AgroSmart - The Smart Agriculture

Mini Project

Bachelor of technology

by

Boddu Sri Pavan - N180606
Bandi Vikram Kumar - N181167
Mannem Pavan Kumar - N180520
Mallidi Sandhyarani - N181108
Pragada Padma Priya - N180272

Under the Guidance of

Mr. T. Chandrasekhar

Department of Computer Science Engineering

Rajiv Gandhi University of Knowledge Technologies

Nuzvid - 521201, India

June 2, 2023

Dedicated to our beloved parents.

Certificate

This is to certify that the mini project report entitled **"AgroSmart- The Smart Agriculture"**, submitted by **Boddu Sri Pavan, Bandi Vikram Kumar, Mannem Pavan Kumar, Mallidi Sandhyarani, Pragada Padma Priya** belonging to **team12** , has been carried out under the guidance of **Mr.T.Chandrasekhar**, Computer Science Engineering, Rajiv Gandhi University of Knowledge Technologies, Nuzvid, Andhra Pradesh. The project is approved for submission requirement for Mini Project in the 6th semester in Computer Science Engineering from Rajiv Gandhi University of Knowledge Technologies.

The results contained in our project thesis have not been submitted in part or in fully to any other University or Institute for the award of any degree or diploma to the best of our knowledge.

....

Internal Examiner :

Date :

Declaration

We hereby declare that the project entitled- **"AgroSmart- The Smart Agriculture"** which is being submitted as a Mini Project of the 6th semester in Computer Science Engineering from Rajiv Gandhi University of Knowledge Technologies is an authentic record of our genuine work done under the guidance of **Mr.T.Chandrasekhar**, Dept of Computer Science Engineering from Rajiv Gandhi University of Knowledge Technologies.

This written submission represents our own ideas in our own words. Where others' ideas and words have been included, we have adequately cited and referenced the original source. We declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/ data/ fact/ source in our submission.

. . . .
Internal Examiner :
Date :

Boddu Sri Pavan
Bandi Vikram Kumar
Mannem Pavan Kumar
Mallidi Sandhyarani
Pragada Padma Priya

Acknowledgements

We express our sincere gratitude towards our guide **Mr. T. Chandrasekhar**, Computer Science Engineering, Rajiv Gandhi University of Knowledge Technologies for his valuable guidance, suggestions, and supervision throughout the work. Without his patronage and guidance, the project would not have taken shape. We would also like to express our regards for his kind approval of the project, time-to-time counseling, and advice.

We owe our sincere thanks to the Dept.of Computer Science Engineering, Rajiv Gandhi University of Knowledge Technologies for providing support to our project. We are grateful to the **Eswara** for the abundant grace and blessings bestowed upon us, which enabled us to successfully complete this project.

Abstract

Agriculture is the primary source in providing food for the entire world. Greater than 45% of the world and 70% of the Indian Population relies on agriculture for their livelihood. Around 50% of the loss in crop yield is reported due to pests and diseases. Soil fertility, climate, pests, and diseases are the four important factors that influence crop yield. Most of the farmers grow the same crops in their lands throughout the years. The wrong selection of crops leads to solid infertility as well as crop failure. Early prediction of diseases can save crops. Our project presents state-of-the-art models to predict the most suitable crop from minimum number of environmental features with an accuracy of 97.04% and to predict disease with a minimum number of parameters with a maximum accuracy of 93.61%.

Contents

List of Figures

Abbreviations

ML	**M**achine **L**earning
DL	**D**eep **L**earning
KNN	**K**- **N**earest **N**eighbors
SVM	**S**upport **V**ector **M**achine
AEFS	**A**scending **E**xhaustive **F**eature **S**election
SCC	**S**parse **Categorical** **C**ross-entropy
TP	**T**rue **P**ositive
TN	**T**rue **N**egative
FP	**F**alse **P**ositive
FN	**F**alse **N**egative
w.r.t	**w**ith **r**espect **t**o

Chapter 1

Introduction

1.1 Background

The agriculture sector is the backbone of Indian society, employing more than 70% of the population. It is primarily engaged in growing crops and livestock management. There are many challenges associated with agriculture including irrigation, soil fertility, temperature, natural calamities, diseases, and pest attacks. All these factors result in a decrease in crop yield.

Machine Learning converts data into models which facilitate predictions. It is the application of algorithms and statistics used to solve a problem accurately by using interfaces from previous data rather than instructions. It uses Supervised, Unsupervised and semi - supervised techniques to solve a problem. Deep Learning refers to the use of artificial neural network architectures that contain a large number of processing layers, as opposed "swallower" to architecture of more traditional neural network methodologies. The now computationally feasible deep learning models have revolutionized sectors such as image processing.Several challenging problems in architecture can be easily solved by Machine Learning and deep learning with loss efforts.

1.2 Problem Statement

i) To predict the best suitable crop for a field with as minimum number of input features as possible.

ii) To predict the disease from the image of a plant leaf with optimized CNN architecture.

1.3 Need of the Study

Generally, farmers grow the same crop year after year which reduces the availability of certain nutrients and degrades the soil. All these factors result in a decrease in crop yield. It severely affects people and causes famines.

Earlier farmers used to examine diseases through visual examination which is not an accurate method as not all farmers can observe and analyze accurately. In recent times agriculture scientists and experts perform certain chemical tests in laboratories and identify various diseases. This method requires a large group of experts and technical equipment. It also consumes time and money. In such conditions, the suggested techniques prove to be beneficial to farmers.

1.4 Study Objective

The main objective of our project is to solve crop selection problem and plant disease prediction in advance with minimum technical requirements so that crop yield can be raised and crop loss can be reduced with less cost.

1.5 Scope and Limitations

The scope of this study is to develop crop recommendation and plant disease prediction models that can help farmers improve their crop yields and reduce crop losses. However, the crop recommender model is only based on a limited dataset of crop yields, soil types, and climates. The plant disease detector model is only trained on a limited dataset of plant diseases and their symptoms.

1.6 Dissertation Organization

This dissertation consists of 6 chapters, including the present chapter (Chapter-1) of introduction to the research topic. This chapter describes the need for the study, the problem statement and objectives of the study, and a brief thesis outline. Chapter Chapter-2 provides a review of past literature, and the research gaps to be filled. Chapter-3 briefly describes the data collection. Chapter-4 presents exploratory data analysis. Chapter-5 describes the study of various algorithms, and the proposed methodology to fulfill the current objectives. Chapter-6 summarizes the results and research contributions of the current work. Chapter-7 concludes our study. Chapter-8 describes future scope.

Chapter 2

Literature Review

2.1 Introduction

Machine learning and deep learning provide better decision-making in real-world circumstances with minimal need for human interaction. It offers a robust structure for decision-making that is driven by data from the past. A comprehensive knowledge of crops, soil, and the surrounding environments is necessary for the development of a crop recommendation system. Several studies have been done but they have to be improved.

