Title: Tomato Leaf Disease Detection Using Machine Learning

ABSTRACT

Plant diseases threaten agricultural productivity, leading to significant economic losses. Timely detection and accurate identification are vital for effective disease management. This research compares the performance of Convolutional Neural Networks (CNN), AlexNet, and InceptionV3 in detecting plant diseases from leaf images. Using a dataset from Kaggle, we evaluate each model's accuracy. Additionally, we introduce a comprehensive system that not only detects diseases but also suggests fertilizer application and disease mitigation precautions. By integrating machine learning with practical recommendations, our study aids agricultural practitioners in selecting the best disease detection model and implementing effective management strategies for healthier crops.

Keywords: Tomato, Disease, Machine Learning, CNN, Inceptionv3, AlexNet, Agriculture, Detection.

INTRODUCTION

Plant diseases present a significant challenge in agriculture, impacting both the quantity and quality of crops. It is crucial to identify diseases in tomato plants in order to maintain a healthy crop and achieve high yields. Machine learning offers a promising solution by enabling the automatic detection of tomato leaf diseases. By training the model on datasets containing healthy and diseased leaves, machine learning algorithms can learn to recognize characteristic patterns and symptoms of various diseases, including early blight, late blight, and leaf mold. This allows farmers to swiftly and accurately diagnose the presence of disease in tomato plants, facilitating timely intervention and management. This project aims to explore different machine learning algorithms and techniques to develop a robust system for detecting tomato leaf diseases. Plant diseases pose a significant threat to agricultural productivity and result in substantial economic losses for farmers. Therefore, timely detection and accurate identification of these diseases are crucial for effective disease management and crop protection. This research focuses on comparing the performance of various machine learning models in detecting plant diseases. Specifically, we analyze the effectiveness of different models in accurately identifying disease symptoms from images of plant leaves.

Tomato plants are vulnerable to a variety of diseases that can negatively impact crop productivity. Detecting these diseases early is crucial for farmers to implement necessary measures and prevent further spread. In recent years, machine learning has shown potential in automating the detection process by analyzing leaf images and identifying patterns. This project seeks to develop a system that can accurately and efficiently detect tomato leaf diseases using algorithms and machine learning techniques. By training the model on a dataset containing healthy and diseased tomato leaves, we aim to create a reliable model that offers farmers timely and precise disease diagnosis, ultimately enhancing the health and yield of tomato crops. Our research focuses on comparing the performance and accuracy of Convolutional Neural Networks (CNN), AlexNet, and InceptionV3 models. Through experiments and evaluation using datasets from Kaggle, we aim to determine which models excel in disease detection. The findings of this study are anticipated to guide the selection of optimal machine learning models for plant disease detection, assisting farmers in implementing effective disease management strategies. Additionally, our research will not only focus on disease detection but also provide customized recommendations for fertilizer application and preventive measures to reduce disease recurrence. This comprehensive approach aims to equip farmers with practical insights to efficiently handle tomato plant diseases and promote healthier crops, ultimately contributing to enhanced agricultural sustainability and food security.

