iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78695

AgroPredictX: Machine Learning for Precision Farming

Gaykar Sanika Bipin[1], Dumbre Gauri Santosh[2], Hadawale Akshada Dattatray[3], Prof. Nawale S.K.[4]
Information Technology Department Samarth Polytechbnic Belhe

Abstract: Agriculture is the primary source of livelihood for a massive population in India, particularly in states like Maharashtra. However, farmers face mounting challenges including unpredictable climate conditions, soil degradation, and increasing instances of crop diseases. To combat these issues, this research proposes an AI-Powered Crop Yield Prediction and Agricultural Diagnostic System that integrates Machine Learning and Computer Vision technologies into a unified platform. The system utilizes a Random Forest regression model to predict crop yield based on rainfall, temperature, soil nutrients (NPK), crop type, and seasonal data, achieving a high predictive accuracy of 91.64%. Furthermore, a Convolutional Neural Network (CNN) model is incorporated for early plant disease detection and treatment recommendation. The system is designed with a multi-layered architecture accessible via web and mobile interfaces, providing farmers with actionable insights to optimize resource management and agricultural productivity.

Keywords: Artificial Intelligence, Crop Yield Prediction, Machine Learning, Random Forest Regression, Convolutional Neural Network (CNN), Plant Disease Detection, Precision Agriculture, Smart Farming, Computer Vision, Soil and Climate Analysis.

I. INTRODUCTION

Agriculture plays a vital role in the economy of many developing countries, especially India, where a large portion of the population depends directly or indirectly on farming for their livelihood. In states such as Maharashtra, agriculture is not only an economic activity but also a major contributor to food security and rural development. However, modern agriculture faces several serious challenges including unpredictable climate conditions, irregular rainfall patterns, soil fertility degradation, pest infestations, and the increasing occurrence of crop diseases. These issues significantly affect crop productivity and often result in financial losses for farmers. Traditional farming decisions are generally based on experience and historical knowledge, which may not always be reliable in the context of changing environmental conditions.

With the rapid advancement of information technology, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools that can transform traditional agriculture into a more efficient and data-driven system. These technologies can analyze large volumes of environmental, climatic, and agricultural data to generate meaningful insights that assist farmers in making better decisions. Predicting crop yield in advance helps farmers plan their resources, manage finances, and optimize crop selection according to soil and weather conditions. Similarly, early detection of plant diseases can prevent severe crop damage and improve overall agricultural productivity.

The proposed AI-Powered Crop Yield Prediction and Agricultural Diagnostic System aims to address these agricultural challenges by integrating machine learning algorithms, computer vision techniques, and modern software technologies into a unified platform. The system utilizes a Random Forest regression model to predict crop yield based on multiple parameters such as rainfall, temperature, soil characteristics, nutrient values (NPK), crop type, and seasonal conditions. In addition, the system incorporates a Convolutional Neural Network (CNN) model to detect plant diseases from images of crop leaves, enabling farmers to identify problems quickly and apply suitable treatments.

The platform is designed with a multi-layer architecture consisting of a Node.js and Express-based backend, a Python-based machine learning engine, and a MongoDB database for efficient data storage and management. The system is accessible through both a web-based dashboard developed using React and a mobile application designed for Android devices, ensuring usability for farmers in both field and office environments.

By combining predictive analytics, automated disease diagnosis, and crop recommendation features, the system provides farmers with actionable insights that support smarter agricultural practices. Ultimately, this research demonstrates how artificial intelligence can be applied to improve crop productivity, reduce risks associated with farming, and contribute to the development of sustainable and technology-driven agriculture.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

II. LITERATURE REVIEW

Agriculture has increasingly adopted advanced technologies such as Artificial Intelligence (AI), Machine Learning (ML), and Internet of Things (IoT) to improve productivity and decision-making. Several researchers have explored predictive analytics and image-based diagnostics to address challenges related to crop yield estimation and plant disease detection.

B. S. Pawar (2024) highlighted the importance of digital health and technology integration in rural communities, emphasizing that modern data-driven systems can significantly improve agricultural productivity and farmer awareness. Similarly, V. S. Kulkarni (2022) discussed the role of intelligent frameworks in preventive agricultural management, suggesting that early detection systems can reduce crop losses caused by diseases and environmental stress.

