Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 3(3) May 2025
DOI: https://dx.doi.org/10.4314/jobasr.v3i3.2

A CNN-SVM Based Model for Detection and Classification of Tomato Leaf Diseases

Aminu Bashir Suleiman[1*], Stephen Luka[2] and Joseph Nda Ndabula[3]
[1]Department of Cyber Security, Federal University Dutsin-Ma, Katsina, Nigeria
[2, 3]Department of Software Engineering, Federal University Dutsin-Ma, Katsina, Nigeria
*Corresponding Author Email: ameenu.basheer10@gmail.com

ABSTRACT
Tomato leaf diseases represent a substantial risk to global agriculture, leading to decreased crop yields and inferior fruit quality. Conventional disease detection techniques depend significantly on manual examination, resulting in delays and inaccuracies. This paper investigates the application of machine learning CNN-SVM methodology to create an automated system for the detection and classification of tomato leaf diseases. The research employs datasets from PlantVillage and Labeled_Features, consisting of more than 18,000 images of high-resolution tomato leaves afflicted by diverse diseases. The images underwent preprocessing, augmentation, and classification into three primary disease categories: Tomato Bacterial Spot, Tomato Leaf Curl Virus, and Tomato Mosaic Virus, in addition to healthy tomato leaves. The evaluation of our model attained the highest classification accuracy at 98.2%. The evaluation metrics, comprising precision, recall, and F1-score, averaged 98%, signifying the model's efficacy in differentiating between the leaves that have diseases and healthy leaves. Moreover, the study emphasizes critical attributes, including color variation, texture descriptors, and shape characteristics, which were essential in enhancing the model's performance. Notwithstanding the success, the study also delineates areas for enhancement, particularly in differentiating analogous diseases and mitigating environmental variability. The results highlight the efficacy of the CNN-SVM model in contemporary agricultural practices, providing efficient and economical solutions for real-time disease detection and management.

Keywords:
Tomato Leaf Diseases, Machine Learning, Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Disease Classification.

INTRODUCTION

Tomato leaf diseases substantially affect global agricultural productivity, presenting a formidable challenge to the agricultural sectorand also the farmers. The increasing demand for sustainable food production necessitates the efficient detecting and classifyingthe tomato leaf diseases to avert crop losses, improve yield quality, and guarantee food security (Wang et al., 2024).The worrisome rise in plant diseases brought on by climate change and changing pathogens has made advanced detection technologies, which can quickly identify and classify these diseases, even more important in guaranteeing agricultural sustainability and food security for a growing world population. Therefore, early and accuracy in the identification of tomato leaf diseases is very important for carrying out quick interventions, minimizing crop losses, maximizing pesticide use, lowering environmental impact, and finally, supporting sustainable agricultural practices in both developed and developing areas (Wang et al., 2024; Jafar et al., 2024). AI filed has made progress in the detection and classification phase (Abubakar et al., 2025).

Among the most often grown vegetables worldwide, tomato plants are vulnerable to a variety of diseases that could compromise their leaves, stems, or fruits (Paret et al., 2013). Widespread ailments such as late blight, early blight, bacterial spot, and leaf curl result in diminished crop yield and inferior produce quality, affecting both smallholder farmers and large-scale producers (Luna-Benoso et al., 2021; GM et al., 2022). Traditional agricultural methods rely on visual inspection for disease detection, a process that is work-intensive and susceptible to inaccuracies. The development of machine learning algorithms and image processing techniques presents an opportunity to improvethe accuracy of disease detection and automate the process, thereby conserving time and resources (Huang et al., 2022).

Effective early identification is the major obstacle in tomato disease diagnosis. While lacking automated systems prevents timely intervention, manual methods cause delays and misdiagnosis. To stop disease spread and crop losses, accurate, quick detection is crucial; thus, modern farming depends on automation (Ullah et al., 2023).Despite progress, reliable tomato disease detection systems remain limited. Challenges include low accuracy in distinguishing similar diseases, environmental variability, and reliance on large datasets. Most methods also need specialized equipment, limiting accessibility. New models must improve accuracy while adapting to diverse real-world conditions for practical farm use (Routis et al., 2024).

