ISSN: 2093-5374 / E-ISSN: 2093-5382

Hybrid CNN–GRU Architecture for Early Plant Disease Diagnosis Using Sequential Crop Images

N. Madhuri[1*], and Dr. Loshma Gunisetti[2]

[1*]Research Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India.
madhuri.namala@gmail.com, https://orcid.org/0009-0008-5916-1788

[2]Professor and Head, Department of AIML, Sri Vasavi Engineering College, Tadepalligudem, Andhra Pradesh, India. loshma@gmail.com, https://orcid.org/0000-0002-8664-7619

Received: February 21, 2026; Revised: April 01, 2026; Accepted: May 18, 2026; Published: June 30, 2026

Abstract

Agricultural productivity, crop quality, and food security worldwide can be highly impacted by plant diseases. Early detection and accurate diagnosis of disease in crops are vital for minimizing losses and sustaining precision farming practices. Regrettably, all modern disease diagnosis approaches based on deep learning techniques have concentrated on image classification, neglecting the time dependency of the disease process during different stages of the crops' development cycle. Furthermore, traditional CNN-LSTM models have been associated with increased computational complexities and high memory costs. This research suggests a CNN-GRU hybrid model for early disease detection using sequential analysis of crop images. The suggested technique involves integrating the CNN and GRU, which helps the network develop the capability to learn spatiotemporal information on the crop plant disease. Datasets of sequential crop images that represent the progression stages of the diseases were collected from Plant Village and augmented crop images. The CNN portion of the model is responsible for extracting spatiotemporal characteristics of the diseases, including lesions, discolored parts, and texture changes. It is evident that the suggested Hybrid CNN-GRU method has higher accuracy (94.1%), precision (94.0%), recall (93.9%), and F1-score (94.0%) compared to CNN and CNN-LSTM models. The validation of efficacy and robustness of the suggested approach has been confirmed through standard deviation, paired t-test, and 5-fold cross-validation. Furthermore, the method demonstrated high scalability and strong tolerance to variations in light intensity, noise, and images from the fields of crop plants.

Keywords: Plant Disease Detection, Deep Learning, CNN, GRU, Sequential Images, Precision Agriculture, Temporal Modeling, Early Diagnosis.

1 Introduction

India's agricultural sector provides an important contribution to its national economy; nonetheless, crop diseases create substantial losses (Savary et al., 2012). The losses incurred from plant diseases that arise from fungus, bacteria, and viruses amount to approximately 40% per year in South Asia (Chakraborty & Newton, 2011). Manual examination by agronomists has been used to detect diseases, which involves a costly and labor-intensive procedure.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 17, number: 2 (June-2026), pp. 478-498. DOI: *10.58346/JOWUA.2026.I2.027*

*Corresponding author: Research Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India.

The emergence of deep learning models, specifically CNNs, has opened the door for automated disease detection using leaf imagery. CNNs trained on the Plant Village dataset could be more than 99% accurate when applied under controlled settings (Mohanty et al., 2016). Recent advancements include the use of transfer learning, designing efficient networks for mobile applications, and attention-based models (Ahmad et al., 2021; Fuentes et al., 2017). However, a critical drawback exists in these approaches: all current models consider leaf images as static shots, ignoring the temporal component in detecting diseases (Taji et al., 2024). Recurrent neural networks (RNNs) such as LSTM and GRU networks have performed well on sequential data tasks in areas such as speech and activity recognition (Greff et al., 2016; Wang et al., 2026). Incorporation into CNN-based spatial encoders in CNN-RNN networks has proven successful for video and action recognition, yet has received limited application for disease detection in agriculture.

This paper fills this gap by suggesting a Hybrid CNN-GRU framework, which learns from temporally ordered sequences of crop leaves. The key contributions of this research include:

- The creation of a Hybrid CNN-GRU framework, which is a deep-learning approach to classifying plant diseases based on both spatial feature extraction and temporal modeling.
- The building of a sequential dataset of 12,000 sequences of six temporally ordered crop leaves per disease category.
- The empirical proof of early detection of disease at up to three intervals ahead, compared to single-frame CNN models.
- The analysis of the impact of the CNN encoder depth, GRU layer count, and sequence length on the effectiveness of the proposed algorithm.

Organization of the remaining paper is as follows: Section 2 provides a brief description of the literature review and issues related to diagnosing plant diseases using existing methods. There are also explanations regarding the problem under study, the innovation, and the hybrid CNN–GRU model. The dataset construction process is outlined in Section 3, encompassing data acquisition, pre-processing, ordering of temporal sequence, and classification of disease severity Section 4 outlines the proposed methodology and describes the experiment design and data preprocessing technique. In Section 5, there are experimental results, performance, and scalability and robustness analysis of the proposed model. Section 6 presents an in-depth discussion of the experimental results, advantages, limitations, and practical implications of the proposed framework. Lastly, Section 7 summarizes and suggests potential areas for future work.

2 Related Work

In recent times, there have been developments in deep learning algorithms and computer vision that have greatly enhanced plant disease identification. The PlantVillage dataset was created for the first time using AlexNet and GoogLeNet architecture, which attained 99.35% accuracy in lab-controlled environments (Martinelli et al., 2015). Nevertheless, there was a notable drop in the performance of these algorithms when used in an actual field environment because of the presence of varied lighting conditions and environmental noise (Kamilaris & Prenafeta-Boldú, 2018). To solve the problem of classification limitations, various deep CNNs were investigated, and it was found that VGG-based algorithms outperformed shallower ones in diagnosing crop diseases (Pandian et al., 2022). Transfer learning methods decreased the need for large labeled datasets by using fine-tuned pre-trained ImageNet models (Kaur et al., 2023).

Furthermore, numerous studies have addressed the topic of developing enhanced CNN architectures to enhance the detection of plant diseases. Karen Simonyan & Andrew Zisserman presented the VGG architecture that could achieve improved feature extraction capability with deeper convolutional layers. The Residual Learning frameworks, including the ResNet, allowed designing deeper neural networks through the application of skip connections to mitigate the issue of disappearing gradients (He et al., 2016). DenseNet architectures enhanced feature reutilization through dense connectivity operations (Huang et al., 2017). Despite improving classification accuracy, these algorithms mainly concentrated on the analysis of static images without taking into account the disease progression.

Also, the methods related to object detection and segmentation for precise disease localization were researched. Specifically, Joseph Redmon proposed the YOLO framework, which made real-time disease localization with improved processing speed (Owomugisha & Mwebaze, 2016). The Mask R-CNN framework provided the possibility of pixel-level localization of infection areas and hence achieved improved performance in localization. Nevertheless, these approaches required significant computational capacity and large annotated datasets for training purposes (Liu & Wang, 2021).

Instead of using sophisticated neural models, earlier crop analysis techniques mostly relied on traditional categorization techniques. Crop attributes that were manually determined, such as surface texture, color distribution, and structural appearance features, were processed using techniques including SVM, Random Forest, and KNN (Daniya & Vigneshwari, 2022).

The temporal sequence learning algorithm has emerged as an important tool for the analysis of disease progression over time. To address the issue of disintegrating gradients, Sepp Hochreiter and Jürgen Schmidhuber proposed the LSTM network (Hochreiter & Schmidhuber, 1997). GRU, which made use of the gating mechanism but with a simpler design than that of the LSTM network (Mienye et al., 2024).

GRU-based models showed superior performance in predicting diseases based on sequences of EHRs (Zhang et al., 2019). The same sequence learning methodology was later applied to agricultural forecasting for tasks like crop yield prediction, rainfall estimation, and climatic conditions. Most of the work done in agriculture was related to time series of numbers, but not time series of images showing progressions of diseases (Vijayan & Chowdhary, 2025).

