E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 15-02-2025
Accepted: 21-03-2025

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 3; Issue 2; 2025; Page No. 224-228

Disease prediction in tomato leaf using convolutional neural network Algorithm

[1]V Dharshini and [2]Dr. AS Arunachalam

[1]PG Scholar, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India
[2]Professor, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India

DOI: https://doi.org/10.5281/zenodo.15586438

Corresponding Author: V Dharshini

Abstract

This study focuses on the application of Convolutional Neural Networks (CNNs) for the detection and classification of diseases in tomato leaves, a critical issue affecting agricultural productivity. Leveraging a dataset of labeled leaf images, we employed a CNN architecture to automate the diagnosis of various diseases such as early blight, late blight, and bacterial spot. The model was trained on a diverse set of augmented images to enhance its robustness and generalization capabilities. Performance metrics, including accuracy and loss, were evaluated on a separate test set, demonstrating the model's efficacy in accurately identifying disease types. The findings indicate that CNNs can significantly aid farmers and agricultural professionals in early disease detection, facilitating timely interventions and improved crop management. This research underscores the potential of deep learning techniques in precision agriculture, paving the way for future advancements in plant disease diagnostics.

Keywords: Disease, Prediction, tomato, convolutional, neural network, Algorithm, CNNs

Introduction

Tomatoes rank among the most extensively grown and consumed vegetables worldwide. Never the less, their yield is considerably impacted by a range of plant diseases, which can result in diminished production and financial setbacks for farmers. Timely and precise identification of these diseases is crucial for implementing effective management and prevention measures.

Historically, the process of diagnosing diseases in tomato plants has depended on the manual assessment by agricultural specialists, a method that is both time-consuming and labor-intensive, with a high likelihood of human error. However, recent advancements in artificial intelligence (AI) and deep learning have led to the development of automated systems that serve as robust tools for disease detection through image analysis. In this initiative, we introduce a deep learning methodology for predicting tomato leaf diseases utilizing a Convolutional Neural Network (CNN) algorithm. CNNs represent a category of deep neural networks tailored for tasks involving image classification and object detection. They autonomously learn to extract relevant features from input images, thereby removing the necessity for manual feature extraction. By training the CNN model on a collection of tomato leaf images-comprising both healthy and diseased specimens-the system can accurately recognize various diseases, including early blight, late blight, and leaf mold, among others. This automated method not only improves the speed and efficiency of disease detection but also aids farmers and agricultural experts in making prompt decisions. The primary objective of this research is to create a dependable and precise model that can facilitate early disease prediction, thereby enhancing crop management and boosting agricultural productivity.

Literature Review

Tomatoes rank among the most crucial horticultural crops cultivated globally, with their yield being significantly influenced by various diseases. Conventional methods for disease identification, which depend on expert visual assessments, tend to be labor- intensive, subjective, and frequently imprecise. However, recent advancements in deep learning and computer vision, particularly through Convolutional Neural Networks (CNNs), have made it possible to automate and enhance the accuracy of disease detection in tomato plants. In their 2016 study, Mohanty et al. employed deep learning techniques to identify plant diseases using a dataset comprised of leaf images. They utilized pre-trained models such as Alex Net and Google Net, achieving high accuracy in classifying 38 distinct categories of plant diseases. This research highlighted the effectiveness of transfer learning in the realm of plant disease detection. Ferentinos (2018) [2] investigated various CNN architectures, including Alex Net, VGG, and LeNet, for recognizing plant diseases. The study encompassed 58 different categories, including diseases affecting tomato leaves, and achieved classification accuracy exceeding 99%. This research confirmed that CNNs can surpass traditional machine learning approaches. Sladojevic et al. (2016) [1] created a deep CNN model capable of recognizing 13 plant diseases from leaf images. Their methodology required minimal preprocessing and utilized raw RGB images, demonstrating that CNNs can autonomously extract significant features from images. Brahimi et al. (2017) [5] advocated for the use of transfer learning with CNNs such as Alex Net and Google Net for detecting tomato diseases. They reported high classification accuracy and underscored the significance of data augmentation to address the challenges posed by limited datasets. Conducted a comparative analysis of various CNN architectures, including Res Net, InceptionV3, and Dense Net, for plant disease classification. Their results revealed that Dense Net outperformed other models in terms of both accuracy and computational efficiency. Zhang et al. (2020) [6] introduced a lightweight CNN model designed for the real-time detection of tomato diseases on edge devices, achieving a balance between accuracy and performance and performance, making it appropriate for implementation in settings with limited resources, such as embedded devices or mobile phones.

