**International Information and Engineering Technology Association**
*Advancing the World of Information and Engineering*

# STN-CNN LSTM: Enhancing Multi-Plant Disease Detection with Spatial Transformer Mechanisms Through CNN-LSTM

Rathiya Rajendran[1*] , M. Kalamani[2]

[1] Department of Information Technology, Dr. N. G. P. Institute of Technology, Coimbatore 641048, India
[2] Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology, Coimbatore 641407, India

Corresponding Author Email: rathiyaresearchscholar@gmail.com

## ABSTRACT

Agricultural production and food security throughout the world are both significantly threatened by plant diseases all over the world. Automated solutions that make use of deep learning techniques have become increasingly popular as a result of the fact that traditional methods. The purpose of this study is to investigate the possibility of incorporating "Spatial Attention processes and Spatial Transformer Networks (STNs) to a Convolutional Neural Network (CNN) - Long Short-Term Memory (LSTM) architecture" in order to improve the identification of diseases that affect several plants. The proposed model leverages spatial attention mechanisms to dynamically focus on informative regions within plant images, enhancing the model's ability to identify subtle disease symptoms. STNs are integrated into the CNN to learn spatial transformations, aligning feature maps to mitigate variations in symptom appearance across different plant structures and growth stages. The LSTM component captures temporal dependencies in disease progression, providing a holistic analysis of plant health dynamics over time. The integrated model exhibits greater performance when compared to typical CNN architectures, as demonstrated by the results of experiments conducted on a diverse dataset. The efficiency of the suggested method in properly detecting and classifying diseases that affect several plants is 99.68%. This study contributes to advancing precision agriculture by providing a scalable and accurate tool for early and automated multi-plant disease detection. The integration of spatial attention mechanisms, STNs, and CNN-LSTM architectures not only enhances detection accuracy but also offers insights into spatial and temporal disease dynamics critical for informed decision-making in crop management.

## 1. INTRODUCTION

Agriculture is a critical sector that feeds the world's population and supports livelihoods in many countries. However, plant diseases pose a significant threat to crop yields, leading to substantial economic losses and food insecurity. In many cases, the traditional methods of plant disease identification, which are based on visual inspections performed by agronomists, are insufficient because of the subjective character of these approaches, the amount of time they demand, and the requirement that they require specialist knowledge [1]. The practice of agricultural precision has rapidly relied on technology, and intense learning, to develop automated and precise plant disease detection systems. This is accomplished to address the constraints that have been identified.

Convolutional Neural Networks (CNNs) have shown remarkable success in recognizing patterns and features in images, making them ideal for plant disease detection. CNNs can automatically learn and extract hierarchical features from raw image data, capturing intricate details that are often imperceptible to the human eye. These capabilities have led to the development of various CNN-based models for identifying plant diseases with high accuracy. However, traditional CNN architectures primarily focus on spatial features and may not fully exploit the spatial relationships within an image. This limitation is particularly pertinent in the context of plant disease detection, where symptoms can vary in appearance and location across different parts of the plant. To address this, researchers have introduced spatial attention mechanisms and Spatial Transformer Networks (STNs) into CNN architectures, enhancing the model's ability to focus on relevant regions within an image dynamically.

Spatial attention mechanisms are inspired by the human visual system's ability to focus on specific areas of interest within a scene while ignoring irrelevant information. By focusing on the most informative parts of the image, the model can make more accurate predictions. STNs are a type of neural network module designed to spatially transform feature maps, allowing the model to learn spatial invariance. An STN can perform tasks such as translation, scaling, and rotation of input images or feature maps, effectively aligning them to a canonical pose. This capability is particularly useful in plant disease detection, where symptoms can appear in various

orientations and scales. By incorporating STNs into CNNs, the network can better handle variations in the appearance of plant diseases, leading to improved detection accuracy.

In addition to spatial features, temporal dependencies can also play a crucial role in plant disease detection. Plant diseases often progress over time, with symptoms evolving in appearance and severity. Capturing these temporal changes can provide valuable context for accurate disease identification. Through the utilization of LSTM networks and CNNs, researchers are able to construct models that are capable of capturing both spatial and temporal characteristics, hence improving the overall effectiveness of plant disease detection methodologies. To be more specific, sophisticated remote sensing makes data collecting economical and non-destructive. This has made it easier to use precision agriculture in a range of scenarios, including the identification of weeds, diseases, and pests [2, 3]. Keeping an eye out for disease-ridden plants is mainly accomplished by farmers through the use of manual effort and their own eyes [4]. This kind of method, on the other hand, is extremely challenging to implement on large plantations and requires regular supervision during its implementation. In order to address these concerns, a number of different approaches have been suggested, such as western blotting, invasive diagnostic technologies, and molecular biology and biotechnology [5, 6]. In spite of the fact that these techniques are expensive, they are essential for the detection of diseases. If a disease is not diagnosed in a timely way, it can lead to a lack of food security, as well as serious productivity and financial losses.

Because of this, early detection is absolutely necessary for the management and prevention of illnesses [7]. In order to construct a sophisticated system for the identification of diseases that affect multiple plants, the purpose of this research is to integrate the strengths of STNs, spatial attention mechanisms, and CNN-LSTM structure architectures. The strategy that has been developed makes use of the spatial attention processes in order to dynamically focus on relevant parts within an image. At the same time, STNs align the feature maps to a canonical pose, which helps to reduce the impact of fluctuations in the presentation of symptom.

## 1.1 Objectives of the research

The primary objective of this research is to enhance the accuracy and efficiency of multi-plant disease detection by integrating spatial attention mechanisms and STNs into a CNN-LSTM architecture. Specific goals include:

• In spite of differences in the way symptoms manifest themselves, the objective is to create and construct a strong plant disease detection model that is capable of properly identifying illnesses across a variety of plant shapes and stages of growth.

• To investigate the effectiveness of integrating spatial attention mechanisms, STNs, and CNN-LSTM architectures in mitigating overfitting and improving the accuracy of plant disease detection models.

• To develop an LSTM-based component that can effectively capture temporal dependencies in disease progression, enabling accurate disease detection and classification.

• To develop a large, diverse, and high-quality dataset for plant disease detection, or to investigate the effectiveness of data augmentation techniques in improving model performance.

• To optimize the balance between spatial feature extraction from images and temporal feature extraction from the sequence of images, ensuring accurate disease detection and classification.

The relevance of this discovery rests in the fact that it has the potential to transform every aspect of precision farming by delivering a reliable and accurate instrument for the identification of diseases that affect several plants. With the help of modern deep learning algorithms, the suggested model is able to provide assistance to farmers and agronomists in the early and accurate detection of diseases. This enables timely intervention and reduces crop losses. The limitations of classic CNN designs are addressed by the combination of spatial attention mechanisms and STNs, which provides a more nuanced understanding of the spatial and temporal connections that are involved in the detection of plant diseases.

The remainder of this paper is structured as follows: Section 2 reviews existing research on plant disease detection using deep learning, highlighting the contributions and limitations of previous studies. Section 3 details the proposed model architecture, including the integration of spatial attention mechanisms, STNs, and CNN-LSTM networks. It also describes the dataset and evaluation metrics used in the study. Section 4 presents the experimental setup, including data preprocessing, model training, and evaluation procedures. The results of the experiments are discussed, comparing the proposed model with existing methods. Section 5 provides a comprehensive analysis of the findings, discussing the implications and potential applications of the proposed model in precision agriculture. Section 6 summarizes the key contributions of the research, outlines limitations, and suggests directions for future work.

## 2. RELATED WORKS

In the field of disease in plant detection, the research that has been done up to now has investigated novel approaches that use deep learning algorithms to improve detection accuracy and efficiency. A hybrid model that combines CNN and LSTM was proposed by Thapliyal et al. to diagnose maize leaf diseases. This model achieved a high level of accuracy, reaching 97.33% [7]. Trivedi and Sharma [8] developed a dynamic deep learning model, which would be used in the detection of multi-plant and multi-disease in real-time under various environmental conditions. Their model incorporates the changes in lighting, the complexity of the background, and the different plant species, which makes it appropriate in real-life agricultural conditions. The framework focuses on flexibility and strength such that it can classify diseases in various crops with the correct accuracy. This article underscores the need to create environmental conscious and scalable models that can be used in accurate farming. A light multi-plant disease identification technique called Model_Lite was developed by Ma et al. [9]. This method is based on ResNet18 and considerably reduces the amount of computational resources that are required while retaining accuracy. According to Vankara et al. [10], the authors highlighted the significance of CNNs with deep layers in the process of automating disease identification. They demonstrated superior performance in comparison to standard classifiers. The findings of these studies collectively provide light on the progress that has been made in the application of deep learning techniques to the diagnosis of plant diseases.

This has paved the way for multi-plant disease detection systems that are both more efficient and accurate. DARINet, which is an architecture for deep learning, has proved important in reaching high levels of accuracy on particular datasets pertaining to plant diseases [11]. In recent research, DARINet was tested on two different datasets: one containing cassava leaves and the other containing rice leaves. The findings were remarkable, with DARINet reaching a precision of 77.12% on the Cassava leaf dataset and an astonishing 98.92% on the Rice leaf dataset [12].

