IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Soil Classification And Crop Suggestion Using CNN Algorithm (K-NN And Gaussian Kernel)

1. Ayush Sagwadiya, 2. Jonak Saikia, 3 Dr. Saranya V S, 1 Student, 2 Student, 3 Assistant Professor
1 Department of Computing Technologies,
1SRM Institute Of Science and technology kattankulathur, Chennai, India

I. ABSTRACT

Agriculture and related industries are vital to India's economy since they provide a primary source of income for a large portion of the people. Accurate crop forecasting is critical for farmers and government agencies to make educated decisions about minimum support prices, crop storage, and trade facilitation through imports and exports. They can act quickly by estimating crop yields before harvest. Given the volume of data involved in crop prediction, the primary goal of this endeavour is to develop a model that can accurately identify different soil types and select acceptable crops.

II. INTRODUCTION

Farmers sometimes struggle to choose the correct crops based on market conditions, and climate change makes it much more difficult to predict weather patterns. Farmers are wondering which crops will produce more. However, anticipating crop yields ahead of time might provide useful insights. Farmers can employ yield prediction algorithms to achieve faster and more accurate outcomes than traditional forecasting approaches. Crop forecasting is difficult because of the multiple variables involved. Although numerous models have been created, their performance is frequently limited. The purpose of this research is to create a model that can classify different soil types and recommend suitable crops using machine learning techniques. These strategies not only assist farmers identify the best crops, but also maximise agricultural yield. Machine learning addresses agricultural concerns by enhancing productivity and crop output.

(a) Clayey sand

(b) Sandy clay

(c) Silty sand

(d) Clay

(e) Humus clay

(f) Clayey peat

(g) Peat

III. RELATED WORK

Namgiri Suresh *et. al*., 2021[1] has proposed a paper Crop Yield Prediction Using Random Forest Algorithm in which they presented the study that showed the practical use of data mining techniques in predicting crop yield based on climate input parameters. D.Jayanarayana Reddy *et. al.*,2021[2] has proposed a paper Crop Yield Prediction using Machine Learning Algorithm in which they have investigated multiple ML methods applied to crop yield estimation. Fariha Shahrin *et. al.*2020[3] has proposed a paper Agricultural Analysisand Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning in which they talked about crop growth and yield prediction with agricultural mapping and monitoring using K-means and Mask R-CNN algorithms in Python and Matlab.

Fatin Farhan Haque *et. al.,* 2020[4] has proposed a paper Crop Yield Analysis Using Machine Learning Algorithms in which they have discussed two machine learning algorithmsto predict crop yield, Support Vector Regression (SVR) and Linear Regression (LR). Prof. A.

V. Deorankar *et. al.*,2020[5] has proposed a paper An Analytical Approach for Soil and Land Classification System using Image Processing which proposed an analytical study of various advanced and efficient classification mechanisms and techniques and stated the proper utilisation of the number of features of remotely sensed data and selecting the best suitable classifier for improving the accuracy

of the classification. Also discussed demerits and meritsof existing methods. Ramesh Medar *et. al.*, 2019[6] has proposed a paper Crop Yield Predictionusing Machine Learning Techniques in which to predict the crop yield rate a java application is created using two ML algorithms, first is Naive Bayes method and second is K-Nearest neighbour method. Yogesh Gandge *et. al.*, 2017[7] has proposed a paper 'A Study on VariousData Mining Techniques for Crop Yield Prediction' covers research on several data mining strategies for forecasting crop yield and suggests that accuracy can be improved by implementing a large dataset.

