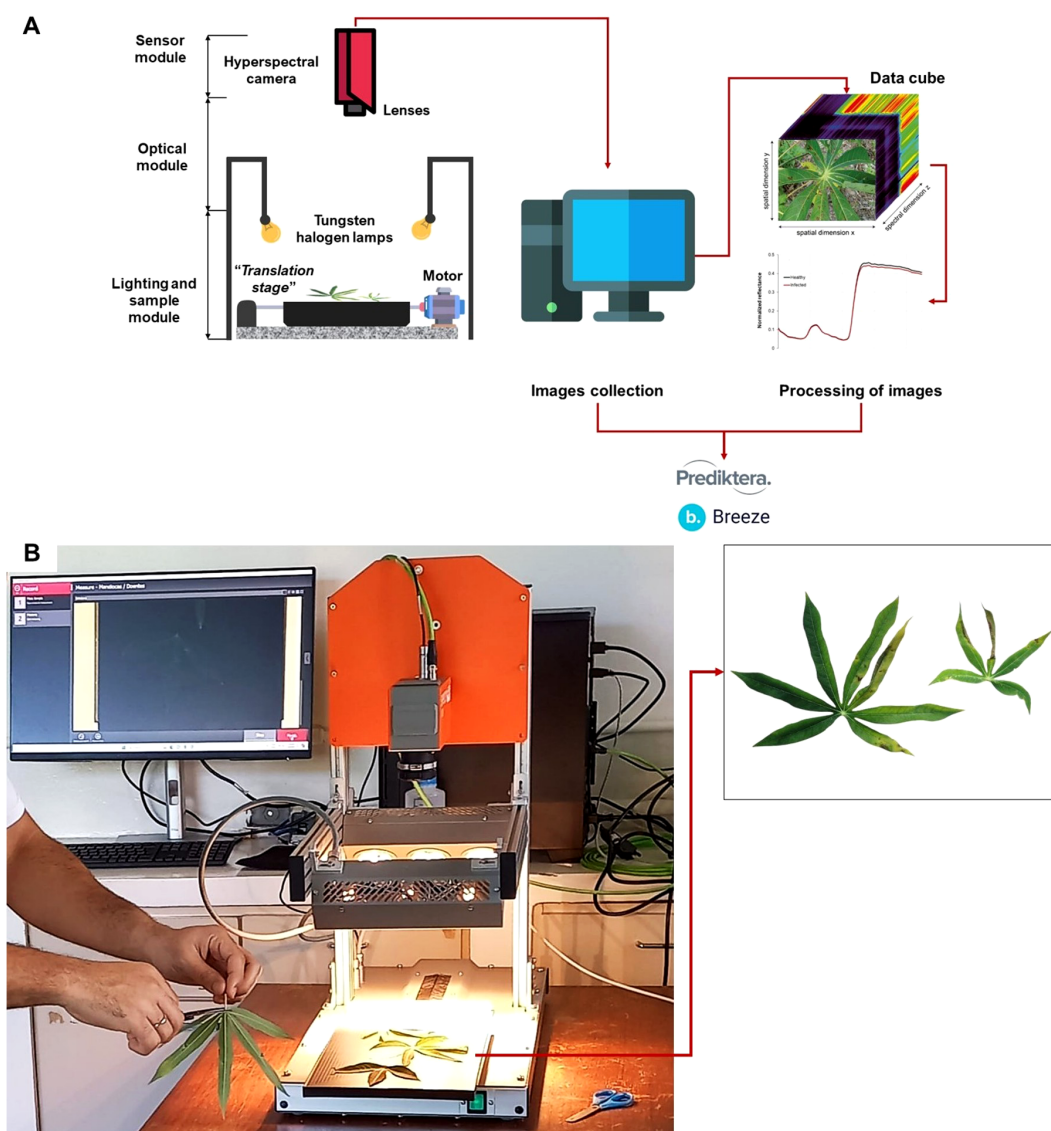


FIGURE 1

Schematic representation and setup of a benchtop hyperspectral imaging system for cassava leaf analysis. (A) Diagram of the hyperspectral imaging system components, including the FX10e camera, halogen lamps for illumination, and a motorized translation stage for sample movement. (B) Practical setup capturing cassava leaf samples under halogen lamp illumination.

Before beginning the image capture, reference images for white and black were obtained with the camera shutter closed. The imaging was performed across all spectral bands within the capabilities of the equipment used for this project. To extract the true spectral response of each sample, the influence of the black and white reference images was removed, resulting in a calibrated image (IR) (Kim et al., 2001).

