

CROP WATCH – Empowering Precision Farming through ML and Image Analysis

Dewanshi Pramani¹, Chandni Yadav², Gargi Balsaraf³, Niyati Wankhede⁴, Ishwar Bharambe⁵

¹Department of Computer Engineering Ajeenkya DY Patil School of Engineering, Pune, India

²Department of Computer Engineering Ajeenkya DY Patil School of Engineering, Pune, India

³Department of Computer Engineering Ajeenkya DY Patil School of Engineering, Pune, India

⁴Department of Computer Engineering Ajeenkya DY Patil School of Engineering, Pune, India

⁵Professor, Department of Computer Engineering Ajeenkya DY Patil School of Engineering, Pune, India

Abstract— "CROP WATCH - Empowering Precision Farming through ML and Image Analysis" uses cutting-edge machine learning and image analysis technology to transform contemporary agriculture. This document gives a summary of the project, emphasizing its main characteristics and functions. Through the simple upload of photos, the app enables users to identify different types of seeds and evaluate their quality, promoting well-informed agricultural decision-making. Furthermore, it helps identify weed plants and detect plant illnesses early in agricultural fields, improving crop health and yield. It provides localized weather predictions with sophisticated machine learning algorithms for better crop planning. The programme also offers recommended pesticides, crop-specific production guidelines, and essential information. "CROP WATCH" is an effective tool at the nexus of agriculture and technology that makes precision farming practices possible and supports the production of food in a sustainable manner.

Keywords: Precision farming, Machine learning, Image analysis, Seed Recognition, Plant Disease Detection, Weed Identification, Convolutional Neural Network.

1. INTRODUCTION

Creative solutions are critical in the quickly changing agricultural sector. In the field of precision farming, "CROP WATCH" stands out as a trailblazing initiative that uses image analysis and machine learning to revolutionize agricultural methods. This all-inclusive Android application, created for our computer engineering senior project, has the potential to completely transform how farmers and agriculture enthusiasts interact with the land. Fundamentally, "CROP WATCH" presents a fresh method for seed analysis. It enables users to make informed decisions from the planting stage onward by identifying different types of seeds, evaluating their quality, and extracting essential seed information from photographs supplied by users.

"CROP WATCH" goes further than that, though. In order to further improve crop management, our application expands its capabilities to include the weed plant identification and the detection of agricultural illnesses, and accurate weather predictions. It's a comprehensive toolset for contemporary farming, giving farmers knowledge and warnings to safeguard their crops and increase harvests.

"CROP WATCH" is unique because of its customized feel. The programme transforms into a virtual agronomist by letting users choose particular crops, providing customized cultivation advice, suggested pesticides, and an extensive inventory of supplies needed for the crop of choice. It's a virtual assistant that makes farming less complicated.

Apart from its distinct functionalities, "CROP WATCH" cultivates a feeling of camaraderie among its users. Through knowledge-sharing and conversation, farmers and hobbyists may connect, share thoughts, and work together to improve farming techniques through the app's "Community" function.

"CROP WATCH - Empowering Precision Farming through ML and Image Analysis is a technological revolution in agriculture, not just a smartphone application. An overview of our study is given in this paper, with a focus on its novel aspects, possible effects, and role in transforming agriculture in the future.

2. LITERATURE SURVEY

[1]The potential of Deep Learning (DL), and more especially Convolutional and Deep Neural Networks (CNN and DNN), to transform crop disease detection is examined in this research. The study intends to improve early disease identification in plants by utilizing these cutting-edge ML techniques, with an emphasis on visualizing disease symptoms. These models are assessed using a range of efficiency indicators, and the study finds important gaps in plant disease detection

that require prompt attention. The suggested method uses CNN and DNN algorithms for routine field monitoring in order to identify diseases early. The model is trained using machine learning techniques to inform choices about disease control, including the application of pesticides. Future upgrades might provide more elements including costs for pesticides and government market data.