Aditya Dhanraj Ramekar *et al.*, 2022[8] has proposed a paper 'Crop Prediction Using CNN Algorithm' in which the research work helps utilising machine learning, one of the mostcutting-edge technologies in crop prediction, work assists the beginning farmer in assisting them in guiding them for sowing the acceptable crops. Giorgio Morales *et al.*, 2022[9] has proposed a paper "Improved Yield Prediction of Winter Wheat Using a Novel Two- Dimensional Deep Regression Neural Network Trained via Remote Sensing" evaluate the performance of proposed Hyper3DNetReg network in comparison to other machine learning- based crop yield prediction techniques. Saeed Khaki *et al.*, 2020[10], "A CNN-RNN Framework for Crop Yield Prediction" proposed a CNN-RNN model along with other well- liked techniques like random forest (RF), deep fully connected neural networks (DFNN), and LASSO to forecast corn and soybean yield across the entire Corn Belt in the United States forthe years 2016, 2017, and 2018 using historical data. The new model produced average yieldswith root-mean-square errors (RMSE) of 9% and 8%, respectively.

IV. PROPOSED WORK

Data Collection

Data collection is defined as the systematic gathering, measurement, and assessment of information from multiple sources in order to draw relevant conclusions. This strategy might draw on a variety of resources, including social media analytics, online polls, customer reviews, and more. The dataset used in this study consists of five separate soil classifications, with 80-20% divided into training and test sets. Each training category comprises over 1,000 photos, but each test category has around 500 images.

Pre-processing

Pre-processing is an important step in putting raw data into a structured format for future analysis. Raw datasets may contain inconsistencies, errors from manual entry, missing values, and schema conflicts. The pre-processing step solves these challenges by organising and putting the data into an efficient format for analysis.

Preparing raw data to be acceptable for a machine learning model is known as data preparation. In order to build a machine learning model, it is the first and most important stage. Real-world data typically includes noise, missing values, and may be in an undesirable format, making it impossible to build machine learning models on it directly. Data preprocessing is necessary to clean the data and prepare it for a machine learning model, which also improves the model's accuracy and effectiveness. While dealing with image dataset, Image input can be resized to fit the dimensions of an image input layer. Additionally, preprocessing data can be used to improve desired characteristics or minimize artefacts that might bias the network. You could, for instance, normalize or eliminate noise from the supplied data. Some of the pre-processing technique includes

- Grayscale conversion

Grayscale is just the process of turning colored images into black and white. It is typically utilized to make machine learning methods less computationally complex. Grayscale is a smart choice because it minimizes the amount of pixels in an image, which lowers the number of computations needed, and most photographs don't need color to be recognized.

- Normalization

Projecting image data pixels to a preset range—typically (0,1) or (-1,1)—is a procedure known as data re-scaling. This is frequently applied to many data types, and you should normalize them all so that you can use the same algorithms on them. Normalization is typically used to translate the pixel values of a picture into a more common or comfortable feeling.

- Data Augmentation

Making minor changes to already-existing data in order to broaden its diversity

without gathering new data is known as data augmentation. It is a method for

increasing a dataset. Standard data augmentation methods include flipping data horizontally and vertically, rotating data, cropping data, shearing data, etc. Data

augmentation can assist stop a neural network from picking up unrelated

features. As a result, the model performs better.

• Image standardization

Standardization is a technique that preprocesses and scales photographs to give them comparable heights and widths. Data is rescaled to have a mean of 0 and a standard deviation of 1 (unit variance). Data consistency and quality are enhanced through standardization.

Feature Extraction

Feature extraction entails converting raw data into a more compact and meaningful representation. The most relevant features from the data are extracted while lowering its dimensionality using several picture pre-processing techniques such as binarization, thresholding, scaling, and normalisation. These collected features are then applied to image classification and recognition tasks.

