Synergizing Random Forest and K-Means Algorithms: An Analytical Study for Precise Crop Recommendation in Southeast Asia

Abstract

Crop detection and classification are pivotal for optimizing agricultural practices and ensuring sustainable farming. This research presents a sophisticated approach to identifying optimal environments for various crops using advanced machine-learning techniques. The study employs a Random Forest classifier framework to categorize crops based on crucial environmental parameters, including soil nitrogen, phosphorus, potassium levels, temperature, humidity, soil pH, and rainfall. Additionally, a K-Means clustering algorithm groups crops with similar growth conditions.

The model demonstrates superior performance compared to existing state-of-the-art approaches. Findings underscore distinct environmental requirements for different crop groups, such as those thriving in arid conditions with minimal rainfall and nutrient content, versus those favouring humid conditions with abundant rainfall and nutrient richness. This study emphasizes the potential of machine learning models to enhance agricultural productivity by aligning crop selection with suitable environmental conditions, facilitating precise agricultural decision-making.

Keywords: Random Forest, Classification, Crop Detection, K-Means Clustering, Sustainable Farming.

1. Introduction

The evolution of agriculture and crop recommendation techniques has been marked by significant technological advancements. Initially reliant on traditional knowledge and experience, modern agriculture has embraced scientific methods and data-driven approaches. Early techniques involved basic soil analysis and climate observation. With the advent of computers and data science, machine learning algorithms such as Random Forest and K-Means clustering have revolutionized crop recommendation [1]. These models analyze vast datasets encompassing soil properties, climate patterns, and historical yields to predict optimal crop choices. This evolution not only enhances efficiency and sustainability in farming but also empowers farmers with precise insights for maximizing yields while minimizing environmental impact [2].

The fusion of Random Forest Classification and K-Means Clustering significantly enhances the optimization of crop recommendation processes. Random Forest Classification excels in handling complex, nonlinear relationships within data by constructing multiple decision trees and aggregating their outputs [3][4]. It effectively categorizes crops based on diverse environmental factors such as soil nutrients, climate conditions, and rainfall patterns.Meanwhile, K-Means Clustering groups crops with similar growth requirements into clusters, identifying distinct environmental niches where specific crops thrive. By combining these methodologies, the model not only categorizes crops accurately but also identifies clusters of crops that share similar environmental preferences. This synergy enables precise matching of crops to their most suitable environmental conditions, thereby optimizing agricultural productivity and resource utilization [5]. Farmers can make informed decisions on crop selection and management practices, fostering sustainable farming practices and maximizing yields in varying environmental contexts.

Crop recommendation is crucial for maximizing agricultural productivity, sustainability, and profitability. It helps farmers select crops best suited to their specific environmental conditions, such as soil quality, climate, and rainfall patterns [6]. By matching crops with optimal growing conditions, crop recommendation techniques enhance yield potential while minimizing resource wastage. This targeted approach also reduces the risk of crop failure and ensures efficient use of land and inputs like water and fertilizers [7]. Moreover, accurate crop recommendations contribute to food security by diversifying crop types and ensuring stable production levels. Overall, effective crop recommendation supports resilient agricultural systems capable of adapting to changing environmental and economic pressures, thereby securing livelihoods and promoting global food supply stability [8].

2. Related Works

In recent years, advancements in the application of ML-based techniques, particularly classification and clustering, have significantly enhanced the performance and efficiency of crop recommendation for various applications and use cases.D. Modi et al. presented an SVM-based crop recommendation system for the formers. In this work, it is required to analyze the profit of a certain crop, which avoids losses for farmers while increasing production [9]. Similarly, T. K. Mishra et al. advocated Crop selection by utilizing machine learning approaches such as K-Nearest Neighbour (KNN) and Random Forest. Both models were fully simulated on the Indian dataset, and an analytical report was produced. The model assists farmers in determining the type of crop before cultivating it on an agricultural field, allowing them to make more informed decisions [10].J. Madhuri et al. presented a new recommendation system that utilizes Artificial Neural Networks (ANN) to select appropriate crops [11].

