IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 04, April 2025 https://www.ijirae.com/archives

A Machine Learning Approach to Crop Recommendation, Plant Disease Identification and Yield Estimation

DrShilpa G V
Assoc professor, Computer Science and Engineering
Vemana Institute of Technology, Bengaluru, India
shilpa.gvs@gmail.com
Chethan S, Jai Kumar B, B Vishwas Reddy, KushalaNayaka
Computer Science and Engineering
Vemana Institute of Technology, Bengaluru, India
chethans20032007@gmail.com,,jaikumarb031@gmail.com,blyvishwasreddy@gmail.com,kushalnayaka875@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.12/Issue04/APAE10089, Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.12/Issue04/APAE10089 Received: 04, April 2025, Revised: 12, April 2025, Accepted: 22, April 2025, Published Online: 06, May 2025. https://www.ijirae.com/volumes/Vol12/iss-04/08.APAE10089.pdf
Article Citation: Dr.Shilpa,Chethan,Jai,Vishwas,Kushala(2025),A Machine Learning Approach to Crop Recommendation, Plant Disease Identification, and Yield Estimation. IJIRAE:: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 03 of 2025 pages 182-193 **Doi:>**https://doi.org/10.26562/ijirae.2025.v1204.08
BibTeX Key:**Dr.Shilpa@2025Machine**

Abstract: The reduction in farm output can be linked to three major causes, i.e., plant diseases, improper crop choices, and wrong yield estimates. The remedy calls for an integrated artificial system that merges disease identification, crop suggestions, and yield estimation functionalities. The proposed system proves to be highly effective in identifying infected plant types and disease types by employing early disease detection with convolutional neural networks (cnns) on the plant village dataset. The crop suggestion model employs Random Forest (RF) as the primary algorithm, prioritizing its performance optimization over methods like k-nearest neighbors (KNN), logistic regression, and decision trees. The crop recommendation helps the farmers to decide on proper crops by looking into their nitrogen, phosphorus, potassium contents in the soil, pH level, and water content. The system utilizes a decision tree model to forecast crop yield, drawing on historical data related to agriculture and weather conditions. The mode enables accurate estimation of yield, which can help the farmer plan effectively and optimize his use of resources. All components and evaluation algorithms exhibited exceptional accuracy in their respective functions during the testing phase. The integrated software system integrates three functions that are critical in agriculture into a single product, offering farmers real-time worth information and improving farm efficiency while maintaining sustainability.

Keywords: Disease Detection, Crop prediction, Yield Prediction, Convolutional neural network (CNN), Machine Learning, Decision Tree Algorithm.

I. INTRODUCTION

Indian agriculture plays a vital role in the nation's economy and holds the distinction of being the world's second-largest producer in the agricultural sector. Indian farming activities have to cope with complex agricultural ecosystems because of the mixture of varied soil, environmental, and climatic conditions. The earlier methods of detection of plant diseases were through visual inspection and laboratory analysis, which were laborious and physically stressful. Additionally, the data obtained carried a significant risk of being misinterpreted. Relying on outdated detection methods leads to incorrect pesticide choices, contributing to pathogen resistance and causing significant crop losses. The accuracy of human inspections significantly decreases when relying on mobile images. Connexins provide accurate results and rapid processing, aiding physicians in the early diagnosis of diseases and the prevention of potential hazards. Crop productivity is highly dependent on the soil quality. Proper soil testing gives beneficial results for crop selection in accordance with the particular land condition. This yields nutritionally rich crops. Phosphorus, nitrogen, pH, and moisture soil parameters can be precisely ascertained by the application of support vector classifiers (svc) machine learning models, which give more accurate crop suggestions. This method improves agricultural planning efficiency and promotes increased production. Effective resource management is strongly dependent on accurate yield predictions, which require dependable data for reliable forecasting. Conventional farming practices are not precise, but machine learning algorithms provide better solutions. The system employs a decision tree-based algorithm to estimate crop yields using historical crop data, soil characteristics, and weather conditions. The transparent outcomes provided by decision trees enable farmers to anticipate production levels in advance, simplifying their decision-making processes. The integrated system that integrates disease detection and crop optimization, with yield calculation, helps farmers maximize sustainability and productivity without any limitation.

II. LITERATURE REVIEW

Over the past decade, machine learning (ML) and deep learning (DL) techniques have made significant progress in identifying plant diseases and predicting crop yields.

Utilizing these models to analyze agricultural data patterns has led to significant progress in improving farm productivity, optimizing crop management, and fostering sustainable agricultural practices. Among the various machine learning models used for detecting plant diseases, Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN) are recognized as the most effective. Statistical methods employing mathematical functions and pattern recognition technologies analyze images of leaves, stems, and roots to detect visible signs of plant health issues.

Jeevaganesh et al., [1] introduced an extensive methodology for identifying plant diseases and making crop selection decisions. The study highlights how support vector machines, random forests, and naive Bayes algorithms perform well in crop yield prediction using structured data but struggle with unstructured data. Deep learning methods significantly improve the accuracy of soil-based disease detection and crop recommendations, with CNNs and RNNs being the preferred analytical techniques. Combining ML and DL models in a hybrid system greatly enhances overall system performance. The goal of this study is to advance sustainable agricultural modeling systems, enhancing their predictive accuracy while improving the efficiency of implementing agricultural practices.

Patki V et al., [2] explores the limitations of traditional crop forecasting methods, proposing a solution that integrates data from farmers and experienced growers. Neural networks, enhanced by environmental factors like temperature, rainfall, and soil type, show improved accuracy but face challenges due to gaps in time-series data. The study introduces a unique data collection approach that focuses on specific time intervals to enhance model accuracy. By considering the effects of weather on crop growth and utilizing more comprehensive data, the authors aim to improve artificial neural networks for more precise yield predictions.

Shahrin et al., [3] uses machine learning algorithms to interpret satellite image data to investigate agricultural production patterns and accurately make harvest predictions. The researchers used k-means clustering combined with mask r-cnn to monitor vegetation but identified that k-means was more accurate in change detection. Researchers apply machine learning algorithms such as random forest and linear regression to predict crop yields, with random forest displaying greater accuracy. Incorporating time series models such as ARIMA and LSTM significantly enhances the accuracy of the modeling process. Vegetation mapping with python is more efficient in crop monitoring than matlab-based methods.

Thirumal et al., [4] developed the sca-wrelm model, combining elm with regularization techniques, weighting functions, and sca optimization to improve the accuracy of rice crop yield forecasting. The model aims to enhance yield precision in rice-producing regions of South Asia, where reliable agricultural planning is crucial. This study utilizes sca optimization of weighted regularized elm (wrelm) parameters to boost prediction performance, creating an effective system for managing rice yields in key areas.