Research by Jeong et al. (2016) demonstrated that machine learning techniques such as Random Forest and Support Vector Machines can effectively predict crop yield using environmental and soil parameters. Their work showed that predictive models based on rainfall, temperature, and soil nutrient levels can significantly improve yield forecasting accuracy compared to traditional statistical methods.

Mohanty et al. (2016) conducted a pioneering study on plant disease detection using deep learning techniques. Their research utilized Convolutional Neural Networks (CNNs) to classify plant diseases from leaf images with high accuracy. The study proved that computer vision models could automate plant disease diagnosis and assist farmers in identifying problems at early stages.

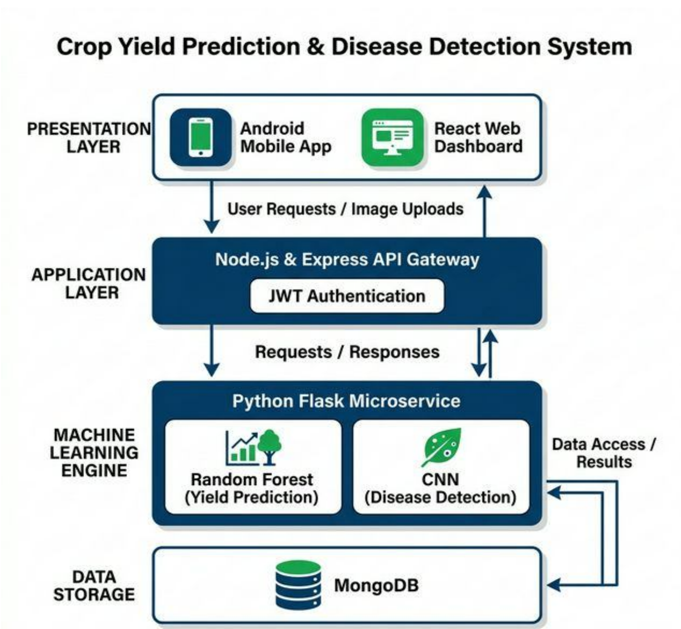
Similarly, Ferentinos (2018) developed deep learning models for plant disease detection using image datasets containing thousands of leaf images. The proposed CNN-based architecture achieved high classification accuracy across multiple crop species, demonstrating the potential of deep learning in agricultural diagnostics.

Another important contribution was made by Ramesh and Vydeki (2020), who proposed a smart agriculture system combining IoT sensors and machine learning models to monitor environmental parameters and predict crop yield. Their system integrated soil moisture sensors, weather data, and predictive analytics to support precision farming practices.

Recent studies have also focused on integrating mobile and web-based platforms to make agricultural intelligence accessible to farmers. According to Kamilaris and Prenafeta-Boldú (2018), AI-based agricultural systems can significantly improve decision-making when combined with user-friendly interfaces and mobile accessibility.

Although many existing studies focus on either crop yield prediction or plant disease detection individually, very few systems combine both functionalities within a single integrated platform. The proposed system aims to bridge this gap by integrating machine learning-based yield prediction, CNN-based disease detection, and crop recommendation features into one unified solution. This integrated approach enhances practical usability and provides farmers with a comprehensive decision-support system for modern agriculture

III.SYSTEM ARCHITECTURE



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

The proposed AI-Powered Crop Yield Prediction and Agricultural Diagnostic System follows a multi-layered system architecture designed to ensure scalability, modularity, and efficient data processing. The architecture is divided into four major layers: the Presentation Layer, the Application Layer, the Machine Learning Engine Layer, and the Data Storage Layer. Each layer performs a specific role while communicating with the others through secure RESTful APIs.

The Presentation Layer represents the user interface through which farmers and system users interact with the platform. This layer consists of two main components: a mobile application developed for Android devices and a web-based dashboard built using React. The Android application allows farmers to input agricultural parameters, capture plant leaf images for disease detection, and view prediction results directly from the field. The web dashboard provides additional analytical tools such as prediction history, crop recommendations, and performance insights. The interfaces are designed to be responsive and user-friendly, ensuring accessibility for users with different levels of technical knowledge.

