IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Smart Harvesting Application for Crop Recommendation and Disease Detection

Suraj Savji[1], Tushar Pawar [2], Gokul Borade [3], Harish Ghodke [4], Avishkar Khude[5]
Guide: Dr. Preeti Patil

Department of Information Technology
Dr. D.Y. Patil, College of Engineering Akurdi, Pune
Maharashtra, India

Abstract — **As we know, Agriculture is the first and foremost factor that is important for survival and India is an agricultural country with the second highest land area occupied by agriculture. This project will help the farmers to know the suitable crop for that particular soil and environmental conditions, before cultivating onto the agricultural field and thus help them to make the appropriate decisions. Implementation of such a system with an easy-to-use Android Application using Machine Learning Algorithms will be carried out. The results of the prediction will be made available to the farmers. This project mainly focuses on predicting the crop by applying various machine-learning techniques. The classifier models used here include various classification techniques like Random Forest is used for the prediction of crops. The prediction made by machine learning algorithms will help the farmers decide on which crop to grow to induce the most yield by considering factors like temperature, rainfall, area, etc. Also, this app contains lots of other features like Disease Detection, Mandi price, Ask AI, Community feature, Agri News, and Multi-language functionality. In this way, this app bridges the gap between technology and the agriculture sector.**

Index Terms – Crop Prediction, Random Forest, CNN Algorithm, Android Application.

I. INTRODUCTION

Agriculture is the backbone of India's economy, with a significant portion of the population dependent on it for their livelihoods. However, Indian farmers face a myriad of challenges, including unpredictable weather patterns, varying soil types, and the uncertainty of crop selection. In a country where agriculture occupies the second-largest land area globally, the need for innovative solutions to support farmers and enhance agricultural productivity is paramount.

This project is designed to revolutionize the way farmers operate by harnessing the power of technology and data-driven insights. Our goal is to create a comprehensive platform that empowers farmers with knowledge, enabling them to make informed decisions that can significantly boost their crop yield and subsequent profits. By addressing the issues of crop prediction and disease detection, this initiative aims to minimize the guesswork in farming, making it more efficient and sustainable.

The core concept of this project revolves around the development of a cutting-edge application that employs machine learning to predict crop yields based on crucial environmental parameters such as temperature, area, rainfall, and soil pH. Moreover, the project includes a robust image recognition system for the early detection of crop diseases. By providing farmers with these critical insights, we aim to equip them with the tools they need to make well-informed choices about their crops, ultimately resulting in more resilient and productive agricultural practices.

Building on the existing project, we are excited to introduce several new and innovative features that will further enhance its capabilities. These enhancements include state-of-the-art technology for precise yield prediction and efficient disease detection. To ensure accessibility to a diverse audience, the project will be available in multiple languages, including Hindi, English, and Marathi, ensuring a user-friendly experience for all farmers.

II. LITERATURE SURVEY

A. Venugopal, Rima Mathew [1], This paper focuses on the prediction of crops and the calculation of their yield with the help of machine learning techniques. Several machine learning methodologies are used for the calculation of accuracy.

Random Forest classifier and Decision tree classifier were used for the crop prediction for the chosen district. The proposed technique helps farmers in decision-making of which crop to cultivate in the field. This work is employed to search out the gain knowledge about the crop that can be deployed to make an efficient and useful harvesting. The accurate prediction of different specified crops across different districts will help farmers of Kerala.

P. Dharmendra Kumar, A. Suhasini, D. Anand [2], The study presents a crop disease detection system that uses 2DCNN to achieve 98% accuracy through sophisticated preprocessing and segmentation techniques. Focusing on automated disease detection highlights the need to grow the database and add 7 DYPCOE, Akurdi Department of Information Technology more training data to improve disease identification in future research.

Nidhi Kundu, Geeta Rani, Kalpit Gupta [3], In This research paper, the authors achieved the object of automating the disease detection, classification, and crop loss estimation for maize crops. The authors pre-processed the collected dataset by applying the K-Means algorithm. they applied both supervised and unsupervised algorithms on pre-processed, and non-pre-processed datasets for classifying images. The authors trained VGG-19, ResNet-50, and finetuned VGG-16 models the proposed DL model MaizeNet, on a dataset comprising 2460 images. These images include four classes namely healthy, TLB, rust, and multi-disease. Among all the above-stated models, MaizeNet reported the highest accuracy of 98.50% on the testing dataset comprising 536 images.

