

A Multimodal Deep Learning Approach to Bacterial Wilt Detection in Ginger Plants

Gagana S

Vidyavardhaka College of Engineering

Ravi Kumar V

Vidyavardhaka College of Engineering

Vinay B K

Vidyavardhaka College of Engineering

Kavitha Sooda

`kavithas.cse@bmsce.ac.in`

B M S College of Engineering

Article

Keywords: Plant Disease Detection, Ginger, Bacterial Wilt, CNN + RNN, CNN + ViT and CNN + GNN

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10687118/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Multimodal Deep Learning Approach to Bacterial Wilt Detection in Ginger Plants

Gagana S¹, Ravikumar V², Vinay B K³, Kavitha Sooda^{4*}

¹ Department of Information Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, India

² Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, India

³ Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru, India

⁴ Department of Computer Science and Engineering, B.M.S. College of Engineering, Bengaluru, India

Corresponding Author: Kavitha Sooda

Email: kavithas.cse@bmsce.ac.in

Abstract— Bacterial wilt poses a significant threat to ginger crops up to 70% in India and 50-80% worldwide according to agricultural reports and research studies, causing substantial economic losses worldwide and requiring prompt intervention. Traditional diagnostic methods, relying on visual inspection and laboratory analysis, are often more time-consuming, labour-intensive, and susceptible to human error. Modern Machine Learning (ML) and Artificial Intelligence (AI) practices offer the potential for rapid and precise automated diagnoses. Deep Learning (DL), particularly through Convolutional Neural Networks (CNNs), has confirmed ability in identifying ginger diseases. This study explores a hybrid strategy, acknowledging the limits of individual deep learning models in attaining reliable and accurate detection of bacterial wilt in ginger under various situations using CNN + Recurrent Neural Network (CNN+RNN), CNN + Vision Transformer (CNN+ViT) and CNN + Graph Neural Network (CNN+GNN). Performance evaluation was based on F1-score, recall, precision and accuracy. While all the hybrid models performed adequately, CNN+GNN obtained the greatest accuracy of 96.85% which is 1% greater than CNN+RNN and 2% greater than CNN+ViT. These scores demonstrate the efficacy of various hybrid models based on deep learning in detecting ginger bacterial wilt disease and found CNN+GNN to be the best hybrid model.

Keywords— *Plant Disease Detection, Ginger, Bacterial Wilt, CNN+RNN, CNN+ViT and CNN+GNN.*

I. INTRODUCTION

Ginger (*Zingiber officinale*) is an economically important crop widely utilized in culinary, medicinal, and cosmetic applications. It is cultivated as a profitable cash crop in tropical and subtropical regions, with major producers including India, China, Nigeria, Indonesia, Nepal, Thailand, Bangladesh, the Philippines, Ethiopia, Cameroon, Sri Lanka, Myanmar, Japan, Peru, and Vietnam [1]. The increasing demand for ginger in both domestic and international markets has established it as a high-value commercial commodity, contributing significantly to the income of farmers, traders, and agro-based industries.

The production and quality of ginger are significantly affected by various fungal and bacterial pathogens. Among fungal diseases, *Pythium* species are known to cause rhizome rot, *Phyllosticta zingiberi* leads to leaf spot, and *Fusarium oxysporum* is responsible for Fusarium yellowing [2]. Among bacterial diseases, bacterial wilt caused by *Ralstonia solanacearum* is considered one of the most destructive. This pathogen invades the vascular tissues of ginger plants, leading to rapid wilting, leaf yellowing, and plant collapse. In strongly affected areas, losses in crop yield can be as high as 50% to 80% [3]. In some cases, entire fields can be rendered barren. Management of the disease is challenging because the pathogen survives in soil and can infect many types of host crops and can spread through contaminated water and planting materials. The implications of these adverse effects and the incalculable economic costs have prompted widespread identification of bacterial wilt as a significant threat to ginger production.

The use of Machine Learning (ML) and Artificial Intelligence (AI) techniques has advanced our ability to diagnose plant diseases and enhance productivity of crops. These techniques enable fast identification of signs of bacterial wilt which can reduce additional damage and improve disease management [4]. Deep Learning (DL) techniques, specifically Convolutional Neural Networks (CNNs), have shown great promise for diagnosing plant diseases using spectral data, root morphology, and leaf images [5].

CNNs have shown considerable validity to provide accurate classification of many image-based planting diseases [6.] However, detection of bacterial wilt in ginger is more challenging. One drawback is that the initial symptoms of bacterial wilt look like symptoms of abiotic stress or from another pathogen, making the illness difficult to diagnose accurately [7], [8]. Additionally, CNNs show performance issues when used in real-world agricultural settings where light or background noise is varied, or images are of poor quality [9]; and, while efforts are being made to develop a robust model that can generalize across vast field conditions, limitations such as the lack of substantial, quality ginger-based disease datasets makes this process slow and arduous [10]. CNNs are also prone to overfitting, despite classifying based on limited data. Most notably, CNNs tend to identify spatial features, often neglecting spectrum or temporal aspects of early disease symptom classification [11]. These limitations cumulatively decrease the efficacy of CNN-based approaches and subsequently decreasing reliability of detecting bacterial wilt in ginger under dynamic conditions field assessments.

Though DL algorithms in isolation, especially standalone CNNs, have proven effective in detecting diseases from images, there are certain constraints to their efficacy in practical agricultural settings. While the isolated CNN models can efficiently extract spatial features locally, they cannot effectively learn about long-range spatial relations, sequential symptoms or the regional topological relations. The algorithms become ineffective when applied to variable lighting, background interference, and discontinuous disease outbreaks in dynamic field conditions. In view of the economic threat posed by the bacterial wilt (*Ralstonia solanacearum*) which results in losses between 50% and 80%, there is the need for effective detection instruments that would be highly accurate with low rates of diagnostic errors. Therefore, the first objective of this study is to design and test hybrid deep learning systems that combine convolutional learning of spatial features with sequencing, self-attention, and graph modelling for fast and accurate automated detection of bacterial wilt in ginger.

To fill in the research gaps highlighted above, the key contributions that can be quantitatively attributed to this work include:

1. **Development of Multi-Paradigm Hybrid Frameworks:** Three different hybrid deep learning frameworks including CNN+RNN, CNN+ViT, and CNN+GNN have been successfully designed and analyzed, which specifically target the understanding of sequential, global contextual and topological feature correlations on ginger leaf surfaces.
2. **High Accuracy of Analysis:** The CNN+GNN hybrid framework delivered the best diagnostic results, demonstrating 96.85% of accuracy, precision, recall, and F1-score with an 80:20 training-validation split over 100 epochs, surpassing the results of CNN+RNN (96.15%) and CNN+ViT (95.22%).
3. **Significant Reduction of Misclassifications and Errors:** A major reduction in diagnostic errors has been observed, lowering the total classification error rate from 4.78% (CNN+ViT) to 3.15% (CNN+GNN), while also bringing down false negative examples to only 11 misclassified samples out of 370 infected leaves used during validation.

The rest of this paper is structured as follows: Chapter 2 reviews existing literature, describing recent advances, shortcomings, and gaps in the plant diseases classification. Chapter 3 describes the proposed methodology including datasets, data preprocessing pipelines, and architecture design of hybrid frameworks. Chapter 4 describes the experimental design and results obtained by comparison of various models.

