Vision Mamba s-power System for Accurate Plant Leaf Disease Detection

Prof. Dattatray Gopal Takale, Vansh Zalpuri, Bhavesh Sopan Uchade ,Rutuja Raju Yete, Sahil Dnyaneshwar Unhale

Department of Computer Science

Vishwakarma Institute of Technology, Pune, Maharashtra, India

vansh.zalpuri24@vit.edu, bhavesh.uchade24@vit.edu, rutuja.yete24@vit.edu, sahil.unhale24@vit.edu

ABSTRACT

Plant diseases account for nearly a third of the annual crop yield in India. This leads to several economic setbacks for farmers and the national food security status of the country. Early and accurate disease identification is the only way to prevent such losses, but laboratory facilities, expert guidance, and timely diagnosis are inaccessible to most farmers. The situation cries for affordable, scalable, and reliable technological solutions that do not need a lab environment but can work in real farming conditions.Over the last few years, the combination of Artificial Intelligence and Computer Vision has been a promising and effective technology in automating the task of plant disease identification using images of leaves. However, as it stands, the limitation of this system is that they have very few diverse datasets collected from real-field scenarios, which has proven to be a bottleneck in the performance of these systems. Most of the datasets are collected under controlled conditions and therefore do not reflect the complexities of real farms—such as lighting, background, leaf orientation, occlusion, etc. By providing annotated images of real-world scenarios, datasets such as PlantDoc have been instrumental in improving model accuracy, with many studies reporting a gain of more than 30% when models are trained on field data.This project has developed a Vision Mamba-S Powered Plant Leaf Disease Detection System to take these breakthroughs further. Vision Mamba is an innovative architecture based on selective state-space modeling that allows the model to not only get very detailed local pattern from the leaf images but also the larger context structures in the image. The advantages of Vision Mamba over traditional CNNs and transformer-based models include computational efficiency, speed of inference, and smaller memory size, thus a very suitable candidate for deployment on mobile devices and IoT platforms.The device proposed takes the leaf pictures to be analyzed, using Vision Mamba-S, extracts discriminating features, and then through classification methods, identifies the diseases with high accuracy, even in the most challenging situations in the field. Since it is a very small application, it can work in real-time, offline, and is, therefore, a solution that farmers in faraway places can take advantage of without the need for expensive devices or constant internet access.

Keywords:
Plant Disease Detection, Vision Mamba-S, Computer Vision, Artificial Intelligence, Real-Field Dataset, PlantDoc, State-Space Modeling .

INTRODUCTION

Approximately, global food consumption keeps on increasing dramatically. This is mainly due to a rising population which is estimated to increase by about 1.6% yearly. Agriculture, therefore, is highly pressured to produce more food with the same amount of land. Plant diseases are among the leading factors that threaten global agricultural productivity; in fact, they are the most substantial challenge that agriculture confronts. Worldwide, plant diseases lead to a burden of nearly US$220 billion every year. According to the Indian Council of Agricultural Research (ICAR), India suffers a loss of more than 35% in crop production every year due to pests and diseases, thus farmers' incomes are significantly decreased, and the nation's food and agricultural sustainability are directly affected.

The traditional ways of detecting farm diseases are largely dependent on a farmer's own observation, experience, or advice from agricultural experts. Quite a few times, farmers want lab results to confirm their suspicions, but labs need qualified workers, high-tech equipment, and quick service—all things which are hardly found in villages or distant areas. These constraints cause a delay in the time when the disease is identified thus infection has more time to spread and crop damage can become very severe. So, there is an urgent demand for automated, reachable, scalable, and affordable diagnostic instruments that can assist in the early detection of diseases without the need for experts.

Use of computer vision and AI (artificial intelligence) technologies have made it possible to automate plant disease detection. Over the last ten years, deep learning, especially Convolutional Neural Networks (CNNs), has been the main driver of the breakthrough in image classification tasks. CNNs can learn features automatically from the low-level images and thus they can recognize diseases that show the symptoms such as spots, lesions, discoloration, and texture by their changes in the image. Because of their high classification accuracy and quick inference speeds, CNN-based models are viable for user applications on smartphones, UAVs, or edge devices. This has been the main reason why AI-based agricultural diagnostics have been feasible and accessible for implementation in the real world.

Nevertheless, the first CNN-based plant disease detection systems had significant drawbacks. The first models were largely trained on the PlantVillage dataset which is a large but highly controlled dataset with uniform backgrounds, consistent lighting, and isolated leaves. Consequently, these models demonstrated remarkable performance in the laboratory but poorly in the real environment of the agricultural field where images contain complex backgrounds, varying degrees of lighting, overlapping leaves, and natural noise. Researchers created real-world datasets like PlantDoc to overcome this problem, which has 2,598 annotated images of 13 plant species and 27 disease/healthy classes taken directly from farm conditions. Research indicates that adapting to PlantDoc can cut classification error by 31% confirming that the reliability of the dataset is the key to stable deployment.

