International Journal of Artificial Intelligence and Mathematical Sciences
June 2025 Vo4. Issue No. 01 pp. xx-xx
e-ISSN: 2663-7073
DOI: https://doi.org/10.58921/ijaims.v4i1.xx

Deep Learning for Precision Agriculture: Tomato Leaf Disease Diagnosis Using Convolutional Neural Networks

Sajjad Ahmed[1], Muhammed Juman Jhatial[2], Inam Ali[1], Abdul Rasheed[1], Abdullah Niaz[1], Saba Akbar[1]
[1]Department of Computer Science, The University of Larkano, (TUL), Larkana, Sindh, Pakistan
[2]Department of Computer Science, Shah Abdul Latif University, Khairpur Mir's, Sindh, Pakistan

E-mail: dr.sajjad@uolrk.edu.pk, j u m a n @uolrk.edu.pk, inamalichandio08@gmail.com, arasheedchandio@uolrk.edu.pk, abdullahniazshaikh@uolrk.edu.pk, saba@uolrk.edu.pk,

Received: 7-04-2025; Accepted: 21-5-2025; Published: 1-7-2025

Abstract
One of the large crops in Pakistan is tomato, which is cultivated on a significant area and yields approximately 0.31 million tons every year. It is a major income earner to agricultural communities, and it is a key contributor to nutritional requirements. Though not underestimated, diseases and unfriendly environmental conditions often lower tomatoes yields particularly at the initial stages of growth. The current research proposes a Convolutional Neural Network (CNN) architectural model of detecting and classifying significant tomato leaf diseases. The methodology involves taking of images, preparation of data, preprocessing, training of CNN, and measurement of performance by conventional accuracy and precision measures. Empirical evidence has shown that the model had a training and validation accuracies of about 90 and 85-88 respectively, showing its efficiency in the detection of diseased and healthy leaves. This suggested framework provides a feasible input to the improvement of the disease management system to enhance sustainable tomato production and to empower the lives of farmers in Pakistan.

Keywords: Tomato disease detection; Convolutional Neural Network; Machine learning; Precision agriculture; Image classification.

1. INTRODUCTION

Tomato is an economically valuable and wide-spread crop all over the world and in Pakistan it is one of the most popular cash crops especially in the provinces of Sindh and Punjab. It is an essential ingredient in fresh consumption, processed food and culinary preparation, whether it comes to household nutritional value or the agricultural economy of a country.

Email: dr.sajjad@uolrk.edu.pk

Nevertheless, the production of tomatoes is recurrently confronted by the diseases of plants consisting of fungi, bacteria, and viruses that significantly diminish the yields and deteriorate the quality of fruits. These losses translate to large economical drawbacks to farmers and disruptions of local food chain supplies [1]. Late Blight, Bacterial Wilt and Tomato Yellow Leaf Curl Virus (TYLCV) are well-known threats in Pakistan, but in other countries all over the world, Early Blight and Fusarium Wilt are also destructive [2]. The exogenous factors that typically enhance these outbreaks include high humidity, low soil fertility, indiscriminate application pesticides and climate variability all of which make conditions conducive to pathogens [3].

A variety of physiological conditions, including healthy tissue or severe infection, are reflected by tomato leaves. The visible effects of the lesions and discoloration in bacteria spot, early, late blight, leaf mold and septoria leaf spot impair photosynthesis as shown in Figure 1. Likewise, attacks, like spider mite and target spot, also progressively undermine the vigor of the plants, whereas viral infections like TYLCV and tomato mosaic virus alter the leaves morphology, causing them to curl, mottle, and grow to a limited extent. Conversely, healthy leaves are a symbol of structural stability and optimum levels of metabolic activity, which guarantee normal growth of the plant and production of fruits. All of these conditions are the primary categories of tomato health and disease, which is why it is essential to properly identify them to manage crops.

To handle these difficulties, this study examines how Convolutional Neural Network (CNN)-based method can be used to classify tomato leaf diseases. Through the ability to use images to analyze ten different classes, the framework will enhance the accuracy of disease recognition, relative to the conventional observation techniques. These developments have a potential to minimize yield losses, help sustain agriculture, and help in food security in Pakistan and other countries [4].