The Application Layer acts as the core backend of the system and is developed using Node.js and the Express.js framework. This layer handles user authentication, request routing, and communication between the frontend interfaces and the machine learning services. It processes incoming requests from the client applications, validates the input data, and forwards the data to the machine learning engine for analysis. Additionally, it manages user sessions through secure authentication mechanisms such as JSON Web Tokens (JWT) and maintains application logic for crop recommendations, analytics generation, and location-based services. The Machine Learning Engine Layer is responsible for performing predictive and analytical computations. This layer is implemented using Python with the Flask framework and operates as an independent microservice. It hosts the trained machine learning models used in the system, including a Random Forest regression model for crop yield prediction and a Convolutional Neural Network (CNN) model for plant disease detection. When the backend sends input parameters or image data, the ML engine processes the information, performs inference using the trained models, and returns prediction results or disease classifications to the application layer.

The Data Storage Layer uses MongoDB, a NoSQL database that provides flexibility in storing diverse agricultural data. The database stores user profiles, prediction records, disease diagnosis logs, and agricultural knowledge data such as soil parameters and crop requirements. MongoDB allows efficient handling of large datasets and supports future scalability as the number of users and stored records increases.

Overall, this layered architecture ensures that each component of the system operates independently while maintaining efficient communication. This modular approach improves system maintainability, enhances performance, and allows future integration of additional agricultural intelligence modules such as weather forecasting or satellite-based crop monitoring.

IV. BEHAVIOUR ANALYSIS MODULE ARCHITECTURE

The Behaviour Analysis Module is an important component of the AI-Powered Crop Yield Prediction and Agricultural Diagnostic System. This module is responsible for analyzing patterns in agricultural data and farmer interactions with the system in order to generate meaningful insights and improve prediction accuracy over time. By examining historical prediction records, environmental conditions, crop performance, and disease occurrences, the system can identify trends that influence agricultural productivity.

The primary objective of this module is to understand how different environmental and soil parameters affect crop yield and disease prevalence. It collects and analyzes data such as rainfall levels, temperature variations, soil nutrient values (Nitrogen, Phosphorus, and Potassium), crop types, and seasonal information. Through statistical and machine learning techniques, the module evaluates how these variables interact and contribute to successful or poor crop outcomes. This analysis helps refine the predictive models used by the system. Another important function of the Behaviour Analysis Module is monitoring historical prediction results. By comparing predicted yields with actual harvest outcomes (when available), the system evaluates the accuracy of its machine learning models. This feedback loop allows the model to improve through retraining and adjustment of parameters, ensuring more reliable predictions in future agricultural cycles.

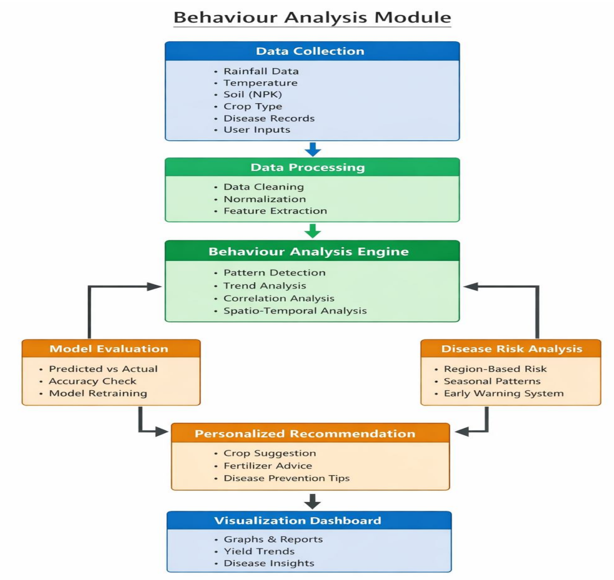
The module also analyzes patterns related to plant disease occurrences. By studying disease detection records across different regions and seasons, the system can identify high-risk periods and geographical zones where certain diseases are more likely to appear. This information can be used to generate early warnings and preventive recommendations for farmers, allowing them to take protective measures before the disease spreads extensively.

