Mini Project Report on

AUTOMATIC PLANT DISEASE PREDICTION

Submitted in partial fulfilment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Student Name : Km Neha Andola University Roll No. : 2018888

Under the Mentorship of

Mr Arnav Kotiyal
Assistant Professor

Department of Computer Science and Engineering
Graphic Era (Deemed to be University)
Dehradun, Uttarakhand

July 2023

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project report entitled **"AUTOMATIC PLANT DISEASE PREDICTION"** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Mr Arnav Kotiyal, Assistant Professor,** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

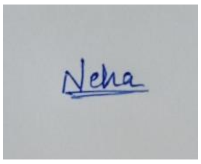
Name	University Roll No.	Signature
Km Neha Andola	2018888	

Table of Contents

Chapter No.	Description	Page No.
Chapter 1	Introduction	**04**
Chapter 2	Literature Survey	**08**
Chapter 3	Methodology	**10**
Chapter 4	Result and Discussion	**14**
Chapter 5	Conclusion and Future Work	**16**
	References	**17**

CHAPTER 1

INTRODUCTION

This mini project focuses on applying machine learning techniques to a plant Disease prediction. The goal is to create a model so that it can be accurately identified and classify various plant diseases based on input data such as plant images leaves. The project uses a dataset containing images of healthy leaves and leaves affected by various diseases. By detecting diseases at an early stage, farmers can adopt appropriate measures to prevent spread and minimize crop losses. Farmers and gardeners can take early action to prevent or mitigate crop damage, leading to higher crop yields and better farming practices.

1.1 ABSTRACT

Agriculture plays a vital role in the Indian economy and serves as the backbone of the Indian economy sustenance and growth of the nation. With a significant portion of the population dependent on agriculture for their livelihood, the sector contributes significantly to employment and countryside development. India is known for its diverse agricultural practices, producing a wide variety of crops, including rice, wheat, sugarcane, cotton, and various fruits and vegetables.

Despite the rapid growth of other sectors, agriculture remains a key contributor GDP of the country, making it a critical element in India's economic landscape. India exports various agricultural products including rice, wheat, spices, tea, coffee and others. It supports millions of farmers, farm workers and their families by contributing to rural livelihood. It creates income and employment opportunities in rural areas, leading to improvements in infrastructure, education and health facilities. India’s most of the population relies heavily on the agricultural sector to meet its food requirements. A robust farming system is essential to ensure food security and sustain the stable prices of basic commodities. Agriculture also serves as a source of raw materials for various agricultural sectors such as food processing, textiles and animal husbandry. The linkages between agriculture and industry further enhance economic growth and diversification. It underlines the dependence of agriculture on the weather and climatic conditions the importance of climate-resilient agricultural practices.

Climate plays a significant role in indirectly influencing the occurrence, development and spread of plant diseases. It has a significant impact on Indian economy and lives farmers.

- When crops are affected by plant diseases, they can lead to reduced yields and lower yields agricultural production. This directly affects the income of farmers and can result in reduced agricultural income for the country.
- Crop failure due to disease can lead to debt accumulation, poverty and migration from rural areas to urban centres in search of alternative livelihoods.
- It limits the country's ability to meet the demand for food for its growing population.
- Severe consequences for vulnerable population groups, worsening hunger and malnutrition.
- To mitigate the impact of plant diseases, investment in agricultural research and education is essential.

1.2 OVERVIEW

The interplay of agriculture and plant diseases greatly affects the Indian economy and the life of farmers. Solving the problems posed by plant diseases requires a multifaceted approach, including improved infrastructure, access to modern agricultural techniques and better disease management practices.

The severity of plant diseases is a major factor that directly affects overall yield crop quality. The degree of severity of the disease determines the extent of the damage caused plants and can significantly affect agricultural productivity. Monitoring and managing the production process requires a thorough understanding of the type of disease and its severity.

