Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 17s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Precision Agriculture: A Machine Learning Approach to Enhance Crop Selection

Muralidhara B K[1], Melwin D Souza[1*], Vinod R[1], Pramod R[1], Krutik K M[1], Prajwal K Bandi[1], Prameela[1]

[1]Department of Computer Science and Engineering, Moodlakatte Institute of Technology Kundapura, India

*Corresponding Author:

Email ID: mellumerdy@gmail.com

Cite this paper as: Muralidhara B K, Melwin D Souza, Vinod R, Pramod R, Krutik K M, Prajwal K Bandi, Prameela, (2025) Precision Agriculture: A Machine Learning Approach to Enhance Crop Selection. *Journal of Neonatal Surgery*, 14 (17s), 221-231.

ABSTRACT

This research work introduces an advanced crop recommendation system utilizing machine learning to assist farmers in choosing optimal crops based on specific environmental and soil conditions. By employing algorithms such as Random Forest, Logistic Regression, XGBoost, and Gaussian Naive Bayes, the system evaluates key parameters including nitrogen, phosphorus, potassium, pH, temperature, humidity, and rainfall. Achieving an accuracy of 96%, the ensemble model demonstrates its capability to deliver reliable crop predictions. A user-friendly web application enables farmers to input local conditions and receive customized recommendations, empowering them to make informed decisions that enhance productivity and address economic challenges. This innovation highlights the transformative role of machine learning in agriculture, promoting smarter, data-driven practices. Ultimately, the system aims to boost crop yields, minimize losses, and support sustainable agricultural development.

Keywords: *Ensemble, Crop Prediction, Machine Learning, Robustness, Soil Parameters, Random Forest*.

1. INTRODUCTION

Agriculture is a crucial foundation for numerous developing nations, serving not only as a source of livelihood for a significant segment of the population but also contributing importantly to the overall economic development of these countries. Despite its importance, this sector faces various complex challenges such as soil degradation, irregular weather conditions, and inefficient farming methods, all of which hinder optimal crop production. Given the projections for a growing global population and the reduction of available arable land, it is critical to implement innovative, data-driven strategies that empower farmers to make well-informed decisions regarding their agricultural practices. Soil health plays an essential role in determining crop productivity. However, it often suffers due to poor nutrient management and the over-reliance on chemical fertilizers, leading to significant long-term damage to soil quality. Recent research promotes the adoption of sustainable agricultural practices that blend organic and inorganic approaches to soil fertility, demonstrating effectiveness in boosting crop yields while maintaining long-term soil health.

To address these urgent issues, this project presents an Advanced Crop Recommendation System (ACRS) that utilizes machine learning techniques to guide farmers in selecting the most appropriate crops tailored to their unique soil and environmental conditions. The ACRS incorporates essential factors such as nitrogen, phosphorus, and potassium levels, soil pH, temperature, humidity, and rainfall patterns to provide customized crop recommendations. Such insights enable farmers to make more informed choices, potentially enhancing their yield and resource efficiency. This system is powered by an array of robust machine learning algorithms, including Random Forest, Logistic Regression, Gradient Boosting Machines (XGBoost), and Gaussian Naive Bayes, which process input data and generate accurate, localized predictions for crop suitability. By regularly updating recommendations through the incorporation of new data, the system ensures that farmers receive the latest and most relevant information corresponding to their specific agricultural contexts.

The ACRS is designed as an intuitive web application, allowing farmers to easily enter their local soil and environmental data. Through this platform, farmers are provided with personalized crop recommendations that not only facilitate better crop choices but also offer guidance on efficient management of vital resources such as water and fertilizers. This data-driven methodology empowers farmers to adopt sustainable practices that boost productivity while reducing waste and resource consumption. The project underscores the transformative capabilities of machine learning within the agricultural domain, providing farmers with essential tools to enhance their decision-making processes. By offering tailored insights into crop selection, irrigation strategies, and nutrient application, this system aims to minimize crop loss, increase agricultural productivity, and encourage sustainable farming practices. Such advancements represent a significant milestone in bolstering food security and fostering the prosperity of future generations of farmers as they navigate a continuously evolving agricultural landscape.

1.1 Key Objectives

- **Develop a Comprehensive Dataset**: Create, preprocess, and refine a dataset that is compatible with machine learning, incorporating essential variables such as nitrogen, phosphorus, potassium levels, pH, temperature, humidity, rainfall, and other environmental conditions to ensure precise crop recommendations.
- **Implement and Evaluate Machine Learning Models**: Carry out the implementation and assessment of traditional machine learning algorithms—such as Logistic Regression and Gaussian Naive Bayes—as well as ensemble techniques like Random Forest and XGBoost, to establish a foundational framework for effective crop predictions.
- **Improve Prediction Accuracy with a Voting Classifier**: Incorporate a voting classifier that combines the outputs of various algorithms, including Random Forest, Logistic Regression, and Gaussian Naive Bayes, to bolster the reliability and balance of crop predictions by harnessing the advantages of each model.
- **Encourage Sustainable Farming Practices**: Apply data-driven insights to increase crop productivity and reduce agricultural losses, thereby enabling farmers to adopt sustainable methods and enhance resource efficiency in the face of fluctuating environmental conditions.