Choudhary, V et al., [5] examined the use of machine learning algorithms in predicting plant diseases, emphasizing the growing focus on applying artificial intelligence to agricultural research. Various kinds of machine learning models that were explored included supervised, unsupervised, and reinforcement learning and their uses to solve issues of agriculture. The research work of kim et al. (2020) along with kranth et al. (2021) shows how svm, neural networks, decision tree, naïve bayes and random forest models efficiently detect crop pests whereas random forest provides the best prediction accuracy among them. Machine learning for agricultural disease prediction is on the path of consistent development.

Mehetre, S et al., [6] emphasized the advancement of crop disease detection research, showcasing the move from traditional models such as ANN and K-NN to the more robust CNNs. This transition was necessitated by the limitations of earlier tools in adaptability and computational efficiency, particularly when working with larger datasets and generalized applications. Although earlier models showed strong performance with smaller datasets, their accuracy significantly decreased when handling larger, more complex datasets. Experiments have demonstrated that convolutional neural networks, especially ResNet, have become popular because of their novel approach to processing data, improved accuracy, and better capacity to deal with big, noisy datasets. Models like vgg16 and efficient models were utilized for certain crop identification, but their efficacy was constrained by the lack of input data and limited focus scope. Convolutional neural networks (CNNs) enhance the accuracy of crop disease detection significantly, with ResNet recognized as a prominent model that plays a key role in advancing the identification process.

Lohi, S.A et al., [7] presented various crop disease recognition models, incorporating advanced classification technologies to improve the accuracy of the results. The unified matrix-based cnn (umbcnn) and leaf wetness sensor (lws) demonstrate their efficacy in crop monitoring of wheat and rice, but they are limited in their capacity to adjust to different agricultural systems. Efficientnet v2 performs better in disease detection applications with greater accuracy and less computational power for various diseases. Models such as dilated multiple scaled AlexNet, ensemble nonlinear SVM, and augmented CNN rely heavily on extensive training datasets, while utilizing image segmentation techniques to improve their overall performance. The lack of feature selection technology slows down the implementation of real-time applications.

Dhande et al. [8] introduced the RoBERTa-BiLSTM hybrid model aimed at addressing challenges in sentiment analysis, including long dependencies, lexical variations, and data imbalances.

The model is based on Roberta for generating word embeddings and bilstm for extracting context semantics from the text. The experimental findings demonstrate enhanced performance compared to baseline models, achieving an accuracy of 92.36% on the IMDB dataset and yielding promising results on the Twitter datasets. Future work includes using a broad variety of datasets and tuning hyperparameters to reach even higher levels of performance.

Jim, J.R., et al.,[9], gives a summary of the recent developments in sentiment analysis, with emphasis on three areas of improvement. The study investigates the application of machine learning methods alongside deep learning approaches, integrating multimodal systems to enhance performance and achieve comprehensive results.

The research looks at the various ways each method is implemented, analyzes real-world examples, and highlights possible challenges that can be encountered during implementation. With their sophisticated architecture, BERT and RoBERTa transform transformer-based models by effectively tackling complex challenges. The study focuses on existing limitations in sentiment analysis, including bias and domain adaptation, while also identifying unresolved issues with current tools. The objective is to tackle these problems and enhance the overall efficiency of sentiment analysis systems. It analyzes the operation of sentiment analysis in different sectors and explores its uses in various fields.

Kancharagunta, K.B., et al., [10] conducted a study aimed at establishing a connection between Internet of Things (IoT) devices and machine learning systems, with the goal of developing intelligent systems. Extrapolated data comes as soon as measurement results. The system uses our system to measure soil quality by testing for nitrogen content, phosphorus, potassium, pH level, humidity, and temperature data. Include temperature and environmental variables. We make use of machine learning algorithms. Our system uses sophisticated processing methods to enhance decision trees and random forests. Decision trees and random forests, as machine learning models, serve as essential tools for exploring data, facilitating pattern recognition, predictive analysis, and the extraction of insights from intricate datasets. Analysts analyze past data together with current data to find relationships between various variables. Through analyzing data, we are able to learn about the factors that influence the choice of appropriate crops. Iot sensors collect environmental sensor data in real-time and send it to our system to be analyzed and used. Our system uses farm data in real-time to formulate personalized instructions depending on the particular conditions. Soil examination optimization results in precision agriculture decisions being boosted. The system offers personalized farm advice to farmers, allowing them to produce higher yields of crops while also increasing sustainable farming practices. The system prioritizes user-friendliness, ensuring seamless connectivity with a wide range of devices to accommodate diverse user needs effectively. The system combines traditional farm methods with computer data to develop personalized farm advice. New agricultural methods have been developed to advance crop production as well as environment conservation.

Yadav et al. [11] conducted research focusing on sentiment analysis and hate speech across multilingual texts. The evaluation and development of bilingual LLMs for English-Hindi and English-Slovene demonstrated that the fine-tuned bilingual model surpassed both multilingual and monolingual models in performance. Future studies aim to explore new language combinations, model variations, and applications in code-mixed scenarios.

Islam,M.R.etal.,[12] The combination of Internet of Things (IoT) and machine learning (ML) in agriculture can transform farming methods by improving soil nutrient management and making precise crop recommendations. Conventional methods of soil analysis and crop decision-making are generally labor-intensive and subjective, based on human intuition and manual effort. IoT-equipped devices, integrated with advanced sensors for measuring moisture, pH, and nutrient levels, provide real-time and precise insights into soil conditions, supporting better-informed decisions in agriculture. Data gathered is analyzed using machine learning models such as random forest, logistic regression, and neural networks to offer tailored crop recommendations. These advances enable precise and timely decision-making, maximizing use of resources, enhancing crop output, and facilitating sustainable agriculture practices. Additionally, field trials and mobile apps reinforce farmers' and consumers' strengths by maximizing visibility in the food supply chain.

Ekanayaka et al.[13] proposed a sophisticated system to monitor crops in real-time, which allows early detection of any health problem that could impact agricultural production and crop yield. The global food supply chain still faces threats due to failed crop production.

Under current circumstances, crops do not function as required due to the influence of diseases on production levels and farm income. There has been conventional disease diagnosis through inspections that has shown both longer time and high reliance on human judgment abilities that incur mistakes. Environmental factors can be sensed by using iot technology sensors to measure different conditions of the environment. The system utilizes temperature readings, soil moisture checks, and humidity tests to get data. Sensors are used in the system to track soil moisture in addition to temperature readings in order to recognize diseases at the earliest stages. Machine learning algorithms analyze vast datasets to uncover distinctive patterns and symptoms that may elude human observation. Random forest classifiers improve prediction accuracy by identifying issues at an early stage. With the support of electronic systems, farmers gain access to critical data points, enabling them to optimize their agricultural practices effectively. Controlled farming environments allow for the monitoring of crop diseases in their initial stages, enabling timely interventions.