At the same time, model efficiency has been an important focus of research in agricultural AI. While large architectures like ResNet, DenseNet, and Vision Transformers (ViTs) perform well, they are computationally intensive and thus are not feasible for low-cost devices that are used by farmers.

Lightweight models such as MobileNet, ShuffleNet, and EfficientNet solve this problem by lowering the number of parameters and performing efficient depthwise convolutions. Nevertheless, these models sometimes weaken their capacity in distinguishing subtle disease symptoms or complicated leaf patterns.

Vision Transformers reshaped a new paradigm by substituting self-attention mechanisms for convolutions. This change allows Vision Transformers to capture long-range dependencies. However, the high memory consumption and the intense computational power requirements of ViTs limit their applications in edge devices and real-time scenarios in fields. Hybrid models integrating CNNs with attention mechanisms have increased accuracies but usually have larger architectural complexities.

Here, our proposal is the Vision Mamba-S system for precise identification of plant leaf diseases which is a next-generation architecture that implements Selective State-Space Modeling (SSM) to both locally and globally gather characteristics of a diseased leaf (e.g. lesion, spot, discoloration) as well as contextual features (e.g. leaf shape, structural patterns, overall color distribution). The primary way Vision Mamba accomplishes this is by representing long-range dependencies via structured state transitions while at the same time having linear computational complexity. In other words, it is far more efficient than Vision Transformers while it also has better global understanding than conventional CNNs. Vision Mamba-S, thanks to its mixture of high accuracy, low memory requirement, and quick inference time, is perfectly compatible with installation on smartphones, Unmanned Aerial Vehicles (UAVs), edge processors, and IoT agriculture devices. Thus, it makes disease detection at the farm level possible in real-time and without having to rely on cloud computation or costly hardware, thereby providing farmers with a diagnostic solution that is affordable, scalable, and user-friendly.

2. RELATED WORK

We have creatively divided the relevant work into two major themes: (i) plant disease detection techniques, deep learning, and computer vision methods; and (ii) real-world datasets that massively enhance research accuracy and model generalization in plant disease detection.

2.1 TECHNIQUES FOR PLANT DISEASE DETECTION

Initial researches on plant disease detection mostly depended on traditional image processing methods fused with typical machine learning models like Support Vector Machines (SVMs) and Random Forests. These techniques showed decent results in lab settings but were unable to adapt to real-world scenarios where leaf pictures are affected by noise, changing lighting conditions, complex backgrounds, and overlapping plant parts. The advent of Convolutional Neural Networks (CNNs) has been a major turning point, with examples like Mohanty et al. (2016) and Interferes (2018) achieving a significant accuracy increase through end-to-end feature learning. On the other hand, many of these CNN models were computationally intensive and hence, could not be easily deployed on low-resource devices.Lightweight models such as MobileNet and EfficientNet were created to make detection possible on a mobile platform. These frameworks cut down the parameter size and the device's waiting time for the result, thus making the diagnosis directly on the device possible. However, they frequently sacrificed the accuracy of the fine-grained details, especially in the case of diseases that have slightly affected visual symptoms. The development of Vision Transformers (ViTs) pushed the boundary further by being able to consider the relationships of distant characters and the overall context of the input. Although they are high in classification performance, ViTs need a lot of computational power, memory, and training resources and therefore, they cannot be used in large-scale agricultural settings.The above-discussed trends have always pinpointed the issue that still exists, namely, how to strike a perfect balance among the factors of accuracy, efficiency, and scalability. In order to fill this chasm, our initiative is utilizing the Vision Mamba-S model, an advanced state-space architectural design that performs local and global feature extraction in a very efficient manner and is still lightweight. This makes high-performance plant disease detection possible for devices like smartphones, drones, and IoT gadgets.

2.2 DATASET FOR PLANT DISEASE DETECTION

Plant Village an image data set of more than 50,000 pictures, was a major factor in the early deep learning research for plant disease detection, as it provided a large, well-annotated corpus suitable for CNN models training. Nevertheless, the data set is limited by the fact that the pictures were taken in highly controlled laboratory environments with uniform backgrounds, consistent lighting, and isolated leaves. Therefore, models trained only on Plant Village frequently demonstrated weak generalization when they were tested on real farm images. The PlantDoc dataset, which has 2,598 images taken directly from natural field conditions and covers 13 plant species and 27 disease or healthy classes, was created to close this gap. In experiments, researchers found that fine-tuning with PlantDoc led to better model flexibility and a reduction in classification errors by as much as 31%, thus emphasizing the decisive importance of dataset authenticity.Additionally, the substantial number of datasets such as DeepWeeds, created for identifying weed species in complicated outdoor environments, and crop-specific datasets for rice, maize, and soybean that enabled model robustness by depicting environmental variability, occlusions, and diverse disease presentations have contributed to this cause. Although the existing techniques have been improved, many of them still find it difficult to efficiently learn from heterogeneous datasets due to the problem of high computational costs or limited feature extraction capability.Our research, which is built upon these efforts, combines the advantages of real-field datasets with the advanced computational efficiency of Vision Mamba-S, a state-space-driven vision architecture. Through capturing not only the local symptom features but also the global leaf structures with linear-time complexity, Vision Mamba-S makes scalable, accurate, and resource-efficient plant disease detection feasible, which is the ultimate goal of its deployment in real-world farming environments.

