Classification and Detection of Tomato Leaf Diseases using Deep Learning Techniques

Sangam Sreeja
Student
Department of CSE
SNIST, Hyderabad,
Telangana, India

Eluka Asmitha
Student
Department of CSE
SNIST, Hyderabad,
Telangana, India

Sri Divya
Student
Department of CSE
SNIST, Hyderabad,
Telangana, India

Dr B Malathi
Associate Professor
Department of CSE
SNIST, Hyderabad,
Telangana, India

Abstract: Tomato is one of the most essential and consumable crops in the world. Tomato plants are susceptible to various diseases and pests, leading to significant yield losses in agriculture. Leaf disease is the primary factor impacting the amount and quality of crop yield. The primary goal of this research work is to propose a novel approach for automated tomato leaf disease detection using deep learning techniques. The proposed method leverages convolutional neural networks (CNNs), a type of deep learning architecture well-suited for image classification tasks. We employ a dataset comprising images of healthy tomato leaves as well as leaves affected by common diseases such as early blight, late blight, and bacterial spot. The results demonstrate the effectiveness of the proposed method in accurately identifying various tomato leaf diseases. By enabling early detection and timely intervention, this approach contributes to sustainable agriculture practices, reducing the reliance on chemical treatments and minimizing yield losses.

Keywords - Deep learning, Convolutional Neural Networks (CNN), Data augmentation

1. INTRODUCTION

A recent area of great success and attention has been deep learning, which is a subset of machine learning and artificial intelligence. Deep neural networks, another name for artificial neural networks with several layers, are trained to learn from enormous quantities of data and make intelligent decisions. Deep learning and artificial intelligence techniques are used in tomato plant disease detection to automatically identify and categorize problems in tomato plants based on pictures of their leaves or other plant parts. With the use of this technology, researchers and farmers would be able to detect diseases early on and take prompt action to stop large-scale crop losses and enhance farming methods.

2. RELATED WORK

Tomato plant disease detection using deep learning has made significant advancements in recent years. Before the widespread adoption of deep learning, tomato plant disease detection predominantly relied on traditional computer vision and machine learning techniques. Some of the existing systems for tomato plant disease detection before deep learning include Rule-Based Systems, Image Processing Techniques, Color-Based Segmentation, Texture Analysis, Machine

Learning with Handcrafted Features, and Expert Systems. Existing techniques for identifying disease include:

1. Visual Inspection by Experts
2. Manual Image Analysis
3. Rule-based Systems
4. Support Vector Machines (SVM)
5. K-Nearest Neighbors (KNN)

However, these techniques may face challenges with large-scale datasets and high-dimensional feature spaces, as they require storing all training samples in memory for classification. For more complex and diverse image datasets, deep learning techniques like CNN have demonstrated superior performance, surpassing traditional approaches.

3. PROPOSED METHOD

The proposed system for tomato leaf disease detection aims to address the limitations of existing methods by leveraging the capabilities of deep learning while considering practical considerations such as scalability, interpretability, and usability. Here's an outline of the proposed system:

- Data Collection and Preprocessing
- Data Augmentation
- Model Development
- Model compilation and Training
- Model Evaluation and Visualization
- Prediction

For tomato leaf disease detection using deep learning, selecting an appropriate model architecture is crucial for achieving accurate and efficient classification. Convolutional Neural Networks (CNNs) are the go-to choice for image classification tasks due to their ability to effectively capture spatial hierarchies in images.

Tomato plant disease detection using deep learning represents a state-of-the-art approach that utilizes Convolutional Neural Networks (CNNs) to automatically identify and categorize diseases in tomato plants from input images. This system is engineered to deliver precise and swift disease diagnosis, aiding farmers and researchers in making well-informed decisions to maintain healthy crop yields. It is built upon a comprehensive dataset of tomato plant images, featuring both healthy and diseased specimens, which reflect various disease manifestations and real-world conditions.