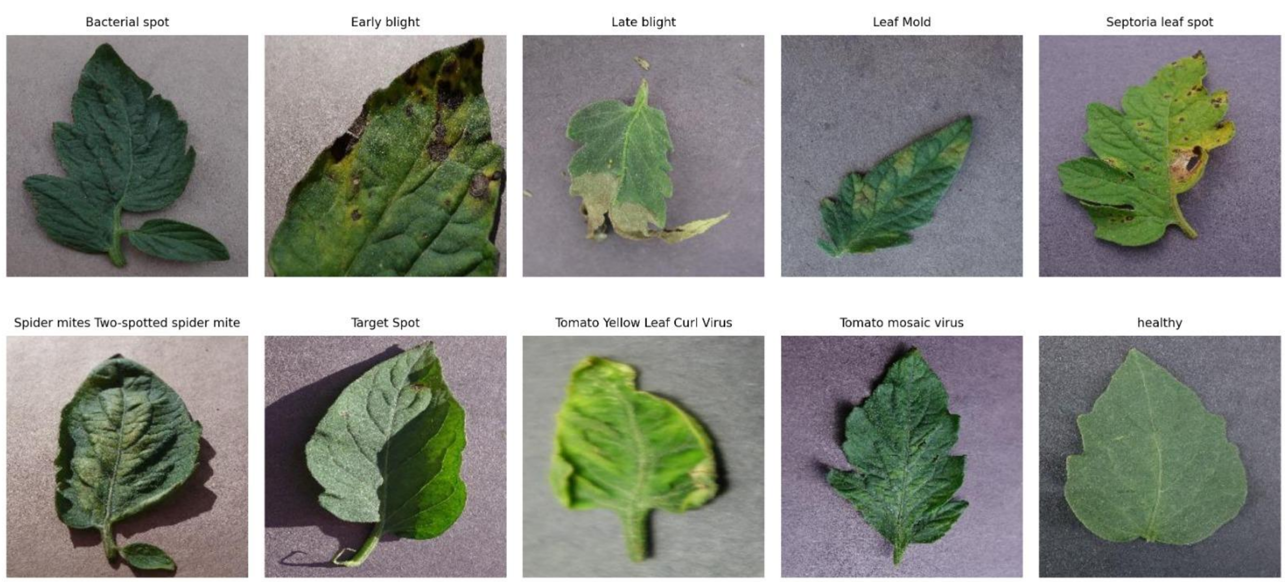

Fig. 1: Images of 10 classes of dataset

Pakistan is experiencing severe losses in the quantity and quality of tomatoes that are due to various fungal, bacterial, and virus diseases. Key outbreaks consist of Early Blight (Alternaria solani) which causes defoliation and poor fruit growth in warm and humid conditions and Late Blight (Phytophthora infestans) one of the most damaging foliar diseases resulting in the destruction of crops in a few days under optimal climatic conditions. Bacterial Wilt (Ralstonia solanacearum), resulting in fast wilting and not disappearing in soil, viral infections including Tomato Yellow Leaf Curl Virus (TYLCV), transmitted by whiteflies, and Tomato Mosaic Virus (ToMV), spread by contaminated seeds, soil, and tools are other significant dangers. The diseases are common in Punjab and Sindh, which are the major tomato growing areas in the country [4]-[5]. Conventional methods of Pakistan comprise the crop rotation, intercropping of the fields, replacement of the infected plants, and application of organic treatments like neem extracts. Although some control is offered through these methods, they are unable to handle large scale outbreaks. The farmers do not have a wide access to modern diagnostic tools, and therefore, the experience of the disease and its visual inspection are mostly used, which postpones effective intervention. This has given rise to a strong urgency to develop automated and dependable detection procedures that will allow an early diagnosis and effective administration of the process to ensure the sustainability of tomato production and minimize economic losses [6]-[7].

In Pakistan, tomato growers, especially in the rural regions, are greatly affected by the fact that proper diagnosis and management of crop diseases cannot be efficiently managed because of inaccessibility to testing centers, agricultural experts, and low-cost resources. Majority of farmers depend on self-diagnosis or anecdotal practices that are not always accurate, expensive, and cause additional economic damage. Farming communities have little awareness, technical skills, scientific assistance to help in early detection of diseases and putting up proper control strategies. Consequently, production of tomatoes is affected by low yields and quality resulting in a direct threat to the livelihoods and food security of farmers.

2. LITERATURE REVIEW

Tomato is a food security crop in Pakistan, but its yield is drastically limited by disasters of nature and the outbreak of frequent diseases. This paper surveys the current approaches to management and outlines the necessity of superior, data-driven management strategies to improve the yield and sustainability. Jhatial et al. [1] examined the contribution of rice to the economy of Pakistan and highlighted the effects of significant leaf diseases that include bacterial blight, brown spot, blast and tungro. In their research, they used the YOLOv5 deep learning architecture to identify rice leaf disease, and the results showed that it was more accurate and able to detect the disease than its predecessors (YOLOv3 and YOLOv4). The model, trained on a dataset of 400 infected rice leaf images, taken on Kaggle and trained during 100 epochs in Google

