Soil Quality Analyzer and Crop Recommendation System Using Machine Learning Algorithms for Optimizing Agricultural Productivity and Sustainability

KAVIPREETHA R[1], KAVINASHRI V[2], NITHISHA N[3], KALAIARASAN T[4]
[1, 2, 3]*Student, Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, India*
[4]*Assistant Professor, Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, India*

Abstract- Agriculture is crucial to India's economy, employing about 48% of the population, yet traditional farming methods and soil depletion pose challenges to crop yield. This project aims to assist farmers by analyzing soil and pH data using machine learning to recommend suitable crops and fertilizers, addressing the lack of precision agriculture. By leveraging categorization strategies, the system enhances crop and fertilizer recommendations, aiding novice farmers. Additionally, it introduces a computer-aided disease recognition model to combat crop losses due to unidentified plant diseases, replacing time-consuming manual inspections with accurate, automated detection. The integration of modern technology in farming not only increases productivity but also optimizes resource utilization, minimizes wastage, and supports sustainable agricultural practices. This approach empowers farmers with data-driven insights, ensuring better decision-making and improved profitability in the long run.

Indexed Terms- ML-Machine Learning, NEHA-National E-Health Authority, PCA-Principal Compound Analysis.

I. INTRODUCTION

Agriculture is a major industry in India, employing a significant portion of the population, and as the world's population expands, agricultural challenges continue to grow. Environmental changes have severely impacted crop production, leading to financial struggles, debt, and even suicide among farmers. Many farmers rely on traditional, non-scientific methods for crop selection, which often results in reduced yield and profit, forcing them to migrate to cities, lease land to industrialists, or abandon farming altogether. Given the limited availability of resources, their efficient utilization is crucial, and technology plays a vital role in mitigating these issues. Machine learning, a branch of artificial intelligence, is being implemented to analyze soil nutrient values and climate conditions for recommending the most suitable crops and fertilizers. The ensemble method integrates multiple machine learning models to enhance the accuracy of crop and fertilizer selection. Furthermore, disease identification is a crucial aspect of a good farming system, as plant diseases often go unnoticed or are misdiagnosed through visual inspection. To address this, image processing techniques are utilized to detect plant diseases using leaf analysis, providing early and accurate diagnosis. Modern farming integrates machine learning to manage the rapidly increasing agricultural data, allowing for better decision-making and resource optimization. The application of artificial intelligence in farming has the potential to tackle various challenges, from categorizing soil information to improving crop yield predictions. Deep learning is becoming increasingly significant in agriculture, offering precise insights into crop yield analysis. Crop prediction remains a critical issue, and in the future, it will be enhanced by incorporating user experiences and detailed data analysis to improve forecasting accuracy. The adoption of different machine learning techniques allows for better estimations of future agricultural production, ensuring more efficient and sustainable farming practices.

