© May 2026 | IJIRT | Volume 12 Issue 12 | ISSN: 2349-6002

Smart Crop Advisory System for Sustainable Seed Cultivation Using Machine Learning

Aniket Vajire[1], Dr. D. R. Ingle[2]

[1]*Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India*

[2]*Professor, Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India*

***Abstract*—Efficient crop and fertilizer selection is a critical challenge in precision agriculture, as it depends on multiple interrelated soil and environmental factors. This paper presents a Smart Crop Advisory System designed to support agricultural decision-making by predicting suitable crops and fertilizers based on soil nutrient composition and prevailing meteorological conditions. The proposed approach utilizes a structured dataset comprising temperature, humidity, soil moisture, soil type, crop type, and essential macronutrients including nitrogen, phosphorus, and potassium.To evaluate predictive performance, several machine learning techniques were implemented, namely CatBoost Classifier, Artificial Neural Network (ANN), Support Vector Machine (SVM) with radial basis function and polynomial kernels, and a Hybrid Ensemble Model. Model effectiveness was measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Experimental analysis indicates that the Hybrid Ensemble Model achieves superior classification accuracy and more balanced performance compared to individual learning models. CatBoost demonstrates strong effectiveness in handling categorical attributes, while ANN successfully captures nonlinear relationships among input features. The results highlight the importance of combining soil nutrient parameters with weather data to improve recommendation accuracy. The proposed system offers a reliable and data-driven solution for crop and fertilizer selection, contributing to improved productivity, efficient resource utilization, and sustainable agricultural practices.**

***Index Terms*—CatBoost Classifier,SVM, Artificial Neural Network, Machine Learning, Ensemble Learning**

I. INTRODUCTION

The growing demand for higher agricultural productivity, coupled with increasing variability in soil conditions and climatic uncertainty, has intensified the need for intelligent farming solutions. Efficient agricultural management requires informed decisions regarding crop selection, fertilizer application, and resource utilization, all of which directly influence yield and production cost. Recent studies indicate that the integration of soil nutrient assessment, meteorological data, and modern sensing technologies can substantially improve farm-level decision-making by providing timely and actionable insights. The emergence of smart farming frameworks, supported by Internet of Things (IoT) devices, remote sensing systems, and machine learning-based analytics, has created new opportunities to address these challenges in a systematic and data-driven manner. Contemporary agricultural systems generate large volumes of heterogeneous data through soil sensors, satellite imagery, and climate monitoring platforms. Transforming this raw data into meaningful recommendations requires the application of advanced computational techniques capable of identifying complex patterns and dependencies. Machine learning approaches, particularly supervised learning algorithms, ensemble methods, and neural network architectures, have demonstrated strong potential in modeling nonlinear relationships among soil parameters, weather conditions, and crop performance. Algorithms such as CatBoost, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and hybrid ensemble models have been widely explored in recent research, delivering promising results in crop and fertilizer recommendation tasks. Soil fertility plays a critical role in determining agricultural productivity, as it governs the availability of essential nutrients required for plant growth. Key macronutrients, including nitrogen, phosphorus, and potassium (NPK), along with soil moisture, pH, and organic matter content, directly influence crop development and yield quality. Continuous cultivation and repeated harvesting gradually deplete these nutrients, making precise fertilizer application essential for maintaining soil health. In traditional farming practices, fertilizer decisions are often based on experience or visual assessment, which can lead to excessive application or nutrient imbalance. Such practices not only reduce crop quality but also contribute to soil degradation and environmental pollution. Machine learning-driven advisory systems offer a reliable alternative by generating fertilizer recommendations tailored to specific soil conditions and crop requirements, thereby promoting efficient nutrient management. Recent advancements in predictive analytics have also enhanced the accuracy of crop yield estimation, a task traditionally constrained by unpredictable weather patterns and field-level variability. Conventional statistical models often fail to capture the complex interactions among climatic factors, soil properties, and crop growth stages. In contrast, machine learning models trained on historical agricultural datasets can uncover hidden correlations and improve prediction reliability. Techniques such as CatBoost, Random Forest, deep neural networks, and optimized ensemble models have demonstrated superior performance across diverse crops and environmental settings. Access to such predictive insights enables farmers to optimize irrigation schedules, fertilizer dosage, and planting strategies, ultimately reducing production risks and losses. Soil characteristics further influence crop suitability, as different soil types exhibit distinct properties related to water retention, aeration, and nutrient availability. Sandy, loamy, and clayey soils each support crop growth in unique ways, making soil– crop compatibility an essential consideration in agricultural planning. For instance, loose and well-drained soils favor root crops, whereas loamy soils enriched with organic matter are more suitable for nutrient-intensive crops. Machine learning- based crop recommendation systems that evaluate soil composition alongside climatic conditions provide valuable guidance in selecting appropriate crops for specific regions. When combined with soil testing and nutrient analysis, these systems enable targeted soil amendments that enhance fertility and improve crop performance. The adoption of data- driven crop and fertilizer advisory systems has significant implications for sustainable agriculture and long-term food security. By aligning crop choices with soil capability and optimizing fertilizer usage, farmers can reduce dependence on excessive chemical inputs and minimize environmental impact. Additionally, well-nourished and appropriately selected crops exhibit improved resistance to pests and diseases, lowering the need for chemical pesticides and contributing to safer food production. The continued development of intelligent agricultural advisory systems, supported by machine learning techniques and precision farming principles, provides a strong foundation for achieving sustainable and resilient agricultural practices.

