Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)

VOLUME:7, ISSUE:2, PAPER NUMBER:51
:3

Intelligent Framework for Crop Data Analytics and Crop Recommender System

Pittala Venkata Dheeraj Reddy[1], Ashwin Raj Pillai[2], Chirala Ashrith[3], Raju Krishnam Mantena[4], B S Anil Kumar[5], Dr Basant Sah[6]

[1,2,3,4] Students, and [5,6]Faculty
Dept. of Computer Science Engineering,
Gokaraju Rangaraju Institute of Engineering
and Technology Hyderabad, India.
ashwinraj1216@gmail.com, anilme87@gmail.com, pvdheerajreddy@gmail.com, ashrithchirala@gmail.com, krishnu2002@gmail.com, basantbitmtech2008@kluniversity.in

Abstract: *Intelligent Framework for Crop Data Analytics [CDA] and Crop Recommender System [CRS] is a Web Application with an integration of different Machine Learning Algorithms and Recommendation Systems. [IDFCDACRS] aims to provide various solutions to the problems faced by Agriculture sector. The Web application is offers various solutions including Crop Disease Detection, Crop Detection and the seasonal analytics. Approaches like Regression techniques are used for Recommendation solutions and Image Pre-processing steps like Contour detection, Normalization along with CNN model are used for Detection solutions. This framework aims to minimize the time for crop diagnosis and crop recommendations for the following seasons using tools from OpenCV, Open AI.*
Keywords: *IDCDACRS, CRS, OpenCV, Open AI, Normalization, Binarization.*

1. INTRODUCTION:

Intelligent Framework for Crop Data Analytics [1] and Crop Recommender System [2] is Flask based Web Application with integration of Machine Learning Algorithms for Detection Systems and Recommender Systems. Users will be provided with Seasonal insights of Crops and the Gross revenue generated with Authentication.

With the recent developments in Flask, there have been multiple web applications being built on the Flask framework. This system is an integration of Machine Learning algorithms and Flask application. Features like User Authentication, [CDA], [CRS] are built using OpenCV, Open AI libraries of python.

2. LITERATURE SURVEY

Data Mining is often used to extract information from data. In the field of Agriculture, there are many applications using Data Mining. A study proves that using parameters like soil characteristics and their types along with percentage of chemicals can be used as training data to build a Crop Recommendation System. Ensemble model is developed with algorithms like KNN, Naïve Bayes, and Random Tree to provide recommendations with higher efficiency [3].

Another citation created a Machine Learning model for Crop recommendation using ensemble techniques with SVM and ANN as learners. The model is trained on dataset consisting of various soil parameters from Polytest Laboratories soil Testing Lab [4]. Various Crops like Banana, Paddy, Cotton, Groundnut were included in the dataset. Attributes like Ph, Soil Color, Texture, were used to build a recommender model.

In the agricultural landscape, the pervasive threat of plant diseases caused by pests, insects, and pathogens poses a substantial risk to crop productivity. A Publication [5] discussed a proposed system addressing the challenge of monitoring extensive cultivated areas, often spanning acres, where manual surveillance becomes impractical. The system employs remote sensing images for continuous monitoring and incorporates automated disease detection mechanisms to promptly alert agriculturists. The primary goal is the early detection of diseases as they emerge on the outer layers of leaves. The system operates in two phases: first, training data sets with both healthy and diseased samples, and second, real-time monitoring and disease identification using Canny's edge detection algorithm [6]. This comprehensive approach aims to revolutionize disease management, providing cultivators with a proactive means of preserving crop health and mitigating losses due to uncontrolled diseases.

The Crop disease detection using image segmentation Tushar H Jaware, Ravindra D Badgujar and Prashant G Patil tells us how K-Means clustering has been applied to solve low-level image segmentation tasks. This clustering algorithm is convergent and its aim is to optimize the partitioning decisions based on a user-defined initial set of clusters that is updated after each iteration. In the first step we identify the mostly green colored pixels. Next, these pixels are masked based on specific threshold values that are computed using Otsu's method, then those mostly green pixels are masked. The other additional step is that the pixels with zeros red, green and blue values and the pixels on the boundaries of the infected cluster (object) were completely removed. The experimental results demonstrate that the proposed technique is a robust technique for the detection of plant leaves diseases.