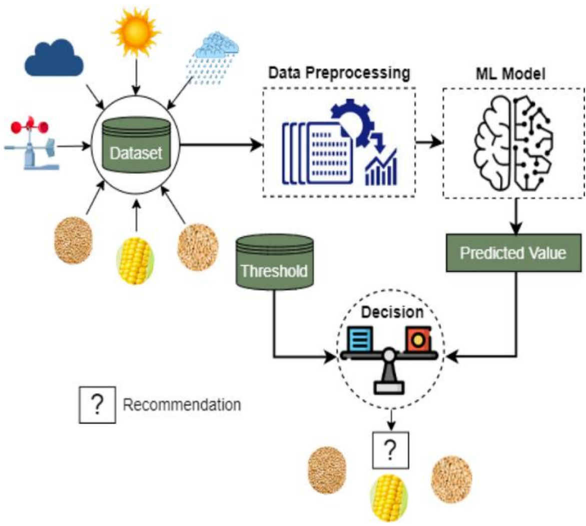


Figure 1 Crop Recommendation System

Machine learning is a subsection of computer science that has evolved through technological discoveries, facilitating automation and reducing the need for manual labor. It is a form of artificial intelligence (AI) that enables computers to acquire knowledge without direct communication guidance. The primary goal of machine learning is to develop computer programs capable of adapting to new data. For untrained farmers, selecting the right crops based on soil conditions can be challenging, and addressing agricultural erosion is equally important. Machine learning tasks are generally categorized into three types based on the learning signal or feedback received. Supervised learning involves training a model using labeled input-output pairs to establish patterns, whereas unsupervised learning requires the model to identify hidden structures within data without labeled outputs. Reinforcement learning, on the other hand, enables a computer algorithm to interact with a dynamic environment and complete specific tasks, such as operating a vehicle, without explicit instructions. Various machine learning algorithms exist, many of which are available in platforms like Azure Machine Learning Studio. These algorithms are broadly classified into four categories, each suited for different types of problems. Among them, anomaly detection algorithms are particularly effective for identifying unusual data points, making them useful in various applications, including fraud detection, predictive maintenance, and agricultural analysis.

.

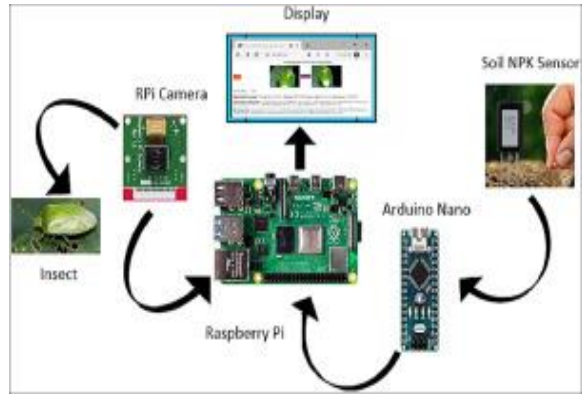


Figure 2 Fertilizer Recommendation System

II. LITERATURE REVIEW

The literature review explores various studies related to data mining, machine learning, and technological advancements in agriculture. Research by Jignasha M. Jethva et al. (2018) highlights data mining techniques for fertilizer recommendations, addressing the issue of incorrect fertilizer application. M.C.S. Geetha (2015) discusses the role of data mining in agricultural decision-making, while S. Srija et al. (2016) introduce an Android-based Agro Nutri application for fertilizer recommendations. Karandeep Kaur (2016) emphasizes machine learning applications in Indian agriculture, showcasing its potential to enhance precision farming. Petr Havlik et al. (2011) analyze the impact of population growth and economic development on food production, and James W. Jones et al. (2016) provide insights into agricultural system modeling, demonstrating its significance in decision-making. These studies collectively highlight the importance of technology-driven solutions in improving agricultural efficiency and sustainability.

[1] Title: A Review on Data Mining Techniques for Fertilizer Recommendation 2018.Authors : Jignasha M. Jethva, Nikhil Gondaliya, Vinita Shah keep up nutrition levels in the soil in case of deficiency, fertilizers are added to soil. The standard issue existing among the Indian agriculturists choose approximate amount of fertilizers and add them manually. Excess or deficient 14 extension of fertilizers can harm the plants life and reduce the yield. This paper gives overview of various data mining frameworks used on cultivating soil datasetfor fertilizer recommendation.

[2] Title: A Survey on Data Mining Techniques in Agriculture, 2015. Authors : M.C.S.Geetha