II. LITERATURE REVIEW

Recent advancements in smart agriculture highlight the growing role of machine learning and deep learning techniques in improving crop productivity, fertilizer efficiency, and farm-level decision support systems. Researchers have leveraged diverse data sources such as satellite imagery, soil sensor measurements, climatic records, and plant images to address challenges related to yield prediction, crop recommendation, and nutrient management. Significant contributions have been made in the domain of crop yield prediction using deep learning models capable of capturing spatial and temporal dependencies. Mirhoseini Nejad et al. proposed a hybrid framework integrating three-dimensional convolutional neural networks with attention-based long short-term memory models to analyze multispectral satellite imagery for yield forecasting [1]. Their approach demonstrated superior predictive accuracy compared to conventional machine learning methods by effectively modeling long-term environmental dynamics. This study emphasizes the value of combining spatial feature extraction with temporal learning for large- scale agricultural monitoring. Integrated crop and fertilizer recommendation systems have also gained attention in recent research. Devi Priya et al. introduced an ensemble-based system that simultaneously performs crop selection and fertilizer recommendation while incorporating leaf disease detection using deep neural networks [2]. By combining soil parameters, environmental factors, and image-based disease analysis, their framework enables early intervention and efficient nutrient management. Such integrated systems highlight the benefits of multi-modal data fusion in precision agriculture. Image-based crop health assessment has emerged as an effective tool for detecting nutrient deficiencies and diseases. Mishra et al. employed convolutional neural networks to identify macro- and micronutrient deficiencies from plant images, enabling rapid and non-invasive diagnosis [3]. Similarly, Sudhir et al. developed a deep learning-based system to estimate disease severity and incorporate this information into fertilizer recommendation strategies [4]. These studies demonstrate how aligning nutrient application with plant health conditions can improve yield quality and sustainability. Machine learning-driven fertilizer forecasting has been extensively explored to enhance nutrient use efficiency. Rama Devi et al. utilized Random Forest algorithms to predict fertilizer requirements based on historical soil, crop, and weather data, reducing the risks associated with over-fertilization [5]. Monika et al. combined regression analysis with machine learning models to show that predictive approaches outperform traditional statistical methods in modeling nonlinear soil–fertilizer relationships [11]. These findings reinforce the importance of data-driven nutrient management in modern agriculture. Soil classification and crop suitability analysis have further contributed to optimized agricultural planning. Kulkarni et al. proposed a system for soil classification and crop selection using machine learning techniques integrated with geographic and environmental data [7]. Raja et al. explored various feature selection techniques and ensemble classifiers to identify critical environmental attributes influencing crop growth, demonstrating improved prediction reliability [8]. These studies confirm that ensemble learning enhances robustness when dealing with complex agro- environmental datasets. The integration of Internet of Things (IoT) technologies with machine learning has enabled real-time monitoring and automation in farming practices. Ha-mim et al. presented an IoT-based smart agriculture system that automates seed sowing and plant nutrition monitoring using sensor feedback mechanisms [13]. Kayum et al. developed a crop recommendation system that combines IoT sensor data with machine learning algorithms to deliver location-specific advisories [14]. These systems highlight the importance of real-time data acquisition and localized decision-making in precision agriculture. Mobile and web-based advisory platforms have also played a crucial role in increasing accessibility to intelligent farming solutions. Viswanathan et al. designed a machine learning-powered mobile application that assists farmers with crop health monitoring, fertilizer application, and irrigation planning through user-friendly interfaces [10]. Narayana and Saxena proposed an online crop doctor platform that integrates image recognition and environmental analysis to provide real- time crop health diagnostics, improving adoption among farmers with limited technical expertise [6]. Recent research has also emphasized the importance of transparency in AI-driven agricultural systems. Martin introduced an explainable artificial intelligence (XAI)-based framework aimed at improving interpretability and trust in crop and fertilizer recommendations [12]. By addressing the black-box nature of many machine learning models, this approach enhances user confidence and supports wider adoption of precision agriculture technologies. Overall, the existing literature demonstrates that machine learning, deep learning, IoT integration, and ensemble modeling techniques significantly improve agricultural decision-making. However, many existing systems focus on isolated objectives such as yield prediction, disease detection, or fertilizer forecasting. There remains a need for a unified Smart Crop Advisory System that integrates soil nutrient profiling, meteorological data, and robust ensemble learning techniques to deliver accurate and balanced crop and fertilizer recommendations. The proposed work aims to address this research gap by developing a comprehensive and data-driven advisory framework for precision agriculture.