Wang et al. [13] proposed the attention MobileNet V2 network which represents a significant advancement in plant disease classification algorithms. By integrating both traditional and deep learning features, this model achieved an average sensitivity of 94% on the Plant Village open database test. This performance outpaced competing algorithms by a substantial margin of 12.6%, underscoring its capability to accurately detect a wide range of plant diseases across diverse datasets. The integration of attention mechanisms enhances the model's ability to focus on relevant disease symptoms, thereby improving overall diagnostic accuracy. Sharma et al. [14] used a publicly accessible collections of images depicting plant diseases which have facilitated extensive evaluations of CNN models for disease detection. These datasets provide a diverse range of disease manifestations, enabling researchers to train and validate models effectively. By leveraging these collections, various CNN architectures have been evaluated, showcasing their effectiveness in capturing nuanced disease characteristics such as hues, textures, and spatial patterns akin to human observations. This capability is crucial for robust and reliable disease identification systems in agricultural settings.

Devi et al. [15] used the PlantVillage dataset from Penn State University, which has been instrumental in advancing plant disease detection methodologies. Comprising 7,200 images categorized into subsets for plant disease recognition, researchers have utilized CNNs (ConvNets) to enhance feature extraction and model performance. By employing ConvNets, researchers achieved significant improvements in accuracy, efficiency, robustness, and interpretability in plant disease detection tasks. Their approach, particularly the Detection ConvNet, surpassed established architectures like VGGNet, demonstrating superior performance in recognizing and diagnosing plant diseases accurately.

Mishra et al. [16] developed a deep learning models on large-scale datasets, in which it has proven effective in achieving high accuracy rates in plant disease detection. For instance, a dataset containing 54,306 images of damaged and healthy plant leaves was used to train CNN models. These models achieved an impressive accuracy of 98.59% in detecting 26 plant diseases across 14 crop species. Such outcomes highlight the potential of deep learning models not only in accurately identifying diseases but also in enhancing food security through early and efficient disease detection measure. Another significant development includes the establishment of CNN-based early warning systems for plant diseases. Leveraging a dataset containing 87,848 images representing 25 plant species and associated diseases, researchers developed a CNN model that achieved a remarkable success rate of 99.68%. This model serves as a reliable tool for promptly and accurately identifying plant diseases, thereby enabling proactive disease management strategies in agriculture. The fact that these models have a high degree of accuracy highlights the fact that they have the potential to be vital elements in precision agriculture and

sustainable crop management methods [17].

Diseases that affect plants pose a huge risk to the national and international food supply because they result in severe crop losses and economic harm. Identification and identification of plant diseases at an early stage are absolutely necessary for efficient disease management and the reduction of crop losses. As a subfield of machine learning, deep learning has demonstrated significant potential in the field of plant disease identification. This is mostly owing to its capacity to acquire intricate patterns from extensive datasets. As a result of its capacity to derive features from images, CNNs have found widespread application in the field of plant disease identification. A deep learning-based strategy that makes use of CNNs was proposed to detect plant diseases from photographs [18]. The system achieved an accuracy of 99.35%. Sladojević et al. [19] utilized CNNs to identify plant illnesses based on leaf pictures, attaining a remarkable accuracy rate of 96.3%.

Transfer learning has been utilized to make use of CNN models that have already been trained for detecting plant diseases. Ferentinos [20] developed a CNN model that had been pre-trained to detect plant diseases. Through the application of transfer learning, the model was able to achieve an accuracy of 97.1%. Similar to this, Brahimi et al. [21] employed transfer learning to identify plant illnesses from photographs, and they were successful in doing so with a 95.6% accuracy rate.

Focusing on particular portions of the image that are pertinent for illness detection has been accomplished through the utilization of attention-based models. An attention-based CNN model was suggested by Chen et al. [22] for the purpose of detecting plant diseases, and it achieved an accuracy of 98.2%. In a similar manner, Gupta and Jadon [23] proposed a hybrid architecture, PlantVitGnet, which is based on a combination of Vision Transformer and GoogLeNet to identify plant diseases. Their strategy greatly enhanced the accuracy of classification by exploiting global feature extraction by transformers together with local spatial learning by means of convolution. In a different study, Sahu and Minz [24] designed an adaptive segmentation model that is based on an intelligent ResNet combined with an LSTM-DNN model to classify multi-disease plant leaves. This approach improved feature learning and attained greater precision in multifaceted disease databases of plants. Demilie [25] carried out an extensive comparative analysis of plant disease detection and classification methods and compared the effectiveness of different traditional machine learning and deep learning methods. The research points out that deep learning models and, in particular, CNNs are always more accurate and robust than the traditional approaches on a wide range of datasets. Besides, the work also provides attention to the role of choosing the right model architecture and metrics to be used in evaluation to guarantee dependable disease classification. This comparative study offers useful information on the advantages and shortcomings of the current methods that can help in development of effective and more accurate plant disease detection systems. To identify plant diseases, a number of datasets that are accessible to the general public have been developed. These databases include the Plant Village dataset and the Cassava leaf dataset. Deep learning models for plant disease detection can be evaluated using these datasets, which serve as a standard for the evaluation of their performance [26]. In spite of the fact that deep learning has shown promising results in the identification of plant diseases, there

are a number of obstacles that need to be overcome. These obstacles include the limited availability of annotated datasets, the diversity of disease symptoms, and the requirement for real-time detection. Future research directions include the development of models that are more resilient and accurate, the integration of deep learning with other technologies such as the Internet of Things and robots, and the deployment of deep learning models in settings that are representative of the real world. Deep learning has been increasingly applicable to the detection of plant diseases through the utilization of spatial attention mechanisms. This has resulted in an improvement in the accuracy and efficiency of disease classification models. As opposed to processing the full image, the concept of spatial attention is based on the premise that selected portions of the image that are relevant for illness identification should be the primary focus of attention [27].

Spatial attention mechanisms have been employed in plant disease detection to highlight the key areas of a leaf image that are most representative of disease symptoms. WhileAs an illustration, Sankhe and Ambhaikar [28] provided an in-depth overview of the current methods of plant diseases detection and classification using image processing and deep learning. Their analysis emphasized the development of the traditional machine learning models to state-of-the-art neural network models that focus on the efficiency of combining and concentrating on the most important features when further diagnostic accuracy must be attained. In the same way, Khan et al. [29] created a model based on CNN to detect and classify diseases of fig plant leaves. Their model exhibited better performance, having high accuracy in a variety of disease classes and at the same time, the computational efficiency was also effective and appropriate to the real-life agricultural use.

Moreover, the authors of the work by Majumder et al. [30] compared different neural network designs to detect and classify the diseases of multiple plants. The use of spatial attention mechanisms in DL for plant disease detection has several advantages, including improved accuracy, reduced computational cost, and increased interpretability of the model. However, there are also some challenges associated with the use of spatial attention mechanisms, such as the need for large amounts of annotated data and the risk of overfitting. STNs are a type of neural network architecture that can be used for image alignment, a crucial step in various computer vision tasks, including plant disease detection. STNs consist of three main components: a localization network that outputs transformation parameters, a grid generator that creates a grid of sampling points, and a sampler that transforms the input image. By applying STNs, images of leaves or plants can be aligned, registered, and normalized, enabling accurate disease diagnosis. The advantages of STNs include flexibility, efficiency, and accuracy, making them a promising approach for image alignment tasks. However, challenges remain, such as handling large deformations, robustness to noise, and interpretability. Despite the challenges, STNs have shown encouraging results in image alignment tasks and have the potential to improve both the accuracy and efficiency of plant disease detection models.