When developing a predictive model, the feature selection procedure entails decreasing the number of input variables. In some circumstances, reducing the number of input variables may reduce the cost of computation while improving the model's effectiveness of modelling. Unsupervised and supervised feature selection approaches are the two basic categories, while supervised methods can be further broken down into intrinsic, wrapper, and filter methods. Input variables that can be filtered to choose the most pertinent features are scored for correlation or dependence using statistical measures in filter-based feature selection methods. The output or response variable's data type and the input variable's data type must be taken into consideration when using statistical measures for feature selection. A feature in computer vision is a quantifiable aspect of your image that is specific to the thing you are trying to recognize. It could be a standout hue in an image or a particular shape like a line, an edge, or a section of an image. To distinguish things from one another, a useful property is used. Feature selection in soil is done in below image

Feature Extraction in Soil sample image

Convolutional Neural Network (CNN) Algorithm

The Convolutional Neural Network (CNN) is used for image classification. CNNs are a sort of neural network design that works particularly well for image recognition and pixel-level data analysis. While there are numerous deep learning models, CNNs are thought to be the best for tasks like as object identification and classification because of their capacity to learn spatial hierarchies of information automatically and adaptively.

One of the most well-known and widely utilized DL networks is the convolutional neural network (CNN). CNN is to blame for the current popularity of DL. The primary benefit of CNN over its forerunners is that it does it automatically and without human supervision, making it the most popular. The most prevalent CNN architectures, starting with the AlexNet network and concluding with the High-Resolution network, have also been explored upon in length. Laith Alzubaidi *et al*. [1] has given deep study of Deep Convolution Network and discussed about its challenges and applications.

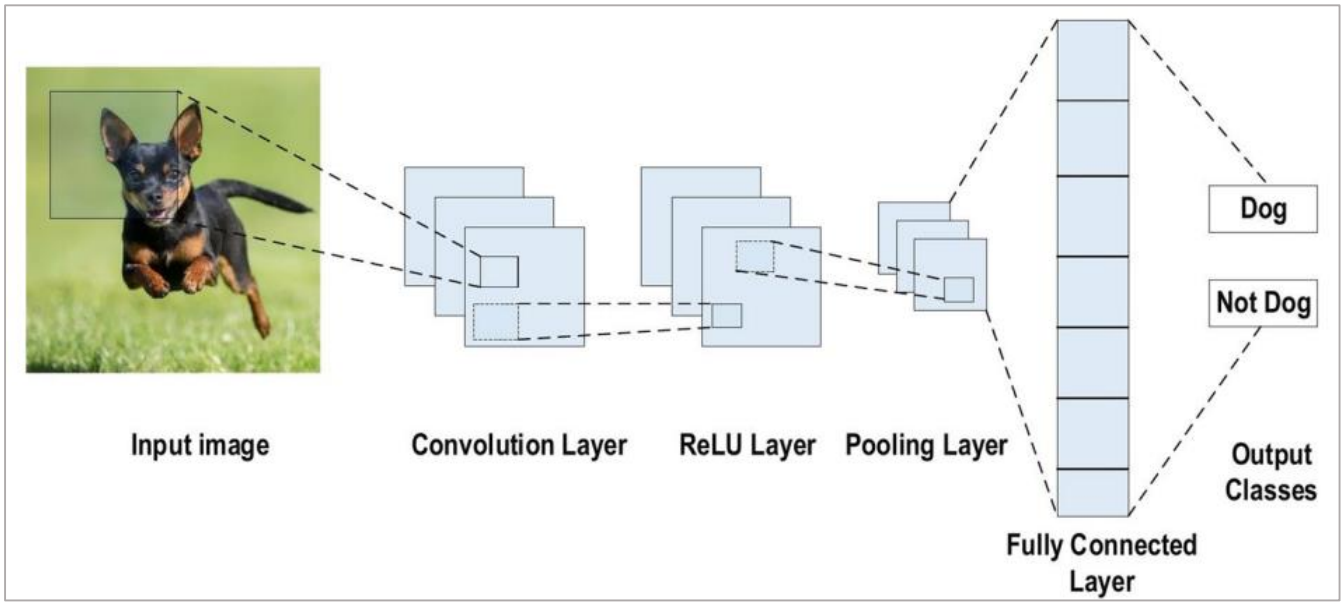
The structure of a CNN is comparable to the connection structure of the human brain. Similar to how the brain has billions of neurons, CNNs also have neurons, but they are structured differently. A CNN's neurons are set up similarly to the frontal lobe of the brain, which processes visual stimuli. This configuration guarantees that the full visual field is covered, avoiding the issue with typical neural networks' piecemeal image processing that requires images to be given to them in low-resolution chunks. A CNN performs better with image inputs and voice or audio signal inputs compared to the earlier networks.

