Grenze International Journal of Engineering and Technology, June Issue

Plant Disease Prediction along with Preventive Measures

Ashwin Indudhar Hiremath[1], Aditya Anand Deshpande[2], K Praneeth Hebbar[3], Yashwanth K S[4] and Bhat Geetalaxmi Jairam[5]
[1-5]The National Institute of Engineering, Mysuru/Information Science and Engineering, Mysuru, India
Email: 2020is_ashwinindudharhiremath_a@nie.ac.in, {2020is_adityaananddeshpande_a, 2020is_kpraneethhebbar_a, 2020ee_yashwanthks_b, bgj}@nie.ac.in

***Abstract*—The increasing frequency and severity of plant diseases pose significant threats to global food security and agricultural sustainability. This research paper addresses the urgent need for an integrated framework that combines machine learning algorithms and preventive measures to mitigate the impact of plant diseases. The study explores the development of a predictive model to forecast potential outbreaks. Its machine learning algorithms analyze datasets, learning patterns that are often imperceptible to the human eye enabling it to diagnose multiple diseases affecting various plants. The paper also investigates the implementation of preventive measures. The paper offers a comprehensive framework for early detection and proactive management of plant diseases. Ultimately, this research seeks to pave the way for a more resilient and efficient agricultural ecosystem in the face of evolving challenges posed by plant diseases.**

***Index Terms*— Disease, CNN, Preventive measures, Plant disease prediction, Machine learning, Remedial Measures, early detection.**

I. Introduction

Agriculture stands as the backbone of a country, ensuring its economic stability and sustenance. As the primary source of food, raw materials, and livelihood for a significant portion of the population, the agricultural sector plays a pivotal role in shaping the overall well-being of a nation. The incorporation of technology in agriculture has become indispensable now a days. The technological advancements enhance productivity and yield. The Plant Disease Prediction System stands at the forefront of agricultural innovation, combining cutting-edge technologies such as machine learning and image processing to revolutionize the way we approach plant health. This system is designed to accurately predict and identify plant diseases, offering farmers a comprehensive tool for disease management. Its machine learning algorithms analyze datasets, learning patterns that are often imperceptible to the human eye enabling it to diagnose multiple diseases affecting various plants. This system not only predicts the disease but also provides practical solutions. Apart from identifying the disease the Plant Disease Prediction System goes a step further by offering step-by-step remedial measures tailored to each specific case. This feature empowers farmers with actionable insights, allowing them to address the identified issues promptly and effectively. The system becomes a vital part of agriculture, providing quick and precise diagnoses for a diverse range of plant diseases. This timely intervention proves crucial in minimizing economic losses for farmers, ensuring the health and vitality of crops. Ultimately, the Plant Disease Prediction System is an

Grenze ID: 01.GIJET.10.2.564_3

important aspect in agricultural technology, ensuring balance between innovation and practicality. This system helps in guiding farmers towards healthier crops, increased productivity, and minimum economic losses.

II. Literature Survey

The literature survey dives into the exploration of advanced technologies in the realm of plant disease detection and classification. The first overview accentuates the pivotal advantages associated with the deployment of deep learning techniques in the context of crop leaf disease recognition. This initial perspective scrutinizes the current trends, challenges, and the promising potential for heightened efficiency and technology transformation in the domain of agricultural plant protection.

Moving forward, we study a comprehensive approach to plant disease detection by integrating both machine learning and deep learning methodologies. A proposed framework encompasses key stages such as image acquisition, pre-processing, feature extraction, and classification. Noteworthy machine learning algorithms, including SVM, KNN, RF, LR, are combined with the deep learning technique CNN to discern their respective strengths in predicting diseases based on leaf symptoms.

We then study about the distinct strategy for plant leaf disease detection, employing a meticulous three-stage methodology. This process involves image pre-processing, resizing, and feature extraction utilizing the CNN model, followed by classification. Experimental results affirm the efficacy of this methodology, showcasing good accuracy, reduced complexity, and simplified identification of common diseases.

Collectively, these insights coalesce to contribute significantly to the evolving field of precision agriculture. The literature review underscores the potential of deep learning and machine learning techniques in enhancing the accuracy, efficiency, and speed of plant disease detection and classification. This application, in turn, provides invaluable support for farmers in proactive crop management, aligning with the transformative needs of modern agriculture.