III. METHODOLOGY

As we know, Agriculture is one of the most important sectors of our life. Agriculture is the primary source of food for the world's population. It provides the necessary resources to sustain human life by producing a variety of crops, fruits, and vegetables.

Algorithms used in this project:

1) Random Forest
2) Convolutional Neural Network (CNN)

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Assumptions for Random Forest:

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output.

Therefore, below are two assumptions for a better Random Forest classifier:

• There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.

• The predictions from each tree must have very low correlations.

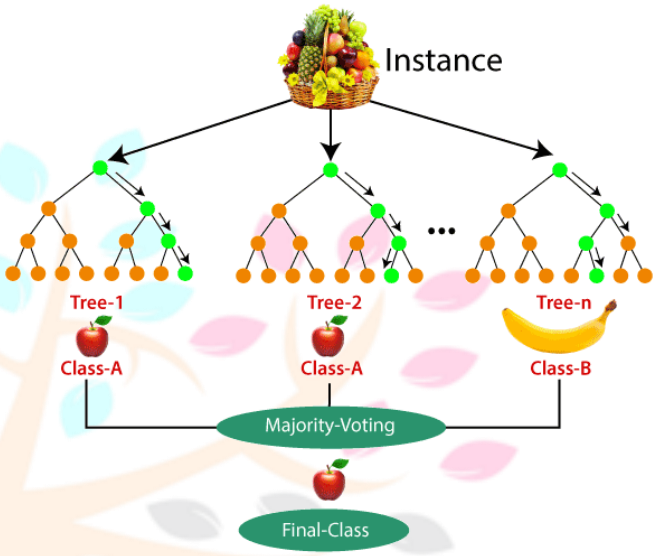

Fig. Random Forest

Working of Random Forest:

Random Forest works in two phases. first is to create the random forest by combining N decision trees, and the second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step 1: Select random K data points from the training set.

Step 2: Build the decision trees associated with the selected data points (Subsets).

Step 3: Choose the number N for decision trees that you want to build.

Step 4: Repeat Steps 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

The accuracy of the Random Forest algorithm for crop prediction is between 80% to 85%.

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms that have proven to be highly effective in various computer vision tasks, such as image classification,

object detection, and image segmentation. CNNs are particularly well-suited for tasks where the input data has a grid-like structure, such as images.

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image.

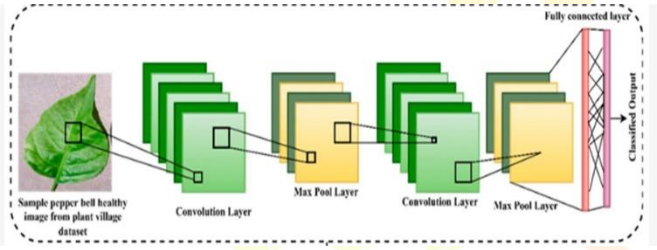

Fig. Convolutional Neural Network

Applications of CNN:

- Decoding Facial Recognition
- Understanding Climate
- Collecting Historic and Environmental Elements

The accuracy of CNN for disease detection lies between 80% to 93%.

This application also helps farmers to detect the various diseases that occur on the leaves of the crop.
Various types of diseases of leaves:

To deal with this kind of disease, first, we need to identify the type of disease. To detect the disease, we have used CNN i.e. Convolutional Neural Network.

We used Kaggle datasets for processing efficiently using the feature extraction method and Random Forest and CNN for the prediction of crops and detection of disease respectively.

Datasets used for the project:

1) Crop Recommendation Dataset
2) New Plant Diseases Dataset

Crop Recommendation Dataset:
This is a dataset that would allow the users to build a predictive model to recommend the most suitable crops to grow on a particular farm based on various parameters.

Data fields:
N - ratio of Nitrogen content in soil
P - ratio of Phosphorous content in soil
K - ratio of Potassium content in soil
temperature - temperature in degrees Celsius
humidity - relative humidity in %
ph - ph value of the soil
rainfall - rainfall in mm

New Plant Diseases Dataset:
This dataset consists of 87,900 images of leaves spanning 38 classes. Each class denotes a combination of the plant the leaf is from and the disease (or lack thereof) present in the leaf. All images are 256*256 in resolution.

