

Intelligent Diagnosis of Potato Leaf Diseases Using NLSCTAN and GPT-4 Powered Multimodal Framework

¹Rupali Maske*, ²B.Bharathi

¹Research Scholar, ²Doctorate

^{1, 2} Sathyabama Institute of Science and Technology, Chennai, India

*Corresponding mail: Rupali Maske

Abstract

To enhance crop production and ensure food security as well as reduce excessive use of chemical pesticides in the modern farming, it is important to detect potato leaf diseases early and precisely. However, environmental changes, the shortcomings of unimodal image based techniques, and visual symptom overlap have posed a hard task in disease recognition. The study offers an intelligent multimodal network of potent segmentation and classification of potato leaf diseases. It is constructed on the basis of the Normalized Long-Short Convolutional Temporal Attention Network and implemented thereof with a decision support system powered by GPT-4. This is aimed at addressing these issues. It has three primary divisions namely, the deep visual feature extraction branch (MSC-ResViT), the semantic and textual feature learning branch (MSC-TextCNN), and the color-texture representation branch (CT-CNN). These features are fused and the NLSCTAN classifies them into complex spatial-temporal connections. The outcomes of the experiments on the PlantVillage and Potato Leaf Disease datasets are promising. The proposed model has 98.2 and 95.6% accuracy rate on the PlantVillage and Potato Leaf Disease datasets respectively. The class-wise analysis shows that, the F1-score of the Healthy class is 0.95, the Early Blight class is 0.90, and the Late Blight class is 0.91. Moreover, the connection with GPT-4 enables the PotatoGPT platform to interpret illnesses and give practical treatment guidance in real-time, which can add greater practicality to the platform to be applied to precision agriculture.

Keywords: Potato Leaf Disease Detection, Multimodal Deep Learning, NLSCTAN, GPT-4 Integration, Precision Agriculture

1. INTRODUCTION

Potato leaf diseases are a collection of diseases that cause adverse impact on the foliage of potato plants and which are caused by a variety of fungi, bacteria and viruses. These diseases are manifested with the symptom of leaf discoloration, lesions, spots, wilting, and deformation. Typical ones are late blight (infested by *Phytophthora infestans*) and early blight [1] (*Alternaria solani*), bacterial wilt (*Ralstonia solanacearum*), and Potato virus Y infections. These diseases may cause drastic losses in yield and quality of crops in case these are not detected and treated early, which explains why it is necessary to detect and identify them in time. Deep learning has proven highly potential in detecting diseases, however, a number of issues concerning the successful segmentation and classification of potato leaf diseases still exist. These are the complexity of the structure of leaves, the variation in presentation of symptoms and the dearth of high-quality large labelled dataset. In addition, the computational capacity required to process the high-resolution images of the agriculture industry is a lot of resource, which may become problematic in field applications in real-time [2,3].

Current deep learning methods of segmentation and classification have a weakness of robustness and generalization. Their complexity, which is highly computationally expensive, limits their application in the real world, especially in resource-limited agricultural systems, misclassification, through the variety of symptoms, and overfitting with small or unbalanced data sets all also limit their usefulness. In order to overcome these shortcomings, this study presents an elaborate plan of segmentation and categorization of potato leaf disease [4,5]. The applied strategy is based on advanced deep learning frameworks, i.e. convolutional neural networks and recurrent neural networks to better represent complex patterns of diseases. The study also aims to archive and grow existing data in order to have a wide diversity of disease types. New data augmentation methods would be used for solving the imbalance of datasets. Moreover, the optimization and lightweight creation of the models would also support their implementation in real time and low resources environment. By means of these collaborative initiatives, the research would come up with a formidable and scalable hybrid deep learning model that

enhances the precision and viability of potato disease detection systems significantly, and eventually resulting in healthier crops and better agricultural yield [6,7].

The application of deep learning methods in plant disease detection has gained interest as artificial intelligence continues to grow rapidly [8,9]. Deep learning models can use image recognition capabilities for detecting crop pathogens and therefore increase the speed and accuracy of detection significantly. These improvements are creating new possibilities in today's farming methods and have a revolutionary impact on the management of diseases. Crop diseases are caused by pathogens like bacteria, fungi, pests, and viruses that disturb essential activities of plants; it pose a big threat to agriculture globally [10]. These directly influence food production with severe effects on the economy and environment. It is estimated that about 60 Among different plant health problems, leaf diseases are most efficient in reducing crop productivity [11,12]. The potato crop has extremely significant importance for domestic food security as well as international trade. Since it is an important staple food, any battle against any disease which can affect its quality and quantity becomes highly relevant. Such problems need solutions not only for improved agricultural output but also for a more reinforced national economy.

To solve this problem, a hybrid deep learning framework that contains several deep learning models is presented here. Also, a high-level comparison on how approaches to plant disease detection have evolved would be given in order to position the proposed framework properly within this evolution. The hierarchical comparison of detection systems is shown in Figure 1. This method uses a feature fusion strategy to extract rich and diverse deep features for better precision and robustness in detecting potato leaf disease [13,14]. The solution aims at improving crop quality reduction in disease prevalence plus encouraging sustainable agricultural development.

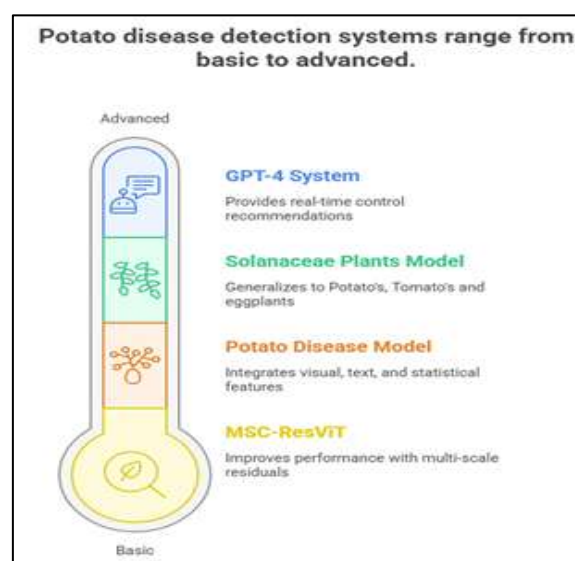


Figure 1: Hierarchical evolution of potato disease detection systems from basic visual models to an advanced GPT-4-enabled multimodal framework.

Deep learning has been extensively used in detecting diseases of potatoes which have brought significant improvements in precision agriculture [15-18]. By making use of advanced image analysis techniques, researchers are now able to identify more accurately and classify various potato diseases thereby improving diagnostic accuracy as well as providing valuable technical support to the farmers. Recently developed intelligent algorithms have greatly enhanced monitoring and detection capabilities related to plant pests and diseases. Initially, machine learning techniques were used for identifying problems like citrus disease. Accurate and effective image-based categorization and disease detection in citrus fruits has been made possible by advancements in deep learning models, particularly Convolutional Neural Networks. For instance, [19] trained a CNN model on a publicly available dataset that contained 87,848 images belonging to 58 classes across 25 plant species and disease combinations. The model achieved almost perfect performance with an accuracy of recognition equal to 99.59%. Figure 2 shows the subtle relationship between the Multimodal AI Model, MSC-ResViT, and its specific task of finding plant diseases within the broader context of the research i.e., GPT-4 System.

Characteristic	MSC-ResViT	Multimodal AI Model	Expert System (GPT-4)
Purpose	Visual Detection model	Potato disease detection	Potato disease diagnosis
Generalizability	Not Specified	Solanaceae plants	Not applicable
Modality	Visual	Multimodal	Language

Figure 2: Comparison of AI models for plant disease detection.

