Abstract:

Today’s technology has brought revolution in many fields including Agriculture. It has a lot of potential to boost crop output and to deal with food security. Tomato being one of the most essential and widely consumed vegetables not only in India but globally, is prone to various diseases which cause severe yield loss. There are 9 different diseases which attack tomato leaves. And so, it demands a technology which will detect the disease and increase the production. In this paper we used DL, Image preprocessing and TensorFlow for detection of disease. The Dataset from Kaggle is used along with some manual data. However, CNN is used for image classification tasks which automatically learns hierarchical features from raw pixel data. Then to mitigate overfitting, a K-Fold cross-validation technique is used and along with it for developing and deploying advanced disease detection algorithm TensorFlow is utilized. Increasing population demands production, so this technique will not only boost the production but also, will reduce the time of production and so the workload.

Introduction:

In financial year 2023, the volume of tomato production in India is estimated to have amounted to over 20 million metric tons, Madhya Pradesh being the highest producer of tomato in India. As it is widely used the production directly affects the market. Increasing population which is directly proportional to demand leads to affect the supply chain. However, the upcoming technology like AI, ML made it easier than before to stable the market by stabilizing the demand supply chain.

Previously it was difficult to meet the demand, as farmers were unable to detect the disease itself, whereas now farmers not only detect but also classify the disease and accordingly treat it. Analyzing the root cause for the infected leaves, we get to know that tomato is warm season crop. So, best temperature for it ranges from 21-24°C. In this, we have worked on 9 diseases of tomato leave namely- Bacterial Spot, Early Blight, L. Mold, Late Blight, Mosaic Virus, Septoria Leaf, Spider mite, Target Spot, Yellow Leaf and also healthy leaf.


a. Bacterial Spot b. Early Blight


c. L. Mold d. Late Blight


e. Mosaic Virus f. Septoria Leaf


g. Spider Mite h. Target Spot


i. Yellow Leaf j. Healthy Leaf

Literature Survey

As we were moving ahead with this project, firstly we started this with image processing method only but didn't get the desired results. Eventually, after much research we got to know using DL along with Image Processing will enhance the output. And that's the reason we opted this method and got satisfactory results.

Er. Saban Kumar K.C. [1] in his research used image processing method. It includes image acquisition, adjusting image ROI, feature extraction and convolution neural network-based classification. To manipulate raw input image, Python programming language, OPENCV library is used. Here, dataset is taken from Plant village and few diseases are detected like, Late blind with training 100 images, testing 21 images (total 121 samples); Gray spot with training 95 images, testing 18 images (total 113 samples); and Bacterial canker with training 90 images and testing 21 images (total 111 samples). To increase the accuracy and to minimize the false detection data with varied variations has been taken. However, an overall accuracy has found to be 89%.

Then, On 6 June 2017, Rupali Khule,[2] in her research, used image segmentation and multi-class SVM algorithm. In these 4 types of diseases are detected. Multi-class SVM algorithm is uses for classification of accurate disease and Image segmentation for parting of damaged area on leaves. Here Multi-class SVM is trained with overall 320 images each 80 images of particular disease. And as a result, accuracy obtained was about 93%. To achieve more accuracy as mentioned above we moved ahead with CNN So, In Wakeel Ahmad [3] research, he did disease detection in plum using CNN with both true field dataset and authenticate source dataset. In this study several network architectures were experimented such as plane networks - AlexNet and VGG-16 and sophisticated networks- Inception and Resnet. Here, dataset was divided prior to data augmentation in the ratio 6:2:2 respectively and later the sourced dataset was also added in this. While training the true field dataset 86% performance was calculated. Further the result was calculated with ImageNet dataset before and after data augmentation. before augmentation, AlexNet noted 53% accuracy, VGG16 was overfit due to smaller dataset to tune large number of parameters. Inception Models due to their superior architecture achieved 69% and 75% accuracy. After augmentation, AlexNet noted 64.06% accuracy whereas this time VGG-16 achieved 78.68%. Inception networks as usual this time too performed well resulting 86.81% due to its capability of processing multiple scales inside the modules.