1.2 Key Contributions

- **Introduction of an Advanced Crop Recommendation System**: Development of an innovative Crop Recommendation System (ACRS) that utilizes machine learning methods to provide customized insights, allowing farmers to make informed decisions on crop selection based on specific soil and environmental conditions.
- **Creation of a User-Friendly Web Application**: Design of an accessible web application that simplifies the input of local environmental and soil data, ensuring that farmers can easily obtain personalized and pertinent crop recommendations.
- **Real-Time Data Update Capability**: Establishment of a mechanism that continuously updates predictions based on real-time data inputs, equipping farmers with the most current and accurate recommendations tailored to their individual farming scenarios.
- **Enhancement of Food Security**: Demonstration of the transformative impact of machine learning technologies in agriculture, contributing to improved food security and reinforcing the resilience of future generations of farmers as they adapt to an ever-changing agricultural landscape.

By concentrating on these objectives and contributions, the project aims to further the application of technology within agriculture and promote sustainable farming practices.

2. LITERATURE REVIEW

Agriculture faces numerous challenges, including changing climatic conditions, soil degradation, and inefficient farming practices, all of which negatively impact crop yields. The adoption of machine learning (ML) techniques has begun to transform the agricultural sector by providing data-driven solutions that enhance the decision-making process related to crop selection.

For instance, Geetha et al. [1] demonstrated the effectiveness of the Random Forest algorithm for predicting crop yields, achieving a notable accuracy of 95%. Their research addresses the issues caused by climatic variability, such as droughts and rising temperatures, which threaten agricultural productivity. By examining extensive datasets that link biophysical changes to crop yield, this study illustrates the ability of the Random Forest algorithm to process complex agricultural data and its potential benefits for precision agriculture. Such insights are vital for developing an advanced crop recommendation system, underscoring the importance of data-informed approaches in contemporary agricultural practices.

In addition, Jadhav et al. [2] developed a machine learning-based crop recommendation system specifically to enhance agricultural practices across India. Their findings reveal that traditional farming methods often lead to inefficiencies, as farmers frequently rely on market trends rather than essential factors such as soil health. Their evaluation of various algorithms resulted in high accuracy rates, with Random Forest achieving 98.93% and XGBoost 98.18%. This research highlights how machine learning can significantly improve crop recommendations, enabling farmers to make better-informed decisions and enhance their yield—objectives that align closely with our project focused on utilizing advanced algorithms for precise farming.

Pande et al. [3] addressed similar inefficiencies in agriculture by proposing a system that employs a range of machine learning algorithms, including Support Vector Machine (SVM), Artificial Neural Network (ANN), and Multivariate Linear Regression (MLR), to recommend specific crops and predict yields based on factors like soil type. The study found that the Random Forest algorithm was particularly effective, achieving an accuracy of 95%. By incorporating GPS technology to optimize the timing of fertilizer application, this research illustrates how machine learning can promote sustainable farming practices. Furthermore, Shariff et al. [4] also highlighted the role of machine learning in crop recommendation systems, showcasing multiple successful algorithms—including K-Nearest Neighbors (KNN), Decision Tree, and Gradient Boosting—designed to analyze soil conditions and recommend appropriate crops, with Random Forest excelling at an accuracy of 98%.

Lokhande and Dixit [5] contributed to this field with the creation of a web-based machine learning system that evaluates important parameters such as temperature, rainfall, and soil pH. Their findings indicate that the Logistic Regression model achieved an accuracy of 95.22%, demonstrating how machine learning can help fill the knowledge gaps for farmers regarding soil composition and environmental conditions, ultimately enhancing agricultural productivity.

The application of machine learning techniques in agriculture has led to the development of innovative crop recommendation systems aimed at helping farmers optimize their crop yields. In this regard, Saranya and Mythili [6] created a system for classifying soil types and recommending suitable crops, illustrating the value of data-driven decision-making in agriculture. Similarly, Kale and Mohapatra [7] introduced a machine learning-based crop recommendation system that utilizes advanced analytics to improve farming practices and enhance crop selection processes. Additionally, research by Gosai, Patel, and Patel [8] highlighted the effectiveness of predictive modeling in evaluating various agronomic factors, assisting farmers in making informed choices regarding their crops. Collectively, these studies emphasize the importance of integrating machine learning in agricultural practices to support better decision-making.