k-Nearest Neighbors (k-NN) algorithm

The k-Nearest Neighbours (k-NN) technique is widely used in classification applications. Unlike deep learning models, k-NN is a straightforward, instance-based learning algorithm that finds the 'k' nearest data points (neighbours) to a given input and classifies it based on the majority label of these neighbours. It is very good at pattern identification and works well with tiny datasets. The technique is highly versatile and does not require an explicit training step because it uses data directly during classification, making it appropriate for applications such as object identification and categorisation. The simplicity and efficacy of k-NN make it a popular choice for a wide range of classification tasks.

Gaussian Kernel Algorithm

The Gaussian Kernel algorithm is a sophisticated approach used in machine learning to perform tasks like classification and regression. It employs the Gaussian function to determine the similarity of data points, essentially changing the input space into a higher-dimensional environment in which linear correlations can be detected. This strategy is especially beneficial for support vector machines and kernel-based learning approaches, as it enables the modelling of complex, non-linear patterns in data. The Gaussian Kernel is distinguished by its ability to emphasise local points while reducing the influence of distant points, resulting in effective decision boundaries and increased classification accuracy. The Gaussian Kernel's flexibility and ability to handle high-dimensional data make it an invaluable tool in a variety of machine learning applications.



Crop Prediction

Using this approach, the best crops for the region and their production may be chosen, hence enhancing the value and profit of farming. Crop genetic innovation has increased the efficiency of farm resource usage, reducing environmental pressures. Crop output predictions under various climatic conditions can help farmers and other stakeholders make critical decisions about agronomy and product selection. Farm production and profitability increase as crops increase and farmland is used more intensively, improving the well-being of farming households.

V. PROPOSED METHOD

Crop prediction systems offer better planning and decision-making, resulting in increased agricultural productivity. The training step of the method involves putting a dataset of soil photographs into a CNN model. The dataset may contain incomplete, redundant, or inconsistent data, hence data filtering is critical for removing such discrepancies. Furthermore, data normalisation is carried out to maintain uniformity throughout the dataset.

A basic way for extracting features from images is to use the raw pixel values as separate features. In greyscale graphics, the background threshold is usually set at (0,0,0), which represents black. In colour pictures, a pixel is considered part of an object if any of its RGB values surpass the background threshold. Following the extraction of features from the photos, feature selection is performed to Determine the most relevant qualities from the dataset.

Once the pre-processing is finished, the dataset is divided into 80% training and 20% testing. The CNN model is then trained to classify soil types using the processed data. Finally, the model is utilised to predict the most appropriate crops for a specific soil type, hence optimising crop selection and increasing agricultural productivity.

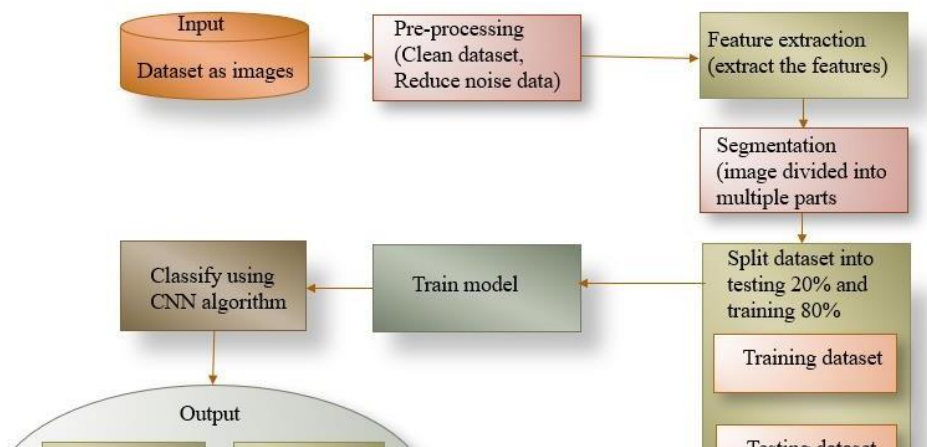


Figure 1. System Architecture

VI. CNN MODEL

Convolutional Neural Networks (CNNs) are specialised deep learning architectures that are commonly employed for tasks like image recognition and pixel data processing. Although deep learning includes numerous types of neural networks, CNNs are especially well-suited for object identification and recognition.