Moving forward, V. Anantha Natarajan [4], in his research used DL technique specifically faster R-CNN with deep feature extractor as ResNet50 to classify and detect tomato disease. He collected datasets consisting of 1090 real time images located at different locations in different seasons. This experiment has taken place on 4 different diseases of tomato i.e., leaf curl, bacterial spot, early blind and septoria leafspot. Here 60% data is used for training, 20% for validation and 10% for testing. Here to overcome the minimizing over-fitting problem which was the challenged face by one of their researched surveys he used validation set. Overall average precision (AP) achieved here is 80.952

However, Feng Qi [5], in his research detection methods based on CNN and Object detection models are used. CNN with 4 deep convolutional neural networks VGG-16, ResNet-50, ResNet-101 and Mobile net which are used to automatically extract original image features and Object detection models like, Faster R-CNN: to identify the types of tomato disease and Mask R-CNN: to detect and segment the locations and shapes of infected areas. Here, including healthy image with 10 diff disease type namely, tomato malformed fruit, tomato blotchy ripening, tomato puffy fruit, tomato dehiscent fruit, tomato blossom-end rot, tomato sunscaled, tomato virus disease, tomato gray mold, tomato ulcer disease and tomato anthracnose total 286 tomato images are used. In this study, dataset is divided into training: validation: test i.e., 6:2:2 ratio. In Faster R-CNN, mean average precision(mAP) of ResNet-101 is 88.53% with performing well in single tomato disease whereas, In Mask R-CNN, the highest rate achieved is 99.64%.

Lastly, Abhishek Singh [6] and his team devised a Convolutional Neural Network (CNN) to detect and classify diseases, showcasing superior performance compared to established models like VGG16, InceptionV3, and MobileNet. Their CNN model consists of three convolutional layers, three max pooling layers, and two fully connected layers. Classification accuracy ranged from 76% to 100%, averaging at 91.2% across ten classes. They utilized a dataset of 10,000 images for training, 7,000 for validation, and 500 for testing, with 1,000 healthy images and 9,000 disease images. Following meticulous analysis, their system achieved an average accuracy of 91.2% on 500 new images.

Research data

The dataset utilized in our study encompasses a comprehensive collection of approximately 87,000 RGB images depicting both healthy and diseased tomato crop leaves. These meticulously curated images serve as a valuable resource for training and evaluating machine learning models aimed at automated plant disease identification within the tomato cultivation domain.

Within this extensive dataset, the tomato subset stands out as a focal point of analysis and investigation. This subset consists of a diverse array of images capturing various stages of tomato plant health and disease manifestations. With meticulous attention to detail, the images are categorized into 38 distinct classes, each representing a specific pairing of tomato crop and disease type.

To complement this extensive dataset, additional images of tomato plants were captured manually. These manually captured images serve to enrich the dataset with real-world variations and nuances that may not be fully represented in publicly available datasets alone. Approximately 350 to 400 images were meticulously acquired, ensuring a broad spectrum of tomato plant conditions and disease states were captured.

The manual acquisition process involved careful selection of tomato plants from different geographical locations and cultivation environments to capture a representative sample of the diversity present within the tomato crop. Special emphasis was placed on capturing images depicting common tomato diseases such as early blight, late blight, bacterial spot, and various fungal infections, alongside images of healthy tomato foliage for comparative analysis.

Each manually captured image underwent rigorous quality checks to ensure clarity, focus, and relevance to the study's objectives. Close attention was paid to factors such as lighting conditions, angle of capture, and overall image composition to maintain consistency and integrity across the dataset.

Furthermore, the manually captured tomato images were meticulously labeled and integrated into the existing dataset framework, thereby expanding the dataset's coverage and enhancing its diversity. This augmentation strategy ensures that the resulting machine learning models are trained on a comprehensive and representative dataset, capable of effectively recognizing and diagnosing tomato plant diseases across different scenarios and environmental conditions.