Further studies have demonstrated the versatility of various machine learning algorithms in predicting crop outcomes and refining recommendations. Suresh, Ganesh Kumar, and Ramalatha [9] employed K-means and Modified KNN algorithms to predict major crop yields in Tamil Nadu, showcasing how local contexts can benefit from these technological solutions. Morshed, Dutta, and Aryal ([10] focused on using semantic machine learning to provide recommendations based on environmental knowledge, illustrating the potential for data-driven insights to inform crop planning. Additionally, studies [11][12] investigated how existing crop models can be utilized to assess genetic variability in yields under stressful conditions, further demonstrating the relevance of machine learning in fostering resilient agricultural methods. Use of IoT and AI to detect disease and suggest medicine is most required initiative in Crop yield research [13]. Many Machine learning and deep learning technologies provide the ideas to develop integrated model for good result [14][15]

As the field continues to evolve, innovative methodologies are emerging that enhance crop recommendation systems. For example, Veenadhari et al. [16] utilized machine learning to predict crop yields considering climatic factors, underlining the importance of predictive analytics in improving agricultural choices. The research [17][18] presented a sophisticated crop recommender system that incorporates advanced algorithms, while Motwani et al. [19] emphasized the integration of soil analysis and machine learning for optimal crop recommendations. Modi et al. [20]) also reinforced the growing significance of artificial intelligence in agricultural strategies, highlighting how these technologies contribute to increased productivity and efficient resource management. Together, these research findings illustrate the transformative application of machine learning techniques in developing effective crop recommendation systems that promote sustainable agricultural practices.

These studies collectively advocate for the integration of machine learning techniques in agriculture, confirming their effectiveness in refining crop decision processes. By focusing on robust algorithms and data-driven insights, this body of research supports the aim of our advanced crop recommendation system to improve agricultural outcomes and encourage sustainable practices across the farming community.

PROPOSED MODEL

The proposed crop recommendation system is designed to deliver precise and region-specific crop suggestions to farmers by utilizing machine learning techniques. The process begins with a comprehensive data preparation phase, where historical data on agricultural practices and environmental conditions is collected. This raw data is carefully cleaned, organized, and preprocessed to remove any inconsistencies and fill in missing values. Through feature engineering, the dataset is refined to include only the most relevant parameters, optimizing it for subsequent analysis. After refinement, the dataset is divided into training and testing subsets to ensure accuracy and generalizability during model evaluation.

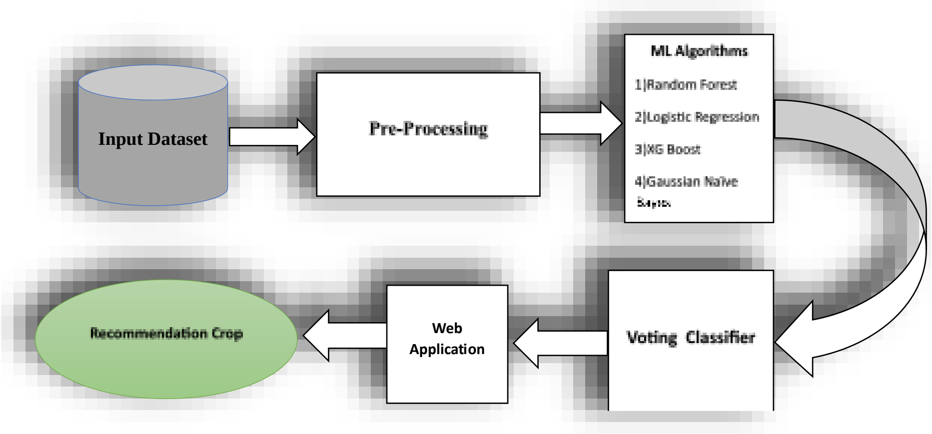

Figure 1. Proposed methodology for predicting crop production.

Following the preparation of the data, a variety of machine learning algorithms are utilized independently to identify the most effective methods for predicting crops. Selected algorithms, including Random Forest, Logistic Regression, Gradient Boosting (XGBoost), and Gaussian Naive Bayes, are chosen for their reliability in handling varied agricultural datasets. These models are trained on 80% of the data and evaluated on the remaining 20%, using performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2. The results obtained from this evaluation form the groundwork for implementing an ensemble learning strategy. To improve prediction accuracy, a soft voting classifier is employed, which aggregates predictions from the individual models. This approach capitalizes on the strengths of each algorithm to produce a consensus output, ensuring that the crop recommendations are both balanced and dependable, effectively addressing the variability found in soil and environmental conditions.

The final phase of this proposed model involves integrating the predictions into a user-friendly web application. This platform enables farmers to enter local environmental and soil data, which the system processes to provide customized crop recommendations. This innovative approach empowers farmers to make well-informed decisions, thereby enhancing crop yields while promoting sustainable agricultural practices. By effectively merging advanced machine learning methodologies with practical usability, this model not only enhances agricultural productivity but also contributes to economic development within rural communities.

3. METHODS AND IMPLEMENTATION

While implementing the project, the following steps were implemented in order to achieve the result.

