RESEARCH ARTICLE OPEN ACCESS

Kisan - Multiservice app for Indian farmers

Shrushti Govindwar
B.E. Computer Engineering
Vivekanand Education Society's Institute of Technology
Mumbai, India

Sonali Bodhwani
B.E. Computer Engineering
Vivekanand Education Society's Institute of Technology
Mumbai, India

Dr. Prashant Kanade
B.E. Computer Engineering
Vivekanand Education Society's Institute of Technology
Mumbai, India

Kanchan Nathani
B.E. Computer Engineering
Vivekanand Education Society's Institute of Technology
Mumbai, India

Shruti Mehta
B.E. Computer Engineering
Vivekanand Education Society's Institute of Technology
Mumbai, India

ABSTRACT
India is one of the most populous nations in the world, and more than 58% of its people work in agriculture. The goal is to provide a tool that farmers may use to ease some of the basic difficulties associated with farming. Farmers regularly grow the same crops, and instead of using the proper quality and quantity of pesticides and fertilizers, they use random pesticides and fertilizers, which reduce crop production and cause soil pollution and acidification. A web application utilizing machine learning methods is created to address these problems and it will benefit farmers in a variety of ways. This web application will suggest the best crop for a certain plot of land based on the content, weather conditions, and time of year. Early diagnosis of plant diseases is the answer to halting the losses in agricultural product productivity and quantity. Examining distinctly observable patterns on plant leaves is part of the research on plant diseases. In order to detect plant illnesses in a sustainable way, image processing is used rather than directly examining the plants themselves.The system will focus on finding solutions to issues, significant yield and quality constraints like plant diseases that result in enormous economic losses, farmers' ignorance of crop yield maximization and crop rotation techniques, etc.
Keywords—Agriculture, Crop prediction, Plant disease detection, Soil, Image processing, Sustainable, News Feed, Expert Assistance, Random forest classifier, Keras, Rest APIs.

Date of Submission: 18-03-2023 Date of acceptance: 03-04-2023

I. INTRODUCTION

In India, agriculture accounts for about half of all occupations. The agricultural sector greatly influences the total development and growth of our nation's economy. Therefore, using modern technology and software to implement a new-age approach in some aspects of agriculture will prove to be quite advantageous for both farmers and consumers. Before crop prediction algorithms were in use, farmers would complete the same work based on their prior knowledge and intuition for a specific place. The yield could be harmed by improper crop rotations and the haphazard application of specific soil nutrients. The system is created with all of these issues in mind in order to provide a solution and meet some agricultural needs. The system will recommend the best crop for a specific land based on various parameters such as weather, rainfall, temperature, humidity, latitude, and longitude, which will be automatically retrieved from Google Weather, once the farmer activates the GPS. Farmers will be able to cultivate profitable crops while avoiding soil pollution if they use this system. The leaves are one of the most sensitive parts of the plant, where we can first detect disease symptoms. It is essential to begin monitoring crops at an early stage of their life cycle and continue until they are ready to be harvested. Initially, plants were examined and monitored to detect diseases using the manual, time-consuming method of traditional "naked eye" monitoring, which calls for extremely careful inspection. The leaves, stem, or fruits are typically where illness symptoms can be noticed. The plant's leaves are frequently taken into account for disease detection. Farmers are frequently adequately and sufficiently knowledgeable about the crops and the diseases that threaten those crops. Without needing to see a specialist, farmers may efficiently boost their output and safeguard their crops from illnesses by using our technology.

A web application is created where a variety of services will be made available to farmers, and further actions will be taken depending on the service that is chosen.Services include crop prediction [1] based on the various parameters like weather, rainfall, temperature, humidity, latitude, and longitude, which will be automatically taken from Google Weather, once the farmer turns on the GPS. The ML model will assist in predicting the most suitable crop to be grown at that specific period using a decision tree regressor. The web app offers a service where you can determine the condition that your plants have by just clicking or choosing a photo from your phone's gallery and submitting it to the database, which uses ML to predict the disease and even suggest a cure for it, Plant Disease Recognition [2] which developed a model using image processing and after receiving an image input of the diseased leaf, it will assist the user in determining the type of illness.