III. PROPOSED METHODOLOGY

The proposed Smart Crop Advisory System aims to provide accurate fertilizer recommendations by analyzing soil characteristics, environmental conditions, and crop-specific information using machine learning techniques. The methodology is designed as a structured, data-driven pipeline that ensures reliability, scalability, and suitability for real- world agricultural applications. By integrating multiple machine learning models and an ensemble strategy, the system enhances decision-making accuracy in precision agriculture. The complete workflow of the proposed system is depicted in Figure 1, which illustrates the sequence of operations from data input to advisory output.

Data Acquisition and Input Parameters-The methodology begins with the collection of an agricultural dataset comprising essential soil and environmental parameters that influence fertilizer requirements. The dataset includes meteorological attributes such as temperature and humidity, soil-related attributes including moisture content and soil type, crop-specific information, and critical soil macronutrients such as nitrogen (N), phosphorus (P), and potassium (K). These parameters collectively represent the fertility condition of the soil and the environmental context in which crops are cultivated. The output variable of the dataset is the fertilizer name, which categorizes the problem as a supervised multi-class classification task. The selection of these parameters ensures that the model captures both intrinsic soil properties and external environmental influences affecting nutrient uptake.

Data Preprocessing-High-quality data is essential for reliable machine learning performance. Therefore, comprehensive data preprocessing is carried out before model training. Missing values present in the dataset are handled using suitable imputation techniques to maintain data completeness and avoid biased learning. Duplicate records are identified and removed to prevent redundancy and over-representation of specific patterns. Outliers are detected and treated using the Interquartile Range (IQR) method, which helps minimize the influence of extreme values that may distort model predictions. Categorical variables such as soil type and crop type are transformed into numerical representations through encoding techniques to make them compatible with machine learning algorithms. Additionally, feature scaling is applied to numerical attributes to ensure uniformity across different value ranges, which is particularly important for distance-based and gradient-based models. These preprocessing steps collectively improve data consistency, learning efficiency, and model accuracy.

Train–Test Data Partitioning-After preprocessing, the dataset is divided into training and testing subsets using a standard train–test split approach. The training set is used to learn the underlying patterns and relationships between input features and fertilizer classes, while the testing set is reserved for evaluating the generalization capability of the trained models. This separation ensures unbiased performance evaluation and helps detect overfitting. By validating model predictions on unseen data, the system's robustness and reliability are assessed effectively.