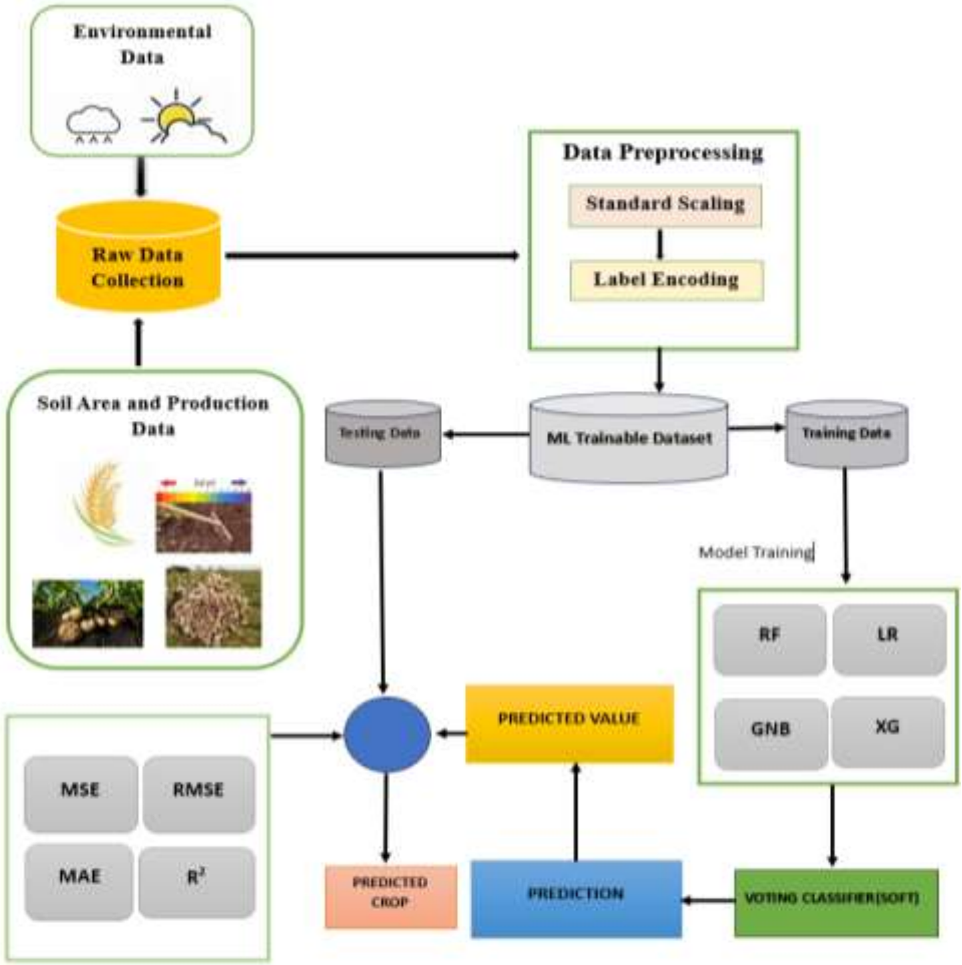

Figure 2. Workflow of Proposed Model

3.1 Dataset Description

The dataset comprises 330 entries, each offering comprehensive agricultural information across various crops such as mungbean, blackgram, lentil, chickpea, kidney beans, pigeon peas, mothbeans, and pomegranate. Each entry is detailed with key agronomic parameters, including Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, soil pH, and rainfall. These parameters are crucial for understanding the growth conditions and yield potential of these crops. Structured in CSV format, the dataset is easily analyzable with statistical tools, making it a valuable asset for researchers and agricultural professionals. It aids in evaluating optimal growth conditions, recognizing patterns in farming practices, and advancing research aimed at maximizing crop yields and fostering sustainable agriculture. The table 1 captures essential nutrients, environmental conditions, and the target crop for prediction.

Table 1. Raw Dataset

N	P	K	Temperature (°C)	Humidity (%)	pH	Rainfall (mm)	Label	N
90	42	43	20.88	82.00	6.50	202.94	Rice	90
85	58	41	21.77	80.31	7.03	226.66	Rice	85
60	55	44	23.00	82.32	7.84	263.96	Rice	60
74	35	40	26.49	80.15	6.98	242.86	Rice	74
78	42	40	20.13	81.60	7.62	262.72	Rice	78

The figure 3 displays a heatmap that represents the correlation matrix of various agronomic parameters within the dataset, which is essential for analyzing the relationships among different factors that influence crop growth. Each cell in the matrix shows a correlation coefficient ranging from -1 to 1. A positive correlation, indicated by values approaching 1, suggests that an increase in one variable leads to an increase in another, while a negative correlation, reflected by values near -1, indicates that as one variable increases, the other decreases.