2.2 Soil Properties and Plant Growth

.Plants are autotrophic organisms that prepare food using the photosynthesis process, soil structure influences plant growth by affecting the movement if water, air, and nutrients to the plants, sandy soils have little or no structure but are often tree drains. Roots can easily penetrate with higher clay contents the soil's structural strength increases but its drainage ability often decrease. Throughout Earth's history, the natural cycling of nutrients has occurred from the soil to plants and animals, and then back to the soil, primarily through decomposition of biomass. Not all soils are suitable for growing crops. Ideal soils for agriculture are suitable for balanced in chemical and mineral components. The balanced combinations of these components allow for water relations and drainage.

N,P,K are the primary macro nutrients, where Nitrogen, Phosphorous, Potassium. Soil organic matter is the partial or well-decomposed residues of organic biomass present in the soil. It gives topsoil its deep black colors and rich aromas. It influences soil pH, and soil structure and improves water, holding capacity. pH, the master variable of soil, significantly influences the availability of plant nutrients, microbial activity, and even the stability of soil aggregates. At low pH, essential plant macronutrients(N, P, K, Ca, Mg, and S) are less bio-available than at higher pH rains near 7 and certain micronutrients(Fe, Mn, Zn) tend to become more soluble and potentially toxic to plants at low pH names. Typically, soil pH values from 6-7.5 are optimal for plant growth however certain plant species adopted for extreme pH values. Nitrogen and Phosphorous are the vital components in photosynthesis where the first forms the structure of chlorophyll and the other forms the structure of ATP(Adenosine Triphosphate), an energy unit, potassium eases the flow the water, nutrients, and carbohydrates through the xylem plants phloem of plants.

2.3 Climatic Factors Affecting Crop Yield

.The climate is the dominating factor influencing the suitability of a crop in a particular region. The yield potential of the crop mainly depends on climate. The most important climatic factors that influence the growth, development, and yield of crops are solar radiation, temperature, and rainfall. Each plant has its own cardinal temperature points (minimum, maximum, and optimum). Lower temperatures affect several aspects of crop growth. Water freezes into ice crystals in the intercellular spaces which results in cell dehydration. At higher temperatures plant cells are severally injured and killed. Formation of a thick cover of ice/snow on the soil surface presents the entry of oxygen which suffers the crop. Atmospheric humidity is the water present in the atmosphere in the form of invisible vapor. Relative humidity influences the water requirement of crops. When relative humidity is high there is a chance for the outbreak of pest and disease. Rainfall and irrigation determine how fast a crop can be grown from seed to harvest. Crops are dependent on water during their entire lifecycle in order to survive and thrive. When rainfall is meaningful, crops can absorb the water and the nutrients they need and the soil around them remains moist and fertile. This allows the plants to grow and produce large, and healthy yields. However, too much rain can be detrimental to crop health. Exercise rainfall can lead to flooding, which can wash away soil nutrients and change crops. Additionally, too much waterlogged can cause root rots and diseases. Insufficient rainfall can lead to drought conditions and negatively impact crop growth and yields.

2.4 Literature Review

Paper[1] consider the features namely soil type, pH, value, humidity, temperature, wind and rainfall and propose a KNN algorithm with an accuracy of 89.4%.

Paper[2] nitrogen, phosphorous, potassium, rainfall, temperature, humidity, and pH features set with 22 crops are considered. Several algorithms are compared and random forest yield the highest accuracy of 93%.

Paper[3] proposes naive bayes and support vector machine algorithms for crop recommendation with an accuracy of 90% by considering features namely temperature, humidity, pH level, and soil moisture.

Paper[4] proposes crop prediction using naive Bayes classifier with an accuracy of 97% ratio of train and test data set is 70:30 temperature, humidity,pH,rainfall,seed,soil moisture are the input features an android application has been developed for farmers.

Paper[5] made a comparative study of several algorithms and concluded Naive Bayes algorithm with the highest accuracy of 94.73%. Nitrogen,phosphorous,potassium, humidity,temperature,pH and rainfall are the features considered.

Paper[6] took Nitrogen,phosphorous,potassium,temperature,humidity,pH,rainfall ,electric conductivity and organic carbon as input features and compared several algorithms.Neural network is proposed with an accuracy of 87%.

Paper[7] worked on a data set containing 35,000 images with 32 different classes of plant varieties and diseases.Various data augmentation adam optimizer and categorical cross entropy loss is used and got training,testing accuracies as 96.5% and 90.0% respectively. Batch size and number of epochs are set to 32,75 respectively. As training accuracy is lesser than testing accuracy,it seems that model is over fitting.

Paper[8] collected a small data set with 500 images and divided it into 70% :30% for training and testing respectively.They constructed a CNN with 3 convolutional layers with RMSProP optimizer.Batch Norm,label smoothing have been included.For 10 epochs an accuracy of 94.80% has been achieved.

Paper[9] worked on a small data set with 500 images of tomato leaves. CNN with LVQ algorithm is proposed with an accuracy of 86%.

Paper[10] considered a data set with 7800 images of 5 different crops with a total of 21 classes. By using CNN with mobileNelv2 pre-trained model an accuracy of 90.38% has been achieved.

Paper[11] considered a data set of 20,636 images with 15 classes. A complex CNN architecture is constructed with 4 convolution layers. By training upto 15 epochs, train,test accuracies are 98.29%, 98.029%.

Paper[12] compared various algorithms and ResNet pre-trained model generated the highest accuracy of 99.92%. A data set with 10 different classes is trained for 50 epochs.This study is only on tomato plant diseases.

Paper[13] proposes a random forest classifier for apple,corn,grape,potato,and tomato plant diseases classification and achieved an accuracy of 93%.

2.5 Research Gaps

Efficiency of a model is determined by its architectural,time and space complexities many papers considered more than five input features to recommend a suitable crop for the field[1][2][3][4][5][6], as input of more features require more hardware equipment and technology which in turn raises cost for recommendation. This causes burden to most of the small farmers across our nation. Training accuracy must be greater than or equal to testing accuracy to avoid under fitting[1]. There is a necessity to extent number of classes as different crops are grown across different parts of out nation[8] [9][11][12][13]. Data set should be large enough to classify diseases more accurately[8][9][10][3].CNN architecture should be simple enough so that time and space complexities are reduced[8][11].

2.6 Summary

After the introduction of convolutional neural networks in deep learning, all the domains are triggered with technology. It provided simple solutions with its complex architecture.This chapter focuses on recent works about crop recommendation and,disease prediction. People used various algorithms and made their contributions. However efficiency of an model is determined by it's architectural , time and space complexities there is a need of highly efficient model which should be beneficial in terms of technology and money to most of the farmers across India.

Chapter 3

Data Collection

3.1 Introduction

Data is the heart of machine learning and deep learning. Models are built from past data. Data must be large enough to capture hidden insights. The collection includes various techniques like interviews, field surveys, and, online forms.

3.2 Data Collection

Crop recommendation[14] to predict crop and PlantVillage[15] datasets to predict diseases are collected from Kaggle repository. The PlantVillage (87,867) is recreated using several augmentation techniques from the original source[16] (54,303).

3.3 Summary

This chapter mentioned the importance of data collection, and resources.

Chapter 4

Exploratory Data Analysis

4.1 Introduction

Exploratory data analysis is an approach to analyzing datasets to summarize their main characteristics often using statistical graphics and other data visualization methods. This phase reveals data distributions and valuable insights hidden in the data.