Standardization of the image data through preprocessing techniques ensures consistent input for the CNN model. At the core of the system lies the CNN architecture, meticulously designed to learn complex features and patterns that signify different diseases from the image data. During the training process, the model adjusts its internal parameters through backpropagation and gradient descent, improving its capability to differentiate between healthy and diseased plants. The trained model undergoes rigorous validation and testing to evaluate its performance and generalization abilities on separate datasets. This step is vital to confirm the model's accuracy and robustness in practical applications. To enhance accessibility for users, a user-friendly web interface or mobile app can be created. Through this platform, users can easily upload images of their tomato plants, which are then analyzed by the deep learning model in real-time.

The system delivers quick and accurate disease classification results, enabling farmers to take prompt and specific measures to reduce the impact of diseases on their crops. This system's benefits extend beyond individual farms, promoting sustainable agricultural practices and supporting global food security initiatives. Regular updates and enhancements to the model, driven by new data and user feedback, ensure its continued effectiveness in identifying and managing emerging disease patterns.

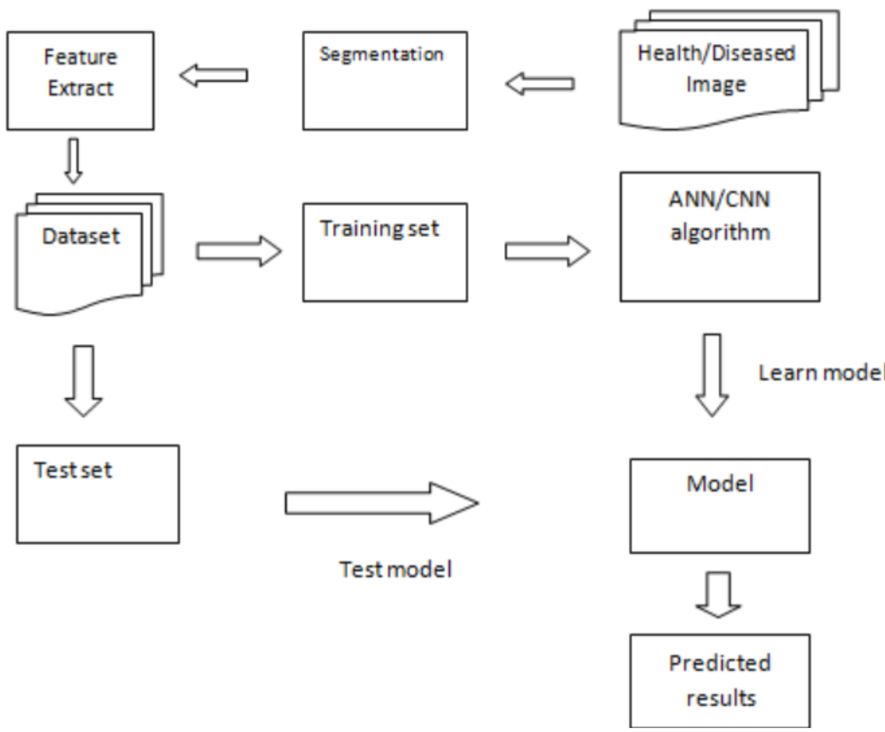


Fig 3.1: Architecture/Design

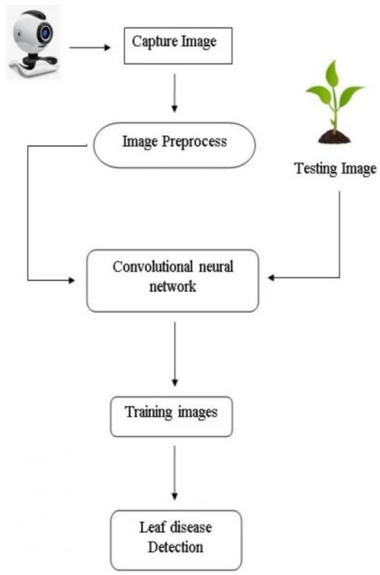


Fig 3.2: Flow diagram

3.1 CNN ALGORITHM

A Convolutional Neural Network (CNN) is an advanced deep learning algorithm tailored for processing grid-like data, particularly images. Its structure includes several distinctive layers, each contributing to its robust image analysis capabilities. The convolutional layers utilize filters to scan the input image, identifying essential features such as edges, textures, and patterns. These filters are learnable, meaning they adapt during training to capture the most relevant aspects of the data. Activation functions, like ReLU, introduce non-linearity, allowing the network to learn and model complex patterns and relationships within the data. Pooling layers, often max or average pooling, reduce the spatial dimensions of the feature maps, which not only decreases computational complexity but also helps in making the representations more robust to variations and