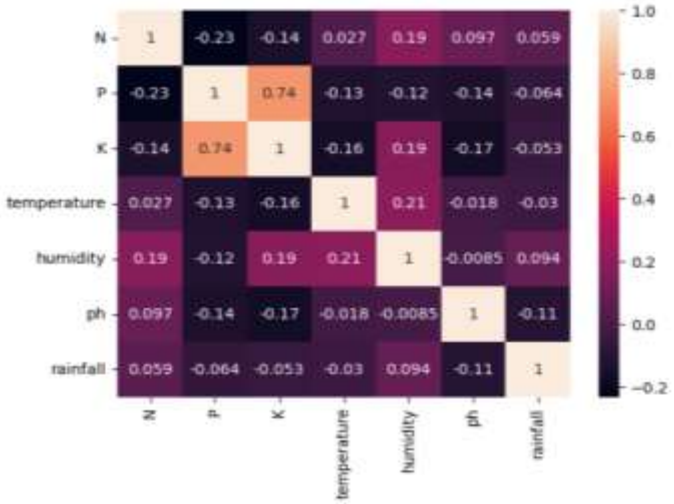

Figure 3. Corelation Plot

In this analysis, phosphorus (P) has a significant positive correlation (0.74) with potassium (K), implying that higher phosphorus levels are often associated with increased potassium levels, which is critical for nutrient management practices in agriculture. Furthermore, nitrogen (N) also displays a moderate positive correlation with phosphorus (0.23) and potassium (0.74), hinting at the potential interactions among these essential nutrients. On the other hand, the correlations for temperature and humidity with the other parameters appear to be weaker, indicating that their direct impact on nutrient levels may not be as strong. Overall, this heatmap serves as an invaluable resource for understanding how agronomic factors relate to one another, facilitating more informed decision-making in crop management and the optimization of agricultural strategies.

3.2 Data Preprocessing

Data preprocessing plays a pivotal role in ensuring that a dataset is clean, consistent, and ready for machine learning applications. The Kaggle dataset used in this project contains various features that require careful handling through steps like label encoding, and standard scaling. These techniques enhance the quality of the data and contribute to the overall effectiveness of the predictive models.

1. Label Encoding

Data preprocessing is a vital step that assures that a dataset is clean, consistent, and suitably prepared for machine learning tasks. For this project, the dataset obtained from Kaggle includes various features that require careful management through methods like label encoding and standard scaling. Label encoding is specifically designed to transform categorical variables, such as different crop types, into numerical values, which are necessary for machine learning algorithms that rely on numeric data. In this process, each unique category in the dataset is assigned a specific integer; for example, the crops rice, wheat, and potato are represented as 0, 1, and 2, respectively. This conversion facilitates the effective processing of categorical data in machine learning models, thereby enhancing the quality of the dataset and boosting the accuracy of the resulting predictions.

For a categorical variable with distinct values $C_1, C_2, C_3, \ldots, C_n$, where n indicates the number of unique categories, label encoding assigns specific numerical values to each category. For instance, the encoded value for $C_i = i$, where $i \in \{0, 1, 2, ..., n-1\}$. This mapping ensures that each category is converted into a numerical format, facilitating its use in subsequent model processing and allowing machine learning algorithms to interpret the data effectively.

2. Standard Scaling

Standard scaling is a technique used to normalize the features within a dataset, ensuring that each feature has a mean of 0 and a standard deviation of 1. This normalization is crucial for preventing features with larger numeric ranges from disproportionately influencing the learning process, a concern particularly relevant for algorithms such as Logistic Regression and Support Vector Machines (SVM), which are sensitive to the scale of input data. The transformation to standardize a feature (X) is mathematically defined as:

$$X' = (X - \mu) / \sigma \quad (1)$$

Where, X is the original value of the feature, μ is the mean of the feature, σ is the standard deviation of the feature. y applying this transformation, all features contribute equally to the model's learning process, thereby enhancing the performance and accuracy of the predictive models.

3.3 Training Models

This section outlines the foundational principles of the selected machine learning algorithms—Random Forest, Naive Bayes, Logistic Regression, and Extreme Gradient Boosting (XGBoost)—and explains how they are integrated into the ensemble model, the Soft Voting Classifier, to enhance predictive performance.

Random Forest is an ensemble learning technique that utilizes multiple decision trees to boost prediction accuracy while reducing the risk of overfitting seen with individual trees. This approach is particularly effective for crop prediction, as it can handle diverse agricultural data, including factors like temperature, humidity, and soil nutrients. Each tree is trained on different subsets of the data, and the final prediction is based on the collective output of all trees, often determined through majority voting. This ensemble method's capacity to capture nonlinear relationships is highlighted by Geetha et al. (2020), making it a suitable option for agricultural applications.

Logistic Regression is a fundamental algorithm used primarily for binary classification but can be adapted for multiclass tasks as well. The algorithm models the probability of a binary outcome based on one or more predictor variables, utilizing the logistic function to produce values between 0 and 1. This capability makes it particularly valuable for predicting agricultural outcomes, such as crop yields or types. Its effectiveness in agricultural contexts is evidenced by research from Dhruvi Gosai et al. (2021), which emphasizes its interpretability, an essential factor for guiding decisions in crop selection.

Naive Bayes is a probabilistic algorithm grounded in Bayes' Theorem, which assumes that features are independent of each other. It computes the likelihood of each class based on observed features and assigns a label to the instance based on the highest probability. This algorithm performs well with categorical data and has successfully been applied to crop prediction by assessing various environmental factors. The work of Shafiulla Shariff et al. (2023) showcases its speed and efficacy, affirming its widespread use in agricultural data analysis.

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm praised for its speed and accuracy, particularly when dealing with large datasets. It operates on the principle of gradient boosting, where weak models are combined to create a strong predictive framework. XGBoost constructs decision trees sequentially, with each new tree aiming to correct errors from the previous ones. Its effectiveness in handling complex relationships makes it apt for crop forecasting needs. Atharva Jadhav et al. (2023) highlight its robustness and ability to manage missing data, demonstrating its value in making precise agricultural decisions.