[2] This research explores the use of image processing for accurate plant disease identification. Image acquisition, extraction, segmentation, and pre-processing are the main processes in the process, which focuses on different plant elements such leaves, stems, and fruits. Utilizing Convolutional Neural Networks (CNNs), a type of Deep Learning, the research presents a model intended to identify illnesses in a variety of plant species.

The model, which was created with the TensorFlow and Keras frameworks and implemented on Android, demonstrates its efficacy. The findings show that the MobileNet architecture provides better illness detection accuracy than other models. Future efforts will expand the project's scope to encompass a greater variety of plant species and diseases as part of its expansion.

[3] This work explores the critical role that machine learning (ML) plays in supporting crop selection and management decisions for Crop Yield Prediction (CYP). The study emphasises the need of data accessibility in feature selection and argues in favour of prioritising quality over number. It emphasises the widespread application of KNN regression, random forests, and neural networks while acknowledging the necessity for further advancements in CYP. The study highlights the potential of machine learning (ML) in the field of agriculture as well as the necessity of improving feature selection, particularly with regard to temperature changes. It also recommends using deterministic crop models for more precise forecasts, such as CO2 fertilization, and explicitly addressing border topographical areas, non-parametric modelling, and other approaches.

[5] This study uses seed classification to solve the problem of plant weeds and their effect on crop productivity. It presents an approach that makes use of the ID3 algorithm to identify unwanted seeds using seed photos and sample seed datasets. The study also examines the crucial component of soil selection for productive crop production, using data from sample datasets to influence soil selection. Using Support Vector Machine (SVM) techniques, the method compares attributes taken from new seeds with those of sample seeds in order to forecast crop growth and any diseases. The study shows that by training the dataset in this way, it is possible to categorize seeds, forecast crop

development, and, depending on those projections, start preventive actions.

[6] This paper addresses the pressing issue of agricultural plant diseases and pests, emphasizing the importance of rapid and precise diagnosis to mitigate financial losses. Employing cutting-edge advancements in Deep Neural Networks, the research focuses on refining the identification and recognition of tomato plant objects. It explores the utilization of deep feature extractors like VGG net and ResNet, coupled with three main deep-learning meta-architectures: Faster R-CNN, R-FCN, and SSD. The study introduces methodologies for enhancing accuracy and diminishing false positives during training through data augmentation and local and global class annotation techniques. The effectiveness of these systems in discerning nine distinct disease and pest categories is showcased through training and testing on a substantial dataset of tomato diseases and pests.

3. RELATED WORK

3.1 Plantvillage

Plant Village stands out as the most extensive plant disease dataset, originating in 2016 and comprising 54,309 images across 14 crop species, including apple cherry, corn, grape, orange, peach, raspberry, soybean, squash, blueberry, strawberry, and bell pepper, potato, tomato. These meticulously selected photographs displaying both healthy and diseased crop leaves are available via the Plant Village website (www.plantvillage.org). The dataset categorizes these crops' illnesses into distinct categories, encompassing 17 fungal diseases, two viral diseases, and four bacterial diseases, two mold (oomycete) diseases one mite disease. It features 38 classes representing various plant diseases and includes an additional class for background images. The dataset's inception marked the commencement of an ongoing crowdsourcing initiative, aiming to harness computer vision methodologies to address yield losses in crops due to infectious diseases. To collect images from fields with infected crops, technicians methodically removed leaves from plants placing them in front of a paper sheet featuring either a grey or black background, all photographs were taken under ideal lighting conditions. Subsequent to collection, the images underwent editing, involving the removal of extraneous background elements and aligning all leaves to point upward. These refined images from the dataset are commonly referred to as laboratory images.

