JOURNAL OF BASIC SCIENCE AND ENGINEERING

Integrated Deep Learning Pipeline for Tomato Plant Disease Recognition and Severity Estimation

Prakriti Saxena[1*], Aditi Sharma[2], Divyanshu Chauhan[3]

[1]Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India
[2]Institute of Engineering and Technology, Lucknow, India
[3]Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India

DOI: https://doie.org/10.10399/JBSE.2026858643

ABSTRACT

Diseases affecting tomato leaves can decrease the yield and quality of crops. Therefore, there is an urgent need for timely and accurate identification of tomato leaf diseases in order to implement precision agriculture management practices. This research reports the implementation of a comparative deep learning framework for monitoring tomato leaf diseases using three different architectural approaches: Fine-tuned Mobile Net V2, CNN+GRU hybrid model; and our new proposed multi-output CNN hybrid model. The study used a dataset from Plantcity which contained images of 11 different types of diseases affecting tomato leaves taken under numerous environmental conditions. The images used for the experiments were pre-processed by resizing, normalizing, augmenting, generating variations based on single input, and splitting the dataset into three parts - training, validation, and testing set aside for future experiments. The Multi-Output CNN Hybrid was identified as the primary model for evaluating the models due to its capacity to classify disease and assess severity and locate infected regions at once, by having a shared convolutional backbone with varied prediction heads for multiple predictions. This model uses a composite loss function that combines sparse categorical cross-entropy and MSE to allow for joint optimization and is able to make better predictions than any of the other models tested, with nearly 98.93% classification accuracy and a strong precision, recall, and F1 Score performance across all test reports. Multi-task learning was used in the creation of this architecture to improve feature sharing, reduce overfitting, and improve generalization. This architecture offers a reliable, scalable and interpretable approach to automated tomato disease monitoring, and is suitable for precision agriculture, as well as real-time crop health assessments.

KEYWORDS

Leaf Disease Detection, Convolutional Neural Network, YOLO, Tomato Leaf Diseases

1. INTRODUCTION

1.1 Background

Tomatoes are widely grown, and their growth is significantly affected by pathogens, including fungi, bacteria, and viruses, without adequate means of detection of these pathogens can result in growers being very limited in the management of their operations; thus, growers will not reach the full production potential and quantity of products that can be produced from each acre of tomato plants. The impact that these pathogens have on farmer incomes, food security or farmer incomes and food security due to the early recognition of pests is critical. In fact, growers typically can produce tomatoes over an average yield of two to three years, depending on the climate in each region, and the producers depend on accurate and timely identification of all diseases to help continue production over this time. As stated in the above paragraph, there are many different sources of pathogens that exist in commercial tomato crops, but ultimately the most significant source is through the environment [1].

1.2 Limitations of Conventional Disease Monitoring Methods

Agricultural professionals or farmers typically perform traditional visual inspections of crops to detect crop diseases. However, such an approach is subjective because of the variation of an inspector's ability to determine whether a crop has a disease depending on their prior experience and how they have visualized the environment.

Many of the early symptoms of disease are also hard to detect and thus may be missed; making it likely that not all of the crops suffering from a disease are detected. In addition, manual crop inspections require personnel and time, thus making them impractical for large-scale farming where extensive field coverage is required. Furthermore, the lack of standardized diagnostics for crop diseases results in inconsistent classification of plant diseases using traditional methods, making them inefficient for precision farming applications [2].

1.3 Motivation for Work

The continuous increase in the complexity of agriculture puts more pressure on the need to identify and react to disease outbreaks quickly due to advances in both AI-based monitoring systems and computer vision technology. Automated extraction of visual characteristics using computer vision can enable a faster and more accurate process for identifying diseases than if left to a human inspector. The ability to perform real-time diagnosis of plant diseases will enable farmers to rely less upon the skills and expertise of individual people assessing for the presence of disease before providing treatment; therefore, they can deploy diagnostic tools at greater distances across entire agricultural areas. The application of deep learning algorithms will improve upon the previous applications by developing a system that is able to define the characteristics of diseased and healthy plant images solely by using the data derived from the images themselves, resulting in better classification outcomes for agriculture, regardless of external factors such as climate.

1.4 Need for Multi-Task Learning Frameworks

Modern agricultural decision-making includes a variety of spatial and severity-related data, so modern disease diagnosis cannot involve just basic classification. Multi-task learning frameworks consider a range of tasks such as disease classification, bounding box-based localisation, and severity estimation, and bring them together into one structure for the purposes of implementing many of the tasks, such as whether a plant is diseased, where the affected part of the plant is located, and the level of risk associated with diagnosing that plant correctly. This facilitates the creation of more effective methods for disease intervention, because all three tasks may be completed together, improving both the efficiency of diagnostics and the number of computations needed for that task as compared with using separate models to complete all three tasks. These recent research studies provide evidence that the use of multi-task deep learning will increase accuracy and practical use in precision agriculture applications [3].