4.2 Exploratory Data Analysis

Crop recommendation dataset consists of a total of 2200 samples with 7 attributes namely Nitrogen, Phosphorous, Potassium, temperature, humidity, pH, and rainfall. All these features are numeric. There are 22 different crops: Rice, Maize, Chickpea, Kidney beans, Pigeon peas, Mothbeans, Mungbean, Blackgram, Lentil, Pomegranate, Banana, Mango, Grapes, Watermelon, Muskmelon, Apple, Orange, Papaya, Coconut, Cotton, Jute, and Coffee. Each crop contributes 100 samples. There are no null, missing, or duplicate entries present in the data set.

Correlation expresses the extent to which two features are linearly related.

Let x,y are two variables and n is the number of samples then Pearson's correlation coefficient is determined by:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

FIGURE 4.1: Formula for Pearson's correlation coefficient

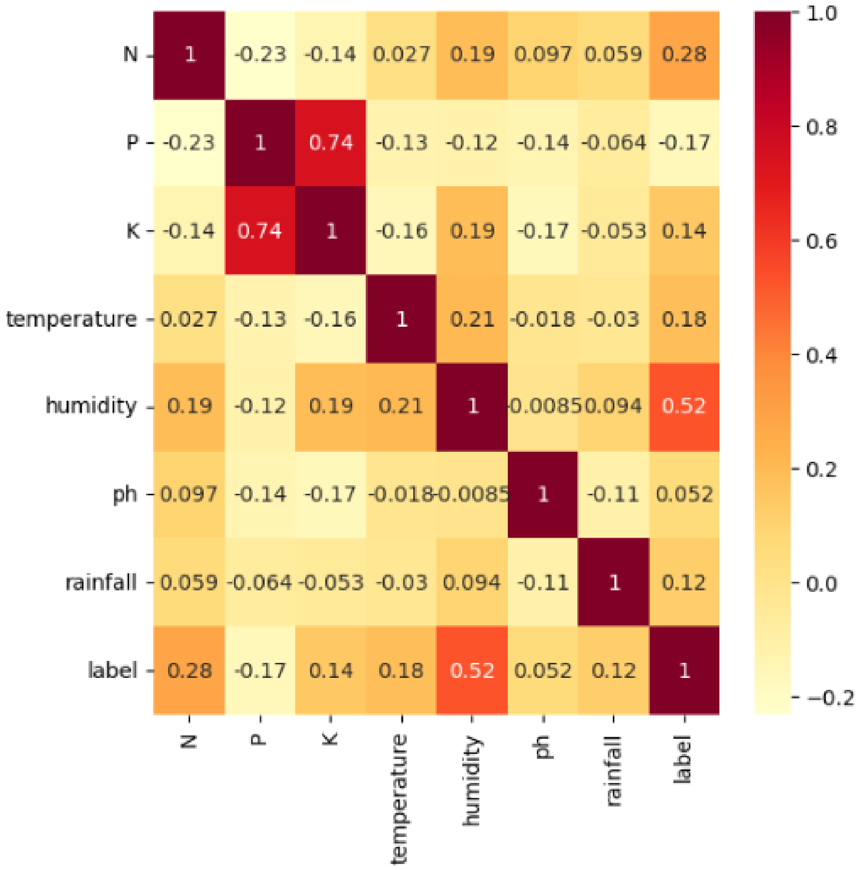
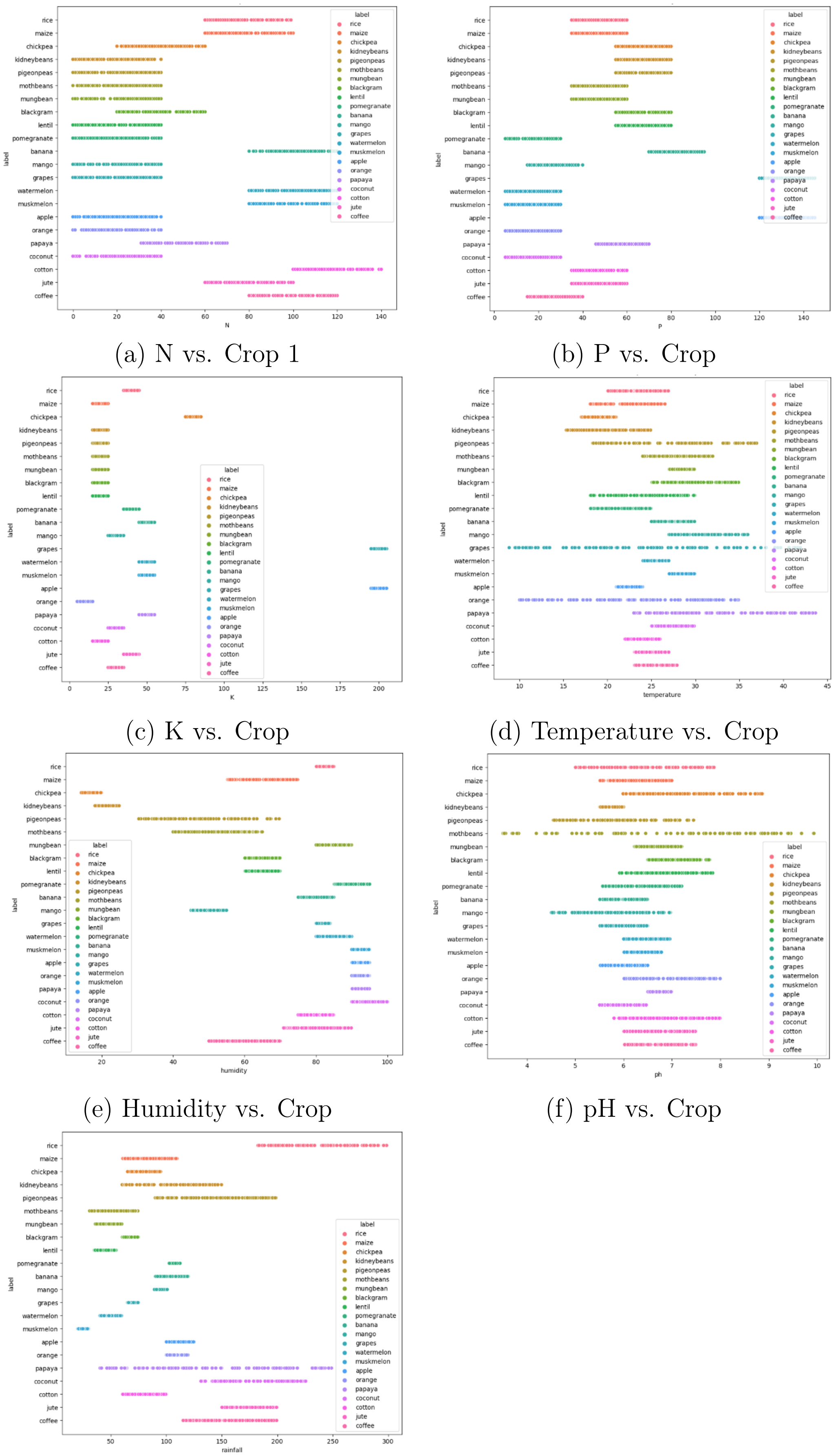


FIGURE 4.2: Heatmap showing the correlation among all the features

The main drawback of correlation is, it can not determine non-linear relationships between two variables.