This paper presents a model that helps in the detection and classification of tomato leaf diseases utilizing deep learning and computer vision techniques. This paper aims to enhance the agricultural sector by developing and evaluating a system that accurately classifies tomato leaf diseases, thereby providing farmers with an accessible tool to improve plant health management.

Recent studies have extensively utilized diverse Convolutional Neural Network (CNN) architectures for the detection and classification of diseases in tomato leaf.

A study by (Chen et al., 2022) utilized a modified AlexNet architecture, attaining remarkable results with an average accuracy of 98%, a precision rate of 0.98, a recall value of 0.99, and an F1 score of 0.98. The researchers employed a comprehensive dataset which contains 18,345 for training images and 4,585 that will be the testing images, classified into ten unique tomato leaf disease categories, with each image normalized to 64 × 64 RGB pixels. The model was refined utilizing the Adam optimizer with a learning rate of 0.0005 and trained for 75 epochs with a batch size of 128.

Another study by (Subramanian et al., 2024) devised a bespoke CNN architecture tailored for the tomato leaf diseases classification. This model included four convolutional layers with ascending filter sizes (32, 64, 128, and 256) and utilized both ReLU and SELU activation functions for effective feature extraction. The network architecture was meticulously constructed with max-pooling layers to diminish spatial dimensions and a fully connected layer comprising 1024 units, culminating in a softmax layer for multi-class classification. This method attained a 94% accuracy rate while preserving a balance between computational efficiency and performance, illustrating the efficacy of meticulously crafted CNN architectures for this particular agricultural application.

A study conducted by (Priyadharshini et al., 2023) has also provided valuable insights into the comparative performance of various CNN architectures. The study assessed multiple CNN variants, including R-CNN, Fast R-CNN, and Faster R-CNN, for the tomato leaf diseases detection. The researchers utilized Visual Geometry Group (VGG-16) for feature extraction and classification in conjunction with the regression boundary box method for disease localization. Following 20 training iterations, the Faster R-CNN model attained approximately 98% accuracy, thereby confirming its efficacy in detecting and classifying diseases from tomato leaf images.

Another study by (Halder et al., 2024) assessed multiple pre-trained models, including ResNet, MobileNet, DenseNet, and InceptionV3, utilizing the PlantVillage dataset for the classification of tomato diseases. The findings indicated that these transfer-learning methodologies could proficiently discern disease-specific patterns in leaf images, thereby enhancing automated disease detection systems that promote improved farm management and sustainability.

A study by (Upadhyay et al., 2024) utilized sophisticated data augmentation methods and transfer learning from an extensive plant disease dataset to create a resilient classifier for tomato leaf diseases. This method achieved exceptional performance, with accuracy exceeding 95% on high-resolution images of tomato leaves affected by various diseases. The model exhibited exceptional resilience under challenging conditions, rendering it suitable for practical agricultural applications where image quality and environmental factors may fluctuate considerably.

MobileNet architectures have become prominent owing to their efficiency and suitability for deployment on resource-limited devices. A research study carried out by (Tarek et al., 2022) assessed various deep learning models pre-trained on the ImageNet dataset, including ResNet50, InceptionV3, AlexNet, MobileNetV1, MobileNetV2, and MobileNetV3, and demonstrated that MobileNetV3 Large attained an exceptional accuracy of 99.81%. The researchers effectively implemented these models on both workstations and Raspberry Pi 4 devices, with MobileNetV3 Small attaining latencies of 66 ms and 251 ms, respectively, illustrating the feasibility of developing Internet of Things (IoT) devices for real-time detection of tomato leaf diseases.