1.5 Problem Statement

The current systems for identifying crop diseases are primarily single-team classification systems that categorize diseases from leaf photographs. Although these systems achieve acceptable levels of success in classification accuracy, they do not consider how a disease will spread within the infected area, which ultimately makes them ineffective in precision agriculture where spatial information and severity of disease are essential to plan for disease intervention.

1.6 Aim and Objectives

This research aims to develop an advanced computer vision-based solution for automating the detection of diseases on tomato plants via deep learning. Beat the productivity of farmers growing tomatoes through providing them with an improved method of identifying the locations of disease-causing organisms as well as quantifying the severity of the disease on individual tomato plants i.e., identifying the amount of the disease affecting the plant using state of the art deep learning techniques. The goal of this project is to develop a robust image recognition method to accurately identify and classify tomato plant diseases, develop algorithms to compute the rectangular area of diseased portions of tomato plants, called bounding boxes, develop a model to estimate the severity of the disease and create a large-scale diagnostic tool that has a low or absent level of subjectivity as well as an ability to support precision agriculture through instantaneous or real-time and accurate assessments of disease-causing pest/disease agents.

1.7 Contribution of This Work

This investigation lays out a multi-task hybrid convolutional neural network architecture that can meet multiple plant disease analysis goals from one computational framework. The network uses common feature extraction layers as a basis for creating separate task-specific prediction branches to provide better efficient representation while lessening computation redundancy. The hybrid multi-task CNN creates a simultaneous prediction of the

Vol. 23, No. 1, (2026)
ISSN: 1005-0930

disease class, disease severity score, and regression for a bounding box which allows all analyses to be conducted from one input image. Disease classification allows a doctor to determine the pathological class, the disease severity gives an estimate of potential yield risk, and the localization provides an approximate area of damage which aids in enhancing the spatial evaluation. The entire hybrid multi-task CNN has been developed to be trained end to end in the TensorFlow deep learning framework, which allows the joint optimization of many objectives while coordinating the loss minimization of multiple image datasets.

2. LITERATURE REVIEW

2.1 Related works

Table 1. Comparative analysis of existing methodologies

Authors & Year	Methodology Used	Key Contribution	Performance
Ramanjot et al. (2023) [4]	Review of CNN, ML, and DL techniques for disease detection	Summarized 75 major studies on plant disease recognition and highlighted preprocessing, augmentation, and CNN architectures	Review-based comparative evaluation
Pacal et al. (2024) [5]	CNN, EfficientNet, MobileNet, Vision Transformers	Provided comprehensive benchmarking of recent DL models for disease classification	Accuracy range: 92–99% across reviewed studies
Foysal et al. (2024) [6]	CNN-based classification integrated with mobile application	Developed mobile-based disease recognition with 98% classification accuracy	98.2% classification accuracy
Kanakala & Ningappa (2025) [7]	CNN and LSTM hybrid architecture	Demonstrated strong CNN-based performance for multi-crop disease classification	97.8% classification accuracy
Journal of Industrial Information Integration (2024) [8]	ML, CNN, Transfer Learning	Discussed industrial-scale plant disease detection pipelines and transfer learning impact	Reported performance exceeding 95% in benchmark datasets
Sahu et al. (2025) [9]	Hybrid CNN + Attention + Mask R-CNN	Combined segmentation and disease detection with attention mechanisms	96.4% segmentation accuracy

2.2 Plant Disease Classification

CNNs, or convolutional neural networks, are quickly becoming one of the most popular methods of automatically identifying specific plant diseases using machine vision due to their capability of learning to understand visual representations’ hierarchical organization of information. For example, VGG, ResNet, and MobileNet have successfully extracted discriminative features from different plant disease images based on texture, colour, and lesion morphology using their specific architectural designs. VGG builds a deeper visual structure, ResNet introduced residual learning to circumvent the vanishing gradient problem and thus makes it possible to use deeper networks, while MobileNet permits the efficient use of extremely low-power devices at or near the farmer to support edge-based agricultural monitoring systems. In addition, another means by which to optimize CNN accuracy when using agricultural data in limited supply is the inclusion of ImageNet pre-trained weights, which enhances the potential benefits associated with transfer learning. As a result, the use of CNN models trained through transfer learning techniques to identify plant diseases can lead to improved efficiency and accuracy when identifying plant diseases across different characteristics in a diverse range of agricultural settings. [10][11].

2.3 Object Detection in Agriculture

Spatial localization technology allows for the identification of diseased areas on crops to aid in detecting disease spread throughout the entire field and in precision agriculture. Both YOLO and Faster R-CNN can be utilized for fast identification of lesions with YOLO producing real-time lesion identification via rapid inference types while Faster R-CNN provides for better accuracy by using only region proposal systems to locate lesions. Although both approaches have shown promise for identifying lesions accurately, they require an extensive amount of computational power and each use extensive annotated datasets for bounding box production. Current agricultural detection algorithms also have limitations within identifying and monitoring the full range of crop destruction because they are based only on locating lesions as opposed to creating a measurement system for establishing disease severity or diagnosis of a disease type [12].