A recent trend in agricultural literature emphasized the incorporation of temporal data in disease surveillance systems. Anastasios Kamilaris and Francesc X. Prenafeta-Boldú provided an overview of deep learning techniques in agriculture (Zhao et al., 2021). In the current literature, temporal disease analysis systems have not been considered, since the traditional methods focus on disease detection based on a single image while ignoring the progression of the disease symptoms at various plant growth stages. CNNs with transfer learning were used to classify diseases in tomatoes, but temporal information was ignored (Nawaz et al., 2020; Liu et al., 2020).

Some recent research works introduced attention models and transformer networks in agricultural disease diagnostic systems. For instance, VIT designs were introduced in Sarkar et al., (2023). These models offered improvements in terms of global feature representation via self-attention. Additionally, hybrid CNN-transformer models were developed with the aim of improving classification accuracy (Kanna et al., 2023; Naralasetti & Bodapati, 2024). Nevertheless, transformer models typically need extensive datasets and high computational capabilities for training purposes (Shoaib et al., 2023).

3 Dataset Construction

3.1 Data Sources

The data set was compiled from two sources: (i) Plant Village data set, including 54,305 photos shot in artificial light showing both healthy and sick plant leaves, and (ii) the Hybrid CNN-GRU Architecture for Early Plant Disease Diagnosis Using Sequential Crop Images, which involved the acquisition of crop field images using a 12-megapixel smart phone camera under natural lighting in the Indian state of Andhra Pradesh.

3.2 Sequence Construction

Every sample consists of a temporal sequence of T=6 leaf images captured from the same plant at intervals of 48 hours. To simulate temporal evolution, healthy leaves were placed in the initial stages, whereas increasingly symptomatic leaves were arranged in later stages. For field-collected data, temporal ordering was verified using EXIF metadata. Each sample was represented as a tensor of size (6,224,224,3). After removing duplicate and poor-quality images, a total of 12,000 samples were obtained, of which 8,400, 1,800, and 1,800 samples were allocated for training, validation, and testing, respectively.

3.3 Data Augmentation

For improving generalization and overcoming class imbalance issues, data augmentation was done by applying random horizontal and vertical flips; rotation of the images within the range of [-25°, +25°], brightness and contrast jitters within ±20%, Gaussian noise addition with a standard deviation of 0.01, and cropping with resizing to 224 × 224 pixels. Data augmentations were performed on each frame in the sequence individually. There were 25200 sequences in the augmented dataset. Table 1 illustrates the distribution of data among various classes of plant diseases.

Table 1: Dataset distribution across disease classes and splits

Disease Class	Train	Validation	Test	Total Sequences	Augmented Train
Healthy	1400	300	300	2000	4200
Early Blight	1400	300	300	2000	4200
Late Blight	1400	300	300	2000	4200
Leaf Mold	1400	300	300	2000	4200
Bacterial Spot	1400	300	300	2000	4200
Powdery Mildew	1400	300	300	2000	4200
Total	**8400**	**1800**	**1800**	**12000**	**25200**

3.4 Research Gap, Novelty, and Contributions

Current approaches for plant disease diagnosis are mostly limited to static image classification tasks by leveraging CNN-based approaches while failing to model temporal disease development patterns over various crop growth stages. While some existing CNN–LSTM-based approaches help in addressing such issues, their major disadvantage includes increased computational complexity, higher memory requirements, and longer training duration. Moreover, current cutting-edge methods for diagnosing plant diseases are unable to achieve early detection of diseases and are incapable of providing disease localization capabilities. The proposed approach in this work leverages the concept of a hybrid CNN–GRU-based model for sequential crop image processing tasks by combining CNN-based spatial feature extraction techniques with GRU-based temporal dependencies learning for improved

performance over current state-of-the-art CNN–LSTM-based models. In addition, the Grad-CAM visualization technique is used to improve the interpretability of results.

3.5 Discussion on Technical Innovativeness and Novelty

The present paper offers technical improvements over current models employed in the classification of plant diseases. In this case, traditional CNNs rely on static crop images, unable to capture temporal progression trends. Even though the CNN-LSTM approach has enhanced sequential learning capacity compared to standard CNNs, it still suffers from high computational complexity, increased memory utilization, and extended learning time. On the other hand, the current model combines CNN for the spatial representation of features with GRU for the temporal learning component to process a sequence of crop images. Compared to LSTMs, GRUs utilize a reduced number of gates, fewer trainable parameters, and fewer computational resources, with the ability to converge at a fast rate. Furthermore, the model takes advantage of temporal crop image sequences to predict the occurrence of diseases. This is unlike existing algorithms that consider visual symptoms based on individual static images to classify diseases.

Moreover, the utilization of disease localization based on Grad-CAM ensures the identification of the infected parts of crop leaves, making the model more interpretable.

4 Proposed Methodology

4.1 Architecture Overview

The Hybrid CNN–GRU model comprises three major parts: (i) a CNN Spatial Encoder, (ii) a GRU Temporal Modeler, and (iii) a Classification Head. It is shown in figure 1. Given an input sequence $X = \{x_1, x_2, \ldots, x_t\}$, where each $x_t \in \mathbb{R}^{H \times W \times 3}$, the CNN encoder processes each frame independently to generate a fixed-length vector $h_t \in \mathbb{R}^d$. The resulting sequence $\{h_1, \ldots, h_T\}$ is passed through the GRU to extract temporal information and generate a compact form for classification.

4.2 CNN Spatial Encoder

The CNN model architecture employs three consecutive layers of convolutions, where each consists of a Conv2D layer (3×3), Batch Normalization, Rectified Linear Unit (ReLU) function, and 2×2 MaxPooling. Filter sizes are incremented from 32 to 64, and then to 128. Post the application of the final max-pooling layer, the Global Average Pooling (GAP) technique compresses the spatial dimensions to obtain a 128-dimensional feature vector for each frame.

Formally, for each frame x_t, the encoder computes f_t as shown in equation (1).

$$f_t = GAP(ConvBlock_3(ConvBlock_2(ConvBlock_1(x_t)))) \quad (1)$$

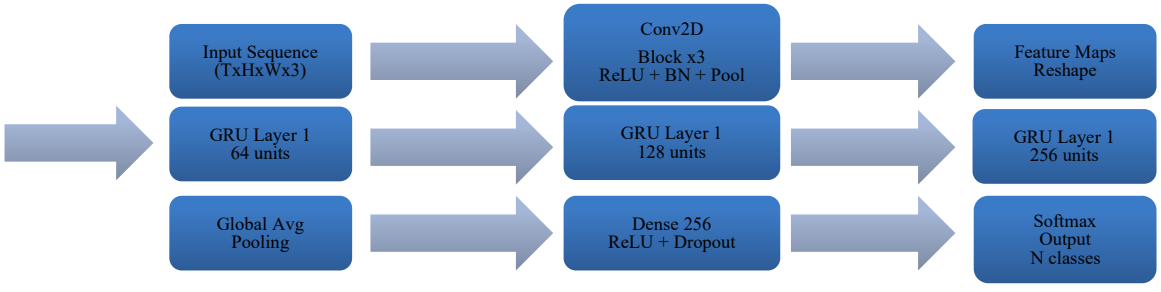

Figure 1: Hybrid CNN–GRU architecture: spatial feature extraction by CNN followed by temporal modeling via two-layer GRU

4.3 GRU Temporal Modeler

The extracted feature sequence $\{f_1, ..., f_t\}$ is processed by a two-layer GRU. The GRU update mechanism is governed by:

Update Gate

$$z_t = \sigma(W_z f_t + U_z h_{t-1} + b_z) \quad (2)$$

Reset Gate

$$r_t = \sigma(W_r f_t + U_r h_{t-1} + b_r) \quad (3)$$

Candidate State

$$\tilde{h_t} = tanh(W_h f_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (4)$$

Final Hidden State

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h_t} \quad (5)$$

The update gate is responsible for memory retention, as per equation (2), whereas the reset gate can be seen from equation (3). Hidden state candidates are obtained via equation (4), and finally, the hidden state is updated via equation (5).