A study by (Srivastava et al., 2023) presented the TLMV2-ELM model, which integrates MobileNetV2 for effective feature extraction with an Extreme Learning Machine for classification purposes. This method attained exceptional outcomes, demonstrating 99% accuracy and a mere 0.06 loss in disease detection. The amalgamation of ELM, recognized for its rapid learning duration and superior generalization abilities, with the feature extraction prowess of MobileNetV2 resulted in a highly efficient system for diagnosing tomato leaf diseases, surpassing numerous existing methodologies.

A study by (Imam et al., 2024) introduced an innovative framework that integrates MobileNet with Support Vector Machine (SVM) for the categorization of tomato leaf diseases. The MobileNet-SVM approach employed transfer learning to extract features via MobileNet's pre-trained capabilities, which were then fed into an SVM classifier utilizing hinge loss to differentiate among nine distinct types of tomato leaf diseases and the healthy leaves. Statistical analysis confirmed the superiority of this hybrid model, achieving an outstanding overall accuracy of 99.37%, as well as notable precision, recall, and Area under the Curve (AUC) value.

(Sundaramoorthi & Kamarasan, 2024) proposed a Chaotic Butterfly Optimization Algorithm Deep Learning Method for the Detection and Classification of the Tomato Leaf Disease (CBOADL-TLDDC). This advanced method utilized Gaussian filtering for preprocessing and implemented the ShuffleNet architecture for feature extraction. The Chaotic Butterfly Optimization Algorithm was integrated with Extreme Gradient Boosting (XGBoost) for hyperparameter optimization and disease classification, achieving an accuracy of 93.75%.

MATERIALS AND METHODS

The techniques, processes, and strategies used in the model's implementation are presented and elaborated in Figure 1.

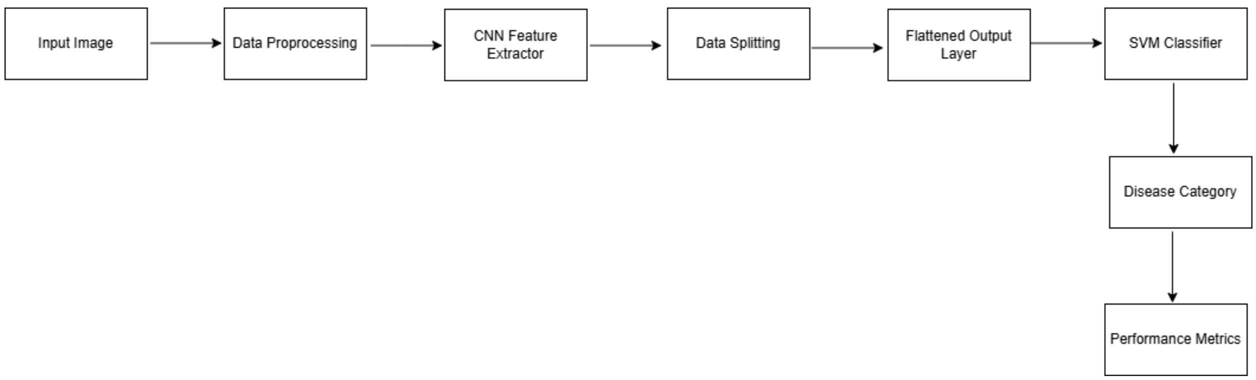

Figure 1: Proposed Model Architecture

Input Image: The system starts with raw images of tomato leaves captured under various conditions. These images may show different types of diseases or represent healthy leaves.

Data Preprocessing

Before feeding the images into the model, preprocessing techniques are applied to improve consistency and model performance. This includes resizing all images to a uniform dimension of 224×224, normalizing pixel values scaling between 0–1, and applying data augmentation (flipping, rotation, zooming) to expand the dataset and prevent overfitting.