2.4 Multi-Task Learning (MTL)

The use of Multi-Task Learning (MTL) has recently shown to be a good way to optimally accomplish related computer vision objectives together in a common neural network structure. It has been demonstrated in agricultural imaging that MTL models will use the same backbone network model for general feature extraction and different head models for specific task predictions like classification, localization, and regression to perform them at the same time. The use of common representations, or features, improves feature reuse and reduces overfitting due to the correlation between the different output representations that are used for a single representation but are being used for different learning objectives. They also remove computational redundancy by not using separate pipelines for training each task model. Research on MTL frameworks with multi-output CNNs has shown that MTL models perform significantly better with regards to generalization and efficacy of the trained model vs. single-task model will be able to perform with very small training datasets e.g., agriculture [13][14].

2.5 Research Gaps Identified

Although there has been progress made in the classification of plant diseases and identification of plant lesions, most existing systems are characterized by separate or unconnected frameworks that are generally focused on individual tasks. Additionally, few if any CNNs employ "learning" mechanisms for integrating the outputs from a total of two or more different algorithms based on CNNs for disease classification and lesion identification to provide an overall diagnostic assessment of a plant disease in the end user [10]. Also, the lack of or little relationship between severity rating systems and classification models leads to no capability to estimate yield losses from crops and, thereby, developing treatments for plant diseases. Additionally, most current lesion identification systems use multiple types of input and require an expansive amount of computer resources, presenting challenges regarding the scalability of any of the existing models to fit into agriculture [12]. Finally, there is little or no literature available to support using a single integrated multi-task learning-based architecture that combines classification, severity regression, and spatial localisation capabilities to monitor a complete plant disease in total. Therefore, the valid need exists for developing a truly integrated multi-task CNN-based framework that will provide complete monitoring of plant diseases within agriculture [10][12].

3. METHODOLOGY

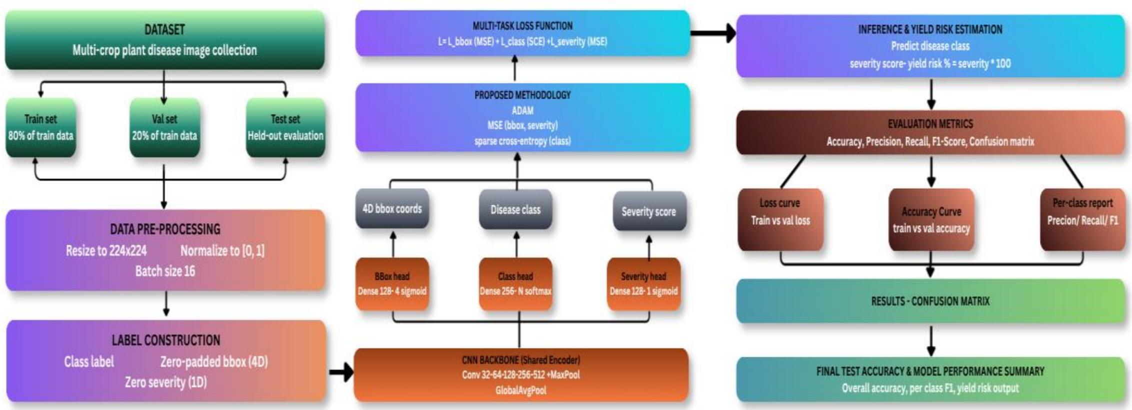

Fig. 1. End-to-End Multitask Deep Learning Framework

3.1 Dataset Collection

The dataset used in this research was taken from the 'PlantCity Dataset' [15]. This dataset contains agricultural images of diseased tomato leaves from a variety of environmental conditions. There are 11 categories of disease included in the dataset: fusarium wilt (tomato), spider mite, verticillium wilt, bacterial spot, early blight, healthy leaf, late blight, leaf curl, leaf miner, leaf mould and septoria leaf spot. The images of the diseased samples were arranged into a structured train and test data directory to facilitate supervised learning workflows. The dataset contains diverse classes, an even representation of diseases, different illumination conditions, different orientations of leaves and a realistic field-like appearance of diseased plants which support the creation of a robust model for training.

3.2 Data Pre-Processing

The dataset dividing process consisted of the Fine-Tuned MobileNetV2, CNN+GRU Hybrid, and Multi-Output CNN Hybrid models to make sure generalization is reliable and evaluations are unbiased. For the MobileNetV2 and Multi-Output CNN Hybrid models 80% of images were used for training and 20% of the images for validation based upon the internal dataset splitting procedure. The CNN+GRU Hybrid Model had a training directory to house all training images and a separate validation directory used for optimizing the models. Independent test data are kept for all models to ensure that an objective performance comparison can be made. The Multi-Output CNN Hybrid model was the primary focus of the development because it had an additional requirement needing to use a structured approach for pre-processing that would fit with using multiple heads output. All images were resized, normalised, batched, and label encoded to create a uniformed approach to their training.