The dataset is divided into three parts as follows:
train - 70,295 images divided into 38 classes with 1,642 to 2,022 images per class.
valid - 17,572 images divided into 38 classes with 410 to 505 images per class.
test - 33 images (These images are not divided into their respective classes but the class can be inferred from the image filename).

Data Flow Diagram is a graphical representation that illustrates the flow of data within a system. It typically consists of processes, data stores, data flows, and external entities.

The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

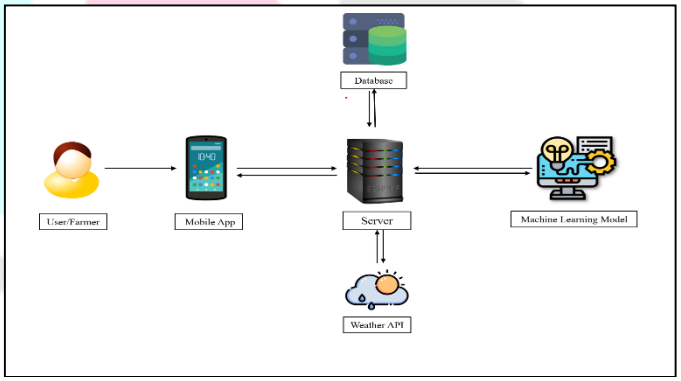

Fig. Data flow Diagram

The above dataflow diagram shows that user can use their mobile phones to enter the details that are necessary for crop prediction and then by using machine learning algorithms we can get information about suitable crops for environmental conditions and soil type.

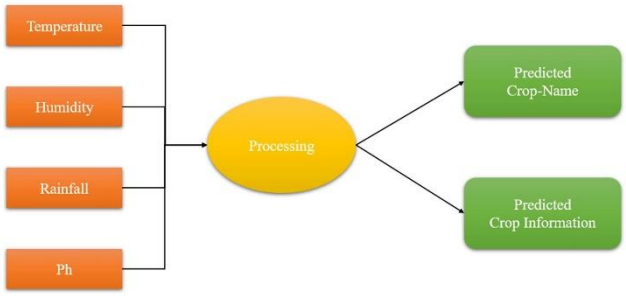

Fig. DFD for Crop Prediction

In the above DFD, we can see that when a farmer enters the necessary details that are required for crop prediction, then by using a machine learning algorithm (Random Forest here), we can predict the crop along with the crop information that is suitable for those environmental conditions.

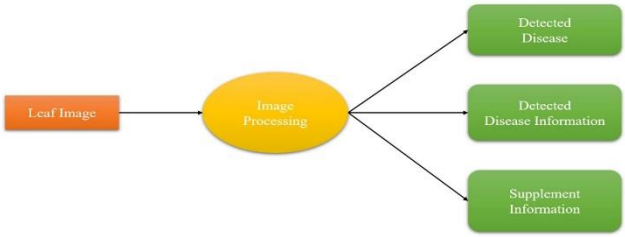

Fig DFD for Crop Disease Detection

Above is the DFD for Crop Disease Detection in which a leaf image can be taken as an input and by using CNN i.e. Convolutional Neural Network, we can perform image processing to identify the disease that happened to that crop. Also, we can suggest some medicines to cure that disease.

Designing a **Proposed system architecture** depends on the specific requirements and goals of the system you are developing.

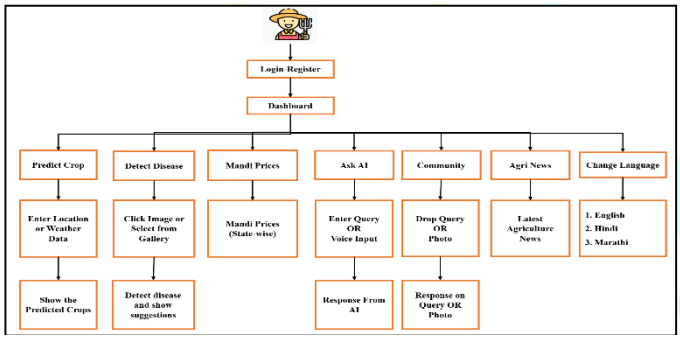

Fig. Proposed system architecture

Above is the Proposed system architecture which displays all the functionalities that are included in our application which can be used by farmers.