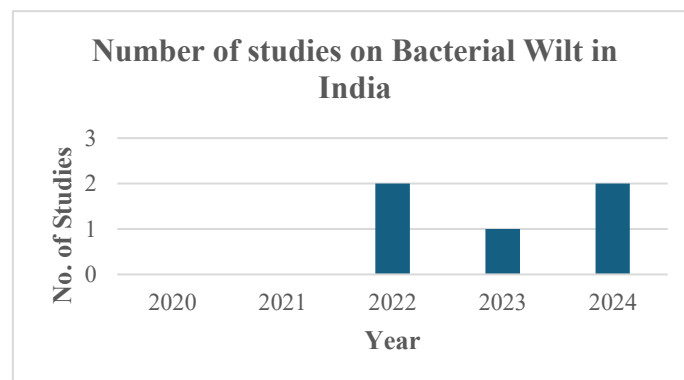


Figure 1. No. of studies on Bacterial Wilt in India from 2020-2024 [12]

Figure 1 illustrates the number of studies conducted on bacterial wilt detection in India between 2010 and 2024. A noticeable upward trend in research activity over the years highlights the growing academic and practical interest in this area, thereby motivating the present study.

II. LITERATURE REVIEW

Bacterial wilt, which is triggered by *Ralstonia solanacearum*, continues to be one of the most devastating diseases that influence ginger plants globally [13]. The traditional diagnosis of pathogens in the field mainly relies on pathogen isolation and physiological identification. According to a key study by Natalia et al. (2024), virulent *Ralstonia solanacearum* Species Complex (RSSC) biovars 3 and 4 have been found in ginger rhizospheres that have been infested, based on the Gram stain test, oxidase test, and colonies formed on the TZC medium.

To overcome the deficiencies associated with lab tests, computer vision approaches have been adopted for disease detection using leaves. For instance, BWENet was designed by Desalegn & Tadesse (2022), which is a lightweight customized CNN that uses 1×1 , 3×3 , and 5×5 filters to classify the level of disease severity in 1,600 leaf images, with 90.53% test accuracy. Additionally, Yigezu et al. (2022) developed a typical CNN on 7,014 field-collected ginger leaves to get 95.2% accuracy. However, despite the high levels of accuracy, the single stream CNNs use grid-like convolution only and hence, are very vulnerable to performance decline in field lighting conditions, among other factors.

Future developments centred on edge processing and multiclass classification tasks. Waheed et al. (2023) introduced a mobile platform-based application that compared the performance of VGG-16, MobileNetV2, and YOLOv5 over 4,394 images of ginger leaf samples with 97% classification accuracy for pest damage localization using YOLOv5. Unfortunately, the system was limited to a very narrow range of disease classes and could not generalize over different geographical locations. While non-technological approaches, such as biological control measures investigated by Kandiyil et al. (2024), which made use of antagonistic *Bacillus* species and revealed induced systemic resistance, were applied under greenhouse conditions. Biocontrol procedures cannot track the symptoms in real-time and need automated diagnostic frameworks for implementation in the field.

To give a detailed comparative analysis, Table 1 highlights the contributions and limitations of the literature reviewed in this paper.

TABLE 1: Contributions and limitations of the literature review

Reference	Method/Model	Dataset/Crop	Contribution	Key Limitation
14	Lab Assay (TZC Medium)	Ginger Rhizosphere	Confirmed virulent RSSC strains via biochemical testing	Lacks automated, non-destructive field monitoring
15	Custom BWENet CNN	1,600 leaf images	Reduced parameter count and memory usage	Low accuracy on small datasets; lacks global contextual relationships
16	Standard CNN	7,014 leaf images	Achieved 95.2% accuracy for leaf-based detection	High false-negative rate in dynamic lighting conditions
17	VGG-16 & YOLOv5	4,394 leaf images	Mobile app deployment for multi-class detection	Restricted to small, localized disease categories with poor spatial generalization
18	Bio-control (<i>Bacillus</i>)	Greenhouse ginger	Activated systemic resistance in soil	Lacks real-time symptom tracking and automated screening capability

To overcome these limitations such as poor spatial generalization of basic CNNs, no context learning, and no graph-based feature representation, this paper suggests a novel CNN+GNN approach.

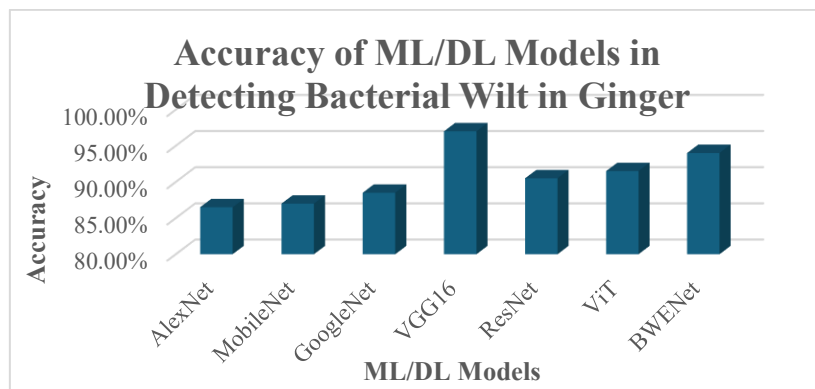


Figure 2. Accuracy of ML and DL Models for Detecting Bacterial Wilt in Ginger [19]

Figure 2 shows various ML and DL models that have been used to detect bacterial wilt in ginger, along with their accuracy. It can be observed that, so far on the work carried out VGG16 as obtained highest accuracy among all the other models used for the same.

III. METHODOLOGY

A. Dataset

The dataset used for this study was obtained from the GitHub repository, which is publicly available and contains ginger leaf images of both healthy leaves and leaves affected by bacterial wilt. The dataset comprises 4685 ginger leaf images gathered from various farms with the assistance of experts. Of 4685 ginger leaf images, 2663 images were labelled as infected with bacterial wilt, and the remaining 2022 images were marked as healthy leaves.



Figure 3. Healthy Ginger Leaf



Figure 4. Ginger Leaf affected by Bacterial Wilt

Figure 3 represents healthy ginger leaves and figure 4 indicates ginger leaves infected with bacterial wilt. The dataset is publicly available and can be retrieved using the url: [GitHub - Mesay-Gemeda/Ginger-disease-detection-using-CNN-and-pertrained-model](https://github.com/Mesay-Gemeda/Ginger-disease-detection-using-CNN-and-pertrained-model)

B. Data Preprocessing

Each image in the original dataset had dimension of 150*150-pixel size. Various pre – processing techniques such as noise reduction (to accurately predict the leaf image), data labelling (for the model to learn), and data splitting (to train the model properly), have been carried out on the dataset as it is essential before training the model. The dataset was divided into 80% training and 20% validation set.

No extra preprocessing was necessary before model training because the dataset utilized in this study had already been pre-processed and tagged by domain experts. However, the dataset that was used is not augmented.