S. Z. Rahman et al. suggested a model that can forecast soil series with land type and, based on the prediction, suggest suitable crops using different machine learning methods such as weighted k-nearest Neighbour (k-NN), Bagged Trees, and Gaussian kernel based Support Vector Machines (SVM) [12]. In a different configuration, M. S. Suchithra et al. emphasized the analysis of agriculture data and the creation of a rank-based recommendation system to find the best-suited crops for a certain place by mining a significant quantity of crop, soil, and geographic data using the clustering approach and the ball-tree algorithm [13]. P. Parameswari et al. contributed to develop a model that uses machine learning algorithms like PART, Decision table, and JRip to provide farmers with crop-related information or crop recommendations based on a variety of attributes like crop details, soil composition, weather conditions that crops can grow in, temperature, soil PH, and rainfall [14]. R. Kumar et al. proposed a system that can usesConvolutional Neural Networks to identify plant illnesses and use ML to analyze many soil characteristics to recommend different crops depending on soil quality. In order to create an accurate database, different plant species are identified and given new names using the Plant Village Dataset, which serves as the source of the dataset for the disease prediction training and test [15].

While these modern, state-of-the-art ML-based techniques are adaptive and cater to various applications, they may not be ideally suited for complex datasets. Our study presents a performance-based comparative analysis, focusing on the fusion of classification and clustering methodologies using Random Forests and K-means clustering. The fusion of clustering and classification offers superior results for crop recommendation and detection by leveraging the strengths of both methodologies. Classification, like Random Forest, accurately categorizes crops based on complex environmental parameters, while clustering, like K-Means, groups crops with similar growth conditions. This dual approach ensures precise crop selection by not only identifying optimal crops for given conditions but also highlighting similar crops that thrive under comparable environments. This synergy enhances prediction accuracy, adaptability, and resource efficiency. By combining these techniques, the model outperforms state-of-the-art approaches, providing more nuanced insights and robust recommendations tailored to diverse agricultural contexts. In the subsequent sections, the detailed methodology of the study is discussed.

3. Methodology

3.1. Data Importing and Outline

Data importing and outlining sets the foundation for any data-driven project, facilitating the transition of raw data from external sources into a programming environment for analysis and modeling. The process typically begins with importing essential libraries such as pandas for data manipulation, numpy for numerical operations, and scikit-learn for machine learning implementations. The next step involves loading data from diverse sources such as CSV files, databases, or APIs using specialized functions like read_csv or read_excel from pandas. Figure. 1 demonstrates the model workflow of the study.

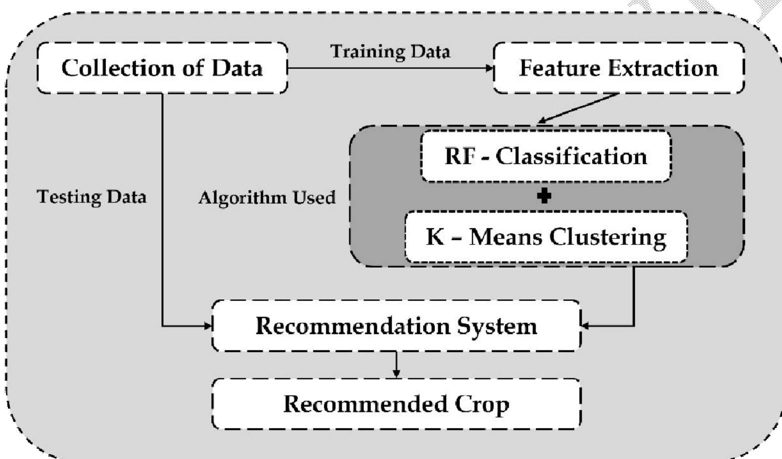

Figure 1: Model Workflow Diagram

Upon loading, initial exploration commences with tasks like displaying the first few rows of data (head()), checking data types (dtypes), and handling missing values (isnull(), fillna()). Subsequently, data transformation techniques may be applied to convert categorical data into numerical formats, ensuring compatibility with machine learning algorithms. Finally, the data is typically split into training and testing sets using train_test_split() to evaluate model performance accurately. This systematic approach ensures that data is imported, prepared, and structured efficiently for subsequent analysis and modeling tasks.