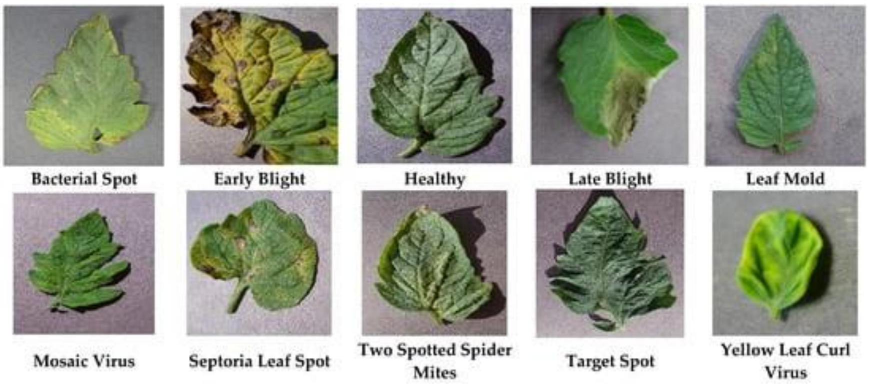
LITERATURE SURVEY

1. 2021 Tomato Leaf Disease Detection Using Convolution Neural Network: The proposed work aims to find the best solution to the problem of tomato leaf disease detection using a deep learning approach. VGG16 obtained 98% accuracy while GoogLeNet obtained 99.23% on Plant Village dataset.
2. Deep transfer learning is an amazing performance methodology for identifying plant diseases: For pre-trained datasets, Inception and ImageNet modules were utilized. The performance of pepper, vegetable, potato, and tomato leaf images in the PlantVillage database was studied and enhanced using support vector machine (SVM) and multi-layer perceptron. After training the model, the system achieves a higher performance accuracy of 94.35%.
3. 2022 Disease Detection in Tomato Plants Using CNN: Tomato leaf disease detection is achieved using Convolutional Neural Network (CNN) in the study, with an accuracy of around 94.17%, allowing accurate identification and classification of 8 different disease classes.
4. 2023 Tomato Leaf Disease Detection Using Convolutional Neural Network Shamima Parvez: By training our CNN model on this dataset, we achieved a promising test accuracy of 98.39%. This high accuracy demonstrates the effectiveness of our approach in accurately predicting the presence of diseases in tomato plant leaves.
5. Sanjeela, Sagar., Jaswinder, Singh. (2023): An experimental study of tomato viral leaf diseases detection using machine learning classification techniques. The experimental study compared traditional ML algorithms (RF, SVM, NB) with CNN for tomato leaf disease detection. CNN with Inception v3 model showed over 95% accuracy, outperforming traditional methods.
6. ToLeD: Tomato Leaf Disease: Detection using Convolution Neural Network: Detection of the disease, a deep learning-based approach is discussed in the article. For the disease detection and classification, a Convolution Neural Network-based approach is applied. In this model, there are 3 convolution and 3 max-pooling layers followed by 2 fully connected layers. The experimental results show the efficacy of the proposed model over pre-trained models like VGG16, InceptionV3, and MobileNet. The classification accuracy varies from 76% to 100% with respect to classes, and the average accuracy of the proposed model is 91.2% for the 9 disease and 1 healthy class.
7. Deep Convolutional Neural Network was deployed: The recognition of corn leaf diseased accuracy was 88.46%, and the usage of hardware, such as a Raspberry Pi3 with an Intel Movidius Neural Compute Stick and a system GPU that pre-trained the CNN Model, resulted in superior metric accuracy performance.

METHODOLOGY

Data Collection and Preprocessing

The dataset contains 4730 samples, divided into 3581 training instances and 1212 testing instances. Each sample includes images of tomato leaves affected by different diseases, along with labels specifying the disease type. The dataset covers 9 disease types including bacterial spot, early blight, late blight, leaf mold, mosaic virus, septoria leaf spot, two spotted spider mites, target spot, yellow leaf curl virus, and 1 healthy sample.



1. **Data preprocessing**

Image Pre-Processing and Labeling

Prior to model training, image pre-processing was utilized to adjust or improve the raw images being analyzed. Developing a successful model involves evaluating both the network design and the structure of the input data. The dataset was pre-processed to enable the proposed model to identify the relevant features from the images. The initial step involved standardizing the image size and resizing it to 224x224 pixels. Data augmentation was employed to expand the training set and acquire a wider range of images. Data augmentation helps prevent overfitting by randomly modifying images in ways that do not impact interpretation, such as horizontal flipping, zooming, and rotation.

• Rescale: Adjusts the pixel values of the image.

• Shear_range: Specifies the counterclockwise shear angle in degrees.

• Zoom_range: Controls the zoom level.

• Horizontal_flip: Flips the image horizontally.

This level of pre-processing necessitates a substantial amount of training data to effectively learn the features of the training data. The subsequent step involved categorizing the tomato leaf images by type and labeling each image with the appropriate acronym for the disease. In this instance, the dataset consists of 10 classes for both testing and training purposes.