distortions in the input data. This hierarchical approach ensures that the network can capture low-level features in early layers and more abstract, high-level features in deeper layers. The fully connected layers, typically at the end of the network, function to integrate all the extracted features and make final predictions. These layers consider the entire set of features holistically, facilitating complex decision-making processes required for classification tasks. Training a CNN involves forward propagation, where input data passes through the network layers to generate predictions, and backpropagation, which adjusts the network's weights based on the prediction errors. This iterative process, guided by optimization algorithms like Stochastic Gradient Descent (SGD) or Adam, refines the model's accuracy over numerous epochs. The data is often divided into mini-batches to balance computational efficiency and model convergence. CNNs are particularly powerful for image recognition and classification due to their ability to learn directly from raw image data, eliminating the need for manual feature extraction. Their applications extend beyond traditional image classification to include tasks like object detection, segmentation, and even video analysis. Innovations in CNN architectures, such as Residual

Networks (ResNet), Inception Networks, and EfficientNets, continuously push the boundaries of what these networks can achieve, enhancing their performance and efficiency for a wide range of practical applications, from medical diagnostics to autonomous driving systems.

3.2 DATASET DESCRIPTION

Our supervised machine learning project starts with data collection process.
There are basically three steps to collect dataset:
1) Collect and annotate data
2) Write web scrapping scripts to collect images from internet.
3) Buy data from third party vendors or use public repositories such as Kaggle.

We have taken the dataset from Kaggle repository. The link for dataset is
https://www.kaggle.com/datasets/emmarex/plantdisease
The dataset consists of images belonging to ten different classes.

```
class_names = dataset.class_names
class_names
```

```
['Tomato_Bacterial_spot',
 'Tomato_Early_blight',
 'Tomato_Late_blight',
 'Tomato_Leaf_Mold',
 'Tomato_Septoria_leaf_spot',
 'Tomato_Spider_mites_Two_spotted_spider_mite',
 'Tomato__Target_Spot',
 'Tomato__Tomato_YellowLeaf__Curl_Virus',
 'Tomato__Tomato_mosaic_virus',
 'Tomato_healthy']
```