Although these studies show great performance on certain datas, these point to a key restriction namely the use of visual data only. In practice, it can be a source of misinterpretation such as the yellowing of the leaves of citrus under water shortage could be confused with nitrogen shortage as these would have similar symptoms. To overcome these difficulties, [20-22] suggested a multimodal system that integrates image and text data with noise information to eliminate the drawbacks in zero-shot and low-sample learning issues when features of the diseases are not visible in the photos or are indefinite. As much as these multimodal approaches are a step in the right direction, the previous models had some critical shortcomings such as the inability to understand the context, inability to allow user interactivity and being computational intensive. This made them unrealistic when it came to mass application by the farmers. In order to address these constraints, this study would incorporate recent developments in large language models, specifically GPT, to augment the levels of context awareness, the level of user engagement, and make the system lightweight and accessible to be applied to real-world farming scenarios.

The use of Large Language Models (LLMs) in the plant protection practice is becoming increasingly common due to the strong potential of such models to process natural language, merge information, and conduct intelligent decisions. Qing et al. (2023) in this work proposed a new method in the field of agricultural pest and disease recognition, which combines the logical image recognition capabilities of GPT-4 with the visual recognition capabilities of the YOLOPC network. The model was also made more flexible and effective due to the application of a new image-to-text mapping method. This study lays down an excellent foundation and also points toward future directions for applying LLMs in intelligent plant pathology diagnosis. Although deep learning approaches to potato disease detection have come a long way, there are still gaps that remain unaddressed. The complex nature of potato diseases makes it impossible to extract all relevant features through visual inspection alone. Most of the literature emphasizes accurate classification of disease; very few works offer any practical treatment or control recommendations, thus limiting their real-world applicability in agriculture.

The framework proposed here is a step toward precision agriculture by providing a system for early, accurate, and interpretable potato disease diagnosis using simple imaging devices. Such systems may help achieve rapid disease surveillance and diagnosis so as to reduce economic losses by enabling timely treatment while promoting sustainable farming through the minimization of unnecessary chemical spraying. This study describes NLSCTAN combined with GPT-4 powered multimodal reasoning as a next-generation approach toward smart crop health monitoring that would be scalable, farmer-centric, and adaptable to various field conditions. Here are the research objectives as shown below:

- I. To design a robust multimodal deep learning framework based on NLSCTAN that effectively integrates visual, textual, and color-texture features for accurate segmentation and classification of potato leaf diseases.

- II. To use the developed image preprocessing, data augmentation, feature fusion and temporal attention mechanisms across heterogeneous datasets to increase the accuracy of disease detection and generalization.
- III. To assess the performance of the proposed framework on the benchmark datasets (Potato Leaf Disease and PlantVillage) through the standard measures of performance, i.e. accuracy, precision, recall, and F1-score.
- IV. To create a decision-support system based on GPT-4 (PotatoGPT) that delivers in-real-time diagnosis of diseases and real-time recommendations in terms of actionable treatment and fertilizer to practice precision agriculture.

2. LITERATURE OF REVIEW

The recent progress in the deep learning-based image classification models has given potato leaf disease detection a new impetus in the level of accuracy. These models consist of convolutional neural networks (CNNs), as well as their variations. Bajpai et al. (2024) [23] provided a rather strong structure including a BFC module based on distillation. This framework obtained 99% accuracy, precision and F1-score with minimal amount of training data as well as strong generalization of 95% accuracy on Kannauj dataset. Bhavani et al. (2025) [24] proposed DarkPotatoleafNet, a hybrid CNN-Vision Transformer, which enhanced the performance in spatial features extraction and high accuracy of 99% with regards to challenges associated with generalizability and computational power. The usefulness of CNN-based methods was confirmed by Abbas et al. (2024) [25] and Ghosh et al. (2023) [26] on the datasets PlantVillage with the reported accuracy over 97% on the multiclass classification of diseases in potatoes and other crops. In 2024, Kolluri et al. extended this paradigm to control images in a total of forty-five thousand and four hundred forty-six images of fourteen plant species and twenty-six diseases with an accuracy of ninety-nine point zero six percent supporting the validity of deep convolutional architecture to recognize plant diseases at scale.

Recent research has discussed multimodal learning and contextual data fusion in order to get over the limitation of unimodal image-based systems. A graph-learning-based multimodal feature interactive fusion model with visual and textual data input and semantic space encoder and self-attention graph convolution network as a significant result, with an accuracy of 96.95% on a large-scale multimodal dataset (Wu et al. 2025) [28] was proposed. In [29], Nasir et al. proposed a multimodal deep learning framework that used EfficientNetB0 to learn how to classify images in addition to recurrent neural networks to learn how to predict disease severity when fed environment-related data, achieving a 96.40% classification and 99.20% disease severity prediction results. Explainable AI tools like LIME and SHAP have led to better interpretability and, therefore, make it more applicable to agricultural decision-making. As depicted by Alhammad et al. (2025) [30], this is how transfer learning can be used alongside explainable AI to gain higher interpretability along with up to 98% test accuracy of potato leaf diseases.

Its model optimization, severity estimation, and feasibility of its real-world implementation were also investigated on a number of occasions. Dame et al. [31] created a CNN-based smartphone application that can identify and determine the level of severity of the diseases with 99% disease recognition and 96% disease severity accuracy between early and late blight diseases. Zhang et al. (2025) [32] optimized the VGG16S by incorporating the global average pooling, CBAM attention, and Leaky ReLU activation features and cut the model by a complete tenfold of what the original VGG16 had still reached an accuracy of 97.87. Thalor et al. (2024) [33] suggested PotatoLeaf Insight System grounded on AI using U-Net to divide the picture and subsequently identify the features using VGG19 that resulted in 94% accuracy and the possibility to identify the disease at an early stage. All these studies demonstrate that there is indeed a shift of high-accuracy CNN-based classifiers to optimized multimodal, explainable, and deployment-ready intelligent potato leaf disease diagnosis systems.

Although the current researches have high accuracies, there are still research gaps in the diagnosis of potato leaf disease. Majority of methods are overly dependent on single modality image representations and thus are not robust to different field conditions like changes in illumination, occlusion and background noise. Despite the existence of some multimodal frames, there is little evidence of deep semantic interaction amongst visual, textual and contextual

knowledge and large language model assisted reasoning remains not well studied. Also, the majority of models solely focus on the classification accuracy without considering real-time implementation, cross-dataset generalization, and adaptive learning with small samples. Besides, explainability is considered more of a post hoc addition instead of a natural part of the diagnostic pipeline, which suggests the necessity of coherent, smart, and scalable multimodal systems.

3. RESEARCH METHODOLOGY

The Figure 2 shows the overall workflow of the proposed architecture used to diagnose the potato leaf multimodal disease. Once pictures have been combined between PlantVillage and Potato Leaf Disease to provide a more diversified dataset, the picture is prepared with the Adaptive Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to enhance the quality of the picture. The second phase is to draw statistical features, including variance, entropy, correlation and contrast. Using encoder-decoder learning, an autoencoder is used to reduce features in a latent space. The model is optimized with a modified walrus optimizer in order to maximize profits. Then apply region-based segmentation, including normalization, convolution, attention, LSTM and ROI pooling layers. A final step is to perform categorization and fertilizer prescription with the help of the Normalized Long-Short Convolved Temporal Attention Network (NLSCTAN).

