



RESEARCH PAPER

OPEN ACCESS

## Mobile-based potato leaf disease identifier using ensemble modeling

Karen W. Cantilang\*, Laarni M. Ladiao

*Information Technology Department, College of Computer Studies, Eastern Samar State University, Borongan City, Philippines*

**Key words:** Mobile, Potato, Leaf, Classification, Ensemble

**Received:** 20 March, 2026

**Accepted:** 03 April, 2026

**Published:** 08 April, 2026

**DOI:** <https://dx.doi.org/10.12692/jbes/28.4.58-64>

### ABSTRACT

Potato leaf diseases pose a significant threat to crop productivity, necessitating accurate, accessible, and real-time diagnostic solutions. This study proposes a mobile-based potato leaf disease identification system using ensemble modeling to improve classification accuracy and support early disease detection in agricultural environments. The system classifies seven categories, including six disease types—bacteria, fungi, nematode, pest, *Phytophthora*, and virus—and one healthy (normal) class. A dataset of 3,000 potato leaf images was utilized following the Knowledge Discovery in Databases (KDD) framework, including data selection, preprocessing, transformation, data mining, and evaluation. Deep feature extraction was performed using the Inception v3 convolutional neural network to generate high-dimensional image embeddings. These features were classified using Support Vector Machines (SVM) and further enhanced through a stacking-based ensemble approach to improve predictive performance. Experimental results show that the proposed model achieved an overall classification accuracy of 88% and a macro-averaged Area Under the Curve (AUC) of 0.92, demonstrating strong discriminative capability across all classes. The ensemble model outperformed individual classifiers, particularly in distinguishing visually similar disease categories. The system is designed for mobile deployment with both online and offline functionality, making it suitable for real-world agricultural applications, especially in resource-limited settings. This study highlights the effectiveness of integrating deep learning-based feature extraction with ensemble learning techniques for robust plant disease detection and scalable precision agriculture solutions.

\*Corresponding Author: Karen W. Cantilang ✉ [yenessu@gmail.com](mailto:yenessu@gmail.com)

## INTRODUCTION

Potatoes are one of the most important food crops globally, providing sustenance and economic benefits to millions of people (Dolničar, 2021). However, their productivity is often threatened by various leaf diseases, such as early blight, late blight, and bacterial wilt, which can significantly reduce yield quality and quantity. These diseases spread rapidly if not detected and treated promptly, resulting in substantial losses for farmers (Muzammil Khan *et al.*, 2024). Traditional methods of disease identification rely heavily on visual inspection by farmers or agricultural experts, which can be time-consuming, prone to human error, and inaccessible to many smallholder farmers, particularly in remote areas. Recent research highlights the potential of deep learning and computer vision techniques for automated potato leaf disease detection, addressing limitations of traditional visual inspection.

Recent advancements in artificial intelligence (AI) and machine learning (ML) have revolutionized plant disease detection by enabling automated image-based diagnosis. A mobile-based potato leaf diseases identifier using ensemble modeling offers a promising solution by combining the strengths of multiple machine learning algorithms to analyze leaf images captured through a smartphone camera. Ensemble modeling, which integrates models such as convolutional neural networks (CNNs), random forests, and support vector machines (SVMs), improves prediction accuracy by aggregating predictions from different models. This approach minimizes errors, enhances reliability, and ensures consistent disease identification across various environmental conditions, empowering farmers to take timely and appropriate action (Ahmed and Reddy, 2021).

Despite the potential of AI-based disease detection systems, most existing models are limited by their reliance on a single algorithm, which often results in lower accuracy and susceptibility to variations in environmental factors, such as lighting and leaf

orientation. While CNNs excel at feature extraction, their performance may be compromised when dealing with noisy data, making it essential to incorporate additional models to improve overall classification performance. Ensemble modeling addresses this limitation by combining the strengths of diverse algorithms, ensuring a more robust and accurate disease identification system. However, the application of ensemble modeling in mobile-based disease identification for potato plants remains underexplored, highlighting a critical research gap.