Deep learning has proven to be a powerful tool for identifying plant diseases and severity assessment. Deep learning, especially convolutional neural networks (CNNs), demonstrated remarkable abilities in analysing and interpreting complex patterns within images, making it an ideal tool for plant disease identification and severity assessment. By training these models on a diverse dataset of plant images showing different grades disease severity, they can accurately classify and quantify disease severity.

1.3 DEEP LEARNING MODEL

Deep learning, a subset of machine learning, involves training artificial neural networks with multiple layers to learn and predict from complex data. In deep learning, neural networks with multiple hidden layers are designed to automatically learn hierarchical data representation. Convolutional Neural Networks (CNNs) are commonly used in computer vision tasks, while recurrent neural networks (RNNs) are popular for sequential data analysis. One of the key benefits of deep learning is its ability to perform end-to-end learning. This means that models can learn directly from raw input data and output the desired results, eliminating the need for manual production function or explicit domain knowledge.

1.4 CNNs

CNN - stands for Convolutional Neural Network, which is a type of deep learning model widely used for computer vision tasks, especially in image recognition and analysis. CNNs are designed to automatically learn and extract relevant features from input images to make accurate predictions or classifications.

The key components are:-

- ✓ Convolution layer - This local patterns and structures, enabling the model to recognize edges, textures and other visual characteristics.
- ✓ Pooling layer -
  - Used to reduce spatial dimensions and control assembly.
  - Common pooling techniques - max pooling: where the maximum value inside an area is preserved, and average pooling, where the average value is calculated.
- ✓ Activation function - Introduces nonlinearities into the network, allowing model to learn complex relationships between input features.
- ✓ Fully interconnected layers-
  - Also known as dense layers which are responsible for high-level representation and making predictions.
  - These layers receive merged feature maps from previous layers and connect each neuron to each neuron in the next layer.
- ✓ Loss Function - The loss function measures the difference between predicted values of the CNN output and real output.
- ✓ Optimizers - update network parameters during the training phase, minimizing the loss function and improving the accuracy of the model.

1.5 VGG19

VGG19 is a deep convolutional neural network architecture developed by the company

Visual Geometry Group (VGG) at the University of Oxford. This is a variant of VGGNet, which has gained popularity for its simplicity and excellent image performance recognition tasks.

The "19" in VGG19 represents the number of layers in the network, inclusive convolutional layers, fully connected layers and pooling layers. VGG19 architecture consists of a stack of convolutional layers, followed by max-pooling layers to reduce spatial dimensions.

1.6 Problem Statement

The main purpose is to detect the diseased part of the plant leaf. Using Python, convolutional neural networks are implemented in order to classify the diseased part. Aim is to detect the diseased part by finding whether a particular image belongs to that particular class of disease.

CHAPTER 2

LITERATURE SURVEY

It reveals several remarkable studies and approaches in this field. Some key papers that contribute to the advancement of automatic plant disease prediction using CNN:

- “Deep Learning-Based Plant Disease Detection” by Mohanty et al. (2016):

This seminal work presents the PlantVillage dataset, which contains more than 50,000 images of 14 crop species with 26 different diseases. The authors propose a CNN-based approach using transfer learning with the VGGNet architecture. They achieved high accuracy in disease classification and demonstrated the potential of deep learning for plant disease detection.

- “Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks” by Pound et al. (2017):

This research focuses on the use of CNNs for complex plant phenotyping tasks, including disease identification. The authors are developing a deep learning platform called Deep Plant Phenomics that uses CNN architectures to automatically analyze and predict plant diseases based on images. The platform demonstrates the potential for deep learning in plant phenotyping and disease prediction.

- "Deep Transfer Learning for Crop Disease Classification and Severity Estimation" by Sladojevic et al. (2018):

This study investigates the effectiveness of deep transfer learning in plant disease classification and severity estimation. Using CNN architectures, including ResNet and Inception, the authors propose a transfer learning approach using pre-trained models on large image datasets. The results highlight the benefits of transfer learning for accurate disease classification and severity estimation.