3.2 Plantdoc

Tomato, apple, bell pepper, blueberry, cherry, maize, grape, peach, potato, raspberry, soybean, strawberry, and squash powder are among the 13 distinct plant

species represented in the 2,569 images of PlantDoc. To facilitate tasks like object detection and picture classification, the dataset is divided into thirty different classes that distinguish between samples that are healthy and those that are sick. These images in PlantDoc predominantly stem from instances of plant diseases in the field, sourced from online platforms, and meticulously annotated to function as training data for models designed to identify crop diseases in real-world scenarios. Notably, due to their internet-sourced origin, the images often exhibit suboptimal quality, with some instances featuring leaves that were not captured on plants, resembling more controlled laboratory settings. The dataset's annotation process faced challenges, potentially introducing errors, as certain plant diseases share striking visual similarities. Despite these challenges, the dataset serves as a valuable resource for training models. However, it is acknowledged that the number of annotated images across different categories may be limited, posing potential constraints for training high-accuracy models. Therefore, considerations for addressing these limitations and enhancing dataset quality are essential for robust model development and accurate disease detection in real-world agricultural scenarios

3.3 Our Model's Dataset

For the Support Vector Machine (SVM) model, a dataset comprising 52 images was acquired from open sources. Of them, fifteen were set aside for testing and the other 37 photos were assigned for training. The SVM model's objective is to discern whether a leaf is classified as healthy or infected. In contrast, the Convolutional Neural Network (CNN) utilized data sourced from the Kaggle website, specifically from the "new plant diseases dataset." This dataset encompasses a substantial 12,949 images designated for training. The images within this dataset represent various crop leaves, encompassing, but not restricted to, the following: potato, tomato, strawberry, grape, peach, corn (maize), apple, cherry and so on. Each image in the dataset is meticulously labeled, attributing different categories to diseases and indicating healthy leaves. This annotation process aids the model in effectively discerning between healthy and affected leaves during its training phase. The soybean seed dataset encompasses five distinct categories of individual seed images, totaling 5513 images, including intact, spotted, immature, broken, and skin-damaged seeds, with over 1000 images per category. These images were categorized based on the Standard of Soybean Classification (GB1352-2009). Captured by an industrial camera, images featuring soybean seeds in physical contact were isolated, with individual soybean images (227x227 pixels) extracted from larger images (3072x2048 pixels) using a segmentation algorithm boasting over 98% accuracy. This dataset offers valuable insights for studying soybean seed classification and

quality assessment. Additionally, it contains 1300 images depicting sesame crops and various weed types, each meticulously annotated with bounding boxes. using both YOLO and Pascal VOC labeling formats. The images are all 512 x 512 pixels in size and in color.

Table-1. Comparison table

<i>Dataset</i>	<i>Plant Village</i>	<i>Plant Doc</i>	<i>New Plant Diseases Dataset</i>
<i>Number of images</i>	54309	2569	87000(approx.)
<i>Classes</i>	38	30	38
<i>Strengths</i>	Extensive dataset, diverse crops, focused initiative	Variety of plant species, real-world instances	Large dataset, diverse classes
<i>Weaknesses</i>	Limited annotated images, potential bias	Suboptimal image quality, limited annotations	Possible limitations due to offline augmentation
<i>Focus</i>	Addressing crop yield losses, computer vision	Real-world disease detection, model training	Image classification, disease detection

4. SYSTEM DESIGN ARCHITECTURE

4.1. Architecture Description

The system design architecture of "CROP WATCH - Empowering Precision Farming through ML and Image Analysis" comprises several key modules to ensure efficient plant disease detection in a web application. The integration module forms the backbone, facilitating seamless communication between different components. It acts as a central hub, coordinating data flow among the decision module, feature extraction module, data preprocessing, data collection, and image upload.