Agriculture is the most critical application area especially in the developing nations like India .Use of information technology in agriculture can change the situation of decision making and farmers can yield in better way.. This paper integrates the work of several authors in a single place so it is valuable for specialists to get data of current situation of data mining systems and applications in context to farmingfield. [3] Title : Agro Nutri Android Application,2016.Authors : S. Srija, R. Geetha Chanda, S.Lavanya, Dr. M. Kalpana Ph.DThis paper communicates the idea regarding the making of AgroNutri an android application that helps in conveying the harvest particular fertilizer amount to beapplied. The idea is to calculate the measure of NPK composts to be applied depend on the blanked proposal of the crop of interest. This application works depend on the product chosen by the farmer and that is taken as input, thus providing the farmers. The future scope of the AgroNutri is that GPRS can be included so that according to location nutrients are suggested. [4] Title: Machine Learning: Applications in Indian Agriculture, 2016.Authors: Karandeep KaurAgriculture is a field that has been lacking from adaption of technologies and their advancements. Indian agriculturists should be up to the mark with the universal procedures. Machine learning is a native concept that can be applied to every field on all inputs and outputs. It has effectively settled its ability over ordinary calculations of software engineering and measurements. Machine learning calculations have improved the exactness of artificial intelligence machines including sensor based frameworks utilized in accuracy farming. This paper has evaluated the different uses of machine learning in the farming area. It additionally gives a knowledge into the inconveniences looked by Indian farmers and how they can be resolved using these procedures.15 [5] Title: Impacts of population growth, economic development, and technical change on global food production and consumption, 2011.Author: Petr Havlik b, Erwin Schmid c, Hugo Valin Throughout the following decades humanity will request more food from less land and water assets. This investigation evaluates the food production effects of four elective advancement situations from the Millennium Ecosystem Assessment and the Special Report on Emission Scenarios. partialy and jointly considered are land and water supply impacts from population development, and specialized change, and forests and agriculture demand request shifts from population development and economic improvement. The income impacts on nourishment request are registered with dynamic flexibilities. Worldwide farming area increments by up to 14% somewhere in the range of 2010 and 2030.Deforestation restrictions stronglyimpact the price of land and water resources but have little consequences for theglobal level of food production and food prices. While projected income changeshave the highest partial impact on per capita food consumption levels, populationgrowth leads to the highest increase in total food production. The impact oftechnical change is amplified or mitigated by adaptations of land managementintensities [6] Title: Brief history of agricultural systems modelling,2016.Author: James W. Jones a,∗, John M. Antle b, Bruno O. Basso c, Kenneth J.Boote a, Richard T. Conant d, Ian Foster e, H. Charles J. Godfray Rural frameworks science creates information that enables analysts to consider complex issues or take educated farming choices. The rich history of this science represents the decent variety of frameworks and scales over which they work and have bee contemplated .Demonstrating, a basic apparatus in agrarian frameworks science, has been expert by researchers from an extensive variety of controls, who have contributed ideas and instruments over six decades. As agrarian researchers currently consider the "people to come" models, information ,and learning items expected to meet the inexorably mind boggling frameworks issues looked by society. before this literature review give me 6 lines para about literature review.

III. RELATED WORK

The literature review explores various studies related to data mining, machine learning, and technological advancements in agriculture. Research by Jignasha M. Jethva et al. (2018) highlights data mining techniques for fertilizer recommendations, addressing the issue of incorrect fertilizer application. M.C.S. Geetha (2015) discusses the role of data mining in agricultural decision-making, while S. Srija et al. (2016) introduce an Android-based Agro Nutri application for fertilizer recommendations. Karandeep Kaur (2016) emphasizes machine learning applications in Indian agriculture, showcasing its potential to enhance precision farming. Petr Havlik et al. (2011) analyze the impact of population growth and economic development on food production, and James W. Jones et al. (2016) provide insights into agricultural system modeling, demonstrating its significance in decision-making. These studies collectively highlight the importance of technology-driven solutions in improving agricultural efficiency and sustainability.

IV. METHODOLOGY

The proposed system leverages Machine Learning (ML) and Deep Learning (DL) to provide accurate crop selection, fertilizer recommendation, and plant disease detection. Initially, datasets are collected from agricultural sources, including soil parameters (pH, NPK levels, temperature, and moisture), climate data (rainfall, humidity), and plant leaf images for disease detection. The data undergoes preprocessing, where missing values are handled, and feature scaling is applied for model efficiency. For crop and fertilizer recommendation, a Random Forest algorithm is trained to predict the most suitable crop and fertilizer based on soil and climate conditions, achieving 95% accuracy. The disease prediction system employs a Convolutional Neural Network (CNN) to classify plant leaves as healthy or infected. The trained models are evaluated using metrics like accuracy, precision, recall, and F1-score, with hyperparameter tuning performed for optimization. A web application is developed using React.js for the frontend and Flask/Django for backend to allow farmers to input soil parameters for recommendations and upload leaf images for disease identification. The system is deployed on a cloud-based server for real-time predictions and tested with real user inputs to ensure accuracy. By integrating ML and DL techniques, the system provides precise agricultural recommendations, helping farmers increase productivity, reduce losses, and improve decision-making in farming.