CNN Feature Extractor

A Convolutional Neural Network (CNN) is used to automatically extract meaningful features from the tomato leaf images. The CNN learns to detect patterns such as spots, blights, or discolorations that are indicative of specific diseases. This stage includes convolutional layers, activation functions (like ReLU), and pooling layers to reduce spatial dimensions while preserving key features.

Data Splitting

The preprocessed dataset was partitioned into training and testing sets to guarantee a dependable model assessment. 80% of the images were designated for the training set, whereas 20% were utilized for the validation phase.

Flattened Output Layer

The multi-dimensional feature maps produced by the CNN are flattened into a one-dimensional feature vector. This transformation is necessary to feed the extracted features into the SVM classifier, which only accepts vector input.

SVM Classifier

The Support Vector Machine (SVM) receives the flattened feature vector and performs the classification. The SVM algorithm identifies the optimal hyperplane that separates different classes (e.g., healthy, early blight, late blight, leaf mold) based on learned features. An RBF or polynomial kernel is commonly used for better non-linear separation.

Disease Category

Finally, the model outputs the predicted disease category for each tomato leaf image. This result helps in early detection and accurate diagnosis, which is crucial for timely intervention and crop management.

Model Evaluation

The developed modelwas assessed using the 20% validation dataset to evaluate the classification accuracy and generalization ability. Performance metrics, such as overall accuracy, precision, recall, and F1-score, were utilized to evaluate the performance, offering comprehensive insights into their capacity to detect specific tomato leaf diseases.

Accuracy assesses the model's proficiency in correctly predicting or classifying data. The formula is as follows:

$$Accuracy = \frac{True\ Positive\ + True\ Negative}{True\ Positive\ +\ True\ Negative\ +\ False\ Positive\ +\ False\ Negative} \tag{1}$$

Precision assesses the accuracy of the positive predictions generated by the model. The formula is presented as follows:

$$Precision = \frac{True\ Positive}{False\ Positive\ +\ True\ Positive} \tag{2}$$

Recall assesses the model's capacity to accurately identify all pertinent instances within the dataset. The equation is;

$$Recall = \frac{True\ Positive}{False\ Negative\ +\ True\ Positive} \tag{3}$$

The F1-score provides a balance between precision and recall, yielding a singular metric that integrates both.

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

RESULTS AND DISCUSSION

Table 1: Performance Evaluation Metrics

Catgory	Precision	Recall	F1-Score
Tomato_Bacterial_Spot	97%	99%	98%
Tomato_Yellow_Leaf_Curl_Virus	99%	95%	97%
Tomato_Mosaic_Virus	98%	99%	98%
Tomato_Healthy	99%	100%	100%

Table 1 above shows a bar chart of the classification report metrics for the model prediction of the tomato leaf disease. The metrics include precision, recall, and F1-score for four classes. The model did exceptionally well throughout all the classes, with the following average metrics:precision at 98%, recall at 98%, and F1-score at 98%. Each bar represents the score for oneof the metrics, demonstrating high accuracy in predicting both diseased and healthy tomato leaves.

Table 2: Confusion Matrix

Class	True Positive	True Negative	False Positive	False Negative
Tomato_Bacterial_Spot	422	1407	12	3
Tomato_Yellow_Leaf_Curl_Virus	467	1351	3	23
Tomato_Mosaic_Virus	445	1385	11	3
Tomato_Healthy	481	1360	3	0

Table 2 above represents the performance of a classification model across four classes: Bacterial Spot, Leaf Curl Virus, Tomato Mosaic Virus, and Tomato Healthy. The diagonal cells show the correct classifications: 422 instances of Bacterial Spot, 467 of Leaf Curl Virus, 445 of Tomato Mosaic Virus, and 481 of Tomato Healthy were correctly predicted. Off-diagonal cells indicate misclassifications, such as 3 instances of Bacterial Spot being misclassified as Leaf Curl Virus and 12 instances of Leaf Curl Virus misclassified as Bacterial Spot, along with 11 instances misclassified as Tomato Mosaic Virus. There are also some misclassifications of the Tomato Mosaic Virus (3 instances as Tomato Healthy). Overall, the model performs well with high accuracy, although there are some misclassifications, particularly between "Bacterial Spot" and "Leaf Curl Virus".