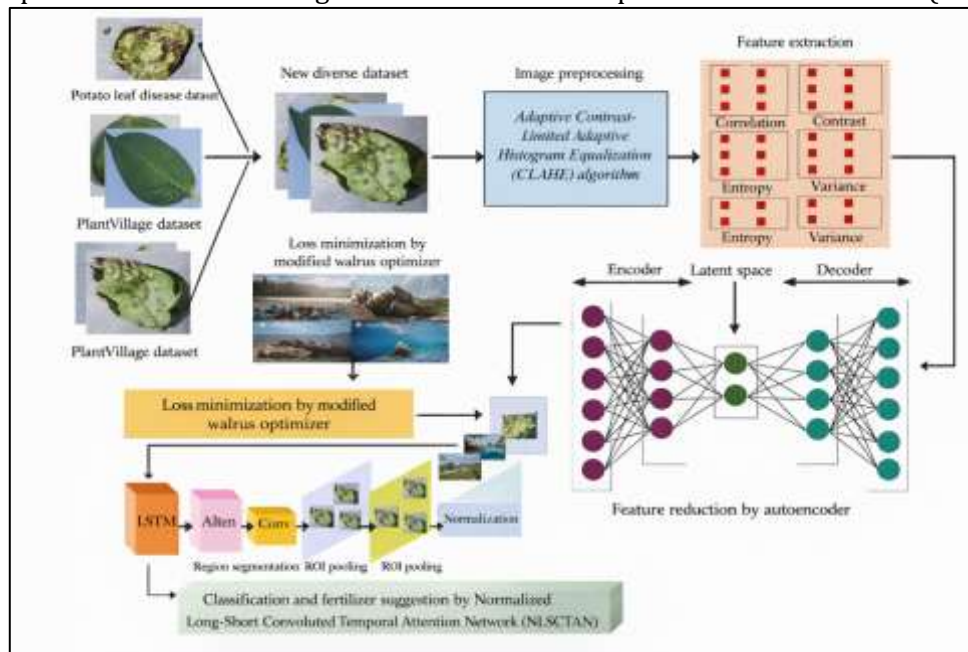


Figure 3: Proposed Methodology

3.1 Dataset Used

The suggested research makes use of a diverse and non-uniform dataset of potato leaf images compiled from a number of open-source resources, including the Potato Leaf Disease Dataset [34] and the PlantVillage Dataset [35]. These collections contain high-resolution images of potato leaves under various lighting, background, and environmental situations, both healthy and diseased. Potatoes can contract early blight, late blight, or healthy leaves, among other common leaf diseases. Data is distributed as follows: 70% for training, 15% for validation, and 15% for testing in order to strike a balance between the learning process and critical testing. The benefit of the combined dataset is that it improves model generalizability to actual field agricultural situations, increases data diversity, and minimizes overfitting. You may see the visual differences between healthy, early blight, and late blight potato leaf classes in Figure 3.



Figure 4: Sample images of healthy and diseased potato leaves from the dataset.

3.2 Image Pre-processing

In this step, raw pictures of the potato leaf undergo processing in such a manner that ensures that these are characterized by uniformity and that they are fit to be analyzed using deep learning. Firstly, images are all resized to some predefined spatial resolution to ensure that the size of the inputs are all the same and to simplify the computation process by reducing the complexity of the computation. Color images are converted to a standard RGB, but their pixel intensity values are reduced to some limit (e.g. 0-1) to ensure that model training is more stable. Smoothing or median filtering procedures remove the noise and the small artifacts that could have been created in the process of capturing the pictures. To this end, contrast enhancement techniques are applied to the disease-afflicted regions thereby rendering the comparison of the change in texture and color on the leaf surfaces easier. Through cropping or threshold-based segmentation the irregularities in the background are minimized to a point where the model can only process the leaf area. The preprocessing step has ensured that the features are more pronounced hence enhancing convergence and the overall accuracy of the classification.

3.3 Data Augmentation

Data augmentation is applied to augment the scale and variety of the training set effectively to enhance the generalizability of the model and decrease overfitting. Due to the imbalance in the number of classes of plant diseases and the small amount of practical samples, the training pictures are subject to a wide range of geometric and photometric transformations. These changes include, among others, random rotations, horizontal and vertical flipping, as well as scaling the latter of which is performed with such small variations that this difference in the size of leaves is quite hard to observe. Moreover, the brightness and contrast of the images are altered to simulate the different lighting conditions that may be present in the field. Slight noise injection is sometimes performed to make the model less sensitive to image distortions. The training set only is subjected to augmentation, so that there would be no bias in evaluation. Consequently, the model becomes better at recognizing the invariant and discriminatory features of diseases, thus, improving classification stability and accuracy.

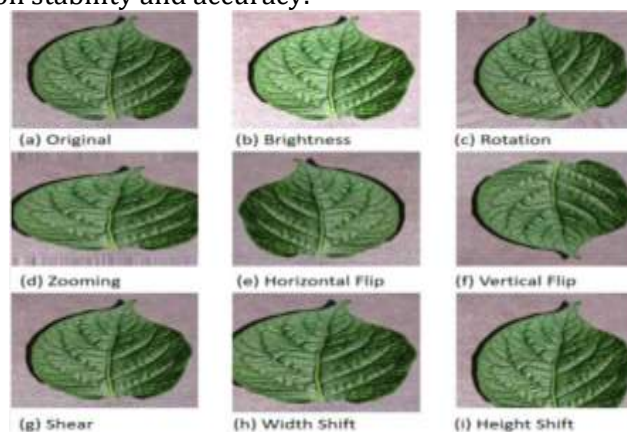


Figure 5: Data Augmentation

3.4 Feature Extraction

This step involves the extraction of discriminative feature in the pre-processed and augmented images of potato leaf through a Color-Texture Convolutional Neural Network (CT-CNN). In contrast to a traditional CNN, CT-CNN incorporates directly the color and texture data to improve representation of disease-specific features. All the input images $I \in \mathbb{R}^H \times W \times C$ are converted to color-texture feature maps and subsequently subjected to convolutional layers with the production of deep feature maps as :

$$F_k = f(X_{CT} * W_k + b_k) \quad (1)$$

Here, X_{CT} denotes the combined color-texture representation of the input image, W_k and b_k weights and bias of the k^{th} convolutional filter, and $f(\cdot)$ is a nonlinear activation function, typically ReLU:

$$f(x) = \max(0, x) \quad (2)$$

To reduce spatial dimensionality while preserving salient disease features, pooling operations are applied:

$$P_k = \text{maxpool}(F_k) \quad (3)$$

Having a hierarchical convolution-pooling architecture, CT-CNN can capture general features such as color variations and texture imperfections as well as more disease-related features such as patterns. The last thing is to reduce the deep feature maps into a feature vector which is then forwarded to the multimodal fusion and classification stage.

3.5 Model Classification and GPT-4 Integration

The Normalized Long-Short Convolved Temporal Attention Network (NLSCTAN) is one of the hybrid deep learning models used to classify feature vectors obtained. Since it is able to model complex relationships within the extracted features, the NLSCTAN model is better than those that model the complex relationship in the original images. Normalization method enhances the stability and effectiveness of learning by integrating a temporal attention method, a sequence of convolutional refinement layers, and Long Short-Term Memory (LSTM) networks. The LSTM network processes the normalized feature vectors in order to capture the long-term dependencies. The network also has a temporal attention mechanism whereby increased weighting is laid on the disease relevant features and reducing the weighting of any redundant or not very important features.

The outcomes of the classification and extracted images are transformed into textual representations of the data to allow real-time intelligent decisions to be taken. The text descriptions are then utilized through the GPT-4 language model to form the PotatoGPT diagnostic platform. GPT-4 offers semantic reasoning on multimodal inputs and can be used in real-time to identify diseases, recommend treatments, and give advice about fertilizers or pesticides, and should be computationally efficient enough to apply in real-world deployments in agricultural environments.

3.6 Model Training and Validation

The division of the entire set of data into training, validation, and testing is a strong measure to guarantee the correct learning and reasonable fairness in assessing performance. The network parameters of the model are trained by updating the weights of the network by using a set of training data containing pictures of leaves of potatoes with labels of their illness. The model is now used to minimize the loss function using optimization and backpropagation. In order to adjust hyperparameters and terminate early to prevent overfitting, the validation set is used to illustrate how a model performs in the course of training. This helps avoid over mesmerization of the training samples on the model by testing it on unknown validation data. You can observe the extent to which it generalizes to real-world situations and is applied to fresh data on the testing set, which has nothing in common with the training set. To evaluate the objective models objectively, the Potato Leaf Disease Dataset has been divided into three subsets, that is, training, validation, and testing (70:15:15, as shown in Table 1).