2.3 Hyperspectral image processing

Image segmentation and region of interest (ROI) selection were conducted using Breeze software. This process was designed to eliminate the background, enabling a clear view of the pixels representing cassava leaves (Baek et al., 2019). Following this, the

spectral data was normalized using the Standard Normal Variate (SNV) method, which adjusted for spectral variations in the numerical data.

2.4 Spectral parameters

Three spectral parameters were used to identify wavelengths with significant differences between healthy and diseased leaves: (i) the mean reflectance values of cassava leaves infected with Xpm compared to healthy leaves; (ii) the spectral difference, calculated by subtracting the mean reflectance of healthy leaves from that of infected leaves at each wavelength; and (iii) sensitivity, determined by the ratio of the mean reflectance of diseased leaves to that of healthy leaves at each analyzed wavelength (Abdulridha et al.,

2020). These parameters provided additional information to support the analysis, complementing the interpretation of the spectral data, and were not directly used in the modeling process.

2.5 Data analysis

The methodology for acquiring and analyzing hyperspectral data followed a structured approach, encompassing data collection, image processing, model selection, data preprocessing, model optimization, training, testing, and performance evaluation. The workflow of the process is illustrated in Figure 2. The analyses were conducted using Python, utilizing the following libraries: Optuna (Akiba et al., 2019), Scikit-learn (Pedregosa et al., 2011), XGBoost (Chen and Guestrin, 2016), Seaborn (Waskom et al., 2017), and Matplotlib (Hunter, 2007).

2.6 Classification methods and hyperparameter optimization

Six supervised learning methods were evaluated for the classification task: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP).

The processed dataset was partitioned into training (70%) and testing (30%) subsets using a stratified sampling approach in order to maintain the original class distribution in each subset. For each machine learning method, hyperparameter optimization was conducted using the Optuna library, with 5-fold stratified cross-validation used to evaluate each hyperparameter combination during training. The mean cross-validation score for each trial was computed to guide the optimization and ensure model robustness, helping to prevent overfitting. After optimization, the

final pipeline for each method was trained on the training dataset using the best-found hyperparameters.

2.7 Model validation

The models' performance was evaluated on the test set using the confusion matrix and the following metrics: accuracy, precision, recall, and F1 Score. These metrics are based on True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) values (Sujatha et al., 2021; Cunha et al., 2023), and are mathematically represented by (Equations 1, 2, 3 and 4):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

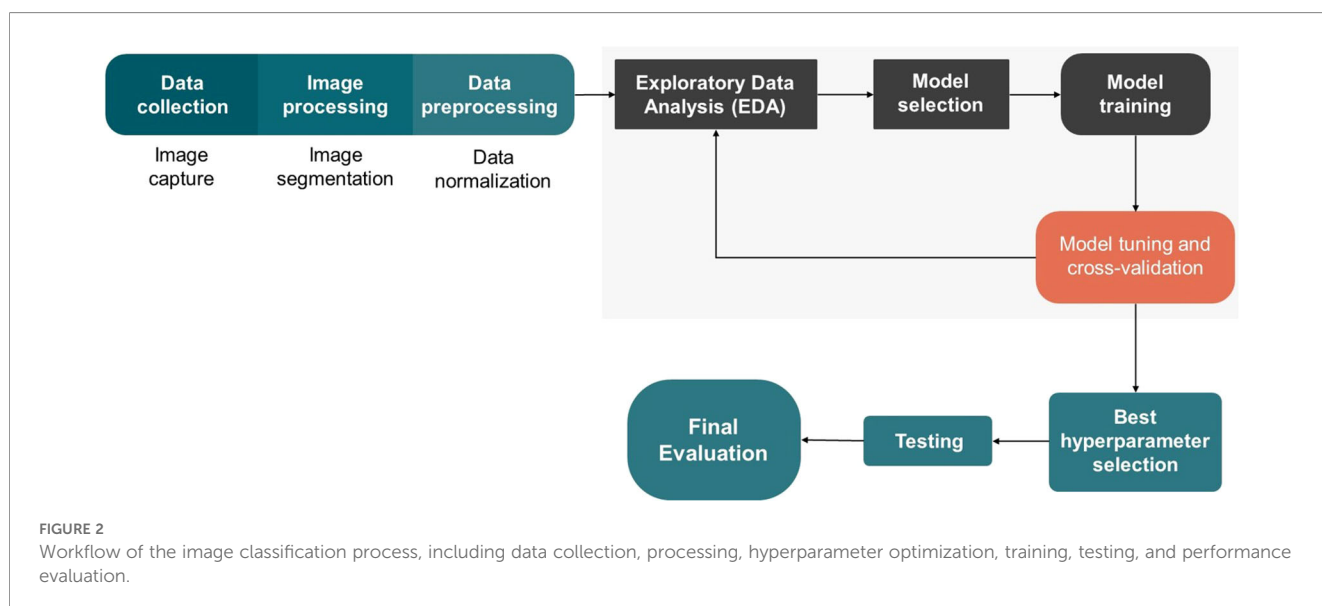
$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ Score} = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Additionally, the ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) metric was calculated to evaluate the models' ability to distinguish between classes (Gonçalves et al., 2014; Li, 2024).