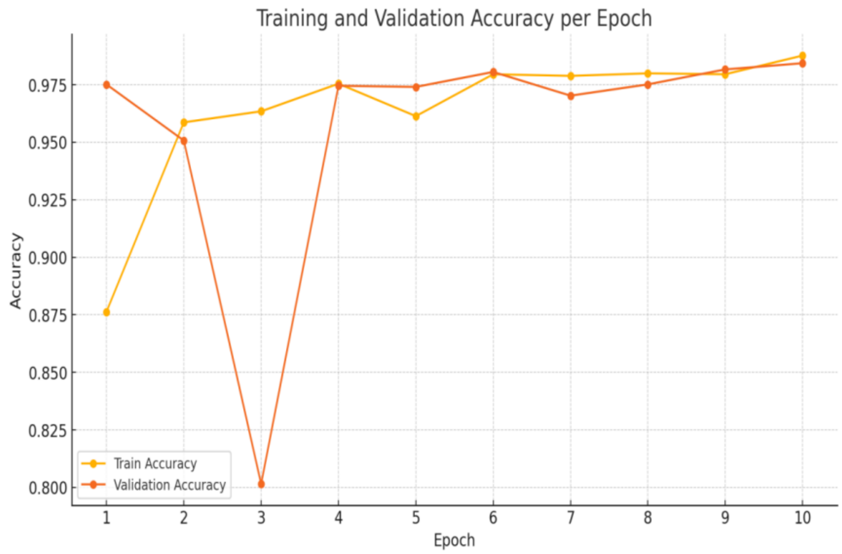

Figure 2: Training and Validation Accuracy per Epoch

Figure 2 above illustrates the training and validation accuracy for each epoch of our model. This visualization monitors the model's performance across 10 training epochs, illustrating the trends in both training and validation accuracy.

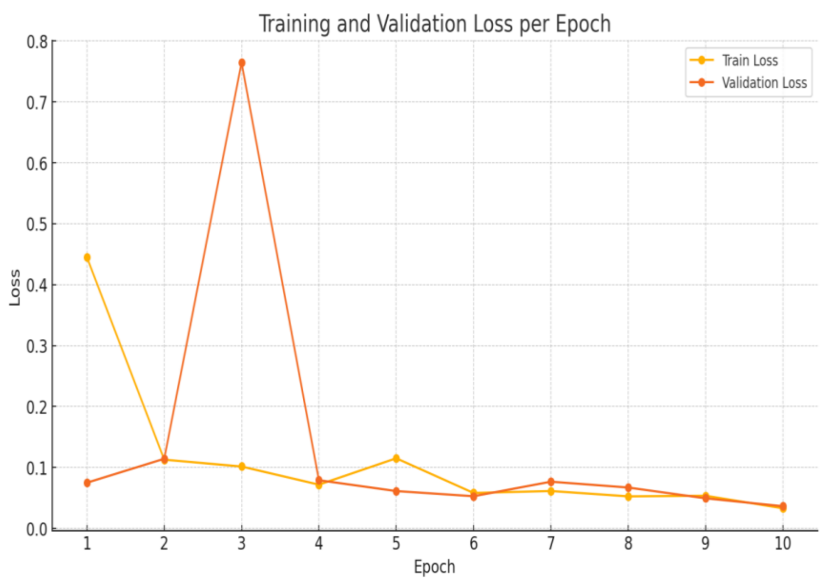

Figure 3: Training and validation Loss per Epoch

Figure 3 above illustrates the training and validation loss for every epoch. This visualization illustrates the reduction of the model's loss across the 10 training epochs, depicting the trends of both training and validation loss.