A CNN's core is the convolutional layer, which employs several filters (also known as kernels). These filters, which are smaller than the input image, learn their parameters during training. When each filter is applied to a picture, it creates an activation map. The pooling layers then lower the dimensionality of the feature maps by aggregating the outputs of small neurone groups from the previous layer into a single neurone in the following layer, thus simplifying the data.

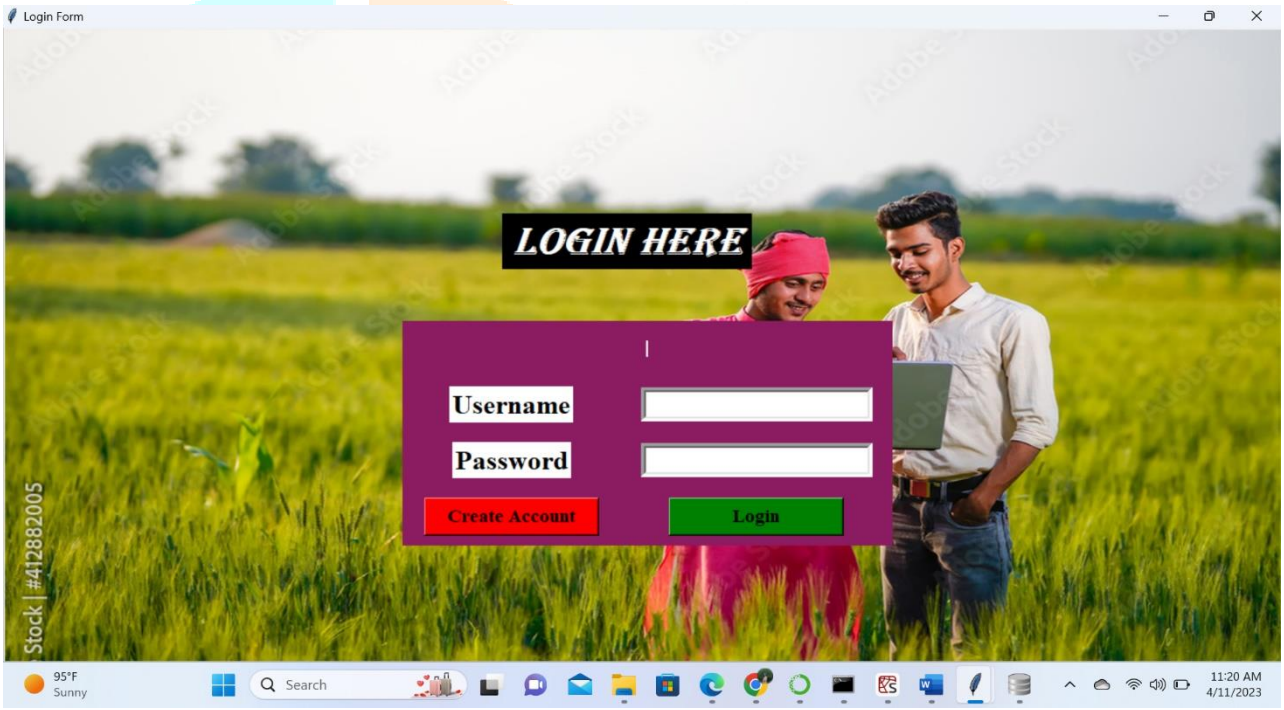
Following the convolution and pooling layers, a fully linked layer is added. This layer multiplies the input by a weight matrix and adds a bias vector, connecting all neurones in the current layer to those in the preceding layer. In a fully connected network, every neurone is linked to the neurones in the layer above it.