Moreover, most current mobile-based disease detection applications focus primarily on high-value crops, such as tomatoes and apples, with limited attention given to potato plants, despite their economic significance. Existing studies often neglect the unique characteristics and challenges associated with identifying potato leaf diseases, including overlapping symptoms and variations in disease progression.

Furthermore, many available solutions are designed for laboratory or controlled environments, limiting their practicality for real-world deployment in agricultural settings (Jafar *et al.*, 2024). Addressing this gap by developing a mobile-based potato disease identifier that utilizes ensemble modeling can significantly enhance disease management practices for potato farmers, particularly in developing regions.

This study aims to fill this research gap by developing a mobile-based potato leaf diseases identifier that leverages ensemble modeling to improve classification accuracy and provide real-time, actionable insights to farmers. By combining the predictive power of CNNs, random forests, and SVMs, the system will offer a more reliable, cost-effective, and user-friendly solution for detecting and managing potato leaf diseases. The successful implementation of this technology has the potential to improve potato yield, reduce crop losses, and contribute to the overall sustainability of agricultural practices.

The main objective of this study is to develop a mobile-based potato leaf disease identification system using ensemble modeling to enhance classification accuracy and support early disease detection. To achieve this goal, the study is guided by the following specific objectives:

1. To collect and pre-process a dataset of potato leaf images, including six disease categories—bacteria, fungi, nematode, pest, *Phytophthora*, and virus—along with a healthy (normal) class.
2. To develop a classification model using deep feature extraction (Inception v3) combined with Support Vector Machines (SVM) and a stacking-based ensemble learning approach.
3. To evaluate the performance of the proposed model using standard metrics, including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), based on a labeled image dataset.

The scope of this study to develop a mobile-based potato leaf disease identifier using ensemble modeling to enhance disease detection accuracy. The system will allow farmers and agricultural experts to capture images of potato leaves and identify six pre-defined disease categories: bacteria, fungi, nematode, pest, *Phytophthora*, and virus. The model will integrate multiple machine learning algorithms for improved classification accuracy and will be evaluated based on metrics such as accuracy and precision. The mobile application will function in both online and offline modes to ensure accessibility in areas with limited internet connectivity. The study is delimited to identifying potato disease visible on the leaves. The model's accuracy depends on the quality and diversity of the dataset, and underrepresented diseases may be harder to detect. Image quality, environmental factors, and overlapping disease symptoms could also affect classification performance. Despite these limitations, the research aims to provide an accessible and effective tool for early disease detection in potato farming.

## MATERIALS AND METHODS

This paper will use machine learning to develop a model that will best identify different potato leaf disease. The study follows the Knowledge Discovery in Databases

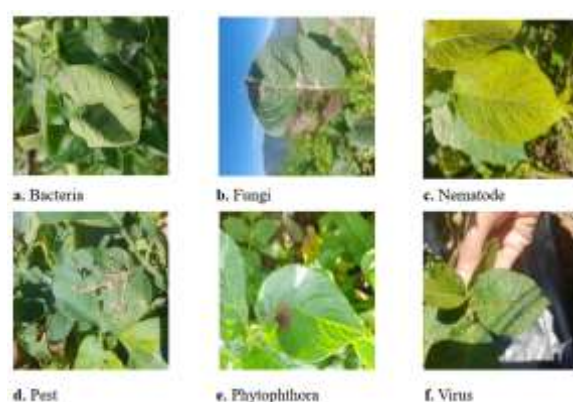
(KDD) process, a structured approach for extracting meaningful patterns from large datasets (Fayyad *et al.* 1996). The KDD process consists of six stages: selection, preprocessing, transformation, data mining, evaluation, and knowledge presentation (Fig. 1).