Data analytics for crop management by Nabila Chergui & Mohand Tahar Kechadi tells us how recent advances in Information and Communication Technologies have a significant impact on all sectors of the economy worldwide. Digital Agriculture appeared as a consequence of the democratisation of digital devices and advances in artificial intelligence and data science. Digital agriculture created new processes for making farming more productive and efficient while respecting the environment. Recent and sophisticated digital devices and data science allowed the collection and analysis of vast amounts of agricultural datasets to help farmers, agronomists, and professionals understand better farming tasks and make better decisions. In this paper, we present a systematic review of the application of data mining techniques to digital agriculture. We introduce the crop yield management process and its components while limiting this study to crop yield and monitoring. After identifying the main categories of data mining techniques for crop yield monitoring, we discuss a panoply of existing works on the use of data analytics. This is followed by a general analysis and discussion on the impact of big data on agriculture.

[7] In a literature survey we found spectral and imaging technology provide acumen about crop health and growth dynamics. This technology has been activated to various crops like soybean, rice, wheat for detecting chlorophyll, nitrogen content, to monitor crop pests and diseases, diagnosing stress factors and predicting crop yield. Data for chlorophyll content detection is acquired using aircraft-based, satellite-based sensing.

While referring to [8] we came across the importance of data mining in agriculture which includes crop yield analysis and prediction. This includes Chameleon clustering and regression for prediction. This is mainly to maximize rice yield production and how different factors play a role in the production. Technologies such as big data is also discussed in this paper.

A research paper showed [9] how machine learning models predicted production of rice in specified areas of Maharashtra. The paper mainly focuses on the application of sequential minimal optimization classifier. The dataset contains data of 27 districts of that state for the season between the months June and November. The parameter which are considered for performing analysis include evapotranspiration, precipitation, temperature etc. The performance of SMO is evaluated using MAE, RAE metrics.

Another survey [10] showed a well organised system for prediction. There are several classification techniques like SVM, Naïve Bayes. This works on adaptive cluster approach. The paper explains the significance of such methodologies in decision making. It compares its performance with other classifiers using the neural networks.

While exploring for machine learning algorithms we came across [11] which mainly focuses on using machine learning algorithms such as Support Vector Machine and Random Forest to analyse agriculture data and recommend suitable fertilizer for particular crops. The main aim was to create a prediction model for future crop yield prediction and it also about discusses the use of different machine learning techniques for estimating future crop production and emphasizes the importance of yield prediction in agriculture. The proposed system involves predicting crop yield from previous data and recommending fertilizer for specific crops. The paper also touches on the impact of modern agricultural practices on seasonal climatic conditions and food security. The research aims to improve crop yield and recommends suitable fertilizer for different crops based on available data.

In the survey for a system for providing pesticide recommendations and amounts for unhealthy crops based on images captured by the farmers we found that [12] the process involves capturing the crop image using an android application or webpage or processing the image and extracting its features using CNN for classification, and retrieving and presenting the pesticide information for the farmer. The system aims to increase the efficiency of crops by recommending the relevant use of organic fertilizers to minimize production loss. A modified CNN architecture is proposed for crop disease detection, achieving an accuracy rate of 89%. The paper concludes by highlighting the potential for future improvements, such as considering other parts of the crop for disease detection.

In a research paper [13] we found a technique using K-Means clustering for low-level image segmentation to detect plant leaf diseases. The K-Means clustering algorithm was applied to optimize partitioning decisions based on initially defined clusters. The paper outlines the process of identifying and masking mostly green-colored pixels using specific thresholds computed with Otsu's method. The technique includes additional steps to remove pixels with zero RGB values and those on the boundaries of the infected cluster. Experimental results show how high are the accuracy rates for the proposed technique for detecting plant leaf diseases. It also discusses the importance of using technology to support the management of perennial fruit crops. Additionally, it presents the process of image segmentation and recognition and the computation of texture features for disease classification. The proposed technique shows promise for efficient disease detection and classification.

Another research [14] which shows the potential of big data and machine learning for crop protection, particularly focusing on weed control and management. The existing research uses big data to predict herbicide resistance, detect invasive weeds, and develop decision support systems. Machine learning plays a very crucial role in employing techniques like supervised and unsupervised learning for weed classification, resistance prediction, and pattern identification used in the large datasets for the experiment. Recognizing the spatial spread of weeds, researchers utilize spatial modelling techniques like geostatistics. This paper also introduces the application of Markov random fields (MRFs) to model herbicide resistance in ryegrass, considering the spatial context. The authors compared their MRF model with other approaches, they also demonstrated its effectiveness, paving the way for further exploration in the domain.

As per a survey [15] it shows that Random Forest is an effective and adaptable machine-learning method for crop yield predictions at regional and global scales for its high accuracy and precision, ease of use, and utility in data analysis.C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. For maximizing the crop productivity various data mining techniques were used to predict the crop yield. The causes of variability of crop growth in an agricultural field might be due to crop stress, irrigation practices, incidence of pest and disease etc. The proposed approach mainly shows us how Random Forest algorithm is applied and the result is calculated Based on the climatic input parameters the present study provides the demonstration of the potential use of data mining techniques in predicting the crop yield.