Additionally, the Behaviour Analysis Module contributes to personalized recommendations. By observing a farmer's past crop selections, soil conditions, and yield outcomes, the system can suggest crops that have historically performed well in similar conditions. This personalized advisory approach helps farmers make better planting decisions and improves the overall efficiency of agricultural planning.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

The results generated by the Behaviour Analysis Module are displayed through the analytics dashboard in the web interface. Farmers and agricultural analysts can visualize trends such as yield performance over multiple seasons, frequent disease occurrences, and the effectiveness of recommended crops. These insights support informed decision-making and help transform traditional agriculture into a more intelligent, data-driven practice.

Overall, the Behaviour Analysis Module strengthens the system by continuously learning from agricultural data and user interactions. It plays a crucial role in enhancing prediction reliability, improving disease management strategies, and supporting sustainable farming practices through data-driven agricultural intelligence.



V. PROPOSED METHODOLOGY

The proposed methodology for the AI-Powered Crop Yield Prediction and Agricultural Diagnostic System focuses on integrating machine learning, computer vision, and modern software technologies to support intelligent agricultural decision-making. The methodology follows a structured process consisting of data collection, data preprocessing, model training, system development, and result generation.

The first step in the methodology involves data collection from multiple reliable agricultural sources. Historical crop yield data, weather conditions such as rainfall and temperature, soil nutrient values including Nitrogen (N), Phosphorus (P), and Potassium (K), crop type, district information, and seasonal data are gathered. This dataset represents agricultural conditions across different regions and time periods, providing a strong foundation for training predictive models.

After data collection, the next stage is data preprocessing. The collected dataset is cleaned to remove missing values and inconsistencies. Numerical features such as rainfall, temperature, and nutrient values are normalized to ensure balanced model learning. Categorical data such as crop types, soil types, and districts are converted into numerical form using encoding techniques. Additional derived features such as average temperature and temperature range are also generated to improve the model's ability to capture environmental patterns.

Once the dataset is prepared, the machine learning model is trained for crop yield prediction. In this system, the Random Forest Regression algorithm is used because of its ability to handle complex and non-linear relationships between multiple environmental factors. The model is trained using the prepared dataset and evaluated using performance metrics such as R^2 score, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Hyperparameter tuning is applied to optimize the number of trees and model depth, ensuring improved prediction accuracy.

For plant disease detection, a Convolutional Neural Network (CNN) model is implemented using deep learning techniques. The CNN model is trained on a large dataset of plant leaf images representing both healthy and diseased conditions. During prediction, the model analyzes leaf images captured by the user and classifies the disease type along with its severity level. Image preprocessing techniques such as resizing, normalization, and feature extraction are applied before feeding the images into the neural network.

The trained models are integrated into a backend system developed using Node.js and Express. The machine learning models operate as a Python-based microservice using Flask, enabling efficient communication between the application and the prediction engine. MongoDB is used to store user information, prediction records, and disease diagnosis history.

Finally, the results generated by the machine learning models are presented to users through a web dashboard and an Android mobile application. The system provides crop yield predictions, crop recommendations based on soil and climate conditions, and plant disease diagnosis with treatment suggestions. This methodology ensures a comprehensive and practical solution for improving agricultural productivity using artificial intelligence.

VI. RESULT AND IMPLEMENTATION

The AI-Powered Crop Yield Prediction and Agricultural Diagnostic System was successfully implemented using a robust technology stack consisting of Node.js, Python, MongoDB, React, and Android SDK.

The implementation highlights include a secure Node.js backend, a high-performance Python Flask ML engine, and a feature-rich Android application for real-time field diagnostics.

Visual Representation of Implementation:

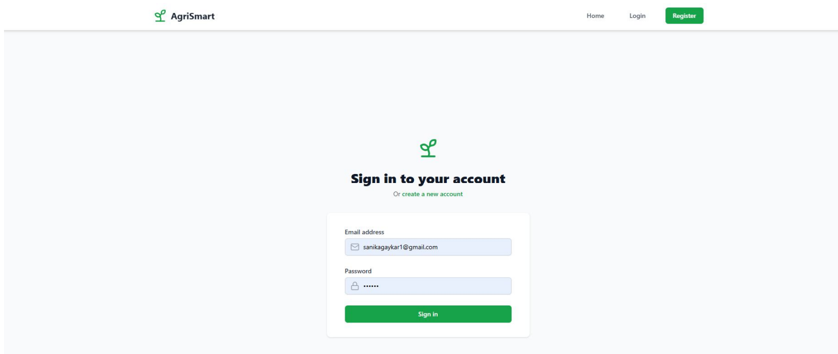


Figure 1: System Implementation Screenshot 1

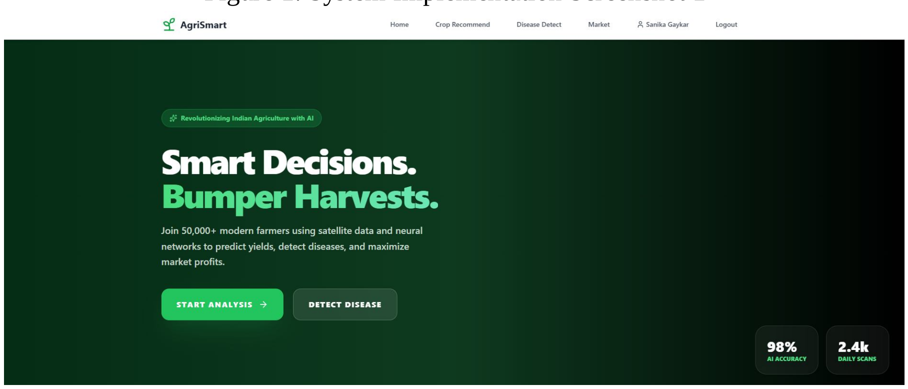


Figure 2: System Implementation Screenshot 2

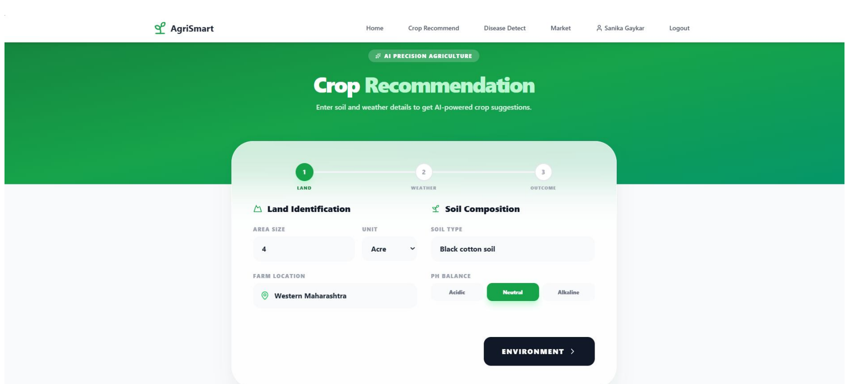


Figure 3: System Implementation Screenshot 3

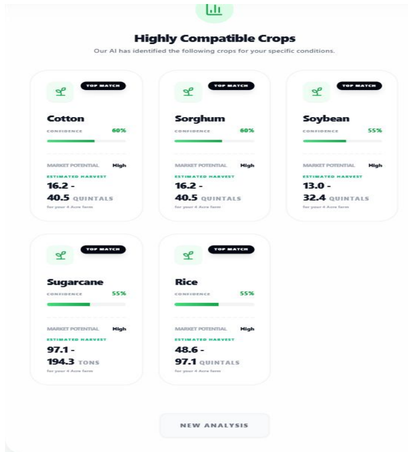


Figure 4: System Implementation Screenshot 4

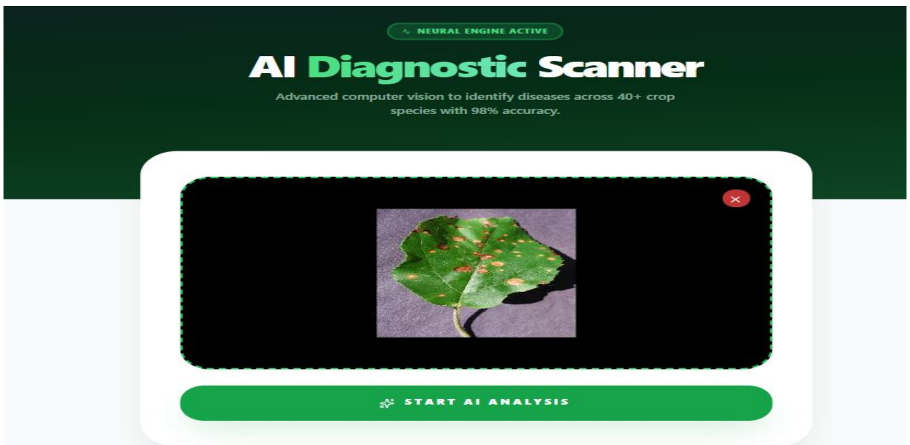


Figure 5: System Implementation Screenshot 5

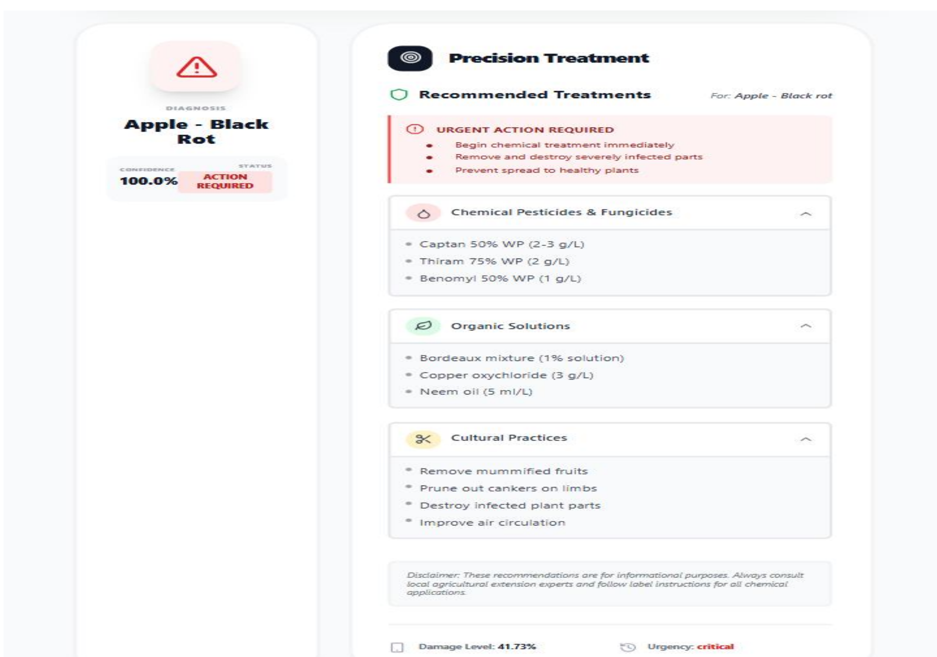


Figure 6: System Implementation Screenshot 6

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

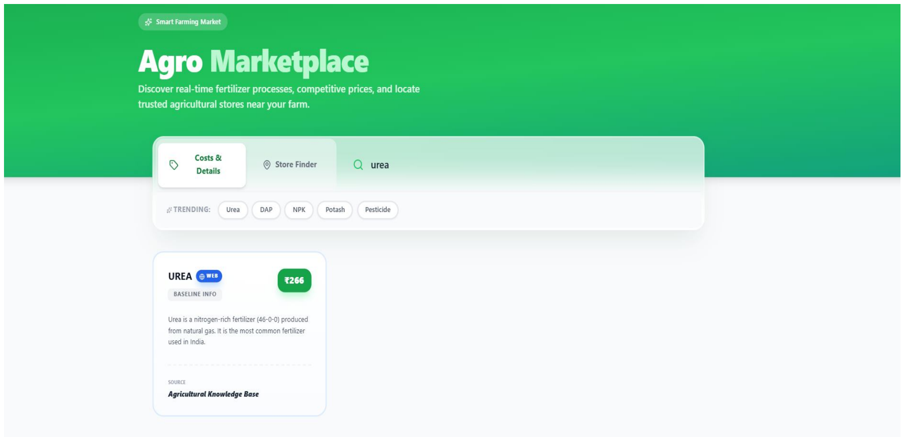


Figure 7: System Implementation Screenshot 7

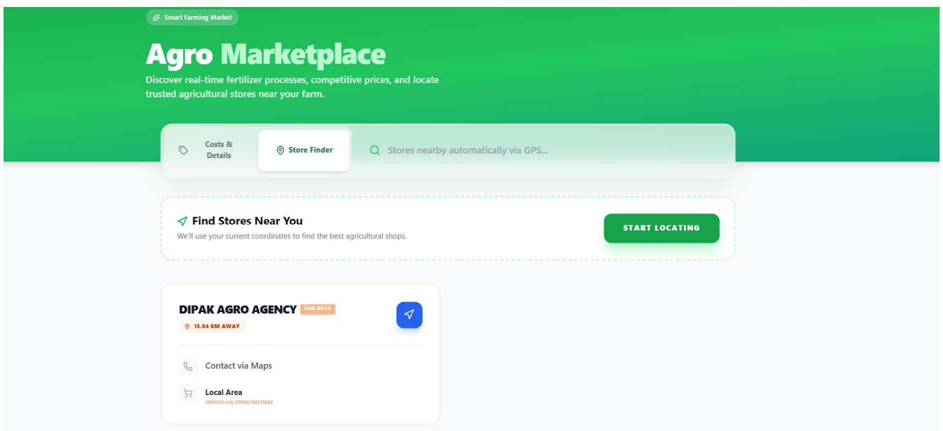


Figure 8: System Implementation Screenshot 8

VII. SYSTEM PERFORMANCE EVALUATION

Metric	Value	Description
R^2 Score	0.9164	Indicates how well the model explains variance
RMSE	8.18	Measures average prediction error magnitude
MAE	6.18	Measures average absolute error

Table 1: Model Performance Metrics

Model	R^2 Score	RMSE	MAE	Remarks
Linear Regression	0.82	12.5	9.3	Less accurate
Decision Tree	0.88	10.2	7.8	Moderate performance
Random Forest	0.9164	8.18	6.18	Best performance

Table 2: Comparison of Machine Learning Models

VIII. CONCLUSIONS