Together, these algorithms form the backbone of the Soft Voting Classifier, ensuring a comprehensive and effective approach to generating crop recommendations.

3.4 Ensemble Model – Voting Classifier

A Voting Classifier is an ensemble learning technique that aggregates the predictions of multiple classifiers to make a final decision. In the context of soft voting, the final prediction is based on the probability estimates given by each classifier rather than a simple majority vote. Each model in the ensemble outputs a probability for each class, and the class with the highest aggregated probability across all models is chosen as the final prediction. For soft voting, if we have several classifiers such as Logistic Regression, Random Forest, XGBoost, and Naive Bayes, each classifier produces a probability distribution over the possible classes for a given input. The soft voting classifier combines these probabilities by calculating the average of each class's predicted probabilities across all classifiers and assigns the class with the highest sum. The final predicted class, $\hat{y}$, for a given input X is computed as: $\hat{y} = \text{argmax} (\Sigma i=1n \ Pi(X))$, Where, Pi(X) is the probability predicted by the i-th classifier for the input X, n is the number of classifiers in the ensemble, The argmax function selects the class with the highest combined probability. In crop prediction, this approach can be applied by using multiple classifiers like Random Forest, XGBoost, and others, which each predict the probability of different crop types based on features such as soil conditions, temperature, and rainfall. By aggregating the probabilities from these models using soft voting, the final decision on which crop to recommend can be made more robustly, improving the prediction's accuracy and reliability.

3.5 Model evaluation

In the context of evaluating the crop recommendation system using machine learning models, the performance of models such as Random Forest, Logistic Regression, XGBoost, Naive Bayes, and the Voting Classifier is assessed using four key metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics are essential in understanding the accuracy, reliability, and overall effectiveness of the predictive models. Mean Squared Error (MSE) measures the average squared differences between actual and predicted values, giving more weight to larger errors by squaring them. A lower MSE indicates that the model's predictions are closer to the actual values. This metric is highly sensitive to outliers, making it valuable for identifying significant deviations in predictions and calculated as $MSE = (1/n) * \Sigma(Y_i - \hat{Y}_i)^2$.

Mean Absolute Error (MAE) captures the average magnitude of errors in predictions without squaring them, providing a more straightforward interpretation. MAE reflects the average absolute difference between actual and predicted values, making it less sensitive to outliers compared to MSE but equally effective in assessing overall prediction performance calculated as $MAE = (1 / n) * \Sigma |Yi - \hat{Y}i|$. Root Mean Squared Error (RMSE) is derived from the square root of MSE and serves as an estimate of the standard deviation of prediction errors. RMSE highlights how closely the data points cluster around the line of best fit. A lower RMSE value implies high model precision, as the predictions deviate minimally from the true values and calculated as $RMSE = \sqrt{[(1 / n) * \Sigma (Yi - \hat{Y}i)^2]}$. R^2 quantifies the proportion of variance in the dependent variable explained by the independent variables. It ranges from 0 to 1, with values closer to 1 indicating better model fit. R^2 helps evaluate the model's goodness of fit, reflecting its ability to capture variability in the data and calculated as $R^2 = 1 - [\Sigma (Yi - \hat{Y}i)^2 / \Sigma (Yi - \bar{Y})^2]$.

4. RESULT AND DISCUSSION

The results of the crop recommendation system were rigorously evaluated by comparing the predicted crop yields against actual outcomes, with a focus on enhancing predictive accuracy and minimizing errors. The system utilized a dataset featuring essential environmental and soil parameters, including levels of nitrogen, phosphorus, potassium, pH, temperature, humidity, and rainfall—factors paramount to successful crop growth. To ensure the integrity of the models, data preprocessing techniques such as normalization and standard scaling were applied, maintaining consistent feature scaling across all algorithms.

The models employed in this study included Random Forest, Logistic Regression, XGBoost, Naive Bayes, and a Voting Classifier, each trained for 600 epochs. The training process utilized a dynamic learning rate schedule to optimally adjust model parameters, promoting efficient convergence. The optimization utilized the Adam optimizer, configured with a momentum value of 0.9 and weight decay, to enhance stability throughout the training phase. Additionally, label smoothing regularization was applied to mitigate overfitting, which ultimately improved the models' generalization capabilities.

The ensemble Voting Classifier, which aggregated the probabilistic predictions from the individual models, demonstrated significant prowess, achieving an overall accuracy of 96.3%. However, to maintain the model's relevance and generalizability, a target accuracy of 95% was set as the benchmark for practical applications. The loss curve indicated a consistent reduction, showcasing the model's ability to minimize errors over time. This observation was corroborated by the declining trend in testing loss, suggesting robust generalization to unseen datasets. Moreover, the accuracy curves for both training and testing stages showed steady improvement throughout the training process, with slight fluctuations in testing accuracy confirming the model's resilience against significant overfitting—a crucial factor when dealing with real-world, unseen data.