(a) N vs. Crop 1

(b) P vs. Crop

(c) K vs. Crop

(d) Temperature vs. Crop

(e) Humidity vs. Crop

(f) pH vs. Crop

(g) Rainfall vs. Crop

FIGURE 4.3: Scatterplots showing distribution of crops w.r.t. features

From the scatterplots, it is observed that each crop clustered in certain intervals.

Data set is divided into 80% and 20% for training, and testing.

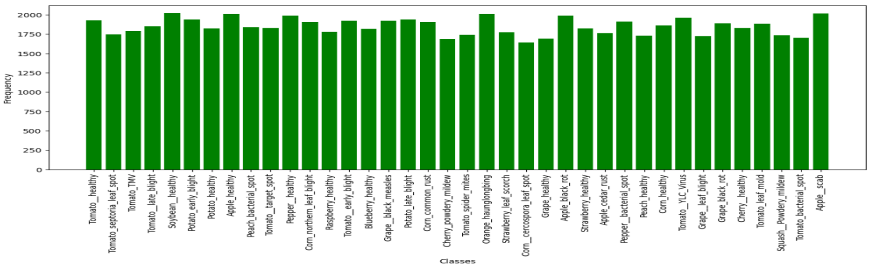


FIGURE 4.4: Frequency distribution in training dataset

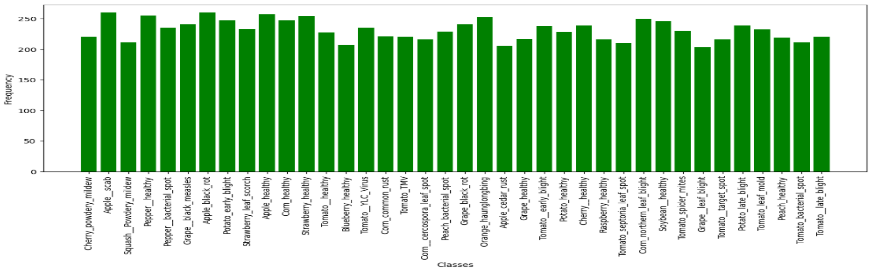


FIGURE 4.5: Frequency distribution in validation dataset

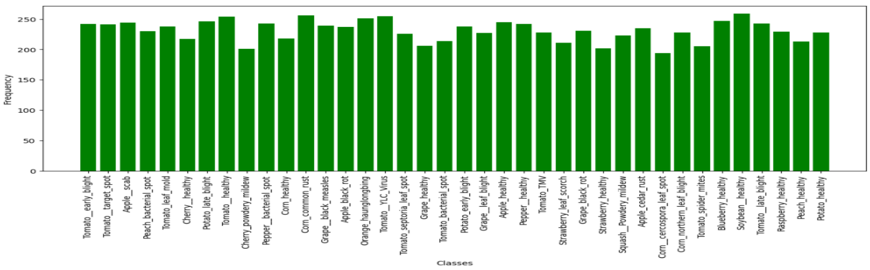


FIGURE 4.6: Frequency distribution in testing dataset

Plant Disease prediction dataset consists of a total of 87,867 RGB images of healthy and diseased crop leaves which are categorized into 38 different classes. The dataset is divided into an 80/10/10 ratio of training, validation, and test sets preserving the directory structure. The batch size is set to 64. Image height and width are set to 100px, and 100px respectively. All classes are uniformly distributed in each split. As the dataset is already augmented explicit augmentation can be omitted.

4.3 Summary

This chapter deeply explores datasets using various statistical visualization techniques such as scatterplots, heatmaps, and histograms.

Chapter 5

Methodology

5.1 Introduction

Machine learning converts data into models which facilitate predictions. Deep Learning refers to the use of artificial neural network architectures that mimic human neural system to make better futuristic decisions.

5.2 Models and algorithms

1) K-Nearest Neighbors algorithm finds K closest data points to a query point and makes predictions based on their labels/values for classification/regression tasks.
2) Support Vector Machines algorithm finds the optimal hyperplane to separate data points into classes in a linearly separable feature space.
3) Decision Tree algorithm recursively partitions the feature space into subsets based on the most informative features, creating a tree-like structure for classification or regression tasks

4) Random Forest algorithm is an ensemble learning method that constructs multiple decision trees and combines their predictions for improved accuracy in classification and regression tasks.

5) Extra Trees Classifier algorithm is an ensemble learning method that constructs multiple decision trees with random splits and combines their predictions for improved accuracy in classification tasks.

6) Logistic Regression is a statistical model that uses a logistic function to estimate the probability of a binary or categorical outcome based on input features.

7) Naive Bayes algorithm is a probabilistic method that applies Bayes' theorem with the "naive" assumption of independence between features to make predictions in classification tasks.

5.3 Methodology for crop recommendation

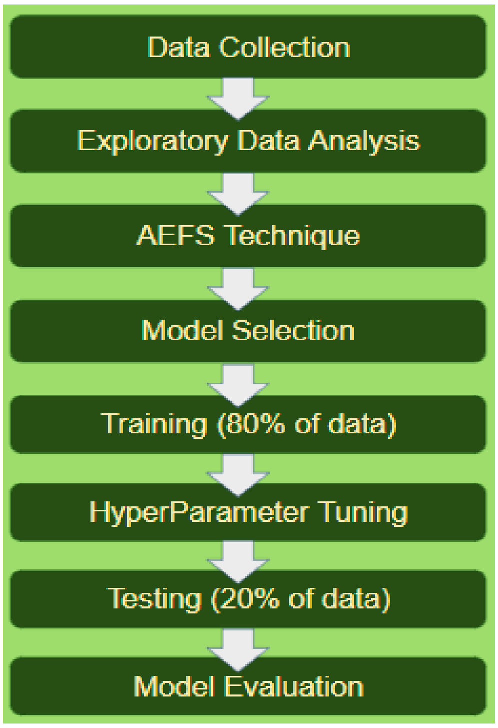


FIGURE 5.1: Proposed methodology for crop recommendation system

5.3.1 Ascending Exhaustive Feature Selection Algorithm

Ascending Exhaustive Feature Selection (AEFS) is a modified exhaustive feature selection technique that finds out the smallest possible subset of features yielding maximum accuracy threshold. Although EFS is a brute force technique AEFS is efficient for data sets with a small number of attributes. In this study, the random forest classifier is considered as the baseline model because it classifies clusters of linearly-separable data efficiently. Threshold accuracy is set to 95%

5.3.2 Random Forest Algorithm

Random Forest classifier is a non-parametric, eager learner, which is an ensemble of randomly subsampled decision trees.

Algorithm :

Step- 1 : Consider no.of decision trees to be built as number of estimators.

Step- 2 : Bootstrap sampling: Generate randomized samples with replacement from the original dataset.

