



COMPARATIVE ANALYSIS OF SVM AND CNN TECHNIQUES FOR TOMATO LEAF DISEASE CLASSIFICATION

Kishori Telaware^{*1}, Rajani P.K.² and Rohini Daund³

^{1, 2, 3}Department of Electronics and Telecommunication, Pimpri Chinchwad College of Engineering, Pune, India.

¹<http://orcid.org/0009-0008-1259-9811>, ²<http://orcid.org/0000-0003-3635-2094>, ³<http://orcid.org/0009-0001-0365-0637>

Email: *kishori.telaware24@pccoepune.org, rajani.pk@pccoepune.org, rohini.daund24@pccoepune.org

ARTICLE INFO

Article History

Received: February 7, 2026

Reviewed: July 31, 2026

Accepted: August 31, 2026

Published: September 30, 2026

Keywords:

CNN,
Deep Learning,
Disease Classification,
Machine Learning,
SVM.

ABSTRACT

Tomatoes are among the most widely cultivated crops worldwide and are valued for their taste, nutritional benefits, and culinary versatility. However, tomato plants are sensitive to various leaf diseases that can significantly reduce yield and quality, making early and accurate disease identification essential for effective management. This study involves two machine learning techniques to categorize tomato leaf diseases using a dataset of 10,000 labelled images representing ten categories, including healthy leaves and bacterial, viral, and fungal infections. A Convolutional Neural Network (CNN) model was developed in Python using TensorFlow with a pretrained DenseNet121 architecture, enabling automatic feature extraction and classification. In contrast, a Support Vector Machine (SVM) model relies on manually extracted features such as color, texture, and shape. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. The CNN outperformed the SVM, achieving an accuracy of 93.35% compared to 78% for the SVM. These results highlight the superiority of deep learning approaches for plant disease detection. The proposed CNN-based system can assist farmers in timely disease diagnosis, reduce crop losses, and minimize pesticide usage. Furthermore, the model can be deployed in real-world agricultural settings through integration with mobile-based platforms or Internet of Things systems for on-site disease monitoring and decision support.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Increasing crop production is the main aim of an agricultural activity. The quality of soil, meteorological processes, and successful management of plant diseases are the few factors that drastically affect growth and productivity of crops [1]. Periodic outbreaks of diseases have led to tremendous crop losses and economic sufferings in the agribusiness sector in the past [2]. In the modern practice, farmers use some sophisticated techniques to predict and curb plant diseases, which increases the health and productivity of crops. Conventional methods of disease detection are based on the experience of the eye and observation of the phenotypic characteristics of the organism to detect a pathogen. As stated by FAO, crop diseases may cut down the total agricultural production by about 20 -40 [3]. The traditional diagnostic methods and manual examination used to check tomato diseases tend to be imprecise and inconsistent, based on subjective evaluation and resulting in the disease being mistaken with another, when the symptoms structurally intersect with diseases [4]. In rural areas, specialist advice is not often available and this leads to delays in diagnosis. Symptom visibility is also affected by climatic conditions. Therefore, the manual practices are labor-intensive and create limitations to massive farming practices. There is an urgent need of scalable, quick and accurate diagnostic systems to detect tomato diseases early enough [5]. Automated methodologies reduce the possibility of human error, save time and provide uniformity in different operating settings. Classification model, specifically in machine learning, Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) can help classify tomato diseases using images. These models ensure that there is a better response time and minimization of yield losses through early disease detection; in addition, they can be used in the smooth deployment of edge and mobile platforms in real time [6]. The most commonly used algorithms in the classification of tomato diseases are Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs).

Experimental studies in machine-learning-based tomato diagnostics have highlighted the ability of such algorithms to reliably distinguish between a range of diseases. To improve the classification performance, scholars have discussed a variety of feature-extraction schemes, data-augmentation schemes and architectural variations [7].



Figure 1: A sample of both healthy and diseased tomato leaves.

Source: Authors, (2026).

