DF-716
[1-5]

RESEARCH ARTICLE Agricultural Science Digest

Early Detection of Fungal Diseases in Tomatoes using Convolutional Neural Networks

Betsy Agustina Naomi Pinaria[1], Charles Reijnaldo Ngangi[1], Saartje Sompotan[1]

10.18805/ag.DF-716

ABSTRACT

Background: Tomatoes play a pivotal role in global agriculture fostering economic growth while also significantly enhancing food security. This paper introduces a data-based approach to improve the early detection of fungal infections in tomatoes.

Methods: The study utilises a dataset comprised of 4,362 high-resolution images obtained from Mendeley, showing healthy tomato plants alongside those damaged by leaf mold, early blight and late blight diseases. Deep learning techniques have been applied to develop a Convolutional Neural Network (CNN) for disease classification based on these images.

Result: The CNN model exhibits an overall accuracy rate of 90.83%, underscoring its efficacy in identifying fungal growth in tomato plants. The study proposes the use of machine learning in the detection and treatment of fungal diseases affecting tomatoes, which consequently leads to increased crop yield and quality preservation. It recommends the development of automated tools for farmers to detect and respond to disease outbreaks, thus improving their agricultural practices.

Key words: Convolutional neural networks, Early blight, Late blight, Leaf mold, Tomato.

INTRODUCTION

Tomatoes are the world's second-most significant vegetable crop, but they are also affected by various biotic stressors, environmental variables and diseases. These diseases can spread quickly due to their contagious nature, with fungi being the primary source of infectious illnesses. Tomatoes commonly experience fungal diseases such as leaf spot, Cercospora leaf mould, Fusarium wilt, Grey leaf spot, Powdery mildew, Verticillium wilt, White mould, Alternaria stem canker, Corky root rot, Fusarium crown and root rot, Fusarium foot rot, Southern blight and Buckeye rot. Pests and diseases are the primary cause of agricultural productivity decline and crop losses, resulting in tonnes of crops lost annually (Selvia *et al*., 2014; Hasan *et al*., 2020).

To prevent significant losses and boost yields, early diagnosis of these illnesses is therefore essential. Conventional techniques for identifying plant diseases include the use of diagnostic specialists (Sabri *et al*., 2020), are costly in terms of time and effort, require specialised instruments, have a high diagnostic error rate and cannot be done with just the naked eye (Dhandapani *et al*., 2022). Recent advances in computer technology introduced a new system that can identify plant diseases by images to distinguish different diseases of plants at an early stage. This is a significant advancement in the field of Agrifood which can automate many processes in modern agriculture. The system utilises Artificial Intelligence methods to achieve desired outcomes. The usage of convolution neural networks is the basis of the model utilized as they are a high-performance deep learning network which is feasible for processes in tomato plants.

Related work

Numerous researchers are tackling the problem of disease of tomato leaf identification using various methods that combine image processing and machine learning to automate the process. Many academics have recently turned to deep learning algorithms in their hunt for more accurate results.

Machine learning was known as the most successful approach in detecting diseases in fish, plants and animals, according to some researchers (Cho *et al*., 2024; AlZubi, 2023; Koike, 2023; Wasik and Pattinson, 2024) and also in the manufacturing industry (Porwal, 2024). Kumar *et al*. (2019) utilized two PlantVillage datasets, specifically the 2019 ImageNet dataset, to examine the plots of "Visual Geometry Group (VGG) Net, LeNet, ResNet50 and Xception" to detect nine distinct tomato leaf diseases. Hasan *et al*. (2019) created a pre-trained network model that correctly recognised illnesses affecting tomato plants at a 94% to 55% accuracy level.

Another work adopted transfer learning where a neural network based on AlexNet as a deep learning framework was used for disease classification on tomato plant leaves; it achieved an accuracy rate of 95.75%, reported by Sangeetha and Rani (2021). The use of VGG 19 and AlexNet architectures helped in disease identification related to tomatoes at a striking speed of 13,262 frames per second.

[1]Universitas Sam Ratulangi Manado 95115 Indonesia.

Corresponding Author: Betsy Agustina Naomi Pinaria, Universitas Sam Ratulangi Manado 95115 Indonesia.
Email: betsypinaria@unsrat.ac.id
Orchid: https://orcid.org/0009-0005-5338-5105.