Step- 3 : Randomized feature splits: Random subset of features are chosen.

Step- 4 : Generate decision tree and iteratively construct number of estimators no.of decision trees.

Step- 5 : Predict the test sample from an ensemble of decision trees and take the result of majority voting as the final prediction.

5.3.3 Hyperparameter Tuning

Hyperparameter tuning is the process of choosing a set of optimal hyperparameters for a model.

number of estimators : The number of trees in the forest (value= 500)

max-features : The number of features to consider when looking for the best split (value= sqrt)

max-depth : Maximum depth of the tree (value= 10)

criterion : Function to measure the quality of a split (value= gini)

random state : To generate reproducible results (value= 111)

5.3.4 Model Evaluation

K-fold cross-validation is a technique in which the entire data set is partitioned into K-equal folds. In each iteration, the model is tested with one of the folds, and the remaining folds are used for training.The accuracy obtained in all iterations is averaged. This method controls overfitting and guarantees accuracy for the model.

5.4 Methodology for plant disease prediction

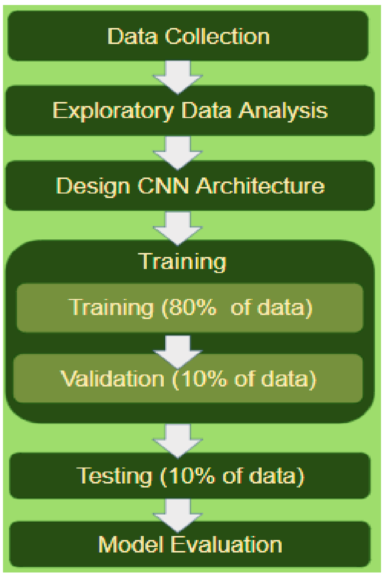


FIGURE 5.2: Proposed methodology for plant disease prediction system

5.4.1 Convolutional Neural Network

Convolutional Neural Network(CNN) is a network architecture for deep learning which is composed of several layers of perceptions. Different layers of CNN are as follows:

1. Convolution layer: The convolution layer acts as a filter to extract a feature using the back-propagation technique. The sliding window moves over the entire image and obtains feature maps. Activation functions are used to introduce non-linearly into the output of neurons. Regularization functions are used to control overfitting. Images are normalized before they are given as input to this layer.

2. Pooling layer: The pooling layer is used to reduce the size of the feature map using the sliding window technique.

3. Dense layer: Dense layer contains a single layer of artificial neurons that act as a hidden layer. Images are flattened into 1-D before this layer.

4. Output layer: The final layer of neurons is called the "Output layer".This layer should contain the total number of neurons equal to the total number of classes.

5.4.2 Regularization techniques

Regularization techniques are used to calibrate the model in order to minimize the adjusted loss function and present overfitting or underfitting.

1. l2 Regularization: It forces the weights to decay towards zero but not exactly to zero.

2. Batch Normalization: It is a normalization technique done between the layers of the neural network along min-batches instead of the full dataset to speed up training and make learning easier by using higher learning rates.

3. Dropout Regularization: It is the practice of disregarding certain neurons in a layer at random during the training phase. It breaks circumstances committed by prior layers making the model more robust. It avoids the overfitting problem.

5.4.3 CNN Architecture

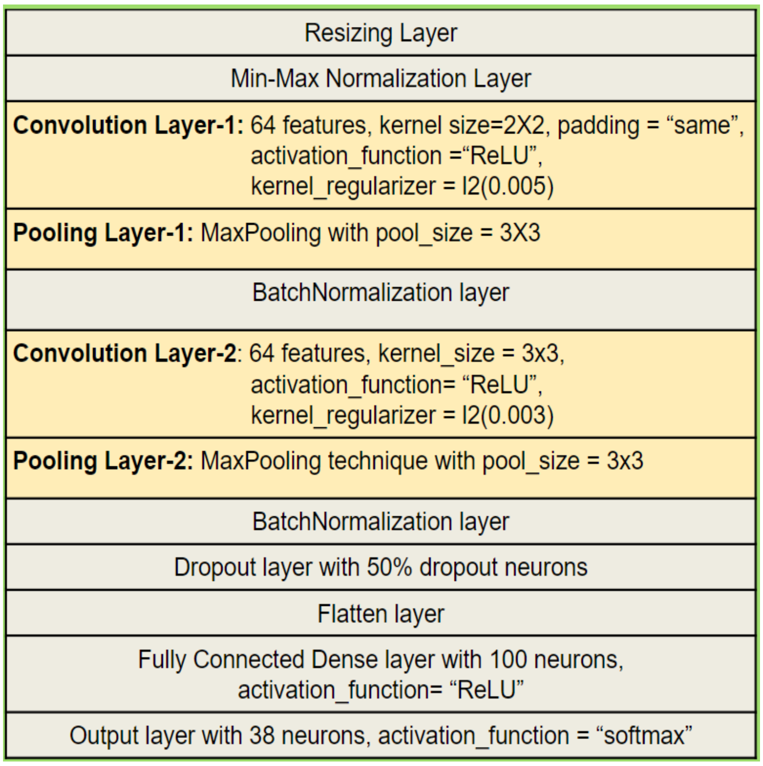


FIGURE 5.3: Proposed CNN architecture

Adam optimizer, sparse categorical cross-entropy loss and accuracy sparse categorical cross-entropy metrics used. The model is trained for 80 epochs. While training validation data set has been used. The remaining parameters are used as default in TensorFlow.

5.5 System Specifications

Processor : 12th Gen Intel(R),Core(TM),i7-12700 2.10GHz

Installed RAM : 16.0GB (15.7 GB usable),

System type : 64-bit operating system, x64-based processor,

Operarting System : Windows 11

Programming Language : Python

Tool(s) : Jupyter Notebook

5.6 Summary

This chapter clearly discussed our proposed system to recommend crops and to predict diseases.

Chapter 6

Observations and Results

6.1 Introduction

All the results are tabularized. Several visualization techniques are used to analyze the results.

6.2 Metrics

True positive is an outcome where the model correctly predicts the positive class.
True negative is an outcome where the model correctly predicts the negative class.
False positive is an outcome where the model incorrectly predicts the positive class.
False negative is an outcome where the model incorrectly predicts the negative class.
The Confusion matrix visualizes and summarizes the performance of a classification algorithm.

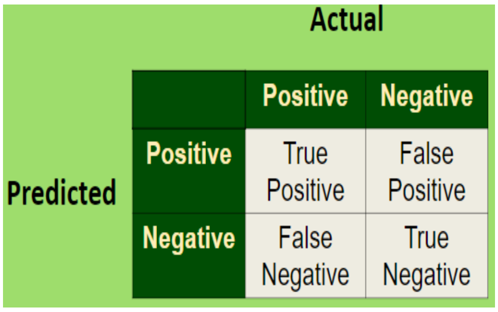


FIGURE 6.1: Confusion Matrix

Accuracy: The fraction of all instances is the classifier's prediction correct (for either positive or negative class).

Accuracy = (TP+FP)/ (TP+TN+FP+FN)

Classification Error: The fraction of all instances is the classifier's prediction incorrect.