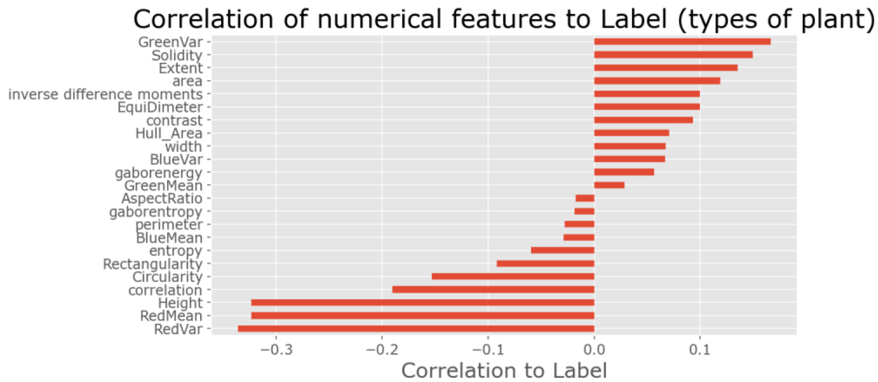

Figure 4: Correlation of numerical features of label

Figure 4, above presents a horizontal bar plot illustrating the correlation between numerical features and plant type labels, a key part of the exploratory data analysis (EDA) phase in machine learning workflows. This analysis identifies features, such as color variance (GreenVar, RedVar), shape descriptors (Solidity, Extent), and texture metrics (entropy, contrast), that show significant correlations with plant classes. Features with strong correlations, like GreenVar and Solidity, are crucial for distinguishing between healthy and diseased leaves. This correlation insight aids the feature selection process for an SVM classifier, where focusing on informative features improves classification accuracy and computational efficiency. Features like RedVar and RedMean, with strong negative correlations, may indicate diseased areas, aligning with methods in the cited paper to enhance model performance for the detection of thetomato leaf disease.

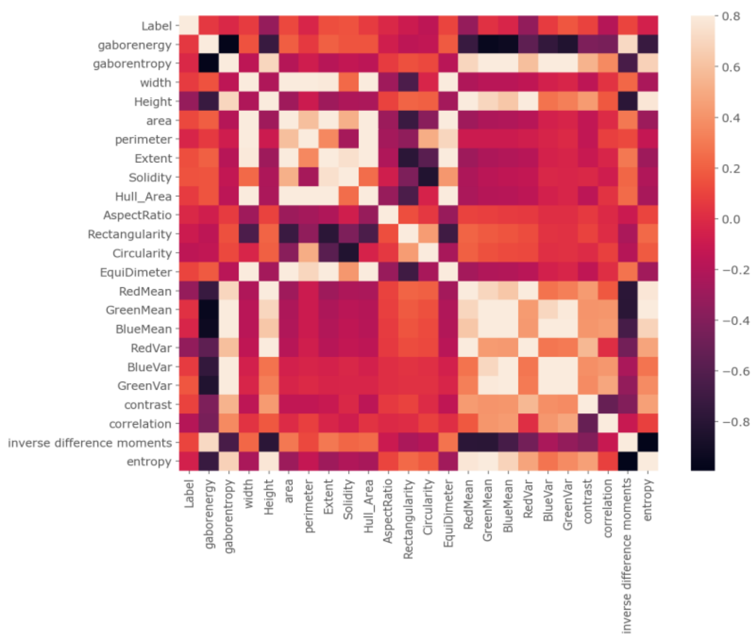

Figure 5: Correlation Heatmap

The figure 5 above is a **correlation heatmap** that visually represents the pairwise correlation coefficients among numerical features used for classifying tomato leaf diseases. Each cell in the heatmap represents the magnitude and orientation of the linear correlation between two features, with color intensity ranging from deep purple (strong negative correlation) to light peach (strong positive correlation). The inclusion of the "Label" row and column allows analysis of how each feature correlates with the disease classification labels, aiding in identifying influential features. This kind of analysis is vital for reducing feature redundancy, detecting multicollinearity, and selecting the most relevant features, which then improves the performance and the SVM classifier accuracy used in the study.