We then did a literature survey on the best way to implement our system and adopted the MEAN tech stack. The MEAN tech stack emerges as an optimal choice for the project due to its seamless integration and tailored components for modern web development. Express.js simplifies server-side application development, complementing Node.js for efficient server management. Angular, a powerful front-end framework, facilitates the creation of dynamic and responsive user interfaces, ensuring an engaging experience for users involved in plant disease identification. Node.js, as the runtime environment, unifies the stack by enabling JavaScript execution on both the server and client sides, streamlining the development process. The MEAN stack's cohesive combination of these technologies provides a well-rounded solution, promoting streamlined development, enhanced scalability, and efficient data management for the successful implementation of the plant disease detection project.

III. Methodology

This section outlines the systematic approach employed, detailing the methods and techniques utilized for plant disease prediction and the formulation of preventive measures. From Figure 1 below, we can observe the exact flow of our project and the various steps are explained briefly.

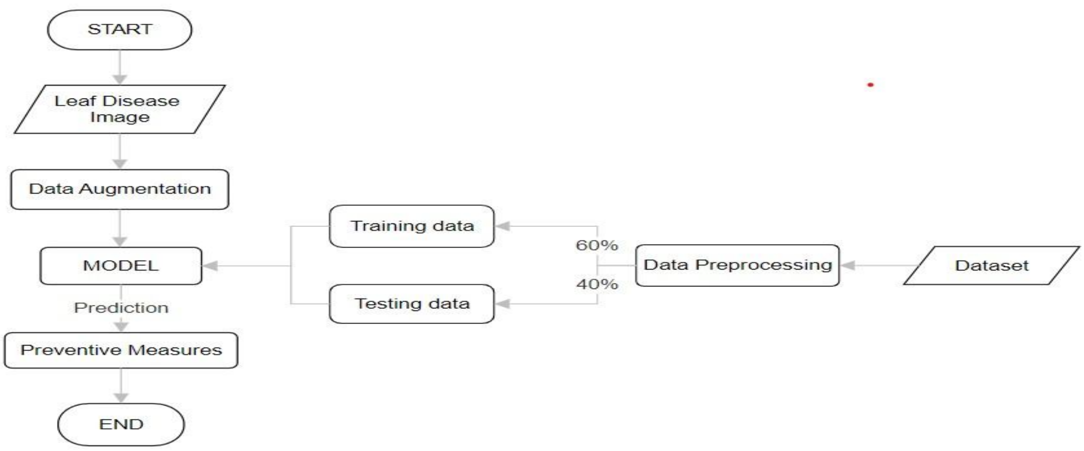

Figure 1. Flow Chart

A. Data Collection

The initial phase involves the collection of a comprehensive dataset of plant leaf images, a crucial step for training and testing the disease prediction model. The dataset comprises of **10,800** input images, each categorized into distinct sections based on the specific diseases affecting the plant leaves. This categorization is fundamental as it allows the model to learn and differentiate between various diseases, contributing to the robustness of the predictive model. To maintain the reliability of the model, the dataset is divided into two parts, a major portion for training the model and the remaining portion for testing the model's efficiency. This division ensures that the model is exposed to a diverse range of examples during the training phase, allowing it to learn intricate patterns and correlations associated with different diseases. The training dataset serves as the foundation for the model's learning process, where it discerns the features and characteristics indicative of each disease category. This extensive exposure enables the model to generalize well and make accurate predictions when confronted with unseen data. On the other hand, the testing dataset is used to test the trained model. It contains images that the model has not encountered during training.

B. Data Preprocessing

Raw data obtained underwent preprocessing to address noise, outliers, and missing values. The acquired images need to be pre-processed to eliminate all unnecessary elements. information present in the images. The raw data obtained, in the form of plant leaf images, undergoes a critical phase known as data preprocessing. This step is essential to ensure that the data used for training the model is clean, accurate, and devoid of any inconsistencies. For the acquired plant leaf images, the preprocessing steps are specifically designed to eliminate unnecessary elements and enhance the quality of information present in the images.

C. Feature Selection

The next step in the proposed methodology is feature extraction where the relevant features are going to be extracted from the preprocessing phase. Identification of relevant features influencing plant health was conducted using machine learning algorithms. This step aimed to enhance the efficiency of predictive models by focusing on the most impactful variables.