Table 1: Potato Leaf Disease Dataset Split

Disease Class	Total Images	Training (70%)	Validation (15%)	Testing (15%)
---------------	--------------	----------------	------------------	---------------

Early Blight	1,000	700	150	150
Late Blight	1,000	700	150	150
Healthy Leaves	152	106	23	23
Total	2,152	1,506	323	323

3.7 Performance Metrics

Some of the evaluation criteria that were used in the study included F1 score, recall, accuracy and precision. In the comparison of performance of the model in all the seven domains, accuracy ranked high and above others. Recall determines the capacity of the model to identify all positive cases where the precision determines the reduction in the number of false positives. The F1 score is a combined measure of recall and precision since it is a geometric mean of the two. In the metrics, TP is used to denote cases in which a positive example is correctly predicted as positive, FP is used to denote cases in which a negative example is incorrectly predicted as positive, FN is used to denote cases in which a positive is incorrectly predicted as negative and TN is used to denote cases in which a negative is correctly predicted as negative. Recall, accuracy, precision and F1 score values can be calculated with the formula presented below as equations (4)-(7).

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{recall} + \text{precision}} \quad (7)$$

4. RESULTS AND DISCUSSION

4.1 Image Pre-processing Analysis

Figure 5 depicts the stepwise image pre, processing and segmentation flow that was utilized for potato leaf disease analysis. At first, the RGB image of the infected leaf is captured and taken as input. After that, the green (G) band is changed from the RGB image in order to raise the contrast between the healthy and the diseased parts. Edge detection is operated to show the structural boundaries of the lesions, and morphological dilation is used to extend the edges that are broken or fragmented. After that, the spaces in the detected edges are filled so that a complete leaf mask can be obtained. The background is removed from the original image with the help of this mask, thus the leaf region is isolated. At last, the processed image is divided into healthy and diseased segments; hence the infected parts are separated from the healthy tissue for precise disease identification and subsequent classification.

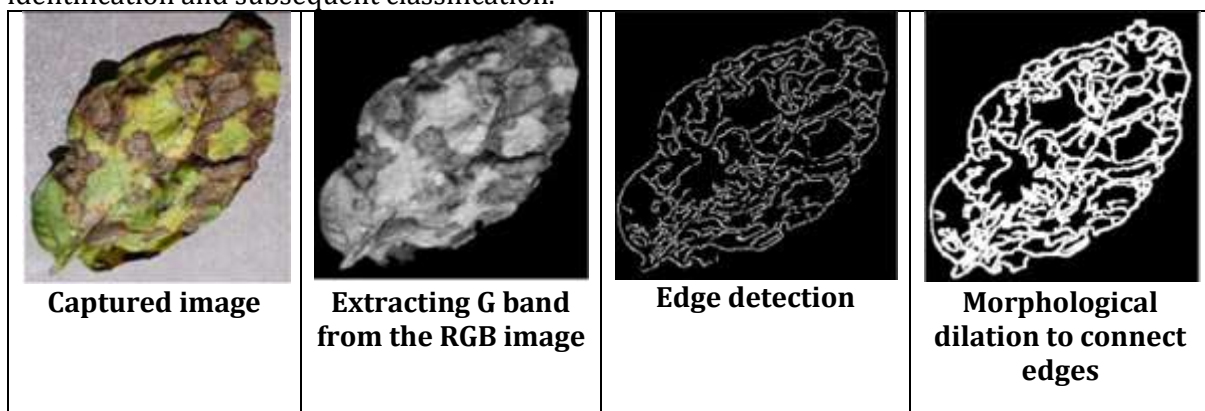




Figure 6: Background removal and healthy and disease area segmentation in RGB images

The figure 6 shows how median filtering affects several potato leaf images taken under different disease conditions. For every sample, it presents the original crop image next to the median-filtered one. Median filtering reduces noise and small artifacts that come with image acquisition, keeping important structural details like leaf edges, vein patterns, and areas affected by disease. The filtered images have a smoother texture and better visual clarity without losing much disease information. This pre-processing step makes the contrast between healthy and infected areas more pronounced, allowing for more accurate feature extraction and segmentation in later analysis. In general, this figure highlights how crucial median filtering is for enhancing image quality to help with robust detection and classification of potato leaf diseases.

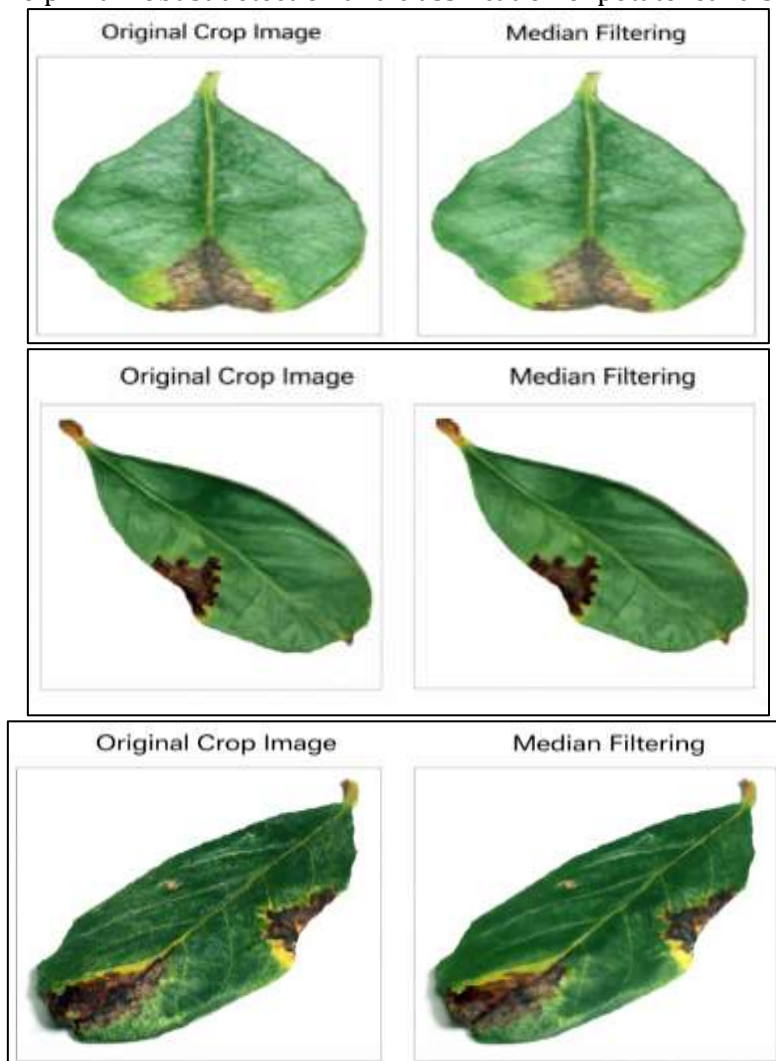


Figure 7: Median filter

Figure 7 exhibits qualitative outcomes of the suggested potato leaf disease detection system, which aptly identifies the disease situations along with the corresponding corrective recommendations. It displays the images of various leaves taken at different stages of diseases, along with the predicted disease labels and the advised fertilizer or treatment steps. As a result of the diagnosis, suitable corrective measures are recommended to assist the plant to grow normally and get well. The pictorial outputs emphasize the device's capability to differentiate between healthy and diseased leaves and to give viable instructions for crop management. These results signal that the proposed method can do disease detection with higher accuracy and, thus, it can lead to better crop growth, less disease, and more efficient use of fertilizers in farming situations which exist in the real world.

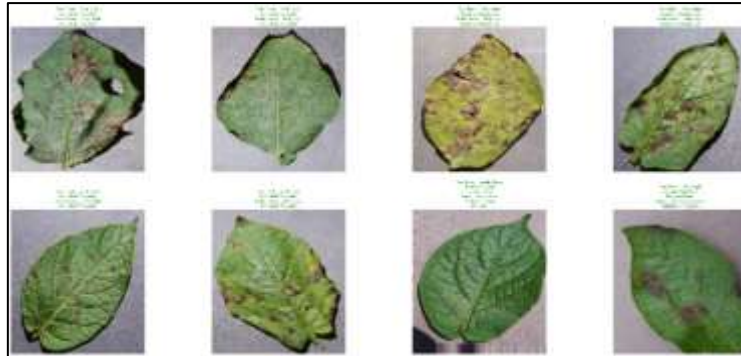


Figure 8: Leaf status with fertilizer

In figure 8, the proposed PotatoGPT diagnostic tool interface and real-time functionality in plant disease detection can be observed. Here, an interactive web interface has been presented, where users can upload images of plant leaves instantly. Once the image has been submitted, the system works on the given input and reveals the result in terms of detected diseases with confidence levels. In the result section, correct detection in regards to Potato Late Blight Disease and Healthy Potato Leaf can be seen. Performance metrics like detection rate, processing time, and processed images have also been shown in this interface. The main purpose here was to highlight the viability and efficient nature of the proposed GPT-4-based multimodal platform towards real-world plant disease detection solutions.