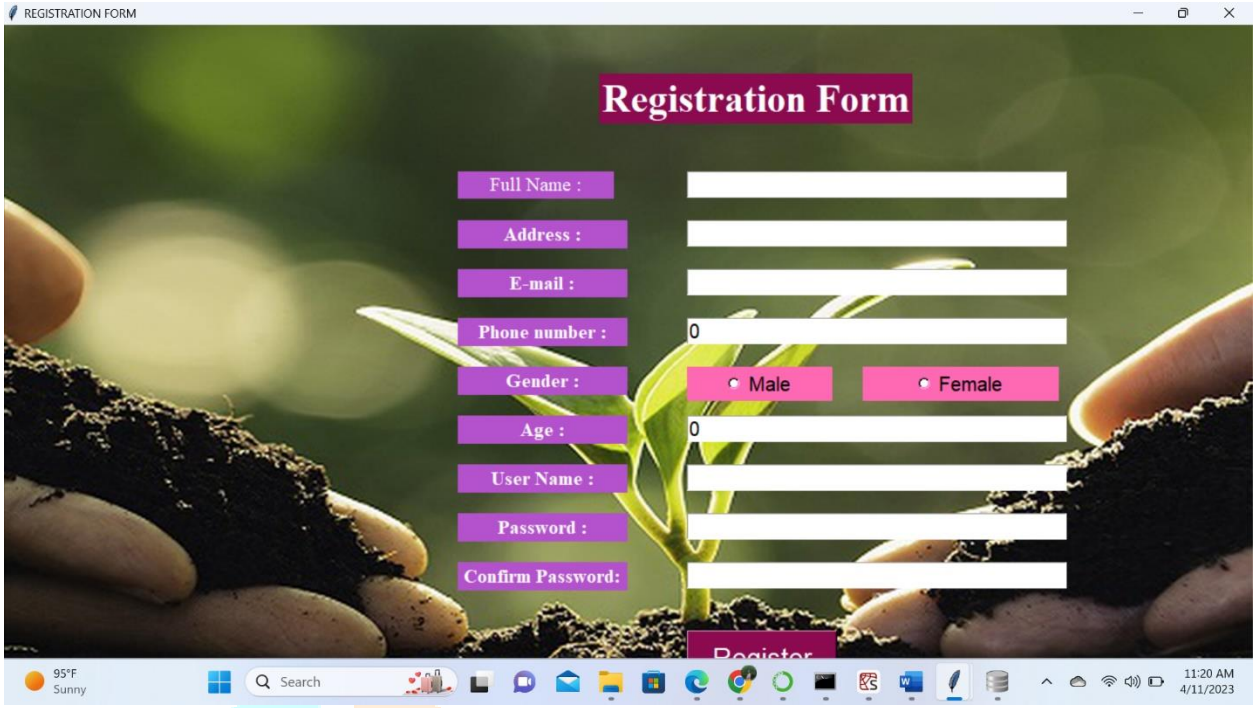
Another important component is the dropout layer, which randomises the contribution of specific neurones to the next layer during training, hence preventing overfitting. In this multi-layer neural network, input values are transmitted directly from one layer to the next, with higher-layer outputs functionally dependent on lower-layer inputs.

This suggested CNN model consists of six convolutional layers with the ReLU activation function, followed by a flatten layer and a fully connected layer. This network consists of a dense layer, a dropout layer, and a dense layer that performs classification using the SoftMax activation function.