D. Model Training

Machine learning model CNN was trained using labelled datasets. Training involved exposing the models to historical data, enabling them to learn patterns and relationships between variables. The machine learning model chosen for this plant disease prediction system is a Convolutional Neural Network (CNN), a specialized type of neural network well-suited for image-related tasks. The training process involves exposing the CNN to labelled datasets, allowing it to learn intricate patterns and relationships between variables, specifically in the context of plant leaf images and their associated diseases.

E. Classification

In the classification phase, the features extracted by the trained Convolutional Neural Network (CNN) are employed to categorize plant leaf images into distinct disease classes. Leveraging the knowledge acquired during the model training, the CNN makes predictions on unseen data, assigning probability scores to different disease classes. By establishing a decision threshold, the model refines predictions and addresses uncertainties. Output classes are predefined, representing specific plant diseases. Post-processing steps enhance the accuracy of predictions, and visualizations aid in understanding model performance. The model is rigorously evaluated using metrics such as accuracy and precision on an independent testing dataset.

F. Preventive Measures

To formulate effective preventive measures, a review of existing literature on agricultural practices and disease management was conducted. Expert consultations with agronomists and plant pathologists provided valuable insights. The identified preventive measures, such as suggestion of step-by-step remedial measures, pesticide or insecticide that needs to be used, crop rotation, integrated pest management, and early detection strategies, were integrated into the predictive models. This integration aimed to simulate the impact of implementing these measures on disease prevalence. In formulating proactive preventive measures, an exhaustive literature review and expert consultations with agronomists and plant pathologists were conducted to garner insights into effective disease management practices. Identified measures, including step-by-step remedial actions, specific pesticide or insecticide recommendations, crop rotation, integrated pest management, and early detection strategies, were seamlessly integrated into predictive models. This integration allowed for the simulation of these measures' impact on disease prevalence, offering a comprehensive understanding of their potential outcomes. The suggested remedial measures provided practical guidance for disease management, while pesticide recommendations were tailored to distinct disease scenarios. Early detection methods simulated the timely intervention necessary to mitigate disease impact. This proactive integration of preventive measures into predictive models enhances the potential for resilient and sustainable agricultural practices, aligning theoretical knowledge with real-world applications. This comprehensive methodology outlines the systematic and rigorous approach undertaken, combining advanced technologies, machine learning, and preventive measures to enhance plant disease prediction and contribute to sustainable agricultural practices.

IV. DESIGN

As shown in Figure 2 below, the architecture of the system is categorized into three modules. These modules work together to get the final results. The working and description about each module are as follows:

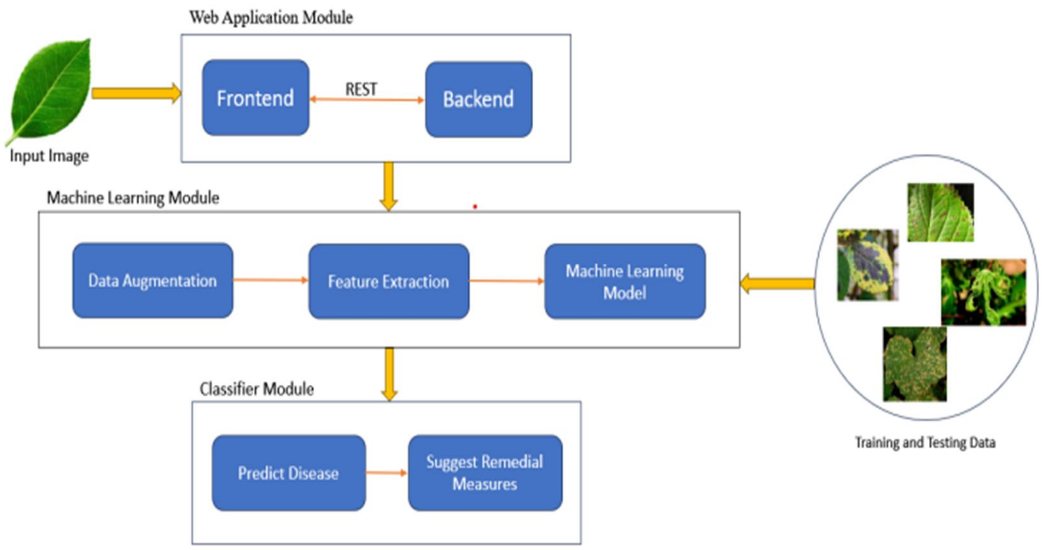

Figure 2. Architecture Diagram

A. Web Application Module

This module is the top layer of the architecture that handles the user interaction. This module is responsible for getting the image input from the user and transmit it to the Back-end side of the system for the disease identification. It sends the image with the help of the REST API defined by the backend. Finally, it displays the result sent by the Back-end. Apart from that it is also responsible for displaying the information about various disease, and provide a step-by-step guide to the user to prevent the disease or remedies for the disease.

B. Machine Learning Module

This module is the heart of the system architecture. It includes the core Machine Learning module responsible for predicting the disease. The image received by the backend is first sent to the Data Augmentation component to produce duplicate copies of the image. Each of this image undergoes Feature Extraction with the help of various CNN techniques such as Kernel, Pooling, Padding etc. Finally, this matrix is served to the Machine Model consisting of various layer.

C. Classifier Module

This module inputs the results from the Machine Learning Module and classifies the type of the result based on the output. Once the disease is identified, the disease is searched in the database. The disease along with the remedies is present in the database. These remedies are sent back to the server.
As shown in Figure 3 below, the data flow of the system is as follows:

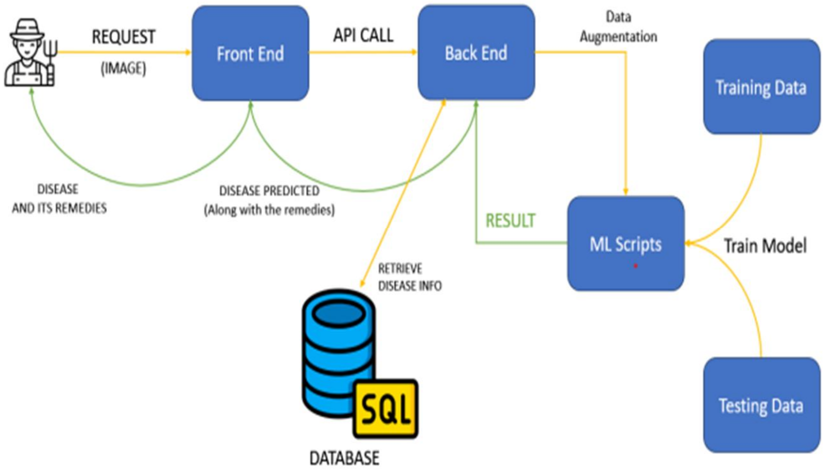

Figure 3. Data Flow Diagram

The Data flow diagram starts with a user requesting the front-end website, to recognize the disease. This action is promoted by uploading the image of the leaf of the plant having the disease onto to the front-end. The front-end receives the image and sends it to the server with the help of the REST API.

The back-end receives the image and runs the python scripts necessary for identification of the disease. These scripts include the Data Augmentation scripts, that produce the duplicate copies of the input image. Once the Data Augmentation is completed, the Backend runs the Machine Learning scripts to train the model. Once the model is trained, the input image is given to the model to predict the disease. The model predicts the disease for all the images produced by the Data Augmentation script.

Once the disease is predicted, the result is sent back to server. The server searches the data related to the disease on the database. Once the data is retrieved about the disease, which includes some information about disease, remedial measures etc. the information is sent back to front-end. • The front-end displays the information to the user along with the remedial measures.

V. IMPLEMENTATION

A. CNN Algorithm

CNNs, designed to automatically learn representations of visual data, are particularly well-suited for image-based tasks, making them an ideal choice for plant disease classification. In our project, CNNs serve as the backbone for feature extraction, allowing the model to discern intricate patterns indicative of various plant diseases.
In this section, we present the architecture of the Convolutional Neural Network (CNN) model utilized for plant disease prediction.