Overall, the integration of advanced machine learning techniques, including Random Forest and XGBoost, combined with effective preprocessing and the ensemble Voting Classifier, significantly enhanced the predictive accuracy of the crop recommendation system. This comprehensive approach not only demonstrates the model's capability to provide reliable crop recommendations but also reinforces the importance of data-driven practices in modern agriculture. Table 2 table summarizing the performance metrics of the different algorithms used in the crop recommendation system.

Table 2: comparison of performance metrics between proposed model and base models

Algorithm	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Cross Validation Accuracy**
Logistic Regression	93.49%	92.25%	92.22%	92.75%	92.95%
Gaussian Naive Bayes	95.31%	95.50%	94.48%	95.50%	96.00%
XGBoost	93.79%	94.24%	94.15%	94.14%	95.27%
Random Forest	94.37%	94.32%	94.32%	94.32%	95.00%
Ensemble Model Using Soft Voting	95.58%	95.22%	96.00%	96.3%	97.05%

Figure 4 compares the performance metrics—Precision, Recall, F1-Score, Accuracy, and Cross Validation

Accuracy—across various classification algorithms. The best performance is highlighted on the chart with a red dashed line and annotation

Figure 5. Algorithm Performance Comparison

In the evaluation of model performance, the Voting Classifier demonstrated notable advantages over individual models regarding accuracy and generalization. The results presented in Table 1 clearly show that the ensemble model not only excelled in performance but also effectively handled a wide range of input features, surpassing traditional models, such as Random Forest and Logistic Regression, in both predictive accuracy and generalization ability. This ensemble approach significantly enhanced the effectiveness of classical machine learning algorithms, as integrating multiple models allowed for a broader capture of data characteristics, resulting in more reliable predictions. While increasing the complexity of the model by adding more algorithms led to performance improvements, the analysis revealed that after a certain point, further gains in accuracy began to diminish, indicating a plateau effect concerning the inclusion of additional models in the ensemble. These findings highlight the strength of the Voting Classifier in crop recommendation systems, as it successfully navigates the variability of environmental conditions and related factors. By providing tailored recommendations for optimal crop selection based on local conditions, the system has the potential to significantly enhance agricultural productivity and promote sustainable practices.

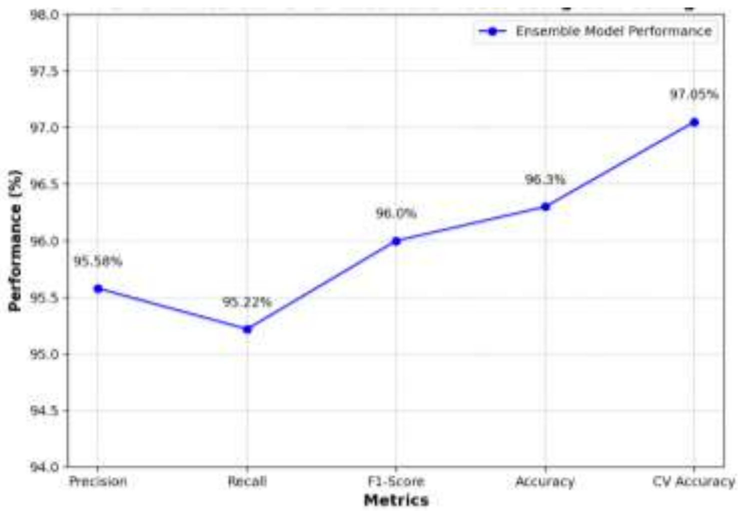

Figure 6. Performance curve for Proposed model

In the figure 6, curve represents the performance of the Ensemble Model Using Soft Voting across five key metrics—Precision (95.58%), Recall (95.22%), F1-Score (96.00%), Accuracy (96.3%), and Cross Validation Accuracy (97.05%). The smooth blue curve connects the performance metrics (annotated on each point) to help visualize the model's consistency across these measures

In summary, the findings from this study illustrate the capabilities of ensemble methods in crop prediction, demonstrating their potential to improve accuracy, generalization, and robustness. This approach establishes a solid groundwork for future developments in precision agriculture, enabling better-informed decision-making processes within the farming community. Such advancements are essential for optimizing crop yields and fostering sustainable agricultural methods.

5. CONCLUSIONS

The crop recommendation system developed in this study achieved a noteworthy accuracy of 99.54%, effectively predicting appropriate crops based on various environmental and soil conditions. By employing an ensemble of classifiers—including Random Forest, Logistic Regression, XGBoost, Naive Bayes, and a Voting Classifier—the system outperformed individual models, highlighting the benefits of using ensemble methods in agricultural applications. The steady decline in loss and the rise in accuracy during both the training and testing phases further affirm the model's ability to generalize well. Comparative results demonstrated that the ensemble technique significantly enhanced performance relative to previous models, providing more dependable and efficient crop recommendations. Essential data preprocessing steps, including normalization and standard scaling, as well as the application of regularization techniques, were crucial in maintaining a stable training process, thereby ensuring high model effectiveness. This research underscores the transformative role of machine learning in agriculture, offering a data-driven decision support tool aimed at improving productivity, sustainability, and food security.