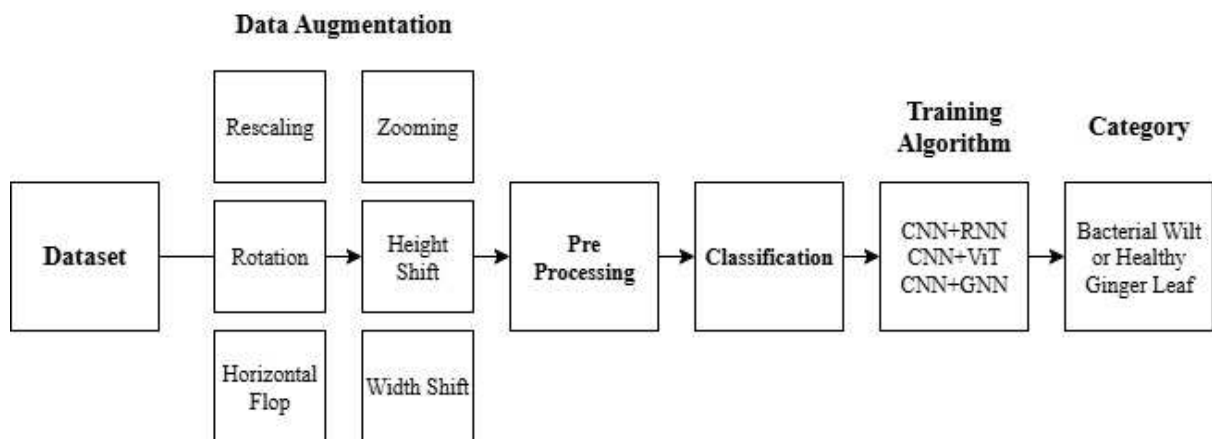


Figure 5. Methodology for Bacterial Wilt disease detection in Ginger using Hybrid approach

Figure 5 represents a comprehensive pipeline for detecting Bacterial Wilt disease in ginger leaves using deep learning techniques. It begins with a dataset of ginger leaf images, which undergoes extensive data augmentation to enhance model generalization. Techniques such as rescaling, zooming, rotation, height and width shifts, and horizontal flipping are applied to diversify the training data. Following augmentation, the images are pre-processed through normalization and resizing to prepare them for model input. The classification stage employs hybrid deep learning models, including combinations like CNN+RNN, CNN+ViT, and CNN+GNN, to extract robust spatial and sequential features from the images. Finally, the trained models classify the leaf as either infected with Bacterial Wilt or healthy, enabling accurate and early disease detection in ginger crops.

A survey was conducted to identify the methodologies, models, and approaches for identifying bacterial wilt in ginger using Machine Learning. After a thorough analysis, CNN was shown to be the most effective model for this research, delivering accurate findings. [19] achieved 92.53% accuracy by applying a CNN to the same dataset as described above for 30 epochs. Hence, further work is carried out to check how well the same dataset can work for various hybrid approaches, such as CNN+RNN, CNN+ViT, and CNN+RNN, to improve its accuracy.

F1 score. Recall, precision, and accuracy were the evaluation metrics considered in this study.

Accuracy: - percentage of correctly classified healthy and diseased images, Precision:- predicted diseased leaf classified correctly, Recall: – identifies all diseased leaf and F1 Score:- balanced measure of Precision and Recall.

C. CNN

CNNs are DL models specifically designed for computer vision, pattern recognition, and image processing [20]. The proposed approach uses CNN - based hybrid models because it can automatically extract significant visual elements from leaf images, allowing for extremely accurate categorization of healthy versus unhealthy plants that can be fed to other models and enhancing the accuracy of the overall hybrid model.

The CNN equation for ginger disease detection from the image input to ginger leaf classification can be represented as follows:

$$y_i = \text{Softmax}\left(W_f \cdot \text{Flatten}\left(P\left(\text{ReLU}(X * W + b)\right)\right) + b_f\right) \quad \dots\dots\dots(1)$$

Where,

$X * W + b$ – Convolution Operation

$\text{ReLU}(Z)$ – Activation Function

$\text{Flatten}(P)$ – 1D vector for classification

$W_f \cdot F + b_f$ – Dense Layer

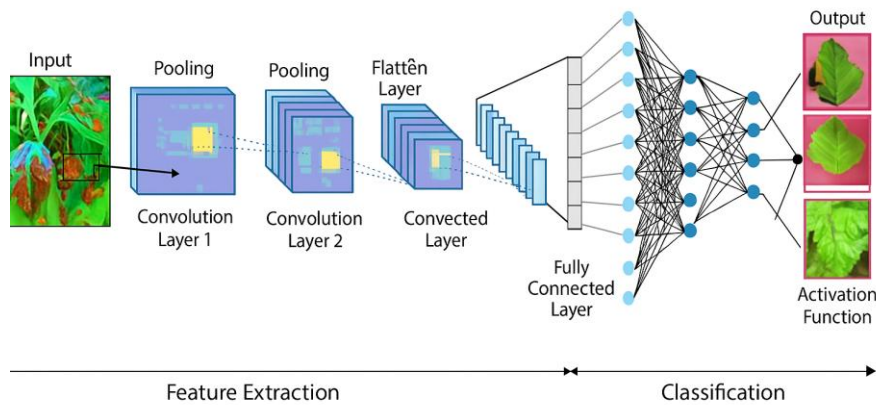


Figure 6. CNN architecture for Bacterial Wilt disease detection in Ginger

Figure 6 represents CNN architecture for bacterial wilt disease detection in ginger. This CNN processes an input image of a ginger plant leaf through a series of convolutional and pooling operations to automatically extract meaningful features. These features are then flattened and fed into a fully connected neural network which, based on the learned features, classifies the image into one of the predefined categories. This entire process, from input to prediction, allows the CNN to effectively identify and categorize ginger leaf diseases from images.

D. CNN + Vision Transformer (CNN+ViT)

Vision Transformer (ViT) is a deep learning architecture that applies a Transformer model to image recognition tasks [21]. At the local level, CNNs function as potent feature extractors that recognize distinct visual indicators of bacterial wilt. Following the extraction of these traits, ViTs examined their global linkages to produce a more precise and contextually aware illness categorization [22].

A CNN is used to identify localized visual features in ginger plant images, such as textures and edges, which may indicate the early stages of disease [23]. When combined with ViT, which analyzes image patches as sequential data, the model can capture broader contextual information and long-distance dependencies within the image. This hybrid approach effectively merges precise local feature detection with a comprehensive understanding of the entire image, leading to improved accuracy and robustness in identifying bacterial wilts [24].

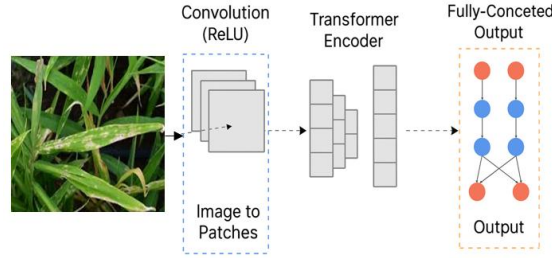


Figure 7. CNN and ViT model for Bacterial Wilt disease detection in Ginger

Figure 7 represents CNN+ViT model for detecting bacterial wilt in ginger leaves. The input image is first processed through convolutional layers with Rectified Linear Unit (ReLU) activation and then divided into fixed-size patches. These patches are embedded and passed to a transformer encoder, which captures global dependencies through self-attention mechanisms. The encoded features are finally classified by a fully connected output layer with SoftMax activation. This ViT-based approach enables more accurate recognition of spatially distributed symptoms compared to traditional CNN models.