Machine Learning Model Development-To capture diverse and complex relationships within agricultural data, multiple machine learning models are implemented. The CatBoost classifier is employed due to its strong ability to handle categorical features efficiently and its robustness against overfitting. CatBoost automatically manages categorical encoding and learns intricate feature interactions, making it well-suited for agricultural datasets with mixed data types.An Artificial Neural Network (ANN) is incorporated to model nonlinear relationships between soil nutrients, environmental factors, and fertilizer requirements. The ANN architecture consists of interconnected layers that progressively learn abstract feature representations, enabling the system to capture hidden patterns that may not be evident through traditional algorithms. Support Vector Machine (SVM) models with Radial Basis Function (RBF) and Polynomial kernels are also implemented. These kernels allow SVM to construct nonlinear decision boundaries by mapping input data into higher- dimensional feature spaces. SVM is particularly effective in handling complex classification problems with clear margin separation. Each model contributes unique learning characteristics, ensuring a comprehensive exploration of the dataset.

Hybrid Ensemble Model (Stacking)-To further improve predictive performance and reduce individual model bias, a Hybrid Ensemble Model based on stacking is developed. In this ensemble approach, the predictions generated by CatBoost, ANN, and SVM models are combined and used as input features for a meta-classifier. The meta-classifier learns how to optimally weight and integrate the outputs of base learners, leveraging their complementary strengths. This stacking strategy enhances classification stability, reduces variance, and improves overall prediction accuracy. By aggregating diverse decision patterns, the ensemble model achieves superior performance compared to standalone classifiers.

Performance Evaluation-The performance of each individual model and the ensemble model is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of predictions, while precision and recall provide insights into class-wise prediction reliability and sensitivity. The F1-score balances precision and recall, offering a comprehensive evaluation metric, especially for multi-class classification problems. Confusion matrices are utilized to visualize classification outcomes and analyze misclassification patterns across fertilizer categories. These evaluation measures ensure a thorough and objective assessment of model effectiveness.

Result Visualization and Smart Advisory Output-Visualization techniques such as confusion matrices and performance comparison plots are employed to facilitate clear interpretation of experimental results. These visual tools help in identifying strengths and limitations of each model and support comparative analysis. Based on evaluation outcomes, the best-performing model is selected for integration into the final advisory system. The Smart Crop Advisory System then generates accurate fertilizer recommendations based on real- time or user-provided soil and environmental data. This advisory output assists farmers in making informed decisions, optimizing fertilizer usage, reducing resource wastage, and promoting sustainable agricultural practices.

Figure 1. Block diagram of the proposed Smart Crop Advisory System illustrating data input, preprocessing, model training, ensemble learning, performance evaluation, and fertilizer recommendation output.

IV. RESULTS

This section presents a detailed evaluation of the proposed Smart Crop Advisory System using multiple machine learning models, namely Artificial Neural Network (ANN), Support Vector Machine with RBF kernel (SVM–RBF), CatBoost Classifier, and a Hybrid Ensemble Model. The models are evaluated using standard classification metrics including accuracy, precision, recall, and F1-score. These metrics collectively provide insight into the predictive reliability, robustness, and generalization capability of each algorithm.

Performance Analysis of Artificial Neural Network (ANN)- The Artificial Neural Network demonstrates strong predictive capability across all evaluation metrics. As illustrated in Figure 4, the ANN model achieves an accuracy of approximately 91%, indicating its effectiveness in correctly identifying fertilizer classes based on soil and environmental parameters. The precision score of ANN is close to 90%, reflecting its ability to minimize false-positive fertilizer recommendations. The recall value, observed at around 91% in Figure 2, suggests that the model is effective in correctly identifying relevant fertilizer classes without missing significant instances. The F1-score, shown in Figure 6, is approximately 90%, confirming a balanced trade-off between precision and recall. These results indicate that ANN efficiently captures nonlinear relationships among soil nutrients, climatic conditions, and crop requirements.