3.2. Dataset Preparation

Preparing the dataset involves several critical steps to ensure its readiness for analysis and modeling. Initially, the dataset's structure is reviewed, confirming the presence of essential columns: Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH_Value, Rainfall, and Crop [16]. The next step involves addressing data quality issues such as missing values, outliers, and inconsistencies. Techniques like imputation for missing values and statistical methods or domain knowledge for outliers are applied. Numerical data, including Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH_Value, and Rainfall, may undergo normalization or standardization to bring them to a consistent scale, ensuring fair comparisons across features during modeling [17]. Categorical data, specifically the Crop column, requires encoding into a numerical format using techniques like one-hot encoding to enable machine learning algorithms to process it effectively. Table. 1 gives a brief idea about the contents of the dataset used in this study.

Table. 1: Dataset Description

Nitrogen	Phosphorus	Potassium	Temperature	Humidity	pH	Rainfall	Crop
90	42	43	20.879744	82.002744	6.5	202.934	Rice
85	58	41	21.770462	80.319644	7.0	226.655	Rice
60	55	44	23.004459	82.320763	7.8	263.964	Rice
74	35	40	26.491096	80.158363	6.9	242.864	Rice
78	42	42	20.130175	81.604873	7.6	262.717	Rice

Finally, the dataset is split into training and testing sets to assess model performance accurately. This meticulous preparation process is crucial for maximizing the dataset's utility, enabling robust analysis and predictive modeling for crop recommendation based on soil and environmental factors.

3.3. Model Implementation

The dual-model approach of combining Random Forest Classification and K-Means Clustering offers a powerful methodology for analyzing agricultural data comprehensively.Random Forest Classification is a robust and widely used technique in supervised learning, particularly suited for predicting crop types based on comprehensive datasets that encompass crucial soil and environmental attributes such as Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH_Value, and Rainfall. Figure. 2 demonstrates the model architecture diagram.