- “Plant Disease Identification Using Deep Learning Techniques: A Review” by Singh et al. (2019):

This comprehensive review paper provides an overview of various deep learning techniques for plant disease identification. It discusses the application of CNNs, including popular architectures such as AlexNet, VGGNet, and ResNet, to various plant disease datasets. The paper also highlights the importance of data augmentation, transfer learning and model interpretability in plant disease prediction using deep learning.

- “Automatic Plant Disease Identification Model Based on Deep Learning” by Abdullah et al. (2020):

This paper presents a model for automatic plant disease identification using deep learning, specifically a CNN-based approach. The study evaluates different CNN architectures, compares the performance of different optimization techniques, and discusses the importance of dataset size and quality. The research provides insight into the implementation and challenges of deep learning for plant disease prediction.

These studies represent a fraction of extensive research on automatic plant disease prediction using CNN deep learning. They demonstrate the effectiveness of CNN architectures, transfer learning, and ensemble methods in accurately identifying and classifying plant diseases based on image data. Continued progress in this area holds great promise for improving crop management practices and promoting sustainable agriculture.

CHAPTER 3

METHODOLOGY

CNNs are capable of automatically learning and extracting meaningful representations from visual data. The hierarchical structure of the layers allows CNNs to capture both low- and high-level features, enabling them to achieve impressive performance in various computer vision tasks.

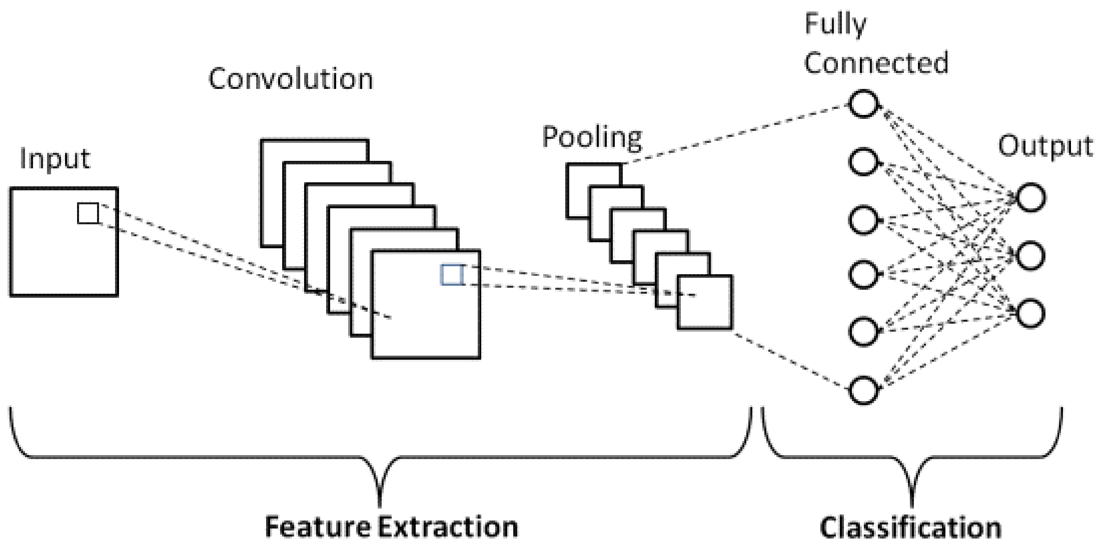


Figure 01: CNN architecture

The following steps were included in the development of the plant disease prediction system:

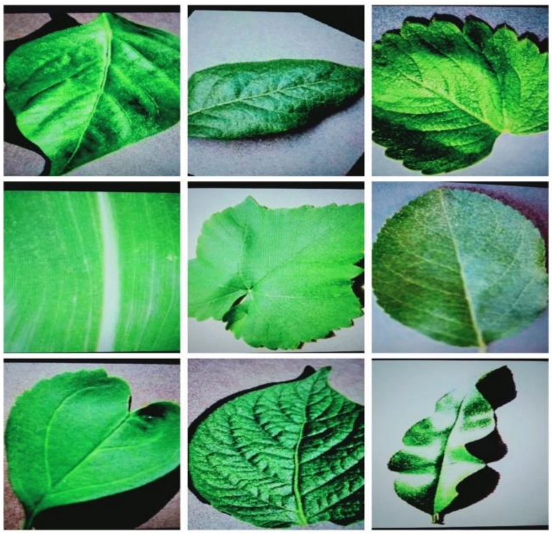
I. Importing important tools and framework and libraries -
   a. Tool - Google Colaboratory :
      - Cloud-based platform provided by Google.
      - Allows us to write and execute our Python code.
   b. Framework and Libraries -
      - TensorFlow: This is an open source machine learning library developed by Google that provides a wide range of tools and features for building and training machine learning models, especially deep learning models.