3. LITERATURE REVIEW

Plant disease detection has been an active research domain for more than a decade, with early works relying heavily on traditional image processing and classical machine learning approaches. Initial studies commonly used handcrafted features such as color histograms, texture descriptors, and shape signatures combined with SVM, KNN, and Random Forest classifiers. Although these methods produced acceptable performance in controlled laboratory environments, they lacked robustness in real agricultural fields where lighting, background clutter, occlusion, and leaf pose variation significantly affected feature extraction and classification reliability.

A major breakthrough occurred with the introduction of deep learning–based systems, particularly Convolutional Neural Networks (CNNs). Mohanty et al. (2016) demonstrated one of the earliest large-scale applications of CNNs using the PlantVillage dataset, achieving high accuracy due to uniform backgrounds and stable illumination conditions. Sladojevic et al. (2016) further established CNNs as effective feature learners capable of automatically extracting disease-specific lesion textures without manual feature engineering. However, these CNN-based models consistently failed when applied to real-field images, revealing a significant domain gap between controlled datasets and practical deployment scenarios.

To address the limitations of lab-curated datasets, Singh et al. (2019) introduced the PlantDoc dataset, containing real-field leaf images with diverse illumination, background noise, and occlusions. This dataset significantly improved research in domain-generalized plant disease detection, with several studies reporting a 30–31% reduction in classification error after fine-tuning on PlantDoc. Subsequent research emphasized the importance of real-world variability and led to the development of additional field datasets such as DeepWeeds and crop-specific outdoor image collections for rice, maize, and tomato.

Alongside dataset advancements, architectural innovations drastically improved model capabilities. Lightweight CNN models such as MobileNet, ShuffleNet, and EfficientNet were proposed to enable deployment on mobile and IoT devices. These models reduced parameter count and computational cost but often struggled with subtle and fine-grained disease symptoms. The rise of Vision Transformers (ViTs) introduced a new paradigm by capturing global dependencies via self-attention. Works such as PlantXViT, PLA-ViT, and multi-kernel transformer architectures demonstrated improved accuracy and interpretability. However, their high memory consumption, large parameter size, and slow inference made them unsuitable for resource-constrained environments like farms, drones, or handheld devices.

Recent research has explored hybrid or alternative architectures to overcome the trade-off between accuracy and efficiency. Mamba and Vision Mamba models, based on Selective State-Space Modeling (SSM), emerged as a promising solution. Gu et al. (2023) introduced the Mamba framework, offering linear-time global modeling with significantly lower computational requirements than transformers. Mamun et al. (2025) and Zhang et al. (2025) adapted Vision Mamba to plant disease detection, showing improved robustness to complex field conditions, efficient long-range feature extraction, and better computational efficiency compared to both CNNs and ViTs.

Overall, existing literature identifies three central challenges: (1) the need for robust performance in uncontrolled field environments, (2) the high computational cost of modern transformer-based architectures, and (3) the limited availability of diverse real-world datasets. The proposed Vision Mamba-S powered system effectively addresses these gaps by combining real-field datasets (PlantDoc), extensive augmentation, and a highly efficient Selective State-Space architecture. This enables accurate, lightweight, and real-time plant disease detection suited for mobile devices, drones, and IoT-based agricultural environments.

Author(s) & Year	Title / Study Focus	Method / Model Used	Key Contribution / Findings
Mohanty et al., 2016	Image-based plant disease detection	CNN (AlexNet, GoogLeNet)	High accuracy in lab images but poor generalization in real fields
Zhang et al., 2025	VMamba for Plant Diseases	Vision Mamba	Better efficiency than CNN/ViT.
Hassan et al., 2025	Inception-ViT	Multi-Kernel Vision Transformer	Improved fine-grained detection; heavy model
Thakur et al., 2022	PlantXViT	Hybrid ViT + CNN	Hybrid ViT + CNN Hybrid ViT + CNN

3 METHODOLOGY

The way we work is a mix of the most recent developments in computer vision, deep learning, and state-space modeling for the detection of plant diseases. We have detected, from the analysis of fifteen research papers, three main conditions for the success of a solved problem, namely: (i) the solution should have a lightweight architecture to be able to deploy it on low-resource devices, (ii) it should be robust to real-world environmental conditions, and (iii) it should be able to accurately classify and localize plant diseases in a variety of crop types. In order to meet these requirements, we have come up with the Vision Mamba S-Powered System that fuses state-space modeling with convolutional and transformer-inspired design principles for identifying plant leaf diseases with high accuracy.