Deshmukh, et al., [14] gives a summary of the machine learning methods that can be applied in classification-based prediction systems. Based on this study, conclude the following. Currently, there is no existing system in place that can promptly notify all stakeholders in the agricultural sector about the emergence of new crop diseases as they occur. Second, investigate other resources to learn about crop diseases.

A well-designed system would provide timely notifications containing crucial information such as symptoms, vulnerable crop varieties, disease classifications, and their mechanisms of transmission. This could enable swift responses and better coordination among agricultural stakeholders. This information could empower agricultural stakeholders to take preventive measures promptly and mitigate risks effectively. Lastly, read the most relevant research papers and examine those based on the application method of machine learning algorithms. The inference from that examination is that using an ensemble model made up of several ml classification models produces more accurate results than using a single ml classification model.

Kumar et al. [15] made an extensive study in order to observe the interlink between soil, water, and air, as well as find the most appropriate crops to plant considering soil condition and nutrient requirement. Fertilization strategies are derived from soil testing results, enriched by analyzing historical farming trends and rainfall patterns.

The system achieves a notable 85.55% accuracy in selecting the best crops for the forthcoming season and accurately predicts harvest times. Through detailed analysis of soil nutrients, it supports sustainable agriculture and contributes to increased crop yields. The technology is critical in the protection of crops from loss, reduction of wastage of resources, and increasing the effectiveness of farming processes. Through machine learning models like random forests and neural networks, the system enables farmers to select crops that are best suited for their specific conditions. This individualized method leads to higher profitability and efficient use of resources.

In addition, the solution provides actionable advice that helps farmers choose crops, thereby promoting improved yields and sustainable farming practices for the future. The research underscores the significance of soil, climate, and water as key elements in successful agricultural planning. It also points out how new developments in machine learning are transforming crop selection, providing intelligent solutions to overcome the challenges of contemporary farming.

GirishL [16]discusses the application of machine learning methods for predicting crop yield and rainfall patterns. The study presents a variety of machine learning techniques used to forecast crop yields and rainfall, providing important information for resource optimization and agricultural planning. It presents a comparison between several techniques, notably Decision Trees, K-Nearest Neighbors, Support Vector Machines, and Linear Regression. Support Vector Machines (SVM) emerged as the most effective algorithm for rainfall prediction, offering superior accuracy and reliability compared to other methods.

Ashwanikumar Kushwaha [17] outlines crop yield prediction techniques and suggests a suitable crop so that it will improve the profit for the farmer and quality of the agriculture sector. This study utilizes extensive datasets, referred to as big data, encompassing soil and weather information, processed through the Hadoop platform and agro algorithms. Utilizing this repository, the system predicts the compatibility of crops for particular conditions and contributes to improving the overall quality of yields.

Rahul katarya [18] describes the different machine learning methods used for accelerating crop yield. This study delves into multiple artificial intelligence approaches, including machine learning algorithms and big data analysis, aimed at enhancing precision agriculture. The authors provide insights into the development of a crop recommender system utilizing methods such as k-nearest neighbors (KNN), ensemble-based models, neural networks, and others to optimize agricultural decision-making.

Fatma M. Talaat [19] presents the Crop Yield Prediction Algorithm (CYPA), a cutting-edge system that improves precision agriculture by utilizing IoT techniques. This method combines data-driven tactics with cutting-edge technology to maximize crop yield forecasts. To help policymakers forecast annual crop yields, the authors integrate climate, meteorological, chemical, and agricultural data into CYPA. To assist in predicting crop yields, the researchers designed a tool to support decision-making for both farmers and policymakers. By analyzing regionally unique meteorological conditions, this technology allows for more individualized advice and better planning for agricultural productivity. The authors suggested a highly advanced machine learning technique for agricultural harvest estimation. By reducing the quantity of labeled training, active learning techniques were incorporated into the Crop Yield Prediction Algorithm (CYPA) to increase the model's efficiency.

III. PROPOSED WORK

The study utilizes an advanced agricultural framework combining IoT technologies, deep learning methods, and machine learning algorithms. IoT sensors continuously monitor crops in real time, enabling timely insights into plant health and effective identification of potential diseases for proactive agricultural decisions. Techniques like Random Forest and Support Vector Machines (SVM) are used to analyze soil properties like temperature, humidity, pH balance, moisture content, and NPK concentration. The system uses a Convolutional Neural Network to examine leaf photos, enabling effective and accurate plant disease identification. The integration of plant images, soil information, climatic patterns, and past yield data substantially improves the system's forecasting precision and decision-making effectiveness. Better agricultural planning and more efficient farming methods are supported by this all-encompassing strategy. A single smart framework that promotes sustainable farming methods houses these interrelated agricultural capacities. Information from sensors tracking key soil and environmental parameters moisture, temperature, humidity, pH, and NPK is used by the Plant Recommendation System.

IV. METHODOLOGY

The Plant Recommendation System gathers data from sensors that monitor soil moisture, temperature, humidity, pH, and NPK levels. The collected data undergoes pre-processing to ensure it is cleaned, normalized, and formatted appropriately for use with the machine learning model. Once the data is processed, the model examines it to produce customized recommendations, which are then made accessible to users through a user-friendly online platform or website. The Disease Detection System progresses through multiple steps, beginning with the user uploading an image of the plant for analysis. This image is preprocessed through steps such as resizing, normalization, and enhancement to ensure optimal quality. A dedicated deep learning model is utilized to examine the image and identify potential diseases. The results are then conveniently presented to the user through an accessible online platform or website.

The mentioned factors play a significant role in determining crop productivity, serving as vital contributors to achieving agricultural efficiency and enhancing overall yield performance. The system feeds the gathered data through a Decision Tree algorithm for pattern analysis based on historical facts and environmental elements. The system provides estimated crop yield predictions, enabling farmers to make well-informed decisions and effectively plan their agricultural activities for optimal results. The Yield Prediction System begins by acquiring essential input variables which include year, country, crop type, pesticide usage, average rainfall and temperature.