The research gaps that exist include lack of adequate training data, issues with real world application, need to have stronger models which can generalize over heterogeneous environmental factors and tomato species. The future aims should be devoted to the creation of more effective disease-detection algorithms in real-time and introducing the multispectral imaging modalities into tomato disease classification systems [8]. One of the promising directions is the integration of visual data with supplementary sensor information temperature, humidity and soil measurements in order to obtain multimodal data synthesis. Also, the interpretable artificial-intelligence framework that will provide actionable insights about the decision-making processes will be designed in a way that will allow farmers and researchers to comprehend, assess, and rely upon the classification results and better [9]. CNNs can automatically extract and build important features using a series of specialized layers. Initial layers provide the edges and features of the visual image. When the data is processed through deeper layers, it produces an abstract concept that gives the disease pattern. Numerous leaf diseases impact the growth of tomato plants. Early and late blight create dark spots and can quickly cause leaves to die. The disease called Septoria causes spots on the leaves. These spots are usually small and produce marks on the leaves and fruits. A similar effect is caused by bacterial infection. Powdery Mildew creates a white dust on leaves, and Spider Mites are tiny pests that create yellow spots on the leaves, weakening the plant. Good monitoring and control of these conditions allows farmers to produce healthier tomatoes [10]. It must define the problem and importance of the research carried out, it presents a (not very extensive) review of the literature on the subject of the article, including the authors' contributions to the state of the art. If you use abbreviations or acronyms, first write the words that identify them and then, in parentheses, the acronym. This set also establishes the research question, the objectives of the work and hypothesis, if necessary, the importance and limitations of the study. Establishes the method used at work. It is written in the present tense.

II. THEORETICAL REFERENCE

An image dataset of the tomato plant was selected to capture a good representation of both healthy and diseased conditions during sensing events.

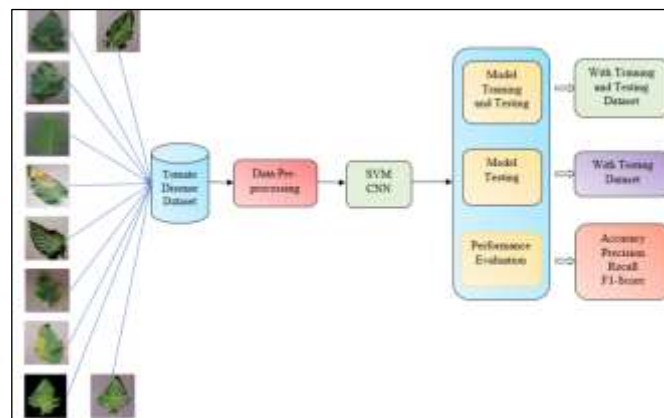


Figure 2: Suggested Tomato Disease Detection Process.

Source: Authors, (2026).

Before the model training, all images were pre-processed. All images were resized to 224 x 224 pixels. The pixel range of 0-1 was created. Then the data augmentation processes were applied. It includes techniques like brightness variations, random rotations, and random flips. Applying the conventional 80:20 ratio, the complete data set was partitioned accordingly [11]. The proposed methodology will utilize the convolutional neural network (CNN) approach. We used Adam optimizer and set the learning rate at 0.001 to train the model. we chose a batch size of 32. we employed early stopping processes to prevent the possibility of overfitting [12]. This technique is represented in fig.2.

II.1 MACHINE LEARNING TECHNIQUES

Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) are also typically used in the detection of plant disease in leaf images. The CNN-based models automatically learn features including lesion shape, color variation, texture anomaly and spatial pattern using plant leaf images and allow high accuracy of identifying complex diseases even with varying lighting and background conditions. Thus CNNs are commonly useful with large size of agricultural image data and real time field monitoring systems. In contrast, SVM-based methods depend on manually crafted features obtained by color histograms, Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) and then the features are classified by a separating hyperplane optimal. Although SVMs are fine with smaller and medium-sized datasets with reduced computational cost, their performance is very much affected by feature selection and parameter optimization. In general, CNNs are always better than SVMs in plant disease detection tasks of large scale, and SVMs are always a viable option when the system has low costs and limited training data.