3 Results

Healthy and infected cassava plants were evaluated, with symptomatic leaves exhibiting evolving symptoms from water-soaked spots to necrotic lesions. A total of 852 leaves, comprising 402 healthy and 450 symptomatic samples across various disease



stages, were analyzed and divided into training and testing sets. For each leaf, 448 bands in the visible and near-infrared spectrum (400–1000 nm) were captured using a hyperspectral camera.

The accuracy of the machine learning models applied to the spectral data of healthy cassava leaves and those infected by Xpm is presented in Figure 3. Among the evaluated models, SVM achieved the highest accuracy (91.41%), effectively distinguishing between healthy and infected leaf samples. The MLP model also yielded strong results (87.89%), indicating the neural network's ability to detect subtle variations in the spectral data.

XGBoost and RF, both decision tree-based algorithms, attained accuracies of 79.69% and 77.34%, respectively, showing intermediate performance compared to SVM and MLP. DT achieved an accuracy of 71.88%, reflecting the model's lower discriminative power. KNN exhibited the lowest accuracy (70.31%), performing below all other models.

The model validation parameters indicate that SVM demonstrated the best overall performance (Table 1). It achieved the highest values for F1 Score (0.9209), Recall (0.9481), and AUC-ROC (0.9684), suggesting that SVM effectively combines precision and sensitivity, making it the most suitable choice for classifying spectral data from healthy and diseased plants. MLP also performed well, particularly in Recall (0.9111) and ROC-AUC (0.9473), demonstrating strong detection of the positive class.

In contrast, XGBoost yielded more consistent results, with a Precision of 0.8074, an F1 Score of 0.8074, and an ROC-AUC of 0.8924, indicating a lower level of discrimination compared to SVM and MLP. RF, while exhibiting a relatively high Recall (0.8148), underperformed relative to the more complex models, with an F1 Score of 0.7914 and an ROC-AUC of 0.8493.

DT displayed limited performance, with lower metrics such as an ROC-AUC of 0.7638. Finally, KNN showed the weakest

performance, with a Precision of 0.7360 and Recall of 0.6815, indicating it was the least suitable model for this dataset.

The spectral signature derived from the normalized reflectance of healthy and Xpm-infected leaves is shown in Figure 4A. In the near-infrared (NIR) region (700–1000 nm), a difference was observed between healthy and infected leaves. The highest sensitivity was recorded in the 640–700 nm range, which encompasses the red region (640–680 nm), known for chlorophyll absorption (Figure 4B). The largest spectral differences between healthy and infected leaves occurred around 760 nm in the NIR range (Figure 4C).

4 Discussion

The integration of machine learning techniques with hyperspectral data for diagnosing cassava bacterial blight has yielded promising results, particularly with the SVM. This model emerged as the most effective model for distinguishing between healthy leaves and those infected with *Xanthomonas phaseoli* pv. *manihotis*, based on all performance metrics. Following closely was the MLP, which also demonstrated itself as a viable alternative. XGBoost and RF exhibited intermediate performances, while Decision Tree and KNN were the least suitable for this task.

The SVM superiority can be attributed to its capability to manage complex, high-dimensional data, an advantage highlighted in various plant disease classification studies (Nagasubramanian et al., 2018). The use of kernel functions in SVM enables the transformation of data into a higher-dimensional space, which facilitates class separation and mitigates the risk of overfitting (Pathak et al., 2022). The MLP achieved an accuracy that was only 3.52% lower than that of the SVM, showcasing its strong