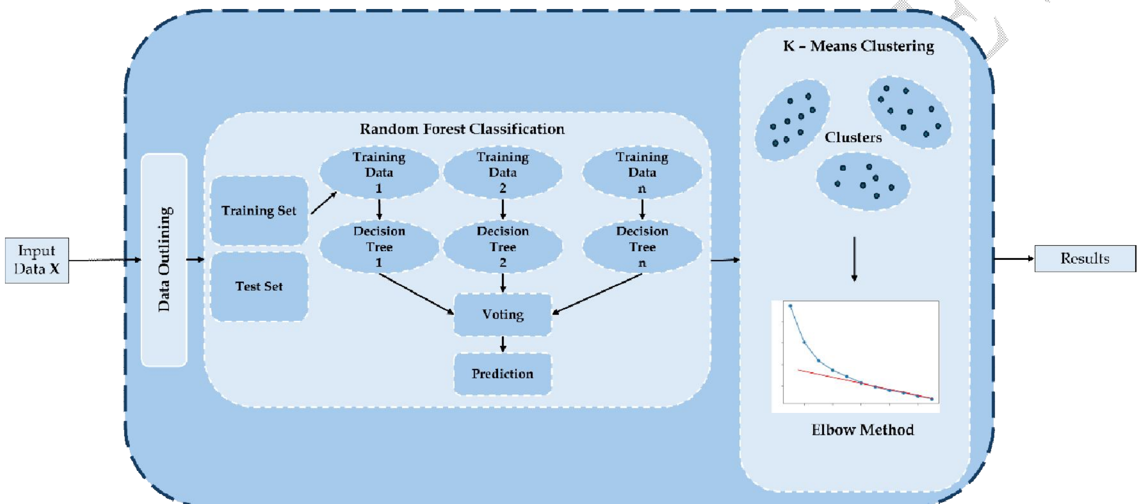

Figure 2: Architecture of the Proposed Model

Initially, to facilitate machine learning algorithms, categorical crop labels are transformed into a numerical format using LabelEncoder [18]. This preprocessing step ensures that the data is compatible with the Random Forest algorithm, which excels in handling both numerical and categorical data effectively.During training, each decision tree independently learns from the dataset by recursively splitting nodes based on the most discriminative features, such as Nitrogen, Phosphorus, Potassium levels, Temperature, Humidity, pH_Value, and Rainfall. This process allows the model to capture complex relationships between these attributes and the target variable—the type of crop.Equation (1)and (2) explains the working of the RF classifier.

$$P\ (y = c|x) = \frac{1}{T}\sum_{t=1}^{T} p_c(x) \qquad (1)$$

This equation illustrates that the Random Forest combines the probabilistic outputs of each decision tree to obtain a more robust estimate of the class probabilities. In equation (1), T is the total number of decision trees in the forest and $p_c(x)$denotes the probability assigned by the t-th tree to the class c for the input x. Each decision tree contributes its probability distribution over the classes, and averaging these distributions across all trees helps to smooth out noise and improve the overall prediction accuracy [19][20].For classification tasks, the final predicted class label is typically determined by choosing the class with the highest average probability in equation (2).

$$y_{RF}(x) = \arg max_c \left(\frac{1}{T}\sum_{t=1}^{T} p_c(x)\right) \qquad (2)$$

Once trained, predictions are made by aggregating the outputs of all trees through a voting mechanism. The majority vote determines the predicted crop type for new data instances.Evaluation metrics like accuracy, precision, recall, and F1-score are employed to assess the model's performance. These metrics provide insights into how well the Random Forest classifier can accurately classify crops based on the provided attributes, crucial for agricultural decision-making and optimizing crop management strategies.

Further, Complementing Random Forest Classification, K-Means Clustering serves as a pivotal tool in unsupervised learning for exploring patterns and groupings within agricultural datasets based on environmental attributes [21]. Unlike supervised methods that predict crop types, K-Means focuses on segmenting crops into clusters based on similarities in their environmental conditions.The goal of K-means clustering is to minimize the within-cluster sum of squares (WCSS), which is defined in equation (3).

$$WCSS = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||\, x_i - \mu_k||^2 \qquad (3)$$

Where, it represents the squared Euclidean distance between data point and the centroid. Before applying K-Means, the dataset undergoes preprocessing to standardize or normalize features like Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH_Value, and Rainfall. This step ensures that all attributes contribute equally to the clustering process.The Elbow Method is then employed to determine the optimal number of clusters, iterating through a range of cluster numbers and plotting the Within-Cluster Sum of Squares (WCSS) [22][23]. Equation (4) gives the mathematical representation of the elbow method.

$$Elbow\ Point = \arg min_K \sum_{k=1}^{K} \sum_{x_i \in C_k} || x_i - \mu_k ||^2 \quad (4)$$

The point where the decrease in WCSS levels off suggests the appropriate number of clusters, indicating how many distinct groups of crops exist based on their environmental attributes.With the optimal cluster count identified, K-Means assigns each data point to a cluster based on its proximity to the cluster's centroid in the feature space. This allows for the exploration and visualization of cohesive groups of crops that share similar environmental requirements. Such insights are invaluable for understanding crop diversity, optimizing resource allocation, and tailoring agricultural practices to maximize productivity and sustainability.

By combining both models, the authors gain a holistic view of the dataset. Random Forest Classification provides predictive insights into which crop types are likely to thrive under specific environmental conditions, facilitating precise crop recommendation systems. Meanwhile, K-Means Clustering offers a deeper understanding of how crops naturally group together based on their shared environmental attributes, revealing potential insights for agricultural management practices and resource allocation. Together, this dual-model approach enhances decision-making processes in agriculture, leveraging machine learning to optimize crop selection and improve productivity sustainably.