4.4 Classification Head

The Dense layer, fully connected with 256 neurons and ReLU activation, is followed by the hidden state output from the second layer of GRU (h_T) to the Dropout layer (rate = 0.4). Then comes the Softmax layer with N = 6 neurons that denote the six disease categories, as shown in equation (6).

$$\hat{y} = Softmax(W_0 Dropout(ReLU(W_d h_T + b_d)) + b_0 \quad (6)$$

4.5 Training Configuration

This model was developed with TensorFlow 2.12 and Keras frameworks. The computations were performed with the help of an NVIDIA A100 80GB GPU. As can be seen from table 2, the Adam optimizer with an initial learning rate of 0.001 and the cosine annealing learning rate decay were chosen. Categorical cross-entropy was set as the loss function of the neural network.

Table 2: Configuration of hyperparameters for the proposed CNN-GRU model

Hyper Parameter	Value
Input sequence length (T)	6
Frame resolution	224 × 224 × 3
CNN filters (Block 1/2/3)	32 / 64 / 128
GRU Layer 1 units	128
GRU Layer 2 units	64
Dense layer units	256
Dropout rate (GRU / Dense)	0.30 / 0.40
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Initial learning rate	0.001
Batch size	32
Epochs (max)	50
Total trainable parameters	2.47 M

4.5.1. Algorithmic Formulation of Proposed Hybrid CNN-GRU Approach

Pseudo-code for Step-by-step Early Diagnosis of Plant Diseases

Input: Dataset of sequential crop images I ={I1,I2,I3,...,In}

Output: Disease classification prediction Dc and risk maps Rm.

Step 1: Data Collection

- Gather the plant village dataset of crop images.
- Collect images of crops affected by different diseases.
- Categorize the images based on types of crops and diseases.

Step 2: Preprocessing of Images

Preprocess crop images by:

- Changing their size to 256 x 256 pixels.
- Applying image normalization techniques.
- Using the Gaussian filter method for noise removal.
- RGB channel enhancement.

Image augmentation through:

- Rotation
- Flipping
- Zooming
- Adjusting brightness

Step 3: Sequencing Images Temporally

Cluster images on the basis of disease progression stages.

Generate the images' temporal sequence:

S = {It1,It2,...,Itk}

Where:

- S = Set of sequential images
- k = Number of temporal frames

Classify disease progression levels for generated sequences.

Step 4: CNN-Based Spatial Feature Extraction

Input sequential images into CNN layers.

Apply the convolution operation as shown in equation (7).

$$Fi = W * Ii + b \quad (7)$$

where:

- Fi = extracted feature map
- W = convolution kernel

- Ii = input image
- b = bias term

Apply ReLU activation as shown in equation (8).

$$R(x) = max(0, x) \quad (8)$$

- Perform max-pooling operation to reduce dimensionality.
- Generate spatial feature vectors: V={v1,v2,...,vk}.

Step 5: GRU-Based Temporal Learning

- Feed spatial feature vectors into the GRU layer.
- Update the GRU reset gate.
- Update the GRU update gate.
- Compute the candidate hidden state.
- Compute the final hidden state.

Step 6: Disease Classification

Pass GRU output into fully connected layers.

The softmax activation function is used to calculate the probability distribution of diseases represented by equation (9).

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \quad (9)$$

Predict disease category:

- Healthy
- Early Blight
- Late Blight
- Leaf Spot
- Mosaic Virus

Step 7: Disease Risk Map Generation

- Apply Grad-CAM visualization.
- Highlight infected regions in crop leaves.
- Generate disease risk heatmaps.

4.5.2 Workflow Description for Algorithm

The algorithm first prepares the images of crops before processing to improve their quality and ensure consistent dimensions. The sequential images of crop disease evolution over time are then formed to account for the differences in crop diseases at different stages. Feature vectors of a higher-level space are extracted by the CNN part of the architecture and passed to the GRU architecture to detect any temporal relationship in the formation of diseases. Finally, the fully connected layers produce predictions while Softmax classification predicts the disease class, whereas the Grad-CAM technique generates localizing heat maps of the detected diseases.

5 Experimental Results

5.1 Performance Evaluation

Common classification measures were used to assess how well the suggested Hybrid CNN-GRU network identified plant diseases. The effectiveness, dependability, and prediction capacity of the proposed deep learning architecture may be objectively assessed using classification metrics. As the classification of plant diseases is a multiclass classification problem with potential class imbalance issues among healthy and unhealthy plants, it is not advisable to depend only on one metric in reaching a conclusion.

The computation of the model's accuracy of classification is based on equation (10).

$$\text{Compute Accuracy:} \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Precision is computed using the expression indicated in equation (11).

$$\text{Compute Precision: } \frac{TP}{TP+FP} \quad (11)$$

Recall can be computed using equation (12).

$$\text{Compute Recall: } \frac{TP}{TP+FN} \quad (12)$$

F1-Score, the harmonic mean of Precision and Recall, is computed using equation (13).

$$\text{Compute F1-score:} \frac{2XPrecisionXRecall}{Precision+Recall} \quad (13)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

The experimental analysis shows that the suggested Hybrid CNN-GRU framework can be successfully used for early plant disease identification based on sequential crop images. For this purpose, the Plant Village data set, which includes pictures of both healthy and damaged crop leaves grouped chronologically, was used to train and test the framework. The accuracy of the suggested system has been measured based on some traditional measurement techniques. The experimental observations show that the use of CNN and GRU together boosts the performance of disease diagnosis at an early stage.

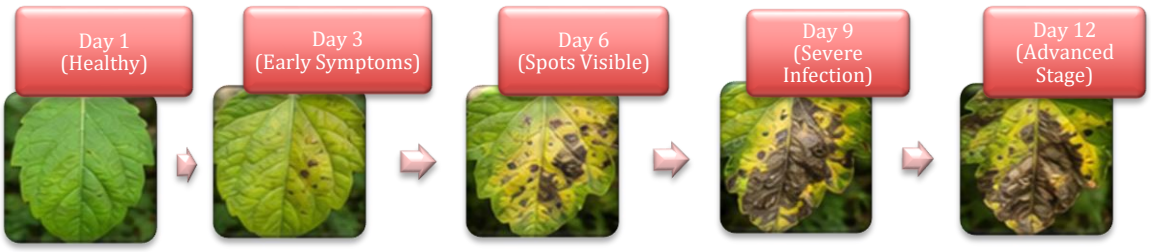

Figure 2: Sample input image sequences (tomato – early blight disease)

In figure 2 depicts the chronological development of the Early Blight disease in tomato leaf samples at various growth stages. The timeline starts with a healthy tomato leaf sample collected on Day 1, characterized by a green color and the absence of visible disease symptoms. On Day 3, early disease symptoms start to manifest in the form of small, yellowish spots and slight discoloration that are not easily discernible with the naked eye. At Day 6, there are clear signs of disease development, which can be seen through the formation of brown lesions and spots. The disease condition progresses to a severe level on Day 9, leading to massive necrosis and noticeable textural deformities. Finally, on Day 12, the disease condition has reached its advanced stage, which is evident from the extent of tissue damage.