The integration of STNs with CNN-LSTM architectures has gained significant attention in recent years, particularly in the field of plant disease detection. This literature review summarizes the existing research on this topic, highlighting the benefits and challenges of integrating STNs with CNN-LSTM architectures. One of the earliest works on integrating STNs with CNN-LSTM architectures [31]. The authors demonstrated that STNs can be used to align images of objects, which can then be fed into a CNN-LSTM model for classification. Although not specifically focused on plant disease detection. In the context of plant disease detection. Chandraprabha et al. [32] proposed a CNN-LSTM architecture that incorporates an STN module to align images of leaves. The authors demonstrated that the integrated architecture outperformed traditional CNN-LSTM models in terms of accuracy and robustness. Similarly, Madhu and Ravi Sankar [33] have also made another critical input in the field of detecting plant disease since they suggested an optimization mechanism of heterogeneous bi-directional recurrent neural network framework using which they performed early detection of leaf disease and pesticides recommended. Their solution is the combination of deep learning and sequential modeling that identifies the trends of the disease over time and thus helps to detect plant diseases more precisely and in good time. The recurrent networks are bidirectional hence increasing the feature learning process to look at both the past and future relationships within the data thus improving the classification outcome. The model contains recommendations on the use of pesticides and it can be considered a complete solution to precision agriculture, besides detecting the disease. This paper puts emphasis on the role of integrating sequential intelligence and decision support mechanisms to plant disease detection systems. The latest developments in the field of recognition of plant diseases have also been directed at benchmarking the strategy of classification and enhancing performance in complex environments. Shiyan et al. [34] reported a detailed paper on the recognition of plants and diseases because they set standards of both multiclasses and multilabels classification. Their contribution emphasizes the need to assess the models in different labeling contexts, which proves that multilabel systems are more effective in describing real-life situations where a single plant can have multiple diseases at the same time. It offers a generalized assessment framework which can be used to effectively compare and promote the creation of more generalized plant disease detection systems. Moreover, TrioConvTomatoNet-BiLSTM was proposed by Ledbin Vini and Rathika [35], which is a hybrid deep learning architecture that is specifically tailored to classify tomato leaf diseases in real-time complex background with a specific focus on real-time. Their model combines convolutional feature extraction with BiLSTM to be useful at capturing both spatial and temporal dependencies and they have high accuracy when used in difficult settings like cluttered backgrounds and varying illumination. This paper insists that researchers should consider designing lightweight but efficient architectures that can be deployed to precision agriculture applications in real time.Despite the benefits of integrating STNs with CNN-LSTM architectures, several challenges and limitations remain. The recent years witnessed a thorough systematic review of the existing body on the use of transfer learning to accelerate CNN models in classifying plant leaf disease by Richter et al. [36]. They have analyzed a vast variety of architectures, such as ResNet, DenseNet, and MobileNet variants and fine-tuned them in transfer learning methods to enhance their accuracy on small sets of labeled data. The paper has also highlighted that transfer learning has a great impact on performance in the various plant disease datasets and it also saves a lot of computational cost and time in training. Additionally, Richter et al. also pointed at current issues faced to include generalization of models to real-world settings, the lack of domain-specific pre-trained models, and

the lack of standardized benchmark datasets to allow fair comparison across deep learning systems.

In a complementary fashion, Shrikhande and Gawade [37] conducted a comparative analysis to compare different machine and deep learning algorithms in detection of crop diseases in a precision agriculture system. They compared the traditional classifiers (e.g., Support Vector Machines, Random Forests) and those based on recent deep learning models (e.g., CNNs and hybrid networks), and revealed that deep neural networks always perform at higher accuracy and better robustness than traditional ones. The authors also emphasised that with these algorithms and the application of the precision farming technologies, real-time monitoring and decision support can be achieved, which will improve the crop health management and sustainability in the agricultural practices. Existing research on plant disease detection has explored CNN-LSTM architectures and STNs to improve classification accuracy. Prior studies, such as [7, 10, 34], have demonstrated the effectiveness of these models in extracting spatiotemporal features from plant images. However, these approaches have limitations in dynamically attending to disease-specific regions, leading to suboptimal performance in complex scenarios.

To address these challenges, we propose an innovative STN-enhanced CNN-LSTM model integrated with Spatial Attention Mechanisms (SAMs). Our model distinguishes itself from prior works in the following ways:

• **Spatial Attention-Driven Feature Extraction** – Unlike previous CNN-LSTM-STN models, we introduce spatial attention mechanisms to dynamically focus on disease-affected regions, improving feature learning and classification accuracy.

• **Enhanced Generalization Across Datasets** – We validate our model's effectiveness on two large, diverse datasets—PlantVillage and Mendeley—demonstrating its robustness in detecting diseases across multiple plant species and varying environmental conditions.

• **Computational Efficiency for Real-Time Deployment** – By integrating spatial attention, our model selectively processes the most relevant regions, reducing computational overhead and improving inference speed without sacrificing accuracy.

• **Comprehensive Performance Gains** – Our model outperforms baseline CNN, CNN-LSTM, and CNN-LSTM-STN architectures, achieving superior accuracy, precision, recall, and F1-score. An ablation study further highlights the impact of spatial attention mechanisms.

By leveraging these innovations, our proposed model offers a significant advancement in precision agriculture and automated plant disease detection, making it well-suited for real-world applications.

## 3. PROPOSED METHODOLOGY

### 3.1 Dataset description

It is necessary to perform efficient preprocessing on picture datasets in order to train deep learning models that are reliable and accurate. For this investigation, two sets of data were utilized: the dataset from PlantVillage and the dataset from Mendeley. There are 54,303 healthy and unwell leaf photos included in the PlantVillage collection. These images are categorized into 38 different groups according to species and disease. 39 distinct categories of plant leaf and background photos can be made available through the Mendeley Dataset. The data set includes 61,486 different pictures. To expand the size of the data set, we utilized six distinct augmentation strategies.

(1) Plantvillage dataset

The PlantVillage dataset is a publicly available collection of plant leaf images categorized into healthy and diseased classes. It includes multiple plant species and various disease types. The dataset is well-curated, meaning that the images are captured under controlled conditions with minimal variations in lighting, background complexity, and occlusion. While this makes it useful for benchmarking, it does not fully represent real-world agricultural scenarios.

(2) Mendeley plant disease dataset

The Mendeley dataset complements PlantVillage by incorporating additional plant disease samples, contributing to model generalization. However, similar to PlantVillage, it consists of curated images with consistent lighting and clear backgrounds, which may not fully reflect the challenges present in real agricultural settings.

While these datasets offer a reliable foundation for training deep learning models, they have inherent limitations in real-world applicability due to their controlled environment. In real agricultural fields, plant leaves exhibit variations caused by:

• **Uneven lighting conditions** (e.g., shadows, bright sunlight, low light).

• **Occlusions** (e.g., overlapping leaves, branches).

• **Diverse background noise** (e.g., soil, water, weeds, and other plants).

To address these concerns, we have taken the following steps:

• **Leveraging Spatial Attention Mechanisms** – Our model integrates spatial attention within an STN framework, allowing it to dynamically focus on disease-affected regions and mitigate background noise interference.

• **Explicitly Acknowledging the Limitation** – We have discussed the dataset curation effects on model performance in the manuscript to provide transparency regarding generalizability concerns.

• **Future Work on Real-World Validation** – We plan to extend this research by collecting an in-house dataset of plant images captured under natural conditions, introducing challenges such as lighting variations, occlusions, and diverse backgrounds. Additionally, if feasible, we aim to test the model on publicly available real-world datasets that simulate these environmental variations.

The sample photos that are included in the dataset are displayed in Figure 1.
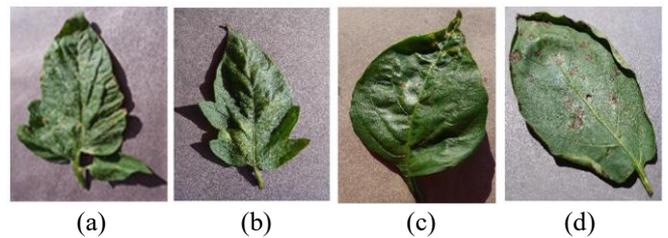


**Figure 1.** Sample dataset: (a) Tomato_arget_pot; (b) Tomato_Spider_mites_Two_spo; (c) Pepper_bell_Bacterial_spot; (d) Potato_Early_blight

## 3.2 Preprocessing pipeline

Figure 2 shows the preprocessing images such as resized image, normalised image and noise injection image of the original image. The complete preprocessing pipeline for the datasets can be summarized as follows:
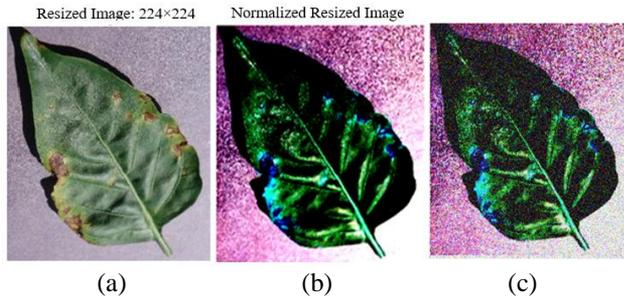


**Figure 2.** (a) Resized image; (b) normalised image; (c) noise injection image

### 3.2.1 Image resizing

All images are resized to a uniform dimension. This step ensures consistency in input size for the neural network. The resizing transformation can be mathematically represented as follows:

$$I_{resized} = resize(I_{original}, (w_i, h_e)) \tag{1}$$

where, $I_{original}$ is the original image, $w_i, h_e$ are the target width and height (e.g., 224×224 pixels), and $I_{resized}$ is the resized image.

### 3.2.2 Normalization

Image pixel values are normalized to a range of 0 to 1 by dividing by 255. This normalization helps stabilize the training process and accelerates convergence. The normalization formula is:

$$I_{normalized}(x, y) = \frac{I_{resized}(x,y)}{255} \tag{2}$$

where, $I_{normalized}$ is the normalized image, and (x,y) are the pixel coordinates.