Classification Error = (FP+FN)/ (TP+TN+FP+FN)= 1- Accuracy

Recall: The fraction of all positive instances does the classifier correctly identify as positive.

Recall = TP/(TP+FN)

Precision: The fraction of positive predictions are correct

Precision = TP/(TP+FP)

F1-score: F1-score is the harmonic mean of recall and precision

F-1 Score = (Precision . Recall)/ (Precision+Recall)

6.3 Crop Recommendation Model

Ascending exhaustive feature selection algorithm resulted in feature set

= {k, Humidity, Rainfall}. Several Machine Learning algorithms are trained and the results obtained are tabularized as follows:

Sl. No	Algorithm	Training Accuracy	Testing Accuracy	F1- Score
1.	KNN	95.80%	92.95%	92.95%
2.	SVM	90.74%	89.55%	89.55%
3.	Decision Tree	100%	93.41%	93.41%
4.	Random Forest	97.73%	97.05%	97.05%
5.	Extra trees	100%	95.23%	95.23%
6.	Gradient Boosting	100%	96.59%	96.59%
7.	Logistic Regression	96.25%	94.77%	94.77%
8.	Naive Bayes	68.58%	68.18%	68.18%

FIGURE 6.2: Model Selection

Among all the algorithms Random forest classifier performed best with an accuracy of 97.05%. To validate the model k-fold cross-validation with k=50 is executed, which gave almost the same accuracy as 96.64%. As both values are closer the model is guaranteed. The confusion matrix for random forest classifier is as follows:

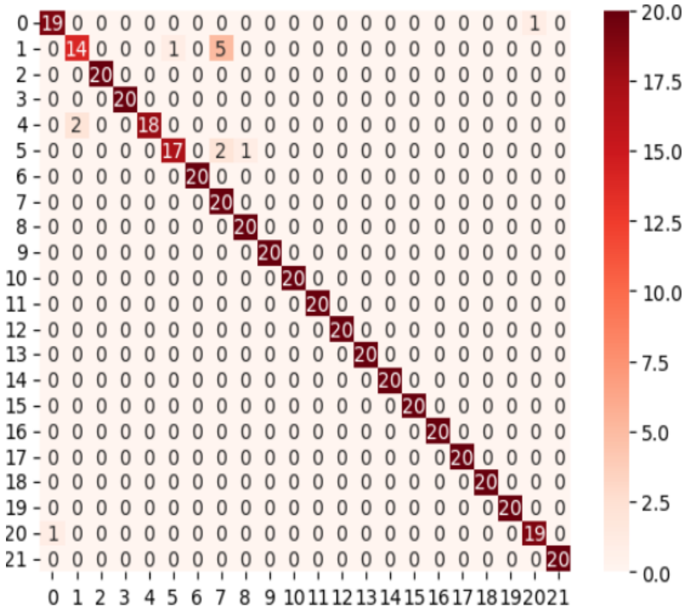


FIGURE 6.3: Confusion matrix for Random forest classifier

6.4 Plant Disease Prediction Model

Epoch-wise observatory curves are as follows:

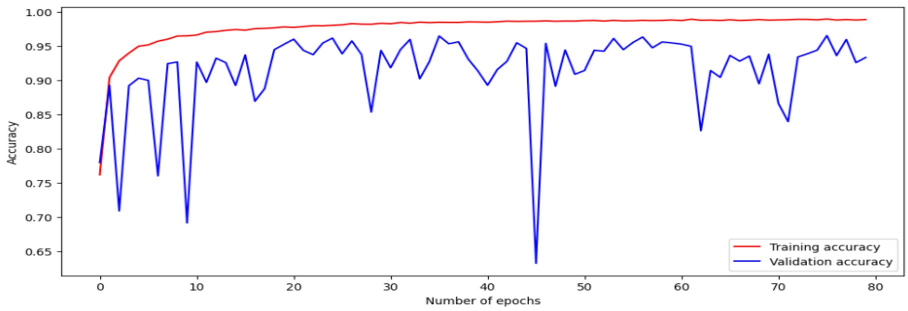


FIGURE 6.4: Epoch-wise accuracy curves for training and validation sets

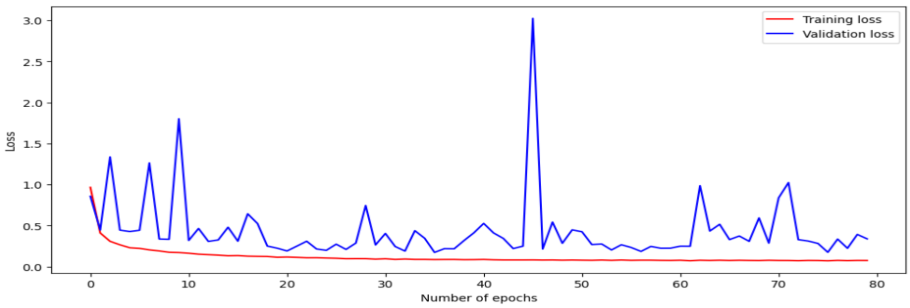


FIGURE 6.5: Epoch-wise loss curves for training and validation sets

As it is an ideal case to nullify the loss, it is minimized to a great extent. Validation accuracy is good when compared to training accuracy. So, we conclude that our model is a good fit.

6.5 Summary

This chapter outlines the findings and conclusions of the study, as well as the practical implications that can be drawn from the results.

Chapter 7

Implementation and Deployment

7.1 Introduction

Web application development is the process of creating software applications that are delivered over the Internet or a local network. The first step is to design the web application. This includes defining the features of the application, the user interface, and the data model. Once the design is complete, the web application is developed. This involves coding the application in a programming language, such as HTML, CSS, JavaScript, and a backend framework.

7.2 Technologies Used

1) HTML : HTML stands for Hypertext Markup Language which provides various tags to create headings, input fields, buttons etc.

2) CSS : CSS stands for Cascading Style Sheets which adds and feels to the pure HTML pages through various colors, margins, decoration, and effects.

3) Bootstrap : Bootstrap is a library built on CSS that provides different types of classes with the same functionalities of CSS to optimize code. It increases reusability.

4) JQuery : Jquery is an optimized library it performs the same as naive javascript. It contains AJAX, Asynchronous javascript XML notations which are used to develop single-page web applications.

5) Flask : Flask is a Python lightweight backend engine that serves users' requests at ease. It is highly scalable in nature.