Colab, showed good experimental results, with a precision, recall, and mAP value of 1.00, 0.94, and 0.62, respectively. The results of this work demonstrate that YOLOv5 can be a trusted disease detector, and it provides a starting point in improving agricultural disease detection systems. Hoque et al. [2] highlighted the shortcomings of the old-fashioned disease detection techniques in the agricultural sector including manual inspection and microscopic analysis that are more labor-intensive and prone to errors. To address these issues, they suggested a multi-stage pipeline to detect tomato leaf disease including image enhancement, region of interest (ROI) detection, K-means clustering, and a combination of feature extraction. They have a framework that integrates discrete wavelet transform (DWT), gray level co-occurrence matrix (GLCM) and principal component analysis (PCA) in order to allow features of both the spatial and frequency domain. The study based on a dataset of 9,014 tomato leaf images with 4 most common diseases: curl virus, bacterial spot, late blight, and Septoria spot, obtained a classification accuracy of 99.97%. The findings emphasize the possibility of using clustering as well as hybrid feature extraction along with neural network models in accurately and scalably diagnosing tomato disease. Assaduzzaman et al. [3] created XSE-TomatoNet, a powerful tomato leaf disease classification framework that is based on EfficientNetB0 with Squeeze-and-Excitation (SE) blocks and multi-scale feature fusion. They paid attention to the problem of identifying minor differences between disease classes and handled it with the help of enhancing multi-scale features and combining them with the Global Average Pooling. The model had an accuracy of 99.11% with 99% precision and recall, which is better than the traditional architectures of MobileNet (87.44) and VGG19 (95.50). The high generalization of the model was validated by the high level of cross-validation (10-fold) and the model reached a high training accuracy (99.41 per cent) and validation accuracy (98.88 per cent). Besides it, interpretability tools, including LIME, SHAP, Grad-CAM, and Grad-CAM++, were incorporated to give a better understanding of the decision-making process, and emphasize disease-relevant areas on tomato leaf images. The research ended with the implementation of the model on a web-based platform to be put into practice in the agricultural sector.

Ahmed et al. [4] addressed the problem of deep learning application to public health compliance during the COVID-19 pandemic through mask-wearing detection. The experiment trained AlexNet on the Masked Face-Net dataset, with transfer learning, which is then fine-tuned on a large and varied set of face images with accurate and inaccurate use of masks. Their model has proven to be effective in classifying mask-wearing compliance providing a valuable instrument that can help reduce airborne disease spread in open areas. This paper will emphasize the flexibility of deep neural networks in practice outside the agricultural field, which validates the ability of AI-based solutions to suit social and health-related issues. Jhatial et al. [5] offered an IoT-based approach to managing nursery using cloud-based infection of data collection of real-time data on the environment in rural Sindh, Pakistan, to improve nursery growth, and the data were analyzed on the Azure cloud system, and the results were visualized using Power BI. A built-in notification system also assisted in making decisions in time notifying the farmers and agricultural experts via email. It was proven that the trial positively influenced the results of precision, efficiency, and yield, which proved that the IoT-based solutions had the potential to encourage sustainability and lower the cost of production in the rice nursery practice.

3. RESEARCH METHODOLOGY

Figure 2 shows flow chart of the research methodology, which includes data preprocessing, CNN model design and performance evaluation as its major components. Preprocessing of data includes resizing, normalization, and augmentation to increase the generalization of the model. The CNN model is conditioned to detect diseases on tomato leaves effectively. Accuracy, precision, and other evaluation indicators are used to measure model performance to achieve strong classification outcomes [8]-[11].

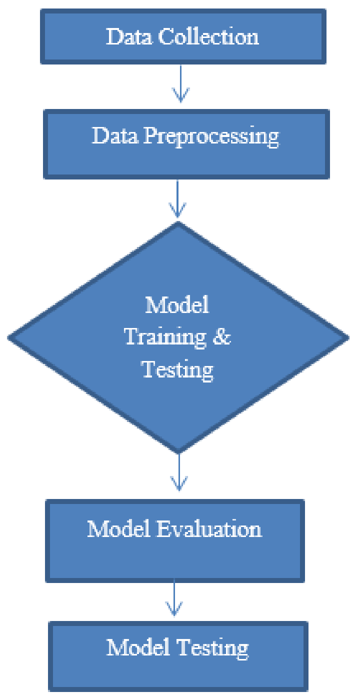

Fig. 2: Research methodology flow chart

4. DATA COLLECTION

The data used in this research was obtained in kaggle site, which included high-resolution photos of normal and infected tomato leaves in various stages of growth. The table 1 provides the dataset of the tomato leaf images in this paper where all of the ten classes are represented with their respective training and validation samples. The dataset contains a total of 10,000 training images and 1,000 validation images in all the categories.

Table 1: **Dataset Classes, Train and Validation images**

Class	Train Images	Val Images
Tomato___Bacterial_spot	1000	100
Tomato___Early_blight	1000	100
Tomato___Late_blight	1000	100
Tomato___Leaf_Mold	1000	100
Tomato___Septoria_leaf_spot	1000	100
Tomato___Spider_mites_Two-spotted_spider_mite	1000	100
Tomato___Target_Spot	1000	100
Tomato___Tomato_Yellow_Leaf_Curl_Virus	1000	100
Tomato___Tomato_mosaic_virus	1000	100
Tomato___healthy	1000	100
TOTAL	10000	1000

4.1. DATA SPLITTING

Figure 3 shows the dataset splitting process, where 10,000 images dataset was divided into a training and validation set in the ratio of 80:20. The division led to the training of 8,000 images with validation of 2,000 images, which guaranteed the balanced assessment of the model performance. The data was divided into a training and validation set that guaranteed an even assessment of the learning and accuracy of the model.