II. RELATED WORK

In their approach, M. Kalimuthu et al. [1] suggested employing the machine learning algorithms to forecast crops that are inexpensive based on the input data. They employed both hardware and software, and to train the data with given input and output data, they used a Naive Bayes Gaussian classifier. With more time for data collection, DHT11 (a digital temperature and humidity sensor) was used to accurately measure the temperature and humidity of the local air. The soil's moisture content was determined using the soil moisture sensor. The crop prediction was done by taking into account the data that was collected by the DHT11 sensor and carrying out the performance evaluation of the naive Bayes classifier method. Parameters like temperature, humidity, wetness, and soil pH were manually entered. Data gathering, feature extraction, classifier training, and classification were the focus of S. Ramesh et al.'s [2] study.To train the random forest classifier to distinguish between photos of diseased and healthy leaves, datasets of diseased and healthy leaves were employed. The features of an image were extracted using the histogram of an oriented gradient (HOG). Huge datasets were trained using machine learning techniques to identify plant diseases. The idea was to use machine learning to identify the plant's sickness from its morphology. To avoid occlusion, the image is initially taken with a plain background. A random forest classifier is then employed, and 160 images of papaya leaves are used to train the model. The procedures to be taken were to preprocess the images, extract the image's features using HOG, add them to the training set, train the model using Random Forest method after this classification has been completed, and then extract the results. The system's overall accuracy was 70%. Climate-smart agriculture was the main topic of M. Gulzar et al.'s study [3]. Recent studies suggest that by 2050, food production must expand by 70 to 90 percent due to population growth and steady growth.To accomplish the following goals, CSA relies on three key goals: Increasing agricultural output will increase crop health, nutritional value, and sufficiency, ensuring food security and revenue. To reduce the risk of global warming, traditional farming practices are being replaced with resource-sustainable adaptive smart approaches. Developed nations are more aware of climate change and its mitigation options. On the other hand, emerging nations have challenges related to rapid population expansion, poor incomes, and food insecurity.India needs CSA because it is a developing country. The fluctuations and trends of environmental parameters, including temperature and precipitation, were identified using linear regression analysis. The authors concentrated on assessing the financial gains smallholder families in India's Indo-Gangetic Plain experienced after implementing CSA practices. The authors identified a primary climatic risk that contributed to low agricultural production and market hazards in the area as rising temperatures, depleting soil carbon, pests and diseases, and the intensity of floods. Net returns on the cost of adoption and changes in yields were evaluated using multiple regression analysis. In order to investigate the effects of the technology, 32 variables in total were computed, and climatic risk scenario analysis was performed. The main problems in the area were determined to be severe climate hazards from droughts, low and unpredictable rainfall, and scarcity of water resources. Convolutional neural networks (CNNs) were used to identify plant diseases using Y. Yadhav et al.'s algorithm [4]. A real artificial intelligence classification system requires non-linearity since the activation function is the most crucial component of the CNN model. As a result, a new mathematical activation function was developed and compared to existing activation functions on the basis of accuracy, leading to the discovery that the new activation function had a peak accuracy of 95%. The K-means clustering technique was used to identify the disease-affected area in order to maximize the use of fertilizer. Furthermore, they investigated the most commonly used activators and optimizers. A new mathematical activation function was created as a result, and its accuracy was compared to that of other activation functions. The comparison revealed that the new activation function had a peak accuracy of 95%. In order to maximize the use of fertilizer, the disease-affected area was identified using the K-means clustering technique. They also looked into the most widely employed activators and optimizers. The overall accuracy of the method was 95%. Crop yield estimation measures included the consideration of sensitivity and specificity. They merged three separate datasets—the soil, rainfall, and yield datasets—and utilized them as the basis for applying various supervised methods for estimating costs and gauging the accuracy of various methodologies. Three supervised techniques—K-Nearest Neighbor, Support Vector Machine, and Least Squares Support Vector Machine—are employed in this study. By comparing the efficacy of the suggested model's training and its mistake rate, it may be determined. This system has regression and classification capabilities. The categorization method divides the data into three categories (low, mid, and high), and the regression step estimates the real cost of yield production. The accuracy of KNN is 60%, SVM is 83%, and LS-SVM is 92%, demonstrating the excellent performance of LS-SVM in this situation. Their long-term goal was to create a farming recommendation system that would aid farmers in choosing the best time of the year for a given crop in order to prevent crop damage. Usman and others [6], When environmental concerns are taken into account, climate change is a major problem for everyone. According to estimates, climate change would also have a significant detrimental influence in India. It is already having a significant negative impact on the natural ecology globally. The most impacted industry is agriculture, which focuses on carbon dioxide emission patterns, climate change mitigation from India's perspective, and the sustainable growth of the agricultural bioeconomy.