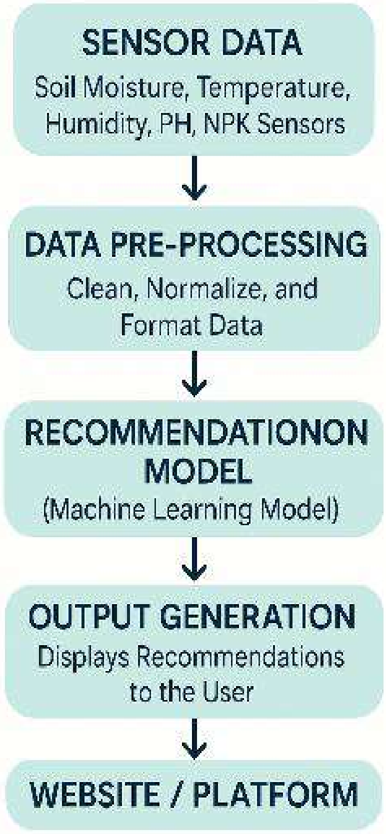

Fig 4.1 Plant recommendation

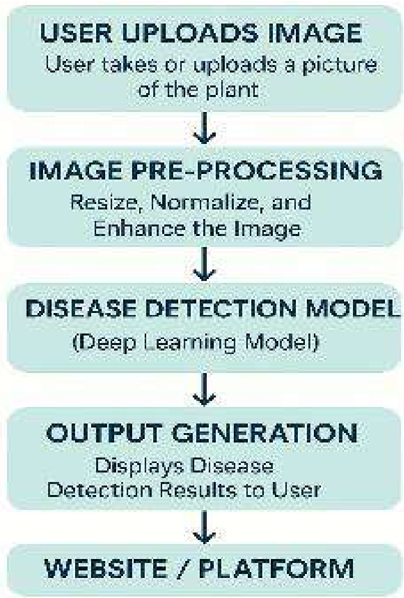

Fig 4.2 Plant Disease detection

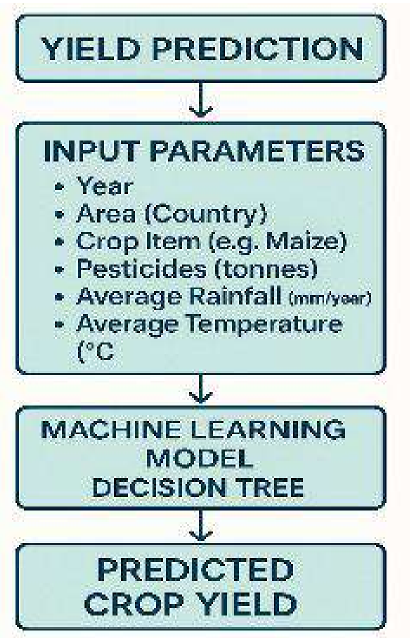

Fig4.3 CropYieldPrediction

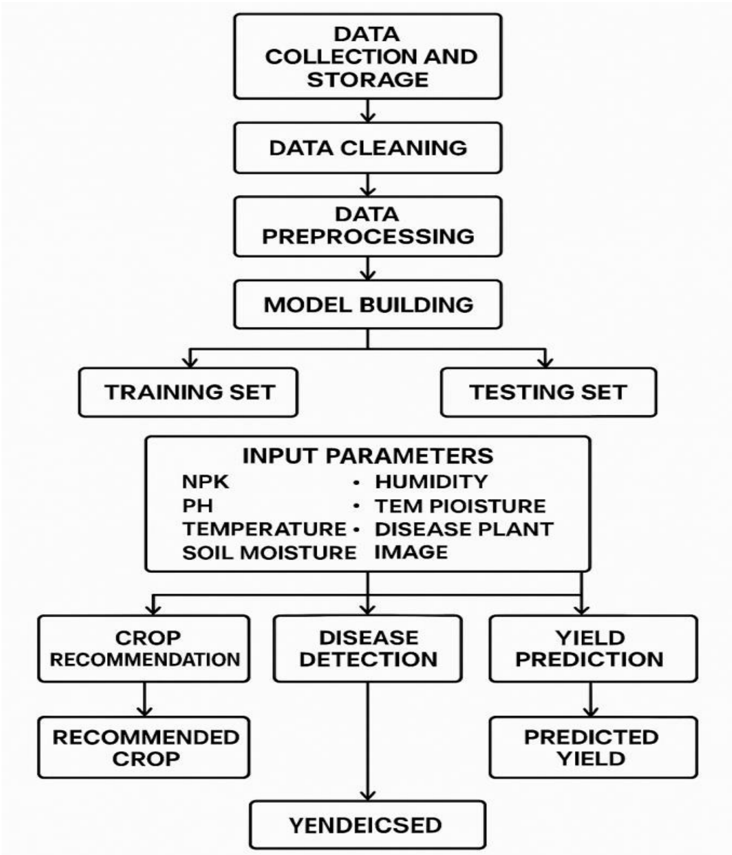

Fig 4.3: Integration of Crop Recommendation and Plant Disease Detection

IoT sensors initially gather data in real time from the soil and surroundings, documenting vital elements such as temperature, pH balance, nutrient levels, and plant photos for in-depth analysis and evaluation. The data preparation process begins with data cleansing, which involves eliminating invalid and irrelevant data from the input. In order to create a stable and effective state of data conditions for subsequent processing, the data preparation step includes normalization and augmentation.

A training set and a test set are generated from the supplied data in order to build efficient machine learning models. Five essential input parameters are used to construct these models: temperature, soil moisture, humidity, nutrient levels, and images of plant diseases.

In the subsequent stage, the system adopts a three-component functionality model.

- Soil properties and climatic conditions act as key inputs in crop recommendation systems, enabling the identification of the most suitable crops for specific agricultural environments.
- The Disease Detection system implements Convolutional Neural Networks (CNNs) to analyze plant leaf images so it can detect plant infections at their earliest stage.
- The Yield Prediction module applies Decision Trees to make yield estimates that rely on environmental factors together with cultivation parameters.
- Every component of the system works together as a unified unit to provide accurate agricultural decisions by offering full farming support, including timely disease warning, appropriate crop selection, and predicted yield values. The system makes advantage of the Flask platform deployment to offer farmers that desire increased productivity and ecological sustainability a solution that is easy to use, has real-time capability, and has scalable capabilities.