3.3 Multi-task Data Pipeline

The Hybrid Multi-Output CUDA Deep Learning or MOCNN architecture was developed to leverage a specialized multitask data pipeline that enabled processing multiple prediction targets in parallel with only one Single Output prediction target model. The Modular MOCNN architecture provided an output for three different types of diagnostic results at once, rather than solely providing the disease classification, used by both MobileNetV2 and the Hybrid CNN+GRU models. Each sample image from the training set generated three outputs: a disease classification label, the bounding box location of the disease, and an estimation of how severe the disease's pathology is; both MobileNetV2 and the Hybrid CNN+GRU models only produced a single output for the disease classification. Furthermore, to ensure that images were able to be utilized as inputs into the MOCNN model, each image was passed through a pre-processing function to normalize the pixel intensity values in each sample image and create structured labels for each output type of the image sample including classification label, bounding box, severity. Specifically, the disease classification labels were encoded as sparse integers while the bounding box location and severity were encoded as numerical tensors. By providing such a streamlined method for multitasking on heterogeneous output types in a common pooled database, the MOCNN model has the potential to learn features from all three tasks namely classification, localization, and severity estimation synchronously.

3.4 Model Architecture

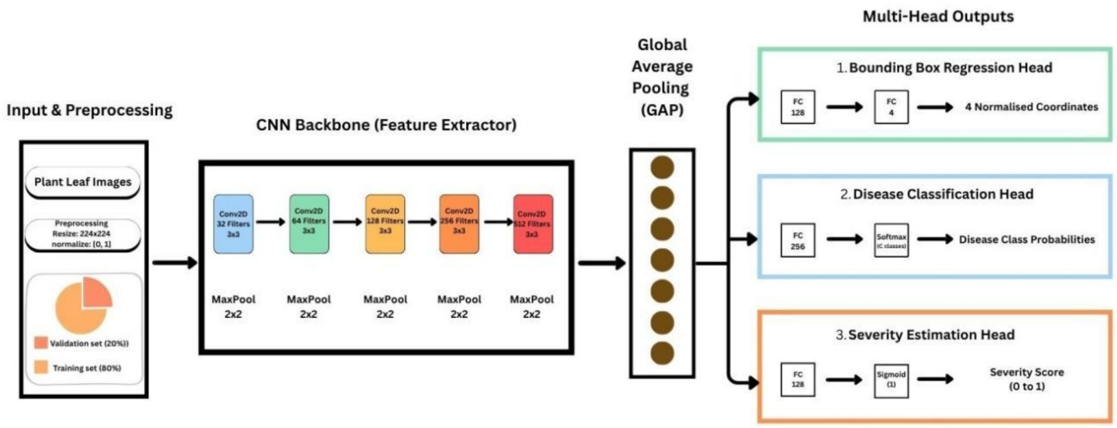

Fig 2. Multi-Head CNN Architecture pipeline

A deep convolutional neural network framework was used to extract hierarchical features from agricultural crop images and automate disease detection. The model starts with layers of convolutional blocks that progressively increase in number like 32, 64, 128, 256, and 512 to create a representation of both low-level and high-level textures/symptoms that identify a specific disease. Each layer contains 3x3 convolutional kernels followed by max-pooling layers that help reduce the amount of data by using various levels to perform this operation. After producing the last set of features from the original input images through the final or 5th convolutional block, the resulting feature map helps create a compact representation of image attributes using a global average pooling layer which will be used to build a multi-head architecture and make disease detection predictions based on the 3-headed prediction approach of the use of three-identical trained networks - one for the detection of disease, one for creating a bounding box on where the disease is located, and one for estimating severity so that all three networks can learn simultaneously using a common extracted feature set.

3.5 Loss Function Design

A composite loss strategy was developed to achieve an optimum Multi-Output CNN Hybrid model by learning multiple objectives simultaneously. For disease classification, we used sparse categorical cross-entropy as it fits naturally with integer-based class labels and multi-class prediction. For setting the bounding box and estimating the severity, we used the Mean Squared Error (MSE) loss function because the outputs required a continuous number for their predictions. All three losses were combined into one objective function in order to provide balanced optimization across outputs. The MobileNetV2 and CNN+GRU Hybrid models used single-task categorical loss functions that were only focused on determining classification. The multi-loss configuration provided better sharing of representations, and reduced redundant learning, while improving the overall generalization of the network.

3.6 Optimization Algorithms and Hyperparameter Tuning

Across all three of our architectures, we have used Adam Optimizer within TensorFlow to provide adaptive gradient updates for stable convergence. We have varied our batch size to be between 16 and 32 depending on the amount of model complexity and the amount of memory available to operate on. MobileNetV2 has been trained with transfer learning as well as fine-tuning; conversely, the CNN+GRU hybrid has been trained in an end-to-end manner with sequential feature extraction. The Multi-Output CNN hybrid has had the greatest amount of emphasis placed on it, as it has been trained over 20 epochs for the joint optimization of classification, localisation, and the severity estimation. Performance during validation of the models was constantly monitored during training through accuracy and loss curves to evaluate for convergence, generalizability, and overfitting.