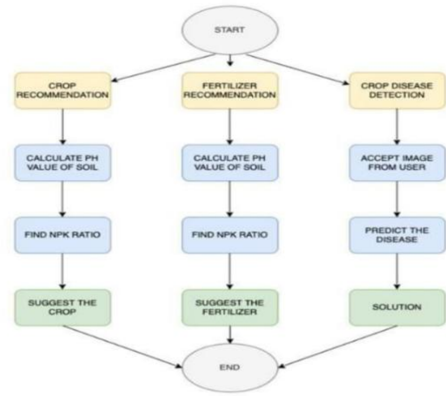


Figure 3 System Flow Diagram

A. Crop Recommendation System

This module aims to recommend suitable crops based on soil analysis and environmental factors such as temperature and humidity. By considering various parameters, including crop compatibility with soil type, climate suitability, and market demand, farmers can make informed decisions to maximize their yield and profitability.

B. Fertilizer Recommendation System

This module focuses on analyzing the quality of soil to determine its suitability for various crops. By assessing key properties such as nutrient levels, pH balance, and moisture content, farmers can make informed decisions about crop selection and soil management practices.

C. Disease Detection and Diagnosis

This module focuses on detecting and diagnosing plant diseases using computer vision techniques and deep learning algorithms. By analyzing leaf images and identifying disease symptoms, farmers can take timely action to prevent disease spread and minimize crop losses

Figure 4.Crop disease prediction

V. ALGORITHM DETAILS

Random Forest is a widely used supervised machine learning algorithm that excels in both classification and regression problems. It operates on the principle of ensemble learning, which enhances model performance by combining multiple classifiers. Instead of relying on a single decision tree, the Random Forest algorithm constructs multiple trees using different subsets of the dataset and aggregates their predictions to improve accuracy while minimizing overfitting. The final output is determined based on the majority vote of all decision trees, making the model robust and reliable. In the context of crop and fertilizer recommendation, the system collects essential soil and environmental parameters, including Nitrogen (N), Phosphorus (P), Potassium (K), rainfall, city, and state details of the user's agricultural land. The ML model is trained with 23 different crops, each analyzed using 100 different nutrition parameters to ensure precise recommendations. When a farmer inputs their desired crop along with soil nutrient levels, the system predicts the most suitable fertilizer composition to optimize growth. By leveraging Random Forest's high accuracy and adaptability, this approach aids farmers in maximizing crop yield, minimizing resource wastage, and improving overall agricultural productivity. Additionally, this model can be expanded to include real-time sensor data, allowing dynamic updates for more accurate recommendations. The system also considers seasonal variations to improve its predictive performance. Historical yield data is integrated to enhance future crop planning and sustainability. Moreover, weather conditions are factored into the recommendation process, ensuring adaptability to climate changes. With its ability to process large datasets efficiently, this model serves as a powerful decision-making tool for modern agriculture.

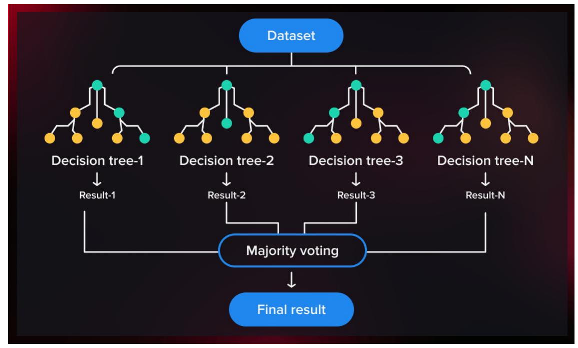


Figure 5 Random Forest Diagram

Naïve Bayes is a supervised learning algorithm that operates based on Bayes' theorem, making a fundamental assumption that all features are independent of each other. Despite its simplicity, it has been proven to outperform complex classification algorithms in various real-world applications. According to Bayes' theorem, the relationship between an input variable xxx and an output variable yyy is expressed as:

p(y|x)=p(y)·p(x|y)p(x)p(y|x) = {p(y) \c p(x|y)}{p(x)}p(y|x)=p(x)p(y)·p(x|y)

In this project, the Gaussian Naïve Bayes algorithm is implemented, which assumes that the features follow a Gaussian (normal) distribution. For continuous data, the system first segments the dataset based on the class label (y), then calculates the mean (μ) and variance (d^2) for each feature in that class. When a new observation is encountered, its probability is estimated using the Gaussian probability density function. This approach allows for efficient classification of agricultural data, such as identifying disease patterns in crops based on leaf images or predicting suitable fertilizers based on soil nutrient levels. By leveraging Naïve Bayes' fast computation and low resource requirements, this model aids in making quick and accurate predictions, supporting better decision-making in agriculture.