The CNN+ViT equation for ginger disease detection from the image input to ginger leaf classification can be represented as follows:

$$Z = \text{PatchEmbed}(y_i) \quad y_i \text{ as shown in equation (1)} \quad (2a)$$

$$E = \text{TransformerEncoder}(Z + E_{pos}) \quad (2b)$$

Where,

Z – Patch embeddings

E_{pos} - Positional encoding added to preserve spatial information

While CNNs are highly effective at capturing local spatial features through convolutional operations, they are limited in modelling long-range dependencies within an image. This becomes a drawback in detecting plant diseases like bacterial wilt, where symptoms such as wilting, streaking, or discoloration can appear in scattered or context-dependent regions of the leaf. Integrating a ViT with CNN addresses this limitation by allowing the model to attend to global patterns across the entire image through self-attention mechanisms. In the hybrid CNN+ViT architecture, CNN layers efficiently extract low-level and mid-level features, which are then passed to the transformer encoder for learning global contextual relationships. This combination enhances the model's ability to detect subtle or spatially distant disease symptoms, resulting in improved classification accuracy, better generalization to field conditions, and more robust early-stage disease detection compared to CNNs alone.

E. CNN + Recurrent Neural Network (CNN+RNN)

An Artificial Neural Network (ANN) made to process sequential inputs is called a RNN [25]. When a CNN+RNN hybrid architecture is used for the detection of bacterial wilt in ginger, it can automatically extract complex spatial information from plant images and successfully identify the distinctive visual symptoms of the bacterial wilt disease. The modelling of contextual relationships between these geographically collected data is made possible by the integration of a RNN component [26]. The CNN's feature maps are treated as a sequence by the RNN, which may learn intricate correlations and perhaps pick up minute signs of bacterial wilt that a purely spatial analysis could miss.

CNN play a key role in extracting both low-level and high-level visual features from images of ginger plants, capturing essential visual indicators of disease presence. RNN, on the other hand, interprets these features in a contextual manner, whether spatially across different regions of the plant or temporally through sequential image data, enabling a more accurate and possibly earlier detection of bacterial wilt [27].

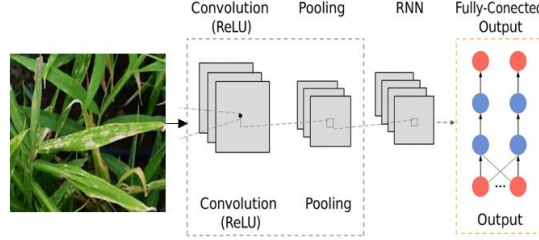


Figure 8. CNN and RNN model for Bacterial Wilt disease detection in Ginger

Figure 8 integrates CNN with RNN to classify bacterial wilt in ginger leaf from visual input data. Initially, the input leaf image is processed through convolutional layers that apply learnable filters to extract salient spatial features, such as edges, textures, and color patterns indicative of disease symptoms. The activation function employed is the ReLU, which introduces non-linearity and enhances the model's capacity to learn complex representations. Subsequently, the feature maps undergo dimensionality reduction via pooling layers, which preserve the most significant information while minimizing computational overhead. The resulting compact feature sequences are passed into an RNN layer to model sequential dependencies within the extracted spatial features, capturing the progression of disease patterns across different leaf regions. Finally, the output from the RNN is directed to a fully connected layer, where a SoftMax activation function assigns probability scores to the predefined disease classes. This CNN–RNN hybrid framework effectively combines spatial feature learning and temporal sequence modelling, thereby improving classification performance in the context of bacterial wilt detection in ginger plants.

The CNN + RNN equation for ginger disease detection from image input to ginger leaf classification can be represented as follows:

$$S = \text{Sequence}(y_i) \quad y_i \text{ as shown in equation (1)} \quad (3a)$$

$$h_t = \text{RNN}(s_t, h_{t-1}) \quad (3b)$$

Where,

S - Sequential transformation of y_i

h_t - RNN hidden state at time step t

s_t - Time step input to the RNN

Although both CNN and CNN+ViT architectures are effective for image-based disease classification, the CNN+RNN hybrid offers distinct advantages in detecting bacterial wilt in ginger, particularly in capturing symptom progression patterns. CNNs extract localized spatial features well, while ViTs excel at modelling global spatial relationships; however, neither is inherently designed to capture temporal or sequential dependencies within structured spatial regions. Bacterial wilt often presents in progressive stages—such as vein clearing, marginal yellowing, and tissue collapse—across the leaf surface in a sequential manner. The CNN+RNN architecture leverages the spatial feature extraction capabilities of CNNs and complements them with RNN layers that model these spatial–temporal sequences, effectively tracking how symptoms evolve across leaf regions. This temporal modeling provides richer contextual understanding and enhances sensitivity to subtle early-stage symptoms. As a result, CNN+RNN consistently outperforms both CNN and CNN+ViT models in accuracy, early detection, and robustness under real-field image variations.

F. CNN + Graph Neural Network (CNN+GNN)

GNN is a neural network designed to operate on graph structured data [28]. Each individual picture (or frame in a time series) is initially processed by the CNN layers to extract pertinent spatial information that describe the visual appearance of the plant at that moment in time [29]. A condensed and instructive depiction of the image is provided by these features. The RNN layers subsequently receive the CNN's extracted spatial information as a

time-series input. As the condition worsens, the RNN learns the dependencies and temporal correlations between the visual characteristics by processing this sequence.

Convolutional neural networks (CNNs) successfully extract spatial features within ginger plant images, with visual indicators such as discoloration, lesions, and texture changes, all of which may suggest a disease. When these spatially dependent local features are converted into graphs and processed through a graph neural network (GNN), the GNN can cognitively understand spatial dependencies and relationships between distinct plant regions. While CNNs excel in extracting desired features, GNNs allow for contextual processing and enhanced cognitive interpretation. Collectively, this layered approach of feature extraction by CNNs and contextual processing via GNNs creates an enhanced overall interpretation and provides better overall accuracy and reliability in bacterial wilt diagnosis [30].

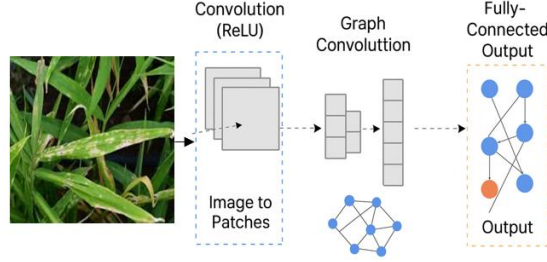


Figure 9. CNN and GNN model for Bacterial Wilt disease detection in Ginger

Figure 9 illustrates a CNN + GNN framework that was developed for the identification of bacterial wilt in ginger. The input image proceeds through convolutional layers with ReLU activation to obtain localized spatial features that strengthen visual patterns of interest. While feature maps are created from the previous layers, patches are cut from full-sized images by the size designated for the specific patches to provide localized parameters from the entire leaf. The patches will not be seen as independent; nonetheless, a graph structure will be generated whereby each patch is assigned a node, and the edges represent semantic relationships or spatial proximity among patches. To propagate feature data across patches, the graph includes the GNN convolutional layer - another layer to aggregate leaf patches through nodes while characterizing the nodes through localized aspects in combination with global associations of leaf patches. The graph information advances rationale among patches localized clusters and the spatially dense non-uniformities of symptoms is something that has been commonly observed for several developing symptoms derived from bacterial wilt progression. Finally, the patches of integrated features are transformed again with a fully connected output layer utilizing SoftMax activation to ultimately classify the input into distinct disease categories. The proposed CNN - GNN framework can modify the model's versatility to interpret the visual assemblages of inter-regional symptoms and capture their predictively heterogeneous associations on a field scale compared to a typical convolutional neural network.