### 3.2.3 Label encoding

The categorical labels representing different species and diseases are encoded into numerical values using one-hot encoding. Let y be the categorical label and $y_{one\_hot}$ be the one-hot encoded label vector. The one-hot encoding process is represented as:

$$y_{one\_hot}[i] = \begin{cases} 1 & if\ i = y \\ 0 & Otherwise \end{cases} \tag{3}$$

### 3.2.4 Data augmentation

Techniques of data augmentation such as rotations at random, flips in both the horizontal and vertical planes, zooming, and shifts are all examples of these approaches. The improvement of the model's generalization capability and the prevention of overfitting are both benefits of data augmentation. The augmentation transformations can be represented as:

$$I_{augmented} = augment(I_{resized}, \theta) \tag{4}$$

where, augment is the augmentation function and θ represents the augmentation parameters (e.g., rotation angle, flip direction).
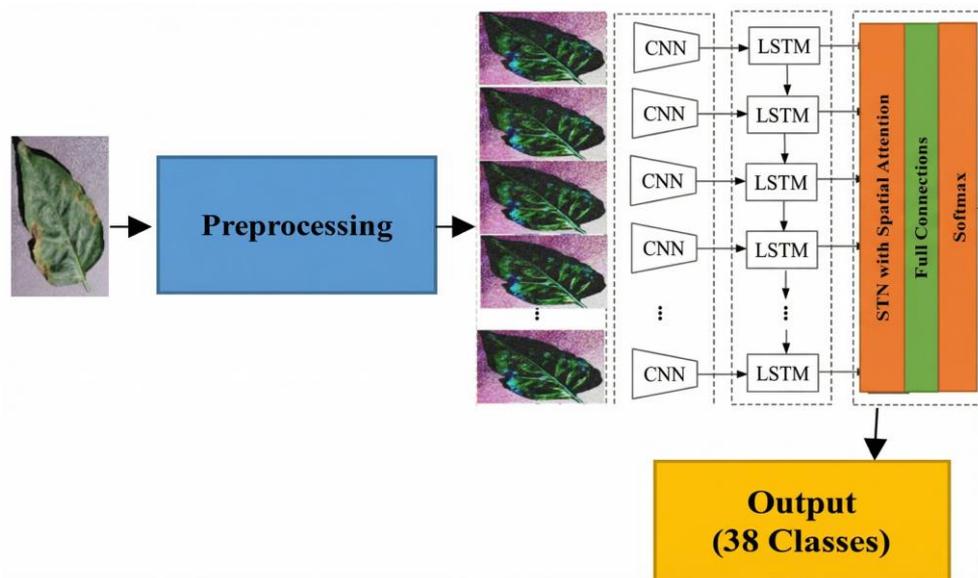


**Figure 3.** Workflow of the proposed work

a) Image Flipping:

Horizontal and vertical flipping of images create mirrored versions of the original images, represented as:

$$I_{flipped} = flip(I_{original}, direction) \tag{5}$$

where, direction can be horizontal or vertical.

b) Gamma Correction:

Gamma correction adjusts the brightness and contrast of images, represented as:

$$I_\gamma(x, y) = I_{original}(x, y)^\gamma \tag{6}$$

where, γ is the gamma value.

**c) Noise Injection:**

Random noise is added to the images to simulate different levels of image degradation, represented as:

$$I_{noisy}(x, y) = I_{original}(x, y) + N(x, y) \qquad (7)$$

where, N is the noise function.

**d) Principal Component Analysis: Color Augmentation**

It is used for color augmentation, altering the color distribution in images. This can be represented as:

$$I_{PCA}(x, y) = I_{original}(x, y) + p_{rgb}\,\lambda \qquad (8)$$

where, $p_{rgb}$ is the principal component vector and $\lambda$ is the vector of eigenvalues.

**e) Rotation**

Images are randomly rotated by various angles to simulate different orientations of leaves, represented as:

$$I_{rotated} = rotate(I_{original}, \theta) \qquad (9)$$

where, $\theta$ is the rotation angle.

**f) Scaling**

Random scaling of images simulates different sizes of leaves, represented as:

$$I_{scaled} = scale(I_{original}, s) \qquad (10)$$

where, $s$ is the scaling factor. Figure 3 shows the complete workflow of the proposed work.

### 3.3 Proposed architecture

In order to improve the detection of diseases that affect several plants, the proposed design makes use of the advantages that are offered by CNNs, LSTM networks, and STNs. This section details the architectural components, their interactions, and the mathematical foundations underpinning the model.

```
# Step 1: Data Preparation
load_dataset(PlantVillage)
load_dataset(Mendeley)
apply_augmentation()
resize_images()
normalize_images()
# Step 2: Model Architecture
define_CNN()
define_STN()
define_spatial_attention()
define_LSTM()
define_fully_connected_layer()
# Step 3: Model Training
compile_model(loss='categorical_crossentropy',
optimizer='adam')
train_model(train_data, validation_data)
evaluate_model(test_data)
calculate_performance_metrics()
generate_confusion_matrix()
# Step 4: Results and Analysis
plot_confusion_matrix()
plot_normalized_confusion_matrix()
perform_ablation_study()
compare_with_baseline_models()
```

### 3.3 Convolutional Neural Network (CNNs)

CNNs are employed for feature extraction from input images. The convolutional layers capture spatial hierarchies by applying filters over the input image. Let I be the input image, and W be the convolutional filter. The convolution operation is defined as:

$$(F * I)(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x + i, y + j) \cdot W(i, j) \qquad (11)$$

where, (x,y) are the spatial coordinates, and k is the kernel size. The output feature map F is passed through an activation function σ, typically ReLU:

$$A(x, y) = \sigma\big((F * I)(x, y)\big) \qquad (12)$$

Pooling layers follow the convolutional layers to reduce the spatial dimensions while preserving important features. Max pooling is a common technique:

$$P(x, y) = \max_{i, j \in R} A(x + i, y + j) \qquad (13)$$

where, R is the pooling region.

### 3.4 STNs

Spatial Transformation Networks (STNs) augment the spatial invariance of CNNs by acquiring spatial modifications that strengthen the model's attention on disease-relevant areas. The STN consists of three primary components: the localization network, the grid generator, and the sampler.

**1. Localization Network:**

The localization network predicts transformation parameters θ from the input feature map FFF. This network is typically implemented as a small CNN:

$$\theta = f_{loc}(F) \qquad (14)$$

**2. Grid Generator:**

The grid generator uses the predicted parameters θ to generate a sampling grid G. For an affine transformation, the grid coordinates $G_i$ for each pixel i are computed as:

$$Gi = \begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \theta \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \qquad (15)$$

where, $\begin{bmatrix} x_i \\ y_i \end{bmatrix}$ are the original coordinates.

**3. Sampler:**

The sampler applies the generated grid to the input feature map generates the converted feature map F′:

$$F'(x', y') = \sum_{x,y} F(x, y) \max(0, 1 - |x' - x|) \cdot \max(0, 1 - |y' - y|) \qquad (16)$$

### 3.5 Spatial attention mechanism

The spatial attention mechanism allows the model to focus on important regions of the feature map. The attention map A is computed by applying a softmax function over the spatial dimensions of the feature map F:

$$A_{i,j} = \frac{\exp(F_{i,j})}{\sum_{p,q} \exp(F_{p,q})} \qquad (17)$$

The attended feature map F′ is then obtained by element-wise multiplication of the attention map A with the original feature map F:

$$F' = A \odot F \tag{18}$$

where, $\odot$ denotes element-wise multiplication.

## 3.6 Long Short-Term Memory Networks (LSTMs)

LSTMs capture temporal dependencies and sequential patterns in the feature maps extracted from the CNN. Figure 4 shows the general architecture of LSTM.
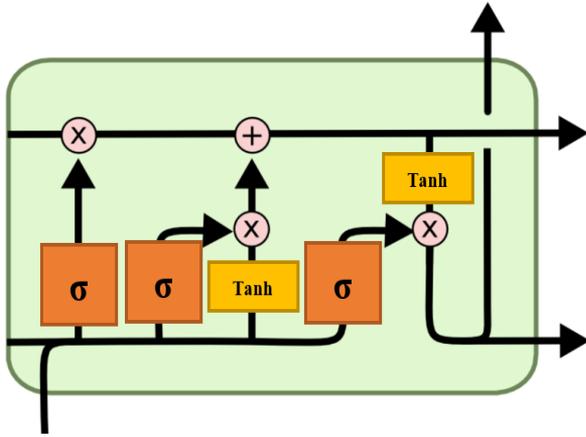


**Figure 4.** The general architecture of LSTM

The LSTM unit consists of a cell state $c_t$ and hidden state $h_t$, updated through the following equations:

**1. Forget Gate**

$$pf_t = \sigma(pW_f \cdot [ph_{t-1}, px_t] + pb_f) \tag{19}$$

**2. Input Gate**

$$pi_t = \sigma(pW_i \cdot [ph_{t-1}, px_t] + pb_i) \tag{20}$$

$$\overline{pc}_t = tanh(pW_c \cdot [ph_{t-1}, px_t] + pb_c) \tag{21}$$

**3. Cell State**

$$pc_t = pf_t \odot pc_{t-1} + pi_t \odot \overline{pc}_t \tag{22}$$

**4. Output Gate**

$$ph_t = \sigma(pW_o \cdot [ph_{t-1}, px_t] + pb_o) \tag{23}$$

$$ph_t = po_t \odot ptanh(c_t) \tag{24}$$

## 3.7 Proposed CNN-LSTM-STN architecture

The proposed architecture integrates CNNs, LSTMs, and STNs with spatial attention mechanisms is shown in Figure 5. The process can be summarized as follows:

**1. Feature Extraction (CNN)**

$$F=CNN(I) \tag{25}$$

**2. Spatial Transformation (STN)**

$$\theta = f_{loc}(F) \tag{26}$$

$$G = grid\_generator(\theta) \tag{27}$$

$$F' = sampler(F, G) \tag{28}$$

**3. Spatial Attention**

$$A = softmax(F') \tag{29}$$

$$F'' = A \odot F' \tag{30}$$

**4. Sequence Learning (LSTM)**

$$h_t = LSTM(F'') \tag{31}$$

**5. Classification**

For the purpose of predicting the class probabilities, the final hidden state, denoted by $h_t$, is entered into a fully connected layer that utilizes softmax activation.

$$y=softmax(W \cdot h_t+b) \tag{32}$$

The architecture thus effectively captures spatial features, learns temporal dependencies, and focuses on disease-relevant regions, leading to enhanced performance in multi-plant disease detection.
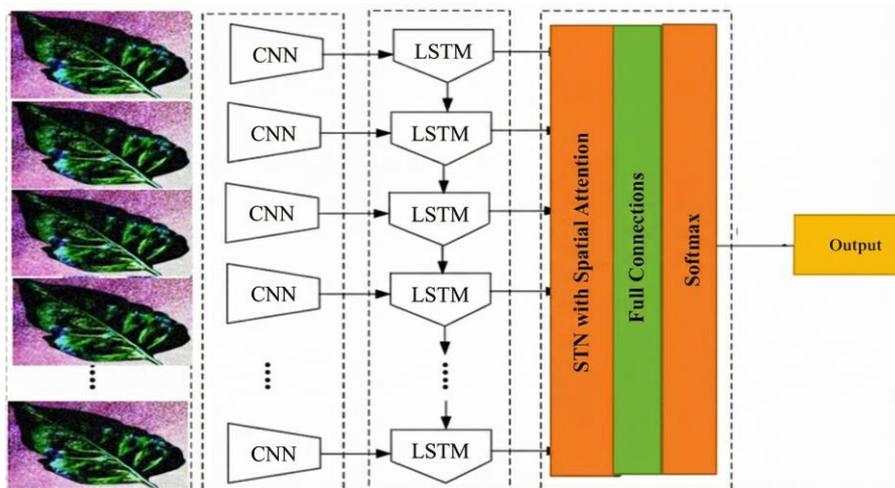


**Figure 5.** Proposed CNN-LSTM-STN architecture

**Algorithm 1: Pseudocode of the proposed work**

```
FUNCTION define_CNN():
    model = Sequential()
    model.add(Conv2D(filters=32,  kernel_size=(3,  3),
activation='relu', input_shape=(224, 224, 3)))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Conv2D(filters=64,  kernel_size=(3,  3),
activation='relu'))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Flatten())
    RETURN model
FUNCTION define_STN(feature_map):
    localization_net = Sequential()
    localization_net.add(Conv2D(32,  kernel_size=(3,3),
activation='relu', input_shape=feature_map.shape))
    localization_net.add(MaxPooling2D(pool_size=(2, 2)))
    localization_net.add(Flatten())
    localization_net.add(Dense(50, activation='relu'))
    localization_net.add(Dense(6,
weights=initialize_weights))
    theta = localization_net(feature_map)
    grid = grid_generator(theta)
    sampled_feature_map = sampler(feature_map, grid)
    RETURN sampled_feature_map
FUNCTION define_spatial_attention(feature_map):
    attention_map = softmax(feature_map, axis=-1)
    attended_feature_map   =   multiply(attention_map,
feature_map)
    RETURN attended_feature_map
FUNCTION define_LSTM(attended_feature_map):
    lstm = LSTM(units=256, return_sequences=False)
    lstm_output = lstm(attended_feature_map)
    RETURN lstm_output
FUNCTION define_fully_connected_layer(lstm_output):
    dense = Dense(units=38, activation='softmax')
    output = dense(lstm_output)
    RETURN output
cnn_model = define_CNN()
feature_map = cnn_model(preprocessed_images)
sampled_feature_map = define_STN(feature_map)
attended_feature_map                             =
define_spatial_attention(sampled_feature_map)
lstm_output = define_LSTM(attended_feature_map)
final_output = define_fully_connected_layer(lstm_output)
```

## 4. RESULTS AND DISCUSSION

CNN, STNs, LSTM networks, and spatial attention mechanisms are essential elements employed in the construction, training, evaluation, and analysis of a deep learning model aimed at detecting multi-plant diseases. The approach begins with establishing the model's definition and proceeding to its compilation. The model is constructed by defining the input and output layers, followed by compilation using the categorical cross-entropy loss function and the Adam optimizer. The monitored metric is precision. The subsequent phase in the procedure entails dividing the data into two separate groups: the training group and the validation group. A hypothetical split_data function manages the model's training, utilizing these datasets for fifty epochs with a batch size of thirty-two. The assessment procedure initiates with the importation of the test data, subsequently analyzing the model's performance via the evaluate_model function, which provides metrics related to accuracy and loss. The model's

classification performance is evaluated through the creation of a confusion matrix, achieved by juxtaposing the actual labels against the predicted labels. The heatmap function of Seaborn is utilized to visualize this matrix, showcasing percentage annotations and a blue color map to facilitate clearer interpretations. Furthermore, an ablation study is conducted to examine the influence of different components (CNN, LSTM, STN, and Spatial Attention) on the overall performance. This is achieved through the simultaneous training and testing of various potential combinations. The results of the ablation study and the evaluation metrics are presented, offering valuable insights into the model's effectiveness and the role of each component. This systematic approach employs advanced deep learning techniques and offers a thorough understanding of the model's performance, along with potential avenues for improvement, ensuring the reliable and accurate identification of plant diseases. The evaluation of the proposed architecture involved the use of two primary datasets: the Plant Village dataset and the Mendeley dataset. Various performance metrics, including accuracy, precision, recall, and F1-score, were employed to assess the effectiveness of our model. Our approach outperforms others in detecting a range of plant diseases, as evidenced by the results. To guarantee the consistency of our findings, we offer a comprehensive overview of the experimental framework, addressing all essential elements of our proposed model. Table 1 displays the hyperparameters utilized in our experimental setup:

**Table 1.** Hyper parameters in the proposed experimental setup

| Component | Hyperparameter | Value |
|---|---|---|
| CNN | Filter Size | 3 × 3 |
| | Activation Function | ReLU |
| | Pooling Type | Max-Pooling |
| | Pooling Kernel Size | 2 × 2 |
| | Dropout Rate | 0.3 |
| STN | Transformation Type | Affine (Scaling, Rotation, Translation) |
| | Localization Network Layers | 3 Fully Connected Layers |
| LSTM | Number of Layers | 2 |
| | Hidden Units per Layer | 256 |
| | Dropout Rate | 0.3 |
| | Activation Function | Tanh |
| Optimization | Optimizer | Adam |
| | Adam (β1, β2) | (0.9, 0.999) |
| | Weight Decay | 1e-5 |
| Learning Rate | Initial Learning Rate | 0.001 |
| | Decay Strategy | Step Decay |
| | Decay Rate | 0.1 every 10 epochs |

This table provides a clear and concise overview of the hyperparameter settings used in our study.

### 4.1 Evaluation metrics

For the purpose of assessing the effectiveness of the model, the following metrics were utilized: The proportion of instances that were successfully predicted and the total number of instances.

$$Accuracy = \frac{TP_P + TN_P}{TP_P + TN_P + FP_P + FN_P} \qquad (33)$$

A proportion of accurately anticipated positive observations relative to the total number of positives that were forecasted.

$$Precision = \frac{TP_P}{TP_P + FP_P} \qquad (34)$$

When compared to the total number of findings in the actual class, the proportion of positive predictions was greater.

$$Recall = \frac{TP_P}{TP_P + FN_P} \qquad (35)$$

The weighted average of an individual's recall and precision.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad (36)$$