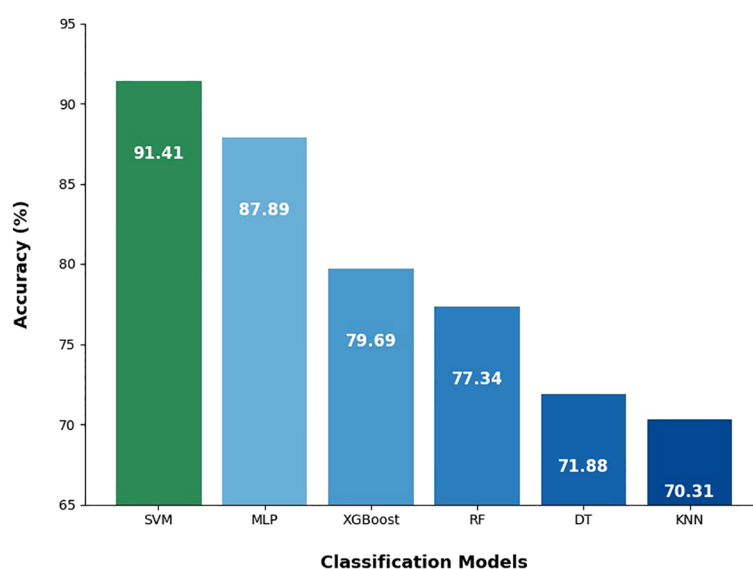


FIGURE 3

Accuracy results of classification methods differentiating healthy cassava leaves from those infected with Xpm. K-Nearest Neighbors (KNN); Decision Tree (DT); Random Forest (RF); Extreme Gradient Boosting (XGBoost); Multi-Layer Perceptron (MLP); Support Vector Machine (SVM).

performance in classifying healthy and infected cassava leaves based on hyperspectral data. Its proficiency in processing large volumes of complex hyperspectral data has been corroborated by other studies on plant diseases (Abdulridha et al., 2020; Lee et al., 2022).

XGBoost and RF are ensemble learning models based on decision trees, designed to combine the predictions of multiple decision trees to enhance accuracy, resulting in more robust and reliable outcomes

(Breiman, 2001; Chen and Guestrin, 2016; Nti et al., 2023). The adoption of this strategy effectively mitigates overfitting, and despite the distinct characteristics of each model, both approaches yielded satisfactory results. These findings highlights the effectiveness of ensemble techniques in the analysis of spectral data.

The DT and KNN models were considered the least suitable for the classification task, exhibiting low precision and poor generalization.