III. PROPOSED SYSTEM

A. Architecture

Referring to Figure 1's architecture as demonstrated, after launching the programme, the user must sign up for a user account and log in to access the system's functions.The farmer needs to give permission for accessing his/her location. Parameters are analyzed according to the city such as values of nitrogen, phosphorus, etc in soil. Also, temperature, rainfall, and pH can be analyzed and based on all these parameters, the farmer will get a suitable crop for that region. Plant Disease Recognition, where plant diseases are forecast.

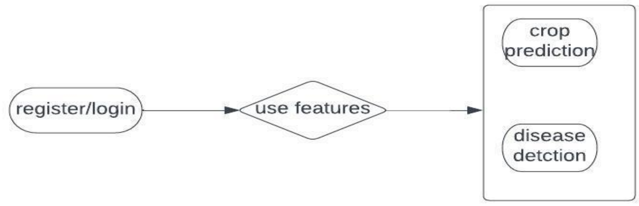

Fig. 1 System Block Diagram

B. Methodology Used :

Two submodules make up the system in total: crop prediction and plant disease detection.

1. Plant Disease Detection module:

The detection of disease was done on many kinds of plant leaves, including potato, tomato, bell pepper, etc. With the help of Keras, which is an open-source software framework that offers developers an ANN Python interface, this disease identification of leaf photos is accomplished. The Sequential API was used to create an ANN model. First, set various settings with default values, such as epochs, picture size, and the image's width and height. It was necessary to use a sizable dataset that included photos of both sick and healthy leaves from diverse plants. These photos from the dataset were retrieved from the directory in order to undergo additional processing. The processing phase comprised array conversion for the photos. Using Keras's preprocessing function "img to array," read the images with OpenCV, and then turn them into arrays. Label Binarizer was used to convert multi-class leaf disease labels to binary labels, while the Numpy module in Python was used to change the datatype of pictures to float. The dataset was subsequently divided into training and testing halves. In order to strengthen the model and reduce overhead memory usage, a real-time image augmentation was used while also training our model. Adam optimization, a stochastic gradient descent technique, was applied to enhance the model's output. Batches of the dataset for 25 iterations were trained, and it was seen that the base accuracy of 93.54% rose steadily with each iteration, peaking at 98.60% in the last epoch. Test accuracy for the model was assessed, and the resulting accuracy was 94.94%.

2. Crop Prediction Module:

Imported all the required libraries. The crop recommendation dataset is taken from Kaggle [5]. The dataset includes soil N, P, and K values, temperature, humidity, pH, rainfall, and a label column with the name of the advantageous crop. Once the farmer turns on the GPS, the latitude and longitude of that area is calculated with the help of free weather API. The crop is predicted on the basis of features such as soil's N, P, and K values as well as the temperature, humidity, soil pH, and rainfall which are present in the dataset along with latitude and longitude of that area and output features, which contain the names of the crops that will grow well. The data is then divided into a training set (80%) and a testing set (20%).Different classification techniques, including Decision Tree Classifier, Random Forest Classifier, Naive Bayes, SVM, and logistic regression are applied to the training dataset. The testing dataset's favorable crops are eventually identified using the trained models. The accuracy of the decision tree classifier, the random forest classifier, the naive bayes, the SVM, and the logistic regression are 90%, 99.09%, 99.05%, 97.95%, and 95.22%, respectively. This indicates that the random forest classifier provides the best accuracy and is therefore chosen for crop prediction.