V. IMPLEMENTATIONS

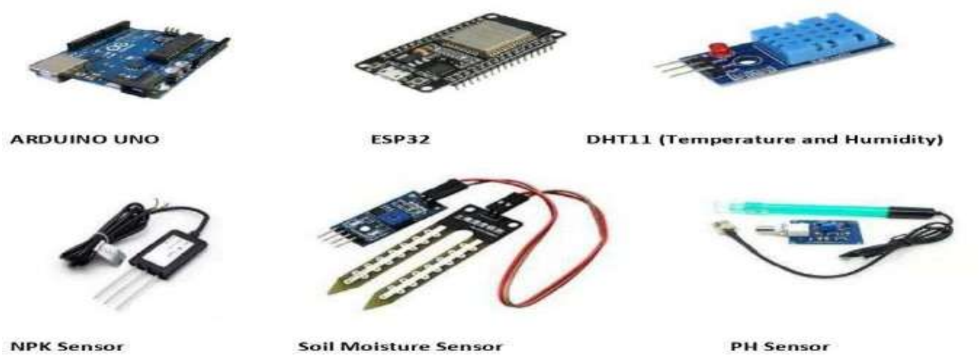

Fig 5.1: Sensors Input Modules (Sensors Used)

1. Soil Moisture Sensor
   - Irrigation needs are determined by checking the moisture level in the soil.
   - Aids in water consumption optimization for better crop health.
2. NPK Sensor
   - In order to improve crop development and preserve soil quality, it analyzes soil samples to determine the amounts of nitrogen (N), phosphorus (P), and potassium (K).
   - Assessing soil quality and offering targeted fertilizer recommendations enables farmers to sustain steady crop yields while guaranteeing the soil's long-term health and fertility.
3. pH Sensor
   - Assesses the soil's pH to identify its acidity or alkalinity, providing key information for soil health evaluation and the selection of suitable crops or corrective actions.
   - Crucial for figuring out whether soil is suitable for a given crop.
4. Temperature & Humidity Sensor (DHT11/DHT22)
   - Helps in climate-based crop recommendations.

CONTROL UNITS

1. Arduino
   - Processes data from all sensors.
   - Operates motors and processes input data to make immediate, adaptive decisions, ensuring efficient and responsive system functionality.
2. ESP32
   - Manages the web interface for remote monitoring and control.
   - Utilizes GPS technology to monitor and record the system's location, ensuring accurate tracking and efficient navigation.
   - Controls servos to handle various mechanisms like irrigation and spraying systems, enabling accurate and effective agricultural operations.

OUTPUT MODULES

1. LED Display / LCD Screen
   - Shows real-time sensor readings (NPK values, pH, moisture, temperature, and humidity).
   - Displays crop recommendations based on collected data.
2. Mobile / Web Interface
   - Users can remotely monitor soil conditions and crop recommendations.
   - Provides alerts and notifications for irrigation and fertilization needs.

Here, Fig. 5.1 Offers detailed sensor readings that play an essential role in gathering data, serving as a foundation for training and refining machine learning models to improve agricultural predictions and practices. The sensor data serves as the input, influencing the overall model performance and accuracy. The accuracy of each algorithm is influenced by its ability to analyze and process sensor data, resulting in the differences reflected in Table 6.1.Smart agriculture employs model representations, which consist of an NPK sensor, pH sensor, soil moisture sensor, temperature and humidity sensor, and LED display. The sensors gather live data about soil and environmental conditions, which is then analyzed and presented on the LED screen. The collected data is then uploaded to the aws cloud for storage and detailed analysis, enabling efficient monitoring and well-informed decision-making for better crop management.

Fig 6.1: Model Representations

VI. RESULTS

The Crop Recommendation dataset undergoes steps such as data preprocessing, exploratory data analysis (EDA), and the potential application of machine learning models for classification and recommendation. The next step involves deriving results, including model accuracy, evaluation metrics, and essential insights, to present a detailed analysis.

Dataset Overview

- The dataset comprises 2,200 entries distributed across 8 distinct columns, containing the following details:
  - N,P,K: macronutrients in the soil
  - Temperature, Humidity, pH, Rainfall: Environmental conditions
  - Label: Crop type classification
- The dataset is well-prepared, with no missing values identified.

Model Performance & Accuracy

The crop recommendation system was evaluated using various machine learning models, achieving the following accuracies**:**

Model	Accuracy(%)
Logistic Regression	96.36
Naïve Bayes	98.54
Support Vector Machine	96.81
K-Nearest Neighbors	95.90
DecisionTree	98.63
Random Forest	**99.09**
Bagging	98.86
Ada Boost	**14.09** (Poor Performance)
Gradient Boosting	98.18
Extra Trees	92.95

Table 6.1 Accuracy Comparisons of Different Models Best

Performing Model: Random Forest (99.09%)

Worst Performing Model: Ada Boost (14.09%)Overall, ensemble models (Naïve Bayes, Decision Tree, and Bagging) performed exceptionally well. Table 6.1 illustrates the accuracy levels of various machine learning models. Random Forest attained the best accuracy, reaching 99.09%. Bagging recorded an accuracy of 98.86%, with Decision Tree following closely at 98.63%, and Naïve Bayes achieving 98.54%. Gradient Boosting achieved 98.18%, while Support Vector Machine showcased a solid performance with 96.81%.In contrast, AdaBoost showed notably low accuracy at 14.09%, suggesting its unsuitability for this particular task.

Plant Disease Prediction Using Convolutional Neural Networks (CNN)

Dataset Overview: The Kaggle project makes use of the Plant Village Dataset, a vast collection of tagged photos showing both healthy and damaged plant leaves.

MODEL ARCHITECTURE

Convolutional Neural Network (CNN): A Convolutional Neural Network (CNN) model is employed in this study to extract features from leaf images and classify them based on the associated plant diseases.

Model Performance & Accuracy: While the exact accuracy metrics of the model are not provided, prior studies have consistently shown that CNNs are highly effective in accurately detecting and categorizing plant diseases as an example:
Early Prediction of Plant Diseases using CNN and GANs by adopting Generative Adversarial Networks (GANs) for data augmentation, this study's classification accuracy increased from 97% using a CNN model to 98%.With an average accuracy of 92% for lesion segmentation and up to 99.04% for identifying particular lesion subtypes in some plants, deep learning models for plant lesion segmentation have shown remarkable results. These findings emphasize the significance of CNN-based models in agricultural diagnostics, as they consistently and accurately identify and classify plant diseases.
Data Augmentation: The use of GANs has shown that enhancing the dataset with synthetic images can lead to improved model accuracy.
Model Optimization: The model's capacity to precisely identify and categorize plant illnesses can be further improved by using cutting-edge strategies like squeeze-and-excitation blocks and residual skip connections. Convolutional Neural Networks (CNNs) are highly efficient in predicting plant diseases, providing precise detection and classification of different disease types. Leveraging extensive datasets and cutting-edge deep learning methods, these models facilitate the early diagnosis of plant health issues, supporting improved agricultural practices and crop management. Accurate diagnostics provided by CNN-based methods play a crucial role in advancing crop management techniques and boosting agricultural productivity, while also promoting environmentally sustainable farming practices.