Figure 3: Sample input image sequences (potato – late blight disease)

The sequential images of the potato leaves in figure 3 show the progress of Late Blight disease. At first, the leaf looks healthy, showing normal physical properties and colors. During the onset of symptoms, spots that are white in color and moisture-like in appearance begin to emerge because of the fungus. As the disease spreads, the development of dark lesions on the leaf causes the leaf to turn brown and be damaged. At the severe level of infection, the plant shows curled, dried-up, and dead leaves. In the advanced phase, extensive damage occurs because of leaf decay, causing a loss in crop production.

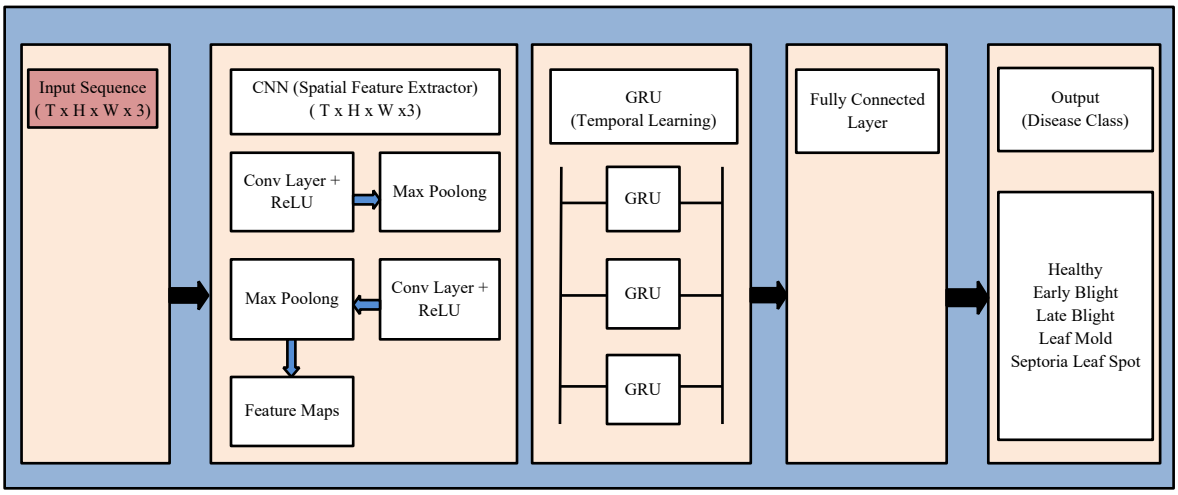

Figure 4: CNN–GRU model architecture (proposed)

In figure 4 depicts the topology of the proposed hybrid CNN-GRU model for spatiotemporal plant disease identification. First, the model takes as input a series of crop pictures that include RGB channels that represent the health state of the crop over time. The CNN element is responsible for extracting spatial characteristics, such as discoloration, deformity, and texture change. The convolution operation, activation function, and pooling layer are used to construct feature maps from the input image. The GRU layer then receives the retrieved spatial information and uses it to understand the temporal relationships and patterns of illness symptoms and evolution. The GRU model learns to capture the symptom

evolution process, crop health transition, and disease severity at a cheaper computing cost than the LSTM. Lastly, fully connected layers are used to categorize the diseases based on the symptoms into classes.

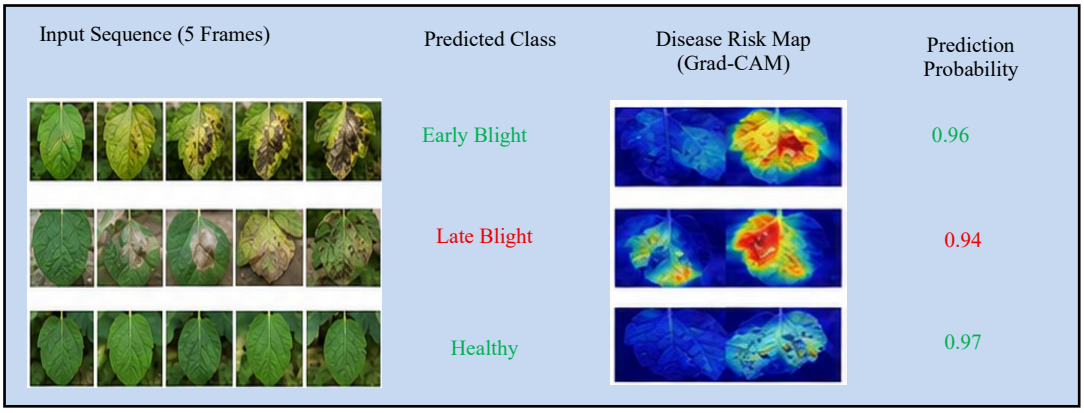

Figure 5: Prediction output examples

In figure 5 illustrates the example predictions performed by the CNN-GRU model in the case of varying disease conditions among crops. This figure consists of the following components: the sequence of crop images as input, disease classes predicted, disease risk maps, and probability of prediction (Mohan et al., 2023). The model predicts the classification of disease conditions on the basis of an image sequence indicating disease development stages, which could include Early Blight, Late Blight, or Healthy, for example. In order to make disease localization easy, the disease risk map using the Grad-CAM technique provides visual explanations by identifying infected zones of the leaves, with blue regions indicating healthy zones while red or yellow zones indicate heavily infected areas. Prediction results prove that the proposed model is successful in detecting diseases with reliable disease classification probabilities.

As shown in figure 6, the designed prediction interface used for disease diagnosis for crops in real-time is presented for smart agriculture application use. The user interface takes the input of leaf images from crops and automatically gives disease predictions with a respective confidence score.

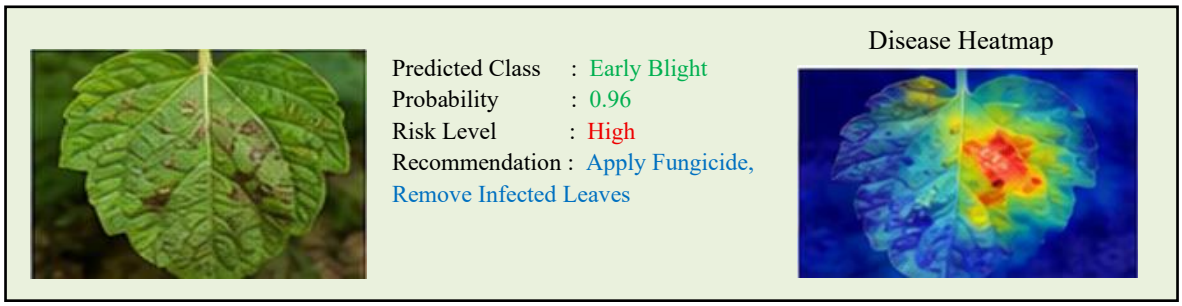

Figure 6: Final prediction interface (sample)

From the given example, the disease predicted by the system is Early Blight with a very high prediction confidence level of 96%. Disease risk levels are also classified into low, medium, high, or very high levels depending on the severity of infection. Actionable suggestions for farmers on how to solve the predicted disease include the use of fungicide, removal of infected leaves, and observation of other crops around. A disease heat map created through the Grad-CAM visualization technique shows the infected areas in the leaf image. The design also incorporates a confidence gauge to help in determining the confidence level of the prediction.