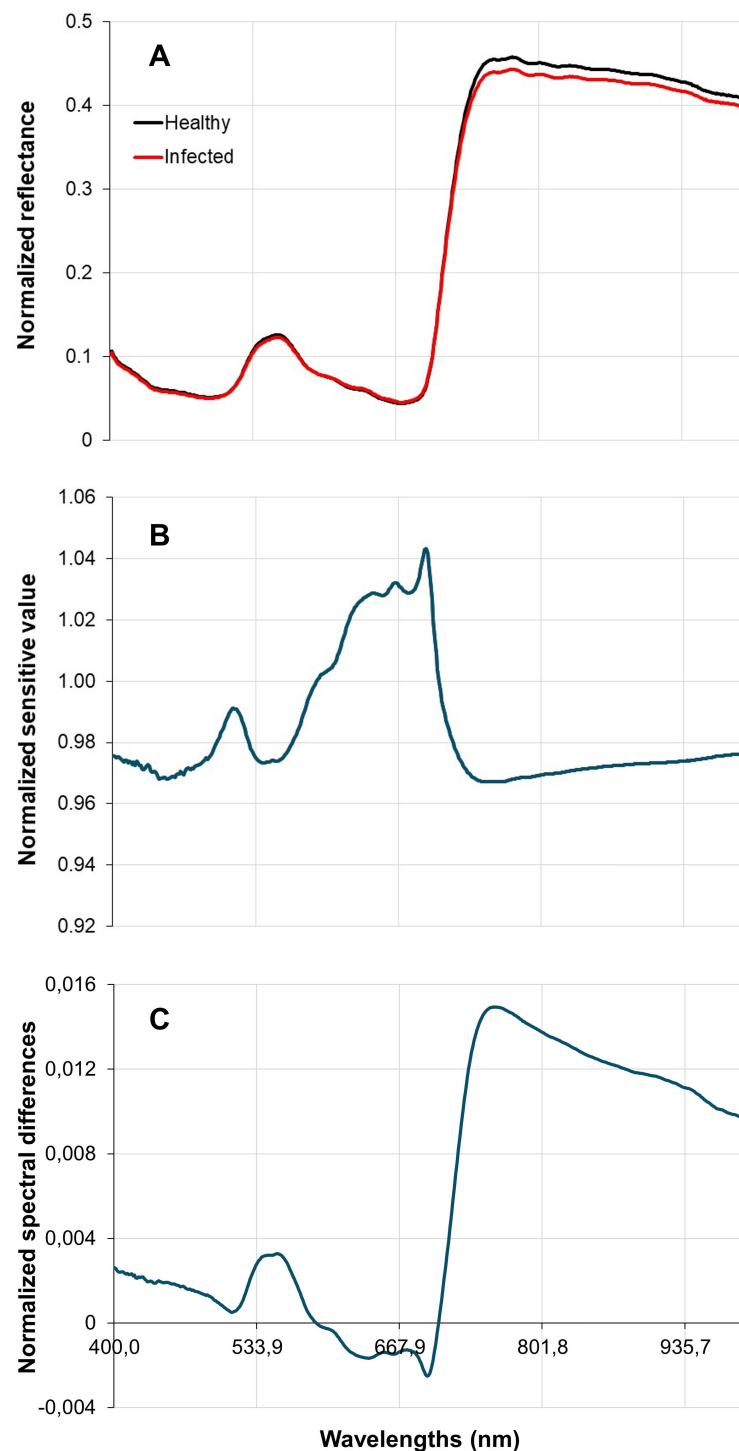


FIGURE 4

Cassava leaves spectral reflectance characterization: normalized spectral reflectance curves (A); normalized sensitivity value (B); normalized spectral difference (C).

TABLE 1 Result of the performance evaluation metrics of the models on the test set.

Classification methods	Confusion matrix		Precision	Recall	F1 Score	ROC-AUC
SVM	TP = 128	FN = 7	0.8951	0.9481	0.9209	0.9684
	FP = 15	TN = 106				
MLP	TP = 123	FN = 12	0.8662	0.9111	0.8881	0.9473
	FP = 19	TN = 102				
XGBoost	TP = 109	FN = 26	0.8074	0.8074	0.8074	0.8924
	FP = 26	TN = 95				
RF	TP = 110	FN = 25	0.7692	0.8148	0.7914	0.8493
	FP = 33	TN = 88				
DT	TP = 103	FN = 32	0.7203	0.7630	0.7410	0.7638
	FP = 40	TN = 81				
KNN	TP = 92	FN = 43	0.7360	0.6815	0.7077	0.7607
	FP = 33	TN = 88				

K-Nearest Neighbors (KNN); Decision Tree (DT); Random Forest (RF); Extreme Gradient Boosting (XGBoost); Multi-Layer Perceptron (MLP); Support Vector Machine (SVM); True Positive (TP); False Positive (FP); False Negative (FN); True Negative (TN).
Receiver Operating Characteristic - Area Under the Curve (ROC-AUC).

Their limited capacity to process high-dimensional and complex datasets, including hyperspectral data, results in reduced performance (Bramer, 2002; Halder et al., 2024; Zhang et al., 2025). These findings indicate that lower-complexity models like DT and KNN are unable to extract truly meaningful information from such spectral data. Recent research supports this, showing that one-dimensional convolutional neural networks (1D-CNNs) outperform traditional algorithms such as PLS-DA, KNN, and RF in distinguishing rice bacterial blight caused by different pathogens (*Xanthomonas oryzae* pv. *oryzae*, *Pantoea ananatis*, and *Enterobacter asburiae*), due to their more robust architectures that enhance feature extraction and generalization (Zhang et al., 2025).

These differences in algorithm performance indicate that each method has distinct capabilities when processing hyperspectral data for binary classification of healthy and diseased plants. In a previous study, differences in machine learning algorithm performance were observed when comparing hyperspectral data of a fungal disease (*Corynespora cassiicola*) and a bacterial disease (*Xanthomonas euvesicatoria* pv. *perforans*) causing leaf spots in tomato plants, under both benchtop and unmanned aerial vehicle conditions. The multi-layer perceptron (MLP) method achieved higher accuracy values compared to the stepwise discriminant analysis (STDA) method (Abdulridha et al., 2020). Another publication analyzed four fungal diseases in tomatoes (*Botrytis cinerea*, *Fusarium oxysporum*, *Alternaria alternata*, and *Alternaria solani*) using hyperspectral and RGB images with a RF model. Hyperspectral imaging proved more accurate, revealing distinct spectral signatures for effective disease differentiation (Javidan et al., 2024). Therefore, the effectiveness of each technique can be influenced by the nature of the data and the interaction between the plant pathogen and the host.