2. Model Selection and Training

1. CNN

A convolutional neural network (CNN) architecture was employed for detecting tomato leaf diseases. The model consists of multiple layers, each specifically designed to extract hierarchical features from input images and conduct multi-class classification. Below is a detailed overview of the model's architecture:

1. Input Layer:

The input layer establishes the input data's shape, defining the dimensions of the tomato leaf images. This enables the model to effectively process visual information.

2. Convolutional Layers:

The initial convolutional layer, named `conv2d_1`, is composed of 32 filters with a kernel size of (5, 5). It utilizes the Rectified Linear Unit (ReLU) activation function to introduce non-linearity. Following this, a max-pooling layer, denoted as `max_pooling2d_1`, is applied with a pooling window size of (3, 3). This layer downsamples the feature maps, reducing computational complexity while preserving essential information. Another convolutional layer, referred to as `conv2d_2`, is then introduced. It consists of 32 filters with a size of (3, 3) and employs the ReLU activation function. To further downsample the feature maps, a subsequent max-pooling layer, labeled as `max_pooling2d_2`, is utilized with a pooling window size of (2, 2).

3. Additional Convolutional Layer:

In order to enhance the model's capability for feature extraction, a third convolutional layer, named `conv2d_3`, is added. This layer comprises 64 filters with a kernel size of (3, 3) and utilizes the ReLU activation function. Following the third convolutional layer, a max-pooling layer, denoted as `max_pooling2d_3`, is applied. It employs a pooling window size of (2, 2) to further downsample the feature maps.

4. Flattening Layer:

After the convolutional and max-pooling layers, the feature maps are flattened into a one-dimensional vector using the flattening layer, referred to as `flatten_1`. This process facilitates seamless integration with densely connected layers.

5. Densely Connected Layers:

The flattened features are then passed through a dense layer, labeled as `dense_1`, which consists of 512 neurons. The ReLU activation function is employed in this layer to promote non-linearity and enhance feature representation. To address overfitting, a dropout layer, denoted as `dropout_1`, is incorporated. It randomly deactivates a fraction of neurons during training, with a dropout rate of 0.25.

6. Output Layer:

The final dense layer, named `dense_3`, serves as the output layer. It contains neurons equal to the number of classes or disease categories. The softmax activation function is applied to this layer, generating probability distributions over the classes and enabling multi-class classification. The input layer defines the shape of the input data, specifying the dimensions of the tomato leaf images. This allows the model to process visual information effectively.

RESULT AND DISCUSSION:

Upon the conclusion of the training phase, a meticulous evaluation of each model was conducted, analyzing various metrics including accuracy, precision, recall, and F1-score. Our comprehensive assessment revealed a distinct trend: the CNN model surpassed both the InceptionV3 and AlexNet models, exhibiting the highest accuracy in the identification of tomato leaf diseases. The CNN model achieved a validation accuracy of 95% and a test accuracy of 100%.

