A

PROJECT REPORT ON

A GEO-INTELLIGENT CROP ADVISOR AND MARKET PREDICTOR SYSTEM FOR FARMERS

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
FOR THE FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND TELECOMMUNICATION

SUBMITTED BY

ADITYA VILAS CHAVAN	**Exam No. B190073025**
TEJAS GOPAL KULKARNI	**Exam No. B190073084**
DARSHAN VIJAY PAKHALE	**Exam No. B190073108**
ANUSHRAY NITIN PINGALE	**Exam No. B190073119**

UNDER THE GUIDANCE OF

Prof. A. S. Najan

Department of Electronics and Telecommunication Engineering Pune Vidyarthi Griha's College of Engineering and Technology & G K Pate(Wani) Institute of Management, Pune -411009

Academic Year 2023-2024

Undertaking

We, the students of Department of Electronics and Telecommunication Engineering, PVG's College of Engineering and Technology & G. K. Pate (Wani) Institute of Management, Parvati, Pune-411009.

ADITYA VILAS CHAVAN	Exam No.: B190073025
TEJAS GOPAL KULKARNI	Exam No.: B190073084
DARSHAN VIJAY PAKHALE	Exam No.: B190073108
ANUSHRAY NITIN PINGALE	Exam No.: B190073119

do hereby undertake on 10.04.2024 the following:

1. We know the **University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations, 2018, and Plagiarism Policy of Savitribai Phule Pune University Pune.**
2. We also know that **Department of Electronics and Telecommunication Engineering, PVG's College of Engineering and Technology & GKPIM, Parvati, Pune has established Departmental Academic Integrity Panel (DAIP) as per UGC regulations 2018** as mentioned above in (1).
3. We are aware of the consequences if found guilty of any act of plagiarism defined in the UGC regulations 2018 as mentioned above in (1).
4. We shall abide by the rules, regulations, and code of conduct for the students of UGC, SPPU, and Department of E&TC.
5. We further undertake and declare that the thesis submitted by us is scanned using anti-plagiarism software as decided by the department and report of the same has been submitted and is free from any kind of plagiarism mentioned above (1) in **UGC REGULATIONS, 2018.**
6. I understand that non-compliance with the **Academic Integrity and Prevention of Plagiarism may result in disciplinary action on us as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations, 2018.**

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION ENGINEERING PUNE VIDYARTHI GRIHA'S COLLEGE OF ENGINEERING AND TECHNOLOGY & G K PATE(WANI) INSTITUTE OF MANAGEMENT, PUNE -411009

CERTIFICATE

This is to certify that the preliminary project report entitled

A GEO-INTELLIGENT CROP ADVISOR AND MARKET PREDICTOR SYSTEM FOR FARMERS

Submitted by

ADITYA VILAS CHAVAN	**Exam No. B190073025**
TEJAS GOPAL KULKARNI	**Exam No. B190073084**
DARSHAN VIJAY PAKHALE	**Exam No. B190073108**
ANUSHRAY NITIN PINGALE	**Exam No. B190073119**

towards the fulfillment of the degree of Bachelor of Engineering in Electronics and Telecommunication Engineering to be awarded by the Savitribai Phule Pune University, Pune at Pune Vidyarthi Griha's College of Engineering and Technology & G K Pate (Wani) Institute of Management, Pune during the academic year 2023-2024.

Prof. A. S. Najan
Project Guide

Dr. K. J. Kulkarni
Head
Department. of E&TC

DR. M. R. Tarambale
I/C Principal

Place: Pune
Date: -

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on **'A GEO-INTELLIGENT CROP ADVISOR AND MARKET PREDICTOR SYSTEM FOR FARMERS'**.

We would like to take this opportunity to thank our guide **Prof. A. S. Najan** for giving us all the help and guidance we needed. We are really grateful to him for his kind support. His valuable suggestions were very helpful. We are also thankful to the non-teaching staff for their valuable suggestions. Also, we are thankful to the institute for providing us valuable resources.

We are also grateful to Head of **Department of E&TC, Dr. K. J. Kulkarni, Pune Vidyarthi Griha's College of Engineering and Technology & G K Pate(Wani) Institute of Management** for her indispensable support and suggestions.

ADITYA VILAS CHAVAN
TEJAS GOPAL KULKARNI
DARSHAN VIJAY PAKHALE
ANUSHRAY NITIN PINGALE

ABSTRACT

The work begins with a brief overview of types, design techniques, research challenges, and objectives discussed. The previous work done by researchers related to the crop prediction system using different Machine Learning approaches. It is a web-based application which is helpful for the farmers. Over the most recent couple of years analysts have been keen ashore planning and its arrangement for different reasons. The intention behind the expansion is that soil strength is fundamental, so growing interest in farmland and soil condition research is fundamental to solid yield generation. The image sequence is one of his methods of studying soil and land health. It’s an amazing way to take into account the influence of different components. This paper proposes the investigation of flow and explores the issues it tends to and its possibilities. The focus is on a logical analysis of numerous cutting-edge grouping systems and techniques. To improve the precision of these methodological characterization, an attempt was made to look at the factors that gave rise to them. A convolutional neural network algorithm is used to classify the soil images in 4 categories Red, Black, Alluvial, Clay. A random forest algorithm is used to suggest crops based on soils also, we are predicting the humidity, rainfall, and temperature. Using KNN we are recommending the shop to the user. In this project, we achieve 81.25% accuracy by using random forest and 99% by the known model.

Keywords: Convolutional Neural Network, Image Processing, Deep Learning, Machine Learning, Random Forest, K-Nearest Neighborhood.

Contents

List of Figures

List of Tables

LIST OF ABBREVIATION

ABBREVIATION	**ILLUSTRATION**
CNN	Convolutional Neural Network
RF	Random Forest
KNN	K Nearest Neighbourhood
SVM	Support Vector Machine
AI	Artificial Intelligence
ML	Machine Learning
ReLu	Rectified Linear Unit
pH	Potential of Hydrogen
N, P, K	Nitrogen, Phosphorus, Potassium
RGB	Red Green Blue
Sqrt	Square Root of
SLR	Systematic Literature Review
SDLC	Software Development Life Cycle

CHAPTER 1

INTRODUCTION

1.1 Introduction of the Proposed System

Farming plays an essential part in the Indian economy. Over 70% of rustic families depend on farming. Farming is a huge piece of the Indian economy, representing around 17% of all Gross domestic product and utilizing approximately 58% of the populace. Throughout many years, India's farming has developed at a quick speed. Alongside the development of a quick populace, these inventive advances are fundamental to fulfilling the requirements of every individual. From the beginning of recorded history, farming has been the essential and most significant movement connected with each civilization and civilization in mankind's set of experiences. To take care of the issue through data investigation, information mining involves isolating stowed-away examples from huge datasets and laying out a relationship among them. Research has profited from the introduction of data mining in the horticultural area. To lay out the essentials, portrayal is pivotal in all logical fields. Finding variety among the items and ideas may be useful. It likewise gives essential data that empowers methodical investigation. One of the critical parts of a cultivating field's capacity to deliver crops is the dirt. The presence of information and good judgment conditions guide soil game plan thinking. Gathering of soil associates soil tests with a few sorts of particular substances on the world's territory surfaces. characterizing soil has turned into a profoundly popular issue in PC vision and picture handling. To make the calculation as exact as could be expected, various new calculations are being created utilizing convolutional structures. The extraction of even pixel-level highlights is currently feasible because to convolutional structures. The objective of this venture is to make a double-face classifier that can extricate highlights like edges, variety, and surface paying little heed to arrangement. This review presents a procedure for precisely characterizing soil from input pictures of any size. Utilizing different substance properties, CNN strategies recognize soil pictures, and potential harvests for that dirt series are proposed utilizing geographic qualities and SVM. It is an electronic instrument that is incredibly gainful to ranchers. The rancher will sell all things considered. His item is accessible online too without entering the market during this pandemic.

Machine learning (ML) approaches are used in many fields, ranging from supermarkets to evaluate the behavior of customers to the prediction of customers' phone use. Machine learning has also been used in agriculture for several years. Crop yield prediction is one of the challenging problems in precision agriculture, and many models have been proposed and validated so far. This problem requires the use of several datasets since crop yield depends on many different factors such as climate, weather, soil, use of fertilizer, and seed variety. This indicates that crop yield prediction is not a trivial task; instead, it consists of several complicated steps. Nowadays, crop yield prediction models can estimate the actual yield reasonably, but a better performance in yield prediction is still desirable. Machine learning, which is a branch of Artificial Intelligence focusing on learning, is a practical approach that can provide better yield prediction based on several features. Machine learning (ML) can determine patterns and correlations and discover knowledge from datasets. The models need to be trained using datasets, where the outcomes are represented based on experience. The predictive model is built using several features, and as such, the parameters of the models are determined using historical data during the training phase. For the testing phase, part of the historical data that has not been used for training is used for performance evaluation purposes. An ML model can be descriptive or predictive, depending on the research problem and research questions. While descriptive models are used to gain knowledge from the collected data and explain what has happened, predictive models are used to make predictions. ML studies consist of different challenges when aiming to build a high-performance predictive model. It is crucial to select the right algorithms to solve the problem at hand, and in addition, the algorithms and the underlying platforms need to be capable of handling the volume of data. To get an overview of what has been done on the application of ML in crop yield prediction, we performed a systematic literature review (SLR). A Systematic Literature Review (SLR) shows the potential gaps in research on a particular area of a problem and guides both practitioners and researchers who wish to do a new research study on that problem area. By following a methodology in SLR, all relevant studies are accessed from electronic databases, synthesized, and presented to respond to research questions defined in the study. An SLR study leads to new perspectives and helps new researchers in the field to understand the state-of-the-art. Crop yield prediction is an essential task for the decision-makers at national and regional levels for rapid decision-making. An accurate crop yield prediction model can help farmers decide on what to grow and when to grow. There are different approaches to crop yield prediction. This review article has investigated what has been done on the use of machine learning in crop yield prediction in the literature. During our analysis of the retrieved publications, one of the exclusion criteria is that the publication is a survey or traditional review paper.

The designed system will recommend the most suitable crop for a particular land. Based on region, season and soil content such as Nitrogen, Phosphorus , Potassium, and ph. The system takes the required input from the user such as Soil Image, Region, Season, Nitrogen, Phosphorus , Potassium, and ph. All input data applies to machine learning predictive algorithms like CNN, Random Forest and KNN to identify the pattern among data and then process it as per input conditions. The system recommends the crop for the farmer based on the input given and also gives the required rainfall, humidity and temperature.