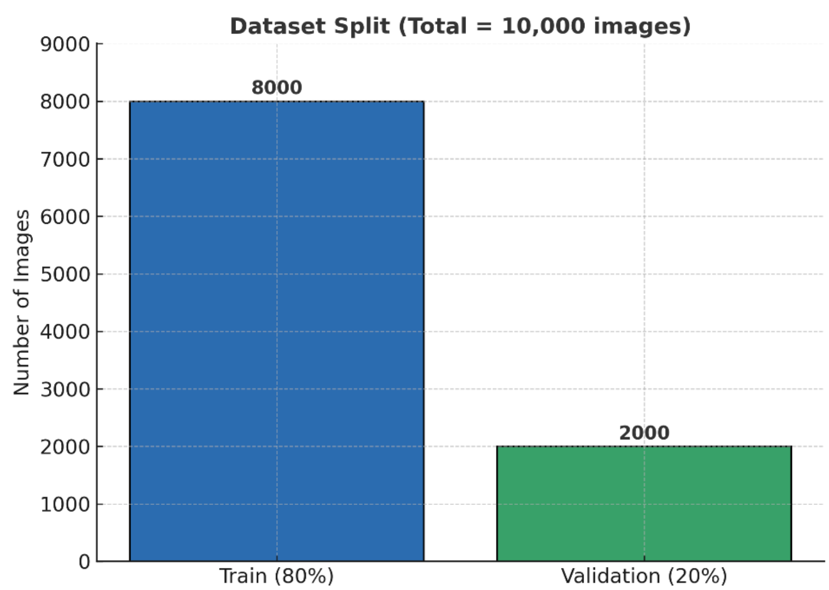

Fig. 3: Dataset splitting

4.2. CNN MODEL SUMMARY AND TOTAL PARAMETERS

Figure 4 shows the CNN model architecture, comprising several convolutional, pooling, and dense layers, with a total 111,050 trainable parameters. This is a lightweight design that guarantees effective training and proper classification of tomato leaf diseases.

CNN Classifier Architecture Summary

Layer (type)	Output Shape	Parameters
Input Layer (InputLayer)	(None, 256, 256, 3)	0
Sequential (Sequential)	(None, 256, 256, 3)	0
Rescaling (Rescaling)	(None, 256, 256, 3)	0
Conv2D (Conv2D)	(None, 256, 256, 32)	896
MaxPooling2D (MaxPooling2D)	(None, 128, 128, 32)	0
Conv2D (Conv2D)	(None, 128, 128, 64)	18,496
MaxPooling2D (MaxPooling2D)	(None, 64, 64, 64)	0
Conv2D (Conv2D)	(None, 64, 64, 128)	73,856
MaxPooling2D (MaxPooling2D)	(None, 32, 32, 128)	0
GlobalAveragePooling2D	(None, 128)	0
Dropout (Dropout)	(None, 128)	0
Dense (Dense)	(None, 128)	16,512
Dense (Dense)	(None, 10)	1,290

Fig. 4: CNN Classifier Summary

4.3. MODEL PARAMETERS SUMMARY

Figure 5 shows the model parameters summary, showing 111 050 trainable and 0 non-trainable parameters meaning that no weights are non-contributory to the learning process.

Model Parameters Summary

Description	Count
Total Parameters	111,050
Trainable Parameters	111,050
Non-Trainable Parameters	0

Fig. 5: CNN Model parameters Summary

5. RESULTS

The CNN model proposed attained accuracy of about 90 percent and a precision of more than 0.9 which indicates a strong performance in classifying ten tomato leaf disease types.

5.1. ACCURACY

The accuracy of the training improved consistently with the increase in the epochs and was as close as 90 percent at the last training run as shown in Figure 6. The accuracy of validation also showed the same tendency with a uniform increase with stabilization of accuracy at 85-88. This implies that the model was capable of learning the underlying patterns effectively and at the same time generalize to unseen validation samples.

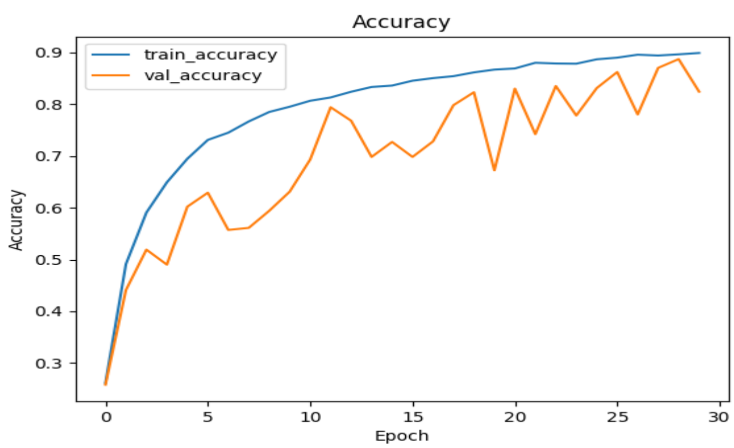

Fig. 6: Model Accuracy

5.2. PRECISION

Precision results also increased dramatically in the course of training, and the convergence of training and validation curves became above 0.9 in the later epochs as shown in Figure 07. It indicates that the model obtained a high level of true positive detection but with a low rate of false positives therefore, demonstrating the efficiency of the model in classifying the ten tomato leaf classes.