A survey for big data analytics architecture that uses machine learning to predict crop yield explains us about the limitations of traditional methods, they also propose us a system that analyses various data sources like weather, soil, and market trends to guide agricultural decisions. The architecture itself involves data collection, storage, analysis, and visualization, utilizing machine learning algorithms like Incremental K-Means clustering to identify patterns and make predictions. The main aim is to recommend suitable crops based on specific seasonal conditions and geographical factors, aiming to increase yield, improve quality, and reducing the costs. However, there were many challenges faced in data collection, processing complexity, and predicting the accuracy based on data quality and quantity. While it is not suitable for all crops or conditions, the paper [16] offers a promising approach to utilize big data and machine learning for improved agricultural practices and food security.

A survey which shows how deep learning shows promising results for tackling corn diseases and helps us know the major cause of crop loss in India. The paper proposes a system using a pre-trained deep learning model on a Raspberry Pi to identify corn diseases directly from leaf images. The model utilizes transfer learning for better accuracy with limited data and is optimized for real-time use on devices like smartphones and drones. Initial tests with diverse images show promising results, further the work aims to expand the disease library and explore even lighter models for wider device compatibility. This research [17] mainly paves the way for real-time, on-device corn disease detection, potentially improving crop management and minimizing losses.

A new study [18] proposes a streamlined CNN model for tackling tomato crop diseases. Unlike bulky pre-trained models, this lightweight version needs less storage and processing power, making it ideal for resource-constrained settings. It was trained on the PlantVillage dataset which was used for the experiment, it effectively classifies nine diseases despite uneven sample distribution across classes. This research paves the way for efficient and accessible plant disease detection, mainly benefiting resource-limited agricultural operations.

A study [19] shows how demand for food grows alongside climate challenges, it also proposes a smart solution for optimal crop selection and placement on farms. By utilizing sensors and data analysis, the system monitors environmental factors like temperature and humidity, then applies 3D clustering to understand their relationships. This allows it to recommend suitable crops based on the farm's unique conditions and even predict future suitability based on weather forecasts. The real-world test results confirm the system's feasibility, helps the farmers in making informed decisions, maximizing yields, and adapting to changing environments. While further research can explore its effectiveness across diverse crops and refine recommendations, this approach holds promise for smarter and more reliable agricultural practices.

A study [20] about battling crop diseases proposes a versatile method utilizing modified CNN models like ResNet-152 and Inception-v3. The study achieved a high accuracy (97.48% - 99.10%) on diverse datasets, it effectively detects diseases in both rice and corn, demonstrating its adaptability across multiple crops. Further analysis shows rice diseases through major and minor subsets, it also showcases its ability to handle diverse disease data types. This robust method holds a promising approach for improving crop yields, reducing losses, and empowering farmers with efficient disease detection, paving the way for smarter and futuristic agricultural practices.

3. METHODOLOGY:

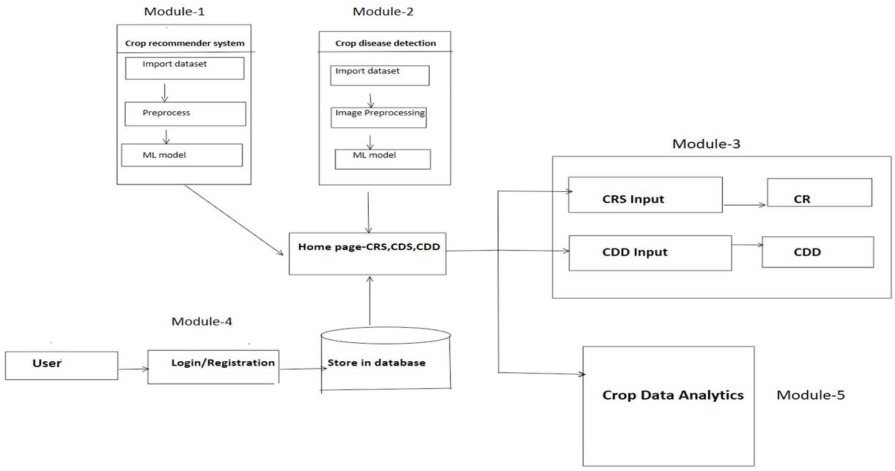
The "Intelligent Framework for Crop Data Analytics and Recommender System" project represents a forward-looking initiative in the agricultural domain. By recognizing and confronting the challenges inherent in traditional farming practices, the project aims to usher in a new era of efficiency and sustainability. Through the strategic integration of data analytics and artificial intelligence, the project seeks to empower farmers and agricultural stakeholders with the tools they need to make intelligent, data-driven decisions.