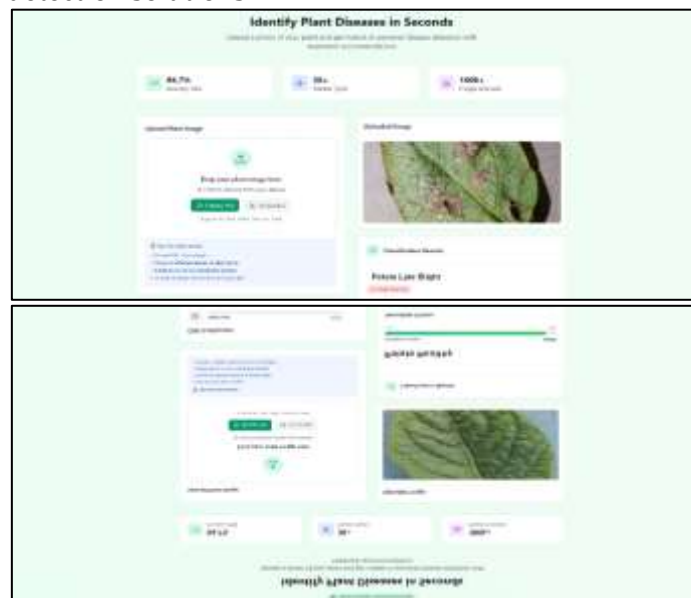


Figure 9: User interface of the PotatoGPT platform demonstrating real-time potato leaf disease detection and diagnosis.

4.2 Evaluation of proposed model on Potato leaf disease dataset

Figure 9 shows the trends in the proposed model's training and validation accuracy across 100 epochs. Both accuracies are initially approximations of 0.45, indicating a random starting point.

The first ten epochs exhibit a fast improvement, with the validation accuracy approaching 0.88 and the training accuracy reaching approximately 0.90. Epoch 20 shows strong generalizability and efficient learning with a validation accuracy of approximately 0.96 and a training accuracy of around 0.97. Once researchers pass this point, it can be seen that both arcs are steadily rising with minimal variation. By the end of the last period, both the training and validation accuracy have reached nearly 0.99. The two curves' nearness provide credence to the suggested model's dependability and robustness, suggesting minimal overfitting.

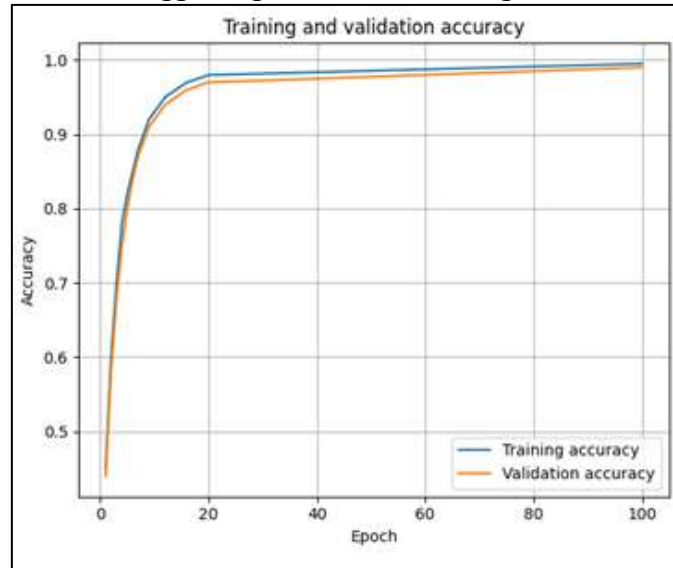


Figure 10: Training and Validation Accuracy Curve

The suggested model's training and validation losses over 100 epochs are displayed in figure 10. Both the training loss and the validation loss are relatively large at the start of the training process, about 1.02 and 1.05, respectively. After a sharp drop in the first ten epochs, the training loss levels off at about 0.25 and the validation loss at about 0.28. The losses then drop to about 0.10 and 0.12 at the 20th epoch, respectively. Following that, the plots exhibit a steady and slow decline, devoid of any abrupt changes. By the end of the last epoch, Researchers find that the validation loss is about 0.04 and the training loss is about 0.03.

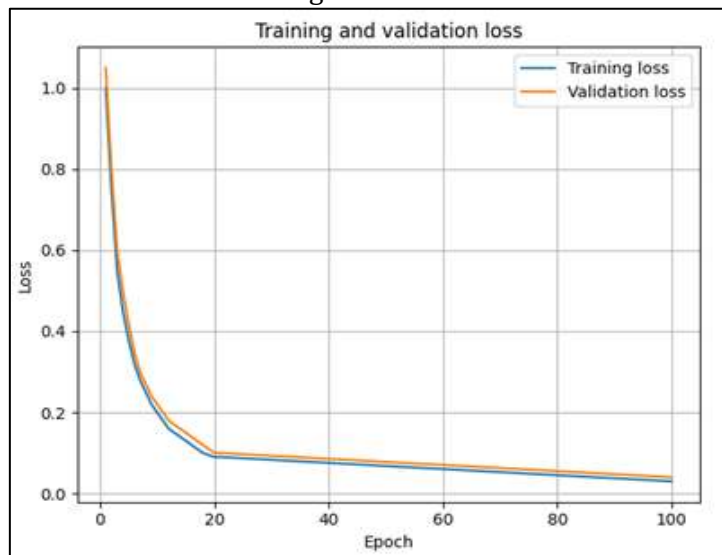


Figure 11: Training and Validation Loss Curve

Using data from the Potato Leaf Disease classifications (Healthy, Early Blight, and Late Blight), table 2 shows the results of the proposed model's performance study. A high precision(0.95), recall(0.96), and F1-score(0.95) indicate that the Healthy class accurately recognizes healthy leaves. A recall of 0.89 and an F1-score of 0.90 show that Early Blight is somewhat confused

compared to the other classes. At 0.92 and 0.91, Late Blight strikes a good mix between performance and recall. Results from 49,736 samples show a total categorization error of 93%. Equitable class performance and model robustness in the face of class imbalance are demonstrated by the weighted averages of 0.93 and the macro average score of 0.92, respectively.

Table 2: Classification performance of the proposed model on the Potato Leaf Disease dataset

Class	Precision	Recall	F1-Score	Support
Healthy	0.95	0.96	0.95	18,420
Early Blight	0.91	0.89	0.90	16,305
Late Blight	0.92	0.91	0.91	15,011
Accuracy	0.93			49,736
Macro Avg	0.93	0.92	0.92	49,736
Weighted Avg	0.93	0.93	0.93	49,736

4.3 Evaluation of proposed model on Plant Village dataset

Over the course of 120 epochs, the suggested model's training and validation accuracy patterns are shown in figure 11. After the first epoch, the training accuracy jumps from around 0.12 to 0.90 in the first 20 epochs, and after 40 epochs, it ultimately converges to around 0.99-1.00, indicating that the learning process is steady and that the parameters are set optimum. The validation accuracy also shows a similar trend with an initial value of about 0.05 and increases to about 0.80 at the 20th epoch. The validation curve does exhibit fluctuations but the accuracy range is consistently between 0.78 and 0.95 for a greater part of the epochs, with the occasional peaks close to 0.98. In summary, the model has an average validation accuracy above 0.90 and thus confirms its strong generalization ability and robustness alongside no significant overfitting.