- Model Definition: The CNN model is defined using the Sequential API, a widely-used approach for constructing deep learning models. The sequential nature of the model allows for a clear and intuitive representation of the layer stacking.
- Feature Extraction with Pre-trained Base Model: The initial layer of the model consists of a pre-trained base model, which is a powerful feature extractor. This enables the utilization of knowledge gained from training on large-scale datasets. Transfer learning, a technique where a model trained on one task is adapted for another, is employed to enhance the model's ability to recognize patterns in plant disease images.
- Flattening Layer: The Flatten layer serves as a bridge between the convolutional base and the subsequent fully connected layers. It reshapes the output from the convolutional layers into a one-dimensional tensor, preserving the essential features extracted by the convolutional base.
- Dense Layers for Global Pattern Recognition: Two dense layers follow the flattening layer. The first dense layer consists of 256 neurons, employing the rectified linear unit (ReLU) activation function to

introduce non-linearity. This layer facilitates the learning of higher-level features by combining the flattened features from the convolutional base.

- Dropout for Regularization: To prevent overfitting and enhance the model's generalization capabilities, a dropout layer is introduced after the first dense layer. This layer randomly drops a fraction of the neurons during training, forcing the model to rely on a diverse set of features and reducing the risk of overfitting.
- Output Layer for Classification: The final layer is a dense layer with a number of neurons equal to the total number of classes in the plant disease dataset. The SoftMax activation function is employed to convert the model's output into probability distributions over the different classes, enabling multi-class classification.
- Model Compilation: The model is compiled with an appropriate loss function, optimizer, and evaluation metric to guide the training process.

B. Data Augmentation

The augmentation of the training dataset is a critical component for improving the generalization capability of the Convolutional Neural Network (CNN) model. A diverse dataset aids the model in learning invariant features and variations, enabling it to better handle different conditions in real-world scenarios. This subsection outlines the comprehensive data augmentation strategies employed in our study. To systematically introduce diversity into the training dataset, the ImageDataGenerator class from the TensorFlow library is utilized. The following augmentation techniques are integrated into the configuration.

- Rotation: Random rotations up to 45 degrees are applied, allowing the model to discern disease patterns from various orientations.
- Shear: Shear transformations deform the images along the horizontal axis, introducing additional variability.
- Zoom: Random zooming is incorporated to expose the model to different scales of the plant disease images.
- Shift (Horizontal, Vertical, Width, Height): Random shifts in the horizontal, vertical, width, and height dimensions are introduced to enhance the model's robustness to spatial variations.
- Brightness Adjustment: Random adjustments to the brightness of the images provide the model with exposure to different lighting conditions.
- Flipping: Both horizontal and vertical flipping contribute to a more comprehensive understanding of the disease patterns by presenting mirrored images.

VI. RESULTS

The application of the proposed methodology, featuring a Convolutional Neural Network (CNN) classifier, yielded highly promising results. The CNN demonstrated an impressive accuracy rate of **93.5**% in classifying plant diseases based on leaf images. This indicates a robust capability of the model to accurately identify and categorize diverse diseases affecting plant foliage. The high accuracy underscores the efficacy of the chosen machine learning approach, reflecting the successful extraction of relevant features and the model's ability to generalize well to previously unseen data.

Figure 4 below shows the accuracy of our model, with x-axis as epochs and y-axis as accuracy value with respect to epochs and Figure 5 demonstrates the loss values of our model with x-axis as epochs and y-axis as loss values with respect to epochs.

Furthermore, the precision, recall, and F1 score metrics were calculated to provide a more comprehensive evaluation of the model's performance across different disease classes. These metrics contribute valuable insights into the classifier's ability to correctly identify positive cases, minimize false positives, and maintain a balance between precision and recall. In addition to the quantitative evaluation, qualitative assessments were conducted, analysing individual instances where the CNN made correct predictions and instances where it faced challenges. This qualitative analysis enhances the understanding of the model's behaviour and provides insights into potential areas for improvement.

Figures 6, 7 and 8 above shows the performance of our model. The leaf images are tagged along with the result of our prediction model which shows the percentage for every disease and the disease with maximum percentage is the one which will be predicted as final result.