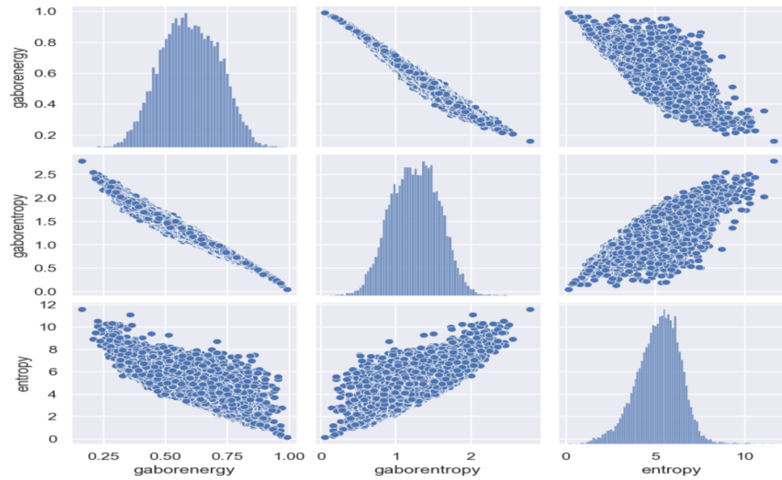

Figure 6: Three Texture-based features

The Figure 6, above is a **pairplot** that visualizes the relationships and distributions of three texture-based features—**gaborenergy**, **gaborentropy**, and **entropy** used in the tomato leaf diseases classification, as described in the study. The diagonal plots show the distribution of each feature, while the scatter plots in the off-diagonal cells depict pairwise relationships, revealing strong negative correlations between gaborenergy and both gaborentropy and entropy, as well as a strong positive correlation between gaborentropy and entropy. These insights are critical for understanding the interdependence between texture descriptors, guiding feature selection, and dimensionality reduction to enhance the performance of the model by minimizing redundancy and improving the discriminative power of the model.

CONCLUSION

This research illustrates the effective implementation of machine learning models, particularly the Support Vector Machine (SVM), for classifying tomato leaf diseases. Through exploratory data analysis (EDA), the study identified key features such as color variance, shape descriptors, and texture metrics that significantly correlated with plant type labels. The model attained 98.2%of the highest accuracy.The model's performance was further validated through training and validation accuracy and loss plots, showing consistent improvement across epochs. Evaluation metrics, including precision, recall, and F1-score, averaged at 98%, reflecting the model's exceptional ability to predict both diseased and healthy tomato leaves. The confusion matrix further highlighted the areas for further improvement and then the strengths of the model, with high accuracy in most classes, though occasional misclassifications occurred, particularly between Bacterial Spot and Leaf Curl Virus. Overall, the findings underscore the impact of the model in the classification of the tomato leaf disease, with the potential for further optimization through feature selection and model tuning.

REFERENCE

Abubakar, M., Surajo, Y. & Tasiu , S. (2025). An Explainable Deep Learning Model for Illegal Dress Code Detection and Classification. Journal of Basics and Applied Sciences Research, 3(1), 10. DOI: https://dx.doi.org/10.4314/jobasr.v3i1.1

Chen, H. C., Widodo, A. M., Wisnujati, A., Rahaman, M., Lin, J. C. W., Chen, L., & Weng, C. E. (2022). AlexNet Convolutional Neural Network for Disease Detection and Classification of Tomato Leaf. Electronics 2022, 11, 951.

GM, S. K., Sriram, S., Laxman, R. H., & Harshita, K. N. (2022). Tomato late blight yield loss assessment and risk aversion with resistant hybrid. Journal of Horticultural Sciences, 17(2), 411-416.