At its core, the project focuses on aggregating and processing a diverse set of data sources critical to effective agricultural decision-making. By harnessing information such as weather patterns, soil quality, satellite imagery, and historical crop performance, the framework aims to provide a comprehensive view of the agricultural landscape. This holistic approach enables farmers to move beyond reliance on anecdotal evidence and intuition, fostering a more accurate and nuanced understanding of their crops' needs. Machine learning algorithms play a vital role in the project's strategy, contributing to the development of predictive models. These models leverage historical and real-time data to forecast crop yields, predict disease outbreaks, and identify potential pest infestations. By offering foresight into these factors, the project equips farmers with the ability to proactively manage their crops, optimize resource allocation, and mitigate risks associated with unpredictable agricultural conditions.

One of the project's key innovations is the integration of an intelligent recommender system. By tailoring recommendations to individual farmers based on their specific crops, soil conditions, and climate, the system ensures that the advice provided is not generic but rather personalized and actionable. This feature aims to bridge the gap between generic agricultural guidelines and the nuanced requirements of individual farming scenarios, fostering a more effective and sustainable approach to crop management.

In summary, the "Intelligent Framework for Crop Data Analytics and Recommender System" aspires to revolutionize agriculture by embracing cutting-edge technologies. Through the amalgamation of data analytics, artificial intelligence, and predictive modelling, the project strives to empower farmers with insights that go beyond traditional practices, enabling them to navigate the complexities of modern agriculture with precision and sustainability.

Fig. 1. Architecture diagram for certificate automation



4. MODULES:

A. MODULE – I

In the development of the crop recommendation module, the first crucial step involves importing the dataset that serves as the foundation for model training and recommendation generation. Leveraging the pandas library in Python, the module reads the dataset, assuming it contains relevant agricultural features such as soil type, climate conditions, and historical crop performance. Following the dataset import, the scikit-learn library is employed to split the data into training and testing sets. A Random Forest Classifier, chosen for its versatility in handling classification tasks, is then initialized and trained on the training set. The model's accuracy is evaluated using the test set, providing insights into its performance. Once the model is trained, the module encapsulates a recommendation function. This function takes a new set of agricultural features as input and employs the trained model to predict the most suitable crop for cultivation. The module thus encapsulates the entire process, from dataset importation and model training to the generation of crop recommendations, offering a holistic solution for aiding farmers in making informed decisions about crop selection based on data driven insights.

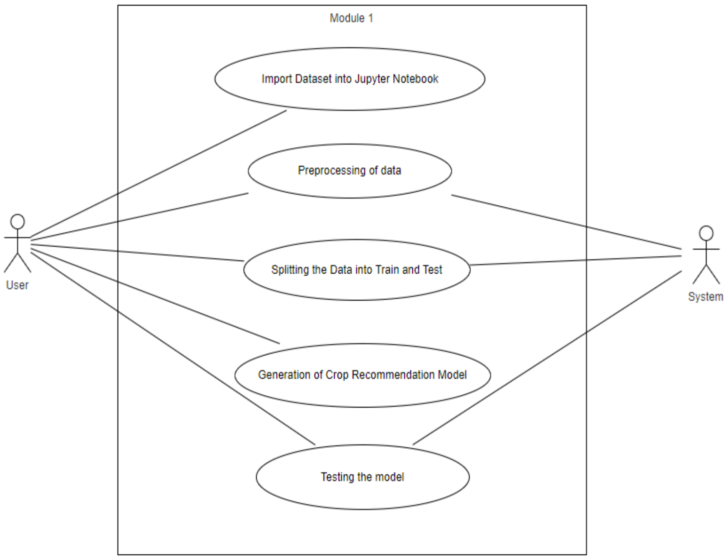


Fig. 2. Use case Diagram for Module 1

B. MODULE - II

Incorporating a crop disease images dataset into our project commences by securing data from reputable sources like agricultural research repositories. After acquisition, we categorize the dataset into folders, each representing a distinct plant disease class. Utilizing TensorFlow or PyTorch, we efficiently load images for seamless integration. Subsequently, a comprehensive pre-processing phase ensues, involving resizing images to a standardized dimension for essential uniformity in deep learning models. Normalization techniques ensure consistent pixel values, promoting model convergence during training. Augmentation strategies, such as rotation and flipping, expand the dataset for improved model generalization. To counter overfitting, the dataset is judiciously split into training, validation, and testing sets, maintaining balanced class representation. These meticulous pre-processing steps establish a well-organized and standardized plant disease image dataset, perfectly poised for effective training and evaluation of our machine learning models for precise disease detection.