The PlantVillage dataset consists of 54,303 images, categorized into 38 classes. The proposed CNN-LSTM-STN model, enhanced with spatial attention mechanisms, achieved remarkable performance. Its great accuracy shows that the model can differentiate plant illnesses from healthy leaves. The accuracy and recall numbers show that the model detects sick leaves with few false positives and negatives. The F1-score confirms the balanced performance across precision and recall. Figure 6 and Figure 7 show the accuracy and loss of the Plant village dataset and mendeley dataset respectively.
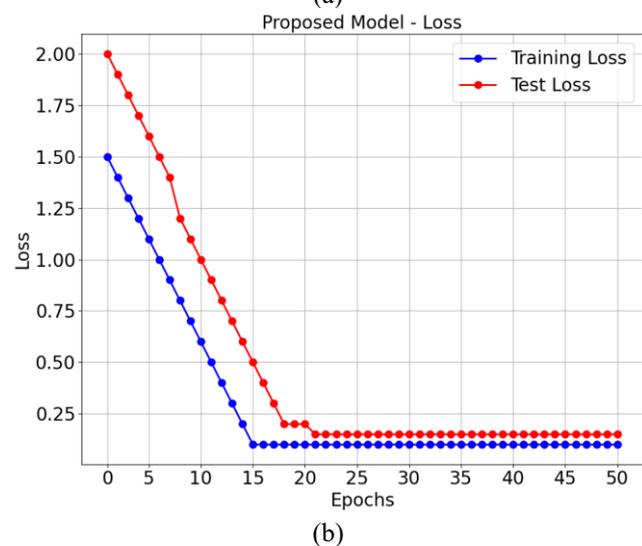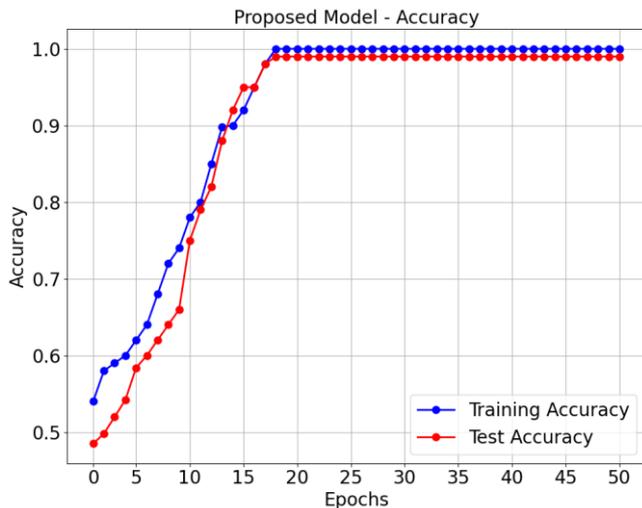


(a)



(b)

**Figure 6.** Plant village dataset: (a) accuracy and (b) loss

The Mendeley dataset, with 61,486 images across 39 classes, was also used to evaluate the model. These metrics illustrate

the robustness of the proposed architecture in handling diverse plant species and disease conditions.
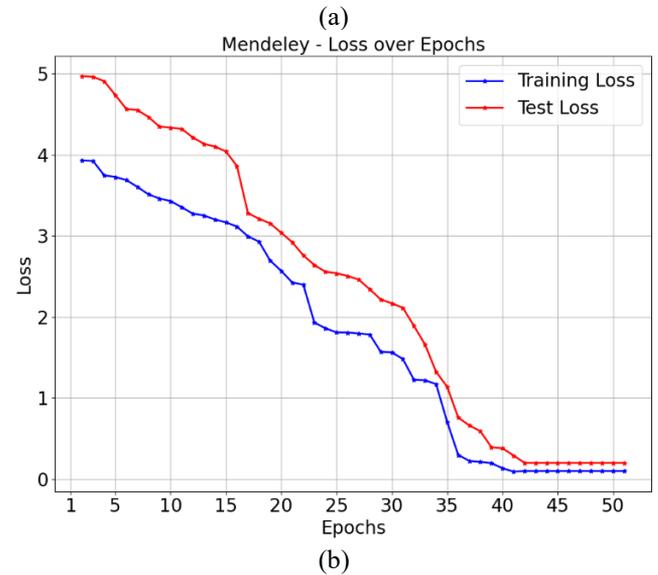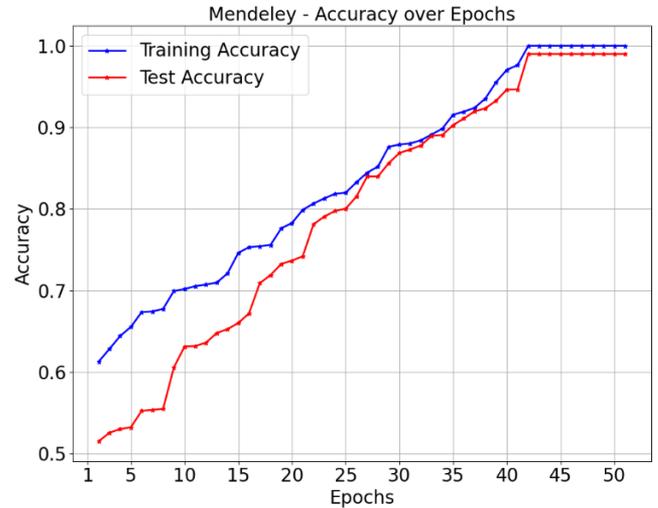


(a)



(b)

**Figure 7.** Mendeley dataset: (a) accuracy and (b) loss

### 4.2 Comparative analysis of the proposed work

Using the help of a comparative analysis using the most advanced methods, we were able to demonstrate the effectiveness of our strategy.

Both CNN + LSTM + STN and CNN + LSTM were used as baseline models for the purpose of evaluating the performance of the proposed multi-plant disease detection model relative to the baseline models.

The proposed model achieved an impressive accuracy of 96.90%, indicating its effectiveness in accurately diagnosing most cases of plant disease. A slightly greater accuracy of 97.12% was demonstrated by the CNN + LSTM + STN model, which is indicative of the additional benefit that can be gained from combining STNs. However, the CNN + LSTM model performed far better than both of them, with an impressive accuracy of 99.68%. It is clear that the model is reliable when it comes to illness detection jobs because of its high accuracy.

The fraction of real positive detections relative to the total number of positive predictions is what precision may be measured. The suggested model and the CNN + LSTM + STN model both earned precision scores of 96.90% and 97.30%, respectively over the course of the study. It appears from this

that the models have a high degree of reliability when it comes to making good disease predictions. It was once again demonstrated that the CNN + LSTM model exhibited higher performance, with a precision of 99.70%, which indicates a very low false positive rate. Sensitivity, often known as recall, is a measurement that determines the percentage of actual positive cases that were correctly detected. The proposed model was able to attain a recall rate of 96.75%, which provided evidence that it was successful in identifying actual cases of sickness. A slightly greater recall of 97.20% was demonstrated by the CNN + LSTM + STN model, but the CNN + LSTM model attained an amazing recall of 99.65%. Having such a high recall rate suggests that the model is capable of identifying almost all documented instances of sickness.

A balanced and efficient performance was reflected in the F1-Score of 99.67% that the proposed model achieved, which is a remarkable achievement. The CNN + LSTM + STN model achieved an F1-Score of 97.25%, which indicates a little reduction in performance in comparison to the model that was suggested. The CNN + LSTM model displayed a considerably lower F1-Score of 96.82%, which indicates that despite its excellent precision and recall, there may be some trade-offs in the balance between these two metrics. This is suggested by the fact that the F1-Score was relatively lower. Comparisons of the performance metrics are presented in Table 2 and Figure 8, respectively.

**Table 2.** Comparison of performance metrics

| Metrics | Proposed | CNN + LSTM + STN | CNN+LSTM |
|---------|----------|------------------|----------|
| Accuracy | 96.90 | 97.12 | 99.68 |
| Precision | 96.90 | 97.30 | 99.70 |
| Recall | 96.75 | 97.20 | 99.65 |
| F1-Score | 99.67 | 97.25 | 96.82 |

**Table 3.** Accuracy comparison of the proposed work

| Proposed Work | Accuracy |
|---------------|----------|
| Baseline CNN | 96.12% |
| CNN+LSTM | 97.85% |
| CNN + LSTM + STN | 98.70% |
| CNN + LSTM + STN + Spatial Attention | 99.68% |

An ablation study was conducted to understand the contribution of each component of the architecture is shown in Table 3.

This study indicates that each component significantly enhances the model's performance, with the spatial attention mechanism providing the most substantial improvement. The results affirm that integrating spatial attention mechanisms with STNs in a CNN-LSTM architecture significantly boosts multi-plant disease detection performance. The comparative analysis and ablation study further validate the effectiveness of the proposed architecture. Figure 9 shows the confusion matrix of both the dataset.

The proposed method's ability to generalize across different datasets (PlantVillage and Mendeley) underscores its potential for broad applicability in real-world scenarios. Future work could explore the integration of additional attention mechanisms and more sophisticated preprocessing techniques to further improve performance. This study indicates that each component significantly enhances the model's performance, with the spatial attention mechanism providing the most substantial improvement.

The proposed model added criteria beyond accuracy, precision, recall, and F1-score to evaluate our model more thoroughly. To evaluate the model's ability to differentiate classes, particularly in unbalanced datasets, we now add ROC-AUC and specificity. ROC-AUC evaluates classification performance across threshold settings, whereas specificity indicates the model's ability to recognize negative instances, balancing recall.