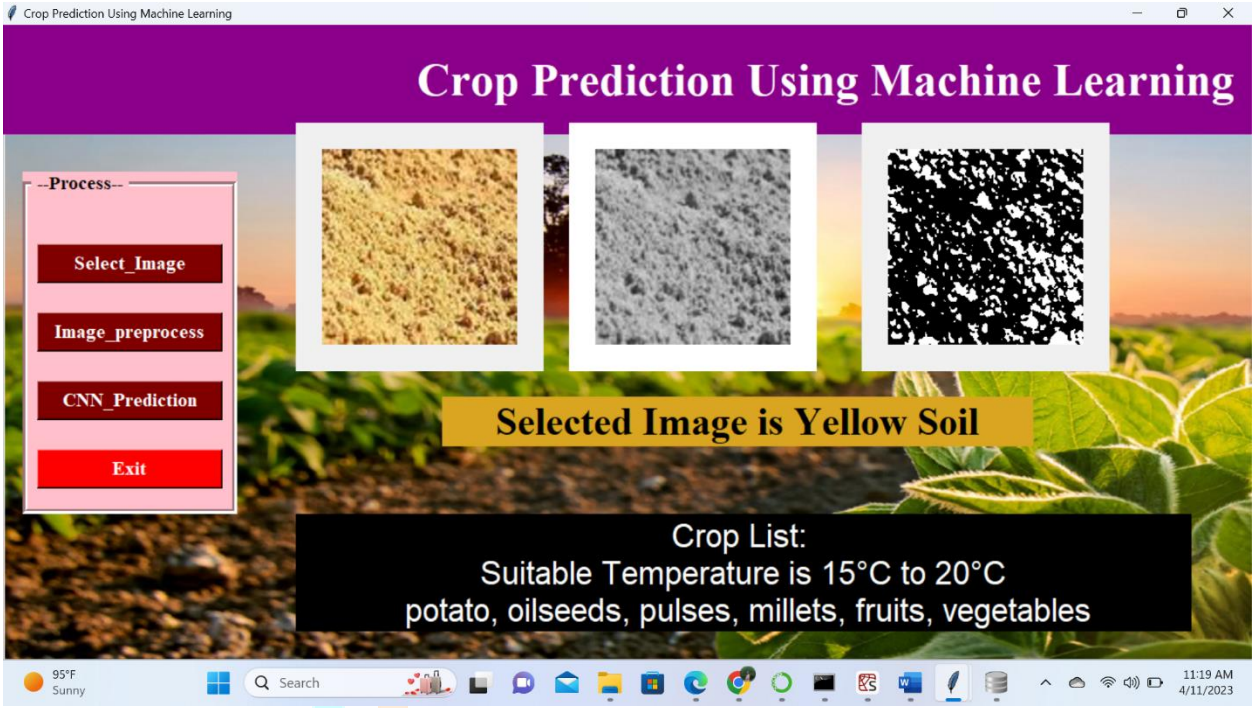
VII. OUTPUT


Crop Prediction Using Machine Learning
Crop Prediction Using Machine Learning
Crop Prediction
Login
Register
Show hidden icons
95°F
Sunny
Search
11:15 AM
4/11/2023


Login Form
LOGIN HERE
Username
Password
Create Account
Login
Stock | #412882005
95°F
Sunny
Search
11:20 AM
4/11/2023


REGISTRATION FORM
Registration Form
Full Name :
Address :
E-mail :
Phone number :
0
Gender :
Male
Female
Age :
0
User Name :
Password :
Confirm Password:
95°F
Sunny
Search
11:20 AM
4/11/2023


Crop Prediction Using Machine Learning
Crop Prediction Using Machine Learning
--Process--
Select_Image
Image_preprocess
CNN_Prediction
Exit
95°F
Sunny
Search
11:19 AM
4/11/2023



VIII. RESULTS

Once collected, the data is divided into an 80% training set and a 20% test set for use in the Spider IDE. After the data has been accurately partitioned, soil photographs are entered into the algorithm, as shown in Figure 2, to forecast crop production. This model outperformed previous models, with an astonishing 95% accuracy. The curve depicted in Figure 3 depicts the gradual rise in accuracy with each epoch.