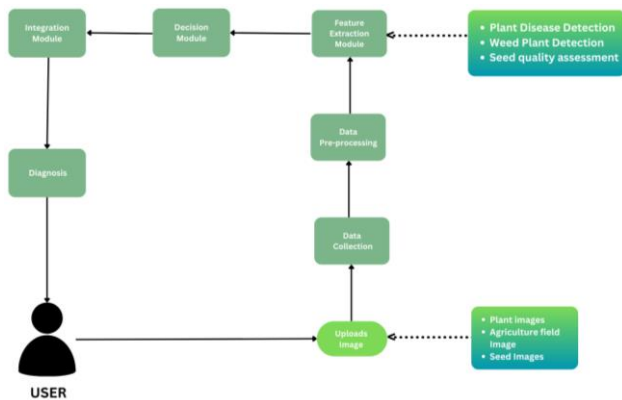


Fig- 1: System Architecture design flowchart

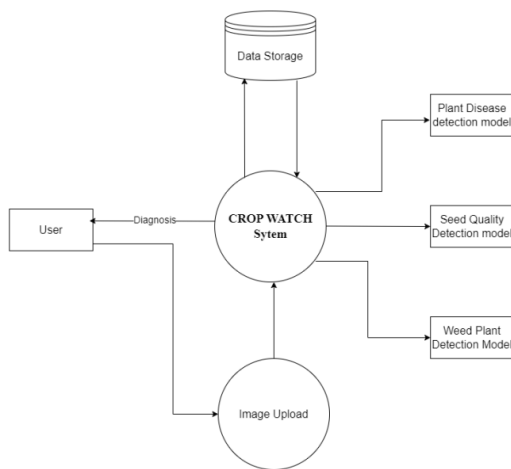


Fig-2: Workflow

4.2 Module

The image upload module allows users to contribute to the dataset by uploading images of diseased plants. This crowd sourced data enriches the system's knowledge base, enhancing the model's accuracy over time. Together, these modules form a cohesive architecture that empowers the "CROP WATCH" app to deliver accurate and real-time plant disease detection, supporting precision farming practices for improved agricultural outcomes.

Data collection: Assembling a diverse dataset through collaborations, open datasets, and user contributions via an image upload feature within the web application.
Image uploading; Allowing users to contribute to the dataset by uploading images, fostering a collaborative approach to data enrichment.

Data preprocessing ensures that incoming data is formatted and standardized, enhancing the efficiency of subsequent analysis. The data collection module gathers information from various sources, including user inputs and external databases, creating a comprehensive dataset for model training and validation.

The feature :The extraction module is in charge of gathering pertinent data from the input photos, capturing distinctive patterns indicative of various diseases.

The decision module lies at the core of disease identification, employing advanced machine learning algorithms, and Convolutional Neural Networks (CNNs). This module analyzes features extracted from images to make informed decisions regarding the presence and type of plant diseases.

Integration Module: This central hub ensures seamless communication among modules. It facilitates data flow between the decision module, feature extraction module, data preprocessing, data collection, and image upload.

4.3 Technology Used

The "CROP WATCH" project leverages cutting-edge technologies to empower precision farming through ML and image analysis. The core technology employed is machine learning, specifically Convolutional Neural Networks (CNNs), for robust and automated plant disease detection. CNNs have revolutionized image analysis, enabling the app to discern intricate patterns and features in plant images, leading to accurate disease identification.

Image analysis techniques are pivotal in the project, encompassing a spectrum of processes such as feature extraction, enabling the system to capture and analyze distinctive traits indicative of various plant diseases. Data preprocessing techniques are utilized to ensure standardized and optimized data input for efficient model training, enhancing the overall accuracy of disease detection.

The integration of cloud technology facilitates real-time data processing and storage, ensuring scalability and accessibility of the system. Additionally, the web app capitalizes on web platform features for seamless user interaction, creating an intuitive interface for farmers. Furthermore, the project harnesses the power of crowd sourced data through an image upload feature. This collaborative approach enriches the dataset, allowing the model to continuously learn and improve its accuracy over time.

Technology Stack:

TensorFlow: TensorFlow served as the core deep learning framework for developing Convolutional Neural Network (CNN) models. Leveraging Tensor Flow's extensive capabilities, we constructed models for plant disease detection, weed detection, and seed identification. Its rich ecosystem and support for machine learning tasks made it a pivotal choice for model development.

Python: Python, a versatile and widely-used programming language, was employed for coding the CNN models. Its popularity in the machine learning community, vast libraries, and ease of integration with Tensor Flow made it the preferred language for developing and training the models.