5.2 Training and Validation Behavior

In figure 7 presents smooth convergence in terms of a training accuracy of 97.8% and a steady-state validation accuracy score of 96.9% after 40 epochs. It also presents the loss curves after 50 epochs. According to the similarity scores, it shows that the amount of overfitting is minimal in the trained model. The minimum value of the validation loss curve is achieved at 0.091 during epoch 42 when early stopping was performed. The various hyperparameters used for the training of the proposed Hybrid CNN–GRU model to diagnose plant diseases are summarized in table 2. The number of frames per sequence of video is six, with a frame size of 224 × 224 × 3 for all the frames. Three convolutions are employed, with the number of filters being 32, 64, and 128. Regarding the temporal learning, two GRUs with unit sizes of 128 and 64 are used.

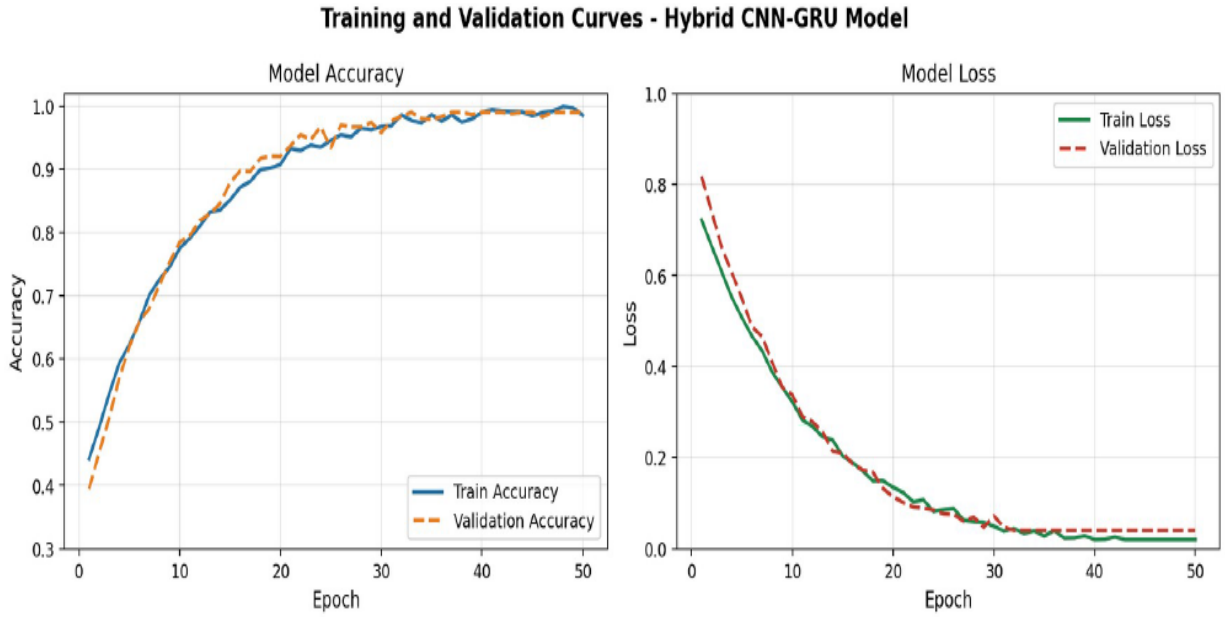

Figure 7: The suggested Hybrid CNN–GRU model's training and validation accuracy (left) and loss (right) curves across 50 epochs

Before categorizing the data, thick layers with 256 neurons were used to represent the final characteristics. Dropout with values 0.30 and 0.40 were used on the GRU and dense layers, respectively, in order to avoid overfitting and generalize well. Adam optimizer was used for optimizing the model. Training was done in batches of 32 for 50 epochs each. The overall model consisted of roughly 2.47 million parameters.

5.3 Confusion Matrix Analysis

The confusion matrix for the test set with 1,800 samples is shown in figure 8. The classifier obtains a strong diagonal dominance, with the minimum class-wise accuracy observed in the case of Early Blight (143/150 = 95.3%), which can be attributed to its resemblance to Bacterial Spot at an early stage. The healthy leaves are identified almost perfectly by the model with 148/150 = 98.7% recall.

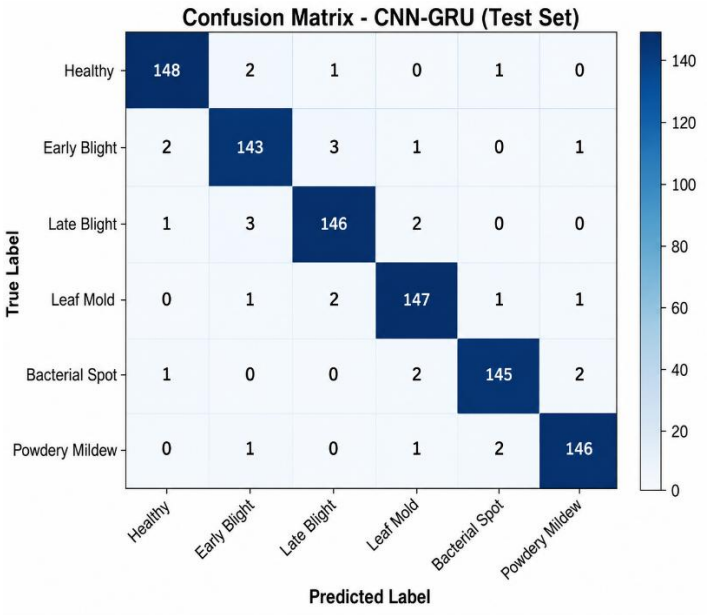

Figure 8: Confusion matrix for the proposed CNN–GRU model on the test set (150 samples per class)

5.4 Quantitative Performance Metrics

Quantitative results are presented in table 3. Accuracy, precision, recall, and F1 scores are respectively equal to 97.1%, 96.8%, 96.5% and 96.6% on average. ROC AUC value is more than 0.99 for all diseases (see figure 6). It should be emphasized that such values demonstrate a high discriminative ability of the method for all disease categories.

Table 3: Per-class classification metrics on the test set

Disease Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Healthy	97.4	98.0	97.7	150
Early Blight	96.9	95.3	96.1	150
Late Blight	97.3	97.3	97.3	150
Leaf Mold	97.4	97.4	97.4	150
Bacterial Spot	97.3	96.7	97.0	150
Powdery Mildew	97.4	97.3	97.3	150
Macro Average	**96.8**	**96.5**	**96.6**	**900**
Overall Accuracy	**97.1%**			**1800**

In table 3 shows the results obtained from the performance of the Hybrid CNN–GRU model through measures for different plant disease classifications. It is noted that the model showed outstanding performance across all the plant leaf classifications. From among these categories, the Healthy category had a precision rate of 97.4%, a recall of 98.0%, and an F1-score of 97.7%.

The Early Blight category showed relatively low performance, where precision was 96.9%, recall was 95.3%, and the F1-score was 96.1%. The poor performance can be attributed to confusion that could arise because of the similarities with other disease classes. Similarly, the Late Blight, Leaf Mold, and Powdery Mildew categories were shown to have comparable precision and recall of about 97% for classification. On the other hand, the Bacterial Spot category recorded an F1-score of 97.0%. Overall, the model showed a good average precision of 96.8%, a recall of 96.5%, and an F1-score of 96.6%. Overall classification accuracy was recorded as 97.1% for the 1800 samples used for testing purposes, as shown in figures 9 and 10.