Dataset Used:

Images of plant leaves from several species, including pepper bell, potato, and tomato, make up the dataset utilized for plant disease identification. It was possible to find this dataset on Kaggle.com. It has both pictures of healthy leaves and pictures of leaves with diseases like bacterial spots in pepper bells, leaf mold in tomatoes, late blight in potatoes, etc. It includes a total of 15 directories. On average, each directory has 1,300 images[7]. The training and testing portions of this dataset are split 80/20.To predict the most appropriate crop from Kaggle, we use the crop recommendation dataset. The soil values in the dataset are as follows: N (ratio of nitrogen content in soil), P (ratio of phosphorus content in soil), and K (ratio of potassium content in soil), temperature in degrees Celsius, relative humidity in percent, soil ph value, rainfall in millimeters, and a label column containing the names of the favorable crop [1]. The dataset comprises 2200 alternative circumstances for determining the best crop among wheat, mungbean, tea, millet, maize, lentil, jute, coffee, cotton, groundnut, peas, rubber, sugarcane, tobacco, kidney beans, moth beans, coconut, blackgram, adzuki beans, pigeon peas, and pomegranate.

Classifiers Used:

1) Random Forest : Random Forest is a famous and largely used machine learning algorithm used for dealing with classification and regression issues. This algorithm works by combining various decision trees. Each one of the trees, which is used in this algorithm, generates a classification for a particular set of qualities. “sklearn.Ensemble” is used to fit the training model. Random Forest algorithm is found to be most accurate for this project with accuracy of 99.09%. In this project this algorithm classifies crops according to their region of growth and suitable factors.

2) Support Vector Machine (SVM) : SVM is a machine learning algorithm which focuses on classification problems. SVM divides the huge dataset into n classes based on the best decision boundary . The data which will be arriving in the future will be categorized in these classes. The accuracy obtained by using SVM for this project is 97.95%. This project has used many images of leaves to detect diseases in the plants. SVM categorizes these diseases in various different categories.

3) Decision Tree Classifier : Decision tree classifier is a machine learning tool used for classification and prediction. It is a form of tree structure that mimics a flowchart, in which each internal node represents a test on an attribute, each branch a test result, and each leaf node (the terminal node) a class label. The accuracy obtained using this algorithm for the given project is 90%.

4) Naive Bayes :A group of classification algorithms founded on the Bayes' theorem is known as naive Bayes classifiers. It is not a single algorithm but a group of algorithms that all adhere to the same fundamental idea that each pair of features being classified is independent of the others as their guiding principle. The accuracy obtained with this algorithm is 99.05%.

5) Logistic Regression : For categorization and predictive analytics, this kind of statistical model is frequently utilised. Based on a collection of independent variables and a given dataset, logistic

regression calculates the likelihood that a particular event will occur, such as voting or not voting. The dependent variable's range is 0 to 1, so the result is a probability.

IV. RESULTS

A. Web App:

To exhibit the working of the Kisan- A multiservice app for farmers (https://github.com/Shrutimehta1211/kisan), we constructed a Web app.The Web App is designed using the Node Js framework.

Fig. 2 Home Page

Figure 2 shows the home page of the Web App, where the user can see both the features available i.e. Crop prediction and Disease detection.

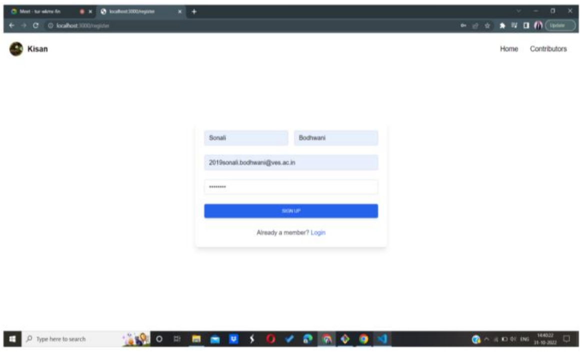

Fig. 3 Sign Up Page and Login page

Figure 3 shows the Sign Up and Login page where the new user first needs to sign up and then login into the web app.