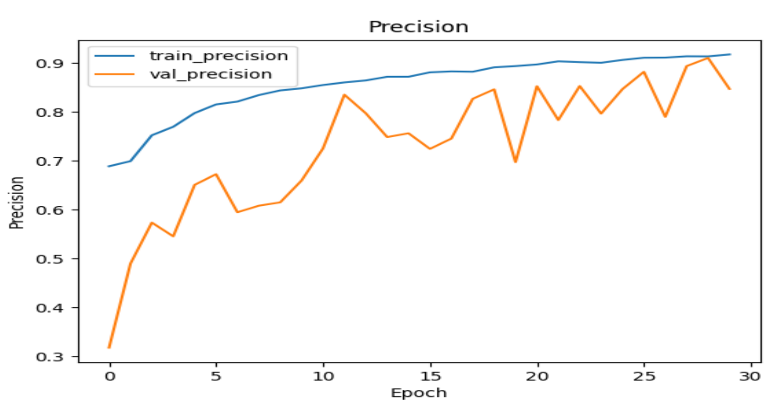

Fig. 7: Model Precision

5.3. RECALL

Figure 8 shows the trend in the recall measure between epochs in both the training and validation data sets. The model is performing well with the training recall steadily growing, but the validation recall is volatile and indicates that it might be overfitting as the difference between the two curves grows after around epoch 10.

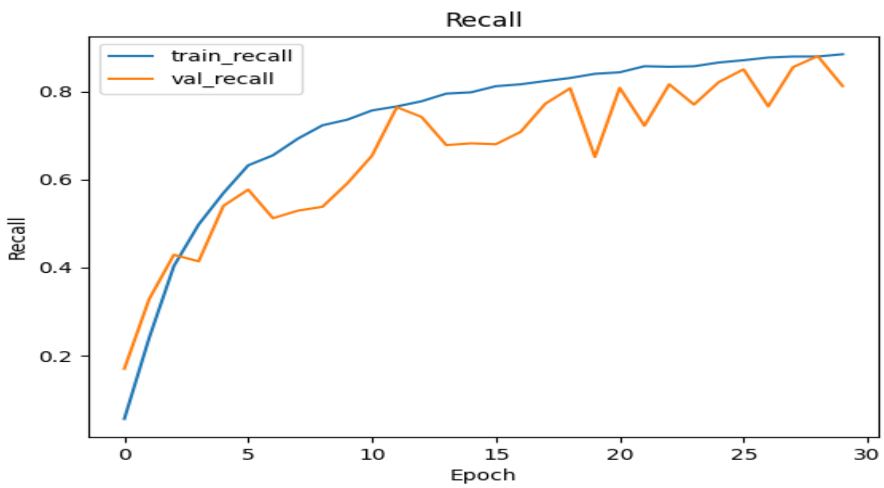

Fig. 8: Recall

5.4. PR CURVE

The figure 9 shows the Precision-Recall (PR) Curve of a multi-class classification problem, which is likely to be detecting various tomato plant diseases, which is followed with a One-vs-Rest approach. The results show very good performance in virtually all classes, with the Area Under the Curve (AP) values between 0.91 and 1.00 which demonstrates that the high precision can be achieved even at high recall levels.

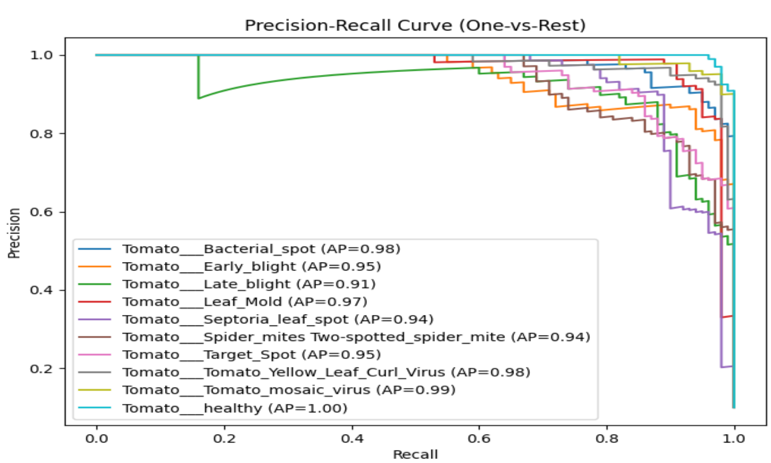

Fig. 9: Model Precision-Recall

5.5. ROC CURVE

Figure 10 shows depicts the Receiver Operating Characteristic (ROC) Curve of the multi-class classification model, once more involving application of One-vs-Rest strategy. The model shows a very high level of discriminatory power of all ten classes, with the Area Under the Curve (AUC) indicator being 0.98 or above. The almost-perfect curves, where some of the classes have an AUC of 1.00, indicate that the classifier is very good at distinguishing between all the target classes and all other classes, with as few false positives and false negatives as possible at different thresholds.