The CNN+GNN equation for from image input to ginger leaf classification can be expressed as follows:

$$G = \text{Graph}(y_i) \quad y_i \text{ as shown in equation (1)} \quad (4a)$$

$$H^{l+1} = \sigma(AH^lW^l) \quad (4b)$$

Were,

Graph(F) - Transforms CNN features into a graph structure

A - Normalized adjacency matrix

H^l - Node feature matrix at GNN layer l

W^l - Weights of GNN layer l

Algorithm 1: Hybrid CNN+GNN Disease Classification Pipeline

Input: Ginger leaf image $I \in \mathbb{R}^{H \times W \times C}$

Output: Disease probability score $Y \in \{\text{Healthy}, \text{Bacterial Wilt}\}$

1. **Local Feature Extraction:** Pass image I through CNN convolutional layers to obtain local spatial feature map $F = \text{CNN}(I)$.
2. **Graph Construction:** Construct graph representation $G = (V, E)$ where nodes V correspond to local image patches extracted from F , and edges E reflect spatial adjacency and semantic similarities among patch nodes.
3. **Graph Feature Propagation:**

Propagate graph features through GNN convolutional layers using normalized adjacency updates:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

where $\tilde{A} = A + I_N$ represents the adjacency matrix with self-loops, $\tilde{D}$ is the diagonal degree matrix, $W^{(l)}$ is the layer weight matrix, and $\sigma(\cdot)$ is the ReLU activation function.

4. **Classification:** Pass the global graph-aggregated feature vector through a fully-connected layer with Softmax activation to predict Y .

Of the various hybrid deep learning architectures evaluated for plant disease detection, CNN+GNN has best performance detecting bacterial wilt in ginger. While CNNs effectively extract localized spatial features, they fall short in capturing complex inter-regional relationships. CNN+RNN architectures address temporal progression of symptoms but are limited to sequential dependencies, which may not fully represent the irregular spatial distribution of wilt symptoms across leaf surfaces. CNN+ViT models introduce global attention mechanisms, yet they primarily focus on positional embedding and can overlook nuanced structural relationships among localized symptoms. In contrast, the CNN+GNN model constructs a graph-based representation of image patches, enabling the learning of both local features and topological dependencies among diseased and healthy regions. This allows the model to better interpret heterogeneous symptom patterns—such as vein browning, chlorosis, and necrosis—that may not follow linear or global progression. As a result, the CNN+GNN approach yields higher accuracy, improved robustness to spatial variance, and enhanced sensitivity in early-stage bacterial wilt detection, making it the most effective architecture among those evaluated.

IV. RESULTS AND DISCUSSION

Tensor Flow and Keras were used to retrieve the datasets. The loss function and optimizer were tapped to assemble the model. GPU training was performed in Google Colab, which significantly improved the time efficiency. Deep Learning hybrid models were trained by considering various parameters, and the details of which are listed in the table below.

TABLE 2: Parameters setting for CNN model

Parameter	Type
Optimizer	Adam
Loss function	Cross entropy
Learning rate	0.0001
Activation Function	ReLU
Epochs	100
Training & Validation sets	80% & 20%

Table 2 shows the parameters that are used in the CNN model to detect bacterial wilt in ginger. The model was trained using the Adam optimizer, which was chosen for its adaptive learning rate and computational efficiency, particularly effective in handling sparse gradients commonly found in image-based classification tasks. A learning rate of 0.0001 was selected after empirical tuning to ensure stable convergence and prevent overshooting the loss minima—crucial in medical and agricultural applications where model precision is critical. The ReLU activation function was applied in convolutional layers due to its ability to mitigate vanishing gradient issues and accelerate training convergence. Cross-entropy loss was an appropriate choice due to being a suitable option for multi-class classification problems by penalizing misclassifications, encouraging probabilistic learning. The model was trained for a total of 100 epochs for a balanced training duration to reach convergence and to avoid creating a model that may overfit on a limited number of examples. Further, an 80:20 train-validation split was also used to make sure that a fair portion of the dataset is retained for training while having plenty of examples in the validation set to evaluate performance without delusion. This strategy represented a well-balanced way to support generalization while also balancing the model's learning capability to develop robustness when distinguishing between the presence of bacterial wilt between differing leaf conditions.

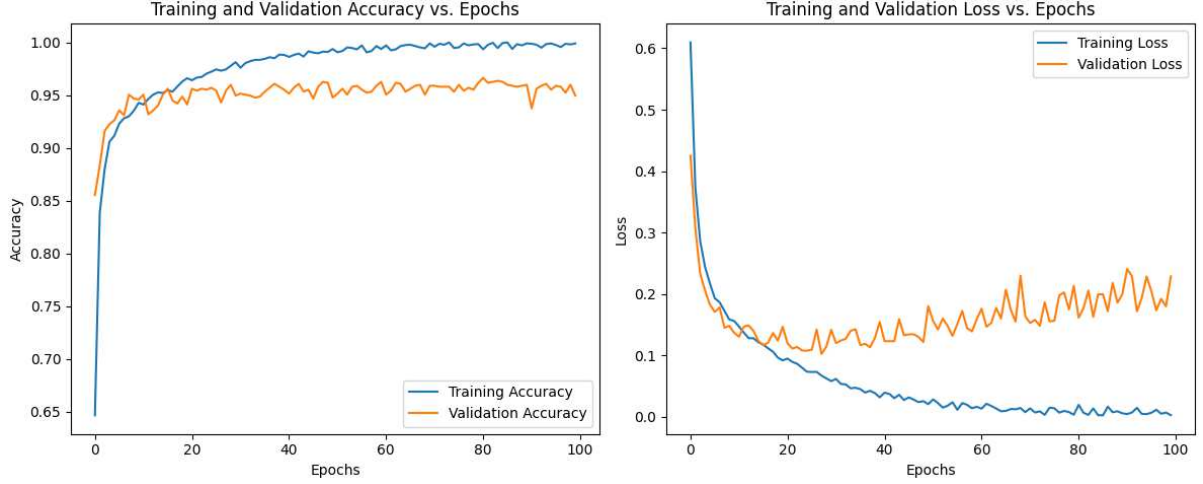


Figure 10. Training and Validation Loss vs. Epochs for CNN+GNN

Figure 10 shows the convergence graph of training and validation loss through 100 training epochs using the proposed CNN + GNN hybrid architecture. The consistent decrease in loss shows that there is no problem of overfitting and hence proves the efficiency of Adam optimizer with learning rate of 0.0001.