CROP YIELD PREDICTION USING DECISION TREE

Dataset Overview: Plant diseases can be accurately detected and classified using Convolutional Neural Networks (CNNs), which are quite effective at this task. By utilizing large datasets and state-of-the-art deep learning techniques, these models promote better crop management and agricultural practices by enabling the early diagnosis of plant health problems. In addition to improving agricultural production and crop management strategies, accurate diagnostics from CNN-based approaches are essential for encouraging ecologically friendly farming methods.

MODEL ARCHITECTURE

Decision Tree Algorithm: In addition to historical environmental parameters such as year, nation, crop types (including maize), and pesticide use, the yield prediction model makes use of structured agricultural data that includes statistics on average annual rainfall (in millimeters) and average temperature (in degrees Celsius). The data points' information is crucial for creating models that look at different factors affecting crop yields
Input Parameters:
Year: Indicates the growing year.
Area: Denotes the nation or geographical area where the crop is grown. Average Rainfall: expressed in millimeters per year. Average Temperature: Indicates the average temperature in degrees Celsius.
Model Output: The system predicts agricultural output by analyzing historical data alongside current weather trends. Finding patterns and correlations offers insightful information that can help farmers and resource managers make well-informed decisions. In order to guarantee sustainable and ideal farming practices, farmers and agricultural planners can use these forecasts as useful tools to help with proactive risk management techniques and effective resource allocation.
Model Performance & Accuracy: Decision Tree models are well known for their usefulness in agricultural production forecasts, albeit precise performance measures aren't given for this use case. The project's success aligns with the positive outcomes observed in prior studies from similar domains. For instance, a study titled "Machine Learning Techniques for Agricultural Yield Prediction" showed that Decision Tree algorithms could predict crop yields with an accuracy ranging from 85% to 90%. The study also showed that Decision Trees required less computing power than other regression methods while performing similarly to Random Forests in yield prediction.