How to cite this article: Pinaria, B.A.N., Ngangi, C.R. and Sompotan, S. (2025). Early Detection of Fungal Diseases in Tomatoes using Convolutional Neural Networks. *Agricultural Science Digest.* 1-5. doi: 10.18805/ag.DF-716.

Submitted: 23-01-2025 **Accepted:** 14-10-2025 **Online:** 16-10-2025

According to Rangarajan *et al*. (2018), the model was 97.49% correct. Kibriya *et al*. (2021) used two (CNNs), VGG16 and GoogLeNet, to classify tomato leaf diseases. This analysis's chief objective is to identify the greatest method for recognizing diseases of tomato leaves by employing deep-learning techniques. Working with 10,735 leaf pictures from the Plant Village dataset, it was observed that VGG16 had an accuracy rate of 98.8%, while GoogLeNet fetched 99.23% accuracy, with the latter having a higher result.

MATERIALS AND METHODS

Dataset collection

- A dataset comprising images of tomato plants affected by leaf mold, early blight and late blight diseases was obtained from Mendeley.
- The dataset consists of 4,362 images, including samples of healthy tomato plants for comparison (Fig 1).

Data preprocessing for CNN models

Data preprocessing for Convolutional Neural Network (CNN) input involves resizing images to 256×256 for consistency, normalizing pixel values, using data augmentation techniques like rotation, flipping and cropping to create training samples and incorporating additional information like labels or metadata to enhance input and model convergence. These steps significantly impact learning and inference processes, as well as the model's ability to operate in various applications.

Architecture of convolutional neural network (cnn) model

Many components of CNN's architectural paradigm promote learning efficiency. Convolutional layers are utilised to extract features, pooling layers to reduce features and softmax layers to classify them.

The hyperparameters are determined outside the algorithm and are established before training. There is no widely accepted methodology for determining the appropriate hyperparameters, leading to numerous experiments. Table 1 presents the hyperparameters utilised during the training of the model.

Convolution operations involve filters in convolutional layers, gradually covering the input picture or feature map from the preceding layer. Information is extracted from sections that overlap the filter at each stage. To get a value at a specific place on the feature map, the filter elements are multiplied by matching picture elements and the results are added. The convolution operation is defined over a two-dimensional kernel (K) and input image (I) as follows:

$$S_{(i,j)} = I * K_{(i,j)} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I_{(i+x,j+y)}$$

Where

i and j = Represent the image.

(I) = Coordinates.

m and n = Designate the kernel.

(K) = Coordinates.

The model uses pooling layers to reduce computational complexity and manage overfitting. The max pooling layer extracts the highest value from feature maps. The Fully Connected Layer then classifies the extracted features. Here, this layer was applied with Softmax function which predicts classes using numerical values based on information from the previous layer. The softmax function converts the numerical values $\chi 1$, $\chi 2$, $\chi 3$, ... χn.of the neurons in the next layer into probabilities P1, P2, P3, ...Pn. The probabilities are given as

$$Pi = \frac{e^{\chi i}}{\Sigma^{n}_{j=1} e^{\chi j}}$$

Where

P_i = Probability of class.

i = After applying softmax.

χ_i = The numerical value of the.

ith = Neuron in the preceding layer.

Next, the training model will be assessed with test data to determine how it performs with new data.

The evaluation matrix using classification parameters

The confusion matrix is a widely used approach for dealing with classification issues such as binary and multiclass classification. Accuracy, precision, recall and F1 score, commonly used measurement measures, were employed to analyse the proposed model's performance.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$\text{Precision} = \frac{TP}{(FP + TP)}$$

$$\text{Recall} = \frac{TP}{(FP + TP)}$$

$$\text{F1 - Score} = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

RESULTS AND DISCUSSION

The research developed Convolutional Neural Network (CNN) architectures to identify diseases in tomatoes by analyzing a dataset of healthy and diseased leaves. The CNN model achieved significant performance metrics, with a recorded loss of 0.1725 and an accuracy of 93.45% on the training dataset. It also achieved a validation loss of 0.2141 and an accuracy of 92.31% (Fig 2). Evaluated using an independent validation dataset, these measures demonstrate the model's capacity to apply its learned patterns to new, unseen data. Although there is a small decline in accuracy compared to the training set, the validation performance remains impressive, indicating the model's efficacy in detecting diseases in tomatoes through image analysis.