Validation Accuracy	**95%**
Test Accuracy	**100%**

Fig.1 Accuracy Comparision

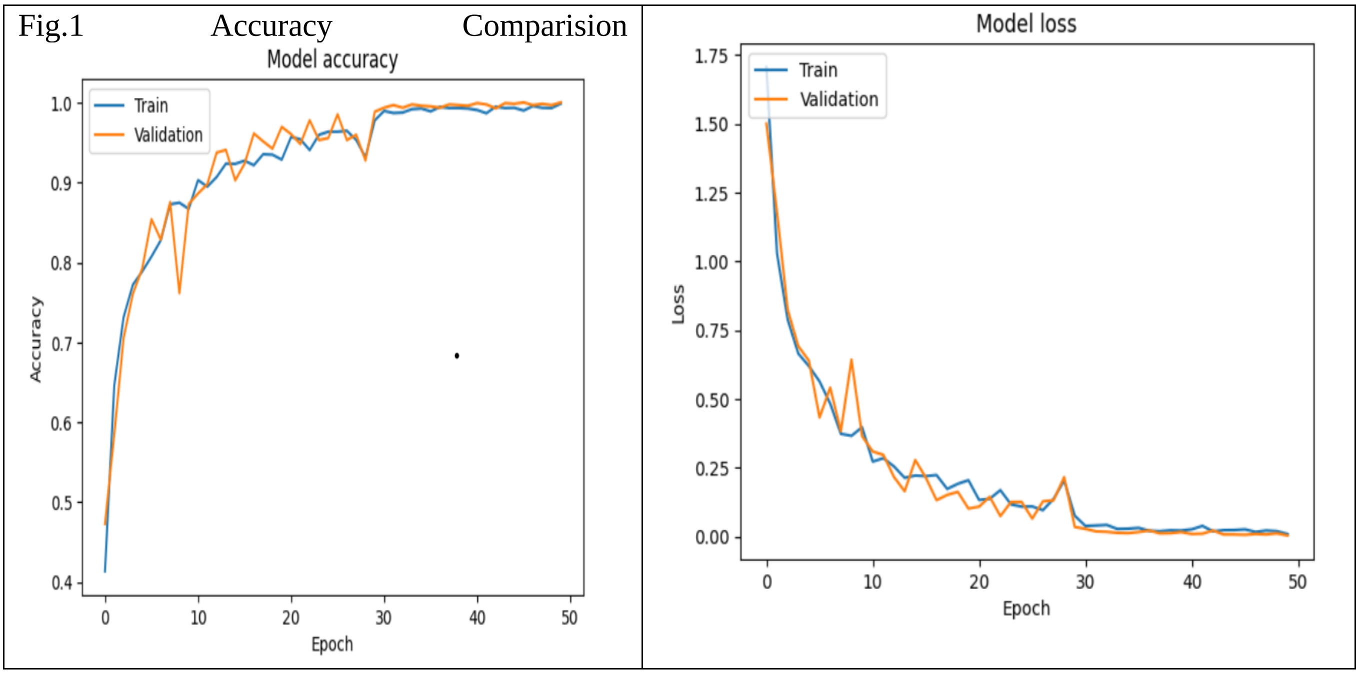
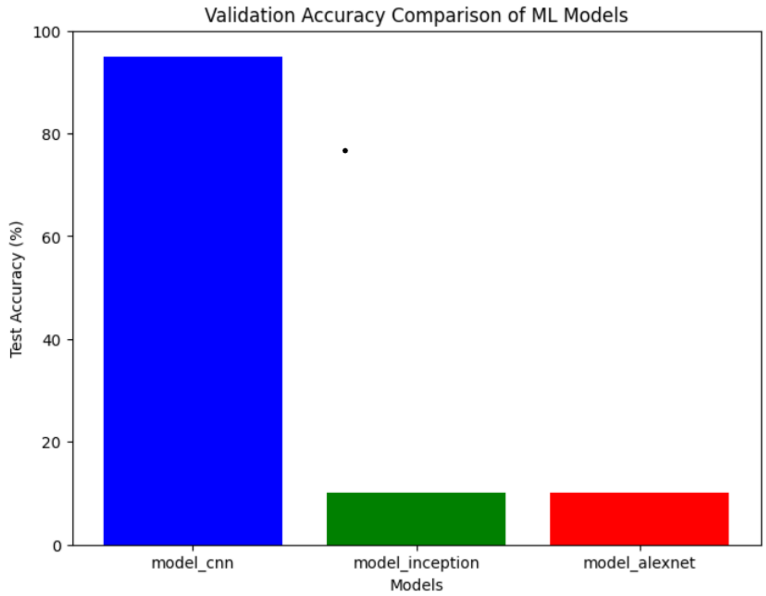
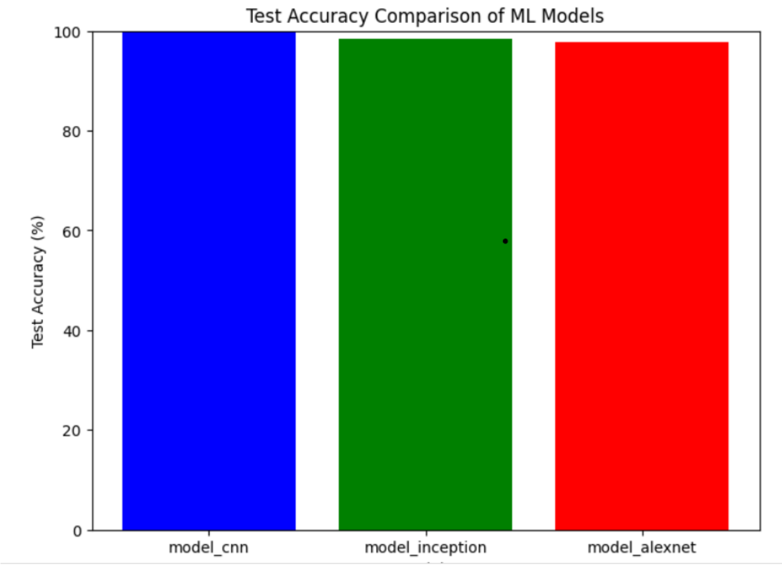