II.1.1 Support Vector Machine

It is a way of classifying data in machine learning to analyze data and find relations in images. It obtains its classifications by separating hyperplanes in a particular n-dimensional space of multiple data groups. Those data points nearest to the points of separation, which support vectors are important to calculate the optimal separation [13]. The equation for the linear hyperplane described as:

$$wTx+b=0 \quad (1)$$

Where:

- w represents the normal vector to the hyperplane, indicating the direction perpendicular to it.
- b denotes the bias (offset) term, which determines how far the hyperplane is from the origin along the direction of w . In a two-dimensional space, this hyperplane becomes a straight line [14].

In the case of two dimensions, this hyperplane is merely a line; however the decision boundary forms a plane in higher dimensional spaces called as hyperplane. A model that does manual feature extraction and manual classification is used to classify tomato leaf diseases.

II.1.2 Convolutional Neural Network

This deep learning CNN architecture is used to classify images. It is very effective and efficient in image classification tasks, such as classifications in object detection, or disease identification in leaves, because it can learn salient features in the images, like edges, shapes, and patterns, independently.

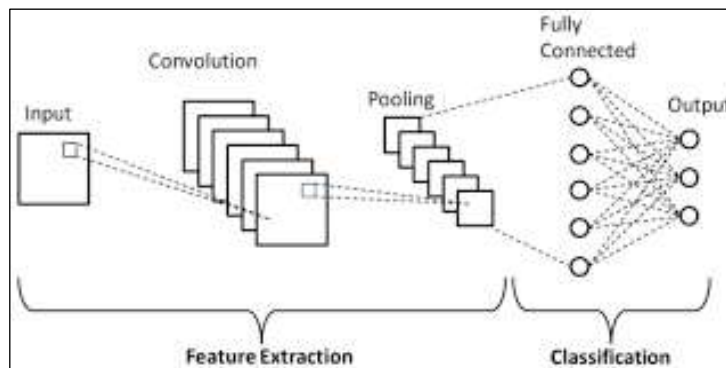


Figure 3: Architecture of CNN.

Source: Authors, (2026).

The formula is:

$$(F * K)(x, y) = \sum_i \sum_j F(x + i, y + j) K(i, j) \quad (2)$$

Where: X = input image

K = kernel/filter

S = feature map

This algorithm is also applied in the sphere of the classification of tomato diseases where there is no need of manual feature extraction since the algorithm will automatically learn the patterns and features based on the images of the leaves. The architecture of convolutional neural networks consists of layers of features. These layers include a series of convolutional, activation (ReLU), and pooling. These layers function together to identify low and high-level features. An important ability of CNNs is to extract and build features using a series of specialized layers. Initial layers provide the edges and features of the visual image. When the data is processed through deeper layers, it produces an abstract concept that gives the disease pattern [15].

III. MATERIALS AND METHODS

In the conventional Support Vector Machine (SVM) approach, preprocessed images are used to manually extract the features. It's a manual process. These features include color, texture, and shape. These features are used to establish a decision boundary for classifying the leaf diseases. The deep learning CNN model operates on its own. It does not need a human operator to identify the features.

It artificially expands the data set. The DenseNet121 architecture is employed. It is pre-trained on a large image library. This model directly reads the features from the original image data. The data augmentation helps the model to work well under various conditions, like lighting, leaf angles, etc. This process expands the data set by adding rotations, flips, and shifts [16-18]. Adam optimizer was applied to the model with a fixed learning rate of 0.001 to train it. There were 32 mini-batches which were used in training. The threat of overfitting was avoided by using early-stopping strategy [19], [20].

IV. RESULTS AND DISCUSSIONS

As a visual representation of the results of the classification, the confusion matrix is a summary of the results in the shortest possible form. It presents a count of correct and incorrect classifications, providing deeper insights into model behavior and enhancing the overall reliability of the evaluation process. It also exudes the nature of performance of the support vector machine classifier (SVM) and convolutional neural network architecture (CNN). The aggregate accuracy measure presents an efficient analysis of the general model performance. It is however misleading particularly when the data set is not balanced. In this instance, model effectiveness is measured with the help of the recall, precision, and F1 score. On top of such methods, there are other performance evaluations that can be acquired that define the models completely. The other method that can be used to measure the effectiveness of the models is with cross-validation, and this will assist in assessing the robustness of the models and also their versatility in a wide range of data landscapes; similarly it will also allow your performance measure to be reflective of its ability as opposed to an echo of a single dataset.