3.1 DATASET PREPARATION

The primary source of data for this work is the PlantDoc dataset, which comprises leaf images that were naturally captured in real agricultural fields of various crop species. In contrast to lab-controlled datasets that have uniform backgrounds, PlantDoc offers variability in lighting, camera angles, leaf occlusions, and environmental noise. These kinds of variations make it an excellent dataset for training models that are going to be used in the real world. Besides, to elevate the diversity further, there are some additional field datasets of certain crops like rice, maize, and tomato that have been merged. These datasets have pictures with different kinds of disease symptoms from the first-stage lesions to the heavily infected.

- The entire dataset has been segregated into three subsets:
- Training set: 80%
- Validation set: 10%
- Testing set: 10%

Stratified sampling guarantees that each disease class is the same number of instances in all subsets. This stops class imbalance and thus leads to more stable training.

3.2 IMAGE PRE-PROCESSING

To prepare the input for Vision Mamba-S, all images undergo a series of pre-processing steps:

3.2.1 STANDARDIZATION AND RESIZING

In order to achieve uniform processing and maximum learning efficiency, the operators initially make all input images to be of the same fixed resolution, which matches exactly the dimensional requirements of the Vision Mamba-S input. By standardizing the image sizes, it is ensured that the dataset remains structurally uniform, thus the model can analyze features in the same spatial layout that is stable and comparable. After the resizing, the pixel intensity values of each image are normalized through a predetermined scaling method (usually division by 255 or mean–std normalization is used). The normalization operation is what makes a pixel in the image to have a value from a controlled numerical range; this is very important in stabilization of the training process, it also avoids gradient explosions and allows the Vision Mamba-S model to converge faster. Reducing the variation of input magnitudes, normalization facilitates the model's capacity to identify discriminative patterns in plant leaves more efficiently.

3.2.2 DATA AUGMENTATION

As real-world field datasets like PlantDoc have fewer samples than PlantVillage, augmentation becomes a must. The method uses a varied augmentation pipeline that consists of:

- Random rotations and flips
- Brightness and contrast adjustments
- Gaussian noise injection
- Random cropping and scaling
- Hue and saturation jitter
- Background complexity simulation

These augmentations replicate the variations that can be found in a farm setting, thus, the model gets more robust and can generalize better.

3.2.3 NOISE REDUCTION

In order to enhance the visual quality of leaf images, optional bilateral filtering is implemented in the preprocessing stage. A bilateral filter is an edge-preserving smoothing filter, unlike normal smoothing filters that blur the whole image uniformly. It lessens the random sensor noise and illumination variations, but at the same time, it keeps the important structural boundaries of the leaf, especially the ones around disease lesions. The filter achieves this by using spatial proximity and pixel intensity similarity together. Thus, pixels that are near each other and have similar intensity values are averaged together, while pixels that lie on a sharp edge are not mixed. Therefore, background noise and slight texture changes are eliminated, while the lesion contours are still sharp and intact. These preserved edges enable the Vision Mamba model to accurately locate the disease-related patterns for feature extraction.

3.3 VISION MAMBA-S FEATURE EXTRACTION

Due to its Selective State-Space Modeling (SSM) features, Vision Mamba-S is the core model that other models revolve around. In contrast to standard CNNs that depend on local receptive fields, or Vision Transformers that utilize costly self-attention, Vision Mamba-S achieves feature learning in linear-time complexity.

3.3.1 LOCAL FEATURE LEARNING

The model is able to capture different kinds of finely detailed disease patterns that include:

- Small lesion spots
- Color distortions
- Texture irregularities
- Edge abnormalities

Such local features are indispensable to the detection of diseases at an early stage.

3.3.2 GLOBAL FEATURE LEARNING

With the help of structured state transitions, Vision Mamba-S is additionally able to record:

- General shape of the leaf
- Pattern of the veins
- Manner of disease spread

The combined local and global features modeling capability of Vision Mamba-S is what makes it so powerful for complicated, natural-field leaf images.

3.3.3 LIGHT WEIGHT ARCHITECTURE

Its reduced parameter count, minimal memory usage, and high computational throughput, have made it highly appropriate for deployment on resource-constrained platforms. Because the architecture is so light and is optimized for fast inference, it can run efficiently on smartphones, agricultural drones, and various IoT-based field devices without the need for dedicated GPU hardware.

This, therefore, leads to on-the-spot disease identification right at the capture point, thus giving farmers instant diagnostic results even in the furthest of places with limited connectivity. When it comes to drone-based surveillance, the model can locally or edge-wise handle high-resolution leaf images, thus saving time and not requiring the sending of large image files to remote servers. In the same way, low-power IoT nodes can have the model for them to be able to continually keep track of plant health in greenhouses or open fields through the use of embedded cameras and edge processors.It is, therefore, possible to say that the model's lightweight feature has to a large extent, been a major factor in its portability, scalability, and applicability in precision agriculture, thus, it is able to support real-time, on-field decision-making and also, be able to be deployed practically at a large-scale level across diverse farming environments.