To build a Crop disease detection model, we utilize Convolutional Neural Networks (CNNs). After acquiring a carefully curated plant disease image dataset, we employ popular deep learning frameworks like TensorFlow or PyTorch. The CNN architecture is chosen for its ability to automatically learn relevant features from images. Through successive convolutional and pooling layers, the model captures hierarchical representations, enabling accurate predictions.

HTML pages for the four different routes are generated. We add event handlers for buttons in home page, recommend page and detection page. The previously defined module functions are now connected with the frontend by rendering the HTML pages. CSS styling, Bootstrap styling are done to the HTML pages.

The saved model files in .pkl file format are imported into the app.py file in Backend repository and the model object is initiated in the event handler functions. When the input is taken from forms in the Frontend it is communicated back to Backend and the corresponding output is transmitted to the event handler functions.

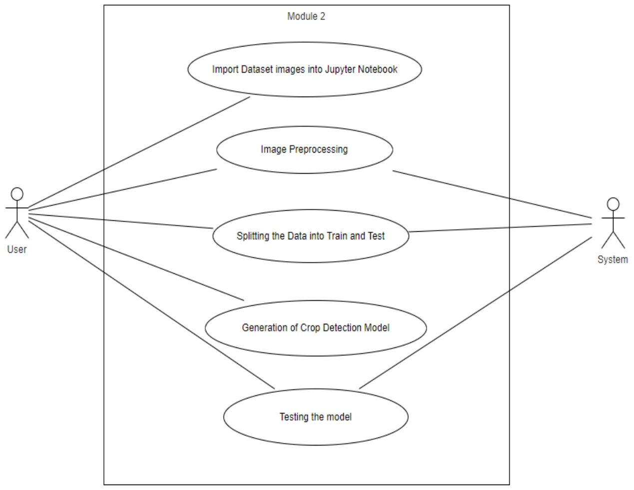


Figure 1:Use case Diagram for Module-2

C. MODULE-III

In this module, there are 3 main phases:

1. Creating Backend repository
2. Creating Frontend repository
3. Integrating ML models with the backend.

We begin this module by installing Flask package using pip commands. Two repositories for backend and frontend were created. Now create a file named app.py. Import the necessary flask libraries into the app.py file. Different routes for the website - “/”, “/home”, “/recommend”, “/detect” and their module functions are added. Module functions for each route are defined.

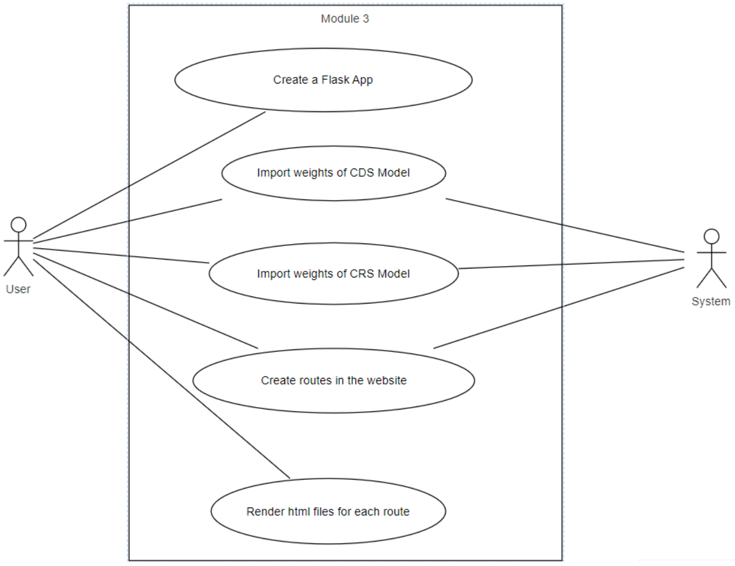


Figure 2:Use case Diagram for Module-3

5. PICTORIAL REPRESENTATION:

Fig. 3. Homepage for [IFCDACRS]

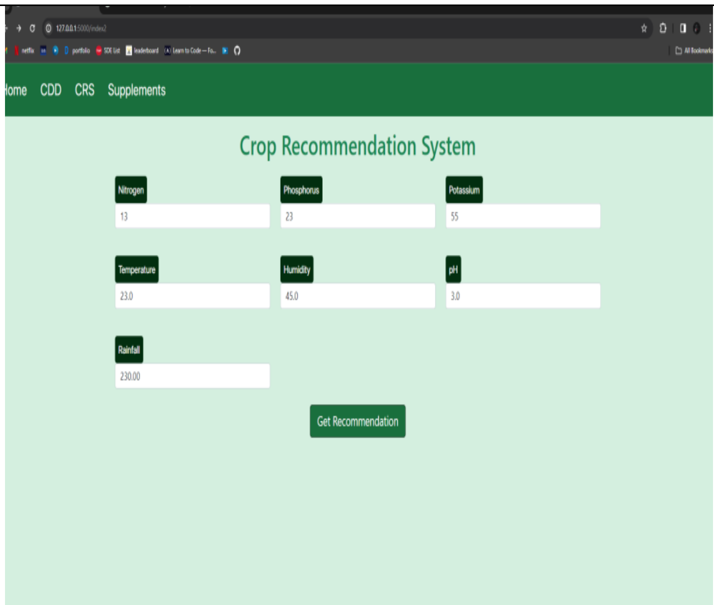


Fig. 4. Frontend for [CRS]