Further, the predictions and corresponding confidence scores for multiple instances are presented in Fig 3. The model consistently performs well across different classes. However, slight variations in accuracy and confidence scores offer useful insights into its ability to distinguish between each class.

The confusion matrix presents an analysis of the model's classification accuracy for three specific categories: healthy leaf, leaf mold and early and late blight.

Table 1: Hyperparameter for CNN model.

Parameters	Description
No. of convolution layer	6
No. of max pulling layer	6
Dropout rate	0.5
Network weight assigned	Uniform
Activation function	Relu
Flatten	(32, 2048)
Epoch	25
Batch size	32

Each of the rows in the matrix contains the actual labels for classes, whereas every column represents the expected class labels (Fig 4).

The CNN architectures have been evaluated for their effectiveness in detecting unhealthy leaves from various images, indicating their potential for disease diagnosis and plant healthcare. The model correctly detected both diseased and healthy leaves, indicating its capacity to distinguish between healthy and unhealthy tomato leaves.

The precision, recall and F1-score metrics are presented in Table 2. The model in this instance exhibits high accuracy in all classes, with precision ranging from 0.8137 to 1.0000, recall from 0.8646 to 0.9924 and F1-scores from 0.8384 to 0.9962. This demonstrates the model's ability to accurately determine samples for each class. The model's overall accuracy is 0.9083, indicating that it correctly predicts the class in 90.83% of instances. The model's performance in all classes is measured through macro-average and weighted average accuracy, recall and F1-score values. These statistics offer a thorough assessment that accounts for the varied distribution of examples among classes.

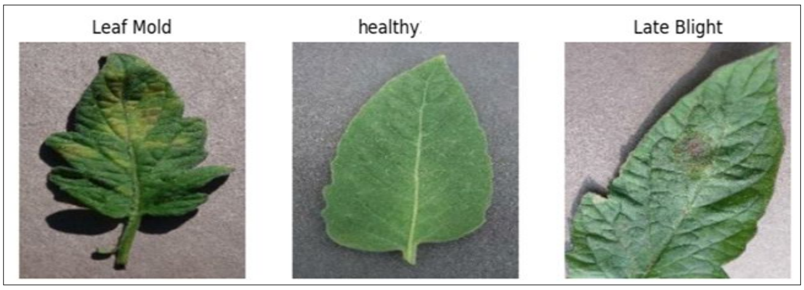

Fig 1: Images from the dataset include healthy and diseased tomato leaves.

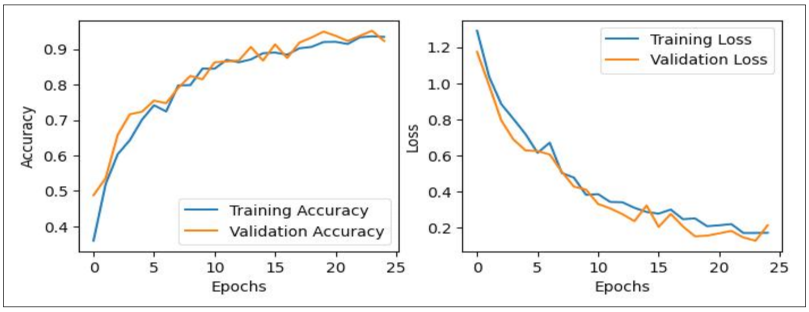

Fig 2: Performance in terms of loss and accuracy over epochs.

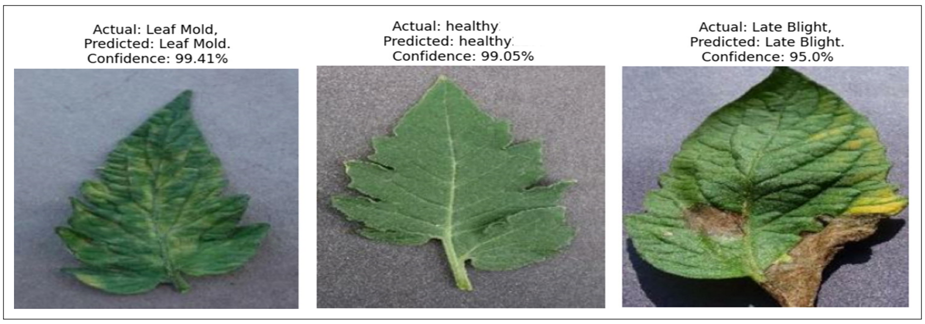

Fig 3: Actual and predicted diseased images using the suggested CNN model.