4. Experimentations and Results

The dataset used by the authors comprises 5000 data points, each characterized by the distribution of 7 features: Nitrogen, Phosphorus, Potassium, Temperature, Humidity, Rainfall, and pH. These features are predominantly derived from agricultural regions in South-East Asia and India. The illustration of the distribution of different features within the dataset is depicted in Figure. 3. Each graph's x-axis delineates the range of values for the respective variable, while the y-axis denotes the frequency of each observed value. Analysis reveals prevalent temperature readings between 20 and 40 degrees Celsius, predominant humidity levels ranging from 60 to 80 percent, and frequent occurrences of rainfall volumes between 50 and 100 millimetres. These insights provide a valuable understanding of the typical climatic conditions within the city, aiding in various environmental analyses and planning endeavours.

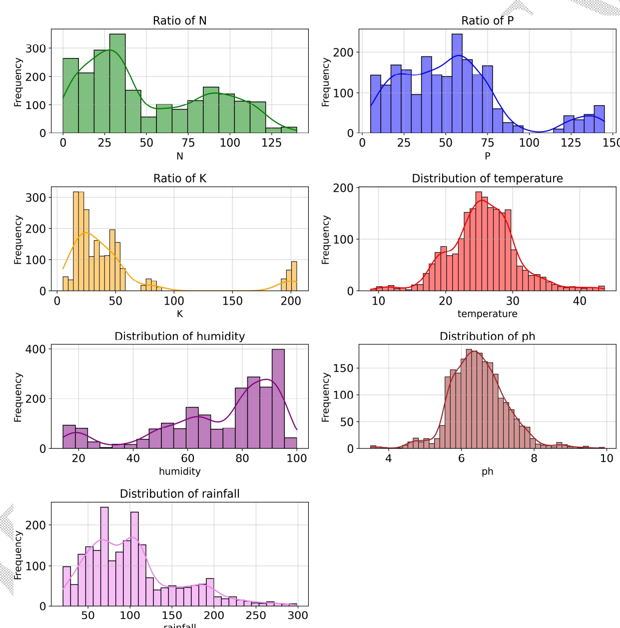

Figure 3: Distribution Graph of Different Features

The performance metrics for the modelare shown in Table. 2 for all the crops. It comprises several parameters, including precision, recall, and F1-score for distinct crop classes (apple, banana, blackgram, etc.), along with accuracy. By taking into account the support (number of examples) for each class, these metrics assess how well the model can predict each class. Overall performance across several crop categories shows good precision, recall, and F1 scores together with excellent accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{F1 Score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (8)$$

Equation (5), (6), (7), and (8) gives the mathematical forefront to calculate the performance metrics for the individual classes and crops from the confusion matrix. Figure. 4 represents the correlation matrix obtained for the model.

Table. 2: Performance Metrics of the Proposed Model

Crop	Precision	Recall	F1-Score	Support
Apple	1.00	1.00	1.00	23
Banana	1.00	1.00	1.00	21
Blackgram	1.00	0.96	0.99	20
Chickpea	1.00	1.00	1.00	26
Coconut	0.98	0.97	1.00	27
Cotton	0.97	0.99	0.96	17
Grapes	1.00	1.00	1.00	14
Jute	0.92	1.00	0.96	23
Lentil	0.92	1.00	0.96	11
Maize	1.00	1.00	1.00	21
Mango	1.00	1.00	1.00	19
Muskmelon	1.00	1.00	1.00	17
Orange	1.00	1.00	1.00	14
Papaya	1.00	1.00	1.00	23
Rice	1.00	0.89	0.94	19

Watermelon	1.00	1.00	1.00	19
Accuracy			**0.97**	**440**
Macro Avg	**0.94**	**0.95**	**0.94**	**440**
Avg	**0.95**	**0.95**	**0.04**	**440**

In the correlation map of Figure. 4, each diagonal element is 1, indicating a perfect correlation between each variable and itself. This is expected as it measures a variable's correlation with itself. Additionally, the correlation matrix is symmetric, meaning the correlation between variables A and B is identical to that between B and A [24]. This symmetry confirms the consistency of relationships between variables, independent of comparison order.