Fig 3.2 : Class names of the dataset

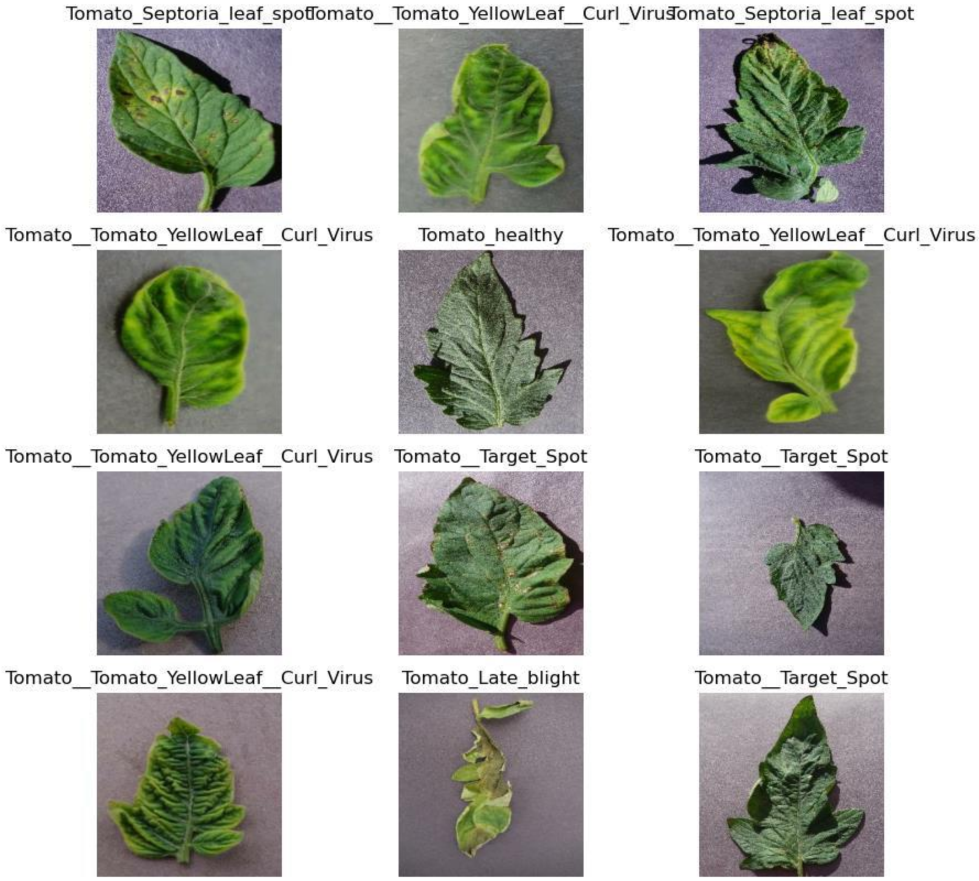


Fig 3.3: Images for the classes in the dataset

4. RESULTS

4.1 Actual Vs Predicted and Confidence

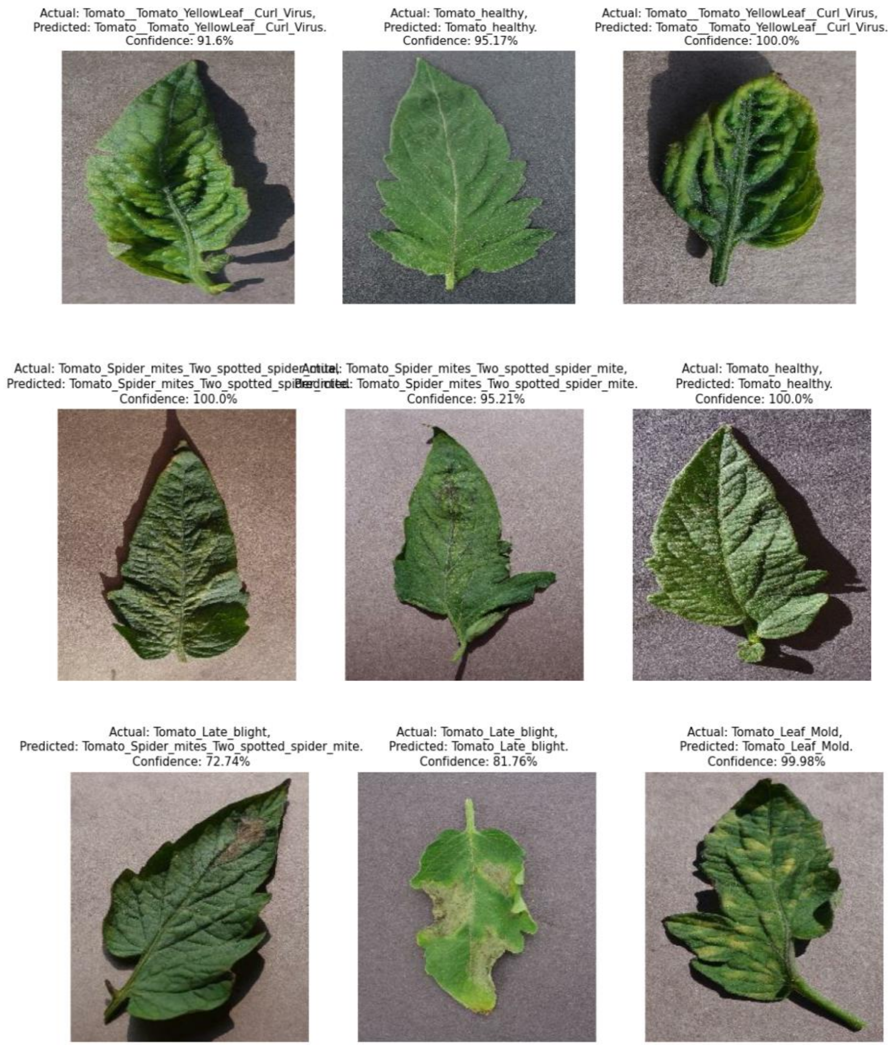
```
plt.figure(figsize=(15, 15))
for images, labels in test_ds.take(1):
    for i in range(9):
        ax = plt.subplot(3, 3, i + 1)
        plt.imshow(images[i].numpy().astype("uint8"))

        predicted_class, confidence = predict(model, images[i].numpy())
        actual_class = class_names[labels[i]]

        plt.title(f"Actual: {actual_class},\n Predicted: {predicted_class}.\n Confidence: {confidence}%")

        plt.axis("off")
```