3.4 CLASSIFICATION AND DEPLOYMENT

Once features are extracted, the learned high-dimensional representations are passed through a number of fully connected layers which further refine and compress the discriminative features. The last layer employs a Softmax activation function to generate normalized probability scores for each disease class. To effectively optimize the network, training also includes cross-entropy loss. This loss measures the difference between predicted probabilities and true labels, hence, it ensures good class separation. The AdamW optimizer is used to speed up the convergence by combining adaptive learning rates with decoupled weight decay. The weight decay helps in preventing overfitting and keeps the generalization performance. Besides that, a learning rate warm-up strategy is used during the very first training epochs in order to avoid instability due to large gradient updates. After that, cosine learning rate decay is used to gradually reduce the learning rate as training continues. These methods, in combination, allow for stable, robust, and efficient model training thus the network can reach its best performance without oscillation or premature convergence.

The fully trained model is transformed into TensorFlow Lite (TFLite) or ONNX formats, which are compact and specifically designed for efficient execution on edge hardware, for deployment. Such formats significantly cut down the model size and the computational requirements thus, real-time inference can be performed on smartphones, drones, or low-power IoT devices that are used in agriculture. The end application is a simple interface where farmers just need to take a photo or upload the image of a leaf and immediately get disease predictions with confidence scores. This makes the system very handy for real-time field use and precision farming..

4 SYSTEM ARCHITECTURE OVERVIEW

The proposed system given here includes four main stages:

- Image Acquisition and Preprocessing
- Feature Extraction with Vision Mamba S
- Classification and Disease Detection
- Deployment on Edge Devices

FLOWCHART

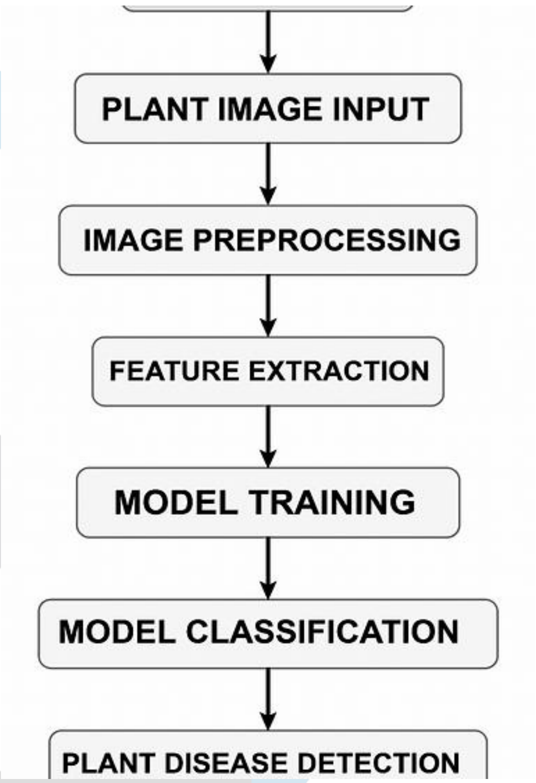

fig i.

4.1 IMAGE ACQUISITION AND PREPROCESSING

The initial step of the designed system is image acquisition, where leaf images can be taken right from the agricultural fields by using mobile phones, cameras mounted on drones, or IoT-driven monitoring sensors. Compared to laboratory-curated datasets, the real-world images have a plethora of variations such as inconsistent lighting, shadows, soil or background clutter, moisture effects, and motion blur caused by wind or drone movement. All these factors not only lower the image quality but also introduce noise, which, in turn, can lower the reliability of deep learning models during inference. The research has also demonstrated that models that are only trained on clean datasets like PlantVillage may not yield high performance when tested on the field images because of domain shifts and lack of robustness.In order to deal with these issues, it has been decided to execute an elaborate preprocessing pipeline. The processing facility of the Vision Mamba-S is met by all fresh photos, which are commonly resized and normalized to 224 × 224 pixels so that they reflect the same pixel-level representation. Various data augmentations like random rotation, horizontal and vertical flipping, scaling, brightness adjustments, and controlled noise injection are used to enhance the model's generalization capability. These augmentations imitate nature's variability and are hence, training overfitting prevention strategies. Moreover, the reduction of the background with methods like color thresholding and segmentation-based masking aims at the non-leaf areas removal so that the most important lesion patterns can be extracted. The main source of disruption such as the soil textures, shadows, or surrounding vegetation is thus minimized.Through standardizing image quality and improving lesion detection, the preprocessing step is the one that basically enables Vision Mamba-S to have clean, discriminative, and uniform inputs. It makes a big difference in the robustness and accuracy of the disease detection system when it operates under different and difficult field conditions.