Fig. 2. (a) Model Accuracy (b) Validation loss of proposed model

The CNN model's remarkable superiority can primarily be attributed to its exceptional capacity to identify intricate features found in the input images. In contrast to conventional machine learning models, CNNs excel at automatically extracting relevant features from raw data, thereby facilitating more accurate classification tasks. This capability proved to be crucial in the context of identifying tomato leaf diseases, enabling the CNN model to distinguish between healthy and diseased leaves with unparalleled precision.





Moreover, the unique architecture of CNNs significantly contributes to their outstanding performance. By employing convolutional layers, pooling layers, and fully connected layers, CNNs can effectively capture spatial relationships and hierarchical features within the input images. This hierarchical feature extraction process empowers CNNs to detect subtle variations in leaf texture, color, and shape, which are indicative of different disease conditions.

Additionally, CNNs' adaptability to diverse datasets and environmental conditions enhances their effectiveness in agricultural applications. Through iterative learning, CNNs continuously refine their internal representations, thereby improving their ability to generalize to new data. This adaptability was particularly advantageous in our research, where the CNN model demonstrated strong performance across various disease presentations and environmental variables.

CONCLUSION :

In summary, the outstanding performance of the CNN model underscores its exceptional potential in agricultural image analysis, specifically in the detection of tomato leaf diseases. Through the adoption of deep learning and convolutional neural networks, farmers and agricultural experts can leverage cutting-edge technology to improve crop health monitoring initiatives, ultimately enhancing agricultural productivity and sustainability.

REFERNCES:

1. (2022). Disease Detection in Tomato Plants Using CNN. doi: 10.1109/gcat55367.2022.9972186
2. Sanjeela, Sagar., Jaswinder, Singh. (2023). An experimental study of tomato viral leaf diseases detection using machine learning classification techniques. Bulletin of Electrical Engineering and Informatics, doi: 10.11591/eei.v12i1.4385
3. Tomato leaf disease detection using deep learning https://www.researchgate.net/publication/376886444_Tomato_Leaf_Disease_Detection_Using_Deep_Learning_Ensemble_Approach
4. H. Kibriya, R. Rafique, W. Ahmad and S. M. Adnan, "Tomato Leaf Disease Detection Using Convolution Neural Network," 2021 , doi: 10.1109/IBCAST51254.2021.9393311.

5. Kurmi, Y., Gangwar, S., Agrawal, D., Kumar, S., & Srivastava, H. S. (2021). Leaf image analysis-based crop diseases classification. *Signal, Image and Video Processing*, *15*(3), 589-597.
6. Chen, Junde, et al. "Using deep transfer learning for image-based plant disease identification." *Computers and electronics in agriculture* 173 (2020): 105393.
7. Sun, Jun, et al. "Northern maize leaf blight detection under complex field environment based on deep learning." *IEEE access* 8 (2020): 33679-33688.