Table 1: Performance comparison of SVM and CNN.

Sr.No.	Algorithm	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score
1	SVM	87%	78%	0.78	0.78	0.78
2	CNN	96%	93%	0.93	0.93	0.93

Source: Authors, (2026).

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{(Precision+Recall)} \quad (6)$$

Where:

True Positive Where TP denotes model correctly predicting the diseased leaf is diseased

TN is True Negative Where the model falsely declares the healthy leaf as the healthy one.

FP is False Positive Where model estimates the diseased leaf, but it is healthy.

FN is False Negative Where model predicts the healthy leaf, but it is diseased.

This is a direct comparative analysis of the values of the trained SVM and CNN model to identify and categorise tomato leaf diseases, which applied four measures of performance: Accuracy, Precision, Recall, and F1-Score. CNN model has a better performance on all measures with about 93-93.35 percent accuracy, precision, recall, and F1-score. The overall SVM model performance was low (78% in all metrics). The findings also reveal that CNN model is more effective in classification, as compared to SVM model. The bar charts have been analysed and it can be seen that a high disparity is observed between CNN and SVM models. The graphical assessment confirms that the deep learning (CNN) outperforms the traditional machine learning (SVM) when it comes to tomato diseases recognition under the prerequisite of leaf pictures [21].

IV.1 RESULT ANALYSIS FOR SVM MODEL (BASED ON CONFUSION MATRIX):

A confusion matrix is a performance assessment tool that is universal and is applied to measure how effective a classification model has been using predicted class labels and actual ground truth labels. It is usually modeled as a square-sized matrix where each row in the matrix represents the real class and each column represents the predicted one. The four important results obtained by the use of the matrix are true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) that together present a detailed information of the classification behavior of the model. In contrast to overall accuracy, the confusion matrix discloses the performance on each of the classes, and thus it is possible to identify the pattern of misclassification, especially with multi-class problems such as tomato leaf disease detection, when some diseases manifest themselves with similar symptoms. The confusion matrix produces metrics like precision, recall, specificity and F1-score which are directly derived, and it is therefore a crucial tool of assessing the robustness of a model, its sensitivity to disease detection and its reliability in real-world agricultural applications.

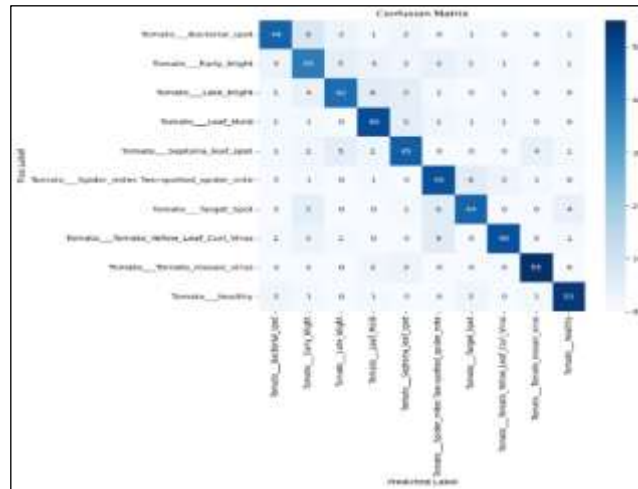


Figure 4: Confusion Matrix of SVM Model Classifier.

Source: Authors, (2026).

According to the confusion matrix analysis of the Support Vector Machine (SVM) model shows good overall performance in terms of the classification, although there are still some drawbacks, especially those that are not visually distant in their manifestations. The model showed interesting performance in the category of Healthy Leaves (53 correct predictions) and Tomato Mosaic Virus (55 correct predictions) showing that the SVM can be used to effectively distinguish between categories where visual differences are evident when trained using features which are sufficiently distinct. However, in fungal infections like Early Blight, Late Blight, Septoria Leaf Spot and Leaf Mold, the model is classified with a lot of misclassification. Such classes often perplex each other mainly because of the visual similarities among the groups, which include beauty in presence of spots of leaves, atmospheric varieties and lesions. The Support Vector Machine model confusion matrix analysis depicts good results in terms of the classification, though with flaws, particularly in the classes that represented visually close symptoms. The model has been shown to perform quite well in such categories as Healthy Leaves (53 correct predictions) and Tomato Mosaic Virus (55 correct predictions), which highlights the fact that the SVM can be a useful tool in distinguishing between categories with specific visual properties when the features used are the best.