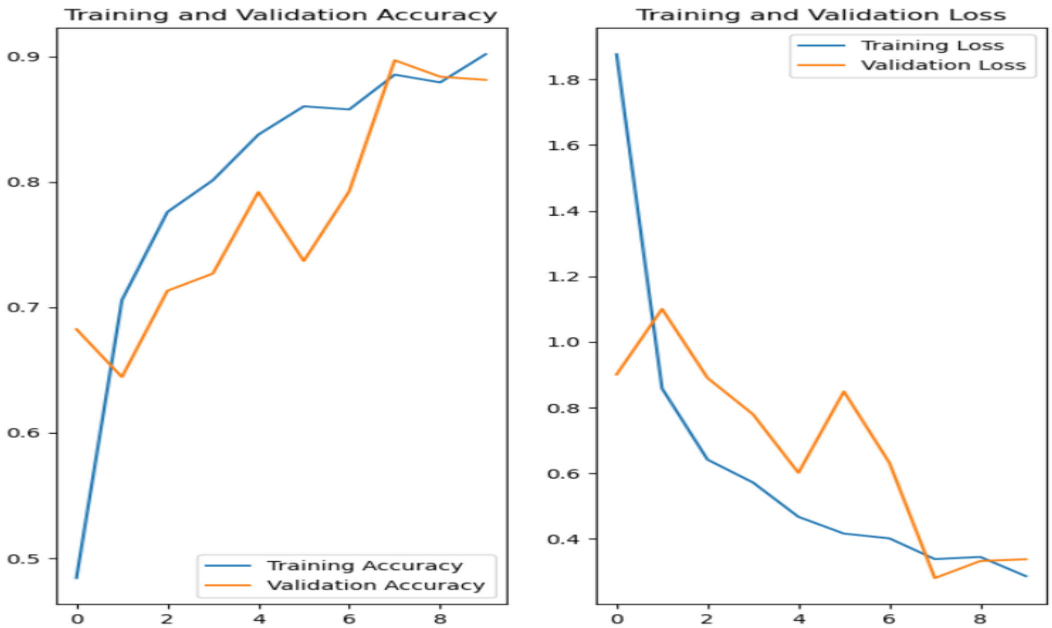
```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 104ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 85ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 120ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 168ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 96ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 161ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 88ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 169ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 169ms/step
```