4. EVALUATION METRICS

4.1 Classification Performance Analysis

The classification performance of three different models namely Fine-Tuned MobileNetV2, CNN+GRU Hybrid and Multi-Output CNN Hybrid was evaluated using standard supervised-learning metrics. The accuracy of a model refers to the number of correct classifications made relative to the total number of samples. Precision defines how accurate the model's predictions of positive disease are when compared to the actual amount of positive disease. Recall is a metric used to determine how many of the actual occurrences of disease are correctly predicted by the model and the F1-score provides an unbiased metric weighting the model's accuracy to both precision and recall. Each of the support values correspond to how many instances of each of the diseases occurred within each of the disease categories. The Multi-Output CNN Hybrid was the most highly supported as it was able to both classify and determine where on the image the disease occurred as well as how severe the disease was giving it the best accuracy with consistent predictions and the best ability to combine similar and differing tasks.

4.2 Confusion Matrix Analysis

The confusion matrix for Fine-Tuned MobileNetV2, CNN+GRU Hybrid and Multi-Output CNN Hybrid architectures provide class-wise prediction behaviour visualisation. For every matrix we can see the comparison of actual versus predicted class labels, with rows containing the ground-truth classes and columns containing the predicted outputs. Analysing the confusion matrices allowed us to identify trends in correct classifications as well as the specific classes of disease that have high levels of misclassification. Of particular interest were the visually similar tomato diseases which had many overlapping textures and colours of lesions. The Multi-Output CNN Hybrid was used as the principal focus due to its significantly improved ability to share features across classes. This enhanced inter-class confusion and provided superior separability of tomato disease classes compared to all other models tested.

5. PREDICTION AND DEPLOYMENT LOGIC

The architecture of the prediction and deployment framework supports inference across all model architectures previously fine-tuned using MobileNetV2, a CNN/GRU hybrid model and a multi-output CNN hybrid architecture to be primarily used for relating to the Multi-Output CNN architecture. All images processed for each respective model will be pre-processed in an identical way as was used to prepare images during training for all incoming plant images during model deployments. All images processed include resizing, normalizing, converting into tensors, and running through one of the trained models to produce a prediction. MobileNetV2 and CNN and GRU primarily produce outputs related to the classification of disease; whereas, the multi-output CNN hybrid model produces an output related to the disease classification/identification, disease state/severity and disease location all at the same time. By utilizing an integrated model approach to produce the various output criteria, the number of computationally redundant processes is minimized while permitting real time monitoring of agricultural crops through cloud computing platforms, edge devices, and mobile device-based deployment.

5.1 Single-Image Inference Pipeline

To start the single image inference process, a plant image can be taken with a camera or mobile device or uploaded from an existing dataset. One way all three architectures perform pre-processing is to ensure that the images comply with the configuration of the model they were trained on. MobileNetV2 and the Multi-output CNN Hybrid are resizing to 224 × 224 pixels, whereas the CNN + GRU Hybrid is resized to 150 x 150. To minimize the effects of light conditions when using the images, pixel intensities are normalised in the interval of [0,1]. The next step includes converting each image into tensor format and passing each of these on through the neural network. The Multi-Output CNN Hybrid will conduct a single occurrence of all inferences, while the output will be produced in multiples at the same time by using one Convolutional Feature Extractor and one multi-headed prediction structure.

5.2 Outputs

The deployed algorithm provides five distinct outputs because of a single inference operation. The classification branch provides a predicted disease class by taking the highest predicted probability score across all learned classes, while the severity branch provides a numerical percentage score that estimates the risk associated with infection to support prioritising treatment decisions. The localisation branch provides bounding box coordinates that delineate the approximate infected region of the leaf in the image. The use of this additional spatial information supports visualising the areas affected by disease as opposed to just providing a categorical output, thus enhancing interpretation of the model outputs. The combination of these three outputs provides an overall diagnostic framework to support decision-making, facilitate visualisation, and provide for automated agricultural monitoring in real-life deployment scenarios.