1.2 Motivation

• The main purpose of the proposed work is to create a suitable model for classifying various kinds of soil data.

• Along with suitable crop suggestions as well as providing the factory details for selling the crops which is very much helpful for the farmer.

• The Motive behind the proposed work is to achieve higher accuracy over existing work by using machine learning.

.

1.3 Objective

• To detect the soil type using CNN.

• To predict the crops using Random Forest.

• To improve the performance of the model.

• To predict shops using KNN.

• To develop the most effective crop-predicting system using a machine learning algorithm.

1.4 Outcomes

• To recommend the crops.

• To recommend the Ph, rainfall, and Temperature.

• To recommend the industry.

1.5 Organization of the Report

The work begins with an abstract in which we define the purpose of the project where design techniques, research challenges, and objectives are discussed. The first chapter is the introduction to the project in that motivation, an overview of the project, problem statement objective, and outcomes of the system. The next chapter is the literature survey, this document gives an overview of the functionality of the product. It describes the informal requirements and is used to establish a context for the technical requirements specification in the next chapter. The third chapter, the Requirements Specification section, of this document, is written primarily for the developers and describes in technical terms the details of the functionality of the product. Both sections of the document describe the same software product in its entirety, but are intended for different audiences and thus use different language.

CHAPTER 2

LITERATURE SURVEY

D.Jayanarayana Reddy et al.[1] stated that, a systematic review that extracts and synth-size the features used for CYP, and there are a variety of methods that were developed to analyze crop yield prediction using artificial intelligence techniques. The major limitations of the Neural Network are a reduction in the relative error and decreased prediction efficiency of Crop Yield. Similarly, supervised learning techniques were incapable of capturing the nonlinear bond between input and output variables and faced a problem during the selection of fruit grading or sorting. This paper explores various ML techniques utilized in the field of crop yield estimation and provides a detailed analysis in terms of accuracy using the techniques.
Kryzhevsky et al. [2] state that a deep Convolutional Neural Network can classify his 1.2 million high-resolution images of the ImageNet LSVRC-2010 competition into many of his classes. In our test data, the author achieved top 1 and top 5 error rates of 37.5 and 17. This is also a significant improvement on the previous state-of-the-art. A neural network with 60,444 million parameters and 650,000 neurons consists of 5 convolutional layers, some of which are followed by max-pooling layers, and 3 fully connected layers are finally It becomes a 1000-way softmax. To speed up the training, we used highly efficient GPU implementations of unsaturated neurons and convolution operations. To reduce the overfitting of the fully connected layers, the author used a recently developed regularization method called “dropout”. This has proven to be very effective. We also submitted a variant of this model to the ILSVRC-2012 competition and achieved a top 5 test error rate of 15.3 compared to 26.2 achieved by the second-best entry.
According to Dharesh Vadalia’s suggestion [3], farmers plant crops using a conventional approach without understanding the soil’s composition or quality. Farmers won’t make enough money from their farms as a result. its current approach to soil.
According to M. P. K., Anthiyur et al. [4] consider a variety of factors, including the soil, climate, rainfall, composts, and pesticides. Contrastingly, a few factors have an impact on farming, which may be assessed by using systems that fit measurable variables. The work’s objective is to examine several information mining techniques that offer the highest degree of exactness.
According to Sk Al Zamnur Rahman et al. [5] predict soil series about land type and, by prediction, suggest appropriate crops. several machine learning techniques
Madhuri Shripathi Rao et al. [6] expressed that this paper expects to find the best model for crop expectation, which can assist ranchers with choosing the sort of yield to develop in light of the climatic circumstances and supplements present in the dirt. This paper looks at famous calculations, for example, K-Closest Neighbor (KNN), Choice Tree, and Arbitrary Timberland Classifier utilizing two unique rules Gini and Entropy. Results uncover that Irregular Woodland gives the most noteworthy exactness among the three.
Madhuri Shripathi Rao et al. [7] expressed that this paper expects to find the best model for crop expectation, which can assist ranchers with choosing the sort of yield to develop

in light of the climatic circumstances and supplements present in the dirt. This paper looks at famous calculations, for example, K-Closest Neighbor (KNN), Choice Tree, and Arbitrary Timberland Classifier utilizing two unique rules Gini and Entropy. Results uncover that Irregular Woodland gives the most noteworthy exactness among the three.

CHAPTER 3

DESIGN OF THE PROPOSED SYSTEM

3.1 Block Diagram of the System

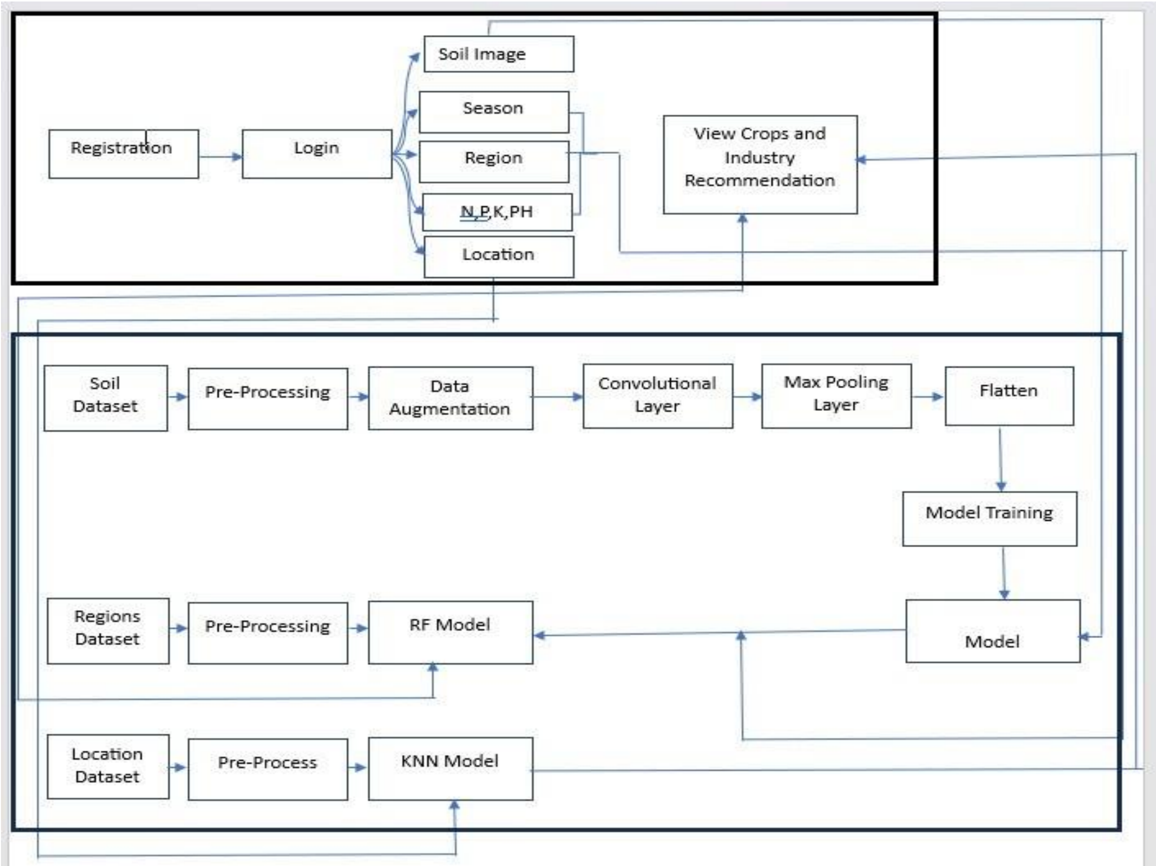


Figure 3.1: Block diagram

In above block diagram it consists of two phases one is Training part and second is the testing part. The Training part contains in a proposed system, we are proposing experiment on the prediction of crops and temperature, Humidity, and rainfall with a limited set of supervised data. Here a framework is proposed for soil arrangement, which is subsequently portrayed in the square chart. Also, we are predicting the industry for farmers by using the KNN model. For that, we are considering the farmer's location. In the system we have built three different algorithms one is CNN for which we passed the soil images dataset which passed to the pre-processing block. In that, we simply resize the image and then it goes to the data augmentation block. Results increase the size of the dataset. Then we built a CNN model and saved it. The second model is a random forest that model we have passed the N, P and K, pH data and based on that it will predict the crops. The final model we have used is the KNN. For that, we have passed the latitude and longitude data of specific shops based on that it will predict the shops. Simply we have created a website using the Django framework. This website predicts the crops and shop information to users. To access this website user wants to create an account on a website. Once login the user needs to pass the soil image, season, Region, N, P, K, Ph and Location. Based on that system will predict the crops and shop information.

3.2 Assumption and Dependencies

- **Assumptions: -**Reliable identification and classification of speech involves human judgment of several factors and a combination of experiences, a decision support system is desirable in this case.
- **Dependencies: -**Algorithm The algorithm uses structured data from the online platform. To the best of our knowledge, none of the existing work focused on both data types in the area of big data analytics. It is reducing violence by giving accurate diagnoses.

3.3 Functional Requirement

3.3.1 System Feature 1(Functional Requirement)

- Soil Photo
- Region/place
- Location/address

3.3.2 System Feature2 (Functional Requirement)

Database Requirements :

- Database –SQLite3

3.4 External Interface Requirement

3.4.1 User Interfaces

- Python

Python interface is being actively developed right now. There are many algorithms and many functions that compose or support those algorithms. Open CV is written natively in C++ and has a template interface that works seamlessly with STL containers.

- Image Processing

Read and Write Images. Detection of images and their features. Detection of shapes like circles, rectangles, etc. in an image, Detection of coins in images. Text recognition in images. e.g. Reading Number Plates. Modifying image quality and colors.

3.4.2 Hardware Interfaces

• To run our project, we required a hardware system that is feasible for our project like an Intel I3 processor, 2 GB RAM, and 20GB Hard disk. We also need a standard keyboard, Mouse, and LED Monitor.

3.4.3 Software Interfaces

• The system can use Microsoft as the operating system platform. The system also makes use of certain GUI tools. To run this application, we need Pycharm and above as a Python platform. To store data, we need an SQLite database.

3.4.4 Communication Interfaces

1] Crop recommend System
2] User-soil image data set item
3] Pre-processing unit
4] Classified answer in the form of predictions
5] Open-CV for image process

3.5 Nonfunctional Requirements

3.5.1 Performance Requirements

Performance: The performance of our system is fast as compared to other systems and response-se time is quick.
Availability: Availability of data is also a requirement for performing any operations.
Maintainability: The system is reliable for crop prediction.
Security: In this system user information is stored in the form of audio, so our system is secure.