The difficulties in applying CNN models in the actual environment, including different illumination conditions, background noise, and leaf occlusion, were emphasized by Barbedo (2019) [8]. The study recommended combining data augmentation and picture preprocessing methods to increase model robustness.

KeyTakeaways: BecauseCNNs automatically extract spatial data, they perform better than conventional machine learning models.

Using pre-trained models for transfer learning works well when there aren't many labeled datasets.

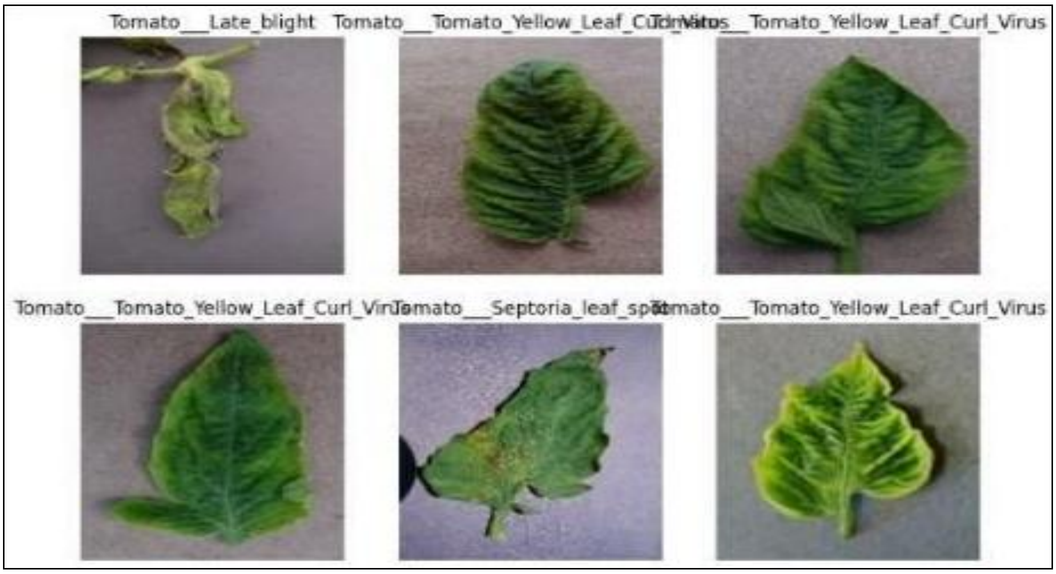
Proposed Solution

The suggested approach for predicting tomato leaf diseases utilizes the power of Convolutional Neural Networks (CNNs) to effectively recognize and categorize different tomato diseases based on images. This system aims to automate the detection process, delivering prompt and dependable results to farmers and agricultural specialists. The overall procedure is organized into structured phases to guarantee the creation of a high-performance and practical solution The objective of this project is to develop an efficient and accurate system for predicting diseases in tomato leaves using image processing and Convolutional Neural Networks (CNN). The proposed solution involves the following major steps:

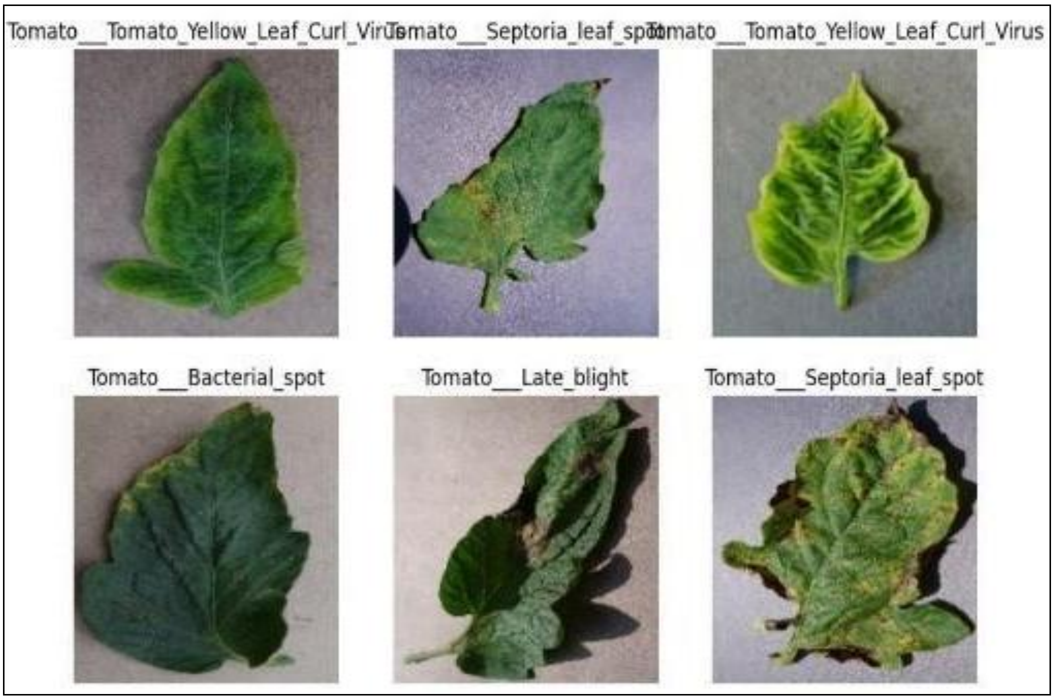
Data Acquisition: The initial phase entails gathering a comprehensive and varied dataset of tomato leaf images. For this project, we employ the PlantVillage dataset, which comprises highresolution, annotated images of tomato leaves impacted by various diseases. The dataset encompasses conditions such as:

- Tomato Early Blight
- Tomato Late Blight
- Tomato Yellow Leaf Curl Virus
- Tomato Mosaic Virus
- Leaf Mold
- Septoria Leaf Spot
- Healthy Tomato Leaves

This dataset is selected for its substantial size, balanced representation of classes, and its prevalent application in agricultural AI research, rendering it suitable for training a deep learning model.

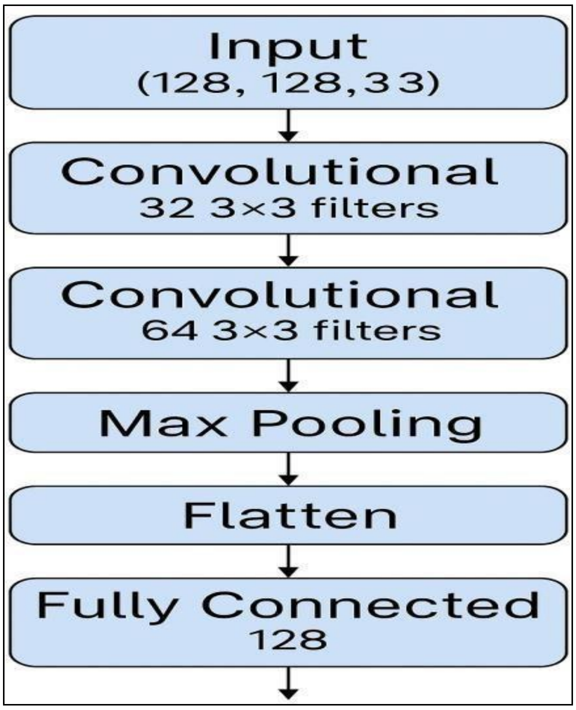


3.2 Data Preprocessing



Raw images cannot be directly input into the CNN. Consequently, the subsequent phase involves preprocessing the data to render it suitable for training.
Image Resizing: All images are adjusted to a uniform dimension (e.g., 224x224) to standardize the input size throughout the network.
Normalization: Pixel values are scaled to a range of 0 to 1 to maintain consistency and enhance convergence speed during the training process.
Data Augmentation: To mitigate overfitting and improve the model's generalization capabilities, we implement random transformations, including horizontal and vertical flips, rotations, zooming, brightness adjustments, and translations.
Label Encoding: Disease categories (e.g., 'Late Blight') are transformed into numerical labels through one-hot encoding for multi- class classification.
These preprocessing techniques ensure that the data is varied, consistent, and informative, which significantly aids in enhancing model learning.

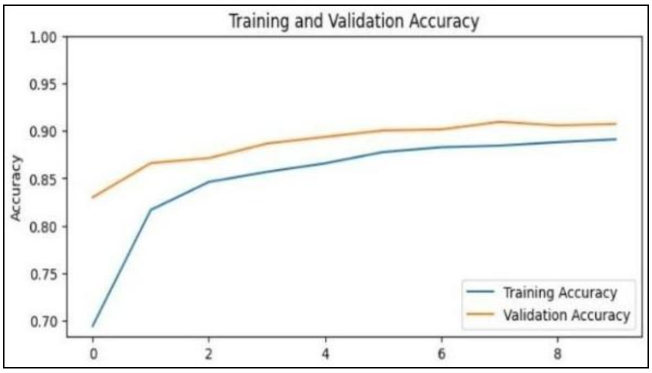
CNN Model Architecture
The Convolutional Neural Network (CNN) serves as the fundamental element of the solution. It is engineered to autonomously extract intricate features from images, eliminating the necessity for manual feature engineering. The architecture proposed comprises the following components:
Input Layer: Accepts images that have been preprocessed.
Convolutional Layers: These layers are responsible for extracting features such as edges, textures, and patterns from the input images. They are succeeded by ReLU (Rectified Linear Unit) activation functions, which introduce non-linearity into the model.
Pooling Layers: Typically employing Max Pooling, these layers down-sample the feature maps, thereby decreasing the computational burden.
Dropout Layers: These layers mitigate the risk of overfitting by randomly omitting a selection of activations during the training process.
Fully Connected Layers (Dense Layers): These layers integrate the features extracted for the purpose of classification.