6. RESULTS AND DISCUSSION

Vol. 23, No. 1, (2026)
ISSN: 1005-0930

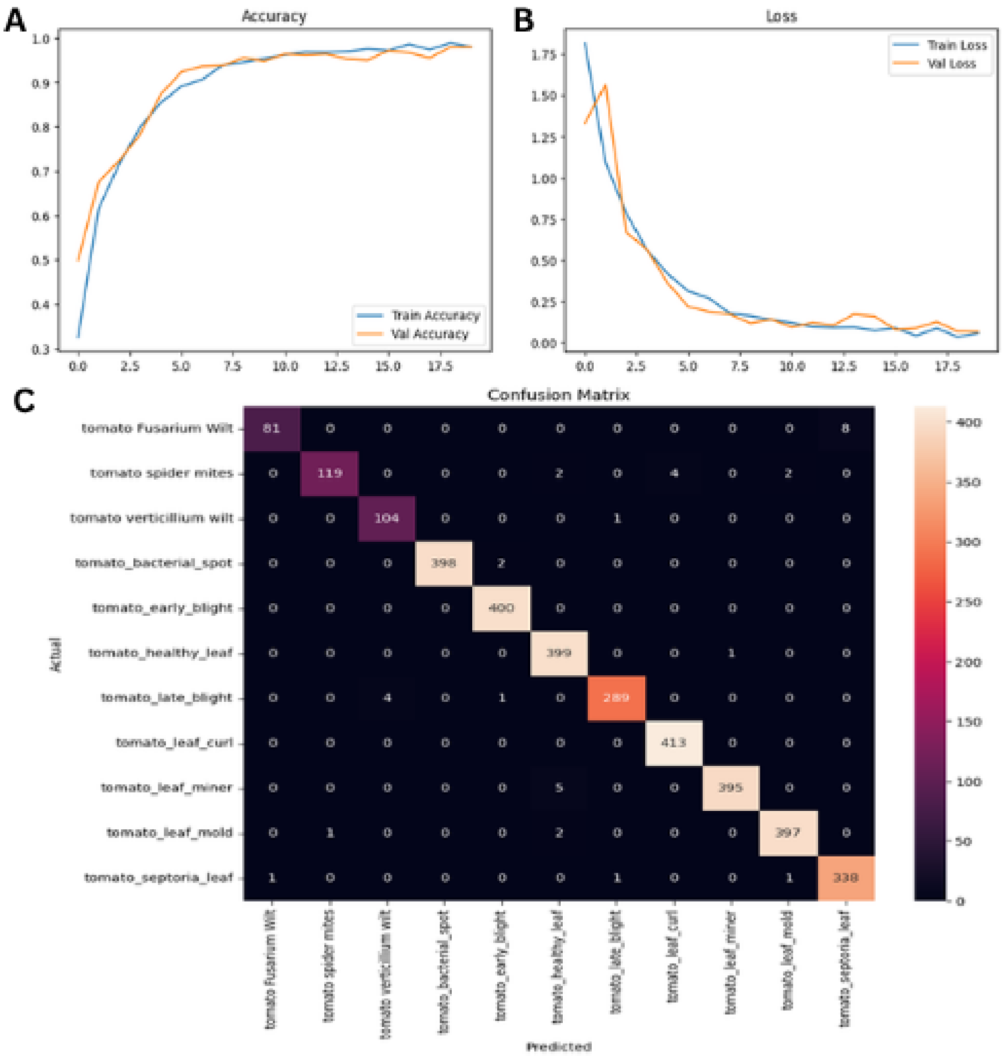

Fig 3. A) Accuracy Curve, B) Loss Curve, C) Confusing Matrix of Multi-Output CNN Hybrid Model

6.1 Results

For tomato disease diagnosis on the PlantCity Dataset, our Multi-Output Convolutional Neural Network (CNN) Hybrid Model achieved approximately 98.93% Class Accuracy, while providing predictions of Disease Severity and infected regions. Also, we used a single shared convolutional backbone to achieve discriminative lesion patterns, colour variations, and texture-based symptoms of disease. The Multi-task Learning improved feature re-use across multiple output data sets allowing for a stable convergence of model training and significantly reducing overfitting. Finally, this model performed very well in terms of precision, recall, and F1 scores across all classes of disease, which along with bounding box detection and severity scoring, provided for overall greater interpretability of the model and made it particularly appropriate for use in both Precision Agriculture and Automated Crop Monitoring applications.

6.2 Limitations

Though the performance of the Multi-Output CNN Hybrid Model is impressive, there are limitations in the model's ability to replicate real-world variability including lighting, background clutter, camera quality, and site conditions, as it used the PlantCity dataset for training. The initiation of the bounding box supervision was also not done through manual labeling, but rather using a weakly initialized method which could potentially decrease accuracy of localization. Due to the simultaneous use of multi-task optimization, higher computational resources were needed for the model. As well, there was class imbalance, and sometimes lesions in symptoms from different diseases were so visually similar that there were low confidence outcomes for predictions. In addition, there was synthetic regression targeting used for estimating degree of severity with no expert annotations available thus limiting usable clinical interpretation of estimations and requiring further validation before being used in large-scale use cases in agriculture.

6.3 Discussion

By combining three tasks - classification, localization, and severity estimation - into a single architecture called the Multi-Output CNN Hybrid, the Hybrid architecture has shown to outperform single-task architectures in terms of training time and model accuracy on the PlantCity dataset by learning shared representations which produce higher levels of disease discrimination while producing accurate predictions; reduced redundant computations when using multiple models as they all share the same feature representation; and providing shared feature representations that can provide generalised and reduce overfitting within the same disease category, thus enabling the model to have a more robust performance in terms of its ability to produce interpretable outputs to assist with decision-making for precision agriculture. Nevertheless, the quality of the dataset, depth of annotation, and type of environment where the model is deployed remain crucial to how well the model will perform and will determine if it can be deployed in full scale in real farm environments.