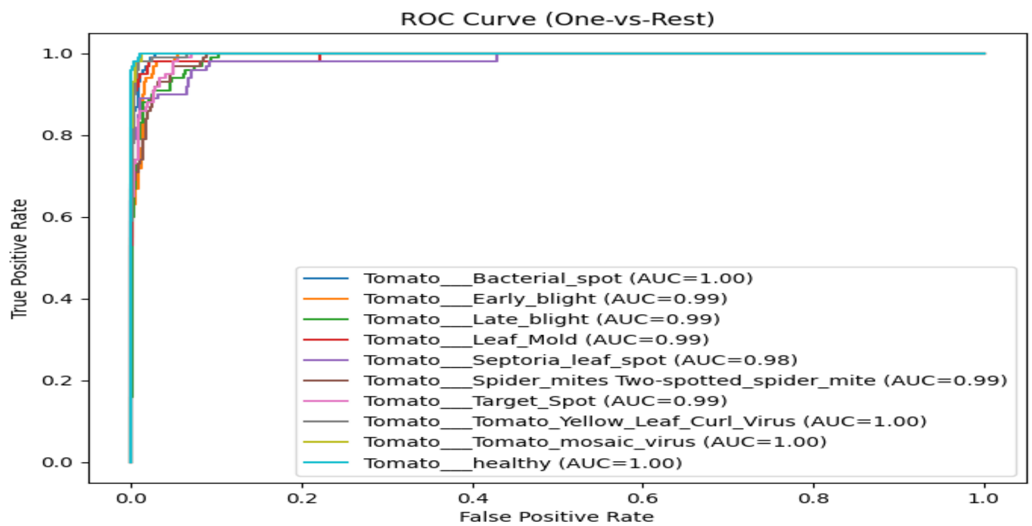

Fig. 10: Model Receiver Operating Characteristic (ROC)

5.6. CONFUSION MATRIX

The confusion matrix in Fig. 11 provides a fine-grained evaluation of the model classification of each of the ten different disease and healthy classes. The concentration of values around the main diagonal is strong, which indicates a strong discriminatory power, which validates the outstanding outputs as indicated by the ROC and PR curves in the previous figures. A majority of the classes like the one called 'healthy' (97) and ' YellowLeafCurlVirus' (98) has almost perfect true positive values. Nevertheless, the matrix also shows certain cases of inter-class confusion. Every misclassification is most pronounced in case of the misdesignation of the 'Spidermites Two-spottedspidermite' (78 TP) versus the misidentification as the 'TargetSpot' (15 FP), which implies that the subtle visual findings distinguishing the two conditions are also not easy to assign the name to the classifier. In the same way, there is a significant level of non-specificity between the mosaicvirus and Septorialeaf spot classes, in which 18 of the former is falsely predicted as the latter. Such off-diagonal values identify the location where the refinement of the feature extraction or division of the boundaries is required to remove diagnostic ambiguities.