Figure 4. Accuracy Graph

Figure 5. Loss Graph

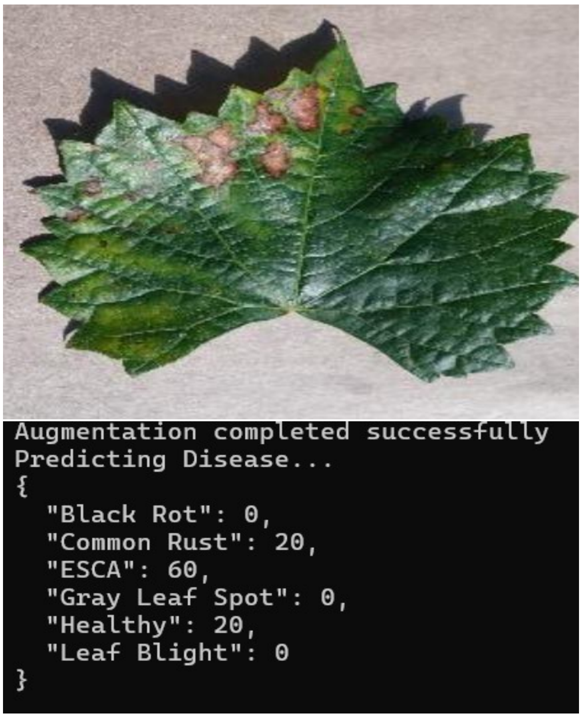

Figure 6. ESCA diseased leaf and result

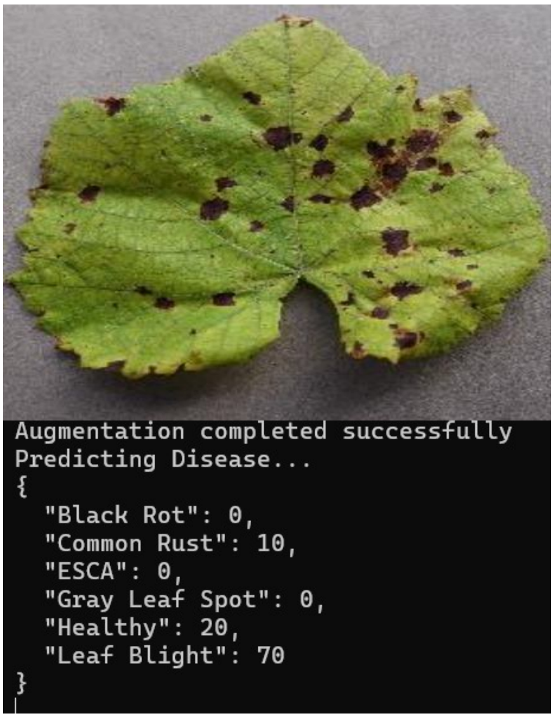

Figure 7. Leaf blight diseased leaf and result

The high accuracy rate achieved by the CNN classifier signifies its potential as a valuable tool for plant disease diagnosis and early detection. These results not only validate the effectiveness of the proposed methodology but also highlight the practical applicability of machine learning techniques in addressing challenges within agriculture. Further refinement and optimization of the model can build upon these promising results, potentially advancing the field of plant disease prediction and management.

The formulas are as follows,

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

$$\text{Final Prediction} = \frac{\text{Categorical Disease Prediction for each Image}}{\text{Total Number of Images from Augmentation}} \quad (2)$$