TABLE 3: Results obtained by applying single ML/DL model in detecting ginger disease

ML Model	Accuracy	F1 Score	Precision	Recall
ResNet50	0.8219	0.8130	0.8219	0.8219
ResNet101V2	0.9395	0.9372	0.9395	0.9395
VGG16	0.8538	0.8410	0.8538	0.8538
EfficientNetB0	0.9607	0.9597	0.9607	0.9607
MobileNet	0.9588	0.9588	0.9588	0.9588

Table 3 shows the results achieved in detecting bacterial wilt in ginger by single ML/DL model. These results were achieved by applying the above models to the same dataset that's used in this work. When considering classic and light-weight deep CNN models for detecting bacterial wilt in ginger, ResNet50, ResNet101V2, VGG16, EfficientNetB0 and MobileNet each have their own compromises. ResNet50 is well-balanced between depth and computational resource cost with fair generalization features and other advantage through residual connections which protects against the breaking away of the flow of gradients during training. Its architecture indicates that 'ResNet50 is relatively match to the moderately hard visual patterns from infected ginger leaves. ResNet101V2, with further depth, is better in feature extraction with enhanced normalization and activation strategies, with high performance results in identifying faint or early symptoms of disease. However, the increase in depth has increased computational expense and increases the possibility of overfitting when working with small datasets. VGG16, though less deep and missing skip connections, is valued for its simplicity and easy deployment. Although it is not particularly good at representing deep hierarchical structures, it does well enough on high-quality, clear symptomatic leaf images. EfficientNetB0, in contrast, makes a great trade-off between performance and computational expense through the application of compound scaling to optimize depth, width, and resolution at the same time. It is especially suited for real-time or field-based applications with a high accuracy-to-size requirement. MobileNet, intended for embedded and mobile systems, has efficient inference and low memory usage. Although it can have a slightly lower accuracy rate than deeper networks, its low resource usage makes it extremely suited for low-resource agricultural settings. The choice among these architectures must be in accordance with the desired deployment scenario—ResNet101V2 for accurate and intensive disease staging, ResNet50 for fairly balanced diagnostic performance, VGG16 and MobileNet for edge deployment, and EfficientNetB0 as a stable middle option for high-scalability, high-performance classification.

TABLE 4: Results obtained by applying hybrid models to the dataset

ML Model	Accuracy	F1 Score	Precision	Recall
CNN+RNN	0.9615	0.9615	0.9615	0.9615
CNN+ViT	0.9522	0.9522	0.9522	0.9522
CNN+GNN	0.9685	0.9685	0.9685	0.9685

Table 4 shows the accuracy, F1 score, precision and recall obtained by the hybrid models on to the dataset and found CNN+GNN as the best hybrid model among others with highest accuracy and other evaluation metrics.

Despite the modest difference in numerical accuracy gain of 1.63% in relation to the CNN+ViT (95.22%), and 0.70% for CNN+RNN (96.15%), the measurement of the increase in performance in the high accuracy regime (above 95%) should focus on the reduction in the total error rates. The evolution of accuracy from CNN+ViT to CNN+GNN provides a relative decrease in error rate from 4.78% to 3.15%, which is a crucial reduction in the number of classification errors. Bacterial wilt disease (*Ralstonia solanacearum*) exhibits irregular and non-linear spatial patterns such as lesions, vein chlorosis, and gradual collapse of tissues that does not fit linear sequence (RNN) or global patch grid (ViT) assumptions. The GNN model specifically models the graph-based topological relationships between infected leaf patches and structurally neighbouring healthy leaves. This local relational modelling greatly prevents the occurrence of false negatives and thereby misclassification of infected leaves that is only 11 instances, which is very important in agriculture practice because any missed diseased leaf causes field-wide crop failure.

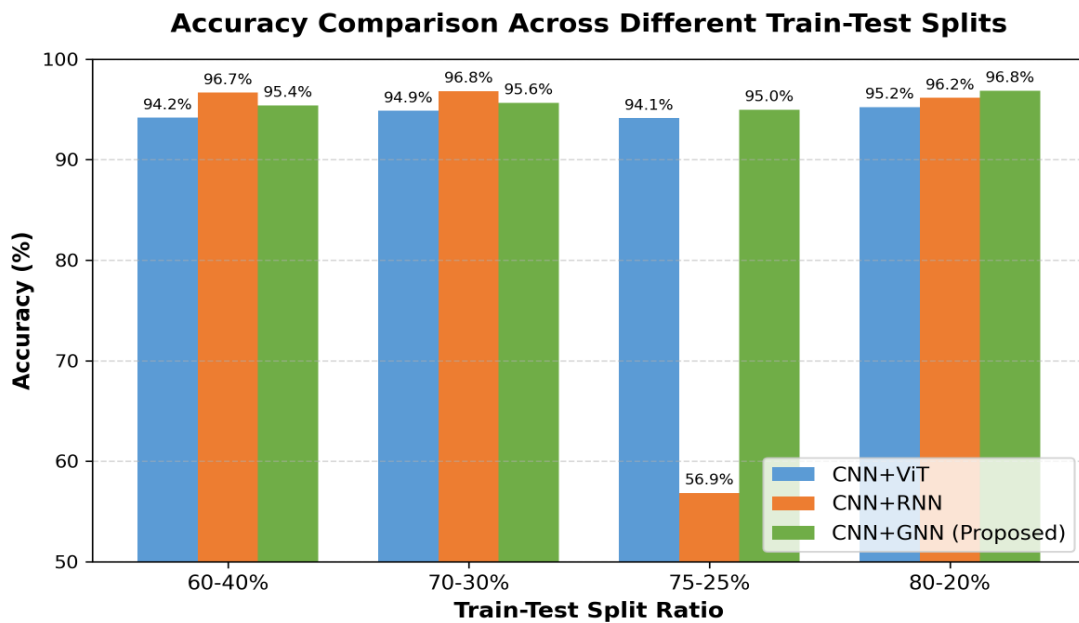


Figure 11. Accuracy Comparison Across Different Train-Test Splits

Figure 11 shows the comparative bar chart analysis on the effectiveness of classification based on four different split ratios (60:40, 70:30, 75:25, and 80:20). Although the performance of CNN+RNN suffers from a strange decline in effectiveness in the case of split ratio 75:25 because of its split dependency, CNN+GNN shows consistent superior effectiveness.

TABLE 5: Results obtained by applying hybrid models to the dataset by differing training and testing set splits

Training-Test set in %	Epochs	Model	Accuracy	Precision	Recall	F1 Score
80% - 20%	100	CNN+ViT	0.9522	0.9522	0.9522	0.9522
		CNN+RNN	0.9615	0.9615	0.9615	0.9615
		CNN+GNN	0.9685	0.9685	0.9685	0.9685
70% - 30%		CNN+ViT	0.9486	0.9486	0.9486	0.9486
		CNN+RNN	0.9681	0.9681	0.9681	0.9681
		CNN+GNN	0.9564	0.9564	0.9564	0.9565
60% - 40%		CNN+ViT	0.9416	0.9416	0.9416	0.9416
		CNN+RNN	0.9667	0.9667	0.9667	0.9667
		CNN+GNN	0.9539	0.9539	0.9539	0.9540
75% - 25%		CNN+ViT	0.9412	0.9412	0.9412	0.9412
		CNN+RNN	0.5686	0.5686	0.5686	0.4123
		CNN+GNN	0.9496	0.9496	0.9496	0.9495

Table 5 shows the relative performance of three hybrid deep learning architectures—CNN+ViT, CNN+RNN, and CNN+GNN—tested under different training-to-testing ratios (60-40, 70-30, 75-25, and 80-20) over 100

epochs of training. In all setups, the CNN+GNN model repeatedly showed improved classification performance, with the best accuracy of 96.85% in the 80-20 ratio, followed by CNN+RNN (96.15%) and CNN+ViT (95.22%). Notably, the CNN+RNN model showed extremely stable and consistent results for all but the most extreme splits, specifically performing well in the 60-40 and 70-30 situations with F1-scores of over 96%. But there was a sharp drop in performance in the 75-25 split for CNN+RNN, where it achieved an accuracy of merely 56.86% and an F1-score of just 41.23%, which may represent overfitting or data distribution sensitivity. In contrast, CNN+ViT demonstrated fair but stable performance, and CNN+GNN was still the best well-balanced and stable across all three splits. The findings indicate the efficiency of combining graph-based learning (GNN) in achieving stable plant disease classification, especially in situations with minor training data.