Output Layer: This layer utilizes the Softmax activation function to generate probabilities for each class of disease.
To improve accuracy and expedite training, we also explore the application of transfer learning techniques. Pre-trained CNN models such as VGG16, ResNet50, or MobileNetV2 are finetuned on our dataset, leveraging the knowledge acquired from extensive image datasets like ImageNet.

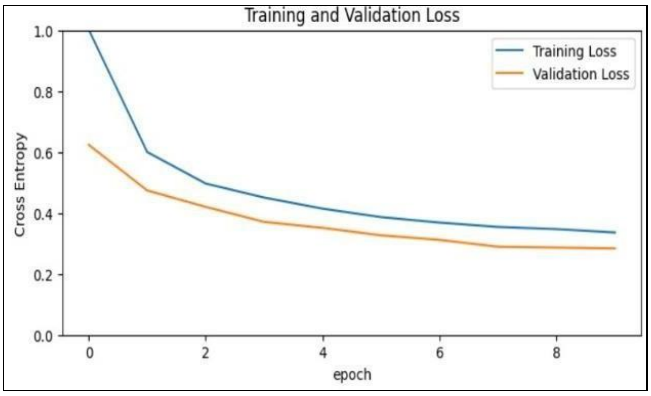
3.4 Model Training

The dataset is divided into three segments: Training Set (70%) – This portion is utilized for instructing the model.
Validation Set (20%) – This segment is employed to adjust hyperparameters and track overfitting.
Test Set (10%) – This part is reserved for assessing the final model's performance.



The model training process involves

Loss Function: Categorical Cross-Entropy Optimizer: Adam, selected for its adaptive learning features.
Batch Size: Commonly set to 32 or 64.
Epochs: The model undergoes training for a range of 20 to 50 epochs, depending on when convergence is achieved.



To ensure effective performance monitoring during training, we incorporate early stopping, which ceases training if there is no improvement in validation loss, thereby minimizing the likelihood of overfitting.

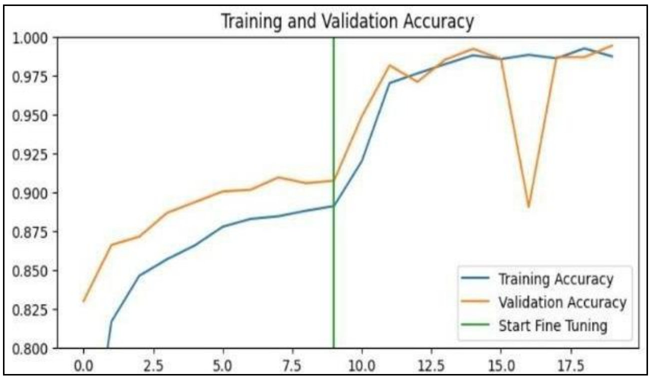
Model Evaluation

After training the model, it is assessed on the test set through a range of performance metrics:

Accuracy: The overall correctness of the predictions made. Precision and Recall: Evaluate the model's effectiveness in identifying particular diseases. F1 Score: The harmonic mean of precision and recall, providing a balance between the two.

Confusion Matrix: A visual representation of classification performance that emphasizes misclassified instances.

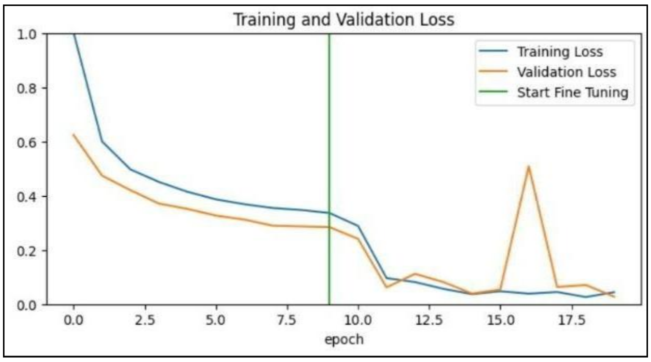
AUC-ROC Curve (optional): Demonstrates the model's capability to differentiate between various classes.