Spectral data have been widely employed to identify physiological and biochemical changes induced by plant pathogens, providing critical insights into the differences between healthy and diseased plants (Abdulridha et al., 2020; Castro-Valdecantos et al., 2024).

In the present study, higher normalized reflectance values in healthy plants in the near-infrared region (NIR, 700–1000 nm) suggest intact cellular structure, essential for physiological functions (Junges et al., 2020; Mevy et al., 2022). Conversely, the lower normalized reflectance observed in infected leaves indicates structural damage and reduced cellular integrity, typical features of bacterial infections (Zhang et al., 2025). This pattern is reinforced by spectral parameters such as sensitivity and spectral differences, with the most pronounced alterations observed within the near-infrared range.

These findings align with observations from research on the physiological changes induced by *Xanthomonas phaseoli* pv. *manihotis* (Xpm) in cassava leaves. Such investigations revealed a reduction in water potential associated with increased stomatal resistance, along with a rise in proline concentration, indicating the plant’s response to disruptions in cellular homeostasis due to bacterial infection (Rubio et al., 2017).

Similarly, a study on *Xanthomonas citri* subsp. *citri* in Sugar Belle mandarins demonstrated that vegetation indices related to chlorophyll and water content effectively detected early-stage bacterial infections (Abdulridha et al., 2019). In a study utilizing hyperspectral data and vegetation indices to assess healthy and *Xanthomonas euvesicatoria* pv. *perforans*-infected tomato plants, significant physiological differences between the groups were detected as early as two hours post-inoculation. At this stage, spectral bands in the ranges of 740–750 nm and 1404 nm were identified as the most critical for distinguishing between healthy and infected plants (Zhang et al., 2024). Furthermore, the spectral data for the four fungal diseases in tomatoes were most effective in the 500–550 nm and 740–950 nm ranges, which encompass the infrared wavelengths, for early-stage identification and diagnosis (Javidan et al., 2024).

This study advanced the understanding of the hyperspectral behavior of cassava leaves under healthy and infected conditions and demonstrated the potential of integrating this technique with machine learning models for the identification of diseased plants. However, because only one susceptible cultivar was used as the reference for analysis, further studies are needed to assess possible variations in spectral behavior among cultivars.

Furthermore, further research will be necessary to refine this technique for detecting infected propagative material, including cuttings for commercial cultivation and seeds for breeding programs. Additionally, the use of drone-mounted cameras and portable sensors emerges as a potential application of this technique in production fields in the future.

5 Conclusions

The integration of machine learning techniques with hyperspectral data has proven effective in detecting physiological alterations caused by cassava bacterial blight in cassava leaves, with SVM achieving the best overall performance. The MLP also exhibited strong performance, while XGBoost and RF produced satisfactory results. The variations in algorithm performance highlight the importance of selecting appropriate methods for the specific pathosystem under evaluation. In addition to that, spectral analysis demonstrated that physiological changes induced by *Xanthomonas phaseoli* pv. *manihotis* can be detected through near-infrared reflectance, reinforcing the significance of spectral techniques in diagnosing plant diseases.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

IC: Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. LF: Formal analysis, Visualization, Writing – review & editing. AN: Visualization, Writing – review & editing. AC: Writing – original draft, Writing – review & editing. HL: Methodology, Writing – original draft. MR: Conceptualization, Formal analysis, Funding

acquisition, Methodology, Project administration, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF), under grant number 00193-00001058/2021-57. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, through scholarships awarded to I.C.B. Carvalho and A.M.S. Carvalho. Additionally, L.C. Ferreira was supported by a scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., and Roberts, P. (2020). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precis. Agric.* 21, 955–978. doi: 10.1007/s11119-019-09703-4
- Abdulridha, J., Batuman, O., and Ampatzidis, Y. (2019). UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sens.* 11, 1373. doi: 10.3390/rs11111373
- Ahmad, A., Saraswat, D., and El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agric. Technol.* 3, 100083. doi: 10.1016/j.atech.2022.100083
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 2623–2631. doi: 10.1145/3292500.3330701