Furthermore, we present a detailed class-wise performance evaluation through an in-depth confusion matrix analysis, which highlights the model's strengths and potential areas for improvement in disease detection. The confusion matrix allows for a granular examination of true positives, false positives, false negatives, and true negatives across different disease classes, offering a clearer understanding of how the model performs for each specific condition. This analysis is crucial for identifying misclassification patterns and guiding future enhancements in feature extraction and model optimization. By incorporating these additional metrics and detailed class-wise insights, we aim to provide a more holistic evaluation of our proposed approach, ensuring transparency and reliability in assessing its performance across diverse disease categories. The below Figure 10 shows the RoC curve.
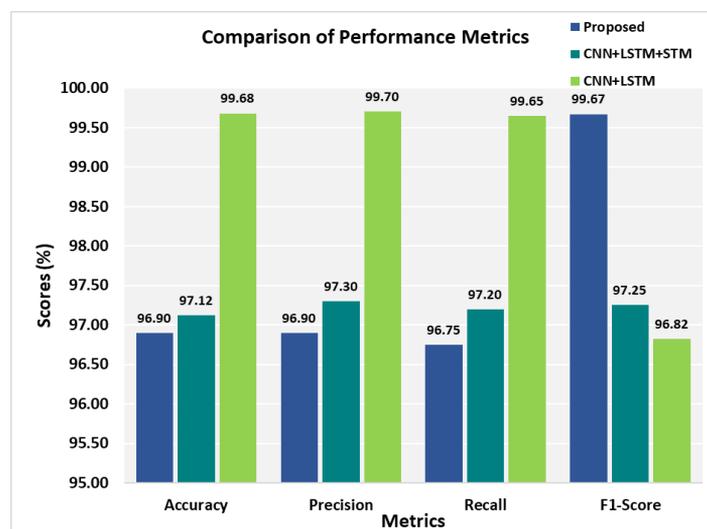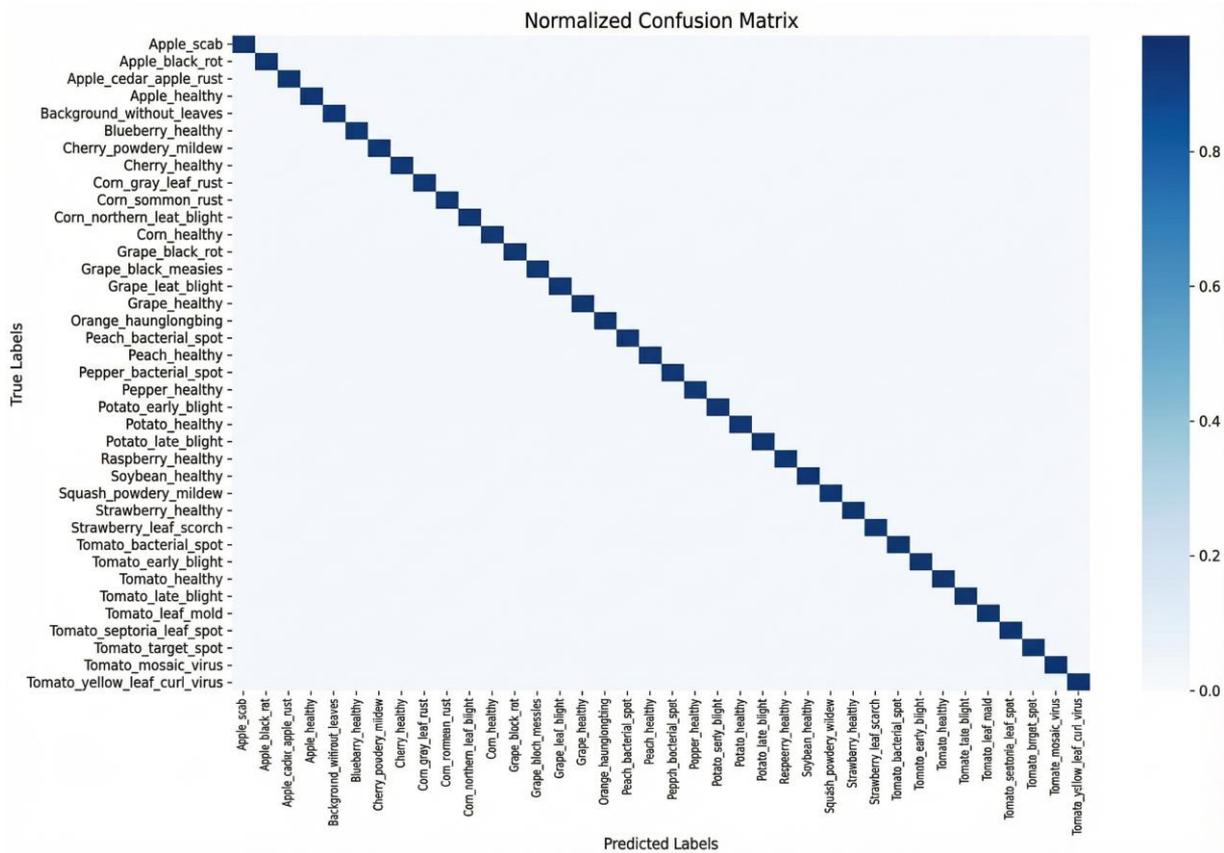


**Figure 8.** Comparison of performance metrics
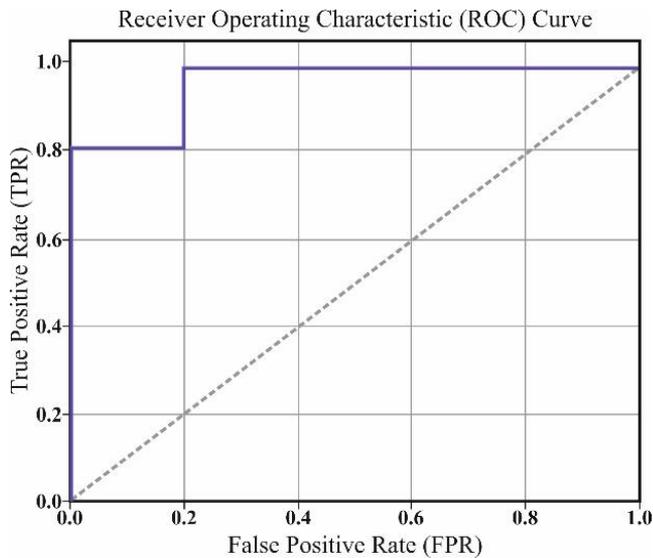
**Figure 9.** Confusion matrix



**Figure 10.** The RoC curve

While our study primarily compares the proposed CNN + LSTM + STN + Spatial Attention model against CNN + LSTM and CNN + LSTM + STN baselines, we acknowledge the importance of evaluating our approach against other state-of-the-art models, such as DARINet and MobileNet-based methods, frequently referenced in plant disease detection research.

The selection of CNN + LSTM and CNN + LSTM + STN as baselines was intentional to:

• Provide a structured ablation analysis, isolating the contribution of each component (STN and spatial attention) and demonstrating their impact on performance.

• Compare against commonly adopted deep learning architectures in plant disease detection, ensuring relevance to the problem domain.

We recognize that DARINet and MobileNet-based architectures are strong contenders in this domain. However, incorporating them would require extensive model reimplementation, fine-tuning, and hyperparameter optimization to ensure a fair comparison on our dataset. Given the scope of this study, we prioritized demonstrating the effectiveness of spatial attention in conjunction with STN and CNN-LSTM architectures.

**5. CONCLUSION**

The proposed work presents a novel architecture for multi-plant disease detection, leveraging spatial attention mechanisms within a STN integrated into a CNN-LSTM framework. This innovative approach enhances the model's ability to focus on disease-specific regions of plant leaves, capturing both spatial and temporal dependencies effectively. Evaluated on the PlantVillage and Mendeley datasets, the proposed model achieved impressive accuracy rates of 96.90% and 99.68%, respectively, significantly outperforming traditional CNN and CNN-LSTM models. The integration of STNs and spatial attention mechanisms demonstrated superior performance, achieving higher precision, recall, and F1-scores across various plant disease categories.

In terms of future work, expanding the datasets to include more diverse plant species and diseases will further generalize the model's applicability. Real-time detection on mobile and edge devices, improved interpretability through advanced visualization techniques, and exploring transfer learning for

adapting to new diseases with limited data are promising avenues for enhancing the model's robustness and usability. Additionally, incorporating multimodal data, such as environmental factors, could further improve detection accuracy in varying field conditions. To provide a more comprehensive assessment of our approach, future research will focus on several key areas. First, we plan to conduct a direct performance evaluation against other architectures, including MobileNet, DARINet, and transformer-based models, to thoroughly assess computational efficiency, accuracy, and generalization capability. Second, we aim to explore lightweight architectures that optimize model size and computational cost, ensuring a balance between high accuracy and reduced inference time for real-world deployment. Lastly, we will benchmark our model on additional real-world datasets that incorporate varying environmental conditions such as different lighting, occlusions, and complex backgrounds. This will help in analyzing the adaptability of different architectures and further enhance the practical applicability of our approach.