Visual Studio Code (VSCode): The integrated development environment (IDE) used to write, debug, and run Python code was called Visual Studio Code. Its features, including extensions for Python development, facilitated efficient coding and model experimentation.

HTML: HTML can be employed alongside Streamlit components to integrate external content such as images, videos, or interactive elements.

CSS: Cascading Style Sheets helped to customize the appearance of web applications created with Streamlit. CSS enables to control the layout, colors, fonts, and other visual aspects of your application, allowing for greater customization and branding.

Streamlit: Streamlit is an open-source Python library that allowed us to create interactive web applications for machine learning and data science projects with ease. It simplified the process of building and sharing custom web apps by providing a simple and intuitive interface for creating interactive components directly from Python scripts.

5. IMPLEMENTATION AND METHODOLOGY

5.1 Dataset

1) Dataset Collection for Plants:

The original dataset, which is accessible on a GitHub repository, was subjected to offline augmentation techniques to create the new dataset. The original dataset comprises approximately 87,000 RGB images depicting healthy and diseased crop leaves, organized into 38 distinct classes. The New Plant Diseases Dataset is a collection of images portraying various healthy and unhealthy crop leaves. Through offline augmentation, the dataset was expanded to include around 87,000 RGB images, representing diverse classes of healthy and diseased crop leaves, suitable for training and validation purposes. These images are classified into 38 different categories, with the entire dataset split into an 80/20 ratio for training and validation sets, maintaining the original directory structure. Additionally, a new directory containing 33 test images is established for prediction purposes. The format of the dataset makes it suitable for image classification tasks related to plant disease detection and analysis.

2) Dataset Collection For Seeds:

The dataset was obtained from industrial-grade image capturing, adhering to the Soybean Classification Standard (GB1352-2009). An industrial camera was used to take pictures of soybean seeds coming into direct contact with one another. An image-processing method is used in the dataset construction procedure to ensure a segmentation accuracy of more than 98%. The five varieties of soybeans included in the soybean seed dataset are intact, spotted, immature, broken, and skin-damaged soybeans. There are 5513 photos in all in the dataset. There are almost a thousand pictures of soybean seeds in each category. An industrial camera took the original high-resolution (3072 x 2048 pixels) photos. Individual soybean images (227 x 227 pixels) were retrieved with a segmentation accuracy of more than 98% using an image-processing technique. This dataset is useful for researching soybean seed classification or quality evaluation, providing diverse examples of soybean seeds under different conditions.

3) Dataset Collection For Crop And Weed Detection:

The dataset was collected through the following steps: 1) Collection of 589 images of weeds and crops through photo capturing. 2) Cleaning the dataset to ensure the removal of any irrelevant or low-quality images, resulting in 546 images. 3) Image processing to resize the photos from the original size of 4000x3000 color to 512x512x3 for efficient training. 4) Utilization of Data Augmentation techniques to increase the dataset size, resulting in a total of 1300 images. 5) Manual labeling of image data by drawing bounding boxes on photos to distinguish between weeds and crops.

The dataset consists of 1300 color images, each with a resolution of 512x512 pixels. The images contain sesame crops and various types of weeds. Labels for the images are provided in YOLO format, allowing for easy integration with detection models. The dataset is a valuable resource for training models for crop and weed detection tasks, with bounding boxes indicating the location of crops and weeds in each image. The dataset's content and format make it suitable for computer vision tasks related to agricultural applications, particularly in the context of crop and weed detection using object detection techniques.

5.2. Model Training And Optimization

New Plant Diseases Dataset:

Model Training: Employ a Convolutional Neural Network (CNN) architecture for image classification. Consider transferring knowledge with trained models like VGG16 or Res.Net. Fine-tune the model on the New Plant Diseases Dataset, adjusting for the 38 different classes.

Optimization: Normalize data, implement dropout layers to avoid over fitting, and adjust learning rates. Experiment with different CNN architectures and hyper parameters for optimal performance on the validation set. Evaluate and optimize for metrics like accuracy, precision, recall, and F1-score.