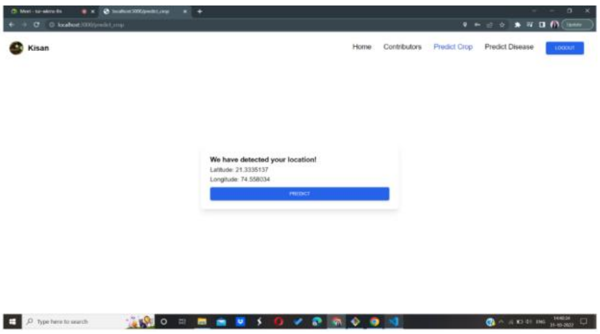

Fig 4: Location Detection

Figure 4 shows once the user login, the location is detected.

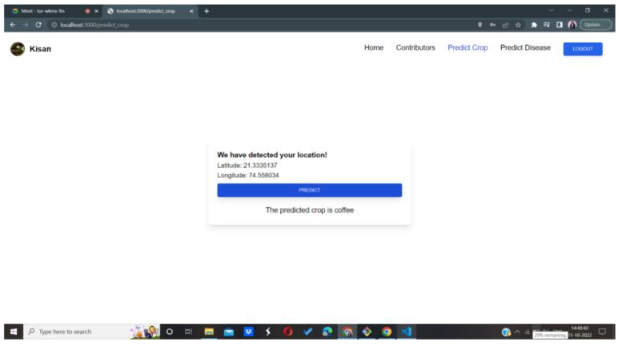

Fig. 5 Crop Prediction

Figure 5 shows once the location is detected, the crop is predicted for that particular region.

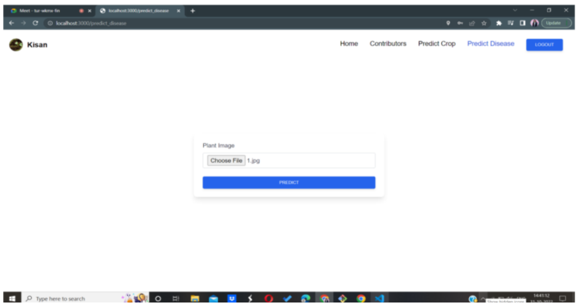

Fig. 6 Disease detection

Figure 6 shows the disease detection, where the user can add the image of the plant that is suffering from a particular disease.

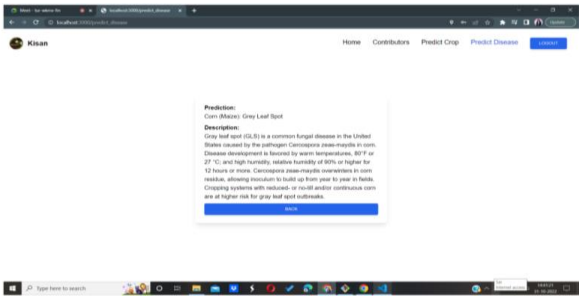

Fig. 7 Disease Detection

Figure 7 shows the user gets the name of the disease and a description for that disease.

B. Performance Evaluation

In order to assess the performance of the system, training and validation accuracy as well as training and validation loss were checked for plant disease detection. These results are shown in Figures 11 and 12. For crop prediction, this involved comparing the accuracy of the various algorithms that were used and discovering a well-suited algorithm, as shown in Figure 13.

Fig. 8 Training and validation accuracy

Fig. 9 Training and validation loss

The accuracy and error during validation acquired for each epoch of plant disease detection are shown on the graph, as can be seen in figures 8 and 9 above. It is obvious to see that accuracy rises and validation loss falls with each run.