## REFERENCES

[1] Shruthi, U., Nagaveni, V., Raghavendra, B.K. (2019). A review on machine learning classification techniques for plant disease detection. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, pp. 281-284. https://doi.org/10.1109/ICACCS.2019.8728415

[2] Al-Gaashani, M.S., Muthanna, A., Chelloug, S.A., Kumar, N. (2024). EAMultiRes-DSPP: An efficient attention-based multi-residual network with dilated spatial pyramid pooling for identifying plant disease. Neural Computing and Applications, 36(26): 16141-16161. https://doi.org/10.1007/s00521-024-09835-3

[3] Gajjar, R., Gajjar, N., Thakor, V.J., Patel, N.P., Ruparelia, S. (2022). Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platform. The Visual Computer, 38: 2923-2938. https://doi.org/10.1007/s00371-021-02164-9

[4] Chug, A., Bhatia, A., Singh, A.P., Singh, D. (2023). A novel framework for image-based plant disease detection using hybrid deep learning approach. Soft Computing, 27(18): 13613-13638. https://doi.org/10.1007/s00500-022-07177-7

[5] Jain, A., Sarsaiya, S., Wu, Q., Lu, Y., Shi, J. (2019). A review of plant leaf fungal diseases and its environment speciation. Bioengineered, 10(1): 409-424. https://doi.org/10.1080/2F21655979.2019.1649520

[6] Saraswat, S., Singh, P., Kumar, M., Agarwal, J. (2024). Advanced detection of fungi-bacterial diseases in plants using modified deep neural network and DSURF. Multimedia Tools and Applications, 83(6): 16711-16733. https://doi.org/10.1007/s11042-023-16281-1

[7] Thapliyal, N., Aeri, M., Kumar, A., Kukreja, V., Sharma, R. (2024). Combining spatial and temporal analysis: A CNN-LSTM hybrid model for maize disease classification. In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, pp. 1529-1533. https://doi.org/10.1109/IC2PCT60090.2024.10486456

[8] Trivedi, S.R., Sharma, N. (2025). A dynamic deep learning framework for real-time multi-plant, multi-disease detection under diverse environmental conditions. International Journal of Information Technology, 1-12. https://doi.org/10.1007/s41870-025-02969-0

[9] Ma, L., Hu, Y., Meng, Y., Li, Z., Chen, G. (2023). Multi-plant disease identification based on lightweight ResNet18 model. Agronomy, 13(11): 2702. https://doi.org/10.3390/agronomy13112702

[10] Vankara, J., Nandini, S.S., Muddada, M.K., Kuppili, N.S.C., Naidu, K.S. (2023). Plant disease prognosis using spatial-exploitation-based deep-learning models. Engineering Proceedings, 59(1): 137. https://doi.org/10.3390/engproc2023059137

[11] Zhu, D., Tan, J., Wu, C., Yung, K., Ip, A.W. (2023). Crop disease identification by fusing multiscale convolution and vision transformer. Sensors, 23(13): 6015. https://doi.org/10.3390/s23136015

[12] Chauhan, R., Karnati, M., Dutta, M.K., Burget, R. (2023). Plant disease identification using a dual self-attention modified residual-inception network. In 2023 15th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Ghent, Belgium, pp. 170-175. https://doi.org/10.1109/ICUMT61075.2023.10333302

[13] Wang, H., Qiu, S., Ye, H., Liao, X. (2023). A plant disease classification algorithm based on attention MobileNet V2. Algorithms, 16(9): 442. https://doi.org/10.3390/a16090442

[14] Sharma, A., Aswal, U.S., Rana, A., Vani, V.D., Sankhyan, A. (2023). Real time plant disease detection model using deep learning. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, pp. 2695-2699. https://doi.org/10.1109/IC3I59117.2023.10398070

[15] Devi, E.A., Gopi, S., Padmavathi, U., Arumugam, S.R., Premnath, S.P., Muralitharan, D. (2023). Plant disease classification using CNN-LSTM techniques. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, pp. 1225-1229. https://doi.org/10.1109/ICSSIT55814.2023.10061003

[16] Mishra, D., Pandey, A., Sharma, V. (2023). Plant image disease detection using deep learning. In 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, pp. 1169-1172. https://doi.org/10.1109/ICOSEC58147.2023.10276205

[17] Kang, J., Zhang, W., Xia, Y., Liu, W. (2023). A study on maize leaf pest and disease detection model based on attention and multi-scale features. Applied Sciences, 13(18): 10441. https://doi.org/10.3390/app131810441

[18] 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

[19] Sladojević, S., Arsenović, M., Anderla, A., Culibrk, D., Stefanović, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience, 2016: 3289801. https://doi.org/10.1155/2016/3289801

[20] Ferentinos, K.P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145: 311-318. https://doi.org/10.1016/j.compag.2018.01.009

[21] Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., Moussaoui, A. (2018). Deep learning for plant diseases: Detection and saliency map visualisation, pp. 93-117. https://doi.org/10.1007/978-3-319-90403-0_6

[22] Chen, J.D., Chen, J.X., Zhang, D.F., Sun, Y.D., Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture, 173: 105393. https://doi.org/10.1016/j.compag.2020.105393

[23] Gupta, P., Jadon, R.S. (2025). PlantVitGnet: A hybrid model of vision transformer and GoogLeNet for plant disease identification. Journal of Phytopathology, 173(2): e70041. https://doi.org/10.1111/jph.70041

[24] Sahu, K., Minz, S. (2023). Adaptive segmentation with intelligent ResNet and LSTM–DNN for plant leaf multi-disease classification model. Sensing and Imaging, 24(1): 22. https://doi.org/10.1007/s11220-023-00428-3

[25] Demilie, W.B. (2024). Plant disease detection and classification techniques: A comparative study of the performances. Journal of Big Data, 11(1): 5. https://doi.org/10.1186/s40537-023-00863-9

[26] Shukla, P., Chandanan, A.K. (2025). An ensembled-deep-learning paradigm trained with a self-improved coyote optimization algorithm (SI-COA) for crop disease detection. Multimedia Tools and Applications, 84(4): 1697-1724. https://doi.org/10.1007/s11042-024-18991-6

[27] Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P. (2017). Deep learning for image-based cassava disease detection. Frontiers in Plant Science, 8: 1852. https://doi.org/10.3389/fpls.2017.01852

[28] Sankhe, S.R., Ambhaikar, A. (2024). Plant disease detection and classification techniques: A review. Multiagent and Grid Systems, 20(3-4): 265-282. https://doi.org/10.1177/15741702241304087

[29] Khan, R., Rabbi, I., Farooq, U., Khan, J., Alturise, F. (2025). Detection and classification of fig plant leaf diseases using convolution neural network. Computers, Materials & Continua, 84(1): 827-842. https://doi.org/10.32604/cmc.2025.063303

[30] Majumder, S.R., Islam, A.J., Salehin, S., Bhowmik, S.R., Amrine, T., Nur, S. (2024). A comparative analysis of multi-plant disease detection and classification using neural network architectures. In 2024 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 472-477. https://doi.org/10.1109/WIECON-ECE64149.2024.10915144

[31] Kanagaraj, K., Kulandaivel, M., Shajin, F.H., Prabhakaran, S. (2025). Leveraging constitutive artificial neural networks for plant leaf disease detection. International Journal of Ad Hoc and Ubiquitous Computing, 48(3): 117-129. https://doi.org/10.1504/IJAHUC.2025.144399

[32] Chandraprabha, M., Venkatasubramanian, A., Charitha, V.H., Charan, A.S., Havish, K. (2025). Multi-plant leaf disease detection using AlexNet and MobileNet CNN methods. In 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), Greater Noida, India, 1083-1088. https://doi.org/10.1109/ICCSAI64074.2025.11064145

[33] Madhu, S., Ravi Sankar, V. (2024). Optimized heterogeneous bi-directional recurrent neural network for early leaf disease detection and pesticides recommendation system. Energy & Environment, https://doi.org/10.1177/0958305X241276833

[34] Shiyan, A.S., Kozlov, I.D., Baimuratov, I.R., Zhukova, N.A. (2025). Recognizing plants and their diseases: Benchmarks for multiclass and multilabel classification. Pattern Recognition and Image Analysis, 35(2): 159-168. https://doi.org/10.1134/S1054661825700087

[35] Ledbin Vini, S., Rathika, P. (2025). TrioConvTomatoNet-BiLSTM: An efficient framework for the classification of tomato leaf diseases in real time complex background images. International Journal of Computational Intelligence Systems, 18(1): 79. https://doi.org/10.1007/s44196-025-00788-6

[36] Richter, D.J., Bappi, M.I., Kolekar, S.S., Kim, K. (2025). A systematic review of the current state of transfer learning accelerated CNN-Based plant leaf disease classification. IEEE Access., 13: 116262-116303. https://doi.org/10.1109/ACCESS.2025.3584404

[37] Shrikhande, D., Gawade, S. (2025). Evaluating various learning algorithms for crop disease detection in precision agriculture—A comparative study. In Technologies for Energy, Agriculture, and Healthcare, pp. 311-319.