By combining meticulously curated publicly available images with a carefully selected set of manually captured tomato images, our dataset provides researchers and practitioners with a robust foundation for developing and validating advanced machine learning algorithms for automated tomato plant disease identification and management.

Methodology

Data Preparation and Splitting

The methodology adopted for this research involved meticulous data preparation and utilization of advanced deep learning techniques for tomato plant disease identification. The dataset, comprising approximately 87,000 RGB images of tomato leaves, was initially split into training and testing sets using a 60-40 ratio. This division ensured a sufficient amount of data for model training while reserving a sizable portion for evaluation purposes.

Convolutional Neural Network Architecture

To develop an effective model for tomato plant disease identification, a Convolutional Neural Network (CNN) architecture was employed. CNNs are a class of deep neural networks particularly well-suited for image classification tasks due to their ability to automatically learn hierarchical features from raw pixel data.

The CNN architecture utilized in this study was CNN 98, a deep neural network model known for its exceptional performance in image classification tasks. This architecture consists of multiple convolutional layers followed by max-pooling layers to extract essential features from input images. Subsequently, fully connected layers and SoftMax activation were employed for classification.

Training and Evaluation

The training process involved feeding the prepared training dataset into the CNN model over ten epochs. Each epoch represents a full pass through the entire training dataset, allowing the model to iteratively learn and improve its performance.

To enhance the robustness of the model and mitigate overfitting, a K-Fold cross-validation technique was applied. This technique involves splitting the dataset into K subsets and training the model K times, each time using a different subset as the validation set and the remaining data for training. This ensures that the model's performance is evaluated o Rn diverse subsets of the data, thereby providing more reliable estimates of its generalization capabilities.

During training, the model's performance was evaluated using various metrics, including accuracy, precision, and recall. Additionally, both training and testing accuracy were monitored to assess the model's ability to generalize to unseen data.

Results

Upon the completion of training and evaluation, the CNN model demonstrated promising performance in tomato plant disease identification. The results indicated a high level of accuracy and precision in classifying tomato leaf images into their respective disease categories.

- *Train Accuracy: * 98.25%

- *Test Accuracy: * 96.72%

- *Precision Score: * 96.72%

- *Recall Score: * 96.72%

These results highlight the effectiveness of the CNN 98 architecture in accurately identifying and distinguishing between various diseases affecting tomato plants. The high accuracy and precision scores indicate the model's capability to assist farmers and agricultural experts in early detection and management of tomato plant diseases, thereby potentially improving crop yield and sustainability.

```
Label: TomatoEarlyBlight1.JPG , Predicted: Tomato___Early_blight
Label: TomatoEarlyBlight2.JPG , Predicted: Tomato___Early_blight
Label: TomatoEarlyBlight3.JPG , Predicted: Tomato___Early_blight
Label: TomatoEarlyBlight4.JPG , Predicted: Tomato___Early_blight
Label: TomatoEarlyBlight5.JPG , Predicted: Tomato___Early_blight
Label: TomatoEarlyBlight6.JPG , Predicted: Tomato___Early_blight
Label: TomatoHealthy1.JPG , Predicted: Tomato___healthy
Label: TomatoHealthy2.JPG , Predicted: Tomato___healthy
Label: TomatoHealthy3.JPG , Predicted: Tomato___healthy
Label: TomatoHealthy4.JPG , Predicted: Tomato___healthy
Label: TomatoYellowCurlVirus1.JPG , Predicted: Tomato___Tomato_Yellow_Leaf_Curl_Virus
Label: TomatoYellowCurlVirus2.JPG , Predicted: Tomato___Tomato_Yellow_Leaf_Curl_Virus
Label: TomatoYellowCurlVirus3.JPG , Predicted: Tomato___Tomato_Yellow_Leaf_Curl_Virus
Label: TomatoYellowCurlVirus4.JPG , Predicted: Tomato___Tomato_Yellow_Leaf_Curl_Virus
Label: TomatoYellowCurlVirus5.JPG , Predicted: Tomato___Tomato_Yellow_Leaf_Curl_Virus
Label: TomatoYellowCurlVirus6.JPG , Predicted: Tomato___Tomato_Yellow_Leaf_Curl_Virus
```

Conclusion

This research paper has highlighted the transformative impact of using deep learning (DL), image preprocessing techniques, and the TensorFlow framework in the realm of plant disease detection. By meticulously investigating and integrating these cutting-edge technologies, significant strides have been made in advancing agricultural practices.