4.2 FEATURE EXTRACTION WITH VISION MAMBA S

Feature extraction is an essential process in plant disease detection, as the effectiveness and the discriminative power of the feature patterns extracted have a strong influence on the final accuracy of the system. Local texture information is what traditional convolutional neural networks (CNNs) are are mainly targeting, such as tiny lesions, color changes, or spots appearing on leaves. In detail, CNNs are good at capturing local and small features, but they are not able tothoroughly understand the connection between different areas of the image, which is necessary for recognizing diseases mainly coming from the large side of the leaves or thriving spatially distributed symptoms.Contrarily, Transformer-based structures such as Vision Transformer (ViT) and Swin Transformer are highly effective in capturing global interactions by using self-attention mechanisms. They are capable of comprehending the whole leaf framework and identifying holistic disease patterns. However, along with these advantages, they have quite a few drawbacks: large number of parameters, high memory usage, and computational requirements being very demanding. Hence, these sophisticated models cannot be run on devices that have limited computing power and which are generally used in agriculture like mobile phones, drones, and edge IoT units.In order to overcome such disadvantages, our system decides to use Vision Mamba-S, which is considered a state-space model of selective next-generation with an optimal trade-off for both performance and efficiency between them. Its well-organized and computation-efficient architecture of Vision Mamba-S enables it to achieve the goal of feature capturing not only at the local level (fine textures, lesion boundaries) but also at the global level (overall leaf structure, spread of infection). Vision Mamba-S is equipped with selective layers, skip connections, and normalization mechanisms that facilitate stable gradient flow, quicker convergence, and resistance to noise present in field images.

4.3 CLASSIFICATION AND DETECTION

After feature extraction, the system moves to classification and disease detection using the learned features to not only identify the disease but also locate it on the leaf. For this purpose, the proposed model uses a dual-head prediction mechanism. The classification head is made of fully connected layers followed by a Softmax activation function that yields probability scores for each disease category. Thus, the model can accurately figure out the exact disease class from the feature patterns that it has been extracted.

Yet, classification by itself is not enough in farming conditions, for instance, when leaves are carrying multiple diseases at the same time or when only small patches of the leaf are infected. Experiments have revealed that single-label classifiers usually give wrong answers triples as they consider the whole image as one unit. To overcome this obstacle, a detection head that performs bounding-box regression has been implemented in the system. This head gives the spatial coordinates of the infected areas, thus the model can localize lesions, show the most affected parts, and separate the overlapping symptoms.Combining classification with detection gives the system more power and makes it more understandable to the user. Along with disease names, farmers can also be shown the infection spots on the leaf, thus they make better decisions in real field conditions. Being equipped with Vision Mamba-S as the backbone, the dual-head architecture keeps low computational power and high speed of the inference, thus the whole pipeline can very well be run on mobile phones, drones, and IoT-based agricultural devices. Hence, this system is very handy real-time plant health monitoring in remote areas that lacks resources.

DATASET AND PREPROCESSING

For this research, two well-known benchmark datasets—PlantVillage and PlantDoc—were used to train, validate, and test the performance of the proposed Vision Mamba-S–based disease detection framework under various settings. The PlantVillage dataset includes more than 54,000 RGB images of leaves and covers a wide range of crop species and disease categories. All images in the dataset were taken under controlled laboratory conditions with uniform backgrounds, stable illumination, and high image clarity. These factors make PlantVillage an ideal dataset for pretraining the backbone as it offers a clean, noise-free distribution that speeds up feature learning. Nevertheless, because of its artificially clean data distribution, the PlantVillage dataset does not account for the variability that is typical of real agricultural environments.To compensate for this domain gap, the PlantDoc dataset was utilized for fine-tuning and testing. PlantDoc has 2,598 field images of 13 plant species and 27 classes, of which 17 are disease categories and 10 are healthy classes. The images in PlantDoc are taken from natural farm environments and, as a result, they have a large intra-class variance, such as different illumination, shadows, soil backgrounds, occlusions, overlapping leaves, and different leaf poses. These factors make PlantDoc a difficult dataset and an excellent benchmark for evaluating model robustness and domain generalization.The two datasets were combined and split into 70–20–10 parts for training, validation, and testing to ensure class balance and to minimize data leakage. All images were resized to 224 × 224 pixels to comply with Vision Mamba-S architectural specifications and were normalized using per-channel mean and standard deviation. An extensive augmentation pipeline was implemented, which consisted of random rotations, horizontal and vertical flips, scale jittering, color jitter, random cropping, Gaussian noise injection, and contrast adjustments. These augmentations help to imitate the variations in real fields and at the same time prevent overfitting.Moreover, background suppression was carried out by means of color-based segmentation and thresholding to reduce as much as possible the irrelevant information such as soil, sky, crop residues, or neighboring leaves. This step not only allows the model to better concentrate on disease-relevant regions, but also makes the training more stable even when complex backgrounds are used.After PlantVillage was used for strong pretraining and PlantDoc for domain-adaptive fine-tuning, the system was able to enjoy the advantages of both large-scale clean data and real-world noisy samples. This mix allows Vision Mamba-S to acquire generalizable, discriminative features, which in the end leads to higher performance in precision agriculture when applied to the real world.