- Asnicar, F., Thomas, A. M., Passerini, A., Waldron, L., and Segata, N. (2024). Machine learning for microbiologists. *Nat. Rev. Microbiol.* 22, 191–205. doi: 10.1038/s41579-023-00984-1
- Baek, I., Kim, M. S., Cho, B.-K., Mo, C., Barnaby, J. Y., McClung, A. M., et al. (2019). Selection of optimal hyperspectral wavebands for detection of discolored, diseased rice seeds. *Appl. Sci.* 9, 1027. doi: 10.3390/app9051027
- Bramer, M. (2002). Using J-pruning to reduce overfitting in classification trees. *Knowl.-Based Syst.* 15, 301–308. doi: 10.1016/S0950-7051(01)00163-0
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Castro-Valdecantos, P., Egea, G., Borrero, C., Pérez-Ruiz, M., and Avilés, M. (2024). Detection of fusarium wilt-induced physiological impairment in strawberry plants using hyperspectral imaging and machine learning. *Precis. Agric.* 25, 2958–2976. doi: 10.1007/s11119-024-10173-6
- Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785–794. doi: 10.1145/2939672.2939785
- Cunha, V. A. G., Hariharan, J., Ampatzidis, Y., and Roberts, P. D. (2023). Early detection of tomato bacterial spot disease in transplant tomato seedlings utilising remote sensing and artificial intelligence. *Biosyst. Eng.* 234, 172–186. doi: 10.1016/j.biosystemseng.2023.09.002
- Demattê, J. A. M., Fiorio, P. R., and Araújo, S. R. (2010). Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. *Remote Sens.* 2, 1998–2016. doi: 10.3390/rs2081998
- Dhakal, K., Sivaramakrishnan, U., Zhang, X., Belay, K., Oakes, J., Wei, X., et al. (2023). Machine learning analysis of hyperspectral images of damaged wheat kernels. *Sensors* 23, 3523. doi: 10.3390/s23073523
- Feng, L., Wu, B., He, Y., and Zhang, C. (2021). Hyperspectral imaging combined with deep transfer learning for rice disease detection. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.693521
- Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., and Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: A review. *Plant Methods* 15, 76. doi: 10.1186/s13007-019-0476-y
- Ferreira, L. C., Carvalho, I. C. B., Jorge, L. A. C., Quezado-Duval, A. M., and Rossato, M. (2024). Hyperspectral imaging for the detection of plant pathogens in seeds: recent developments and challenges. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1387925
- Golhani, K., Balasundram, S. K., Vadamalai, G., and Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* 5, 354–371. doi: 10.1016/j.inpa.2018.05.002
- Gonçalves, L., Subtil, A., Oliveira, M. R., and de Zea Bermudez, P. (2014). ROC curve estimation: An overview. *REVSTAT-Statistical Journal* 12, 1–20. doi: 10.57805/revstatv12i1.141
- Greener, J. G., Kandathil, S. M., Moffat, L., and Jones, D. T. (2022). A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* 23, 40–55. doi: 10.1038/s41580-021-00407-0
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., and Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. *J. Big Data* 11, 65. doi: 10.1186/s40537-024-00973-y
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55
- Javidan, S. M., Banakar, A., Vakilian, K. A., Ampatzidis, Y., and Rahnama, K. (2024). Early detection and spectral signature identification of tomato fungal diseases by RGB and hyperspectral image analysis and machine learning. *Heliyon* 10, e38017. doi: 10.1016/j.heliyon.2024.e38017
- Junges, A. H., Almança, M. A. K., Fajardo, T. V. M., and Ducati, J. R. (2020). Leaf hyperspectral reflectance as a potential tool to detect diseases associated with vineyard decline. *Trop. Plant Pathol.* 45, 522–533. doi: 10.1007/s40858-020-00387-0
- Kado, C. I., and Heskett, M. G. (1970). Selective media for isolation of *Agrobacterium*, *Corynebacterium*, *Erwinia*, *Pseudomonas*, and *Xanthomonas*. *Phytopathology* 60, 969–976. doi: 10.1094/PHYTO-60-969
- Kim, M. S., Chen, Y. R., and Mehl, P. M. (2001). Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Trans. ASAE* 44, 721–729. doi: 10.13031/2013.6099
- Lee, C. C., Koo, V. C., Lim, T. S., Lee, Y. P., and Abidin, H. (2022). A multi-layer perceptron-based approach for early detection of BSR disease in oil palm trees using hyperspectral images. *Heliyon* 8, e09252. doi: 10.1016/j.heliyon.2022.e09252
- Li, J. (2024). Area under the ROC Curve has the most consistent evaluation for binary classification. *PloS One* 19, e0316019. doi: 10.1371/journal.pone.0316019