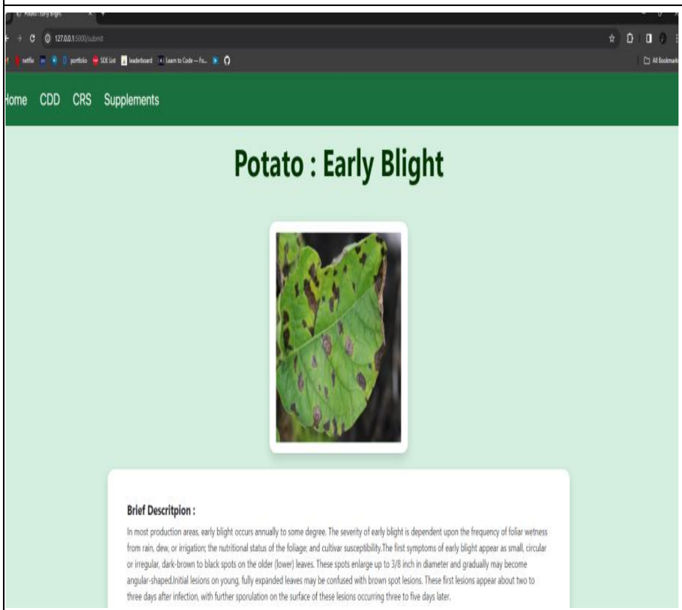


Fig. 6. Frontend for [CDS]