In the proposed system we developed an application. A Random Forest classifier and Convolutional neural network were employed to predict the crop and analyze images of crop leaves that have the disease.

This app contains functionalities like Predict Crop, Detect Disease, Mandi Price, Ask AI, Community feature, Agri News, and Multilanguage functionality.

Building on the existing project, we are excited to introduce several new and innovative features that will further enhance its capabilities. These enhancements include state-of-the-art technology for precise yield prediction and efficient disease detection. To ensure accessibility to a diverse audience, the project will be available in multiple languages, including Hindi, English, and Marathi, ensuring a user-friendly experience for all farmers.

Additionally, we are integrating an OpenAI chatbot to provide instant assistance and answers to farmers' queries, making vital information readily accessible at their fingertips. To foster a sense of community and knowledge-sharing among farmers, we are introducing a social networking feature like Instagram. This community-building aspect will allow farmers to connect, share their experiences, discuss yield-related challenges, and access additional information, creating a sense of unity and progress within the agricultural community.

By combining technological innovation with community-building, our project seeks to empower Indian farmers, ensuring that they are better equipped to navigate the challenges of agriculture. Through this, we aim to not only benefit the farmers but also contribute to a more resilient and productive agricultural sector, ultimately ensuring food security for the nation.

MOTIVATION

The motivations for undertaking this project are as follows:

1. Solving Real-World Challenges: This project is motivated by the opportunity to address real-world challenges faced by farmers in India. It recognizes the significance of mitigating the impact of unpredictable weather, soil variations, and crop selection dilemmas on the agricultural sector.

2. Empowering Farmers: We are motivated by the goal of empowering farmers with knowledge and tools that can make their livelihoods more sustainable and financially viable. By equipping them with data-driven insights, we aim to enhance their decision-making capabilities.

3. Enhancing Efficiency: Agriculture is the backbone of India's economy, and this project is driven by the aspiration to enhance the efficiency of farming practices. Through accurate crop yield predictions and early disease detection, we intend to minimize guesswork and resource wastage.

4. Contributing to Food Security: This project is motivated by the crucial role that agriculture plays in ensuring food security for the nation. By increasing crop yields and reducing losses, we aim to contribute to a more secure and reliable food supply.

5. Inclusivity: The motivation lies in creating a solution that is inclusive and accessible to farmers from diverse linguistic and technological backgrounds. By offering multiple language options and user-friendly interfaces, we aim to bridge the digital divide.

6. Building a Strong Agricultural Community: We are inspired by the prospect of creating a strong sense of community and knowledge-sharing among farmers. The project's community-building feature is motivated by the desire to foster unity and progress within the agricultural sector.

7. Sustainability and Progress: The long-term motivation is to create a sustainable and scalable solution that can adapt to the evolving needs of the agricultural sector. We aspire to contribute to the lasting progress and resilience of Indian agriculture.

In summary, this project is motivated by a deep commitment to addressing agricultural challenges, empowering farmers, enhancing efficiency, ensuring food security, promoting inclusivity, fostering community, and contributing to the sustainable progress of agriculture in India. It aims to be a catalyst for positive change and innovation in the agricultural sector, ultimately creating a brighter and more secure future for Indian agriculture.

RELATED WORKS

Crop forecasting, a data-intensive endeavor, is an ideal application for data mining. Traditionally, agriculture planning has relied on these predictions to forecast crop yield rates, weather conditions, soil quality, and crop classifications. However, one notable limitation has been the selection of the most suitable crop from a myriad of choices available. In response, a Crop Selection Method (CSM) has been proposed to maximize net yield rates for specific crops during a season. CSM considers factors such as sowing time, plantation days, and predicted yield rates to identify the sequence of crops that will maximize daily production.

Plant disease detection is another critical facet of agriculture. Neglecting the early signs of disease can result in significant crop losses with far-reaching economic consequences. Recent developments in plant disease identification have been significant. Notably, Convolutional Neural Networks (CNNs) have played a pivotal role in the precise classification of plant diseases. In one instance, a deep learning model trained on a substantial dataset of 87,000 RGB images achieved a remarkable 93.5% recognition accuracy. This model underwent preprocessing, segmentation, and CNN-based classification, but challenges emerged when dealing with certain disease classes.