These metrics collectively offer a thorough insight into the model's performance across diverse disease categories.

Optimization and Fine-tuning

In cases where performance does not meet expectations, additional enhancements can be implemented through the following methods:
Hyperparameter Adjustment: Modifying learning rates, batch sizes, dropout rates, and layer structures. Model Pruning: Eliminating redundant filters and neurons to decrease model size and enhance prediction speed.
Enhancements in Transfer Learning: Further refining additional layers of the pre-trained model to improve feature adaptation.
Ensemble Techniques: Merging predictions from various models to boost overall accuracy.
These strategies contribute to the development of a more dependable, quicker, and efficient model that is well-suited for practical applications.



These strategies contribute to the development of a more dependable, quicker, and efficient model that is well-suited for practical applications.

Future Prospects: The initiative "Disease Prediction in Tomato Leaf Using CNN Algorithm" showcases a significant application of deep learning within the agricultural sector. While the existing model demonstrates commendable performance, there are numerous opportunities for future improvements and developments:

Dataset Enhancement

Utilizing a more extensive and varied dataset that includes images from different geographical locations, seasons, and lighting conditions can enhance the model's generalizability and precision.

Multi-Disease Identification

Subsequent models could be designed to identify multiple diseases concurrently or to address scenarios where several infections impact a single leaf.

Real-time Mobile Application

Incorporating the model into a mobile or web application would enable farmers to take pictures of leaves and receive immediate disease predictions while on the move.

IoT Integration

Linking the prediction system with IoT-based sensors that monitor environmental factors (such as humidity and temperature) can yield more precise predictions and recommend preventive actions.

Explainable AI (XAI)

Adopting explainability methods (like Grad- CAM) can assist users in comprehending the rationale behind the model's predictions, thereby enhancing trust and transparency.

Automated Treatment Recommendations

The system could be improved to offer disease-specific treatment suggestions or facilitate connections with local agricultural specialists.

Support for Additional Crops

Broadening the model's capabilities to include various crops can enhance its usefulness and encourage wider adoption among farmers. Cloud Deployment Implementing the system on cloud platforms will ensure scalability and accessibility, particularly in remote areas with limited computational resources.

Conclusion

The project titled "Disease Prediction in Tomato Leaf Using CNN Algorithm" effectively illustrates the use of deep learning methodologies within agriculture, particularly in the detection of diseases affecting tomato plants through image classification. Utilizing a Convolutional Neural Network (CNN), the system successfully identified various diseases in tomato leaves, such as early blight, late blight, and leaf mold, among others. The findings demonstrate that CNNs excel in learning and identifying intricate features within leaf images, achieving notable accuracy and reliability in disease prediction. This automated method not only diminishes the reliance on manual inspections but also aids farmers in making prompt and informed decisions regarding plant care, which can lead to improved crop yield and quality.

In summary, the project advances the creation of intelligent agricultural tools that can improve disease management and precision farming practices. Future developments may involve expanding the dataset, incorporating the model into mobile applications, and enabling real-time disease detection in the field through the use of drones or IoT-based systems.

References

1. Mohanty SP, Hughes DP, Salathé M. Implementation of deep learning techniques for detecting plant diseases through image analysis. Frontiers in Plant Science. 2016.
2. Ferentinos KP. Application of deep learning models for the identification and diagnosis of plant diseases. Computers and Electronics in Agriculture. 2018;145:311–318.
3. Kaggle. PlantVillage Tomato Leaf Disease Image Collection [Internet]. Available from: https://www.kaggle.com/datasets/emmarex/plantdisease
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014. Available from: https://arxiv.org/abs/1409.1556
5. Chollet F. Deep Learning with Python. Manning Publications; c2017.
6. Zhang Y, Chen C, Zhu S, Shu C, Wang D, Song J, *et al*. Isolation of 2019-nCoV from a stool specimen of a laboratory-confirmed case of the coronavirus disease 2019 (COVID-19). China CDC weekly. 2020;2(8):123.
7. Angamuthu T, Arunachalam AS. A comprehensive survey on the revolution of plant disease detection and diagnosis through automated image processing techniques with CNN and RNN. Singaporean Journal of Scientific Research (SJSR). 2024;16(1).
8. Barbedo JG. Plant disease identification from individual lesions and spots using deep learning. Biosystems engineering. 2019;180:96-107.