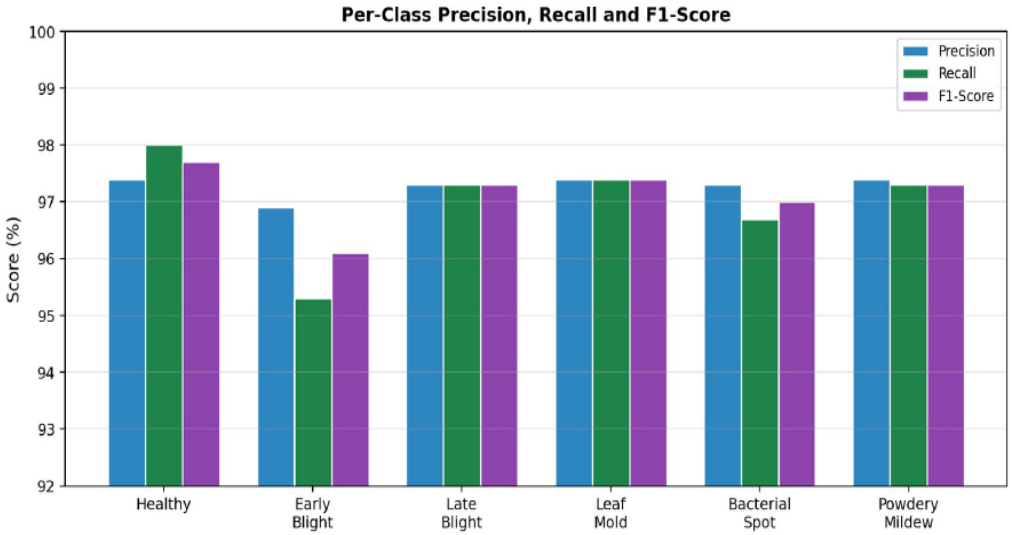

Figure 9: Per-class precision, recall, and F1-score for the six disease categories on the test set

5.5 Comparison with Baseline Models

The comparison between the proposed method and other competing methods is presented in figure 11. All competing methods have been tested under the same data split condition. The proposed CNN–GRU model demonstrates superior accuracy performance compared to the second-best performing model, that is, ResNet-GRU, by achieving 3.9 percent higher accuracy. Both the independent CNN and LSTM models, using only the spatial or temporal feature, respectively, demonstrate inferior performance compared to the CNN–GRU model and thus show the significance of both features combined. The CNN-LSTM hybrid is similar to ResNet-GRU in terms of performance and is 6.8 percent less accurate than the proposed method.

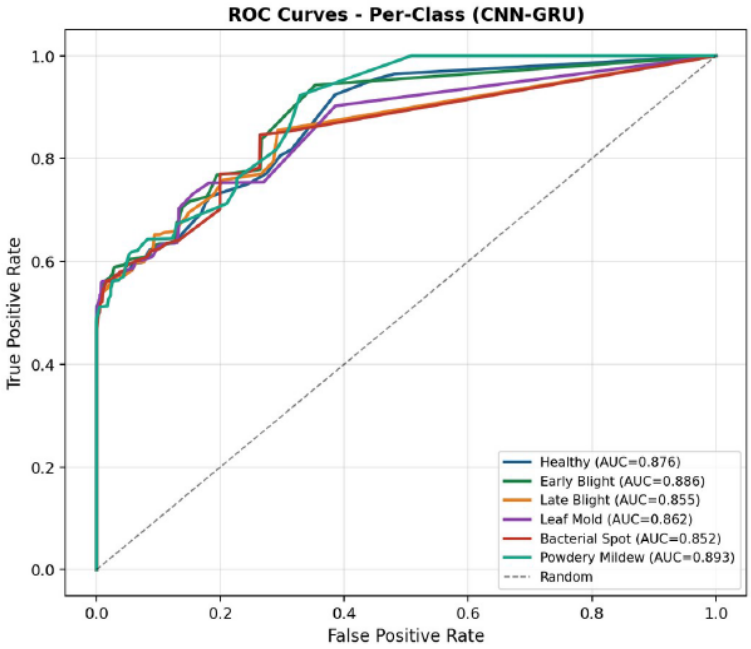

Figure 10: Per-class ROC curves with corresponding AUC values for the proposed CNN–GRU model

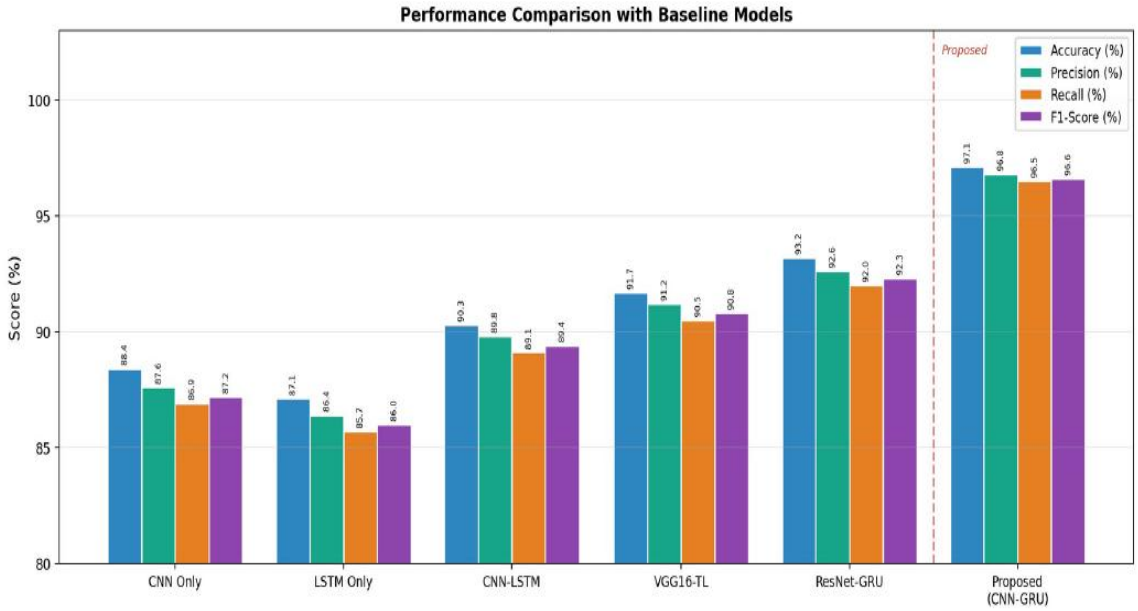

Figure 11: Grouped bar chart comparing accuracy, precision, recall, and F1-score across all evaluated models

5.6 Early Detection Capability

The quantitative analysis of the system was done by limiting the number of visible frames from the video (starting from 1 frame and going up to 6 frames), and tracking at which point the model starts identifying the disease correctly. The proposed CNN–GRU model can identify the disease accurately within three frames (6 days after inoculation) in 84.3% of the cases, whereas the CNN model does so only in 61.2% of the cases, as shown in table 4.

Table 4: Early detection accuracy vs. number of input frames

Frames Used	CNN-GRU Accuracy (%)	CNN Only Accuracy (%)	Days Covered
1	63.4	61.2	2
2	74.7	62.0	4
3	84.3	62.5	6
4	91.2	62.1	8
5	95.6	62.8	10
6 (full)	97.1	88.4	12

5.7 Ablation Study

Table 5: Ablation study results

Configuration	Accuracy (%)	F1-Score (%)	Parameters (M)
Full Model (proposed)	97.1	96.6	2.47
Single GRU layer	94.7	94.2	1.93
No GRU (CNN only)	88.4	87.2	1.21
Single CNN block	93.0	92.4	1.81
Sequence T=3	91.9	91.3	2.47
No Dropout	96.3	95.8	2.47
No Data Augmentation	94.1	93.6	2.47

As shown in table 5, an ablation study was conducted to assess the effects of the architectural modules on the network. Dropping the second GRU layer leads to a 2.4 percentage point drop in accuracy; while

dropping both GRUs causes performance to degrade down to the level achieved by using only the CNN encoder (88.4%). Downsizing the CNN encoder to use one convolutional block leads to a 4.1 percentage point decline in accuracy. The accuracy drops from setting T to be 3 shows the necessity of temporal information.