- Keras: an open-source API for high-level neural networks written in Python that can be easily integrated with other Python libraries such as NumPy, Pandas, and scikit-learn.
- NumPy: Its array manipulation, data preprocessing, and numerical computing capabilities make it an essential tool for processing and manipulating the image data necessary for plant disease prediction.
- Panda: Provides functionality to easily partition a dataset based on specific criteria, ensuring that the model is trained and evaluated on separate, non-overlapping subsets of the data.
- Matplotlib: A popular Python library used for data visualization. It is used to plot and visualize the input images to better understand the differences between healthy and diseased plant leaves.
- Os: This is the core library used in all the models as it allows the user to interact with the operating system and this library comes in python's standard utility modules.

II. Dataset Collection - A diverse and comprehensive dataset of plant leaf images containing healthy leaves and leaves affected by various diseases is obtained from Kaggle. The data set includes multiple plant species and different disease types, ensuring the generalizability of the model.

- The dataset consists of 87000 RGB images of healthy and unhealthy plant leaves having 38 classes.

Some of the images are shown below:-



III. Data pre-processing and feature extraction - pre-processed to remove noise, standardize image sizes, and enhance image quality.
   - Techniques used – data resizing, normalization, and augmentation are employed to create a balanced and informative dataset.

IV. Data augmentation – A technique used in machine learning and computer vision to increase the size and diversity of a training dataset by artificially creating new samples from existing ones.
   - The primary goal is to improve the generalization and robustness of machine learning models by exposing them to a wider range of variations and conditions.
   - By creating new samples, the model becomes more capable of handling different scenarios and reduces the tendency to overfit.

V. Model Selection - Various deep learning algorithms have been considered but Convolutional Neural Network (CNN) is chosen as the deep learning architecture due to its –
   - Effective Feature Extraction
   - Spatial Hierarchical Representation
   - Ability to learn from Large Datasets
   - Transfer Learning Capability
   - Real-time and Automated Analysis
   - Continuous Improvement

VI. Model Training - The selected machine learning model is trained using the pre-processed dataset.
   - The dataset is divided into two parts, one for training and one for testing. 80% of the dataset is for training, and 20% for testing.
   - The training dataset is used to train the model while the testing dataset is kept unseen so that accuracy of the model can be tested.

VII. Model Evaluation - The trained CNN model is evaluated on the validation set to assess its accuracy and performance.
   - Model's effectiveness is calculated using Evaluation metrics such as –
      - ✓ Accuracy
      - ✓ Precision
      - ✓ Recall

VIII. Model Summary - Get the summary of the model by using summary function directly.

IX. Callbacks - We import the Model Checkpoint callbacks to train the model. This will be used to store the weight of the model after training it or when early stopping occurs. Only the best weight will be saved as we specify that save_best_only = True. Using loss metric, the model is trained.