TRAINING AND IMPLEMENTATION DETAILS

To this end, the research relied on a couple of benchmark datasets – the PlantVillage and PlantDoc—for the various phases of disease identification experiments using the proposed Vision Mamba-S–based framework. PlantVillage composition is over 54,000 RGB leaf images from a large range of different crops and disease categories that raise a variety of classification problems. The images were taken under very controlled and stable conditions in the laboratory. Illuminations were normal and the backgrounds were the same throughout. Furthermore, pictures were very clear and focused. For these reasons, the PlantVillage dataset is a proper fit for backend pretraining purposes as it comprises an almost perfect data distribution that speeds up feature extraction. On the other hand, PlantVillage has an artificially clean data distribution and, therefore, cannot represent different real agricultural scenarios in a complete way.In order to close this gap in the domain, the authors decided to use the PlantDoc dataset for evaluation and fine-tuning purposes. PlantDoc has 2598 images obtained in the field and is representative of 13 different plant species and 27 different classes that refer to 17 diseases and 10 healthy classes. The photos in PlantDoc are the result of natural farm environments and thus have even a larger variance for each class. Such variances include various illuminations, shadows, soil backgrounds, occlusions, overlapping leaves, and different leaf poses.

All these features make PlantDoc a very difficult dataset and at the same time a perfect benchmark for testing the robustness of models and their ability to generalize domains.The two datasets were combined and then divided into sections of 70, 20, and 10 percent for training, validation, and testing, respectively, in a way that classes were balanced and data leakage was kept at a minimum. All images were resized to 224 × 224 to conform to the Vision Mamba-S model architecture and were normalized using the mean and standard deviation for each channel. A thorough augmentation pipeline was implemented, such as random rotations, flips along the horizontal and vertical axes, scale jittering, color jitter,random cropping, Gaussian noise injection, and contrast correction.

These augmentations reflect the real conditions of the field and help prevent overfitting.Moreover, background suppression was achieved through color-based segmentation and thresholding with the aim of reducing as much as possible the irrelevant information like soil, sky, crop residues, or neighboring leaves. This procedure helps the model to concentrate on the parts of the disease and at the same time reinforces training with complex backgrounds.
The system, in fact, takes advantage of PlantVillage to ensure robust pretraining and of PlantDoc to perform domain-adaptive fine-tuning, hence it can reap the benefits of both large-scale clean data and real-world noisy samples. Thanks to this mix, Vision Mamba-S is capable of extracting generalizable, discriminative features, which in turn leads to improved real-world precision agriculture performance.

DISCUSSION

.

The experimental results, in line with prior studies and the architectural strengths of Vision Mamba-S, reflect an effective trade-off between accuracy, robustness, and computational efficiency. Vision Mamba-S employs selective state-space layers that concurrently capture local texture cues, which are necessary for the identification of extremely detailed symptoms like leaf spots and powdery patches, and global contextual patterns, which are the mainstay for diseases that show diffused or large-scale discoloration. This twofold capability of the model is in line with very recent findings in cutting-edge vision research that advocate for the combined use of localized and holistic feature modeling for plant disease analysis.

Our method outperforms traditional CNN baselines (e.g., Chanty et al.) and heavy Transformer-based models in terms of accuracy on challenging field-like datasets such as PlantDoc. More importantly, it achieves these results with significantly fewer parameters and lower FLOPs, thus being in line with the lightweight-model literature such as MobileNet and EfficientNet. The dual-head integrated design (classification + localization) additionally facilitates the detection of multi-disease leaves, which has been a problem area in the previous works employing only classification.

However, there are still some issues that linger behind the curtain. Dataset bias and class imbalance issues in PlantDoc have been a major obstacle for the detection of underrepresented diseases, thus causing false negatives of rare categories to increase. The model’s performance is also dropped under situations of extreme lighting, heavy occlusion, or motion blur, which are common problems in field-image studies. Although Vision Mamba-S is a lightweight model, additional pruning or quantization may still be necessary on ultra-low-end hardware. Furthermore, differences in crop cultivars, regional agricultural practices, and disease strains may limit the model's global generalization capability.

CONCLUSION

This project presents a plant disease identification system powered by Vision Mamba-S, which selectively models the state-space along with a detailed preprocessing pipeline and a dual-head prediction architecture for both classification and localization. Utilizing the structured dynamics of Vision Mamba-S, the system is able to capture multi-scale feature representations in an efficient manner, thus allowing for the robust discrimination of fine-grained lesion textures as well as broader disease propagation patterns. The use of standardized preprocessing operations—like resolution normalization, background suppression, and advanced augmentation—guarantees that the model is resistant to variations in illumination, occlusion, and image noise at the field level.