Usability: This system is a very useful assistive tool for the public sector.

3.5.2 Safety Requirements

This study is carried out to check the economic impact that the system will have on the organization. The amount of funds that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system is well within the budget and this was achieved because most of the technologies used are freely available.

3.5.3 Security Requirements

The main thing in our system is, we have to provide end to end security for user and provider signs by using proper authentication login credentials. User has been given rights to upload speech and only view the results .The system is fully secure as well as eco- friendly. We implemented this system by considering security aspects so we divide our system into four different modules for achieving integrity.

3.5.4 Software Quality Attributes

Capacity: Capacity of project according to data it is very less.
Availability: Proposed system will available on python application.
Reliability: System is reliable for crop prediction.
Security User when login to system that time users mail id and password match accurately.

3.6 System Requirements

3.6.1 Database Requirements

- Database - SQLite

3.6.2 Software Requirements

- Operating System:- Windows
- Front End - Python3x
- Database -SQLite3
- IDE - Py-Charm

3.6.3 Hardware Requirements

- Processor - I3
- Speed - 1.1 GHz
- RAM - 2 GB (min)
- Hard Disk - 20 GB
- Floppy Drive - 1.44 MB
- Camera -System camera
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

3.7 Analysis Models: SDLC Model to be applied

Waterfall Model – The waterfall model is a sequential model that is used in the software development processes, where the process is seen flowing steadily downwards through the phases of Requirement Gathering and Analysis, System Design, Implementation, Testing, Deployment, and Maintenance.

1) Requirement analysis:

Here requirements are gathered means which kind of dataset is required. Then what are the functional requirements of the system? A document is prepared, and then use cases are designed. In our system, we gather all information on the admin and user and functionality of each module.

2) System Design:

In this stage, the hardware and software requirements to design the system are decided. It uses the above-mentioned hardware and software requirements. We design the admin and user modules. Design the according to the functionality of each module.

3) Implementation: In this stage, the system is developed module-wise. In this system consist of mainly 2 modules That is

1. Admin
2. User

4) Testing

In this stage, all developed software is installed and they are tested in different ways against the system requirements. In this stage, we check whether all these modules are working properly or not with proper authentication.

5) Deployment:

In this deployment stage we deployed the new functionality of each module like dataset gathering, pre-processing, feature extraction, and classification. We deploy all systems with proper functions.

6) Maintenance:

According to the software's new version and their use, they need to be updated. some predefined machine learning libraries need to be used. This system is easy to maintain

3.8 System Implementation Plan

Sr. No.	Task	Labour days/Hours/Days
1	Topic Selection	2 Weeks
2	Feasibility Study	1 Weeks
3	Project Design	2 Weeks
4	Develop Functional Specification	10 Weeks
5	Develop System Architecture	2 Weeks
6	Develop Detailed Design Specification	2 Weeks
7	Data Collection and Environmental Setup	2.5 Weeks
8	Project Development	10 Weeks
9	Perform Module Integration	3 Weeks
10	Perform Testing	3 Weeks
11	Perform Testing	1 Weeks

Figure 3.2: Project Estimate

3.9 Data Flow Diagrams

3.9.1 Data Flow Diagram Level-0

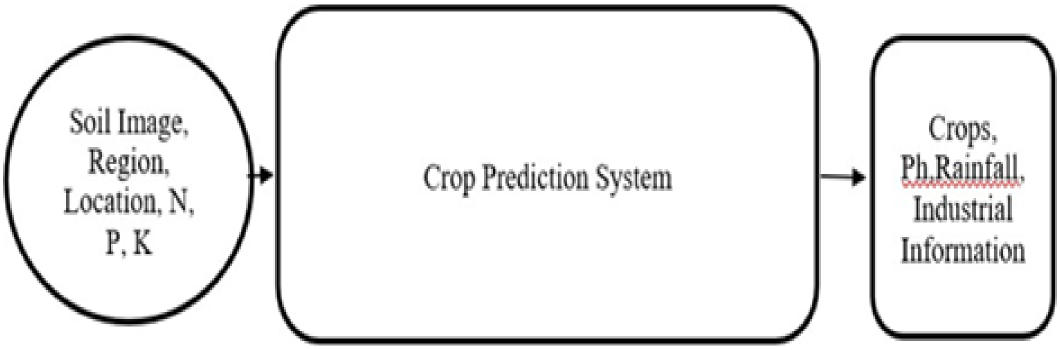


Figure 3.3: Data Flow Diagram level-0

3.9.2 Data Flow Diagram Level-1

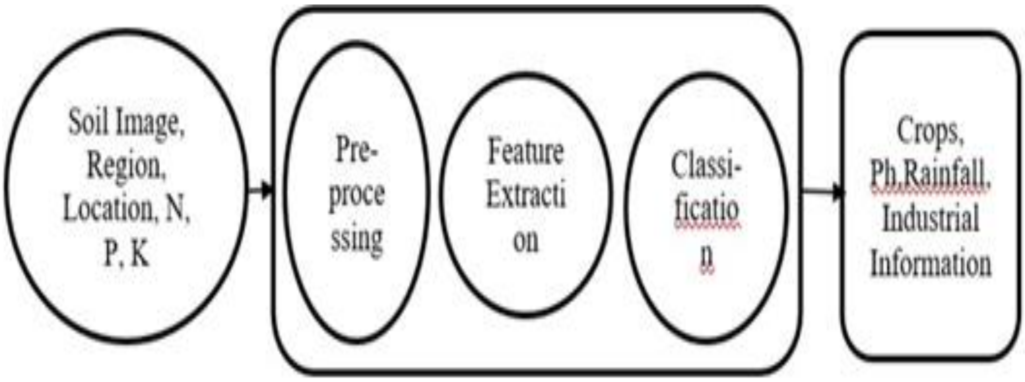


Figure 3.4: Data Flow Diagram-1

3.10 Use Case Diagrams

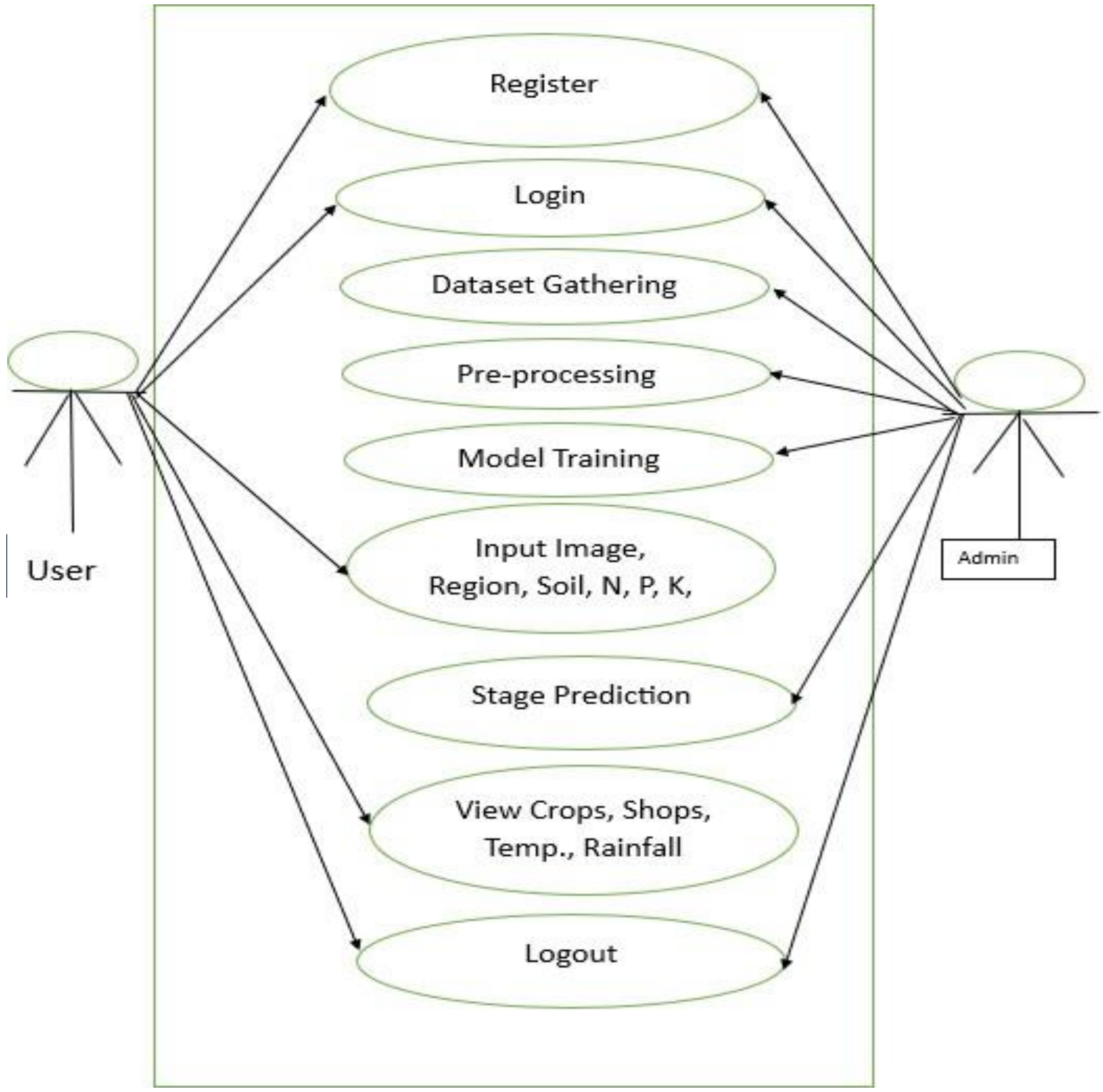


Figure 3.5: Use Case Diagram

In above use case diagram, it is consisting two modules user and an admin module. The user module has access to register the account, log in, then pass the inputs view the result, and finally log out. The admin module can register the account, log in, data gathering pre-processing, model building, and training display the results to the user end, and finally log in.