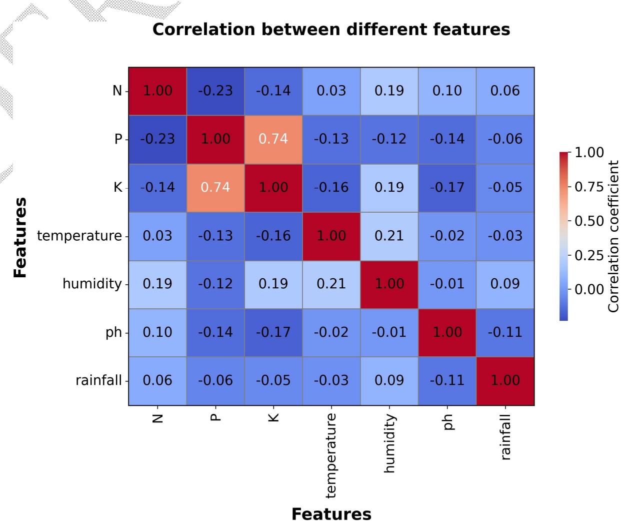

Figure 4: Correlation Between Different Features

The number of clusters determined for feeding into the K-means Clustering algorithm was experimentally determined to be 4 using the Elbow method.From Figure. 5it can be inferred that 4 is the optimal number of clusters for the dataset, as it represents the point where the WCSS reduction begins to taper off, indicating that adding more clusters does not significantly improve the clustering performance. This balance helps in finding a meaningful partitioning of the data without overfitting or underfitting.The PCA plot effectively illustrates the results of the K-Means clustering, showing clear separation between clusters and indicating that the chosen number of clusters (four) is appropriate for this dataset. The distinct clusters reflect the underlying patterns in the data, suggesting that the environmental attributes used in the analysis are effective in differentiating between different crop types. This visualization aids in understanding the similarities and differences in crop requirements, which is valuable for optimizing agricultural practices and decision-making.