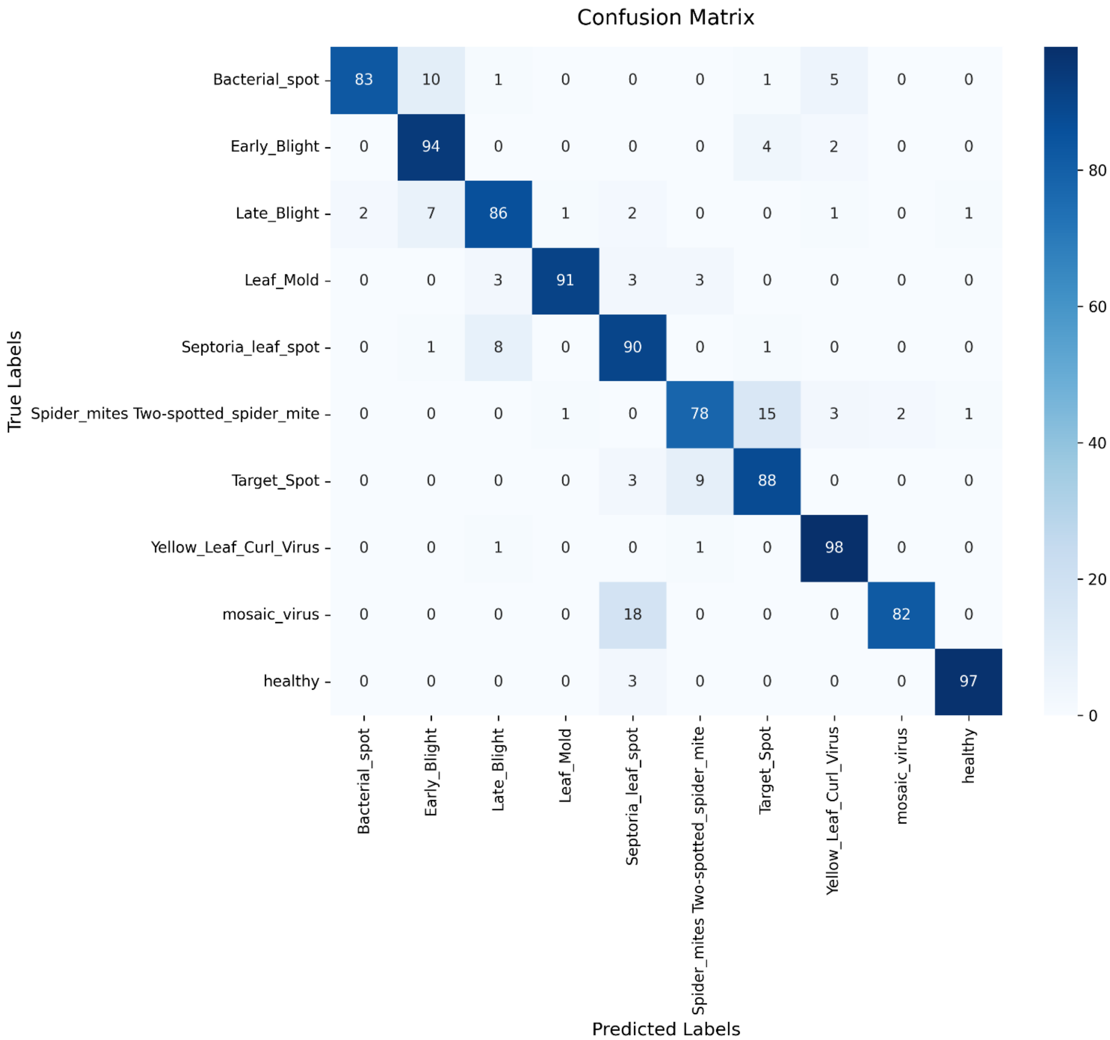

Fig. 11: Model Confusion Matrix

5.7. LOSS

The figure 12 shows the loss curve as a function of training epochs on both training and validation set. The training loss decays smoothly, and continuously, which represents that the model is continually learning on the training data. On the other hand, the validation loss is extremely large in value, and it is characterized by a high volatility that is exhibited by enormous, frequent spikes during the training. This sharp difference with the validation loss being much larger than the training loss is a standard signal of serious overfitting and an indication that the model parameters learned, do not generalize to unknown data.

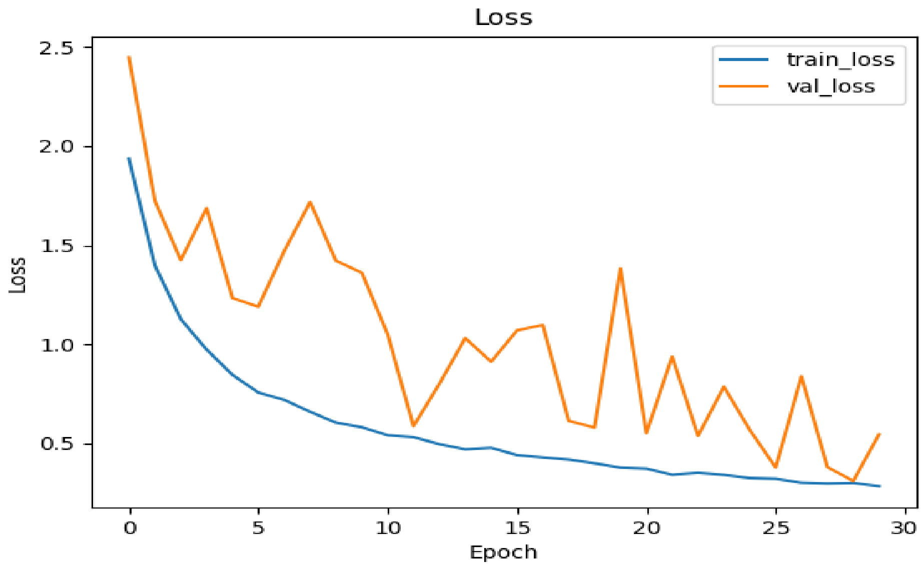

Figure 11: Model Loss

6. DISCUSSIONS

6.1. COMPARISON WITH RELATED WORKS

The primary contribution of this research lies in demonstrating that a relatively lightweight CNN architecture can deliver reliable tomato disease classification across ten classes without the need for complex hybrid frameworks or excessive hardware requirements. The balanced dataset of 10,000 training and 1,000 validation images ensured effective generalization across different disease categories. Moreover, the study provides context-specific insights by focusing on tomato cultivation in Pakistan, a country where agriculture is vital for livelihoods yet often constrained by limited access to advanced technology. By striking a balance between accuracy, efficiency, and practical applicability, this work highlights the feasibility of deploying AI-based disease diagnosis solutions in real-world farming contexts. Similarly, Assaduzzaman et al. [7] introduced XSE-TomatoNet an advanced architecture based on EfficientNetB0, although these studies demonstrate impressive accuracy, they can be expensive to use in an agricultural setting with low resources due to the size of the datasets and the intensive computations used. Our solution, on the contrary, is simpler and needs fewer computing resources yet it remains high-performing, which makes it fit the real-world application in developing countries like Pakistan.