TESTING AND EVALUATION PHASE

The testing and evaluation phase in machine learning is the moment of truth for a trained model. Once the model has been fine-tuned and validated, it is put to the test using a separate dataset that it has never seen before. This unseen data allows us to assess how well the model generalizes to new examples and whether it can make accurate predictions in real-world scenarios. During testing, the model's performance metrics, such as accuracy, precision, recall, and F1 score, are calculated to quantitatively evaluate its effectiveness. These metrics provide insights into the model's strengths and weaknesses, helping to gauge its reliability and suitability for the intended task. The testing and evaluation phase is crucial for ensuring that the model not only performs well on the data it was trained on but also demonstrates robustness and efficiency in handling novel inputs.

It is important to note that the effectiveness of such an app depends on the quality and accuracy of the data inputs, the sophistication of the algorithms used, and the integration of the technology into existing farming practices. Additionally, considerations about data privacy, accessibility, and education for farmers are crucial for the successful implementation of these technologies.

IV. DISCUSSION

In this way, we have created a system that consists of lots of functionalities that can help farmers in decision-making about which crop to grow and help to solve various doubts by using community features.
The applications that are already available in the market do not come with all the functionalities in one place. So, this app is a solution for that problem where we can get all the necessary functions in a single app.

ADVANTAGES:

1. Easily Predict Crops using Weather conditions.
2. Detect diseases and Suggest Supplements spread on crops using images.
3. Multi-Language Translation Platform.
4. Community for Farmers where farmers can connect with each other.
5. Help to improve Agriculture Yield.

APPLICATIONS:

1. Crop Prediction using weather conditions
2. Disease Detection using crop leaf images
3. Multi-Language Platform
4. E-mandi prizes (State-wise)
5. Community Feature
6. Weather Alerts
7. Agri news

V. CONCLUSION

Farmers are not connected to any technology in the current system. Therefore, the likelihood of financial loss is high. Occasionally, choosing the incorrect crops can hinder their ability to grow financially. This Android application assists farmers in choosing appropriate crops for a given area to minimize these losses. This application can also assist in identifying diseases from crop images. If the crop is afflicted with any diseases, image processing methods and machine learning can be used to identify the problem and provide the farmer with an appropriate remedy. By increasing the yield rate of crop production, this strengthens the economy of the nation. This offers a multilingual platform in their mother tongue for farmers to interact with applications and a farmer community platform for them to connect with one another

Furthermore, the application can contribute to sustainable farming practices by promoting efficient use of resources such as water, fertilizers, and pesticides. This, in turn, helps reduce environmental impact and ensures the long-term viability of agricultural systems.

VI. REFERENCES

[1] Website – "Android Studio Tutorial"
Link -https://www.javatpoint.com/android-tutorial

[2] Rushikesh Bhave, Mayuresh Deodhar, Kevin Bhalodia, Prof. Mansing Rathod "Android Application for Crop Yield Prediction and Crop Disease Detection" 2018 International Journal of Innovative Science and Research Technology (ISSN No: -2456-2165).

[3] Website – "Machine Learning Tutorial"
Link -https://www.javatpoint.com/machine-learning

[4] 4] S. Veenadhari, Dr. Bharat, Dr. CD Singh "ML approach for forecasting crop yield based on climatic parameter" 2014 International Conference on Computer Communication and Informatics (ICCCI-2014), Jan. 03-05,2014 Coimbatore, India.

[5] Anakha Venugopal, Aparna S, Jinsu Mani "Crop Yield Prediction using Machine Learning Algorithms" 2021, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181.

[6] P. Dharmendra Kumar, A. Suhasini, D. Anand "Crop Disease Detection using 2D CNN Based Deep Learning Architecture" 2023 ISSN:2147-6799, Feb 12/02/2023, India.

[7] Website – "Language Translation"
Link- https://developers.google.com/ml-kit/language/translation

ACKNOWLEDGMENT

We are highly indebted to the Project Guide Dr. Preeti Patil, for her guidance and constant supervision as well as for providing necessary information regarding the development of the system.