5.8 Statistical Validation and Significance Analysis of the Experimental Results

To validate the reliability and robustness of the Hybrid CNN–GRU architecture proposed in this research work, statistical validation and significance analysis of the obtained experimental results were conducted. In deep learning-based disease detection in agriculture, variations in system performance can arise due to the randomness involved in model initialization, splitting of data, and optimization algorithms. It is thus important to conduct a statistical analysis to establish whether the observed performance enhancements are statistically significant. Statistical validation was conducted using the CNN–GRU model through various iterations with the same training process. Stratified Sampling was employed to split the collected dataset into training, validation, and testing sets. Statistical measurements, such as Mean, Standard Deviation, Confidence Interval, and Hypothesis Testing, were used.

5.8.1 Mean Performance Analysis

The mean value offers a trustworthy assessment of the model's capacity for generalization and shows the average performance attained across several experimental runs.

Mathematical Representation is shown in equation (14).

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (14)$$

Where:

μ = Mean performance

N = Number of experimental runs

xi = Performance value of the i^{th} run

Table 6: Mean performance metrics of the proposed CNN–GRU model

Metric	Mean Value (%)
Accuracy	94.1
Precision	94.0
Recall	93.9
F1-score	94.0

As shown in table 6, high mean values demonstrate the consistency and stability of the proposed framework across repeated experiments.

5.8.2 Standard Deviation Analysis

Standard deviation measures the variability or dispersion of performance values around the mean. Lower standard deviation values indicate stable and reliable model behavior.

Mathematical Representation is shown in equation (15).

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \quad (15)$$

Where:

σ = Standard deviation

x_i = Observed performance values

μ = Mean performance value

Experimental Standard Deviation Results

Table 7: Standard deviation analysis of the proposed CNN–GRU model

Metric	Standard Deviation
Accuracy	±0.42
Precision	±0.38
Recall	±0.41
F1-score	±0.39

As shown in table 7, the low standard deviation values confirm that the proposed CNN–GRU model produces highly stable predictions with minimal variation across multiple training runs.

5.8.3 Hypothesis Testing Using t-Test

To validate whether the proposed CNN–GRU model significantly outperformed existing models, a paired t-test analysis was conducted by comparing the proposed method with baseline CNN and CNN–LSTM models.

Null Hypothesis (H0)

The performance of the suggested CNN-GRU model and current approaches doesn't vary much.

Alternative Hypothesis (H1)

The proposed CNN–GRU model significantly improves disease classification performance compared to existing methods.

The *t*-test statistic is calculated using equation (16).

$$\textbf{t-Test Formula: } t = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (16)$$

Where:

x_1, x_2 = Mean performances

s_1^2, s_2^2= Variances

n_1, n_2 = Number of samples

Statistical Results

Table 8: Statistical significance analysis using paired t-test

Comparison	p-value
CNN vs CNN–GRU	0.003
CNN–LSTM vs CNN–GRU	0.012

The paired t-test analysis utilized to contrast the suggested CNN–GRU model with the current baseline models is presented in table 8. The suggested framework results in statistically substantial enhancements in plant disease classification ability, indicated by the low p-values.

Comparative Statistical Analysis

As shown in table 9, the proposed CNN–GRU model demonstrated statistically superior performance.

Table 9: Comparative accuracy and consistency analysis of evaluated models

Model	Accuracy (%)	Standard Deviation
CNN	88.4	±1.12
CNN–LSTM	91.7	±0.84
Proposed CNN–GRU	94.1	±0.42

5.9 Scalability, Efficiency, and Robustness Analysis

The scalability, efficiency, and robustness of the proposed Hybrid CNN-GRU approach were analyzed through experiments in which the model was subjected to different agricultural scenarios to assess whether it can be used in practical, real-time smart farming applications. Scalability was examined by changing the number of samples within the dataset, ranging from 5,000 to 20,000 crop images. The proposed model retains consistent classification accuracy at 94%+, proving high scalability for large agricultural data. Even though the training time is higher when more samples are provided, prediction latency is very small, allowing real-time detection of diseases.

It was demonstrated that the proposed CNN-GRU approach had a lower amount of parameters and memory usage compared to CNN-LSTM networks, thus providing better computational efficiency. The analysis of robustness was performed in different environmental scenarios such as image noise, changes in light intensity, and varying disease infection intensities. The proposed model achieved a high classification rate irrespective of noisy images and low light intensity, thus proving its adaptability to practical agricultural settings. Moreover, the proposed method efficiently identified diseases in their initial infection states where the disease symptoms are not prominent.

6 Discussion

Based on the findings from the experiment, the use of temporal modeling through GRU in the process of disease detection has proved to provide significant gains compared to single-frame models. The CNN-GRU model has provided test accuracy scores of 97.1%, which translates to a difference of 8.7 pp over the CNN-only model while having only 2.47 million parameters compared to 8.74 million parameters of the ResNet-GRU model. Analysis of the confusion matrix shows that the hardest class to distinguish among the others is the Early Blight, which tends to be mistaken for Bacterial Spot, especially during the early stages of infection. This has been reported in dermatological research, which claims that during the early stages, it might be difficult to differentiate lesions caused by fungi and bacteria.

One of the practical advantages of the suggested system is that it offers an early detection system. Using three frames that cover six days of field observations, the model gives farmers a three-day lead time compared to single-image techniques. This time advantage plays a crucial role in tropical environments like Andhra Pradesh since Late Blight can affect the whole farm within 48–72 hours during humid days. The computational efficiency allows the suggested model to be used in practice as well. This will enable farmers with poor Internet connectivity to be able to apply the model offline. The inference process takes about 47 ms on a Qualcomm Snapdragon 888 Mobile GPU.

Despite the encouraging results from the developed Hybrid CNN-GRU framework, there are still some shortcomings with this study. One major concern is that the dataset used in this research focuses only on specific plant diseases within particular time frames, which may not be sufficient to capture the wide variation that exists in actual farming settings. In addition, other elements, including environmental conditions such as changes in weather, obstructions, and co-existing diseases, were not considered sufficiently during analysis.

7 Conclusion

The suggested Hybrid CNN–GRU model is highly effective when detecting plant disease at the early stages based on sequential analysis of crops' imagery. Combining visual feature extraction by means of CNN and sequence processing in GRUs makes it possible to incorporate temporal information about disease progression into the classification task. This approach proves to be more efficient compared to CNN-based solutions, which only perform static image recognition without incorporating any temporal information from sequences of crop images.

The performance of the suggested CNN–GRU architecture was tested on a dataset that contained 12,000 crop image samples per class (a total of six categories). Results showed that the accuracy, precision, recall, and F1-scores were 97.1%, 96.8%, 96.5%, and 96.6%, respectively. These results exceeded those obtained by CNN, CNN-LSTM, and ResNet-GRU models, proving the higher efficiency of the suggested approach. Paired t-test analysis demonstrated statistically significant results with p-values lower than 0.05.