6.2. KEY CONTRIBUTIONS

The main value of the research is that it has shown that a relatively lightweight CNN architecture can produce trustworthy tomato disease classification in ten categories without the necessity to use complex hybrid models, as well as without the resources of a heavy machine. The balanced dataset (10,000 training and 1,000 validation) gave the ability of effective generalization of the various disease categories. Additionally, the research offers context relevant findings by drawing a particular case on tomato production in Pakistan, a nation where agriculture is a crucial sector of the economy but is in most cases limited due to lack of access to modern technology. Providing a balance between accuracy, efficiency, and practical applicability, this work makes it clear that AI-based solutions to disease diagnostics can be utilized in the real farming conditions.

7. CONCLUSION AND FUTURE WORK

Due to the study, it was proved that automatic detection of tomato leaf diseases based on deep learning can greatly enhance the efficiency and speed of plant health measurement compared to the conventional practices. The experimental analysis proved that the suggested method could classify various tomato leaf conditions with a promising detection rate, which makes it a promising means of assistance to farmers and agricultural experts. Such systems can save on yield losses as it will reduce dependence on manual inspection, which can facilitate more sustainable disease management practices. Despite the positive outcomes of the results, there are still the aspects that should be improved further. The current model was also trained on a small dataset, making it hard to generalise to a very diverse field environment. The dataset may be extended in the future with images that are taken in actual farm setting at different light, background and time of the year. Besides, the connection with mobile-based services or IoT could also enable the real-time diagnostics and timely notices to farmers. It may also be possible to explore hybrid models, strategies of the transfer learning, and lightweight architectures to make the system stronger and deploy it to resource-constrained rural environments.

References

Hoque, M. J., Islam, M. S., & Khaliluzzaman, M. (2025). AI-Powered Precision in Diagnosing Tomato Leaf Diseases. Complexity, 2025(1), 7838841.

Singh, A., Kumar, S., & Choudhury, D. (2024, January). Tomato Leaf Disease Prediction Based on Deep Learning Techniques. In International Conference on Computation of Artificial Intelligence & Machine Learning (pp. 357-375). Cham: Springer Nature Switzerland.

Assaduzzaman, M., Bishshash, P., Nirob, M. A. S., Al Marouf, A., Rokne, J. G., & Alhajj, R. (2025). XSE-TomatoNet: An explainable AI based tomato leaf disease classification method using EfficientNetB0 with squeeze-and-excitation blocks and multi-scale feature fusion. MethodsX, 14, 103159.

Assaduzzaman, M., Bishshash, P., Nirob, M. A. S., Al Marouf, A., Rokne, J. G., & Alhajj, R. (2025). XSE-TomatoNet: An explainable AI based tomato leaf disease classification method using EfficientNetB0 with squeeze-and-excitation blocks and multi-scale feature fusion. MethodsX, 14, 103159.

Wang, X., & Liu, J. (2024). An efficient deep learning model for tomato disease detection. Plant Methods, 20(1), 61.
Pallavi, M. O., & Raj, A. (2024, December). Tomato Disease Prediction Model using Machine Learning Algorithms and

Image Processing Techniques. In 2024 6th International Conference on Computational Intelligence and Networks (CINE) (pp. 1-6). IEEE.

Shakarji, A., & Gölcük, A. (2025). Classification of Tomato Diseases Using Deep Learning Method. Journal of Intelligent Systems & Internet of Things, 14(2).

Joseph, V. A. (2024, March). Precision Agriculture Meets AI: Utilizing Deep Learning Models for Accurate Tomato Leaf Disease Classification. In 2024 International Conference on Recent Innovation in Smart and Sustainable Technology (ICRISST) (pp. 1-6). IEEE.

Ahmed, S., Aminuddin, N. F., Jhatial, M. J., Niaz, A., & Ali, I. (2024). AI-based Detection of Mask-Wearing Compliance Using Deep Neural Networks. Lahore Garrison University Research Journal of Computer Science and Information Technology, 8(3), 47-57.

Jhatial, M. J., Shaikh, R. A., Shaikh, N. A., Rajper, S., Arain, R. H., Chandio, G. H., ... & Shaikh, K. H. (2022). Deep learning-based rice leaf diseases detection using Yolov5. Sukkur IBA Journal of Computing and Mathematical Sciences, 6(1), 49-61.

Jhatial, M. J., Shaikh, R. A., Arain, R. H., Bhutto, K. H., & Talpur, S. A. (2023). Azure-based multi-sensor IoT network for smart rice-nursery field. VFAST Transactions on Software Engineering, 11(2), 187-195