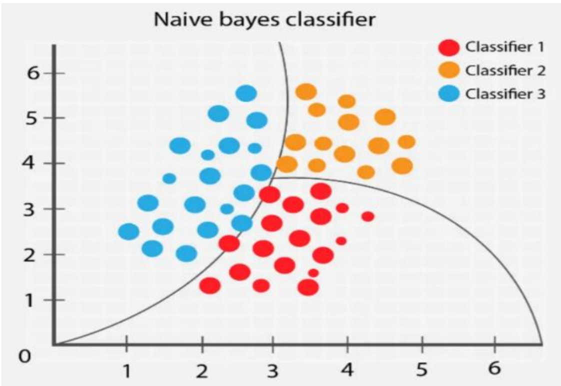
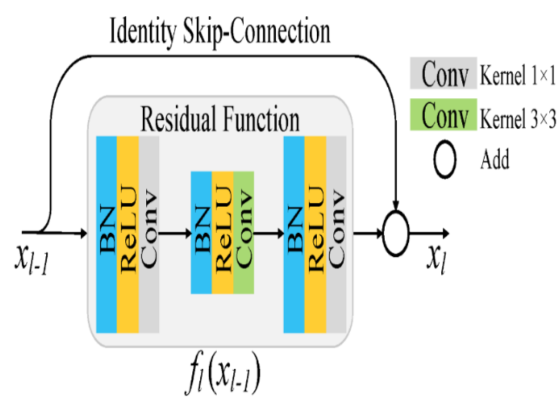


Figure 6: Naïve bayes classifier

VI. RESULTS AND DISCUSSION

The Crop Recommendation System allows users to input N-P-K (Nitrogen-Phosphorous-Potassium) values, along with their state and city, to receive the most suitable crop suggestions based on soil composition and environmental conditions. Users should enter common city names, as remote locations may not be available in the Weather API, which fetches humidity and temperature data. The Fertilizer Suggestion System analyzes soil nutrient levels and the selected crop to identify nutrient deficiencies or excesses and recommends appropriate fertilizers to enhance soil fertility. The Disease Detection System enables users to upload a leaf image, which is processed using deep learning models to determine whether the plant is healthy or diseased. If a disease is detected, the system identifies the cause and provides preventive or curative measures. The dataset used for training the disease detection model contains approximately 87K images of healthy and diseased crops, and it is split in an 80/20 ratio for training and validation, ensuring robust model performance. Each image is processed in RGB format with a 256×256 resolution, enabling accurate feature extraction for disease classification. The system undergoes rigorous testing and optimization in this phase to ensure high accuracy and efficiency in real-world applications.