7. COMPARATIVE ANALYSIS

Table 2. Comparative analysis of models used

Models Used	Accuracy	Precision	Recall	F1-Score
Custom Multi output CNN Model	98.93	99.0	98.0	98.0
Transfer learning using MobileNetV2	98.43	99.0	98.0	99.0
Hybrid CNN+GRU	98.90	99.0	99.0	99.0

Comparative results obtained using Fine-Tuned MobileNetV2, CNN+GRU Hybrid and Multi-Output CNN Hybrid models show that deep learning techniques are effective for the recognition of tomato diseases, as all three architectures produced good results in the extraction of visual patterns related to diseases e.g., texture, lesion shape, colour variation from images of plants. The MobileNetV2 architecture used transfer learning to get benefits from pretrained convolutional features, while the CNN+GRU Hybrid architecture was able to combine the features from the spatial extraction with the sequential extraction to learn the feature patterns associated with the plants and their corresponding diseases. In contrast to the other two architectures, the Multi-Output CNN Hybrid architecture performed the best because its model was combined; therefore, all tasks could be learned together i.e., classification task, severity estimation task, and localization task, thus leading to better feature learning and improving generalisation. Additionally, shared feature learning via multiple classification tasks reduced model overfitting and improved the robustness of the model across all diseases and across different environmental conditions.

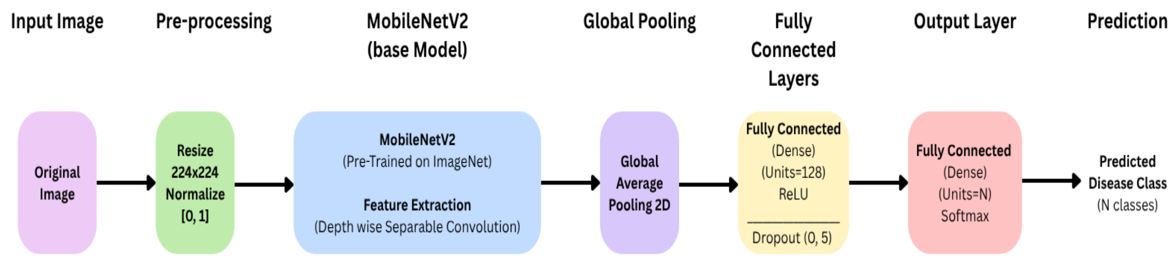

Fig 4. Fine-Tuned MobileNetV2 Architecture

Vol. 23, No. 1, (2026)
ISSN: 1005-0930

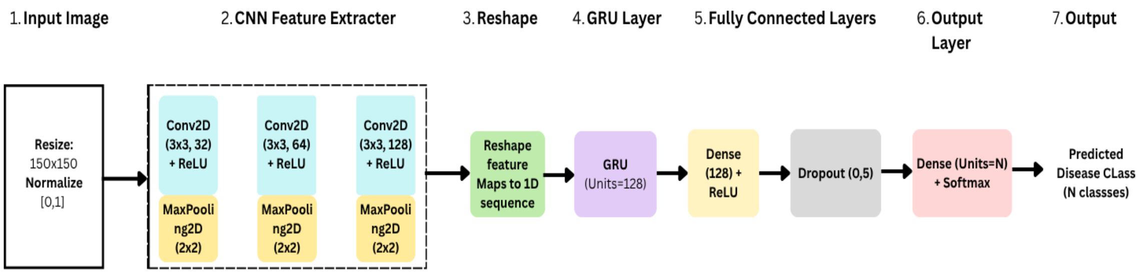

Fig 5. CNN+GRU Hybrid Architecture

Comparatively, all other models, such as MobileNetV2, and the CNN+GRU, had only been designed for supports classification purposes, while in the case of the proposed model a single inference could produce multi-diagnostic outputs during the same run, specifically the prediction of severity of disease by location. Additionally, where all other models relied upon fully annotated datasets or bounding boxes, the Weak Bounding Box Supervision Model reduced those requirements for requiring full annotations while still providing for proper localization outputs. All models exhibited performance, well across all experimental conditions of variation, weather and image quality to a degree where they would have variable performance measurements due to data set diversity and class imbalance, however the multi-task learning strategies used in the Hybrid Model improved the adaptability to each production method as well as supporting dependable/sound agriculturally based decision-making.

8. CONCLUSION

An integrated deep learning system has been proposed to monitor tomato leaf disease in one system by using multiple analytics to accomplish all the data processing in one combined computational architecture. The model processes tomato leaf images through shared feature extraction layers to complete disease classification, localization, and severity estimation all at the same time. By doing this, the multi-tasking nature of the model supports learning efficiency by allowing for correlated or similar representations to be shared across all prediction tasks, which also minimizes redundant calculations, thus reducing overall computational overhead. The unified computational framework is set apart from traditional methods where there are separate models for disease detection and has less complexity for deployment, lower hardware requirements, and increased ease of training and maintenance. This architecture supports accurate disease interpretation, spatially identifying diseased areas of the plant, and quantitatively estimating disease severity, which provide complete assessments of crop health. The scalable design of the architecture allows it to be applicable for precision agriculture where early disease detection is important through automated and rapid monitoring. Additionally, integration with smart farming technology such as Internet of Things or IoT enabled systems and edge devices further improves the decision-making support provided by the system. In sum, this framework represents advances in intelligent agricultural management through development of a more reliable, accurate, and efficient data-based plant disease monitoring system.