IV.2 RESULT ANALYSIS FOR CNN MODEL (BASED ON CONFUSION MATRIX):

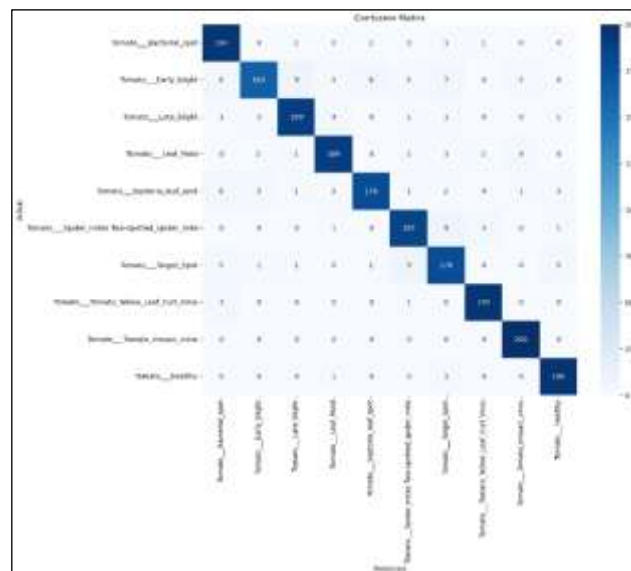


Figure 5: Confusion Matrix of CNN Model Classifier.

Source: Authors, (2026).

The study of the confusion matrix gives an exhaustive evaluation of the results produced by the Convolutional Neural Network (CNN) model. The related results demonstrate that the model has extraordinarily high performance which is reflected by impressive accuracy and consistency with all ten classifications involving tomato leaf disease and healthy foliage. The CNN model was able to determine most samples in each category with a near-perfect accuracy in particular classes (200 / 200 were correct in Tomato Mosaic Virus; 195 / 200 were correct in Tomato Yellow Leaf Curl Virus; 196 / 200 were correct in Healthy Leaves). It proves the ability of the model to extract and learn different patterns from an image. But there is a small overlap in specific classes. It can lead to misclassification. But it is negligibly small [22]. The CNN model substantially outperformed the SVM model in these cases, particularly with classes that showed visual complexity and overlap. The strength of CNN for these classifications relies on the model's ability to learn features automatically through deep learning rather than rely on extracting features by hand.

All in all, the CNN model showed great performance in terms of providing high classification rates, low misclassification rates, and accurate assessments of both distinct and subtly different disease classes. It also confirms CNN's applicability for tasks in the classification of complex images, such as tomato leaf diseases. The CNN model was trained for a total of 7 epochs, with both training and testing.

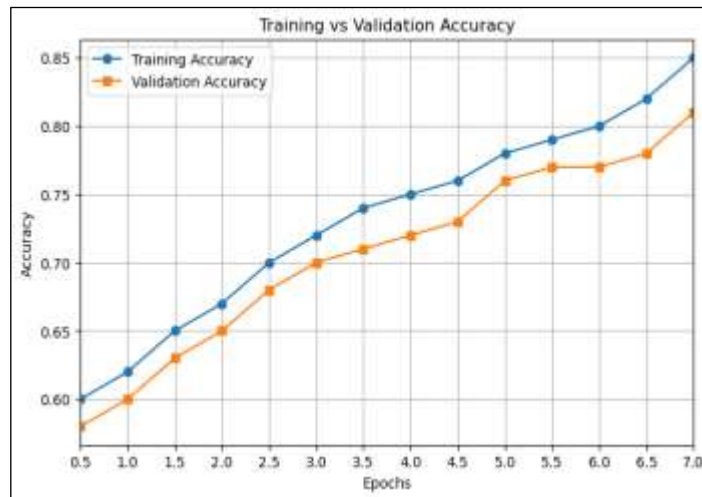


Figure 6: Training vs validation accuracy of the models.