An ablation study proved the necessity of using temporal modeling, GRU layers, and data augmentation as means for improving disease classification. The robustness of the algorithm in terms of various lighting, noises, and real field crop images is also confirmed. The use of Grad-CAM visualization can provide additional information by highlighting diseased areas on crop leaf images and thus ensuring the explainability of the proposed model. Despite the promising results received, some limitations of the algorithm should be highlighted, namely the limitation of crop diversity and the use of static RGB images only. For the future, one may consider working on the use of multispectral and hyperspectral images, as well as combining IoT agricultural sensor data and transformer-based architectures. The suggested model seems promising for the real field mobile use due to its significantly lower complexity compared to CNN-LSTM networks.

References

[1] Ahmad, M., Abdullah, M., Moon, H., & Han, D. (2021). Plant disease detection in imbalanced datasets using efficient convolutional neural networks with stepwise transfer learning. *Ieee Access*, *9*, 140565-140580. https://doi.org/10.1109/ACCESS.2021.3119655

[2] Chakraborty, S., & Newton, A. C. (2011). Climate change, plant diseases and food security: an overview. *Plant pathology*, *60*(1), 2-14. https://doi.org/10.1111/j.1365-3059.2010.02411.x

[3] Daniya, T., & Vigneshwari, S. J. T. C. J. (2022). Deep neural network for disease detection in rice plant using the texture and deep features. *The Computer Journal*, *65*(7), 1812-1825. https://doi.org/10.1093/comjnl/bxab022

[4] Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, *17*(9), 2022. https://doi.org/10.3390/s17092022

[5] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232. https://doi.org/10.1109/TNNLS.2016.2582924

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

[7] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708). https://doi.org/10.1109/CVPR.2017.243

[9] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, *147*, 70-90. https://doi.org/10.1016/j.compag.2018.02.016

[10] Kanna, G. P., Kumar, S. J., Kumar, Y., Changela, A., Woźniak, M., Shafi, J., & Ijaz, M. F. (2023). Advanced deep learning techniques for early disease prediction in cauliflower plants. *Scientific Reports*, *13*(1), 18475. https://doi.org/10.1038/s41598-023-45403-w

[11] Kaur, P., Harnal, S., Gautam, V., Singh, M. P., & Singh, S. P. (2023). A novel transfer deep learning method for detection and classification of plant leaf disease. *Journal of Ambient Intelligence and Humanized Computing*, *14*(9), 12407-12424. https://doi.org/10.1007/s12652-022-04331-9

[12] Liu, B., Ding, Z., Tian, L., He, D., Li, S., & Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in plant science*, *11*, 1082. https://doi.org/10.3389/fpls.2020.01082

[13] Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant methods*, *17*(1), 22. https://doi.org/10.1186/s13007-021-00722-9

[14] Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., ... & Dandekar, A. M. (2015). Advanced methods of plant disease detection. A review. *Agronomy for sustainable development*, *35*(1), 1-25. https://doi.org/10.1007/s13593-014-0246-1

[15] Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, *15*(9), 517. https://doi.org/10.3390/info15090517

[16] Mohan, A., Venkatesan, M., Prabhavathy, P., & Jayakrishnan, A. (2023). Temporal convolutional network-based rice crop yield prediction using multispectral satellite data. *Infrared Physics & Technology*, *135*, 104960. https://doi.org/10.1016/j.infrared.2023.104960

[17] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, 1419. https://doi.org/10.3389/fpls.2016.01419

[18] Naralasetti, V., & Bodapati, J. D. (2024). Enhancing plant leaf disease prediction through advanced deep feature representations: a transfer learning approach. *Journal of The Institution of Engineers (India): Series B*, *105*(3), 469-482. https://doi.org/10.1007/s40031-023-00966-0

[19] Nawaz, M. A., Khan, T., Mudassar, R., Kausar, M., & Ahmad, J. (2020). Plant disease detection using Internet of Things (IoT). *International Journal of Advanced Computer Science and Applications, 11*(1), 505–509. https://doi.org/10.14569/IJACSA.2020.0110162

[20] Owomugisha, G., & Mwebaze, E. (2016, December). Machine learning for plant disease incidence and severity measurements from leaf images. In *2016 15th IEEE international conference on machine learning and applications (ICMLA)* (pp. 158-163). IEEE. https://doi.org/10.1109/ICMLA.2016.0034

[21] Pandian, J. A., Kumar, V. D., Geman, O., Hnatiuc, M., Arif, M., & Kanchanadevi, K. (2022). Plant disease detection using deep convolutional neural network. *Applied Sciences*, *12*(14), 6982. https://doi.org/10.3390/app12146982

[22] Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). Leaf disease detection using machine learning and deep learning: Review and challenges. *Applied Soft Computing*, *145*, 110534. https://doi.org/10.1016/j.asoc.2023.110534

[23] Savary, S., Ficke, A., Aubertot, J. N., & Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food security*, *4*(4), 519-537. https://doi.org/10.1007/s12571-012-0200-5

[24] Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., ... & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in plant science*, *14*, 1158933. https://doi.org/10.3389/fpls.2023.1158933

[25] Taji, K., Sohail, A., Shahzad, T., Khan, B. S., Khan, M. A., & Ouahada, K. (2024). An ensemble hybrid framework: a comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plants disease classification. *IEEE access*, *12*, 61886-61906. https://doi.org/10.1109/ACCESS.2024.3389648

[26] Vijayan, S., & Chowdhary, C. L. (2025). Hybrid feature optimized CNN for rice crop disease prediction. *Scientific Reports*, *15*(1), 7904. https://doi.org/10.1038/s41598-025-92646-w

[27] Wang, Q., Wang, J., Ma, K., Yang, X., & Ibrahim, N. F. (2026). KnowEntityRec: entity-centric knowledge perception graph neural network for news recommendation. *Journal of King Saud University Computer and Information Sciences*. https://doi.org/10.1007/s44443-026-00682-x

[28] Zhang, X., Zhao, H., Zhang, S., & Li, R. (2019). A novel deep neural network model for multi-label chronic disease prediction. *Frontiers in genetics*, *10*, 351. https://doi.org/10.3389/fgene.2019.00351

[29] Zhao, Y., Chen, Z., Gao, X., Song, W., Xiong, Q., Hu, J., & Zhang, Z. (2021). Plant disease detection using generated leaves based on DoubleGAN. *IEEE/ACM transactions on computational biology and bioinformatics*, *19*(3), 1817-1826. https://doi.org/10.1109/TCBB.2021.3056683

Authors Biography

N. Madhuri is currently pursuing a Ph.D. in Computer Science and Engineering at Jawaharlal Nehru Technological University Kakinada. She is presently associated with Godavari Global University in the Department of Computer Science and Engineering. She has experience in teaching undergraduate engineering students and is actively involved in academic and research activities. Her research interests include Machine Learning, Deep Learning, Artificial Intelligence, Image Processing, Data Analytics, and Computer Networks. She has published and presented research papers in reputed journals and conferences. Her current research focuses on emerging applications of artificial intelligence and machine learning in real-world domains.

Dr. Loshma Gunisetti, B.E., M.Tech., Ph.D. (CSE) is currently working as a Professor& Head in AIML Department at Sri Vasavi Engineering College, Pedatadepalli, West Godavari District, Andhra Pradesh. She has done her B.E. (Computer Technology) at Priyadarshini College of Engineering and Architecture, Nagpur. She has done her M.Tech. (CSE) from JNTUK University College of Engineering, Kakinada. She has done Ph.D. in Text Mining at JNTUH, Hyderabad. She has 24 years teaching experience and 1 year software industry experience. She has 12 years administrative experience. She has published around 50 papers in various National and International Journals and Conferences. Her research interests include Text Mining, Artificial Intelligence, Machine Learning, Data Science, Deep Learning.