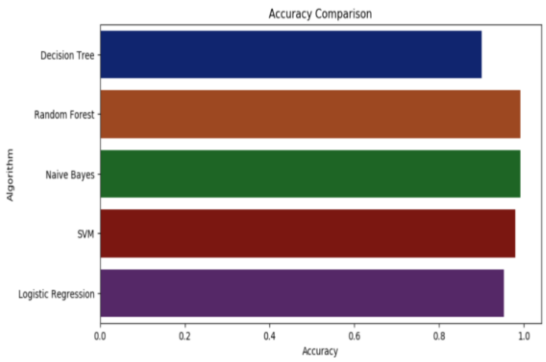

Fig. 10 Accuracy comparison

Figure 10 shows that after training the model using various classification algorithms such as Decision Tree Classifier, Random Forest Classifier, Naive Bayes, SVM, and logistic regression, the accuracies are 90%, 99.09%, 99.05%, 97.95%, and 95.22%, respectively. This demonstrates that Random Forest Classifier gives the best accuracy for the prediction of the most favorable crop based on N, P, and K values of the soil and temperature, humidity, soil pH, and rainfall.

V. CONCLUSION AND FUTURE WORKS

Features like Disease Detection and Cure, Productivity Prediction, and Feed with Tips and Tutorials have been implemented to help farmers who are working with tiny plots of land enhance their yield. To help raise awareness and keep farmers up to date on the most recent developments in the agriculture industry, this web app will have a feed feature.This website application will assist farmers in learning more about the market and serve as a special interface for schemes and payments. The farmers will remain informed about new farming methods and fashions. However, initially, novice users can find it challenging to use. Overall, this system aspires to be quicker, safer, and more comfortable for all users. Agriculture and allied sectors contribute significantly to the Indian economy. In today's digital world, where everything is controlled by electronics, these professions must remain relevant. One of the objectives was to increase awareness of the various e-farming characteristics among our Indian farmers. Although the effort is small, it is not constrained in any way. Both have potential for e-farming in the future and have potential for the system to appear to be very broad. Future initiatives include (1) expanding crop forecast datasets, (2) adding more leaf diseases and improving model accuracy, (3) including renting tools feature, and an online payment method for renting tools, and (4) only presenting information on tools made accessible by farmers in the user's location.

REFERENCES

[1]. M. Kalimuthu, P. Vaishnavi and M. Kishore, "Crop Prediction using Machine Learning," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 926-932, doi: 10.1109/ICSSIT48917.2020.9214190.
[2]. S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 2018, pp. 41-45, doi: 10.1109/ICDI3C.2018.00017.
[3]. M. Gulzar, G. Abbas and M. Waqas, "Climate Smart Agriculture: A Survey and Taxonomy," 2020 International Conference on Emerging Trends in Smart Technologies (ICETST), 2020, pp. 1-6, doi: 10.1109/ICETST49965.2020.9080695
[4]. Y. Yadhav, T. Senthilkumar, S. Jayanthy and J. J. A. Kovilpillai, "Plant Disease Detection and Classification using CNN Model with

Optimized Activation Function," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 564-569, doi: 10.1109/ICESC48915.2020.9155815.

[5]. EFFICIENT CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS, Arun Kumar, Naveen Kumar, Vishal Vats, International Research Journal of Engineering and Technology (IRJET) 2018.

[6]. A. Usman, "Sustainable development through climate change mitigation and biomass agriculture: India's perspective," 2017 IEEE Conference on Technologies for Sustainability (SusTech), 2017, pp. 1-7, doi: 10.1109/SusTech.2017.8333504.

[7]. Plant Disease Detection Using Different Algorithms.Trimi Neha Tete, Sushma Kamlu,DOI: 10.15439/2017R24

[8]. T. N. Tete and S. Kamlu, "Detection of plant disease using threshold, k-mean cluster and ann algorithm," 2017 2nd International Conference for Convergence in Technology (I2CT), 2017, pp. 523-526, doi: 10.1109/I2CT.2017.8226184.

[9]. Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems engineering, 144, 52-60.

[10]. Suresh, V & Krishnan, Mohana & Hemavarthini, M & Jayanthan, D. (2020). Plant Disease Detection using Image Processing. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS030114.

[11]. Pujari, J. D., Yakkundimath, R., & Byadgi, A. S. (2015). Image processing based detection of fungal diseases in plants. Procedia Computer Science, 46, 1802-1808.

[12]. Samer D. M , Subramaniya Raman M. K, 2020, E- Farming: A Breakthrough for Farmers, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 07 (July 2020).