Source: Authors, (2026).

In Figure 6, we can see the training and validation accuracy of the model proposed throughout epochs, specifically between 0.5 to 7. The graph demonstrates consistently increasing training and testing accuracy as the number of epochs increases. Starting with an initial training accuracy of approximately 0.60, the model gains an accuracy of around 0.85 by epoch 7, indicating that the model has effectively learned from the training data. Likewise, the initial testing accuracy was approximately 0.58 and increased to 0.81 by the final epoch, demonstrating that the model is performing well and showing little to no overfitting of the data. Finally, we can see that the slow convergence of the curves in training and validation demonstrates that the model is reliable and effective when it comes to correctly classifying tomato leaf diseases.



Figure 7: Training vs validation loss of the models.

Source: Authors, (2026).

Figure 7 shows the training and validation loss trajectories for the proposed model for 7 epochs. The x-axis is the epochs spanning an interval from 0.5 to 7, and the y-axis displays the corresponding loss values. The blue line denotes the training loss, and the orange line represents the validation loss. The training and validation losses are both clearly decreasing as the number of epochs increases, indicating the model has effective learning and improvement in the training and validation datasets. The training and validation loss trajectories are closely aligned, suggesting generalization well without significant overfitting.

V. CONCLUSIONS

Typically, the performance comparison part of the results and analysis includes a complete comparison of SVM and CNN. Important metrics: accuracy, precision, recall, and F1-score are tested for every algorithm. These metrics help provide a broad perspective on the performance of each algorithm, enabling researchers to analyze an algorithm on all aspects of classification or prediction tasks. Confusion matrices give a detailed description of the type of errors that an algorithm makes. The CNN model resulted in a testing accuracy of 93.35% by epoch seven. While the SVM model reached an accuracy of 78%. Overall, the CNN model proved to be more reliable for the tomato disease classification.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Rajani P.K., Kishori S.Telaware, Rohini V.Daund.

Methodology: Kishori S.Telaware.

Investigation: Rajani P.K., Kishori S.Telaware.

Discussion of results: Rajani P.K., Kishori S.Telaware, Rohini V.Daund.

Writing – Original Draft: Kishori S.Telaware.

Writing – Review and Editing: Kishori S.Telaware, Rohini V.Daund.

Resources: Kishori S.Telaware, Rohini V.Daund.

Supervision: Rajani P.K.

Approval of the final text: Rajani P.K.