CHAPTER 4

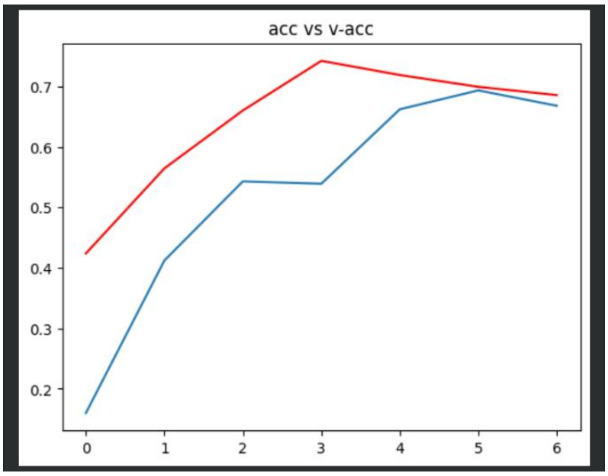
RESULTS AND DISCUSSIONS

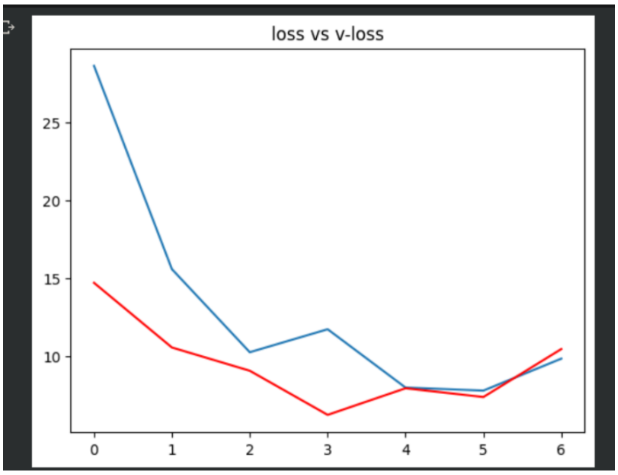
The model is created using the Python programming language. It is a popular, high- level programming language known for its clear syntax, dynamic semantics, and support for object- oriented programming.

The performance of the plant disease prediction system is evaluated using selected evaluation metrics. The accuracy of the model on a test dataset is measured to assess the effectiveness of the system in accurately predicting the disease.

```
Epoch 1/50
16/16 [==============================] - ETA: 0s - loss: 28.6111 - accuracy: 0.1602
Epoch 1: val_accuracy improved from -inf to 0.42383, saving model to best_model.h5
16/16 [==============================] - 28s 867ms/step - loss: 28.6111 - accuracy: 0.1602 - val_loss: 14.7136 - val_accuracy: 0.4238
Epoch 2/50
16/16 [==============================] - ETA: 0s - loss: 15.5905 - accuracy: 0.4121
Epoch 2: val_accuracy improved from 0.42383 to 0.56445, saving model to best_model.h5
16/16 [==============================] - 16s 976ms/step - loss: 15.5905 - accuracy: 0.4121 - val_loss: 10.5724 - val_accuracy: 0.5645
Epoch 3/50
16/16 [==============================] - ETA: 0s - loss: 10.2682 - accuracy: 0.5430
Epoch 3: val_accuracy improved from 0.56445 to 0.66016, saving model to best_model.h5
16/16 [==============================] - 14s 893ms/step - loss: 10.2682 - accuracy: 0.5430 - val_loss: 9.0813 - val_accuracy: 0.6602
Epoch 4/50
16/16 [==============================] - ETA: 0s - loss: 11.7398 - accuracy: 0.5391
Epoch 4: val_accuracy improved from 0.66016 to 0.74219, saving model to best_model.h5
16/16 [==============================] - 14s 900ms/step - loss: 11.7398 - accuracy: 0.5391 - val_loss: 6.2552 - val_accuracy: 0.7422
Epoch 5/50
16/16 [==============================] - ETA: 0s - loss: 8.0042 - accuracy: 0.6621
Epoch 5: val_accuracy did not improve from 0.74219
16/16 [==============================] - 14s 885ms/step - loss: 8.0042 - accuracy: 0.6621 - val_loss: 7.9595 - val_accuracy: 0.7188
Epoch 6/50
16/16 [==============================] - ETA: 0s - loss: 7.8098 - accuracy: 0.6934
Epoch 6: val_accuracy did not improve from 0.74219
16/16 [==============================] - 16s 1s/step - loss: 7.8098 - accuracy: 0.6934 - val_loss: 7.3996 - val_accuracy: 0.6992
Epoch 7/50
16/16 [==============================] - ETA: 0s - loss: 9.8567 - accuracy: 0.6680
Epoch 7: val_accuracy did not improve from 0.74219
16/16 [==============================] - 14s 877ms/step - loss: 9.8567 - accuracy: 0.6680 - val_loss: 10.4707 - val_accuracy: 0.6855
Epoch 7: early stopping
```

We plot a graph to illustrate the maximum accuracy the model achieved during training and validation while minimizing the loss.