6.1 WEB APPLICATION SCREEN SHOTS

Fig 6.2: Home Page

Fig 6.3: Register and Login

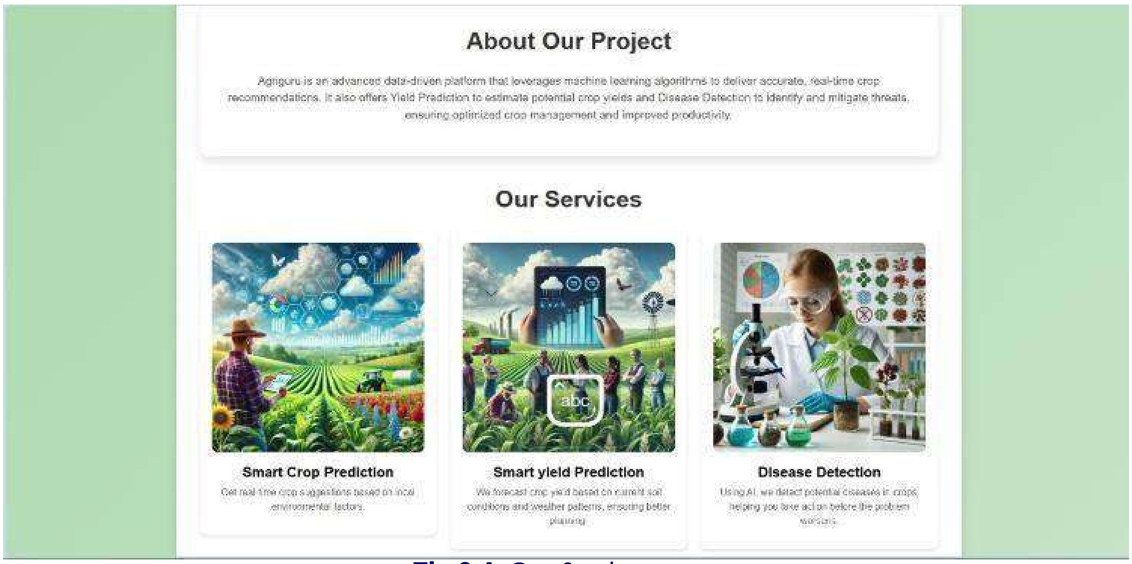

Fig 6.4: Our Services

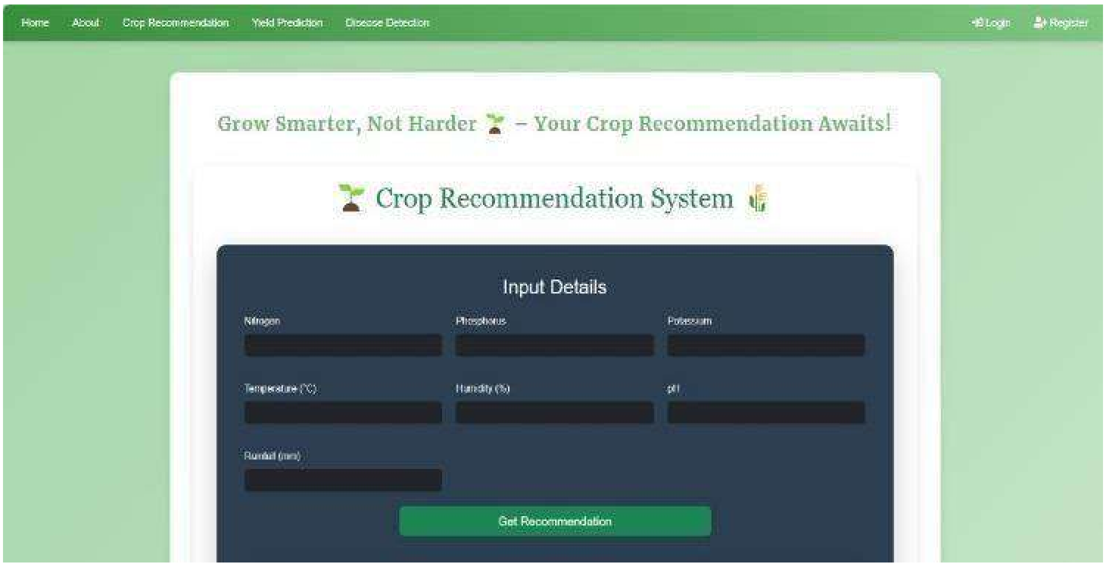

Fig 6.5: Crop Recommendation System

Fig6.6: Plant Disease Detection

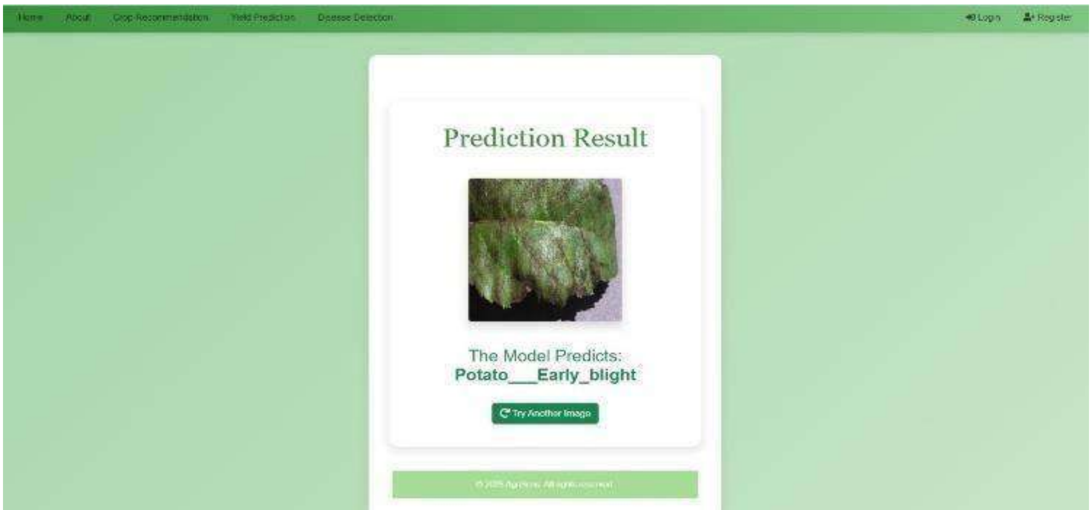

Fig 6.7: Plant Disease Detection Results

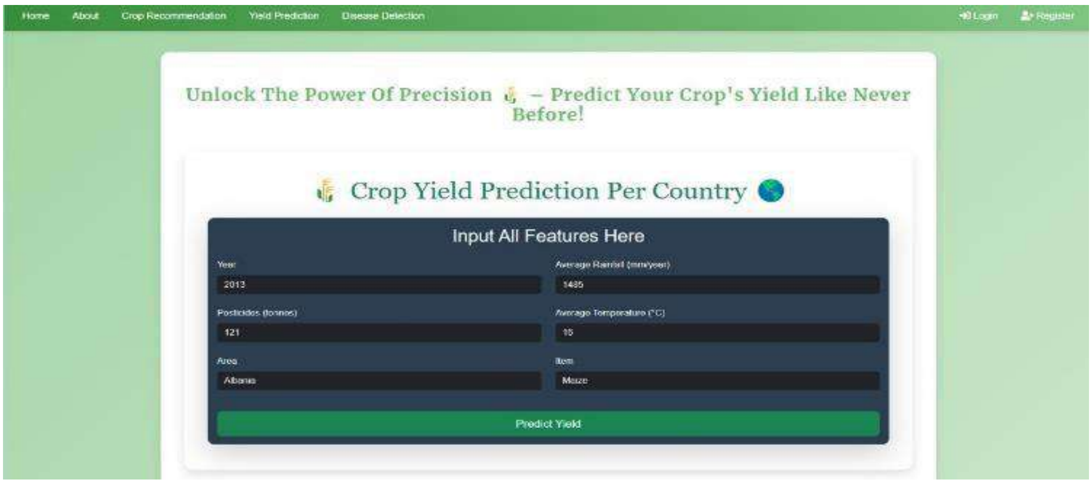

Fig 6.8: Crop Yield Prediction

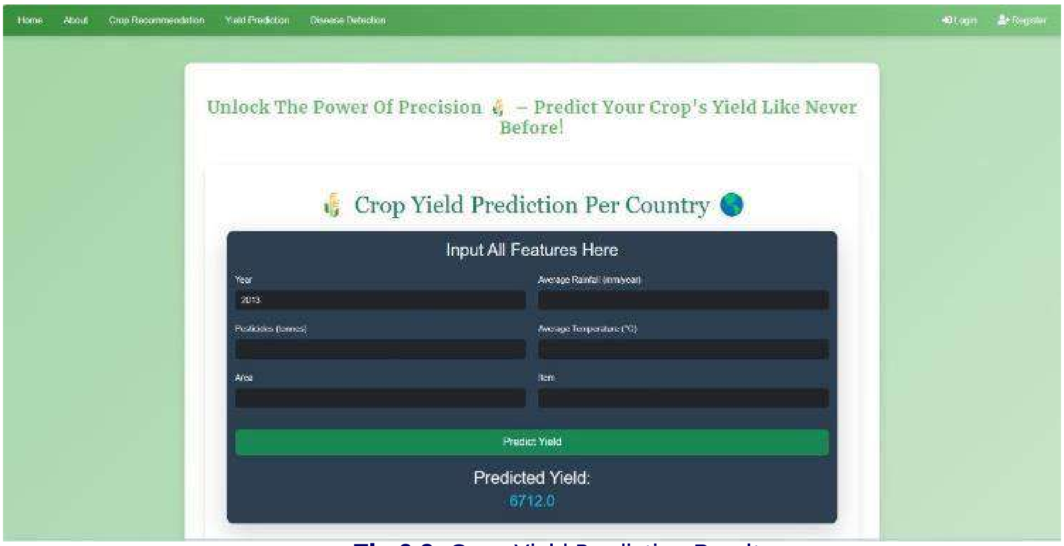

Fig 6.9: Crop Yield Prediction Result

VII. CONCLUSION

Modern advancements in technology drive agricultural growth, fostering economic development in India. The proposed system employs CNNs for plant disease identification, SVC for recommending crops, and Decision Tree models for forecasting yields. IoT sensors collect live data on NPK levels, soil moisture, temperature, and humidity, enhancing the precision of predicted results by providing accurate and timely environmental insights. This system enables farmers to identify plant diseases early, receive tailored crop recommendations, and access precise yield predictions, supporting informed decision-making and efficient resource usage. User-friendly interfaces on both web and mobile platforms ensure widespread adoption among different farmer communities. Future enhancements are designed to incorporate functionalities like disease detection combined with wholesale market price forecasting, fostering sustainable agricultural practices and supporting farmers with informed decision-making.