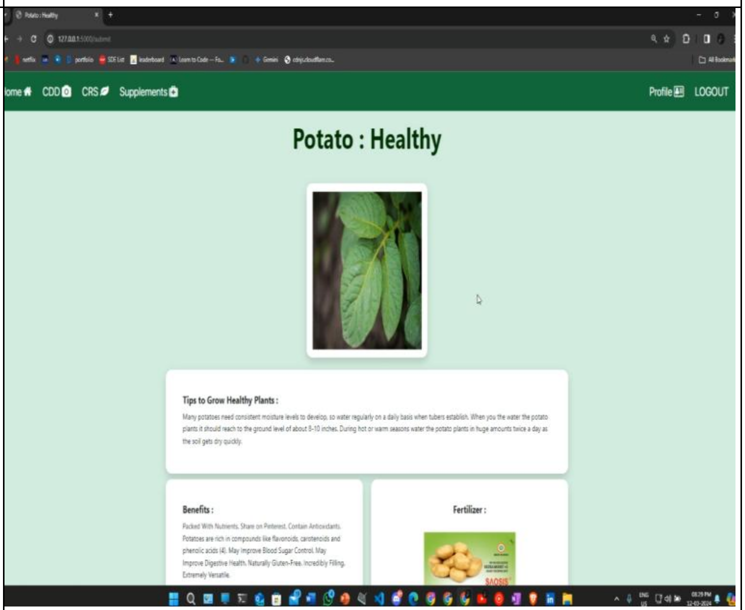


Fig.5. CRS Output

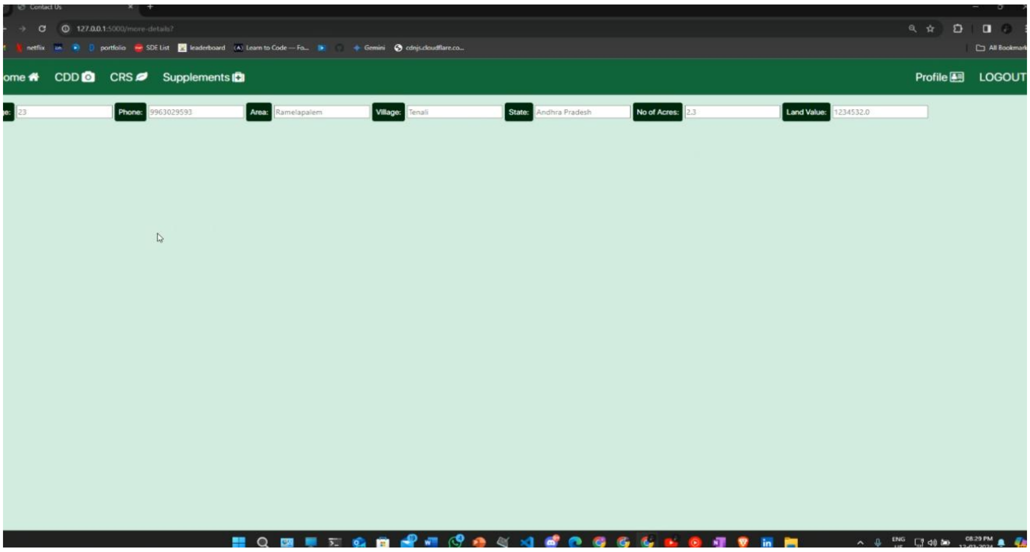


Fig.7. Crop Data Analytics Page

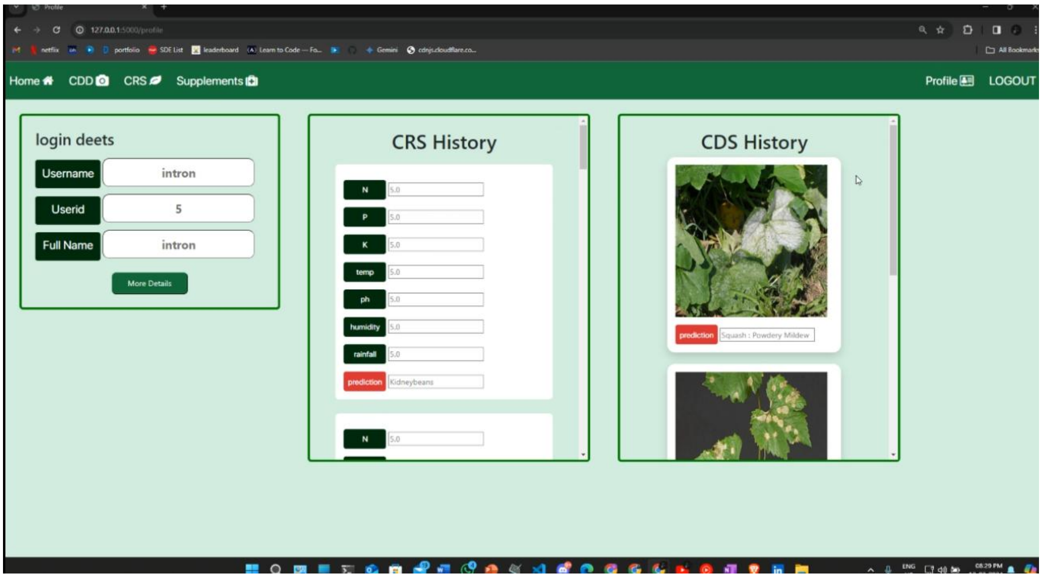


Fig.8. Crop Data Analytics Output Page

6. RESULTS:

The following table demonstrates various testing scenarios of Module-I [CRS] against different parameters for the model which are the soil conditions. The accuracy of [CRS] is 99.08%

TABLE I. TESTING CROP RECOMMENDATION SYSTEM

Testcase Id	Testing [CRS]			
	Test Scenario	*Testcase Number*	*Result*	*Status*
1	N=50 P=34 K=23 Temp=21.1 Humidity=31.0 Rainfall=10.0	CASE-1	The predicted output is Kidney beans	Pass
2	N=20 P=44 K=53 Temp=41.1 Humidity=11.0 Rainfall=103.0	CASE-2	The predicted output is Muskmel on	Pass

Testcase Id	Testing [CRS]			
	Test Scenario	***Testcase Number***	***Result***	***Status***
3	N=20 P=34 K=53 Temp=41.1 Humidity=51.0 Rainfall=120.0	CASE-3	The predicted output is Orange	Pass
4	N=85 P=34 K=78 Temp=10.1 Humidity=50.0 Rainfall=76.0	CASE-4	The predicted output is Apple	Pass

TABLE II. TESTING CROP DISEASE DETECTION SYSTEM

Testcase Id	Testing [CDS]			
	Test Scenario	***Testcase Number***	***Result***	***Status***
1	Input image is cornhealthy.jpg	CASE-1	The predicted output is Corn Healthy	Pass
2	Input image is corncommonrust.JPG	CASE-2	The predicted output is Common Rust	Pass
3	Input image is cherrypowdery mildew.JPG	CASE-3	The predicted output is Cherry Powdery Mildew	Pass
4	Input image is corncorncercosporaleaf.JPG	CASE-4	The predicted output corn Cerpora leaf	Pass
5	Input image is tomatobacterial. JPG	CASE-5	The predicted output is Cherry Powdery Mildew	Fail

The above table demonstrates various testing scenarios of the [CDS] system. Different input images of various crops are passed as inputs to the [CDS] machine learning model. The accuracy of the model is 89%.