From the above graphs, we observe that as the training accuracy increases, validation accuracy increases. Similarly as the training loss decreases, the validation loss decreases too.

A random sample of images is taken from the dataset and predicts the plant image's disease and class.

```
[ ] path = "/content/folder/New Plant Diseases Dataset(Augmented)/test/test/CornCommonRust3.JPG"

    prediction(path)

    1/1 [==============================] - 0s 24ms/step
    The image belongs to Corn_(maize)___Common_rust_
```

```
[ ] path = "/content/folder/New Plant Diseases Dataset(Augmented)/test/test/PotatoEarlyBlight1.JPG"

    prediction(path)

    1/1 [==============================] - 1s 734ms/step
    The image belongs to Strawberry___Leaf_scorch
```

DISCUSSIONS

By implementing a robust machine learning model such as CNN, the system can detect plant diseases at an early stage. Early detection enables farmers to take timely and targeted measures to effectively manage diseases.

Accurate disease classification achieved by the machine learning model ensures that farmers receive reliable information about the specific diseases affecting their crops. Armed with accurate disease predictions, farmers can quickly implement appropriate interventions.

Minimizing crop damage caused by diseases translates into better crop yields and higher profits for agricultural producers.

By improving crop management practices and minimizing the impact of the disease, the project contributes to sustainable agricultural practices. Sustainable agricultural practices promote long-term productivity while reducing negative environmental impacts.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This project has successfully developed a plant disease prediction system using deep learning techniques. By incorporating the power of deep neural networks, the system can accurately classify and identify plant diseases based on input images of plant leaves. One of the key advantages of the deep learning-based system is its ability to handle a wide range of plant species and diseases.

Implementing a plant disease prediction system can bring several benefits for farmers. By early recognition and mitigation of disease impact, farmers can avoid significant crop losses and achieve higher yields. Early disease detection can lead to cost savings as farmers can apply targeted treatments or interventions, reducing the need for broad-spectrum chemical pesticides and unnecessary expenditure.

Since the system can be integrated into user-friendly mobile applications or cloud platforms, it becomes accessible to farmers in remote areas, giving them access to real-time disease forecasts and expert recommendations.

FUTURE WORK

Potential future improvements to the plant disease prediction system include:

a. Data set expansion: Increase the size and diversity of the dataset to cover a wider range of plant species and diseases, allowing the system to handle a wider range of scenarios.

b. Real-time prediction: Develop a system capable of processing live image sources for real-time disease detection and early intervention.

c. Mobile application: Create a user-friendly mobile application that farmers can use to capture pictures of plant leaves and get instant disease predictions and recommendations.

d. Continuous model improvement: Implement techniques such as transfer learning and fine-tuning to increase the model's accuracy and robustness over time.

e. Multi-Sensor Integration: Integrating additional sensors, such as infrared or hyper spectral cameras, can provide additional information for more comprehensive disease diagnosis.

f. Multi-lingual Support: Enabling multi-lingual interfaces can make the system accessible to farmers from diverse linguistic backgrounds and ensure wider adoption.

REFERENCES

[1] GOOGLE

[2] YOUTUBE

[3] UDEMY

[4] DATASET FROM: New Plant Diseases Dataset | Kaggle

[5] https://www.researchgate.net/publication/353336310

[6] Recent advances in plant disease severity assessment using convolutional neural networks | Scientific Reports (nature.com)

[7] G. Shrestha, Deepsikha, M. Das and N. Dey, "Plant Disease Detection Using CNN," 2020 IEEE Applied Signal Processing Conference (ASPCON), 2020, pp. 109-113, doi: 10.1109/ASPCON49795.2020.9276722.