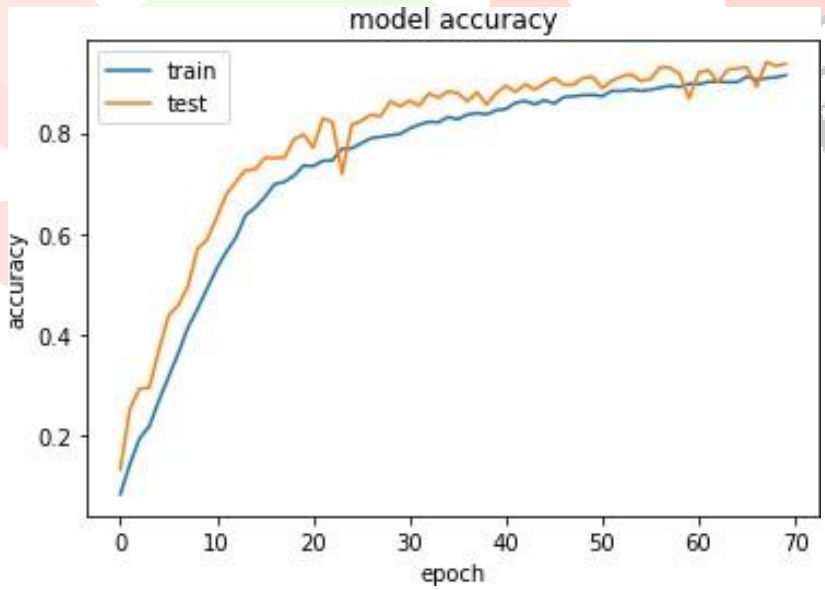


Figure 3. Accuracy graph

IX. CONCLUSION

This study describes a crop prediction system that detects soil types and recommends appropriate crops based on optimal temperature conditions. The suggested approach is based on Convolutional Neural Networks (CNNs), which provide more accuracy than existing methods. The soil picture collection is divided into five categories, with each training class containing over 3,000 photos and each test class containing more than 500. The model attained an impressive accuracy of 95%, significantly higher than previous models.

X. REFERENCES

1. Namgiri Suresh *et. al*., 'Crop Yield Prediction Using Random Forest Algorithm' 2021(Available at: https://ieeexplore.ieee.org/document/9441871)
2. D.Jayanarayana Reddy *et. al.,* 'Crop Yield Prediction using Machine Learning Algorithm', 2021 (Available at: https://ieeexplore.ieee.org/document/9432236)
3. Fariha Shahrin *et. al.,* 'Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning', 2020 (Available at: https://ieeexplore.ieee.org/document/9393066)
4. Fatin Farhan Haque *et. al.,* 'Crop Yield Analysis Using Machine Learning Algorithms'2020 (Available at: https://ieeexplore.ieee.org/document/9221459)
5. Prof. A. V. Deorankar *et. al.,* 'An Analytical Approach for Soil and Land ClassificationSystem using Image Processing', 2020 (Available at: https://ieeexplore.ieee.org/document/9137952)
6. Ramesh Medar *et. al.,* 'Crop Yield Prediction using Machine Learning Techniques', 2019 (Available at: https://ieeexplore.ieee.org/document/9033611)
7. Yogesh Gandge *et. al.,* 'A Study on Various Data Mining Techniques for Crop Yield Prediction'2017(Available at: https://ieeexplore.ieee.org/abstract/document/8284541)
8. Aditya Dhanraj Ramekar *et. al.,* 'Crop Prediction Using CNN Algorithm' (Available at: www.ijraset.com/research-paper/crop-prediction-using-cnn-algorithm)
9. Giorgio Morales *et al.,* 'Improved Yield Prediction of Winter Wheat Using a Novel Two-Dimensional Deep Regression Neural Network Trained via Remote Sensing' (Available at: https://www.mdpi.com/1424-8iee220/23/1/489)
10. Saeed Khaki *et al*., 'A CNN-RNN Framework for Crop Yield Prediction' (Available at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01750/full)