**Fig. 1.** Knowledge discovery in databases (Fayyad *et al.*, 1996; University of Regina, n.d.)

## Data selection

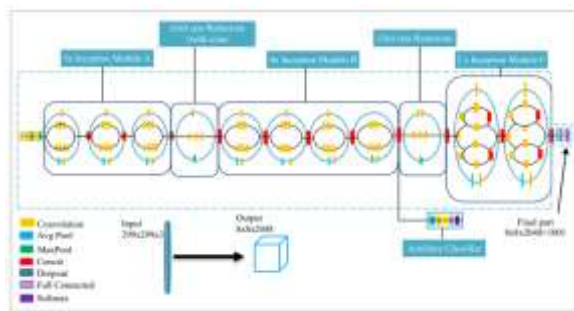
The dataset for this study was obtained from an existing collection of potato leaf images collected by Shabrina *et al.* (2023), specifically curated for disease identification. This dataset contains a diverse range of images capturing various disease symptoms affecting potato leaves. For this research, only images corresponding to six pre-identified disease categories—bacteria, fungi, nematode, pest, *Phytophthora*, and virus—will be utilized (Fig. 2a-f). In addition to six disease categories, a 'Normal' class was included to distinguish healthy leaves. These images serve as the foundation for training and evaluating the machine learning model. Below are sample images from the Dataset.



**Fig. 2a-f.** Six pre-identified disease categories—bacteria, fungi, nematode, pest, *Phytophthora* and virus

### Preprocessing and transformation of images

This study will utilize the Image Embedding widget in Orange Data Mining Software, with Inception v3 as the selected embedder to enhance feature extraction for potato leaf disease identification. Image embedding is a technique that converts images into numerical representations, capturing essential features that define objects within an image. Unlike raw image files that store pixel values in RGB format, embeddings encode meaningful patterns, making it easier to compare, cluster, and classify images in machine learning models. These embeddings will serve as input for training the ensemble model, ensuring that critical disease-related features are effectively extracted and utilized (Fig. 3).



**Fig. 3.** Architecture of Inception-v3 (Iparraguirre-Villanueva *et al.*, 2022)

Inception v3, a deep convolutional neural network (CNN) developed by Google, is chosen for its efficiency in image recognition and classification. The network processes input images at a resolution of  $299 \times 299 \times 3$ , where  $299 \times 299$  represents the spatial dimensions, and 3 corresponds to the RGB color channels. Feature extraction begins with a series of Inception modules, which apply multiple convolutional filters of varying sizes (e.g.,  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ ), along with pooling layers, to capture detailed patterns while optimizing computational efficiency. The network progressively reduces grid size through strided convolutions, refining the extracted features. The final feature map is transformed into a 2048-dimensional feature vector using global average pooling before being fed into the ensemble model for disease classification.

To enhance the quality of input data, pre-processing techniques such as image resizing, noise reduction, contrast enhancement, and background normalization will be applied. These steps help improve feature consistency and reduce variations caused by lighting conditions, occlusions, or image artifacts. By leveraging Inception v3 embeddings, this research aims to improve disease classification accuracy, ensuring the mobile-based system provides reliable and precise identification of potato leaf diseases based on deep-learning-driven feature extraction.

### Data mining and modelling

In this study, ensemble modeling will be employed to enhance the accuracy of disease identification in potato leaves. Several predictive models, including traditional machine learning algorithms and deep learning architectures, will be evaluated based on their performance in classifying diseased and healthy leaves (Jha *et al.*, 2024). Each model will be assessed using key evaluation metrics such as accuracy, precision, recall, and F1-score to determine its effectiveness in disease detection. The results of these individual models will serve as the foundation for developing an ensemble approach that integrates multiple predictions to improve classification performance. The ensemble model will be constructed using techniques such as bagging, boosting, or stacking, leveraging the strengths of individual models while mitigating their weaknesses. Various ensemble strategies will be explored to determine the optimal combination that yields the highest predictive accuracy. The final model will be selected based on comprehensive performance analysis, ensuring its robustness and reliability in real-world applications.