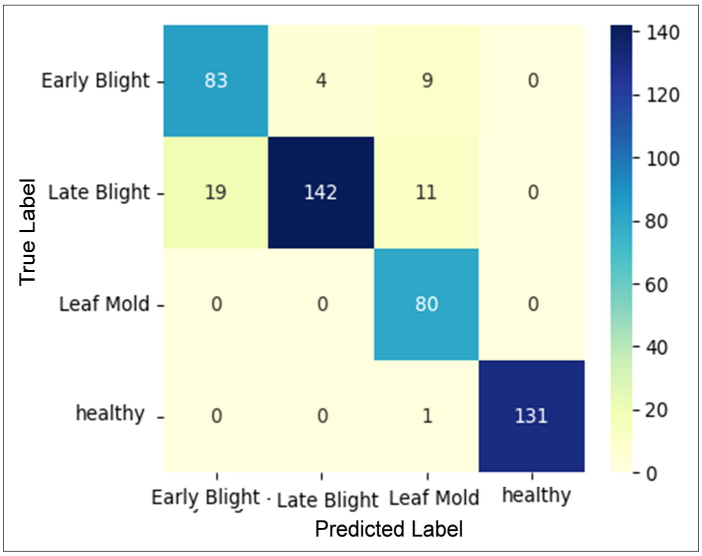

Fig 4: Confusion matrix.

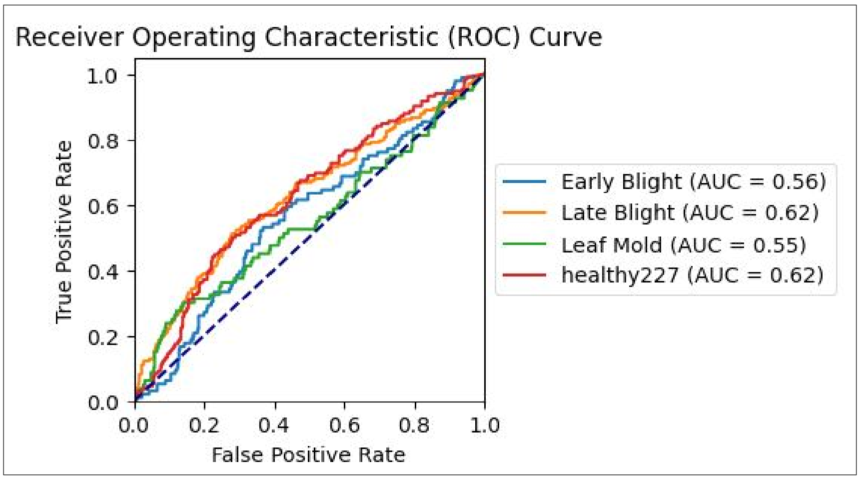

Fig 5: ROC Curve showing AUC for different diseases.

Table 2: Matrices for classification.

	Precision	Recall	F1 score	Support
Early blight	0.8137	0.8646	0.8384	96
Late blight	0.9726	0.8256	0.8931	172
Leaf mold	0.7921	1.000	0.8840	80
Healthy	1.000	0.9924	0.9962	132
Accuracy			0.9083	480
Macro avg	0.8946	0.9206	0.9029	480
Weighted avg	0.9183	0.9083	0.909	480

Fig 5 shows the classifier's ability to identify between many unhealthy classes, including early blight, late blight, leaf mold and a healthy class. Each unhealthy area under the curve (AUC) is displayed, with a greater AUC suggesting better performance in identifying the unhealthy from healthy plants. The AUC is low at 0.56 for early blight, showing the rate of false positives (FPR) or true positive rate (TPR) based on the classification level. In contrast, late blight had a higher AUC of 0.62, showing better distinction between the two.