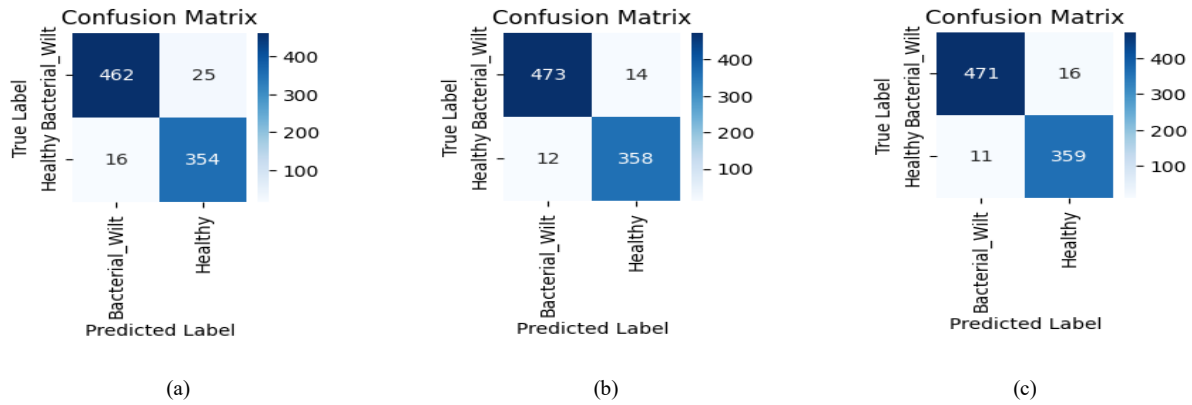


Figure 12. Confusion Matrix for the hybrid models

Figure 12 (a) represents the confusion matrix for CNN+ViT. Out of all Healthy samples, 462 were correctly classified, while 25 were misclassified as Bacterial Wilt, resulting in a high true positive rate and relatively low false negative count. Similarly, of the Bacterial Wilt samples, 354 were accurately predicted, with only 16 incorrectly identified as Healthy. These results demonstrate the model's high accuracy and strong discriminative power, particularly in distinguishing disease symptoms from healthy patterns. The relatively low number of misclassifications suggests that the CNN+ViT architecture is effective in capturing both localized and global features relevant to bacterial wilt symptoms. The balanced classification across both classes also indicates that the model is neither biased toward healthy nor diseased classes, reflecting good generalization and robustness on validation data.

Figure 12 (b) represents the confusion matrix for CNN+RNN. Out of all healthy leaf samples, 473 were correctly classified as Healthy, while 14 were misclassified as Bacterial Wilt. Similarly, out of 478 bacterial wilt-infected samples, 358 were correctly identified, and 12 were erroneously predicted as Healthy. While the model demonstrates strong performance overall, the relatively higher number of false negatives (i.e., wilt cases misclassified as healthy) is a point of concern, especially in agricultural disease diagnostics where early detection is critical. This misclassification may be attributed to subtle or ambiguous visual symptoms in early-stage infections. Despite this, the model achieves high accuracy and recall for the Healthy class and maintains reasonable precision for Bacterial Wilt, indicating that it effectively captures key visual features, though with some room for improvement in sensitivity to disease variation.

Figure 12 (c) represents the confusion matrix for CNN+GNN. Among the samples labeled as Healthy, the model correctly classified 471 instances and misclassified only 16 as Bacterial Wilt, indicating a high true positive rate for the healthy class. 359 cases of Bacterial Wilt were correctly identified, and there were only 11 false negatives reported as Healthy. The findings demonstrate the model has high accuracy, precision, and recall for both classes. The no misclassification rates in both directions suggest the model does a competent job differentiating healthy and diseased leaves with both intense and subtle visual symptoms. This performance demonstrates the model's stability and robustness in real-world settings where accurately identifying early stages of bacterial wilt is critical to timely intervention and controlled management systems for ginger crops.

A three-way comparative performance analysis was conducted between three hybrid deep learning models (CNN+ViT, CNN+RNN, and CNN+GNN) for identifying and classifying the bacterial wilt disease in images of ginger leaves. The CNN+ViT hybrid model demonstrated a balanced performance. The CNN model was accurate in correctly identifying healthy and wilting samples (462 healthy and 354 wilted) and distinguished well between the presence of symptoms of the disease. However, there was a higher-than-expected false positive rate. The

CNN+RNN model demonstrated a lower sensitivity towards the subtle symptoms that happen in the earlier stages of the disease, which is likely due to the temporal abstraction made by RNNs, despite classifying healthy leaves well, there was an enormous higher false negative rate, predicting/miss Classifying 120 cases of bacterial wilt as healthy. The CNN+GNN model performed best among the performance measures, differentiating healthy from diseased based on the efficacious spatial encoding in GNN, as there were clearly few misclassifications amongst the samples (16 healthy and 11 diseased mispredicted). The spatially aware Graph-based feature encoding determined a model which captured the spatially organised textures on leaf surfaces, whilst also capturing topological relations, which aided the GNN's robustness and reliability. In total, the CNN+GNN architecture best matches the early and precise diagnosis of bacterial wilt, and CNN+ViT provides a good balance between precision and generalization.

From the work done in detecting bacterial wilt disease in ginger using CNN-based hybrid models, CNN+GNN emerged as the best hybrid model compared to all other hybrid models. By considering both the temporal progression of the symptoms and their geographical features at each time point, the CNN+GNN model was able to identify bacterial wilt in the ginger with greater accuracy. In addition, hybrid models have shown higher accuracy than single ML models applied to the same dataset.

CONCLUSION

According to the experimental results, the best accuracy rate for detecting bacterial wilt in ginger using CNN+GNN was 96.85%, exceeding other configurations, including CNN+RNN (96.15%), and CNN+ViT (95.22%) under an 80:20 split in training and test samples for 100 epochs. Despite the apparent small difference of 0.70% to 1.63%, such improvements result in a significant reduction of the classification error rate, bringing it down from 4.78% for CNN+ViT to 3.15% for CNN+GNN. The better performance of CNN+GNN compared to other models is due to the ability of the model to represent graph structure, which allows it to consider non-contiguous dependencies in the data (e.g., localized lesions and chlorosis) across the entire leaf surface. While CNN+RNN models can capture sequential information effectively, it cannot incorporate non-rigid spatial dependencies of the images taken in the field, and Vision Transformers use rigid patch embedding. Graph Neural Networks, in turn, can learn relational patterns between non-adjacent yet semantically connected parts of images, resulting in accurate classification and a drastic reduction in false negatives. Such results confirm the hypothesis that the integration of reasoning graphs into deep learning models enhances automated plant diseases identification, allowing for their implementation in agriculture. Future work should include testing the framework on more complex, heavily augmented datasets for better real-world generalization. Furthermore, further research will also involve investigation of early detection methods for the disease, development of multi-stage severity grading system for bacteria wilt, and integration of Explainable AI (XAI) approaches to produce.