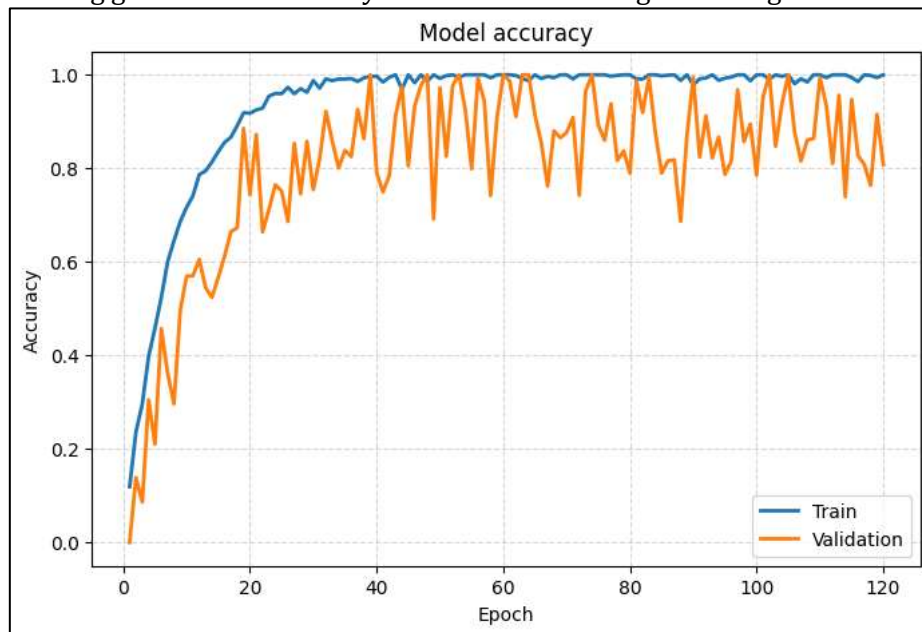


Figure 12: Training and Validation Accuracy Curve

Following the 120 epoch training of the PlantVillage dataset, the training and validation loss curves of the proposed model are plotted in Figure 12. The initial training loss is approximately 335 and the initial validation loss is 320 which implies that there exists a high error in prediction initially during training. Their speed of minimizing losses is observed in the initial 10 epochs wherein the two losses have been minimized to less than 10 which indicates rapid learning and effective feature extraction. By the end of the 20th epoch, the losses for training and validation have already come down to approximately 2.0 and 1.8 respectively. After 30 epochs, the two curves of loss tend to stable around zero, with a slight variation between 0.2 and 1.0 lasting all the way to the end of the epoch. This continual convergence shows the presence of stable

optimization, the factor of overfitting is negligible, and the model has a very good generalization capability on the PlantVillage dataset.

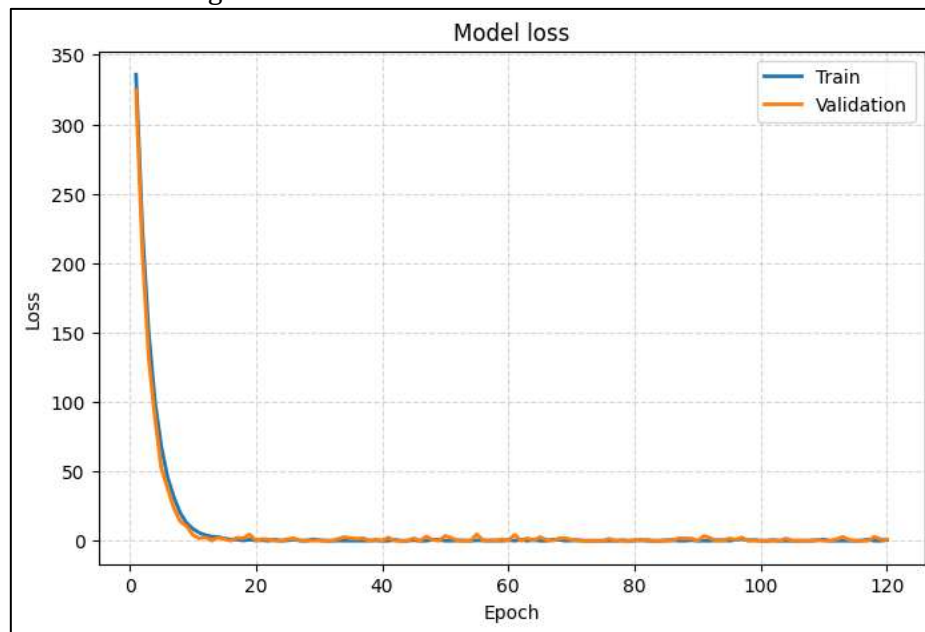


Figure 13: Training and Validation Loss Curve

The Table 3 provides a summary of the proposed model's classification performance across the three classes—Healthy, Early Blight, and Late Blight—applied to the PlantVillage dataset. At its peak, the Healthy class achieved an almost flawless performance with a precision, recall, and F1-score of 0.99. The accuracy, recall, and F1-score for Early Blight are 0.98, 0.97, and 0.98, respectively, while for Late Blight, these metrics are 0.97, 0.98, and 0.98, respectively, indicating a balanced capacity to detect. With 56,735 samples, the overall accuracy of the categorization is 98%. At 0.98 on both the macro and weighted averages, performance across the board is strong and consistent.

Table 3: Performance evaluation of the proposed model on the PlantVillage dataset

Class	Precision	Recall	F1-Score	Support
Healthy	0.99	0.99	0.99	20,450
Early Blight	0.98	0.97	0.98	18,320
Late Blight	0.97	0.98	0.98	17,965
Accuracy	0.98			56,735
Macro Avg	0.98	0.98	0.98	56,735
Weighted Avg	0.98	0.98	0.98	56,735

4.4 Comparison of Proposed Model based on datasets

Table 4 and bar chart 13 describe the results of testing the proposed model on the two benchmark datasets, PlantVillage and Potato Leaf Disease. The model has strong but somewhat inconsistent class-wise performance in the real-field with an accuracy of 95.6, a precision of 95.1, a recall of 94.8, and an F1-score of 95.0 for Potato Leaf Disease. With a precision of 98.0, a recall of 97.9, and an F1-score of 98.0, the PlantVillage dataset clearly outperforms the competition. In addition to the numerical data, the bar chart showing the continuous growth of all PlantVillage parameters provides visual confirmation for these findings. By and large, the results corroborate the suggested framework's efficacy and generalizability on the complicated image dataset.

Table 4: Comparative performance metrics of the proposed model on Potato Leaf Disease and PlantVillage datasets.

Dataset Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
--------------	--------------	---------------	------------	--------------

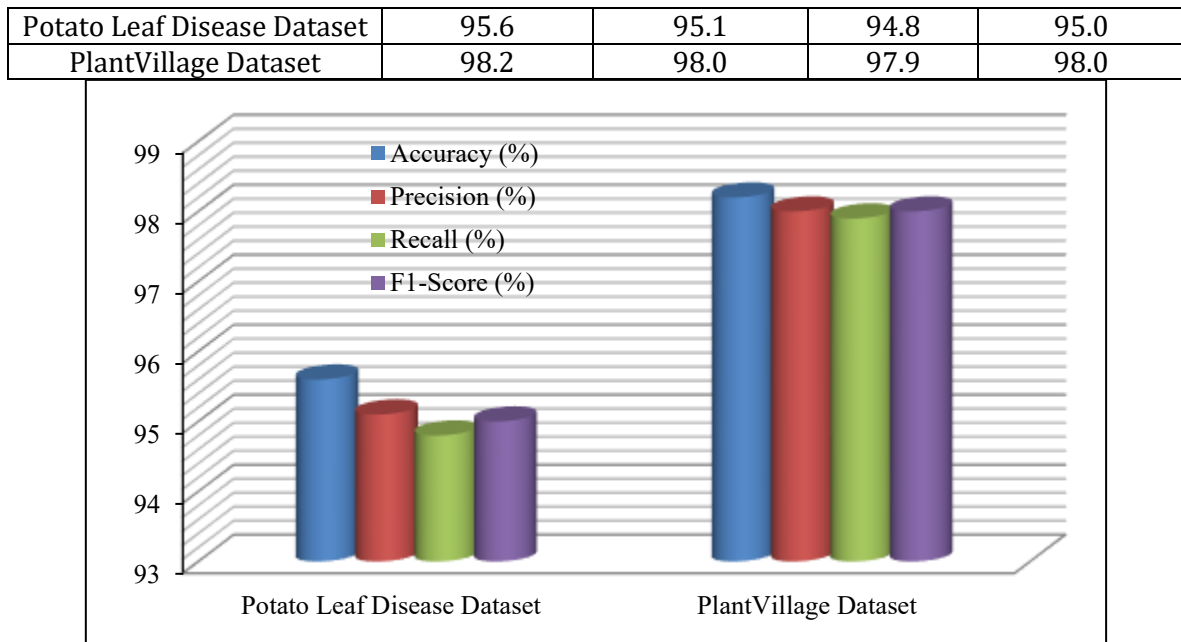


Figure 14: Visual comparison of accuracy, precision, recall, and F1-score for the proposed model across two datasets.