7.3 Interface Screenshots

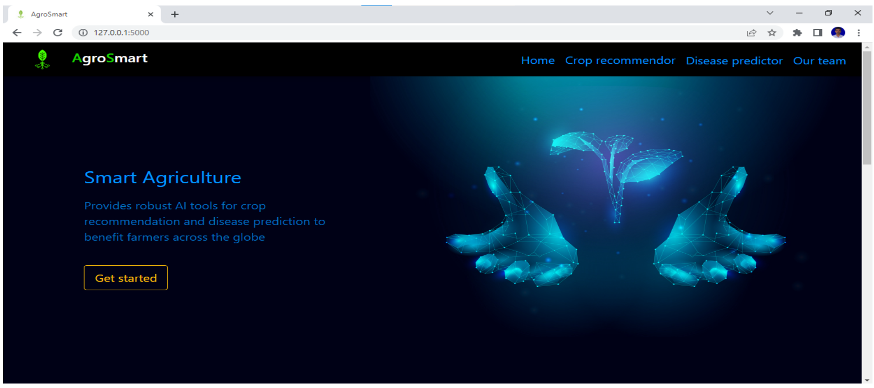


FIGURE 7.1: Homepage

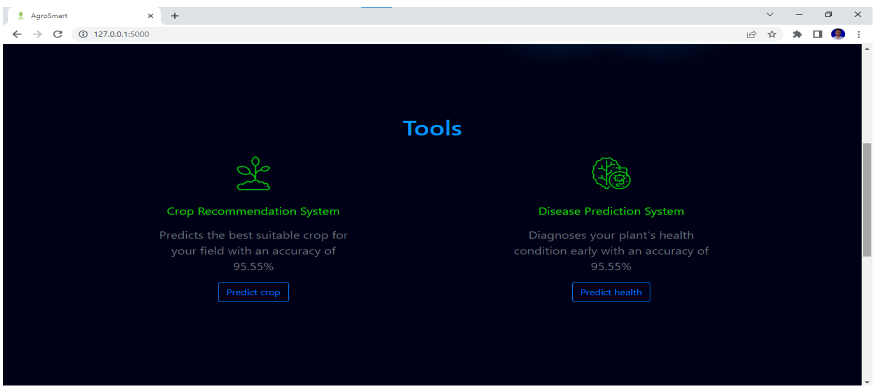


FIGURE 7.2: Homepage (scrolled)

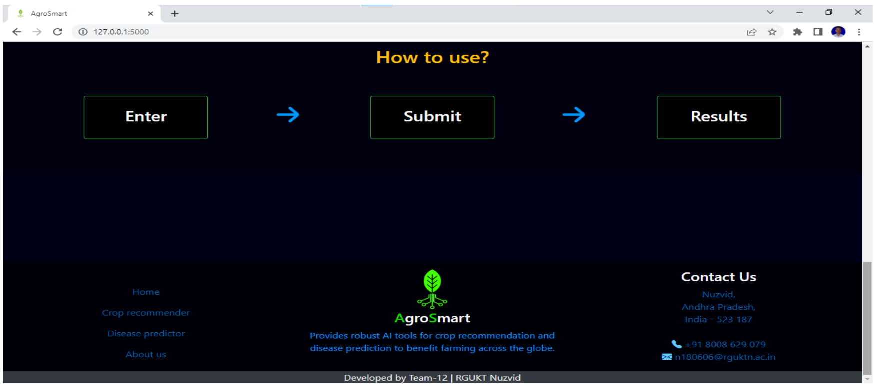


FIGURE 7.3: Homepage (scrolled)

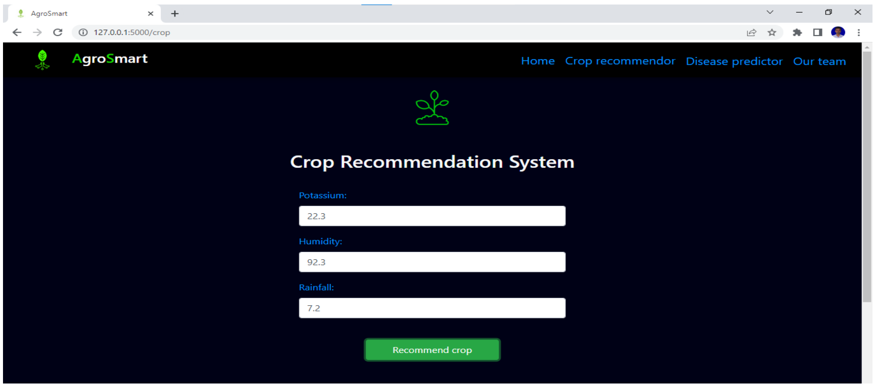


FIGURE 7.4: Crop recommendation page (input view)

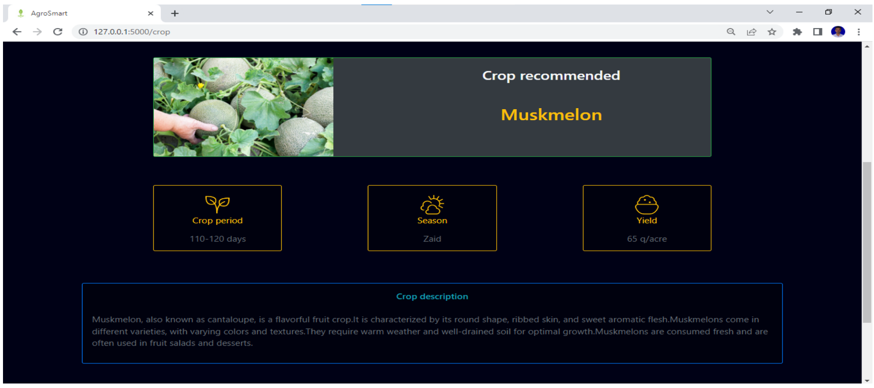


FIGURE 7.5: Crop recommendation page (output view)

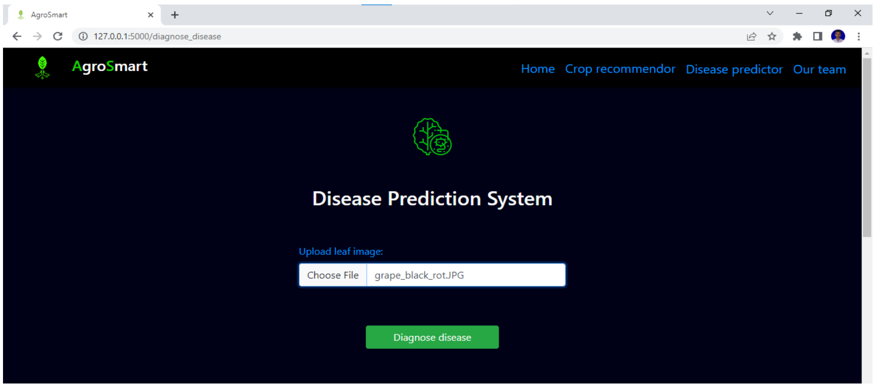


FIGURE 7.6: Disease prediction page (input view)

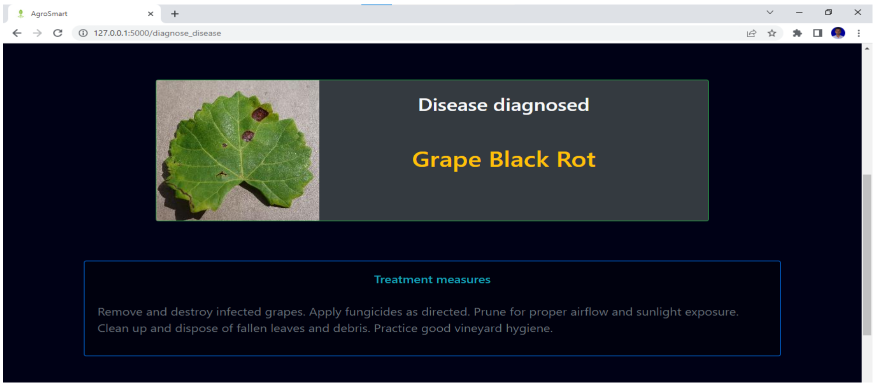


FIGURE 7.7: Disease prediction page (output view)

Chapter 8

Impact and Future Work

8.1 Significance of the Research

Our research has the potential to make a significant impact on agriculture. By helping farmers select the right crops and detect plant diseases early can help them improve their crop yields and reduce crop losses. This will contribute to food security and economic development in many parts of the world.