Halder, A., GuhaNeogi, S., Dutta, S., Khamaru, K., Bhattacharya, A., & Roy, S. (2024). Early Detection and Classification of Tomato Leaf Blight Disease Using Deep Neural Networks. 2024 International Conference on Data Science and Network Security (ICDSNS), 1-10.

Huang, J., Li, J., Li, Z., Zhu, Z., Shen, C., Qi, G., & Yu, G. (2022). [Retracted] Detection of Diseases Using Machine Learning Image Recognition Technology in Artificial Intelligence. Computational Intelligence and Neuroscience, 2022(1), 5658641.

Imam, M. H., Nahar, N., Bhowmik, R., Omit, S. B. S., Mahmud, T., Hossain, M. S., &Andersson, K. (2024, May). A transfer learning-based framework: Mobilenet-svm for efficient tomato leaf disease classification. In 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) (pp. 693-698). IEEE.

Jafar, A., Bibi, N., Naqvi, R. A., Sadeghi-Niaraki, A., & Jeong, D. (2024). Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. Frontiers in Plant Science, 15, 1356260.

Luna-Benoso, B., Martínez-Perales, J. C., Cortés-Galicia, J., Flores-Carapia, R., & Silva-García, V. M. (2021). Detection of diseases in tomato leaves by color analysis. Electronics, 10(9), 1055.

Paret, M., Pernezny, K., & Roberts, P. (2013). Disease control for Florida tomatoes. Univ. Fla. Coop. Ext. Serv. Tech. Bull. EDIS PPP35. http://edis. ifas. ufl. edu/vh056 (verified 23 Feb. 2016).

Priyadharshini, G., Raveena, D., & Dolly, J. (2023). Comparative Investigations on Tomato Leaf Disease Detection and Classification Using CNN, R-CNN, Fast R-CNN and Faster R-CNN. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), 1, 1540-1545.

Routis, G., Michailidis, M., &Roussaki, I. (2024). Plant disease identification using machine learning algorithms on single-board computers in IoT environments. Electronics, 13(6), 1010.

Srivastava, M., Sisaudia, V., &Meena, J. (2023, November). Tomato Leaf Disease Detection and Classification using MobileNetV2 and Extreme Learning Method: A Hybrid Approach. In 2023 10th International Conference on Soft Computing & Machine Intelligence (ISCMI) (pp. 108-112). IEEE.

Subramanian, R. R., Deepalakshmi, M., & Kumar, S. P. P. (2024, December). Deep Learning-Build Detection and Classification of Tomato Leaf Diseases Using Advanced CNN Architectures. In 2024 IEEE Pune Section International Conference (PuneCon) (pp. 1-5). IEEE.

Sundaramoorthi, K., & Kamarasan, M. (2024, April). A Chaotic Butterfly Optimization Driven Deep Learning Approach for Tomato Leaf Disease Detection and Classification. In 2024 International Conference on Computing and Data Science (ICCDS) (pp. 1-6). IEEE.

Tarek, H., Aly, H., Eisa, S., & Abul-Soud, M. (2022). Optimized deep learning algorithms for tomato leaf disease detection with hardware deployment. Electronics, 11(1), 140.

Ullah, Z., Alsubaie, N., Jamjoom, M., Alajmani, S. H., & Saleem, F. (2023). EffiMob-Net: A deep learning-based hybrid model for detection and identification of tomato diseases using leaf images. Agriculture, 13(3), 737.

Upadhyay, L., & Saxena, A. (2024). Evaluation of Enhanced Resnet-50-Based Deep Learning Classifier for Tomato Leaf Disease Detection and Classification. Journal of Electrical Systems, 20(3s), 2270-2282.

Wang, Y., Zhang, P., &Tian, S. (2024). Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion. Frontiers in Plant Science, 15, 1382802.