3.11 Class diagram

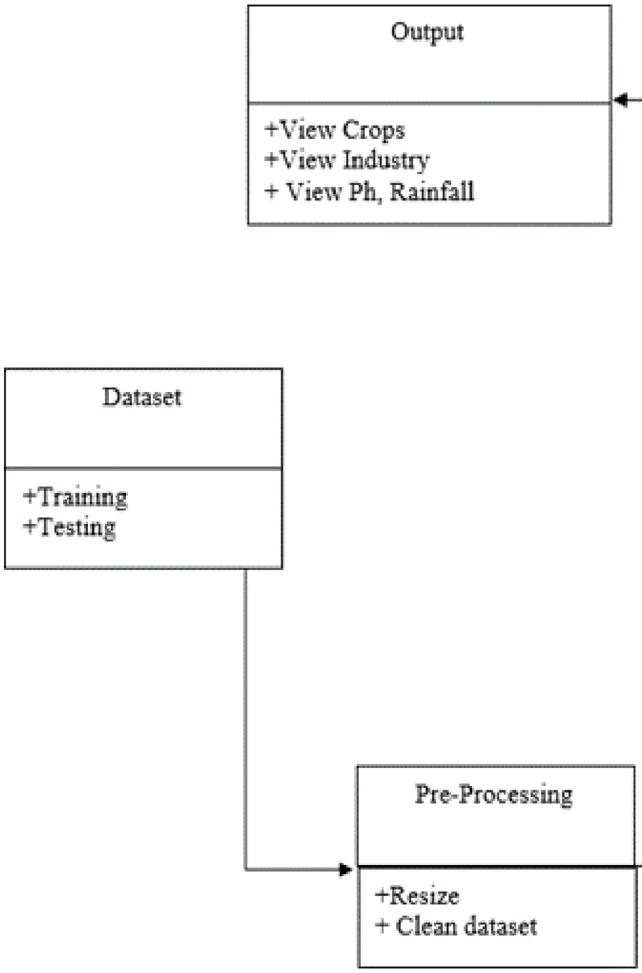


Figure 3.6: Class diagram

Our system contains mainly four classes of datasets in that we have split the dataset into two categories training and testing. The second class is pre-processing in that we have resized the image. The third class represents the models. total 3 models like KNN, CNN, and Random Forest we have used in the system. Final class which represents the output.

3.12 Activity Diagram

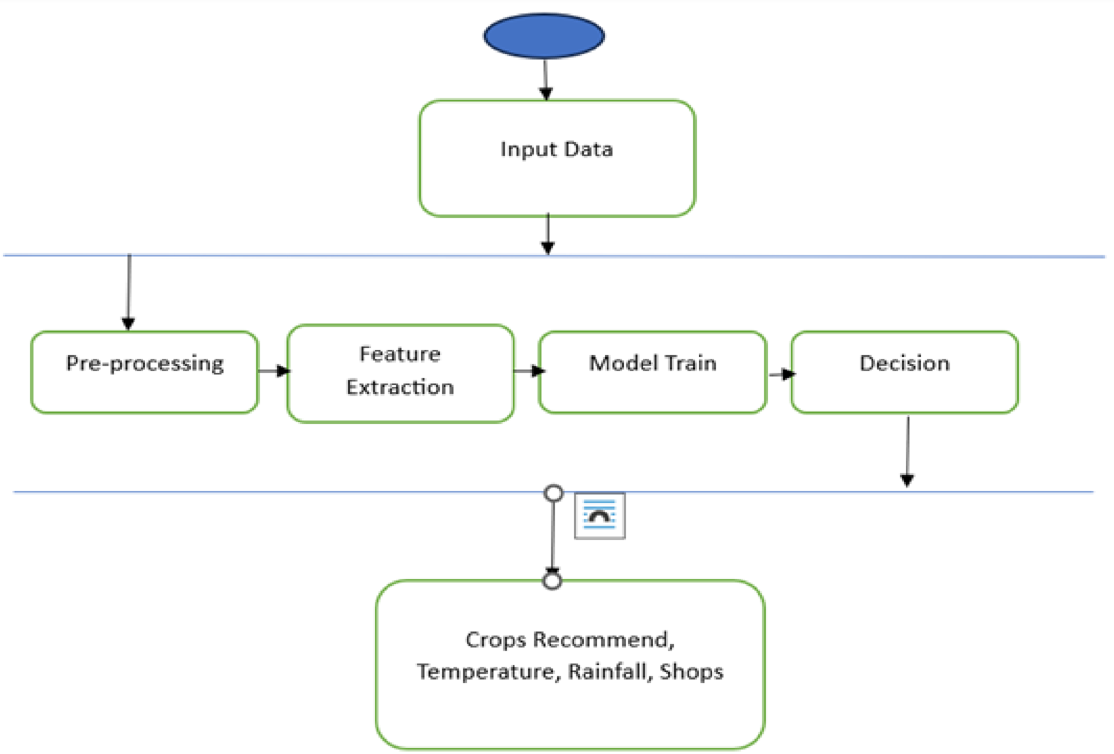


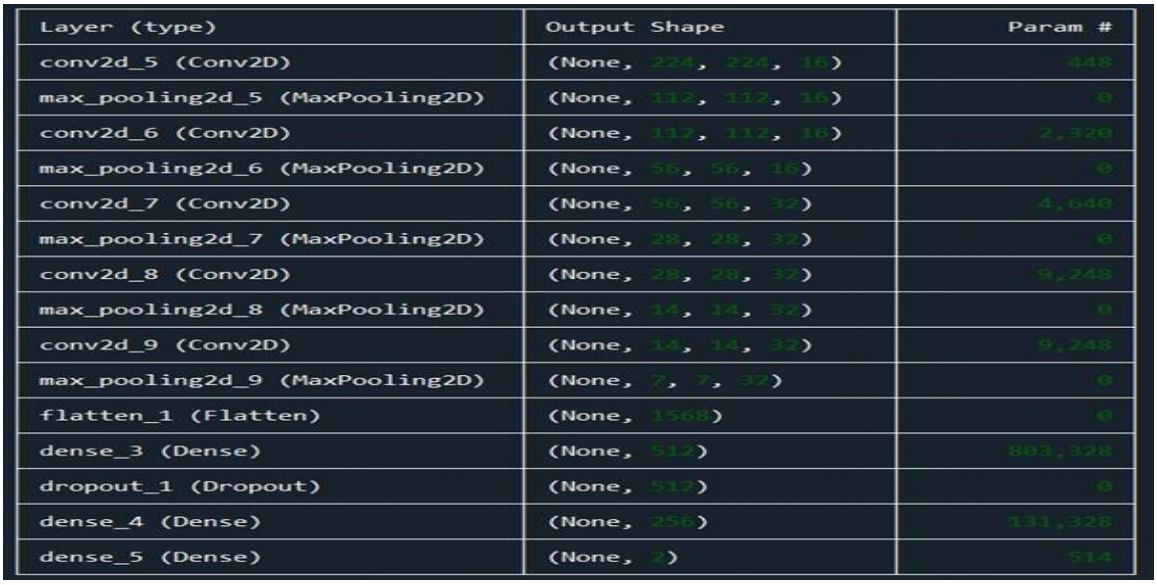
Figure 3.7: Admin Activity diagram

The activity diagram is a graphical representation of the workflow of the system. In that, we represent the flow of the system stepwise. In the above figure, the admin activity diagram input will be passed to the system. It will pass to the pre-processing stage. Then feature will be extracted from the image or check the null values from crop data and models will be trained on these data.

CHAPTER 4

SIMULATION AND TESTING

4.1 Software Testing



Layer (type)	Output Shape	Param #
conv2d_5 (Conv2D)	(None, 224, 224, 16)	448
max_pooling2d_5 (MaxPooling2D)	(None, 112, 112, 16)	0
conv2d_6 (Conv2D)	(None, 112, 112, 16)	2,320
max_pooling2d_6 (MaxPooling2D)	(None, 56, 56, 16)	0
conv2d_7 (Conv2D)	(None, 56, 56, 32)	4,640
max_pooling2d_7 (MaxPooling2D)	(None, 28, 28, 32)	0
conv2d_8 (Conv2D)	(None, 28, 28, 32)	9,248
max_pooling2d_8 (MaxPooling2D)	(None, 14, 14, 32)	0
conv2d_9 (Conv2D)	(None, 14, 14, 32)	9,248
max_pooling2d_9 (MaxPooling2D)	(None, 7, 7, 32)	0
flatten_1 (Flatten)	(None, 1568)	0
dense_3 (Dense)	(None, 512)	803,328
dropout_1 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 256)	131,328
dense_5 (Dense)	(None, 2)	514