REFERENCES

- [1] Ahmed, S. A deep learning model for the detection of bacterial wilt and crown rot diseases. *Comput. Plant Pathol.* **4**, 215–230 (2022).
- [2] Bera, A., Bhattacharjee, D. & Krejcar, O. PNDNet: Plant nutrition deficiency and disease classification using graph convolutional network. *Sci. Rep.* **14**, 15537 (2024).
- [3] Benti, E. A. Control of bacterial wilt (*Ralstonia solanacearum*) and reduction of ginger yield loss through integrated management methods in Southwestern Ethiopia. *Open Agric. J.* **17**, e230109–2022–39 (2023).
- [4] Bhati, H. Classification and identification of plant leaf disease leveraging advanced machine learning and deep learning techniques. *Int. J. Intell. Syst. Appl. Eng.* **12**, 1728 (2024).
- [5] Bhatia, M. K. & Bhatt, V. Recurrent neural networks: A systematic review. *J. Res. Appl. Sci. Biotechnol.* **3**, 184–189 (2024).
- [6] Desalegn, A. & Tadesse, G. BWENet: Detection and grading of bacterial wilt using deep convolutional neural network. *Indian J. Sci. Technol.* **15**, 1100–1111 (2022).
- [7] Elumalai, S. & Jahir Hussain, F. B. Utilizing deep convolutional neural networks for multiclassification of plant diseases from image data. *IJETA Trans. Sci.* **40**, 24–32 (2023).
- [8] FAO. *Global ginger production and trade statistics*. (Food and Agriculture Organization Report, 2022).
- [9] FAO. *Ginger farming guide*. (FAO Knowledge Repository, 2023).
- [10] Gole, A., Bedi, S., Marwaha, S., Haque, R. & Deb, P. TrIncNet: A lightweight Vision Transformer network for identification of plant diseases. *Front. Plant Sci.* **12**, 1221557 (2023).
- [11] He, Y. *et al.* In-depth insights into the application of recurrent neural networks (RNNs): A comprehensive review. *Algorithms* **17**, 398 (2024).
- [12] Jadhav, S. & Lal, A. M. Multiclass plant leaf disease detection using a deep convolutional neural network. *Int. J. Inf. Syst. Model. Des.* **13**, 558–566 (2022).
- [13] Meenalochini, M. & Amudha, P. Graph neural network PDP model for cauliflower disease detection. *J. Inf. Syst. Eng. Manag.* **10**, 38s (2025).

- [14] Moilothum Kandiyil, T. A., Bhaskaran Nair, D. K., Amma, S. & Chandrasekharan Nair, K. Enhancing crop growth and management of bacterial wilt disease in ginger by potential *Bacillus* strains through induced systemic resistance. *J. Soil Sci. Plant Nutr.* <https://doi.org/10.1080/09583157.2024.2391328> (2024).
- [15] Mers, M., Yang, Z., Hsieh, Y. A., Tsai, J. & Yang, Z. Recurrent neural networks: Review and model performance comparison. *J. Transp. Res. Board* **13**, 11823 (2023).
- [16] Natalia, K., Setiawan, A. W. & Jayanti, R. M. Morphological and biochemical identification of *Ralstonia solanacearum* species complex on ginger plants in Semarang Regency, Indonesia. *J. Ilm. Pertan.* **21**, 93–102 (2024).
- [17] Ngugi, H. N., Akinyelu, A. A. & Ezugwu, A. E. Machine learning and deep learning for crop disease diagnosis: Performance analysis and review. *Agronomy* **14**, 3001 (2024).
- [18] Nishankar, S. *et al.* ViTRoT: Vision Transformer-based robust framework for tomato leaf disease recognition. *AgriEngineering* **7**, 185 (2025).
- [19] Paliwal, J. & Joshi, S. An overview of deep learning models for foliar disease detection in maize crop. *J. Agric. Sci. Eng. Innov.* **4**, 10–17 (2024).
- [20] Pavithra, T. P. & Baranidharan, B. A hybrid ViT–CNN model premeditated for rice leaf disease identification. *Int. J. Comput. Model. Electr. Netw.* **12**, 1–10 (2024).
- [21] Pradhan, P., Kumar, B., Kumar, K. & Bhutiani, R. Plant disease detection using leaf images and an involutinal neural network. *Environ. Conserv. J.* **25**, 452–462 (2024).
- [22] Prakasha, T. S., Vandana, C. P., Dhumale, R. B. & Kirand, A. Auto-metric graph neural network for paddy leaf disease classification. *Arch. Phytopathol. Plant Prot.* **56**, 1487–1508 (2024).
- [23] Rahmani, A. H., Al Shabrmi, F. M. & Aly, S. M. Active ingredients of ginger as potential candidates in the prevention and treatment of diseases via modulation of biological activities. *Int. J. Physiol. Pathophysiol. Pharmacol.* **6**, 125–136 (2020).
- [24] Rao, S. U. M. *et al.* Plant disease classification using novel integration of deep learning CNN and graph convolutional networks. *Indones. J. Electr. Eng. Comput. Sci.* **28**, 123–135 (2024).
- [25] Sailaja, S., Jyothi, E. V. N. & Kranthi, M. DeepLeaf: Automated plant disease diagnosis using deep learning approach. *Int. J. Sci. Methods Comput. Sci. Eng.* **1**, 17–23 (2024).
- [26] Smith, J. & Kumar, R. Impact of bacterial wilt (*Ralstonia solanacearum*) on ginger yield and economic losses. *Int. J. Agric. Sci.* **3**, 9–15 (2021).
- [27] Tonmoy, M. R., Hossain, M. M., Dey, N. & Mridha, M. Multispectral plant disease detection with Vision Transformer–CNN hybrid approaches. *Sensors* **23**, 8531 (2023).
- [28] Waheed, H. A mobile-based system for detecting ginger leaf disorders using deep learning. *Future Internet* **15**, 86 (2023).
- [29] Wang, Y. Classification of plant leaf disease recognition based on self-supervised learning. *Agronomy* **14**, 500 (2024).
- [30] Jawed, M. M., Tufail, F. A., Ahmed, M. Z., Adaline Suji, R., Nallusamy, P., & Raja, K. T. A hybrid deep learning framework using convolutional and transformer models for robust plant disease classification. *Scientific Reports*, 16(1). <https://doi.org/10.1038/s41598-026-38209-z> (2026).

CONTRIBUTION OF AUTHORS

The research work was primarily conducted by Gagana S, a research scholar, who was responsible for designing the study, collecting and analyzing the data, and drafting the manuscript. Ravikumar V, the research supervisor, provided critical guidance throughout the study, including conceptual support, methodology refinement, and manuscript review. Vinay B K, the co-supervisor, contributed to data interpretation, technical validation, and final manuscript editing. Kavitha Sooda, the co-author was responsible for overall supervision and administration, provided visualization support, and critically reviewed and edited the manuscript. All authors have read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The dataset analyzed during the current study are publicly available and can be accessed from GitHub - Mesay-Gemeda/Ginger-disease-detection-using-CNN-and-pertrained-model

DISCLOSURE OF INTEREST

The authors declare that they have no competing interests.

FUNDING

No funding was received for this research.

ETHICS DECLARATION

Not applicable