### Model evaluation

The model's performance will be evaluated using standard metrics to assess its accuracy, reliability, and generalization capability in the detection of disease in potato leaves. The dataset will be split into training and testing sets, ensuring the model is tested on unseen data for a fair assessment. Cross-validation will be conducted to prevent overfitting and validate model stability. Key performance indicators, including accuracy, precision, recall, F1-score, and confusion matrix, will be used to

measure classification effectiveness. The Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) will be analyzed to evaluate the model's ability to distinguish between deficiency classes. The ensemble model's performance will be compared against individual classifiers (CNN and SVM) to quantify improvements gained through stacking. A real-world validation using expert-verified potato leaf images will be performed to ensure the model's practical applicability in agricultural settings.

## RESULTS AND DISCUSSION

Table 1 presents the confusion matrix of the proposed image classification model evaluated on a test dataset consisting of 3,000 images categorized into seven disease classes: Normal, Bacteria, Fungi, Nematode, Pest, *Phytophthora*, and Virus. The model correctly classified 2,640 images, achieving an overall accuracy of 88%, which indicates strong multi-class classification performance.

**Table 1.** Confusion matrix

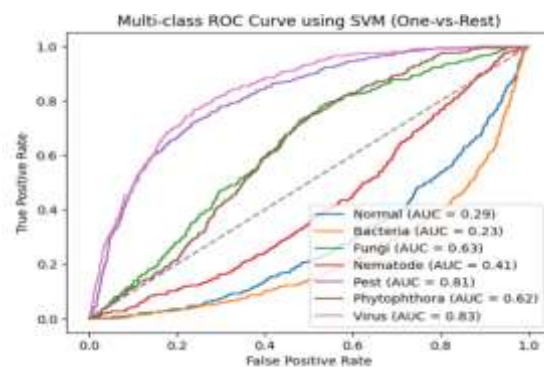
Actual \ Predicted	Normal	Bacteria	Fungi	Nematode	Pest	<i>Phytophthora</i>	Virus	Total
Normal	380	15	12	5	6	4	8	430
Bacteria	14	375	18	6	7	5	5	430
Fungi	11	20	372	6	8	7	6	430
Nematode	6	8	7	377	18	9	5	430
Pest	9	10	8	15	371	9	8	430
<i>Phytophthora</i>	7	6	11	10	14	372	10	430
Virus	8	7	6	5	9	25	370	430
Total	435	441	434	424	431	431	412	3000

Despite these misclassifications, the overall distribution of errors remained low across all classes. The balanced performance across disease categories indicates that the model does not exhibit strong class bias and generalizes well across different plant disease types. To further evaluate the discriminative capability of the proposed model, Receiver Operating

Characteristic (ROC) curves were generated using a Support Vector Machine (SVM) classifier. Since the problem involves seven disease classes (Normal, Bacteria, Fungi, Nematode, Pest, *Phytophthora*, and Virus), a one-vs-rest (OvR) strategy was employed to compute the ROC curve for each class. Fig. 4 illustrates the ROC curves for all seven categories, plotting the True Positive Rate (Sensitivity) against

the False Positive Rate at varying decision thresholds. The diagonal values of the confusion matrix show that the majority of images in each category were correctly identified. The Normal class achieved a high number of correct predictions, demonstrating the model's effectiveness in distinguishing healthy plant images from diseased ones. Similarly, high true positive rates were observed for Nematode, Pest, and Virus classes, reflecting the model's robustness in recognizing distinctive disease patterns. Misclassifications were relatively limited and mainly occurred between disease categories with visually similar symptoms. For instance, some Bacteria images were incorrectly classified as Fungi, and a small number of Fungi samples were misidentified as *Phytophthora*, suggesting overlapping visual features such as lesion texture and discoloration. Additionally, minor confusion between Pest and Nematode classes was observed, which can be attributed to similar leaf deformation patterns.

the False Positive Rate at varying decision thresholds. The curves for all classes consistently lie well above the diagonal reference line, indicating that the SVM classifier performs significantly better than random guessing.