Figure 4.1: CNN Model Summary diagram

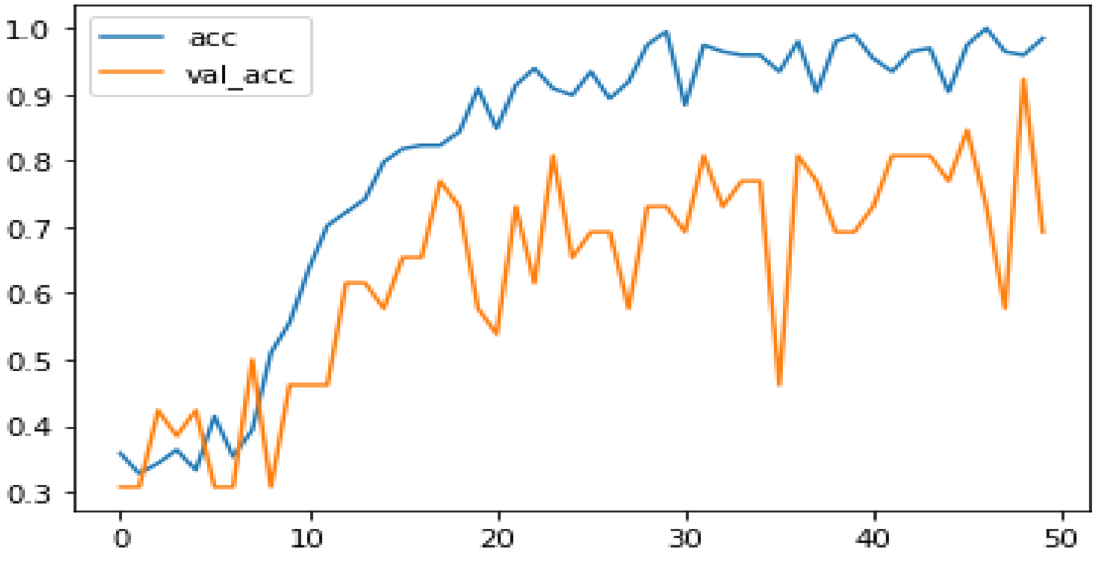


Figure 4.2: CNN Accuracy diagram: 50 Epochs

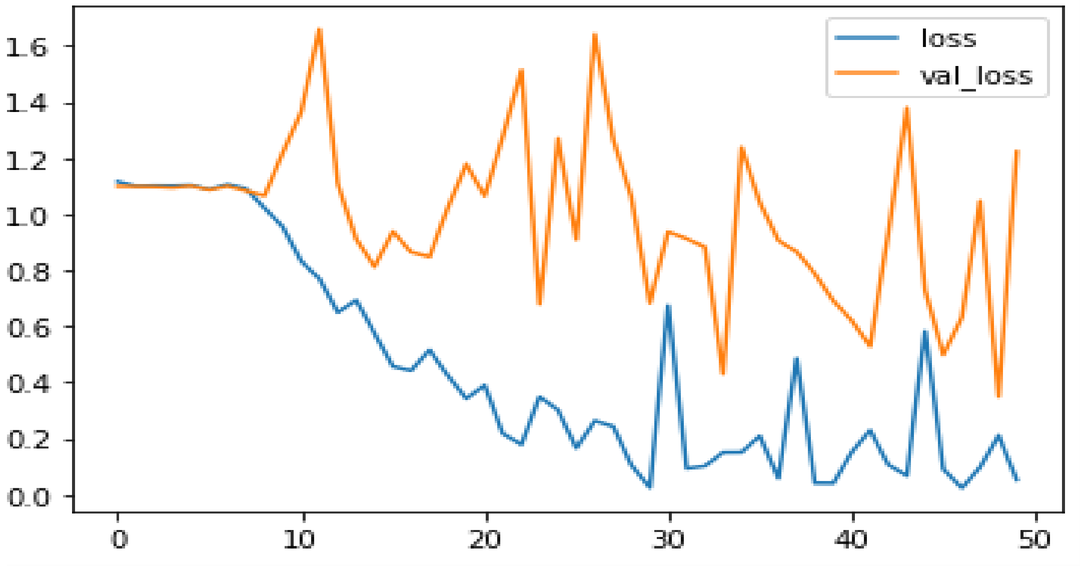


Figure 4.3: CNN Loss diagram: 50 Epochs

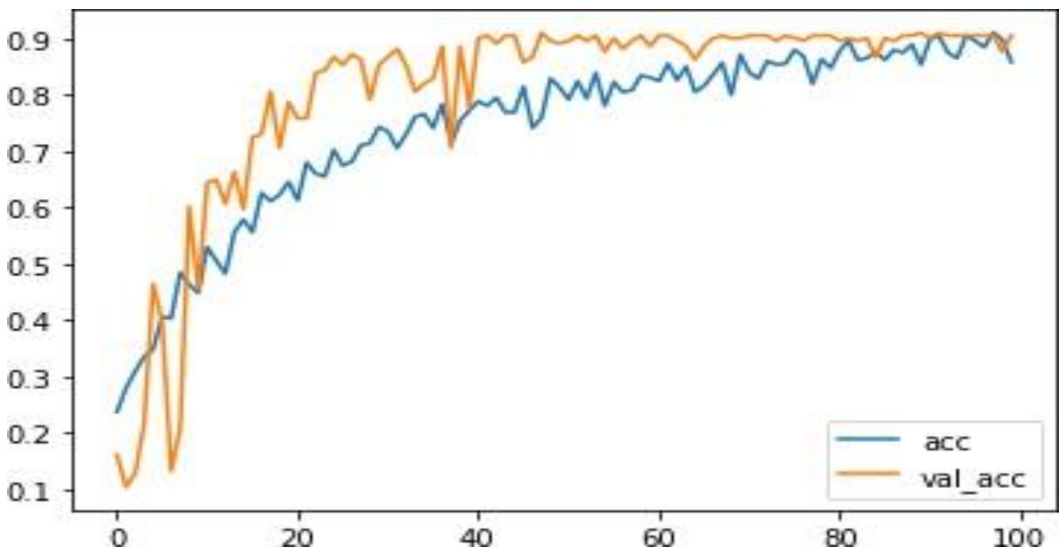


Figure 4.4: CNN Accuracy diagram: 100 Epochs

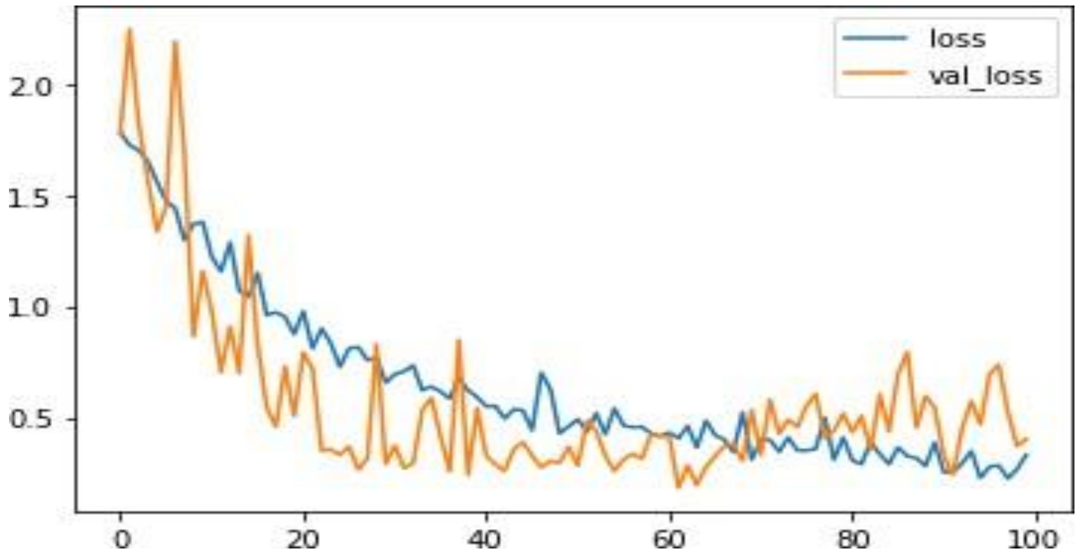


Figure 4.5: CNN Loss diagram: 100 Epochs

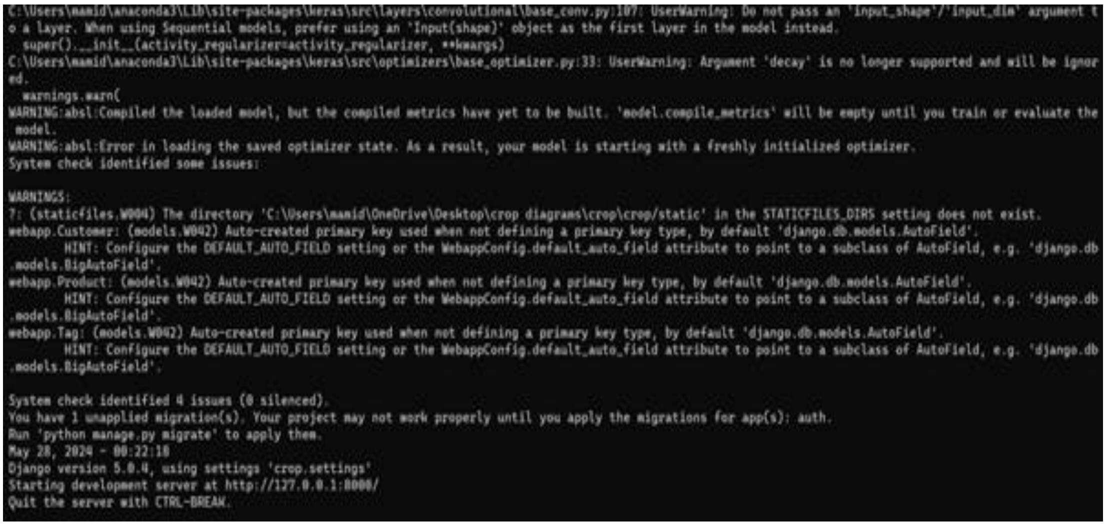


Figure 4.6: Run-Time Sever Screenshot

4.2 Test cases & Test Results

Figure 4.7: Test Case-1

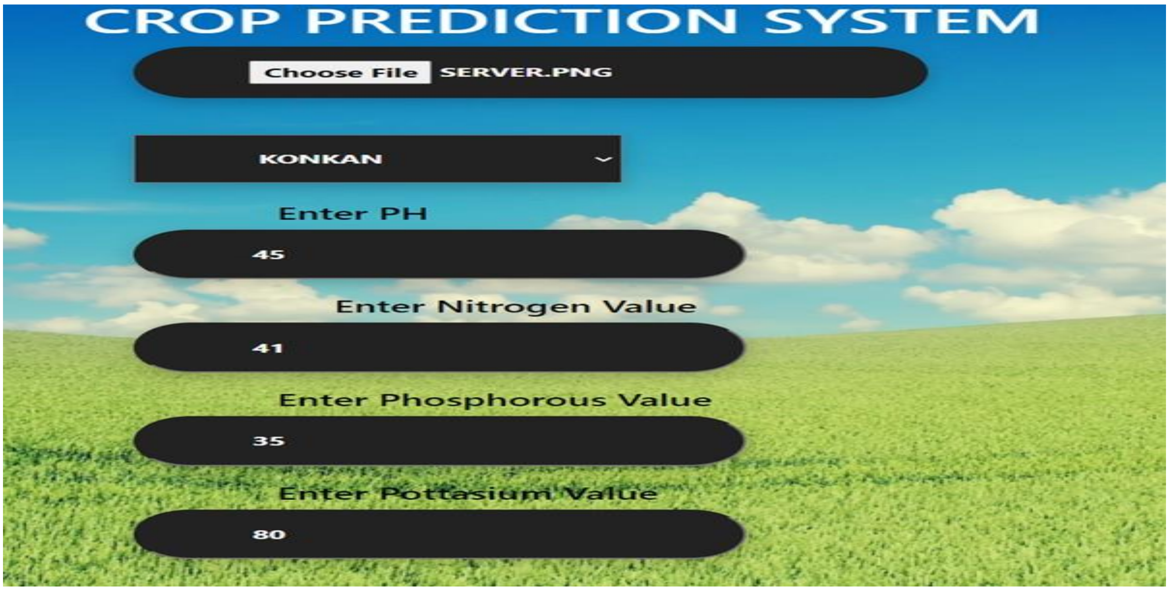


Figure 4.8: Test Case-2

CHAPTER 5

RESULT, ANALYSIS, CONCLUSION

5.1 Discussion

5.1.1 Dataset Gathering

In our project, we gathered the data from the Kaggle platform. There are a total of 440 images with different 4 categories such as alluvial, clay, black, and red. After gathering we split the dataset into two categories training and testing. The training folder contains 400 images and 40 images for testing. In our project, we require three different datasets one is for soil images, soil features such as N, P, K, Temperature, Humidity and PH, etc., and finally shop address. The two datasets we are going to gather are the Kaggle and Crops datasets we are collecting and the shop's dataset we are collecting from Google.