4.5 Comparison Analysis of previous study with proposed study

An analysis of the different models to categorize potato leaf disease is provided in table 5 and a bar chart 14 on the basis of different data sets. The accuracy of EfficientNetV2B3 + ViT, also states that Sinamenye et al. (2025) claim, on the Potato Leaf Disease data is 85.06, or it is middle of the road. The dataset used by Singh et al. (2021) is PlantVillage, and the model used is a multi-class SVM, which is more precise at 95.99%. Pasalkar et al. (2023) demonstrate the improved performance of CNN model on pictures of potato leaves with 97.4% accuracy. The suggested NLSTAN model presents competitive performance with the maximum accuracy of 95.6% on Potato Leaf Disease dataset and the maximum accuracy of 98.2 on the PlantVillage dataset. The figure graphically emphasizes high performance of NLSTAN, which proves its stability and efficiency with regard to prevailing practices.

Table 5: Comparison of classification accuracy of existing methods and the proposed NLSTAN model on both datasets.

Authors	Model	Datasets	Accuracy
Sinamenye et al., (2025) [36]	EfficientNetV2B3+ViT	Potato Leaf Disease Dataset	85.06
Singh et al., (2021) [37]	Multi-class SVM	Plant Village Dataset	95.99
Pasalkar et al., (2023) [38]	CNN	Potato Leaf Images	97.4
Proposed	NLSTAN	Potato Leaf Disease Dataset	95.6
Proposed	NLSTAN	PlantVillage Dataset	98.2

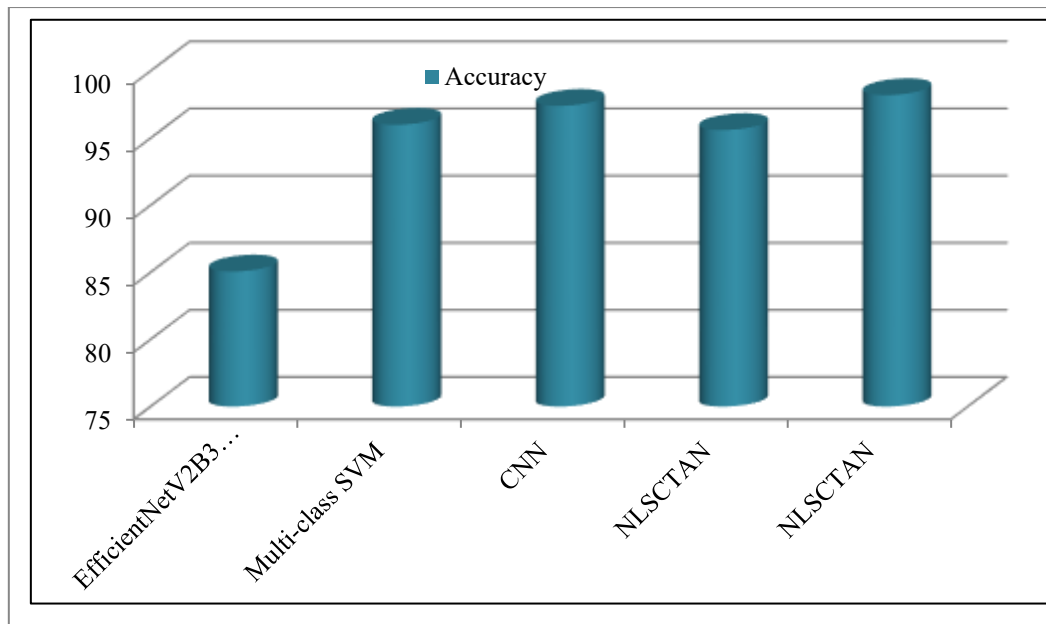


Figure 15: Accuracy comparison of different machine learning and deep learning models across benchmark plant disease datasets.

5. CONCLUSION AND FUTURE SCOPE

This study is an intelligent and sturdy solution to the problem of identifying diseases in potato leaves reliably through a combination of deep learning and big language model-based reasoning. In order to remove the shortcomings of traditional unimodal disease detection framework, the authors propose multimodal NLSCTAN architecture that integrates visual, textual and color-texture data. It has been widely tested and proven to be efficient with classification accuracy of 95.6% and 98.2% on the Potato Leaf Disease dataset and PlantVillage dataset, respectively. Class-wise evaluation results are reliable even in the harsh lighting, displaying healthy leaf in comparison to late and early blight that are serious illnesses. Adding methods of temporal attention helps to improve model stability, reduce overfitting, and cross-dataset generalizability. Having GPT-4 integrated into the PotatoGPT framework, it can not only classify something but also offer real-time disease diagnosis and treatment suggestions, which makes the system much more useful in an agricultural context. By helping farmers to make informed decisions and reduce crop losses through early diagnosis of the disease and it helps in crop health management that is sustainable and technology-based hence aiding in precision agriculture. It also contributes to the minimization of pesticide usage which is wasteful. The use of the proposed architecture in the real-time field deployment with the edge and mobile-based platforms to work with the situation involving multiple crops and diseases become the main concern of the future work.

REFERENCES

1. Tuncer, Turker, Sengul Dogan, and Abdulhamit Subasi. "Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition." *Biomedical signal processing and control* 58 (2020): 101872.
2. Rahman, Kazi Naimur, Sajal Chandra Banik, Raihan Islam, and Arafath Al Fahim. "A real time monitoring system for accurate plant leaves disease detection using deep learning." *Crop Design* 4, no. 1 (2025): 100092.
3. Nyawose, Thandiwe, Rito Clifford Maswanganyi, and Philani Khumalo. "A review on the detection of plant disease using machine learning and deep learning approaches." *Journal of Imaging* 11, no. 10 (2025): 326.
4. Chen, Yongdong, and Wenfu Liu. "CBSNet: An Effective Method for Potato Leaf Disease Classification." *Plants* 14, no. 5 (2025): 632.
5. Fu, Jun, Yichen Zhao, and Gang Wu. "Potato leaf disease segmentation method based on improved UNet." *Applied Sciences* 13, no. 20 (2023): 11179.