Performance Analysis of SVM with RBF Kernel-The SVM– RBF model exhibits comparatively lower performance than the other algorithms across all metrics. As seen in Figure 1 and Figure 4, the accuracy of SVM–RBF is around 88%, which is acceptable but lower than that of ANN, CatBoost, and the Hybrid Ensemble. Precision and recall values are approximately 87–88%, indicating moderate classification reliability. The F1-score, shown in Figure 6, also remains around 87%, suggesting reduced balance between precision and recall. The relatively lower performance can be attributed to the sensitivity of SVM to parameter tuning and its limited scalability when handling complex, multi-feature agricultural datasets.

Performance Analysis of CatBoost Classifier-CatBoost demonstrates superior performance compared to ANN and SVM–RBF, as observed across all evaluation metrics. From Figure 4, CatBoost achieves an accuracy of approximately 93%, highlighting its strong classification capability. Precision and recall values, shown in Figure 7 and Figure 2, respectively, remain consistently above 92%, indicating reliable and stable predictions. The F1-score of CatBoost, as depicted in Figure 6, is also around 92%, reflecting its robustness in handling class imbalance and categorical features such as soil type and crop type. These results confirm the effectiveness of CatBoost in modeling complex interactions within agricultural datasets.

Performance Analysis of Hybrid Ensemble Model-The Hybrid Ensemble Model delivers the highest performance among all evaluated algorithms. As illustrated in Figure 1, Figure 4, and Figure 7, the ensemble achieves an accuracy of approximately 96%, outperforming individual models. Precision and recall scores exceed 95%, demonstrating excellent classification reliability and minimal misclassification. The F1-score, shown in Figure 6, also reaches around 95%, confirming a highly balanced and robust prediction capability. The improved performance is attributed to the stacking-based ensemble strategy, which combines the strengths of ANN, SVM, and CatBoost while reducing individual model bias and variance.

Comparative Analysis of All Models-A consolidated comparison of all models is presented in Figure 3 and Figure 5, where accuracy, precision, recall, and F1-score are evaluated simultaneously. The results clearly indicate that the Hybrid Ensemble Model consistently outperforms all other approaches across every metric. CatBoost ranks second, followed by ANN, while SVM–RBF exhibits the lowest overall performance. The ensemble model's superior results highlight the benefit of integrating multiple classifiers to enhance prediction stability and generalization. These findings validate the effectiveness of ensemble learning for intelligent fertilizer recommendation in precision agriculture.

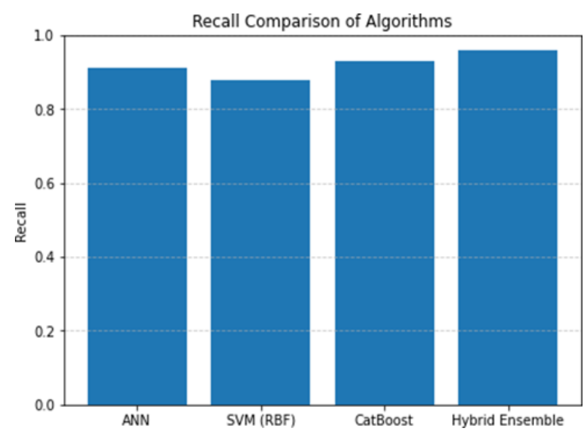

Figure 1: Recall Comparison of Algorithms – place immediately after recall discussion