4.2 Training and validation Accuracy

```
plt.figure(figsize=(8, 8))
plt.subplot(1, 2, 1)
plt.plot(range(EPOCHS), acc, label='Training Accuracy')
plt.plot(range(EPOCHS), val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation Accuracy')

plt.subplot(1, 2, 2)
plt.plot(range(EPOCHS), loss, label='Training Loss')
plt.plot(range(EPOCHS), val_loss, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')
plt.show()
```



4.3 Accuracy : is the ratio of number of correctly classified cases

```
In [33]: # Evaluate the model on the test dataset
         scores = model.evaluate(test_ds)

         # Extract the accuracy from the scores
         accuracy = scores[1]  # Assuming the accuracy metric is the second element in the scores list

         # Print the accuracy
         print(f"Accuracy: {accuracy * 100:.2f}%")

         51/51 ━━━━━━━━━━━━━━━━━━━━ 48s 926ms/step - accuracy: 0.8899 - loss: 0.3795
         Accuracy: 88.60%
```

5. CONCLUSION

The article presents a deep neural network model designed to detect and classify tomato plant leaf diseases into predefined categories. Morphological traits such as color, texture, and leaf edges were considered in the disease detection process. Specifically, the article focused on biotic diseases caused by fungal and bacterial pathogens, including blight, blast, and browns of tomato leaves. The proposed model achieved an impressive detection rate of 88.64% accuracy. This approach marks a significant advancement in identifying tomato diseases. Future plans include expanding the model to incorporate abiotic diseases caused by nutrient deficiencies in crop leaves. The long-term objective is to gather a vast amount of data on various plant diseases and enhance accuracy through further technological advancements

6. FUTURE SCOPE

Enhanced Accuracy: Continued research and development in deep learning algorithms can lead to even higher accuracy rates in detecting and classifying tomato leaf diseases. Fine-tuning models and incorporating advanced techniques like ensemble learning could improve detection performance.

Multi-Disease Detection: Expanding deep learning models to detect a broader range of tomato leaf diseases, including both biotic and abiotic factors, will be crucial. This could involve collecting and annotating large datasets encompassing various diseases and environmental stressors. Real-time monitoring systems: using deep learning enable early detection of tomato leaf diseases, integrating IoT and remote sensing for continuous field monitoring and timely farmer decision-making.

Mobile Applications: Creating user-friendly mobile applications powered by deep learning models can empower farmers to diagnose tomato leaf diseases directly in the field. These applications could provide instant feedback and recommendations for disease management strategies.

Scale-Up and Deployment: Scaling up successful deep learning-based disease detection solutions and deploying them at a larger scale will be essential for widespread adoption. This could involve partnerships with agricultural extension services, government agencies, and agribusinesses to reach farmers globally.

REFERENCES

- Zhang, Keke & Wu, Qiufeng & Liu, Anwang & Meng, Xiangyan. (2018). Can Deep Learning Identify Tomato Leaf Disease?. Advances in Multimedia. 2018. 1-10. 10.1155/2018/6710865. https://www.researchgate.net/publication/327913899
- H. D. GadadE, Dr. D.K.Kirange, Machine Learning Approach towards Tomato Leaf Disease Classification, ijatcse, https://doi.org/10.30534/ijatcse/2020/67912020. https://www.researchgate.net/publication/339552917
- Alhaj Ali, A., Chramcov, B., Jasek, R., Katta, R., Krayem, S., Awwama, E. (2021). Tomato Leaf Diseases Detection Using Deep Learning. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) Data Science and Intelligent Systems. CoMeSySo 2021. Lecture Notes in Networks and Systems, vol 231. Springer, Cham. https://doi.org/10.1007/978-3-030-90321-3_18
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8659659/