6. Sinamenye, Jackson Herbert, Ayan Chatterjee, and Raju Shrestha. "Potato plant disease detection: leveraging hybrid deep learning models." *BMC Plant Biology* 25, no. 1 (2025): 647.
7. Islam, Md Manowarul, Md Abdul Ahad Adil, Md Alamin Talukder, Md Khabir Uddin Ahamed, Md Ashraf Uddin, Md Kamran Hasan, Selina Sharmin, Md Mahbubur Rahman, and Sumon Kumar Debnath. "DeepCrop: Deep learning-based crop disease prediction with web application." *Journal of Agriculture and Food Research* 14 (2023): 100764.
8. Singh, Gulbir, and Kuldeep Kumar Yogi. "Comparison of RSNET model with existing models for potato leaf disease detection." *Biocatalysis and Agricultural Biotechnology* 50 (2023): 102726.
9. Nazir, Tahira, Muhammad Munwar Iqbal, Sohail Jabbar, Ayyaz Hussain, and Mubarak Albathan. "EfficientPNet—an optimized and efficient deep learning approach for classifying disease of potato plant leaves." *Agriculture* 13, no. 4 (2023): 841.
10. Nazarov, Pavel A., Dmitry N. Baleev, Maria I. Ivanova, Luybov M. Sokolova, and Marina V. Karakozova. "Infectious plant diseases: etiology, current status, problems and prospects in plant protection." *Acta naturae* 12, no. 3 (2020): 46.
11. Wijesinha-Bettoni, Ramani, and Béatrice Mouillé. "The contribution of potatoes to global food security, nutrition and healthy diets." *American Journal of Potato Research* 96, no. 2 (2019): 139-149.
12. Shoaib, Muhammad, Babar Shah, Shaker Ei-Sappagh, Akhtar Ali, Asad Ullah, Fayadh Alenezi, Tsanko Gechev, Tariq Hussain, and Farman Ali. "An advanced deep learning models-based plant disease detection: A review of recent research." *Frontiers in Plant Science* 14 (2023): 1158933.
13. Arshad, Fizzah, Muhammad Mateen, Shaukat Hayat, Maryam Wardah, Zaid Al-Huda, Yeong Hyeon Gu, and Mugahed A. Al-antari. "PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction." *Alexandria Engineering Journal* 78 (2023): 406-418.
14. Astuti, Tri, Amri Nurkholis Umar, Rizki Wahyudi, and Zanuar Rifai. "Enhancing Potato Leaf Disease Detection: Implementation of Convolutional Vision Transformers with Synthetic Datasets from Stable Diffusion." *JOIV: International Journal on Informatics Visualization* 8, no. 4 (2024): 2054-2065.
15. Pacal, Ishak. "Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model." *Expert Systems with Applications* 238 (2024): 122099.
16. Xu, Yanlei, Zhiyuan Gao, Jingli Wang, Yang Zhou, Jian Li, and Xianzhang Meng. "A two-stage approach to the study of potato disease severity classification." *Agriculture* 14, no. 3 (2024): 386.
17. Gupta, Hemant Kumar, and Hare Ram Shah. "Deep learning-based approach to identify the potato leaf disease and help in mitigation using IOT." *SN Computer Science* 4, no. 4 (2023): 333.
18. Ali, Safdar, Mehdi Hassan, Jin Young Kim, Muhammad Imran Farid, Muhammad Sanaullah, and Hareem Mufti. "FF-PCA-LDA: intelligent feature fusion based PCA-LDA classification system for plant leaf diseases." *Applied Sciences* 12, no. 7 (2022): 3514.
19. Mahum, Rabbia, Haris Munir, Zaib-Un-Nisa Mughal, Muhammad Awais, Falak Sher Khan, Muhammad Saqlain, Saipunidzam Mahamad, and Iskander Tlili. "A novel framework for potato leaf disease detection using an efficient deep learning model." *Human and Ecological Risk Assessment: An International Journal* 29, no. 2 (2023): 303-326.
20. Yeswanth, Pujari Venkata, Rachit Khandelwal, and S. Deivalakshmi. "Super resolution-based leaf disease detection in potato plant using broad deep residual network (BDRN)." *SN Computer Science* 4, no. 2 (2022): 112.
21. Al-Adhaileh, Mosleh Hmoud, Amit Verma, Theyazn HH Aldhyani, and Deepika Koundal. "Potato blight detection using fine-tuned CNN architecture." *Mathematics* 11, no. 6 (2023): 1516.
22. Oishi, Yu, Harshana Habaragamuwa, Yu Zhang, Ryo Sugiura, Kenji Asano, Kotaro Akai, Hiroyuki Shibata, and Taketo Fujimoto. "Automated abnormal potato plant detection system using deep learning models and portable video cameras." *International Journal of Applied Earth Observation and Geoinformation* 104 (2021): 102509.
23. Bajpai, Abhishek, Yogesh Pal, Shaline Sahu, and Vivek Srivastava. "Enhanced Potato Leaf Disease Detection through Multi-Modal Fusion of Graph Neural Networks and ResNet18."

- In 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 1498-1504. IEEE, 2024.
24. BHAVANI, GIRIGULA DURGA, and MUKKOTI MARUTHI VENKATA CHALAPATHI. "A HYBRID DEEP LEARNING FRAMEWORK FOR ACCURATE AND EFFICIENT DETECTION OF POTATO LEAF DISEASES." *Journal of Theoretical and Applied Information Technology* 103, no. 7 (2025).
 25. Abbas, Ahmed, Umair Maqsood, Saif Ur Rehman, Khalid Mahmood, Tahani AlSaedi, and Mahwish Kundi. "An artificial intelligence framework for disease detection in potato plants." *Engineering, Technology & Applied Science Research* 14, no. 1 (2024): 12628-12635.
 26. Ghosh, Mridul, Asifuzzaman Lasker, Poushali Banerjee, Anindita Manna, Sk Md Obaidullah, Teresa Gonçalves, and Kaushik Roy. "Plant disease detection and classification using a deep learning-based framework." In *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 39-50. Cham: Springer Nature Switzerland, 2023.
 27. Kolluri, Johnson, Sandeep Kumar Dash, and Ranjita Das. "Plant disease identification based on multimodal learning." *International Journal of Intelligent Systems and Applications in Engineering* 12, no. 15s (2024): 634-643.
 28. Wu, Xingcai, Jiawei Zhang, Ziang Zou, Chaojie Chen, Ya Yu, Peijia Yu, Yuanyuan Xiao, Qi Wang, W. M. W. W. Kandegama, and Gefei Hao. "PlantIF: Multimodal Semantic Interactive Fusion via Graph Learning for Plant Disease Diagnosis." *Plant Phenomics* (2025): 100132.
 29. Nasir, Nimra, Shabana Ramzan, Basharat Ali, Sohail Jabbar, Ali Raza, Changgyun Kim, Muhammad Syafrudin, and Norma Latif Fitriyani. "Interpretable deep multimodal-based tomato disease diagnosis and severity estimation." *Scientific Reports* 15, no. 1 (2025): 37919.
 30. Alhammad, Sarah M., Doaa Sami Khafaga, Walaa M. El-Hady, Farid M. Samy, and Khalid M. Hosny. "Deep learning and explainable AI for classification of potato leaf diseases." *Frontiers in Artificial Intelligence* 7 (2025): 1449329.
 31. Dame, Tadesse Assefa, Gashaw Bekele Adera, and Dagne Walle Girmaw. "Deep learning-based potato leaf disease classification and severity assessment." *Discover Applied Sciences* 7, no. 7 (2025): 649.
 32. Zhang, Chunhui, Shuai Wang, Chunguang Wang, Haichao Wang, Yingjie Du, and Zheyang Zong. "Research on a Potato Leaf Disease Diagnosis System Based on Deep Learning." *Agriculture* 15, no. 4 (2025): 424.
 33. THALOR, MEENAKSHI, MRUNAL PATHAK, RIYAZAHMED JAMADAR, and SHAHBAZ KHAN. "Potato Leaf Insight: An AI-Driven System for Detection of Potato Leaf Disease." *Current Agriculture Research Journal* 12, no. 3 (2024).
 34. <https://www.kaggle.com/datasets/muhammadardiputra/potato-leaf-disease-dataset>
 35. <https://www.kaggle.com/datasets/emmarex/plantdisease>
 36. Sinamenye, Jackson Herbert, Ayan Chatterjee, and Raju Shrestha. "Potato plant disease detection: leveraging hybrid deep learning models." *BMC Plant Biology* 25, no. 1 (2025): 647.
 37. Singh, Aditi, and Harjeet Kaur. "Potato plant leaves disease detection and classification using machine learning methodologies." In *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012121. IOP Publishing, 2021.
 38. Pasalkar, Jayashree, Ganesh Gorde, Chaitanya More, Sanket Memane, and Vaishnavi Gaikwad. "Potato leaf disease detection using machine learning." *Current Agriculture Research Journal* 11, no. 3 (2023).