Soybean Seed Classification Dataset

Model Training: Design a CNN for multi-class classification of soybean seed types. Use techniques like data augmentation to enhance generalization. Split the dataset into training and validation sets, ensuring a representative mix of intact, immature, broken, and skin-damaged, spotted seeds.

Optimization: Leverage transfer learning if applicable, and fine-tune the model to reduce false positives/negatives. Optimize for precision, recall, and F1-score, considering the importance of correctly classifying each seed type.

Crop and Weed Detection Dataset:

Model Training: Develop an object detection model using YOLO architecture. Train the model on the Crop and Weed Detection Dataset, considering the 1300 images with bounding box annotations for crops and weeds.

Optimization: Fine-tune the model to enhance accuracy and minimize false detections. Experiment with parameters like anchor box configurations, object confidence thresholds, and non-maximum suppression for optimal object localization. Optimize for real-time performance on resource-constrained devices.

6. RESULT AND DISCUSSION

TABLE -2:. Result table

Model	Training Accuracy	Validation Accuracy
Plant Disease	0.982	0.952
Soyabean Seed Detection	0.999	0.876
Weed Detection	0.985	0.963

7. CHALLENGES AND SOLUTION

Limited Data and Model Training To effectively detect diverse plant diseases and weeds, a substantial dataset is crucial. However, obtaining comprehensive datasets for various crops and regions can be challenging. **Collaborative Data Sharing** Encourage collaboration with agricultural research institutions, local farmers, and global agricultural communities to share diverse datasets. **Implementing federated learning** techniques

can enable model training without centralizing sensitive data.

Real-time Image Processing Performing image analysis for disease detection and other parameters in real-time on mobile devices can strain computational resources, leading to slower performance and increased energy consumption. **Edge Computing and Optimization** Implement edge computing techniques to process data on the device, reducing latency. Optimize algorithms and leverage model quantization to ensure efficient processing without compromising accuracy.

Localization of Cultivation Tips Providing accurate and region-specific cultivation tips requires extensive knowledge of local farming practices, climate, and soil conditions. **Geospatial Integration and Local Expertise** Integrate geospatial data to tailor recommendations based on the user's location. Collaborate with local agricultural experts to incorporate region-specific knowledge, ensuring the cultivation tips align with local conditions