5.1.2 Dataset Pre-processing

In this project, we used two different datasets one is of soil images and the other is N, P, K dataset. Some data processing steps have Kaggle to be performed to prepare data and transform it into sequential integers of the same length. Following are some data preprocessing steps are performed to preprocess data before actual training. We have separate validation sets and test sets, so there is a need to divide the dataset into training sets, validation sets, and test sets. As per requirement dataset is divided into 80:20 ratio. For images in pre-processing, we resized the image into 224*224 size.

5.1.3 Data Augmentation

In data augmentation, we simply increase the size of the training dataset. It means that we rotate the image, zoom, and change the brightness of the image. Data augmentation is useful to improve performance and reduce overfitting. In this step, we perform some transformations on images such as rotating the Image to 40 degrees/angle, zooming the image with a 0.2 scale factor, and changing the Contrast of the original image. It helps to address issues like overfitting and data scarcity, and it makes the model robust with better performance.

5.1.4 Model Creation and Training

In this step, we create the model with 128 kernels with the ReLu activation function. In our system, we get 91.25% accuracy on 50 epochs and 98.23% accuracy on 100 epochs.

5.2 Algorithm Details

5.2.1 Convolutional Neural Network

In this proposed research paper Convolution Neural Network will be used for feature extraction. CNN can fetch exact features from the image data, rather than taking the features one by one. Generated weights are extracted from the different layers of CNN such as convolution layers, pooling layers, activation layers, and fully connected layers. The convolution layer is the key role of this network, which does the extraction of the features from the training image data.

- Convolution:- The principal utilization of the Convolution activity if there should be an occurrence of a CNN is to recognize fitting highlights from the picture which goes about as a contribution to the primary layer. Convolution keeps up the spatial interrelation of the pixels This is finished by the fulfilment of picture highlights utilizing miniscule squares of the picture. Convolution equation. Every picture is seen as a network of pixels, each having its worth. Pixel is the littlest unit in this picture grid. Allow us to take a 5 by 5(5*5) framework whose qualities are just in twofold (for example 0 or 1), for better agreement. It is to be noticed that pictures are by and large RGB with upsides of the pixels going from 0 - 255 i.e. 256 pixels.
- ReLu:- ReLu follows up on a rudimentary level. All in all, it is an activity that is applied per pixel and overrides every one of the non-positive upsides of every pixel in the component map by nothing.
- Pooling or sub-sampling Spatial Pooling which is likewise called sub-sampling or down-sampling helps in lessening the elements of each element map yet even at the same time, holds the most important data of the guide. After pooling is done, in the long run, our 3D element map is changed over to a one-dimensional component vector.
- Fully Connected layer:- The yield from the convolution and pooling activities gives noticeable highlights that are removed from the picture. These highlights are then

used by the Fully Connected layer for consigning the info picture into various classes predicated on the preparation dataset.

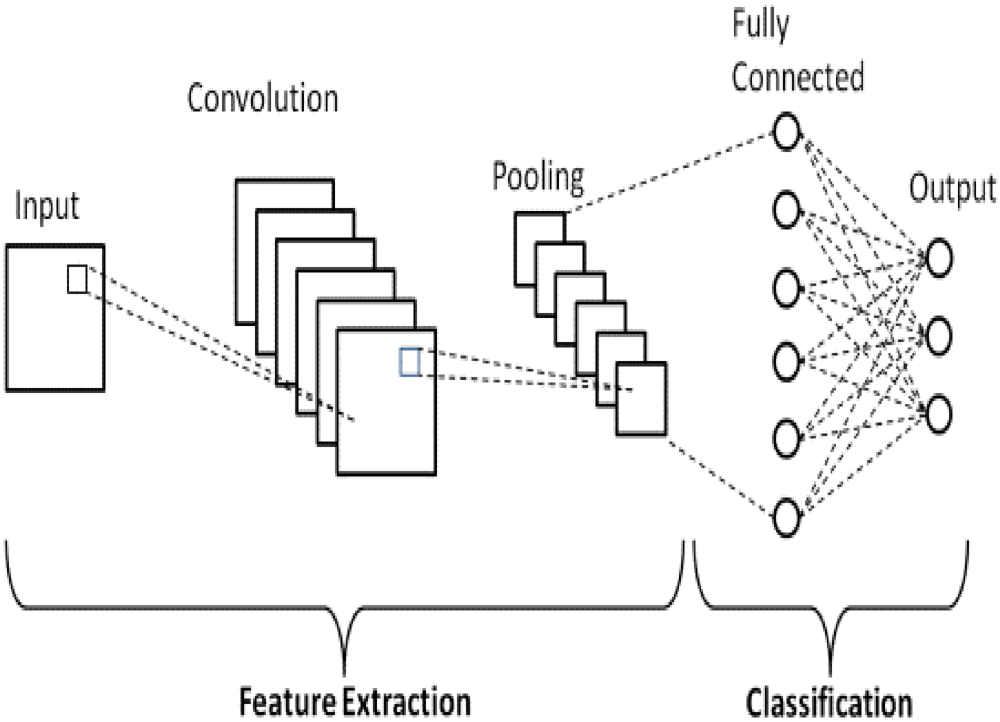


Figure 5.1: CNN diagram

5.2.2 Random Forest

Random Forest:- A conflicting woods locale is a man-made understanding strategy used to oversee coordination and fearlessness concerns. It utilizes pack understanding, a framework that consolidates a few classifiers to manage any outcomes relating to convoluted difficulties. Different choice trees are incorporated into a capricious backcountry gauge. By pulverizing or bootstrapping amounting, tree choice for the extravagant forest region area evaluation is ready. Squeezing is a get-together meta-assessment that makes arrangements with the precision of human grasping evaluations. The end is fanned out by the assessment (inconsistent woodlands) suppositions for the choice trees. It makes expectations by utilizing the typical or mean of the outcomes from various trees. The number of trees that are planted decides the result's accuracy as we showed in the figure.

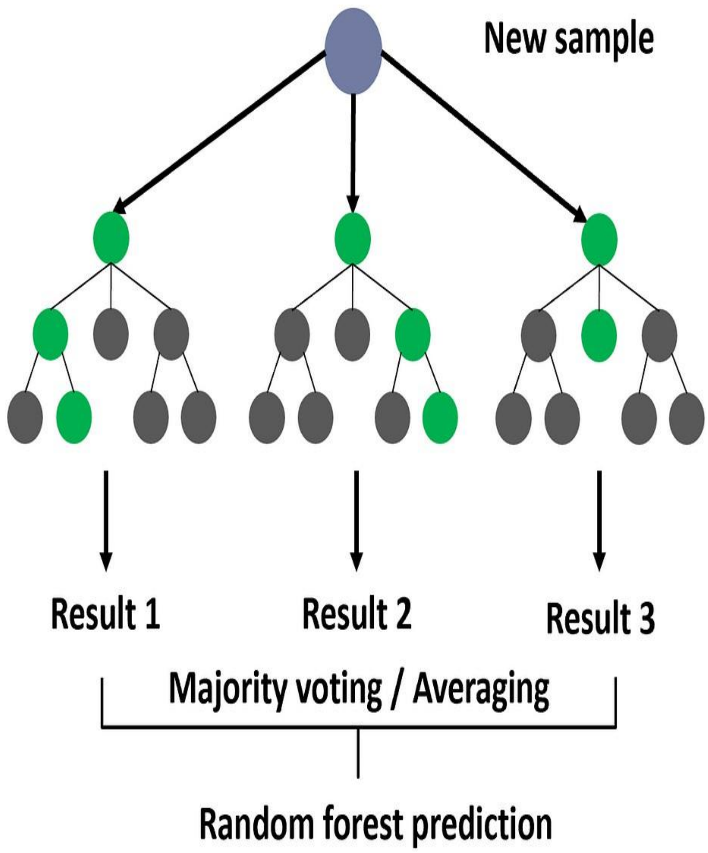


Figure 5.2: Random Forest diagram

5.2.3 K Nearest Neighbor

A refinement of the k-NN characterization calculation is to gauge the commitment of every one of the k neighbors as indicated by their distance to the inquiry point, giving a more prominent load to nearer neighbors. The KNN classifier recommends the emergency clinic subtleties for infection dependent on the closest distance.

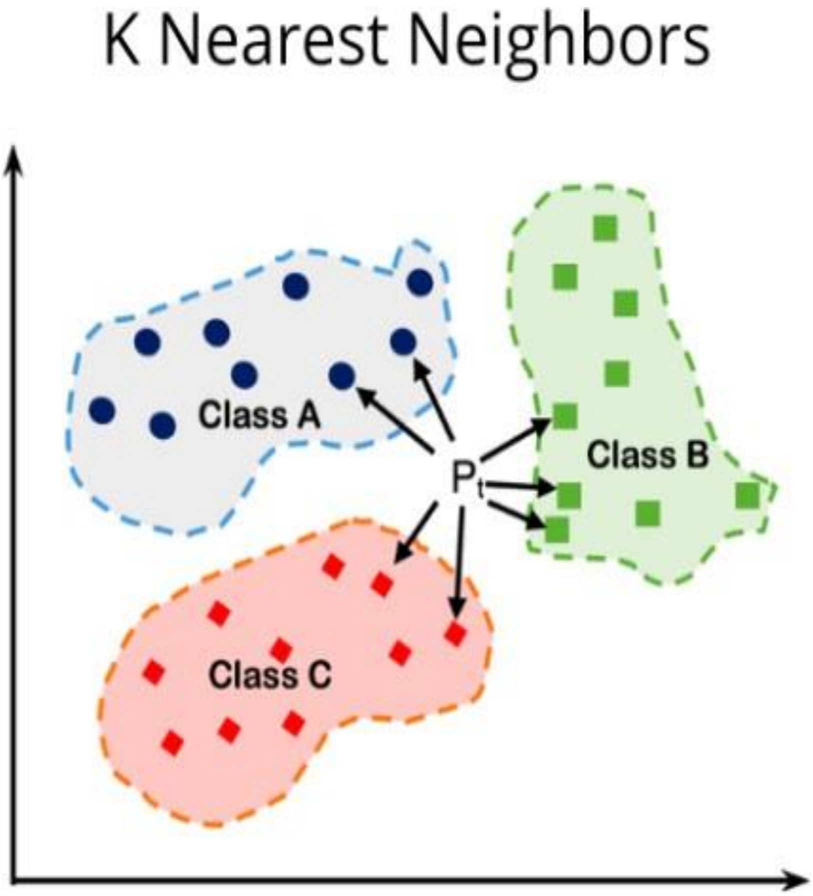


Figure 5.3: KNN diagram

d=sqrt((x2-x1)²+(y2-y1)²)
d=distance
x1, x2, y1, y2 = data points

5.3 Result Analysis

In this project we used three different datasets one is for soil, the second is for soil features such as temperature, PH, Humidity, and Rainfall and the final data is for shops. In our experimental setup, as shown in Table 1, the total number of 400 trained images for 4 categories such as Alluvial, Black, Clay, and Red soil types, and 40 new images were tested. These images go through the CNN framework by following feature extraction using our image processing module. Then our trained model of classification of soil type classifies the image into specified categories. And then this category we are passing to support vector machine again we are taking some features from farmers such as rainfall and PH of the soil and two more we are fetching from API. Based on this the random forest will suggest the crops based on regions. We get an accuracy of 98.23% at 100 epochs for the CNN model. In this project, we are recommending the shops through the KNN model. Figure 5.4 shows the registration form. In that, the user will create an account on the website. Figure 5.5 shows the login page. After creating the account user will log in to the website. In Figures 5.6 and 5.7 the user will upload the image and also provide the PH, nitrogen, phosphorous, region, and location. In fig.5.8 and 5.9, the user gets predicted crops, rainfall, temperature, humidity, and shop address.