VII. REFERENCES

- [1] A. K. Aggarwal, "Biological tomato leaf disease classification using deep learning framework," *International Journal of Biology and Biomedical Engineering*, vol. 16, pp. 241–244, 2022, doi: 10.46300/91011.2022.16.30.
- [2] A. R. Al-Shamasneh and R. W. Ibrahim, "Classification of tomato leaf images for detection of plant disease using conformable polynomials image features," *MethodsX*, vol. 13, p. 102844, 2024, doi: 10.1016/j.mex.2024.102844.
- [3] M. Bhandari, T. B. Shahi, A. Neupane, and K. B. Walsh, "BotanicX-AI: Identification of tomato leaf diseases using an explanation-driven deep-learning model," *Journal of Imaging*, vol. 9, no. 2, p. 53, 2023, doi: 10.3390/jimaging9020053.
- [4] E. Cengil and A. Çınar, "Hybrid convolutional neural network based classification of bacterial, viral, and fungal diseases on tomato leaf images," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, p. e6617, 2021, doi: 10.1002/cpe.6617.
- [5] P. Chaturvedi, D. Bishnoi, S. Manekar, and A. Kumari, "Disease identification and classification from pearl millet leaf images using machine learning techniques," in *Advances in Agricultural Informatics*. Hershey, PA, USA: IGI Global, 2024, pp. 232–243, doi: 10.4018/979-8-3693-1062-5.ch013.
- [6] F. De Vita, S. K. Das, D. Bruneo, D. Giacalone, G. Nocera, and V. Tomaselli, "Quantitative analysis of Deep Leaf: A plant disease detector on the smart edge," in *Proc. IEEE Int. Conf. Smart Computing*, 2020, pp. 49–56, doi: 10.1109/SMARTCOMP50058.2020.00027.
- [7] W. B. Demilie, "Plant disease detection and classification techniques: A comparative study of the performances," *Journal of Big Data*, vol. 11, Art. no. 5, 2024, doi: 10.1186/s40537-023-00863-9.
- [8] M. Francis and C. Deisy, "Disease detection and classification in agricultural plants using convolutional neural networks—A visual understanding," in *Proc. IEEE SPIN Conf.*, 2019, doi: 10.1109/SPIN.2019.8711701.
- [9] H. Ghosh et al., "Advanced neural network architectures for tomato leaf disease diagnosis in precision agriculture," *Discover Sustainability*, vol. 6, no. 1, 2025, doi: 10.1007/s43621-025-01149-1.
- [10] A. E. Hassanién, T. Gaber, U. Mokhtar, and H. Hefny, "An improved moth flame optimization algorithm based on rough sets for tomato diseases detection," *Computers and Electronics in Agriculture*, vol. 136, pp. 86–96, 2017, doi: 10.1016/j.compag.2017.02.026.
- [11] B. Khan et al., "Bayesian optimized multimodal deep hybrid learning approach for tomato leaf disease classification," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-72237-x.
- [12] J. Kolhe, G. Deshpande, G. Patel, and P. K. Rajani, "Crop detection using various machine learning classification algorithms," in *IoT with Smart Systems: Proc. ICTIS 2022*. Singapore: Springer, 2022, pp. 495–502, doi: 10.1007/978-981-19-2159-1_49.
- [13] A. Mir and J. A. Nasiri, "KNN-based least squares twin support vector machine for pattern classification," *Applied Intelligence*, vol. 48, no. 12, pp. 4551–4564, 2018, doi: 10.1007/s10489-018-1225-z.
- [14] V. Pandiyaraju et al., "Improved tomato leaf disease classification through adaptive ensemble models with exponential moving average fusion and enhanced weighted gradient optimization," *Frontiers in Plant Science*, vol. 15, Art. no. 1382416, 2024, doi: 10.3389/fpls.2024.1382416.
- [15] R. P. K. Rajani, V. Deshmukh, S. U. Bhandari, R. Raut, and R. Kharat, "Rice leaf disease detection using convolutional neural network," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 10S, pp. 512–516, 2023, doi: 10.17762/ijritcc.v11i10s.7687.
- [16] A. Pawar, "Brain tumor detection using machine learning," *International Journal of Scientific Research in Engineering and Management*, vol. 9, no. 3, pp. 1–9, 2025, doi: 10.55041/ijrsrem42421.
- [17] S. Saifullah and R. Drezewski, "Redefining brain tumor segmentation: A cutting-edge convolutional neural networks-transfer learning approach," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 2583–2591, 2024, doi: 10.11591/ijece.v14i3.pp2583-2591.
- [18] M. Scortichini, "Sustainable management of diseases in horticulture: Conventional and new options," *Horticulturae*, vol. 8, no. 6, p. 517, 2022, doi: 10.3390/horticulturae8060517.
- [19] L. Tan, J. Lu, and H. Jiang, "Tomato leaf diseases classification based on leaf images: A comparison between classical machine learning and deep learning methods," *AgriEngineering*, vol. 3, no. 3, pp. 542–558, 2021, doi: 10.3390/agriengineering3030035.
- [20] R. Thangaraj, S. Anandamurugan, and V. K. Kaliappan, "Automated tomato leaf disease classification using transfer learning-based deep convolution neural network," *Journal of Plant Diseases and Protection*, vol. 128, no. 1, pp. 73–86, 2020, doi: 10.1007/s41348-020-00403-0.
- [21] N. Zhang, Y. Deng, H. Wu, X. Han, and H. Zhu, "Tomato disease classification and identification method based on multimodal fusion deep learning," *Agriculture*, vol. 12, no. 12, p. 2014, 2022, doi: 10.3390/agriculture12122014.
- [22] H. Sun et al., "Multi-kernel inception aggregation diffusion network for tomato disease detection," *BMC Plant Biology*, vol. 24, no. 1, 2024, doi: 10.1186/s12870-024-05797-9.