Crop recommender model can help farmers select the right crops for their land and climate. This is important because different crops have different requirements in terms of soil type, climate, and water. By selecting the right crops, farmers can improve their crop yields and reduce crop losses. Plant disease detector model can help farmers detect plant diseases early. This is important because plant diseases can spread quickly and cause significant crop losses. By detecting plant diseases early, farmers can take steps to prevent the spread of the disease and protect their crops. Integration of these ML models into a web interface makes them even more accessible to farmers. This means that farmers can use these models to improve their agricultural practices without having to be experts in ML.

8.2 Potential Applications

The web application is a valuable tool for farmers. It can help farmers improve their crop yields and reduce crop losses. The application is also a great way to educate farmers about crop selection and plant disease prevention. It can help farmers select the right crops for their location, soil type, and climate. This can lead to improved crop yields. It further helps farmers detect plant diseases early. This can help farmers reduce crop losses. Farming practices are optimized as well as modernized.

Governments can use this system to provide farmers with information and resources to help them improve their agricultural practices. NGOs can use this system to provide farmers with education and training on crop selection and plant disease prevention. Researchers can use this system to collect data on crop yields, plant diseases, and agricultural practices. This data can be used to improve our understanding of agriculture and to develop new crop varieties and disease-resistant crops.

This system could be used to help farmers in developing countries. These farmers often lack access to the information and resources they need to improve their agricultural practices. It could help them select the right crops for their land, detect plant diseases early, and get the information and resources they need to improve their crop yields. This is also helpful for farmers in developed countries. These farmers often face challenges such as climate change, pests, and diseases. It could help them adapt to these challenges and improve their crop yields.

8.3 Future Enhancements and Extensions

We have developed web application for crop recommendation and disease prediction. As a future work, we are planning to design a full-fledged mobile application where farmers can enter, promote their products and sell their products directly to customers. We want to increase crop recommendation for at least 250 crops and disease prediction for at least 500 classes.

- Improve the accuracy of the crop recommender model: This could be done by collecting more data on crop yields, soil types, and climates. It could also be done by using more sophisticated machine learning algorithms.

- Improve the accuracy of the plant disease detector model: This could be done by collecting more data on plant diseases and their symptoms. It could also be done by using more sophisticated machine learning algorithms.

- Include more features: Extend the crop recommender model to include other factors, such as the price of crops and the availability of water. This would allow farmers to make more informed decisions about which crops to grow.

- Include more diseases: Extend the plant disease detector model to include other types of plant diseases. This would allow farmers to detect a wider range of diseases and take steps to prevent them from spreading.

- Mobile application development: Develop a mobile app for the crop recommender and plant disease detector models. This would make the models more accessible to farmers who do not have access to the internet.

Chapter 9

Conclusion

9.1 Summary of Findings

Agriculture is a vast sector. Although there are many challenges in farming, it is the amount and quality of yield which benefits farmers. Our study mainly focuses on two main challenges of selecting suitable crop for a field and early detection of plant diseases. We have built robust machine learning models for crop recommendation with an accuracy of 97.05%, f1-score of 97.05%, and disease prediction with an accuracy of 93.61%, f1-score of 97.05%. As it is ideal to nullify loss in CNN, we have used regularization techniques to minimize it to a greater extent.

9.2 Contributions to knowledge

- Open-source contribution of optimized models in GitHub
- Developed new plotting technique “IQR Plot” for statistical summary
- Proposed an approach to generate confusion matrix using TensorFlow

References

[1] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate "Crop Prediction System using Machine Learning Algorithms", IRJET volume: 07, issue: 02, pp. 748-753 February 2020,

[2] Mahendra Choudhary, Rohit Sartandel, Anish Arun, Leena Ladge "Crop Recommendation System and Plant Disease Classification using Machine Learning for Precision Agriculture" *Artificial Intelligence and Communication Technologies, pp. 39-49, 2023

[3] Thewahettige Harinditha Ruchirawya, Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P "Crop Recommendation System", Sri Lanka Institute of Information Technology, Volume 175– pp. 22-25 October 2020,

[4] M.Kalimuthu, P.Vaishnavi, M.Kishore "Crop Prediction using Machine Learning" IEEE International Conference on Smart Systems and Inventive Technology(ICSSIT), pp. 926-932 october 2020,

[5] Aman Sinha, Pallavi Sinha, Ritika Rajani, Dr. Sumithra Devi K.A "Crop Recommendation System using Machine Learning", IJARIIE Vol-8 Issue-3 pp. 4471-4474,2022 ,

[6] Varshitha D.N., Savita Choudhary "An artificial intelligence solution for crop recommendation" Indonesian Journal of Electrical Engineering and Computer Science Vol-25, pp. 1688 1695, March 2022,

[7] Sammy V. Militante, Bobby D. Gerardo, Nanette V.Dionisio "Plant Leaf Detection and Disease Recognition using Deep Learning", IEEE pp.579-582, 2019,

[8] Fatma MARZOUGUI, Mohamed ELLEUCH, Monji KHERALLAH,"A Deep CNN Approach for Plant Disease Detection"International Arab conference on Information Technology, IEEE, 2020,

[9] Melike Sardogan, Adem Tuncer, Yunus Ozen "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm" (UBMK'18) 3rd International Conference on Computer Science and Engineering IEEE pp.382-385, 2018,

[10] Raida Moyazzoma, Md. Al Amin Hossain, Md. Hasanuzzaman Anuz and Abdus Sattar "Transfer Learning Approach for Plant Leaf Disease Detection Using CNN with Pre-Trained Feature Extraction Method Mobilnetv2" IEEE, pp.526-529, 2021

[11] Marwan Adnan Jasim, Jamal Mustafa AL-Tuwaijari, "Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques" IEEE, International Conference on Computer Science and Software Engineering (CSASE) pp.259-265, 2020

[12] Ashish Chandak, Anshul Sharma, Aryan Khandelwal,Raunak Gandhi, "Detection of Diseases in Tomato Plant using Machine Learning" International Journal of Next-generation computing, Vol 13, pp.942-952, November 2020

[13] Pranesh Kulkarni, Atharva Karwande, Tejas Kolhe, Soham Kamble, Akshay Joshi, Medha Wyawahare "Plant Disease Detection Using Image Processing and Machine Learning" Vishwakarma Institute of Technology, june 2021.

[14] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[15] https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset

[16]https://github.com/spMohanty/PlantVillage-Dataset

END