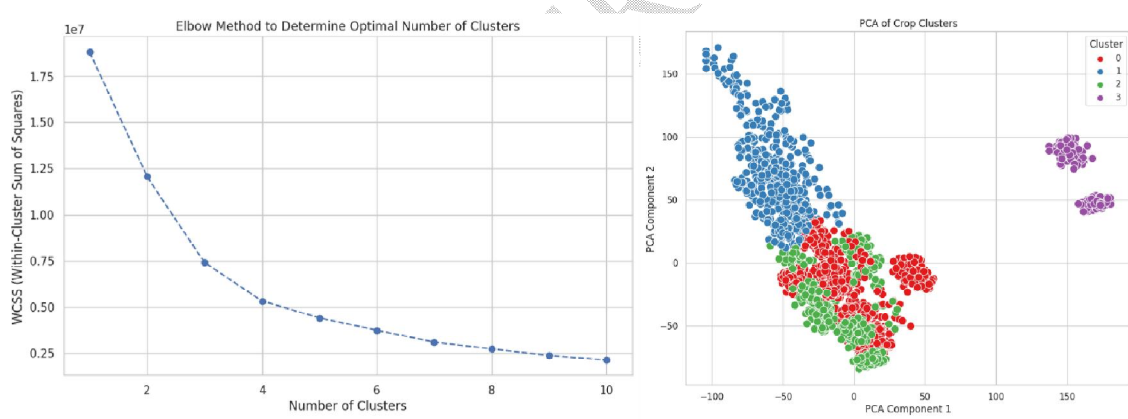

Figure 5: Performance of K-Means Clustering Analysis

The analysis of crop clusters, based on environmental attributes and using the fusion of RF and K-Means Clustering algorithm, reveals distinct groupings that highlight the common characteristics and optimal growing conditions for different crops. Cluster 0 encompasses crops like Mango, Orange, and Chickpea, which generally prefer warm climates and well-drained loamy soils, thriving in regions like India, the United States, and tropical countries. Cluster 1 includes crops such as Coconut, Rice, and Coffee, which favor high humidity and abundant rainfall, and are prominently grown in Southeast Asia, Latin America, and Africa. Cluster 2 groups crops like Banana, Cotton, and Watermelon, requiring warm temperatures and well-drained soils, and are commonly cultivated in tropical and subtropical regions such as Asia and the Americas. Lastly, Cluster 3 features Apples and Grapes, which thrive in temperate climates with well-drained loamy soils, and are produced in regions with distinct seasonal variations, such as China, the United States, and Europe. This analysis provides valuable insights into the environmental requirements of different crops, helping to inform agricultural practices and optimize crop production based on regional climate and soil conditions. The performance of the proposed fused model has been meticulously compared with other state-of-the-art models and architectures. Table. 3 provides a comprehensive evaluation and thorough understanding of the proposed model's effectiveness with existing cutting-edge approaches in the field.

Table. 3: Comparison of the Proposed Model Performance Metrics

Model	Accuracy (%)
RF – KMC (Our Model)	**97.32% (Avg)**
SVM [9]	94.82%
RF – KNN [10]	95.21%
ANN [11]	96.76% (Avg)
DBSCAN [13]	89.54%
PART Algorithm [14]	98.33%
CNN [15]	97.86%

Table. 3 presents a comparison of various machine learning models based on their accuracy in the crop recommendation task. The combined Random Forest and K-Means Clustering (RF – KMC) modelachieves an average accuracy of 97.32%, demonstrating its effectiveness by leveraging both supervised and unsupervised learning techniques. Support Vector

Machine (SVM), known for its optimal hyperplane separation, attains a 94.82% accuracy, while the integration of Random Forest with K-Nearest Neighbors (RF – KNN) slightly improves performance to 95.21%. ANN exhibits strong capabilities with an average accuracy of 96.76%, benefiting from their multi-layered neuron structure. On the other hand, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN), an unsupervised algorithm, achieves a lower accuracy of 89.54%, highlighting its limitations in this context. The PART algorithm (Projective Adaptive Resonance Theory), which uses rule-based classification derived from partial decision trees, stands out with the highest accuracy of 98.33%. Convolutional Neural Networks (CNN), primarily used for image recognition, also perform exceptionally well with 97.86% accuracy, indicating their versatility and robustness in classification tasks. This comparison underscores the varied strengths of different models, with the PART algorithm and CNN leading in accuracy, while the RF–KMC model offers a balanced and highly effective approach by combining supervised and unsupervised learning techniques.

5. Conclusion

In conclusion, the integration of a Random Forest classifier and K-Means clustering algorithm for crop detection and classification significantly enhances agricultural practices by optimizing crop selection based on environmental conditions. The study's sophisticated approach leverages key parameters such as soil nutrients, temperature, humidity, soil pH, and rainfall to accurately categorize and group crops, showcasing superior performance over existing methodologies. The findings highlight the distinct environmental requirements of various crop groups, providing invaluable insights for aligning crop selection with optimal growth conditions. This alignment not only enhances agricultural productivity but also supports sustainable farming by minimizing resource wastage and maximizing yield [25][26].

Future work should focus on expanding the dataset to include a broader range of crops and environmental conditions, enabling the model to apply to diverse geographical regions beyond South-East Asia and India. Additionally, incorporating more advanced machine learning techniques, such as deep learning models, could further improve classification accuracy and robustness. Integration with real-time environmental monitoring systems and IoT technologies

References

1) G. Dhabal, J. Lachure and R. Doriya, "Crop Recommendation System with Cloud Computing," *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2021, pp. 1404-1411.

2) Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal, Crop yield prediction using machine learning: A systematic literature review, Computers and Electronics in Agriculture, Volume 177, 2020.

3) V. Y. Kulkarni and P. K. Sinha, "Pruning of Random Forest classifiers: A survey and future directions," *2012 International Conference on Data Science & Engineering (ICDSE)*, Cochin, India, 2012, pp. 64-68

4) A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha and S. Kundu, "Improved Random Forest for Classification," in *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4012-4024, Aug. 2018.