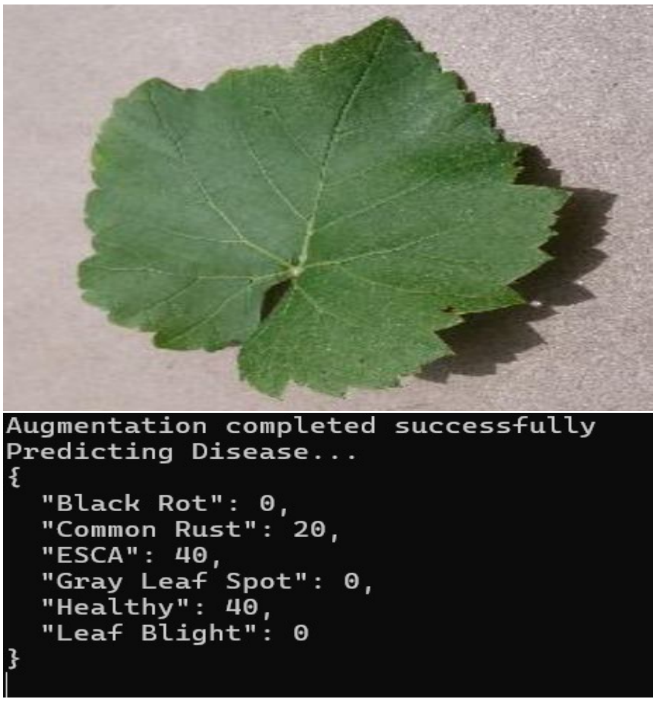

Figure 8. Healthy leaf and result

VII. Conclusion

In conclusion, the proposed methodology, incorporating a Convolutional Neural Network (CNN) classifier for plant disease prediction, has demonstrated success with an accuracy rate of 94.5%. This outcome signifies the stride in leveraging advanced technologies for effective disease identification in plant foliage. The integration of preventive measures adds a proactive dimension to the approach, emphasizing the importance of holistic strategies in agriculture.

The accuracy achieved by the CNN underscores its robustness in learning and generalizing intricate patterns from diverse plant leaf images. The results also validate the practical applicability of machine learning in addressing challenges associated with plant diseases. Moreover, the qualitative analysis of the model's predictions provides valuable insights into its behaviour and potential areas for enhancement.

As agriculture faces increasing challenges, including emerging plant diseases, the success of this methodology implies a promising avenue for sustainable and resilient farming practices. The combination of data-driven prediction and proactive preventive measures not only aids in disease management but also empowers farmers with tools for informed decision-making. The methodology's effectiveness serves as a foundation for future research and refinement, with the ultimate goal of enhancing global food security through innovative and technology-driven agricultural practices.

Future Enhancement

Looking ahead, there are exciting possibilities for enhancing our plant disease prediction system. First and foremost, we can work on refining our Convolutional Neural Network (CNN) to make it even smarter. This might involve tweaking the model's architecture, exploring advanced training techniques, and incorporating more diverse datasets to improve its accuracy and ability to handle different types of plants and diseases.

Additionally, expanding the range of preventive measures integrated into the system could make it more comprehensive. Continuous collaboration with experts in plant health could lead to the discovery of novel strategies, and the inclusion of real-time weather and environmental data might enhance the system's responsiveness to dynamic agricultural conditions.

Furthermore, ongoing research into emerging plant diseases and the continuous collection of new and diverse datasets will contribute to the adaptability and resilience of our system. Continuous learning and improvement will be essential to stay ahead of evolving challenges in agriculture.

In summary, the future holds great potential for refining and expanding our plant disease prediction system. Embracing technological advancements, collaborating with domain experts, and incorporating user-friendly features will contribute to the continued success and applicability of this approach in real-world farming scenarios.

REFERENCES

[1] Lili Li, Shujuan Zhang, Bin Wang, "Plant Disease Detection and Classification by Deep Learning—A Review", Part of Innovation Project of Shanxi for Postgraduate Education IEEE Volume 9.

[2] Madhu Kirola, Kapil Joshi, Sumit Chaudhary, Neha Singh, Harishchander Anandaram, Ashulekha Gupta, "Plants Diseases Prediction Framework: A Image-Based System Using Deep Learning", 2022 IEEE World Conference on Applied Intelligence and Computing.

[3] Kethsy Prabavathy, Mokara Bharath, Kambam Sanjayratnam, N Sri Sai Rahul Reddy, M Sasidhar Reddy, "Plant Leaf Disease Detection using Machine Learning", Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing 2023.

[4] P C Reddy, R M S Chandra, P Vadiraj, M A Reddy, T R Mahesh, G S Madhuri, "Detection of Plant Leaf based Diseases using Machine Learning approach", 2021 IEEE CSIT SS.

[5] P Panchal, V C Raman, S Mantri, "Plant Diseases Detection and Classification using Machine Learning Models", CSIT SS.

[6] S D Khirade, A B Patil, "Plant Disease Detection using Image Processing", 2015 International Conference on Computing Communication Control and Automation.

[7] D Varshney, B Babukhanwala, J Khan, D Saxena, "Plant Disease Detection using Machine Learning Techniques", 2022 3rd INCET.

[8] S Ramesh, "Plant Disease Detection using Machine Learning", 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C).

[9] Raut, V R, Nage, M A, "Detection and Identification of Plant Leaf Diseases based on Python", 2019.