REFERENCES

1. Jeevaganesh, R., Harish, D. and Priya, B., 2022, April. A machine learning-based approach for crop yield prediction and fertilizer recommendation. In 2022 6th International conference on trends in electronics and informatics (ICOEI) (pp. 1330-1334). IEEE
2. Patki, V. and Wazurkar, P., 2021, June. Application of Data Sampling for Crop Yield Prediction using Stochastic Gradient Descent Neural Networks. In 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT) (pp. 685-689). IEEE.
3. Shahrin,F.,Zahin,L.,Rahman,R.,Hossain,A.J.,Kaf,A.H. and Azad, A.A.M., 2020, December. Agricultural analysis and crop yield prediction of habiganj using multispectral bands of satellite imagery with machine learning.In202011thInternational Conference on Electrical and Computer Engineering (ICECE) (pp. 21-24). IEEE.
4. Thirumal, S. and Latha, R., 2023, May. Automated Rice Crop Yield Prediction using Sine Cosine Algorithm with Weighted Regularized Extreme Learning Machine. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 3540). IEEE.
5. Choudhary,V.andThakur,A.,2022,April.Comparative Analysis of Machine Learning Techniques for Disease Prediction in Crops. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 190-195). IEEE.
6. Mehetre, S., Mangle, R., Karvanje, P. and Sarode, R.Y., 2023,June. Crop Disease Diagnosis using Convolutional Neural Network. In 2023 3rd International Conference on Intelligent Technologies (CONIT) (pp. 1-7). IEEE.
7. Lohi,S.A.andBhatt,C.,2023,February.DesignofaCropDiseaseDetectionModelusingMulti-parametricBio-inspired FeatureRepresentationandEnsembleClassification.In2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) (pp. 1-6). IEEE.
8. Dhande,A. Malik,R.,2022,March.Empirical study of crop-disease detection and crop-yield analysis systems: a statistical view. In 2022 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 1-4). IEEE.
9. Jim,J.R.,Talukder,M.A.R.,Malakar,P.,Kabir,M.M.,Nur,K. and Mridha, M.F., 2024. Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. Natural Language Processing Journal, p.100059.
10. Kancharagunta, K.B., Akkem, Y., Vembadi, M., Garalapati, S.T.,HariPratapVarma,A.andRuhaJessica,M.,2024,February. Crop Recommendation System Using Machine Learning and IoT: A Survey. In International Conference On Innovative Computing And Communication (pp. 63-86). Singapore: Springer Nature Singapore.
11. Yadav,A., Garg,T., Klemen,M.,Ulcar,M.,Agarwal,B. and Sikonja, M.R., 2024. Code-mixed Sentiment and Hate-speech Prediction. arXiv preprint arXiv:2405.12929.
12. Islam, M.R., Oliullah, K., Kabir, M.M., Alom, M. and Mridha,M.F.,2023.Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. Journal of Agriculture and Food Research, 14, p.100880.
13. Ekanayaka, E.M.H.I. and Kumara, P.P.N.V., 2022, August. Machine Learning Approach for New Crop Disease Predict

and Alert System: A Review. In 2022 International Conference on Emerging Techniques in Computational Intelligence (ICETCI) (pp. 4549). IEEE.

14. Deshmukh, T., Rajawat, A.S. and Potgantwar, A., 2023, October. Machine Learning Technique for Crop Selection and Prediction of Crop Cultivation. In 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA) (pp. 17). IEEE.

15. Kumar,R.and Shukla, N., 2022, March. Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 168-172). IEEE.

16. Vaishnnave, M.P., Kamaitachi, S.B., Kumar, C.R. and Premkumar, S., 2023, December. Prediction Of Crop: Based On VariousPredictiveAlgorithmicModels.In20235thInternational Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 725-728). IEEE.

17. S.Raj Kumar , S.Sowrirajan," Automatic Leaf Disease Detection and Classification using Hybrid Features and Supervised Classifier", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 5, Issue 6,2016 .

18. Chana, A.M., Batchakui, B. and Nges, B.B., 2023. Real- Time Crop Prediction Based on Soil Fertility and Weather Forecast Using IoT and a Machine Learning Algorithm. Agricultural Sciences, 14(5), pp.645-664.

19. Holzman ME, Carmona F, Rivas R, Niclo`s R (2018) Early assessment of crop yield from remotely sensed water stress and solar radiation data. ISPRS J Photogramm Remote Sens 145:297–308.

20. Shiragudikar, S.K., Bharamagoudar, G., Manohara, K.K. et al. "Insight Analysis of Deep Learning and a Conventional Standardized Evaluation System for Assessing Rice Crop's Susceptibility to Salt Stress during the Seedling Stage", ISSN: 2661-8907, SN COMPUT. SCI. 4, 262 (2023). https://doi.org/10.1007/s42979-022-01656-2

21. SinghA,GanapathysubramanianB,SinghAK,SarkarS (2016) Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci 21(2):110–124.

22. Alhnaity B, Pearson S, Leontidis G, Kollias S (2019) Using deep learning to predict plant growth and yield in greenhouse environments. In International Symposium on Advanced Technologies and Management for Innovative Greenhouses Green Sys 20191296:425–432.

23. Malathi S Y, Geeta R Bharamagoudar (2023) "Classification of Knee X-Ray Images by Severity of Osteoarthritis Using Skip Connection Based ResNet101", in International Journal of Intelligent Engineering and Systems. ISSN:2185-3118

24. Taher F, Abdel-salam M, Elhoseny M, El-hasnony IM (2023) Reliable Machine Learning Model for IIoT Botnet Detection. IEEE Access 11:49319–49336.

25. PourpanahF,LimCP,Wang X, Tan CJ,SeeraM, ShiY (2019) A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification. Neuro computing 333:440–451.

26. Qader SH, Dash J, Atkinson PM (2018) Forecasting wheat and barley crop production inaridand semi-arid regions using remotely sensed primary productivity and crop phenology: A case study in Iraq. Sci Total Environ 613:250–262.

27. PuneethTotad, Geeta R. Bharamagoudar and Shanta S. Kallur "Boosting-based machine learning approaches for diabetes prediction using Indian demographic and health survey-2021 data" , Research Square, 11-April-2023, https://doi.org/10.21203/rs.3.rs-2784266/v1

28. Sharada K. Shiragudikar, Geeta Bharamagoudar, "Enhancing Rice Crop Resilience: Leveraging Image Processing Techniques in Deep Learning Models to Predict Salinity Stress of Rice during the Seedling Stage", ISSN: 2147-6799, Vol. 12 No. 14s (2024), 02.02.2024, https://ijisae.org/index.php/IJISAE/article/view/4643

29. Azzari G, Jain M, Lobell DB (2017) Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. Remote Sens Environ 202:129–141.

30. Sun J, Di L, Sun Z, Shen Y, Lai Z (2019) County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19(20):4363.

31. Geeta R.B, GouriC.Khadabadi, Shashikumar G Totad , "A Survey on identification and classification diseases in crops using image processing techniques", 12th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2021, Volum2021-August 2021, Pages 174-179, Hyderabad.

32. Irita K (2011) Risk and crisis management in intraoperative hemorrhage: Human factors in hemorrhagic critical events. Korean J Anesthesiol 60(3):151–160.

33. Elavarasan D, Vincent PD (2020) Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE access 8:86886–86901.