5) K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in *IEEE Access*, vol. 8, pp. 80716-80727, 2020

6) Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha and V. G. R. R. Devi, "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector," *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2020, pp. 736-741

7) D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using Machine Learning Algorithm," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2021, pp. 1466-1470

8) D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," in *IEEE Access*, vol. 8, pp. 86886-86901, 2020

9) D. Modi, A. V. Sutagundar, V. Yalavigi and A. Aravatagimath, "Crop Recommendation Using Machine Learning Algorithm," *2021 5th International*

Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-5.

10) T. K. Mishra, S. K. Mishra, K. J. Sai, B. S. Alekhya and A. R. Nishith, "Crop Recommendation System using KNN and Random Forest considering Indian Data set," *2021 19th OITS International Conference on Information Technology (OCIT)*, Bhubaneswar, India, 2021, pp. 308-312

11) Javaregowda, Madhuri & Indiramma, M. (2021). Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters. Indian Journal of Science and Technology. 14.

12) S. A. Z. Rahman, K. Chandra Mitra and S. M. Mohidul Islam, "Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series," *2018 21st International Conference of Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2018, pp. 1-4

13) M. S. Suchithra and M. L. Pai, "Data Mining based Geospatial Clustering for Suitable Recommendation system," *2020 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2020, pp. 132-139

14) P. Parameswari, N. Rajathi and K. J. Harshanaa, "Machine Learning Approaches for Crop Recommendation," *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, Coimbatore, India, 2021, pp. 1-5

15) R. Kumar, N. Shukla and Princee, "Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning," *2022 International Mobile and Embedded Technology Conference (MECON)*, Noida, India, 2022, pp. 168-172

16) M. Masrie, A. Z. M. Rosli, R. Sam, Z. Janin and M. K. Nordin, "Integrated optical sensor for NPK Nutrient of Soil detection," *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, Songkhla, Thailand, 2018, pp. 1-4

17) Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A., & Hickson, A. (2002). Data modeling and the application of a neural network approach to the prediction of total construction costs. Construction Management and Economics, 20(6), 465–472.

18) Y. Ma, X. Zou, Q. Pan, M. Yan and G. Li, "Target-Embedding Autoencoder With Knowledge Distillation for Multi-Label Classification," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 3, pp. 2506-2517, June 2024.

19) B. Van Essen, C. Macaraeg, M. Gokhale and R. Prenger, "Accelerating a Random Forest Classifier: Multi-Core, GP-GPU, or FPGA?," *2012 IEEE 20th International Symposium on Field-Programmable Custom Computing Machines*, Toronto, ON,

Canada, 2012, pp. 232-239.

20) E. Jedari, Zheng Wu, R. Rashidzadeh and M. Saif, "Wi-Fi based indoor location positioning employing random forest classifier," *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Banff, AB, Canada, 2015, pp. 1-5.

21) H. U. Dike, Y. Zhou, K. K. Deveerasetty and Q. Wu, "Unsupervised Learning Based On Artificial Neural Network: A Review," *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, Shenzhen, China, 2018, pp. 322-327.

22) F. Liu and Y. Deng, "Determine the Number of Unknown Targets in Open World Based on Elbow Method," in *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 986-995, May 2021.

23) G. Racolte *et al.*, "Spherical K-Means and Elbow Method Optimizations With Fisher Statistics for 3D Stochastic DFN From Virtual Outcrop Models," in *IEEE Access*, vol. 10, pp. 63723-63735, 2022.

24) T. Kohonen, "Correlation Matrix Memories," in *IEEE Transactions on Computers*, vol. C-21, no. 4, pp. 353-359, April 1972.

25) Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., Usman, M. (2015). Principle and application of plant mutagenesis in crop improvement: a review. Biotechnology & Biotechnological Equipment, 30(1), 1–16.

26) Abdallah, N. A., Prakash, C. S., &McHughen, A. G. (2015). Genome editing for crop improvement: Challenges and opportunities. GM Crops & Food, 6(4), 183–205.