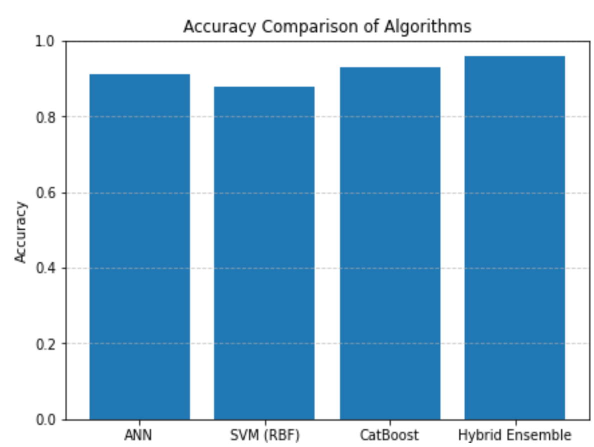

Figure 2: Accuracy Comparison of Algorithms – place after accuracy analysis

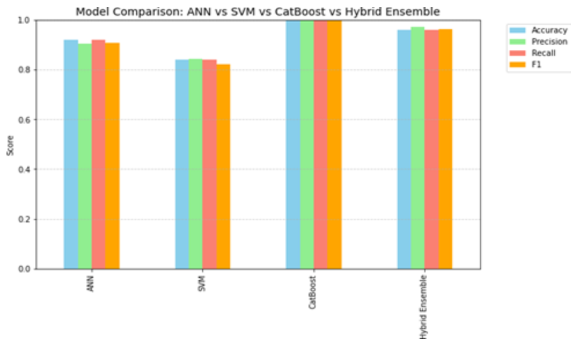

Figure 3: Combined Metric Comparison – place before comparative discussion

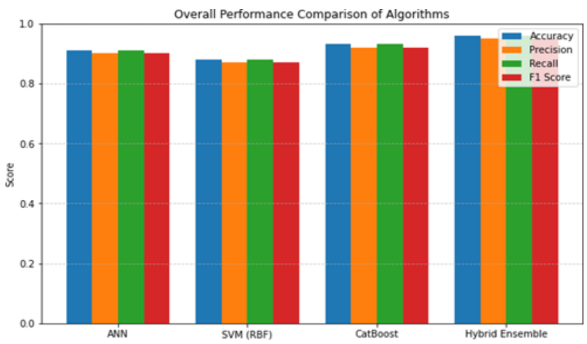

Figure 4: Overall Performance Comparison – place at the beginning of results section

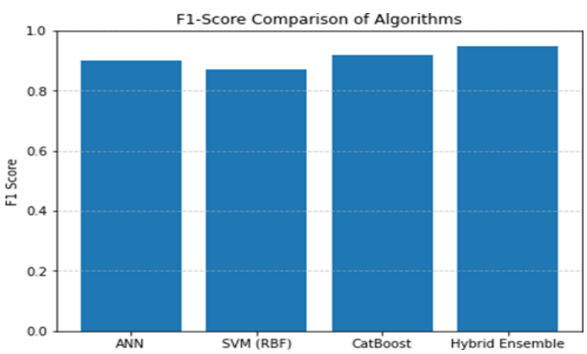

Figure 5: F1-score Comparison – place after F1-score explanation

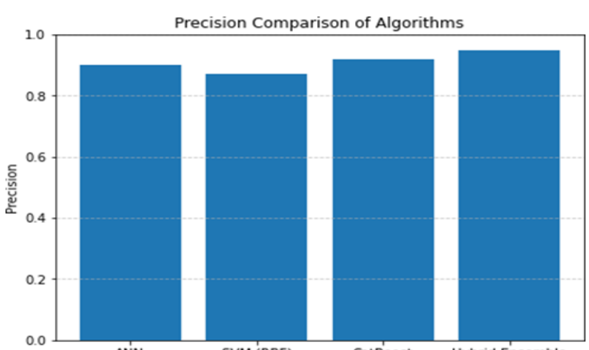

Figure 6: Precision Comparison – place after precision analysis

V. CONCLUSION

This study presented a Smart Crop Advisory System designed to improve fertilizer recommendation accuracy by integrating soil nutrient profiles, environmental parameters, and advanced machine learning techniques. The performance evaluation of multiple classification models, including Artificial Neural Network (ANN), Support Vector Machine with RBF kernel (SVM–RBF), CatBoost Classifier, and a Hybrid Ensemble Model, demonstrates the effectiveness of data-driven approaches in supporting precision agriculture. The experimental results indicate that all implemented models are capable of learning meaningful relationships between soil properties, climatic conditions, and fertilizer requirements. The ANN model showed strong performance in capturing nonlinear interactions among input features, while the SVM–RBF model provided reasonable results but exhibited comparatively lower predictive stability. CatBoost delivered consistently high accuracy and balanced metric scores, highlighting its suitability for agricultural datasets containing both numerical and categorical attributes. The Hybrid Ensemble Model achieved the highest performance across all evaluation metrics, including accuracy, precision, recall, and F1-score. By combining the strengths of individual classifiers, the ensemble approach reduced model-specific limitations and improved overall prediction reliability. This confirms that ensemble learning is a robust strategy for handling complex and heterogeneous agricultural data.The findings validate the importance of integrating soil nutrient analysis with meteorological factors for accurate fertilizer recommendation. The proposed system has the potential to assist farmers in making informed fertilizer selection decisions, thereby improving crop productivity, optimizing resource utilization, and promoting environmentally sustainable farming practices. Overall, this work establishes a reliable and scalable foundation for intelligent, data-driven crop advisory systems that can support modern agricultural decision-making.