8.RESULTS AND IMPLEMENTATION



Fig-3: Output 1-Home Page

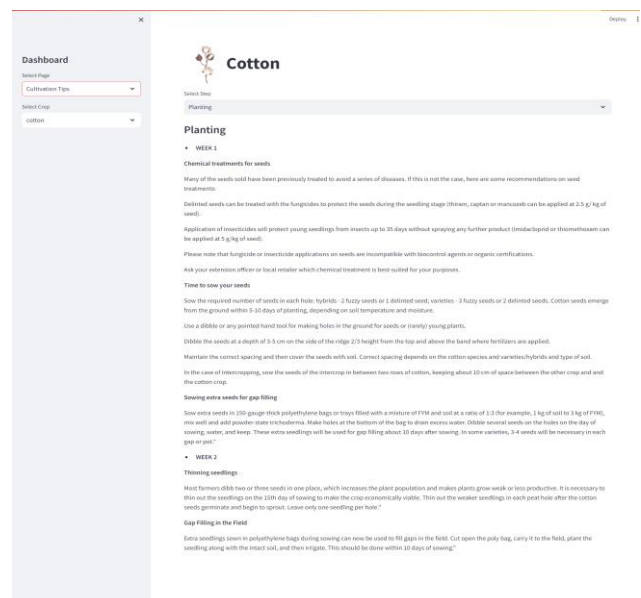


Fig-4: Output 2-Cultivation Tips

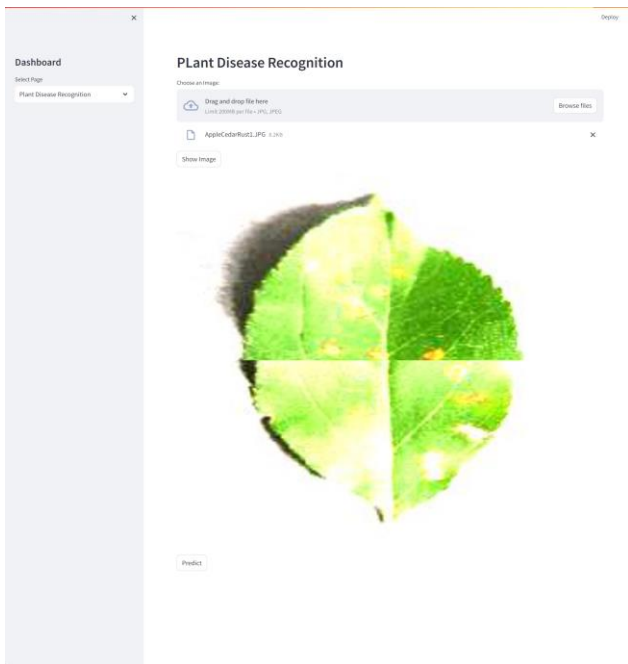


Fig-5: Output 3-Input Plant Image

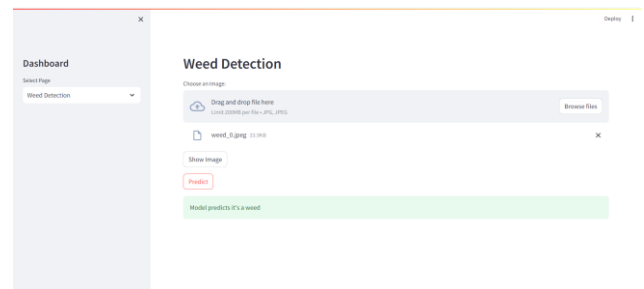


Fig-8: Output 6-Weed Detection

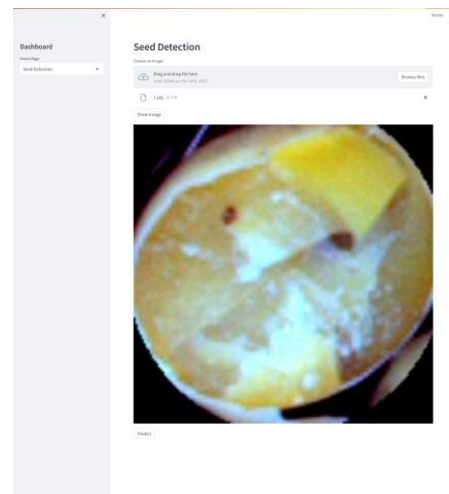


Fig-9: Output 7-Input Seed Image

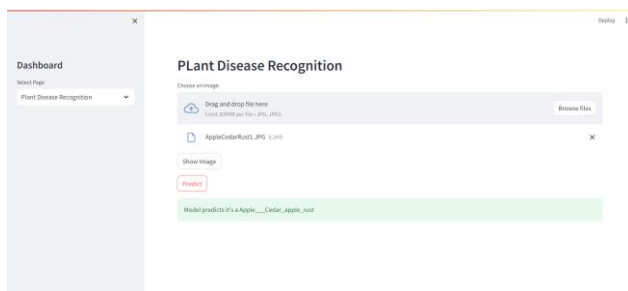


Fig-6: Output 4-Plant Disease Recognition

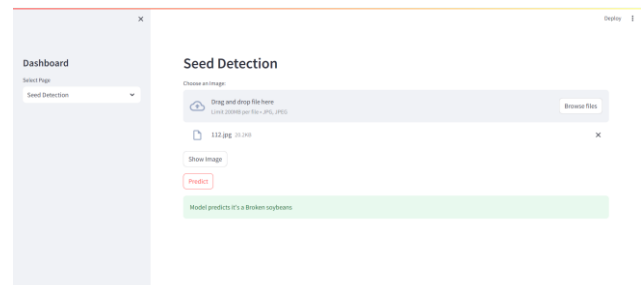


Fig-10: Output 8-Seed Detection

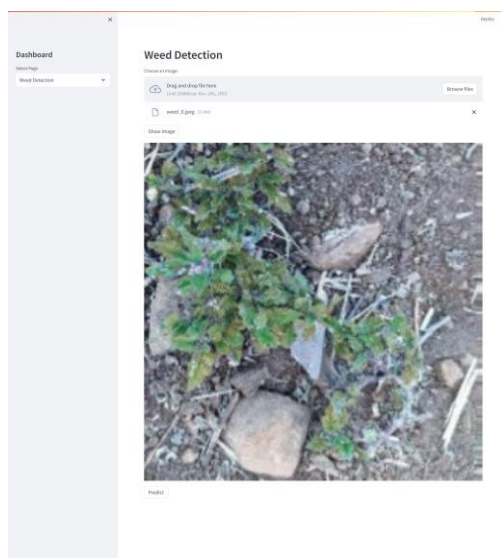


Fig-7: Output 5-Input Weed Image

9.CONCLUSION AND FUTURESCOPE

9.1 Conclusion

In conclusion the full web app "CROP WATCH - Empowering Precision Farming through ML and Image Analysis" is the result of our initiative, which aims to transform the agricultural sector. It gives farmers the necessary tools for better crop management with an emphasis on weed identification, disease detection, quality assessment, and seed recognition. Our weather forecasting feature also helps with farming decision-making. Users can share expertise by using the community platform, pesticide advice, and horticulture suggestions. The initiative showcases the transformative capacity of machine learning and image analysis in

enhancing agriculture and promoting sustainable farming practices.

9.2 Future Scope

As we look towards the future, there are exciting avenues for further enhancement. Future work will focus on expanding the app's dataset with a broader crop spectrum, geographic diversity, and real-world environmental conditions. Continuous refinement of the machine learning model will be pursued to enhance accuracy, with a keen emphasis on incorporating more advanced techniques such as transfer learning. Integration with real-time weather data will be explored for a more holistic understanding of plant