9. FUTURE WORKS

Future enhancements of the intended architecture will implement YOLO-based detection of different objects for accurate bounding box learning as well as the spatial position of areas affected by disease on a leaf level. This will allow for accurate analysis of lesions, multi-object detection, and an overall accurate identification of the types of lesions present. Conversion of YOLO-trained data to transfer learning models such as EfficientNet and ResNet to produce feature extraction to increase classification performance and reduce the potential for overfitting using prior hierarchical representations may also be feasible. Furthermore, the architecture can be deployed on edge-compliant IoT devices to afford the capability of conducting decentralized low-latency crop monitoring, thereby providing a means for real-time disease detection in agriculture in environments with limited resources. The use of drones will also allow for large-scale field monitoring and high-resolution aerial imagery for precision farming. Finally, the use of recurrent or time-series learning, and models lends itself to formal temporal analyses to monitor disease progression over time for identification of infection level, timely intervention measures, and long-term evaluation of crop quality.

REFERENCES

[1] Nawaz, M., Nazir, T., Javed, A., et al. (2022). A robust deep learning approach for tomato plant leaf disease localization and classification. Scientific Reports, 12, 18568. https://doi.org/10.1038/s41598-022-21498-5
[2] Upadhyay, A., Chandel, N. S., Singh, K. P., et al. (2025). Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture. Artificial Intelligence Review, 58, 92. https://doi.org/10.1007/s10462-024-11100-x
[3] Liu, B., Wei, S., Zhang, F., Guo, N., Fan, H., & Yao, W. (2024). Tomato leaf disease recognition based on multi-task distillation learning. Frontiers in Plant Science, 14, 1330527. https://doi.org/10.3389/fpls.2023.1330527
[4] Ramanjot, Mittal, U., Wadhawan, A., Singla, J., Jhanjhi, N. Z., Ghoniem, R. M., Ray, S. K., & Abdelmaboud, A. (2023). Plant disease detection and classification: A systematic literature review. Sensors, 23(10), 4769. https://doi.org/10.3390/s23104769
[5] Pacal, I., Kunduracioglu, I., Alma, M. H., et al. (2024). A systematic review of deep learning techniques for plant diseases. Artificial Intelligence Review, 57, 304. https://doi.org/10.1007/s10462-024-10944-7
[6] Foysal, M. A. H., Ahmed, F., & Haque, M. Z. (2024). Multi-class plant leaf disease detection: A CNN-based approach with mobile app integration. arXiv. https://doi.org/10.48550/arXiv.2408.15289
[7] Kanakala, S., & Ningappa, S. (2025). Detection and classification of diseases in multi-crop leaves using LSTM and CNN models. arXiv. https://doi.org/10.48550/arXiv.2505.00741
[8] Sajitha, P., Andrushia, A. D., Anand, N., & Naser, M. Z. (2024). A review on machine learning and deep learning image-based plant disease classification for industrial farming systems. Journal of Industrial Information Integration, 38, 100572. https://doi.org/10.1016/j.jii.2024.100572
[9] Sahu, K., Tiwari, S., Singh, M. K., et al. (2025). Plant disease detection using a hybrid dilated CNN with attention mechanisms and optimized Mask R-CNN segmentation. Scientific Reports, 15, 42008. https://doi.org/10.1038/s41598-025-26192-w
[10] Saleem, M. H., Potgieter, J., & Arif, K. M. (2020). Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers. Plants, 9(10), 1319. https://doi.org/10.3390/plants9101319
[11] Lu, J., Tan, L., & Jiang, H. (2021). Review on convolutional neural network (CNN) applied to plant leaf disease classification. Agriculture, 11(8), 707. https://doi.org/10.3390/agriculture11080707
[12] Abade, A. S., Ferreira, P. A., & Vidal, F. B. (2020). Plant diseases recognition on images using convolutional neural networks: A systematic review. arXiv. https://doi.org/10.48550/arXiv.2009.04365
[13] Yao, J., Tran, S. N., Garg, S., & Sawyer, S. (2023). Deep learning for plant identification and disease classification from leaf images: multi-prediction approaches. arXiv. https://doi.org/10.48550/arXiv.2310.16273
[14] Hossen, M. I., Awrangjeb, M., Pan, S., et al. (2025). Transfer learning in agriculture: A review. Artificial Intelligence Review, 58, 97. https://doi.org/10.1007/s10462-024-11081-x
[15] Khan, M. S., Nisa, K., Ahmad, I., Zubair, M., & Alshammari, K. (2025). PlantCity: A comprehensive image based on multi crop leaves in Pakistan. Data in Brief, 63, 112130. https://doi.org/10.1016/j.dib.2025.112130