7. CONCLUSION AND FUTURE SCOPE:

"Intelligent Framework for Crop Data Analytics and Recommender System"[IFCDACRS] project presents a transformative solution to the challenges facing modern agriculture. By harnessing the power of Full Stack Web Development and Machine Learning, this framework offers accurate predictions and personalized recommendations to farmers. The anticipated outcomes of this project ensure increased crop yields, resource optimization and adaptability to changing climate conditions. The user-friendly interface of the Centralized System empowers clients with accessible tools for decision-making, contributing to food environmental preservation and agricultural innovation. As this project bridges the gap between technology and agriculture, it holds the potential to reshape farming practices and enhance the well-being of farming communities on a global scale.

8. References:

[1] T. G. Chergui, Nabila, and Mohand Tahar Kechadi. "Data analytics for crop management: a big data view." *Journal of Big Data* 9.1 (2022): 1-37.

[2] Pande, Shilpa Mangesh, et al. "Crop recommender system using machine learning approach." *2021 5th international conference on computing methodologies and communication (ICCMC)*. IEEE, 2021.

[3] Pudumalar, S., et al. "Crop recommendation system for precision agriculture." *2016 Eighth International Conference on Advanced Computing (ICoAC)*. IEEE, 2017.

[4] Rajak, Rohit Kumar, et al. "Crop recommendation system to maximize crop yield using machine learning technique." *International Research Journal of Engineering and Technology* 4.12 (2017): 950-953.

[5] Badage, Anuradha. "Crop disease detection using machine learning: Indian agriculture." Int. Res. J. Eng. Technol 5.9 (2018): 866-869.

[6] Zhou, Ping, et al. "An improved canny algorithm for edge detection." Journal of Computational Information Systems 7.5 (2011): 1516-1523.

[7] He, Yong, et al. "Critical review of fast detection of crop nutrient and physiological information with spectral and imaging technology." *Transactions of the Chinese Society of Agricultural Engineering* 31.3 (2015): 174-189.

[8] Sneha, N., and J. Majumdar. "Big data application in agriculture to maximize the rice yield crop production using data mining techniques." *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)* 5.5 (2017): 2017.

[9] Gandhi, Niketa, et al. "Rice crop yield prediction in India using support vector machines." *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016.

[10] Bhangale, P. P., P. Y. S. Patil, and D. D. Patil. "Improved crop yield prediction using neural network." *IJARIIE* 3.2 (2017): 2395-4396.

[11] Bondre, Devdatta A., and Santosh Mahagaonkar. "Prediction of crop yield and fertilizer recommendation using machine learning algorithms." *International Journal of Engineering Applied Sciences and Technology* 4.5 (2019): 371-376.

[12] Patil, Nikhil, et al. "Crop disease detection using deep convolutional neural networks." *International Journal of Engineering Research & Technology (IJERT)* 8.03 (2019): 2278-0181.

[13] Jaware, Tushar H., Ravindra D. Badgujar, and Prashant G. Patil. "Crop disease detection using image segmentation." *World Journal of Science and Technology* 2.4 (2012): 190-194.

[14] Rejmánek, Marcel, George R. Robinson, and Eliska Rejmankova. "Weed-crop competition: experimental designs and models for data analysis." *Weed Science* 37.2 (1989): 276-284.

[15] Champaneri, Mayank, et al. "Crop yield prediction using machine learning." *Technology* 9.38 (2016).

[16] Bhavanandam, Srideviidev. "Wcp: Weather-based crop yield prediction using machine learning and big data analytics." *technology* 3 (2022): 4.

[17] Mishra, Sumita, Rishabh Sachan, and Diksha Rajpal. "Deep convolutional neural network based detection system for real-time corn plant disease recognition." *Procedia Computer Science* 167 (2020): 2003-2010.

[18] Agarwal, Mohit, Suneet Kr Gupta, and K. K. Biswas. "Development of Efficient CNN model for Tomato crop disease identification." *Sustainable Computing: Informatics and Systems* 28 (2020): 100407.

[19] Tseng, Fan-Hsun, Hsin-Hung Cho, and Hsin-Te Wu. "Applying big data for intelligent agriculture-based crop selection analysis." *IEEE Access* 7 (2019): 116965-116974.

[20] Saeed, Zubair, et al. "A multi-crop disease detection and classification approach using cnn." *2021 International Conference on Robotics and Automation in Industry (ICRAI)*. IEEE, 2021.