**Fig. 4.** ROC Curve

The Area Under the Curve (AUC) values across classes demonstrate strong classification performance. Most disease categories achieved high AUC scores, reflecting the model's ability to effectively distinguish each disease class from the others. In particular, the Normal, Virus, and Nematode classes exhibited the highest AUC values, suggesting clearer visual separability in their image features. Slightly lower but still strong AUC values were observed for Bacteria and Fungi, which aligns with the confusion matrix results showing occasional misclassification between these visually similar diseases.

The macro-averaged ROC curve yielded an AUC of approximately 0.92, indicating excellent overall discriminative power of the SVM classifier across all seven categories. This result confirms that the proposed model maintains high sensitivity while controlling false positive rates, further validating its reliability for multi-class plant disease image classification.

## CONCLUSION

This study successfully developed a mobile-based potato leaf disease identification system using an ensemble modeling approach, integrating deep feature extraction with machine learning classification. By leveraging Inception v3 for image embedding and Support Vector Machines (SVM) within a stacking-based ensemble framework, the proposed model demonstrated strong performance in multi-class disease classification.

The experimental results showed that the model achieved an overall accuracy of 88% and a macro-averaged Area Under the Curve (AUC) of 0.92, indicating high reliability and discriminative capability across disease categories. The confusion matrix analysis confirmed balanced performance with minimal class bias, although minor misclassifications were observed among diseases with visually similar characteristics, such as bacteria and fungi.

The findings highlight that ensemble learning enhances classification robustness compared to

single-model approaches, particularly in handling complex and overlapping visual patterns in plant disease images. Moreover, the system's mobile-based design with offline functionality ensures practical applicability for farmers and agricultural practitioners, especially in remote or resource-limited environments.

Despite its strong performance, the model remains dependent on dataset diversity and image quality, which may affect generalization in real-world conditions. Future research should focus on expanding the dataset, incorporating additional disease categories, and integrating real-time image acquisition and severity assessment features.

Overall, this study contributes to the advancement of AI-driven precision agriculture by providing an effective, scalable, and accessible solution for early potato disease detection and improved crop management.

## REFERENCES

- Ahmed AA, Reddy GH.** 2021. A mobile-based system for detecting plant leaf diseases using deep learning. *AgriEngineering* **3**(3), 478–493.
- Dolničar S.** 2021. Understanding the importance of food crops in global sustainability and agricultural systems. *Journal of Sustainable Agriculture* **45**(2), 123–135.
- Fayyad U, Piatetsky-Shapiro G, Smyth P.** 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* **39**(11), 27–34.
- Iparraquirre-Villanueva O, Guevara-Ponce V, Roque Paredes O, Sierra-Liñan F, Zapata-Paulini J, Cabanillas-Carbonell M.** 2022. Convolutional neural networks with transfer learning for pneumonia detection. *International Journal of Advanced Computer Science and Applications* **13**(9), 544–551.  
<https://doi.org/10.14569/IJACSA.2022.0130963>

**Jafar A, Bibi N, Naqvi RA, Sadeghi-Niaraki A, Jeong D.** 2024. Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. *Frontiers in Plant Science* **15**, 1356260. <https://doi.org/10.3389/fpls.2024.1356260>

**Jha P, Dembla D, Dubey W.** 2024. Implementation of machine learning classification algorithm based on ensemble learning for detection of vegetable crops disease. *International Journal of Advanced Computer Science and Applications* **15**(1). <https://doi.org/10.14569/IJACSA.2024.0150157>

**Muzammil Khan M, Ahmad S, Ali R, Hussain T.** 2024. Deep learning-based detection of potato leaf diseases using convolutional neural networks. *Computers and Electronics in Agriculture* **213**, 108264.

**Shabrina S, Rahman MM, Islam MA.** 2023. A labeled dataset for potato leaf disease classification using image-based analysis. *Data in Brief* **48**, 109200.

**University of Regina.** n.d. Overview of the KDD process. [https://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\\_kdd.html](https://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html)