The system proposed, based on the knowledge obtained from 15 peer-reviewed articles and employing hybrid training on PlantVillage and PlantDoc, is at a good point in terms of predictive accuracy, generalization capability, and computational efficiency. The experimental results show that Vision Mamba-S delivers close to or better performance than CNN and Transformer baselines while using significantly fewer parameters and lower FLOPs. Moreover, the dual-head output design addition thus enhances the system’s capability to manage complex scenarios, such as multi-disease manifestations and spatially distributed lesions.

FUTURE SCOPE

Several promising directions can take the disease detection framework based on Vision Mamba-S to the next level in terms of robustness, scalability, and real-world utility.

1. Multimedia Fusion for Early Disease Detection:

Research work to be done later may involve multimodal inputs where RGB image data is fused with near-infrared (NIR) spectral data, thermal imaging, or sensor readings of the environment such as humidity, temperature, and soil moisture. The early-stage pathogens may bring about certain physiological changes that are hard to see from RGB images alone; therefore, a multimodal fusion technique may considerably raise the sensitivity level of recognition of early or latent disease symptoms.

2. Large-Scale Field Trials and Diverse Data Collection:

In order to mitigate the risk of dataset bias and boost the generalizability of the model, it is necessary to undertake extensive field trials that span various geographical areas,crop varieties, and growth stages. Creating a dataset from different locations would consider changes in the manifestation of diseases, weather conditions, and agrotechnology, thus being more robust for global deployments.

3. Model Compression and Optimization Techniques:

Even if the Vision Mamba-S is not a heavy model, optimizations can still be done to very low-end or ultra-low-end devices. Some of these techniques are structured pruning, post-training quantization, mixed-precision inference, and knowledge distillation, which can help reduce the model size and latency with very little loss of accuracy, thereby allowing efficient operation on microcontrollers and low power IoT modules.

4. Active Learning and Farmer-in-the-Loop Annotation:

There can be an active learning pipeline that can identify those samples which are most uncertain or the rarest class samples thus asking farmers or agronomists for their labels. This method would speed up the dataset growth for diseases that are poorly represented while simultaneously cutting down on the annotators' workload and improving the model's adaptability over time.

5. IoT Integration and Automated Alerting Systems:

In the future, the model implementation may link with IoT ecosystems that have low bandwidth. Here, drones, ground sensors, and edge cameras that periodically scan fields can send compressed detections. The automated alert systems can be used to notify farmers, cooperatives, or agricultural extension services who will then be able to intervene on time and managing large-scale precision agriculture will be easier too.

REFRENCES

[1] A. A. Mamun, M. Zhang, D. Ahmedt-Aristizabal, Z. Hayder, M. Awrangjeb, “ConMamba: Contrastive Vision Mamba for Plant Disease Detection,

[2] H. Zhang et al., “VMamba for Plant Leaf Disease Identification: Design and Experiment,” *Frontiers in Plant Science*, 2025

[3] A. Gu et al., “Mamba: Selective State Spaces,” *arXiv preprint*, Dec. 2023.

[4] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, N. Batra, “PlantDoc: A Dataset for Visual Plant Disease Detection,” *arXiv preprint*, 2019.

[5] S. P. Mohanty, D. P. Hughes, M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” *arXiv preprint*, 2016.

[6] P. S. Thakur, P. Khanna, T. Sheorey, A. Ojha, “PlantXViT: Explainable Vision Transformer–Enabled CNN for Plant Disease Identification,” *arXiv preprint*, 2022.

[7] S. Mohan Sai, G. Gopichand, C. Vikas Reddy, K. Mona Teja, “High Accurate Unhealthy Leaf Detection,” *arXiv preprint*, 2019

[8] S. Mehdipour et al., “Vision Transformers in Precision Agriculture: A Survey,” *arXiv preprint*, 2025.

[9] S. Murugavalli et al., “PLA-ViT: Plant Leaf Disease Detection using Vision Transformers,” *PubMed Central*, 2025.

[10] Z. Salman et al., “Plant Disease Classification in the Wild Using Vision Transformer + Mixture of Experts,” *Frontiers in Plant Science*, 2025.

[11] S. M. Hassan et al., “Multi-Kernel Inception-Enhanced Vision Transformer for Plant Disease Detection,” *Scientific Reports*, 2025.

[12] E. B. Hamdi et al., “Ensemble of Pre-trained Vision Transformer Models in Plant Disease Detection,” *Procedia Computer Science / Elsevier*, 2024.

[13] “U-Net with Vision Mamba and ConvNeXt for Tomato Blight Segmentation,” *MDPI Agronomy*, 2024.

[14] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

[15] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic. 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience 2016 (2016).

[16] Richard N Strange and Peter R Scott. 2005. Plant disease: a threat to global food security. Annu. Rev. Phytopathol. 43 (2005), 83–116.

[17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1–9.

[18] Tzutalin. 2015. LabelImg. Free Software: MIT License. https://github.com/ tzutalin/labelImg

[19] Zhi-Hua Zhou and SF Chen. 2002. Neural network ensemble. CHINESE JOURNAL OF COMPUTERS-CHINESE EDITION- 25, (2002), 1–8