health. Additionally, the app's features will be extended to provide comprehensive cultivation tips for detected crops, fostering a more proactive approach to precision farming. Looking ahead, future iterations of "CROP WATCH" will focus on enhancing user accessibility by incorporating support for multiple regional languages within the app. This inclusivity aims to bridge language barriers, ensuring farmers from diverse regions can seamlessly comprehend and engage with the application.

REFERENCES

- [1] Kowshik B ,Savitha V, Nimosh madhav M, Karpagam G, Sangeetha K, "Plant Disease Detection Using Deep Learning".
- [2] Ms. Nilam Bhise, Ms. Shreya Kathet, Mast. Sagar Jaiswar , Prof. Amarja Adgaonkar, "Plant Disease Detection using Machine Learning",vol. 7 , 2020.
- [3] D.Jayanarayana Reddy, M. Rudra Kumar, "Crop Yield Prediction using Machine Learning Algorithm".
- [4] Emmanuel Moupojou, Appolinaire Tagne,Florent Retraint,Dongmo Wilfried, Hyppolite Tapamo, Marcellin Nkenlifack, "FieldPlant : A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning".
- [5] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed ".
- [6] A. . Fuentes, S. Yoon, S. C. Kim, and D. S. Park , "A robust deeplearning-based detector for real-time tomato plant diseases and pests' recognition,".
- [7] J. Sun, L, Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model" Sensors, vol. 19, no. 20, pp. 4363, 2019.
- [8] Qiong Qiao , "Image Processing Technology Based on Machine Learning".
- [9] Sumit Kumar Chaudhary , Prashant K. Srivastava , Dileep Kumar Gupta , Pradeep Kumar , Rajendra Prasad , Dharmendra Kumar Pandey , Anup Kumar Das , Manika Gupta, "Machine learning algorithms for soil moisture estimation using Sentinel-1: Model development and implementation".
- [10] Prachi Juyal,Amit Kundaliya, "Multilabel Image Classification using the CNN and DC-CNN Model on Pascal VOC 2012 Dataset"