Through comprehensive experimentation and analysis, it has been evidenced that the amalgamation of DL models with image preprocessing methods greatly enhances the accuracy, reliability, and efficiency of disease detection systems. The utilization of DL algorithms, such as convolutional neural networks (CNNs), enables the extraction of intricate patterns and features from plant images, facilitating precise identification of various diseases across different crops.

Moreover, the incorporation of sophisticated image preprocessing techniques, including normalization, augmentation, and segmentation, further refines the input data, mitigating noise and enhancing the discriminatory power of the DL models. This synergy between DL and preprocessing methodologies not only improves the robustness of disease detection systems but also enables them to adapt to diverse environmental conditions and varying image qualities encountered in real-world agricultural settings.

The utilization of TensorFlow as a powerful deep learning framework provides a scalable and flexible platform for developing and deploying advanced disease detection algorithms. Its rich ecosystem of tools and libraries streamlines the model development process, while also facilitating efficient utilization of computational resources, thereby enabling real-time or near-real-time disease diagnosis in the field.

Overall, this research underscores the immense potential of DL, image preprocessing, and TensorFlow in revolutionizing agricultural practices by offering timely and accurate solutions for disease management. As we continue to refine and optimize these technologies, there is a promising trajectory towards achieving sustainable agriculture, ensuring global food security, and mitigating economic losses incurred due to crop diseases. By embracing innovation and collaboration, the agricultural community can harness the transformative power of these technologies to address the challenges facing modern farming and usher in a new era of productivity and resilience in crop protection.

Future scop

Integrating a disease detection model with remedies presents an opportunity to offer comprehensive solutions for identified issues. Building upon this, a system could be developed to suggest tailored remedies and farming techniques for each specific disease detected, enhancing agricultural practices. To further bolster accuracy, captured images could be archived within the model for future reference and analysis, aiding in continuous improvement. Implementing a chatbot feature would enable seamless user-system interaction, thereby improving accessibility to services. To refine the model's precision, expanding the dataset and fine-tuning its parameters will be imperative. Furthermore, integrating weather APIs can furnish real-time weather verification, facilitating enhanced farm management and decision-making processes. This holistic approach promises to revolutionize disease detection and farm management practices, fostering sustainable agriculture.

References

[1] Santosh Adhikari, Er. Saban Kumar K.C, Bikesh Shrestha, Bibek “Tomato Plant Disease Detection Using Image Processing” 1st KEC Conference Proceedings, September 27, 2018

[2] Sagar Vetal, R.S. Khule"Tomato Plant Disease Detection using Image Processing “International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 6, June 2017

[3] Jamil Ahmad, Bilal Jan, Haleem Farman, Wakeel Ahmad and Atta Ullah “Disease Detection in Plum Using Convolutional Neural Network under True Field Conditions” Sensors 2020 20(19)

[4] Dr. V. Anantha Natarajan, Ms. Macha Babitha, Dr. M. Sunil Kumar, “Detection of disease in tomato plant using Deep Learning Techniques,” International Journal of Modern Agriculture, Volume 9, No.4, 2020

[5] Qimei Wang, Feng Qi, Minghe Sun, Jianhua Qu, and Jie Xue” Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques” Juan Carlos Fernandez Volume 2019.

[6] Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, Suneet Gupta “Tomato Leaf Disease Detection Using Convolution Neural Network” Procedia Computer Science Volume 167, 2020