Algorithm	Accuracy
Existing	80
Proposed	95

COMPARISON TABLE

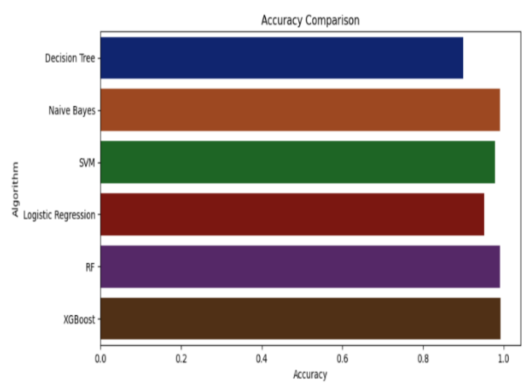


Figure 7: COMPARISON GRAPH

CONCLUSION

The proposed system effectively recommends the most suitable crop for a specific land based on key parameters such as annual rainfall, temperature, humidity, and soil pH. The system predicts annual rainfall using historical data with the Random Forest algorithm, while other parameters are provided by the user. The output includes the recommended crop, required seeds per acre, market price, approximate yield, and necessary NPK (Nitrogen, Phosphorus, Potassium) values for optimal growth. Many farmers lack access to advanced technology, often leading to poor crop selection and financial losses. To address this, a farmer-friendly GUI-based system has been developed, providing precise crop recommendations, required nutrients, expected yield, and market price, thereby enabling better decision-making and boosting agricultural productivity. Additionally, a plant disease prediction system using Convolutional Neural Networks (CNN) and ResNet34 has been implemented, leveraging convolutional layers for feature extraction, pooling layers for sample downsizing, and dense layers for final prediction. The ReLU activation function is used in convolutional layers, while the softmax activation function is applied in the dense layer for classification. The system has been trained on 10,000 plant leaf images from 14 different plants, successfully recognizing 38 types of plant diseases, making it a valuable tool for modern precision farming.

FUTURE WORK

Future enhancements for crop prediction, fertilizer recommendation, and plant disease prediction in machine learning (ML) will focus on integrating IoT and satellite data for real-time monitoring, employing advanced algorithms like deep learning and reinforcement learning, and enhancing disease image recognition for early detection. The implementation of AI chatbots will provide instant guidance to farmers, while personalized recommendations based on historical and real-time farm data will optimize decision-making. ML models will also be adapted to address climate change impacts, ensuring sustainable farming practices. User-friendly mobile applications will be developed for accessibility, encouraging data sharing and collaboration to improve agricultural datasets. Incorporating blockchain technology will enhance supply chain transparency, while the development of low-cost sensors will empower small-scale farmers by providing affordable precision agriculture solutions. Additionally, an admin module will be introduced, enabling data storage and user tracking, allowing administrators to monitor system usage efficiently. These advancements aim to boost agricultural productivity, sustainability, and food security while minimizing environmental impact through data-driven and technology-enhanced farming practices.

REFERENCES

[1] Joyce G. Latimer, Extension Specialist, Greenhouse Crops; Virginia Tech, *"The Basics of Fertilizer Calculations for Greenhouse Crops"*, Virginia Polytechnic Institute and State University, 2015.

[2] Chuan-Sheng Zhou, Software College, Shenyang Normal University, Shenyang, Liaoning, and Li-Hua Niu, College of Basic Medicine, Medical University, He Fei, China, *"Research on Component-Based Online Shopping System Design"*, International Conference on Computational Aspects of Social Networks, 2010.

[3] M.L. Vitosh, Extension Specialist, Crop and Soil Sciences, *"N-PK Fertilizers"*, Michigan State University Extension, Reprint July 1996.

[4] Agricultural Department, *"Soil Test Information | Recommended Fertilizers – Soil Health Card"*, February 2015.

[5] E.W. Wilde, R.L. Brigmon, C.J. Berry, D.J. Altman, J. Rossabi, B.B. Looney, and S.P. Harris, *"D-Area Drip Irrigation Phytoremediation Project: SRTC Final Report"*, WSRCTR Journal, January 2013.

[6] S. Praveena Katharine, R. Santhi, S. Maragatham, R. Natesan, V. Ravikumar, and Pradip Dey, Department of Soil Science and Agricultural Chemistry, Tamil Nadu Agricultural University, Coimbatore, India; Agricultural Engineering College and Research Institute, Kumulur, Tamil Nadu, India; Indian Institute of Soil Science, Bhopal, Madhya Pradesh, India, *"Soil Test Based Fertilizer Prescriptions Through Inductive Cum Targeted Yield Model for Transgenic Cotton on Inceptisol"*, Volume 6, Issue 5 (Nov. - Dec. 2013).

[7] Lloyd Murdock, Extension Soils Specialist, *"Evaluating Fertilizer Recommendations"*, IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS), December 2011.

[8] Nilkamal More, Assistant Professor, Department of Information Technology, K.J. Somaiya College of Engineering, Mumbai-07, *"Recommendation of Books Using Improved Apriori Algorithm"*, International Journal for Innovative Research in Science.