CONCLUSION

Early detection of fungal infections allows the use of preventive treatments that involve antifungal treatments or crop rotation methods, resulting in reduced disease spread and production losses. Further research might focus on improving CNN designs and increasing the broad range of disease detection to include more crop species. Improvements in food security and environmentally friendly farming methods can be made by using deep learning skills and vast image datasets in the face of emerging challenges, such as pathogens, while also incorporating real-time tracking systems to improve precision agriculture.

ACKNOWLEDGEMENT

We would like to thank LPPM Universitas Sam Ratulagi and all parties involved in the completion of this research.

Funding details

This research received no external funding.

Authors' contributions

All authors contributed toward data analysis, drafting and revising the paper and agreed to be responsible for all the aspects of this work.

Data availability

The data analysed/generated in the present study will be made available from corresponding authors upon reasonable request.

Availability of data and materials

Not Applicable.

Use of Artificial Intelligence

Not applicable.

Declarations

Authors declare that all works are original and this manuscript has not been published in any other journal.

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions. The authors are responsible for the accuracy and completeness of the information provided but do not accept any liability for any direct or indirect losses resulting from the use of this content.

Conflict of interest

Authors declare that they have no conflict of interest.

REFERENCES

AlZubi, A.A. (2023). Artificial intelligence and its application in the prediction and diagnosis of animal diseases: A review. *Indian Journal of Animal Research*. **57(10):** 1265-1271. https://doi.org/10.18805/IJAR.BF-1684.

Cho, O.H., Na, I.S. and Koh, J.G. (2024). Exploring advanced machine learning techniques for swift Legume disease detection. *Legume Research*. **47(7):** 1221-1227. https://doi.org/10.18805/LRF-789.

Dhandapani, P. and Varadarajan, A. (2022). Multi-channel convolutional neural network for prediction of leaf disease and soil properties. *International Journal of Intelligent Engineering and Systems*. **15(1):** 318-328.

Hasan, M., Tanawala, B. and Patel, K.J. (2019, March). Deep Learning Precision Farming: Tomato Leaf Disease Detection By Transfer Learning. In Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE).

Hasan, R.I., Yusuf, S.M. and Alzubaidi, L. (2020). Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion. *Plants*. **9(10):** 1302.

Kibriya, H., Rafique, R., Ahmad, W. and Adnan, S. M. (2021). Tomato leaf disease detection using convolution neural network. In 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST) (pp. 346-351). IEEE.

Koike, T., Yamamoto, S., Furui, T., Miyazaki, C., Ishikawa, H. and Morishige, K.I. (2023). Evaluation of the relationship between equol production and the risk of locomotive syndrome in very elderly women. *International Journal of Probiotics and Prebiotics*. **18(1):** 7-13. doi: https://doi.org/10.37290/ijpp2641-7197.18:7-13.

Kumar, A. and Vani, M. (2019, July). Image Based Tomato Leaf Disease Detection. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

Porwal, S., Majid, M., Desai, S. C. Vaishnav, J. and Alam, S. (2024). Recent advances, challenges in applying artificial intelligence and deep learning in the manufacturing industry. *Pacific Business Review* (International). **16(7):** 143-152.

Rangarajan, A.K., Purushothaman, R. and Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Computer Science*. **133:** 1040-1047.

Sabri, N., Kassim, N.S., Ibrahim, S., Roslan, R., Mangshor, N.N.A. and Ibrahim, Z. (2020). Nutrient deficiency detection in maize (Zea mays L.) leaves using image processing. *IAES International Journal of Artificial Intelligence*. **9(2):** 304.

Sangeetha, R. and Rani, M.M.S. (2021). Tomato Leaf Disease Prediction using Transfer Learning. In Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6. 2020, Revised Selected Papers, Part II 10 (pp. 3-18). Springer Singapore.

Selvia, D.S., Betsy, A.N.P., Dantje, T. and Emmy, S. (2014). Insect Species and Populations in Generative Phase Tomato Plants (Lycopersicum Esculentum Mill) Treated With in organic And Organic Fertilizers In Tonsewer Village, Tompaso II District, Cocos Journal, 2014.

Wasik, S. and Pattinson, R. (2024). Artificial intelligence applications in fish classification and taxonomy: Advancing our understanding of aquatic biodiversity. *FishTaxa*. **31:** 11-21.