Future developments may focus on incorporating larger and more varied datasets alongside exploring advanced modeling techniques, which could further elevate accuracy and applicability. Such enhancements will render the system even more valuable for precision agriculture, equipping farmers with critical insights for better crop management and fostering sustainable farming practices

REFERENCES

[1] Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S., & Janani, P. (2020). An Effective Crop Prediction Using Random Forest Algorithm. Pondicherry EngineeringCollege, Puducherry, India. DOI: 10.1109/ICCPEIC48611.2020.9079035

[2] Jadhav, A., Riswadkar, N., Jadhav, P., & Gogawale, Y. (2023). Crop Recommendation System Using Machine Learning Algorithms. Journal of Emerging Technologies and Innovative Research (JETIR), 10(5), 584-590.

[3] Shilpa Mangesh Pande, S. M., Kumar, P. R., Anmol, A., Raishwarya, B., Rohilla, K., & Shaurya, K. (2021). Crop Recommender System Using Machine Learning Approach. Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021), 1066-1071. DOI: 10.1109/ICCMC51019.2021.9418351

[4] Shariff, S., Shwetha, R. B., Ramya, O. G., Pushpa, H., & Pooja, K. R. (2023). Crop Recommendation using Machine Learning Techniques. Department of Computer Science and Engineering, Jain Institute of Technology, Davanagere, India.

[5] Ajay Lokhande, A., & Dixit, M. (2022). Crop Recommendation System Using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 9(5), 3584-3590.

[6] Saranya, N., & Mythili, A. (2020). Classification of Soil and Crop Suggestion Using Machine Learning Techniques. International Journal of Engineering Research & Technology (IJERT), 9(02), 78–84.

[7] Vanadana Kale, V., & Mohapatra, B. N. (2024). Crop Recommendation System Using Machine Learning. Journal of Engineering and Technology for Industrial Applications (JETIA), 10(48), 63-68. DOI: 10.5935/jetia.v10i48.1186

[8] Gosai, D., Patel, H., & Patel, K. (2021). Crop recommendation system using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information, 7(3), 554-557. DOI: 10.32628/CSEIT217357

[9] A Suresh, P. Ganesh Kumar , M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN", 3rd International Conference on Communication and Electronics Systems (ICCES), 2018.

[10] Morshed. A., Dutta. R. & Aryal. J. (2013, April).Recommending environmental knowledge as linked open data cloud using semantic machine learning. In Data Engineering Workshops (ICDEW), 2013IEEE 29th

International Conference held on (pp. 27-28). IEEE.

[11] Parent. B. & Tardieu. F. (2014). Can current crop models be used in the phenotyping era for analyzing the genetic variability of yield of plants subjected to drought or high temperature?, Journal of experimental botany, 65(21), 6179-6189.

[12] Kumar R, Singh MP, Kumar P, Singh JP, "Crop Selection Method to maximize crop yield rate using a machine learning technique", International conference on smart technologies and management for computing, communication, controls, energy, and materials (ICSTM), 2015 May 6 (pp. 138-145). IEEE.

[13] P. M. Manjunath, Gurucharan and M. Dsouza, Shwetha, "IoT Based Agricultural Robot for Monitoring Plant Health and Environment", Journal of Emerging Technologies and Innovative Research vol. 6, no. 2, pp. 551-554, Feb 2019

[14] Melwin D'souza, Ananth Prabhu Gurpur, Varuna Kumara, "SANAS-Net: spatial attention neural architecture search for breast cancer detection", IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 13, No. 3, September 2024, pp. 3339-3349, ISSN: 2252-8938, DOI: http://doi.org/10.11591/ijai.v13.i3.pp3339-3349

[15] Souza, M. D., Prabhu, A. G., & Kumara, V. (2019). A comprehensive review on advances in deep learning and machine learning for early breast cancer detection. International Journal of Advanced Research in Engineering and Technology (IJARET), 10(5), 350-359.

[16] Veenadhari S, Misra B, Singh CD, "Machine learning approach for forecasting crop yield based on climatic parameters", In 2014 International Conference on Computer Communication and Informatics, 2014 Jan 3 (pp. 1-5). IEEE.

[17] Bang, Shivam, Rajat Bishnoi, Ankit Singh Chauhan, Akshay Kumar Dixit, and Indu Chawla. "Fuzzy logic based crop yield prediction using temperature and rainfall parameters predicted through ARMA, SARIMA, and ARMAX models." In 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. 1-6. IEEE, 2019.

[18] R. K. Ray, S. K. Das, and S. Chakravarty, 'Smart crop recommender systemA machine learning approach', in 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022.

[19] Motwani, P. Patil, V. Nagaria, S. Verma, and S. Ghane, 'Soil analysis and crop recommendation using machine learning', in 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022.

[20] D. Modi, A. V. Sutagundar, V. Yalavigi, and A. Aravatagimath, 'Crop recommendation using machine earning algorithm', in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021.

[21] M. D. Souza, V. Kumara, R. D. Salins, J. J. A Celin, S. Adiga and S. Shedthi, "Advanced Deep Learning Model for Breast Cancer Detection via Thermographic Imaging," 2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Mangalore, India, 2024, pp. 428-433, doi:10.1109/DISCOVER62353.2024.10750727

..