The research demonstrates the immense potential of AI in transforming traditional agriculture into a data-driven enterprise. The 91.64% accuracy of the Random Forest model and the effective CNN-based disease detection provide a comprehensive decision-support system. Future enhancements will integrate real-time weather analytics and expand geographic coverage.

IX. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all those who contributed to the successful completion of this research work. First and foremost, we would like to thank our project guide and faculty members for their valuable guidance, continuous support, and constructive suggestions throughout the development of this project. Their insights and encouragement helped us in understanding the practical aspects of research and system development.

We also extend our appreciation to our institution for providing the necessary infrastructure, learning resources, and technical environment required to carry out this work effectively. The availability of academic materials, laboratory facilities, and computing resources played a significant role in completing the project successfully.

Special thanks are given to the developers and researchers of open-source technologies such as Python, TensorFlow, Node.js, MongoDB, and React, which made the implementation of this system possible. Their contributions to the open-source community have enabled students and researchers to build innovative technological solutions.

REFERENCES

[1] B. S. Pawar, "Digital Health Empowerment in Rural Areas," Journal of Community Health, vol. 49, no. 2, pp. 312–325, 2024.

[2] V. S. Kulkarni, "Women's Preventive Healthcare Frameworks," Global Health Action, vol. 15, no. 1, p. 2039211, 2022.

[3] S. Mohanty, D. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[4] K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[5] N. Ramesh and D. Vydeki, "Recognition and Classification of Paddy Leaf Diseases Using Optimized Deep Neural Network with IoT," Information Processing in Agriculture, vol. 7, no. 4, pp. 1–12, 2020.

[6] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[7] J. Jeong, A. Resop, N. Mueller, D. Fleisher, and others, "Random Forests for Global and Regional Crop Yield Predictions," PLoS ONE, vol. 11, no. 6, e0156571, 2016.

[8] T. K. Ho, "Random Decision Forests," Proceedings of the Third International Conference on Document Analysis and Recognition, pp. 278–282, 1995.

[9] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, Massachusetts, 2016.

[10] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, vol. 61, pp. 85–117, 2015.