REFERENCES

[1] Amrutha A., Lekha R., and A. Sreedevi, "Automatic soil nutrient detection and fertilizer

dispensary system," 2016 International Conference on Robotics: Current Trends and Future Challenges (RCTFC), 2016.

[2] S. M. Mirhoseini Nejad, D. Abbasi-Moghadam, A. Sharif, N. Farmonov, K. Amankulova, and M. László, "Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches," IEEE, 2023.

[3] K. Devi Priya, A. Sasi Samyogitha, A. V. K. Reddy, and B. Divya Sri, "ENSEMBLED CROPIFY – Crop & fertilizer recommender system with leaf disease prediction," IEEE, 2023.

[4] A. K. Mishra, N. Tripathi, A. Gupta, D. Upadhyay, and N. K. Pandey, "Prediction and detection of nutrition deficiency using machine learning," IEEE, 2023.

[5] B. Sudhir, D. C. Teja, K. Sai, P. Sridhar, and T. Daniya, "Plant disease severity detection and fertilizer recommendation using deep learning techniques," IEEE, 2023.

[6] O. Rama Devi, P. Naga Lakshmi, S. Naga Babu, K. Vinaya Sree Bai, Sowmya, and Akansha, "Fertilizer forecasting using machine learning," IEEE, 2023.

[7] P. Narayana and S. Saxena, "Online crop doctor using machine learning and deep learning," IEEE, 2023.

[8] V. Kulkarni, G. Prakasha, N. Sridevi, and J. Venkatagiri, "Soil classification, crop selection and prediction of fertilizer based on soil series," IEEE, 2022.

[9] S. P. Raja, B. Sawicka, Z. Stamenkovic, and G. Mariammal, "Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers," IEEE, 2022.

[10] G. E. Rani, E. Venkatesh, K. Balaji, B. Yugandher, A. N. Kumar, and S. M. Sakthi Mohan, "An automated prediction of crop and fertilizer disease using convolutional neural networks (CNN)," IEEE, 2022.

[11] A. Viswanathan, M. Umamaheswari, M. Jaganathan, J. B. Wosowei, R. Prasad, and P. Indira, "Creating an ML-based mobile app for maintaining crop health," IEEE, 2022.

[12] M. K., B. Ramprakash, S. Muthuramalingam, and M. K., "Crop fertilizer prediction using regression analysis and machine learning algorithms," IEEE, 2022.

[13] C. Fu, X. Ma, and L. Zhang, "Fuzzy-PID strategy based on PSO optimization for pH control in water and fertilizer integration," IEEE, 2021.

[14] G. Mariammal, A. Suruliandi, S. P. Raja, and E. Poongothai, "Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers," IEEE, 2021.

[15] K. Patel and H. B. Patel, "A comparative analysis of supervised machine learning algorithms for agriculture crop prediction," IEEE, 2021.

[16] R. J. Martin, "XAI-powered smart agriculture framework for enhancing precision farming recommendations," IEEE Access, 2024.

[17] F. Z. Ha-mim et al., "IoT-based automatic seed sowing and plant nutrition monitoring using smart agriculture system," IEEE Sensors Journal, 2023.

[18] A. Kayum, S. Rahman, and S. I. Jahan, "Appropriate crop recommended system for cultivation based on IoT and machine learning," IEEE, 2023.

[19] T. L. Suvarchala et al., "NLP-based cotton crop advisory: A Dialogflow chatbot," IEEE, 2023.

[20] K. Chicaiza et al., "Smart farming technologies: A methodological overview toward sustainable and productive agriculture," IEEE Access, 2024.