Sr. No.	**Category**	**Samples**	**Training**	**Testing**	**Algorithm**	**Accuracy**
1	Soil Images	817	656	161	CNN	98.23%
2	Crops	91	73	18	RF	81.25%
3	Shops	44	19	5	KNN	99.0%

Table 5.1: Dataset Information & Accuracy

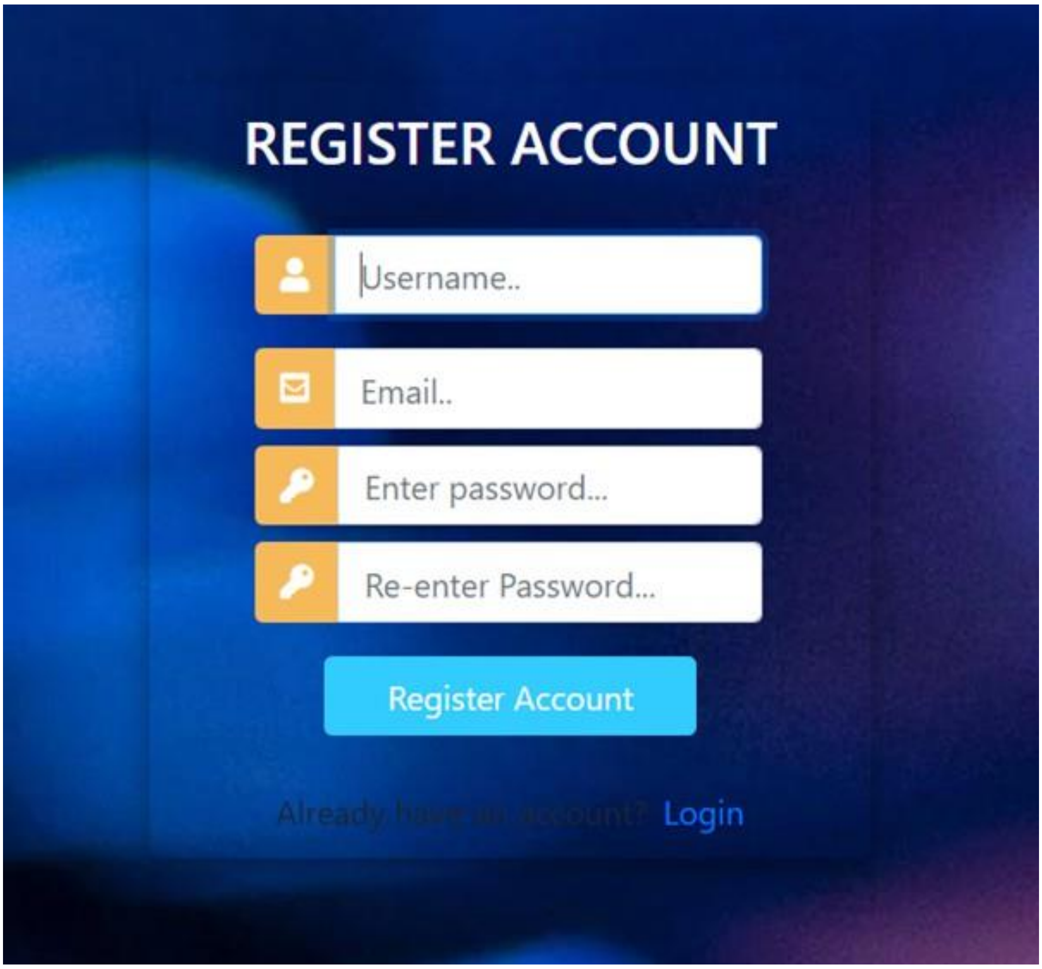


Figure 5.4: Registration Form

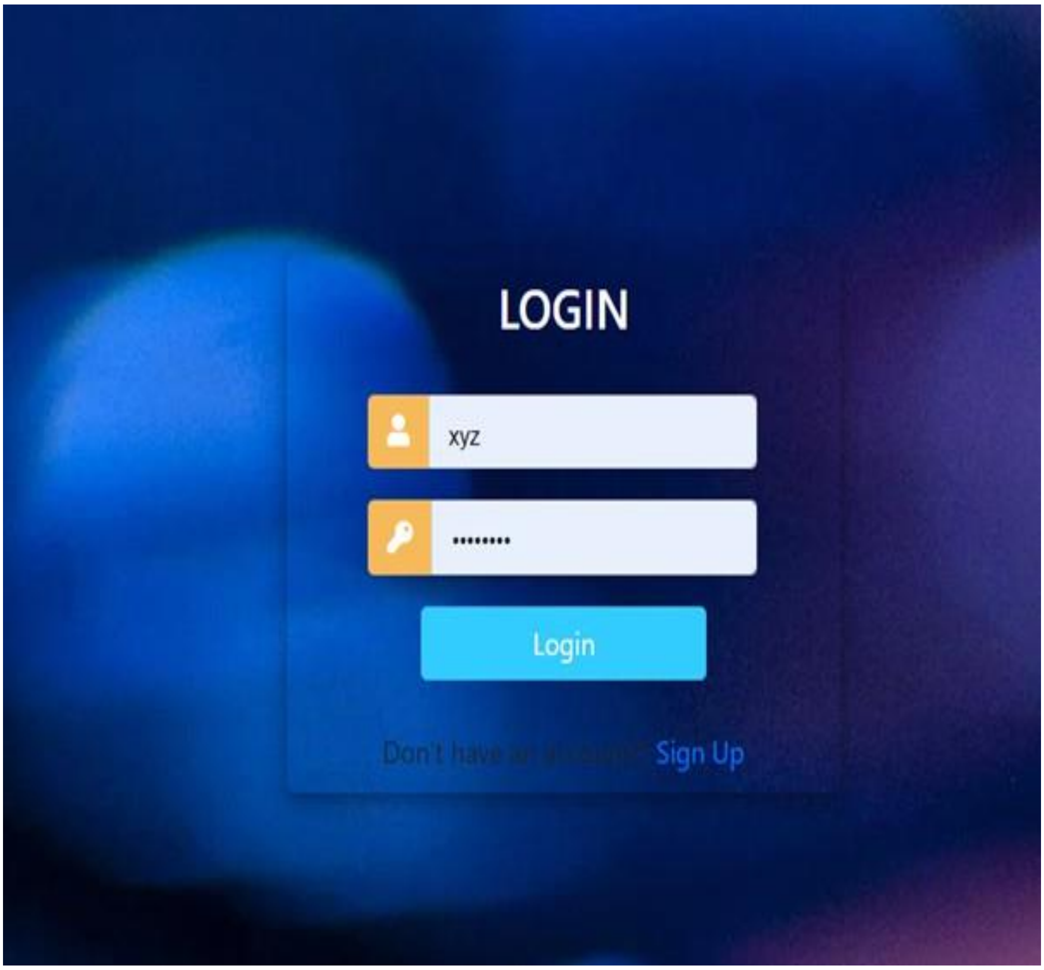


Figure 5.5: Login Form

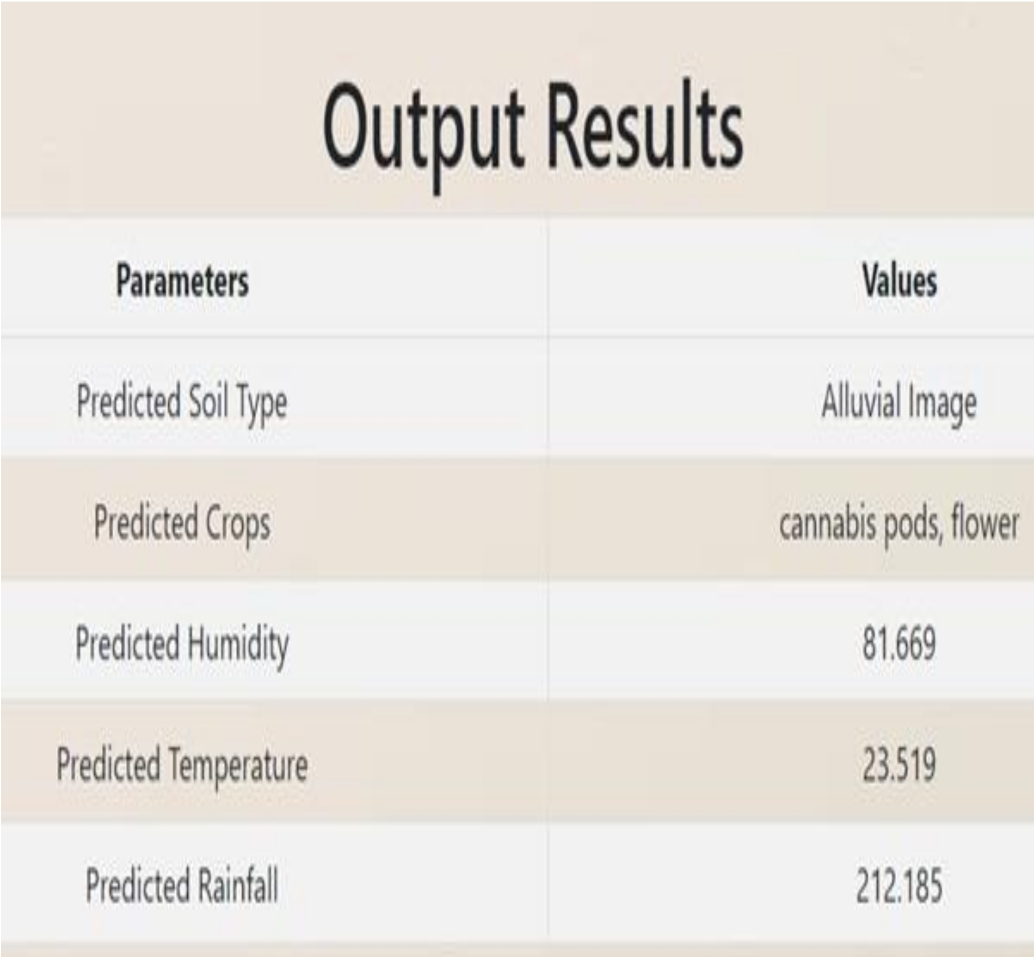
Figure 5.6: Home Page: Crop Prediction

Figure 5.7: Home Page: Industry

Input Values

Parameters	Values
Region	Konkan
Season	Kharip
PH	45
Nitrogen	40.0
Phosphorous	35.0
Pottasium	70.0