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x
- Mevy, J.-P., Biryol, C., Boiteau-Barral, M., and Miglietta, F. (2022). The optical response of a Mediterranean shrubland to climate change: hyperspectral reflectance measurements during spring. *Plants* 11, 505. doi: 10.3390/plants11040505
- Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A. K., Singh, A., and Ganapathysubramanian, B. (2018). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods* 14, 86. doi: 10.1186/s13007-018-0349-9
- Nti, I. K., Zaman, A., Nyarko-Boateng, O., Adekoya, A. F., and Keyeremeh, F. (2023). A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. *Decis. Anal. J.* 8, 100311. doi: 10.1016/j.dajour.2023.100311
- Omaye, J. D., Ogbuju, E., Ataguba, G., Jaiyeoba, O., Aneke, J., and Oladipo, F. (2024). Cross-comparative review of machine learning for plant disease detection: Apple, cassava, cotton and potato plants. *Artif. Intell. Agric.* 12, 127–151. doi: 10.1016/j.iaia.2024.04.002
- Ortenberg, F. (2018). “Hyperspectral sensor characteristics: airborne, spaceborne, hand-held, and truck-mounted; integration of hyperspectral data with LIDAR,” in *Fundamentals, sensor systems, spectral libraries, and data mining for vegetation*. Eds. P. S. Thenkabail, J. G. Lyon and A. Huete (CRC Press, New York), 41–69. doi: 10.1201/9781315164151
- Pathak, D. K., Kalita, S. K., and Bhattacharya, D. K. (2022). Hyperspectral image classification using support vector machine: a spectral spatial feature based approach. *Evol. Intell.* 15, 1809–1823. doi: 10.1007/s12065-021-00591-0
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peng, Y., Dallas, M. M., Ascencio-Ibáñez, J. T., Hoyer, J. S., Legg, J., Hanley-Bowdoin, L., et al. (2022). Early detection of plant virus infection using multispectral imaging and spatial-spectral machine learning. *Sci. Rep.* 12, 3113. doi: 10.1038/s41598-022-06372-8
- Qiao, X., Wang, J., Jing, B., Zhang, X., Jia, Y., Huang, K., et al. (2025). Hyperspectral assessment of bacterial blight disease in red kidney beans by feature selection and machine learning algorithms. *Precis. Agric.* 26, 1–16. doi: 10.1007/s11119-025-10253-1
- Rubio, J. S. R., López Carrascal, C. E., and Melgarejo, L. M. (2017). Physiological behavior of *Manihot esculenta* Crantz in response to infection by *Xanthomonas axonopodis* pv. *manihotis* under greenhouse conditions. *Physiol. Mol. Plant Pathol.* 100, 136–141. doi: 10.1016/j.pmpp.2017.09.004
- Sankaran, S., Mishra, A., Ehsani, R., and Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* 72, 1–13. doi: 10.1016/j.compag.2010.02.007
- Sujatha, R., Chatterjee, J. M., Jhanjhi, N. Z., and Brohi, S. N. (2021). Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocess. Microsyst.* 80, 103615. doi: 10.1016/j.micpro.2020.103615
- Taylor, R. K., Griffin, R. L., Jones, L. M., Pease, B., Tsatsia, F., Fanai, C., et al. (2017). First record of *Xanthomonas axonopodis* pv. *manihotis* in Solomon Islands. *Australas. Plant Dis. Notes* 12. doi: 10.1007/s13314-017-0275-0
- Waskom, M., Botvinnik, O., O’Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., et al. (2017). Seaborn v0.8.1. *Zenodo*. doi: 10.5281/zenodo.883859
- Zárate-Chaves, C. A., Osorio-Rodríguez, D., Mora, R. E., Pérez-Quintero, Á.L., Dereeper, A., Restrepo, S., et al. (2021). TAL effector repertoires of strains of *Xanthomonas phaseoli* pv. *manihotis* in commercial cassava crops reveal high diversity at the country scale. *Microorganisms* 9, 315. doi: 10.3390/microorganisms9020315
- Zhang, M., Tang, S., Lin, C., Lin, Z., Zhang, L., Dong, W., et al. (2025). Hyperspectral imaging and machine learning for diagnosing rice bacterial blight symptoms caused by *Xanthomonas oryzae* pv. *oryzae*, *Pantoea ananatis* and *Enterobacter asburiae*. *Plants* 14, 733. doi: 10.3390/plants14050733
- Zhang, X., Vinatzer, B. A., and Li, S. (2024). Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease. *Sci. Rep.* 14, 27666. doi: 10.1038/s41598-024-78650-6
- Zhang, J., Yang, Y., Feng, X., Xu, H., Chen, J., and He, Y. (2020). Identification of bacterial blight resistant rice seeds using terahertz imaging and hyperspectral imaging combined with convolutional neural network. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00821