Figure 5.8: Result Page: Crops Predicted-1



Output Results

Parameters	Values
Predicted Soil Type	Alluvial Image
Predicted Crops	cannabis pods, flower
Predicted Humidity	81.669
Predicted Temperature	23.519
Predicted Rainfall	212.185

Figure 5.9: Result Page: Crops Predicted-2

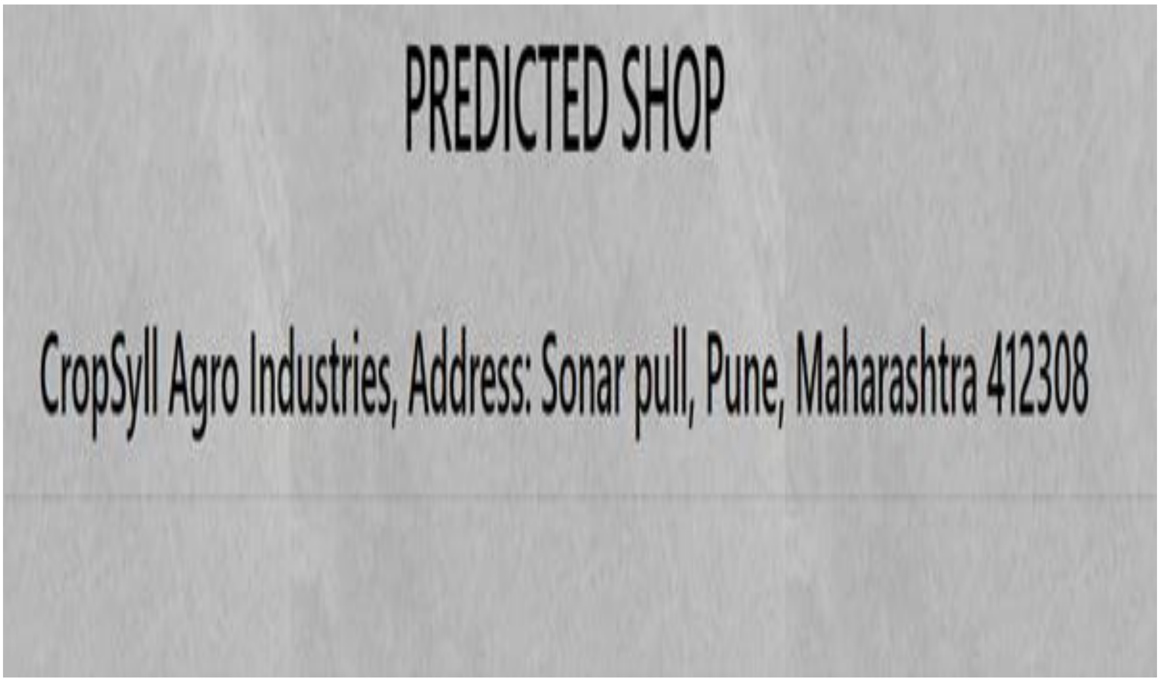


Figure 5.10: Result Page: Industry

5.4 Conclusions

In this research, a wide range of strategies needs to be investigated in Machine Learning and deep learning tailored for soil classification and recommendation of crops. In this paper, we propose a CNN model developed from ConvNet. We are using a CNN model for soil image classification, which has four classes (Alluvial, Clay, Red, Black) of image classification problems. Based on every class we are suggesting the crops, Ph, Rainfall, and Temperature by using random forests. Also, we are going to recommend industrial information to farmers for selling crops. With this system, we will provide a user-friendly application that covers aspects like the name of that crop recommendation. Using the challenging database in which the soil images are taken. At the front end, i.e. the User Interface, the input will be the soil image by the user. The output will be in the form of crop recommendations. In this project, we achieve 98.23% accuracy using a convolutional neural network. In this project we achieve 81.25% accuracy by using random forest and 99% by the known model.

5.5 Future Scope

5.5.1 Enhanced Accuracy

Nonstop enhancements in AI calculations, combined with the accessibility of huge datasets, can prompt more exact yield expectation models. Joining cutting-edge procedures, for example, profound learning and troupe techniques could additionally upgrade expectation exactness.

5.5.2 Integration of Multiple Data Sources

Integrating different information sources, for example, satellite symbolism, weather conditions conjectures, soil quality information, authentic harvest yield information, and market patterns can enhance crop expectation models. The future examination could zero in on creating methods to coordinate and use these divergent information sources.

5.5.3 Real-Time Prediction

Real-time crop prediction systems can provide valuable insights to farmers, enabling them to make timely decisions regarding crop management practices, irrigation, fertilization, and pest control. Future advancements may focus on developing fast and efficient algorithms capable of delivering real-time predictions.

5.5.4 Localized Predictions

Fitting yield forecast models to explicit geographic locales or miniature environments can work on their exactness and pertinence. Future examinations could investigate strategies for building confined expectation models that record for area explicit factors like soil type, territory, and cultivating rehearses.

5.5.5 Crop Recommendation Systems

In addition to predicting crop yields, machine learning can be leveraged to recommend optimal crop varieties and cultivation practices based on environmental conditions and market demand. Future developments may include the integration of crop prediction systems with crop prediction models to provide comprehensive decision support to farmers.

5.5.6 Risk Assessment and Management

Machine learning models can help assess and mitigate risks associated with crop products, such as droughts, pests, diseases, and market fluctuations. Future research could focus on developing risk assessment frameworks that leverage historical data and predictive analytics to identify potential risks and recommend risk management strategies.

5.5.7 Sustainability and Climate Resilience

Machine learning can play a crucial role in promoting sustainable agricultural practices and enhancing climate resilience. Future research could explore how crop prediction models can be used to optimize resource utilization, minimize environmental impact, and adapt to changing climate conditions.

REFERENCES

[1] Madhuri Shripathi Rao1, Arushi Singh1, N.V. Subba Reddy1 and Dinesh U Acharya "Crop prediction using machine learning" Journal of Physics: Conference Series AICE-CS 2021.

[2] Sonal Agarwal1 and Sandhya Tarar2 "A Hybrid Approach For Crop Yield Prediction Using Machine Learning And Deep Learning Algorithms" Journal of Physics: Conference Series CONSILIO 2020.

[3] Pavan Patil1, Virendra Panpatil2, Prof. Shrikant Kokate3 "Crop Prediction System using Machine Learning Algorithms". International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 02 | Feb 2020.

[4] Surili Agarwal, Neha Bhangale, Kameya Dhanure, Shreeya Gavhane, V.A.Chakkarwar, Dr. M.B.Nagori "Application of Colorimetry to determine Soil Fertility through Naive Bayes Classification Algorithm" 9th ICCCNT 2018 July 10-12, 2018, IISC, Bengaluru.

[5] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series" 2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December 2018.

[6] Dhanush Vishwakarma "Crop Prediction using Machine Learning Approaches" International Journal of Engineering Research Technology (IJERT).

[7] Rageena P.M et al. "Agriculture Crop Prediction System Based on Meteorological Information" International Journal of Control Theory and Applications ISSN: 0974-5572.

[8] Prof. D.S. Zingade ,Omkar Buchade ,Nilesh Mehta ,Shubham Ghodekar ,Chandan Mehta "Crop Prediction System using Machine Learning".

[9] Ashwani Kumar Kushwaha, Swetabhattachrya "crop yield prediction using the agro algorithm in Hadoop".

[10] Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K T "Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning".

[11] Rahul Kataria, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper "Impact of Machine Learning Techniques in Precision Agriculture".

[12] Pijush Samui, Venkata Ravibabu Mandla, Arun Krishna and Tarun Teja “Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine”.

[13] Himani Sharma, Sunil Kumar “A Survey on Decision Tree Algorithms of Classification in Data Mining”.

[14] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate “Crop Prediction System using Machine Learning Algorithms”.

[15] David B. Lobell and Marshall B. Burke(2010) On The Use Of Statistical Models To Predict Crop Yield Responses To Climate Change, Elsevier,1.

APPENDIX

Appendix A: Problem statement feasibility assessment using, satisfiability analysis and NP-Hard, NP-Complete, or P-type using modern algebra and relevant mathematical models.

Title:

The project problem statement is a feasibility assessment using NP-Hard, NP-Complete, or stability issues using modern algebra and relevant mathematical models.

1. Theory:

What is P?

- P is a set of all decision problems which can be solved in polynomial time by a deterministic.
- Since it can be solved in polynomial time, it can be verified in polynomial time.
- Therefore, P is a subset of NP.

What is N?

- "N" in "NP" refers to the fact that you are not bound by the normal way a computer works, which is step-by-step. The "N" actually stands for "non-deterministic". This means that you are dealing with an amazing kind of computer that can run things simultaneously or could somehow guess the right way to do things, or something like that.
- So, this "N" computer can solve lots more problems in "P" time - for example, it can just clone copies of itself when needed.
- So, programs that take dramatically longer as the problem gets harder (i.e. not in "P") could be solved quickly on this amazing "N" computer and so are in "NP".
- Thus "NP" means "we can solve it in polynomial time if we can break the normal rules of step-by-step computing".

What is NP?

- "NP" means "we can solve it in polynomial time if we can break the normal rules of step-by-step computing".

Project Status:
Problem: The main problem is how to recommend crops and Industry Using AI

Solution:
In the Feasibility Study stage, the assigned project is analyzed, then information about the project participants is collected, and the requirements for the system are gathered and analyzed. During the Feasibility Study stage, the project's goals, parameters, and restraints are agreed and a conceptual problem solution is prepared. In this system we can predict crops, and techniques over data after allowing the technique, the user can upload an image. If the user uploads an image and provides the location is found in the system. All this procedure for searching crops and industries is possible in real-time, so this project is NP-complete.

The project is NP-Complete.