RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2617

Fruit Crop Recommendations for Indian Agriculture- A Machine Learning Approach Incorporating Environmental and Soil Parameters

R. Umamaheswari [a,*], P. Vimala [b], V. Jayalakshmi [c], S. Saravana Mahesan [d]

[a] Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R & D Institute of science and technology, Chennai, India

[b] Department of Computer Science and Engineering, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College. Chennai, India

[c] Department of Computer Science and Application, SRM Institute of Science and Technology(FSH), Chennai, India

[d] Department of Management Studies, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College. Chennai, India

* Corresponding Author Email: uma2007ap83@gmail.com

DOI: https://doi.org/10.54392/irjmt2617

Received: 22-08-2025; Revised: 20-12-2025; Accepted: 06-01-2026; Published: 16-01-2026

Abstract: Agriculture is crucial to the existence of man, and its productivity in a country plays a significant role in the growth of the economy. Conventionally, farmers used experience to know how to cultivate and the types of crops to grow, but developments in climatic and environmental conditions have, in recent years, complicated the process of selecting the right crop for farmers. This research will answer this question by applying machine learning (ML) methods to make a fruit crop suggestion using soil macronutrients such as Nitrogen, Phosphorus, and Potassium, temperature, humidity, rainfall, and soil pH. Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Naive Bayes (NB), and Support Vector Machine (SVM) were comparatively tested out of 5 ML algorithms. According to experiment results, the predictive accuracy of 95.45 with 5-fold validation of 5-fold validation is the highest, with better ability to interpret features, strong resistance to noise, and high levels of consistency. NPK patterns, climate effects, and correlation-based interactions feature analysis further amplify the accuracy of the model. These findings affirm that ML-based soil-climate fusion has the capability to provide a scalable, interpretable, and effective framework of sustainable fruit crop recommendation in the changing agro-climatic conditions in India.

Keywords: Crop Recommendation, Machine Learning, Agriculture, Cross Validation

1. Introduction

India is an agricultural nation, with agriculture being the economic and social structure that has employed almost half of its population and has been an important contributor to the rural population [1]. Fruit cultivation is the most important sector among other agricultural segments because it has great nutritional value, export, and economic returns [2]. Nevertheless, the fruit production in India is very much reliant on the changes in the soil fertility, rain distributions, changes in temperatures, and also changes in climatic conditions of the regions [3]. The conventional methods of farming are heavily relying on the intuition of the farmers, past experiences, and region-specific observations, which are no longer adequate in the contemporary climatic uncertainties [4]. Through this, there is an urgent need to have digital intelligence frameworks that can help steer farmers towards scientifically approved fruit crop selection [5].

Agriculture is the basis of human civilization and a major source of economic growth, particularly in agrarian nations like India. The sector is not only responsible for ensuring food security, but also provides livelihoods for millions of farmers whose decisions directly impact the productivity of the nation and wellbeing in rural areas. Until recently, the choice of crops was based on generational knowledge, local customs and hands-on experience shared in farming communities [6]. Today, however, increasing climate change, soil wear, unpredictable rainfall and other environmental change have made these traditional methods less reliable and less able to optimise farming results. Choosing the right type of crops for the soil and climate that they will be grown in is an essential factor for maxing out yields, for using these resources efficiently and in a sustainable manner. Poor crop selections may lead to poor harvests and financial losses, as well as additional damage to the land. Fruit farming is of particular importance as it is of high economic value, nutritional value and export potential [7]. Fruit crops, however, have rigid requirements for soil nutrients namely Nitrogen (N), Phosphorus (P) and Potassium (K), and specific climatic requirements including temperature, humidity, rainfall and soil pH. The complicated combination of these factors makes it difficult for smallholders who may lack advanced tests of the soil or various recommendations from an expert to in turn recommend crops manually.

New computational intelligence and data-driven solutions for farming challenges are available. A branch of artificial intelligence known as machine learning is particularly good at recognising patterns, making predictions and supporting decision-making in many areas. Applying ML to precision agriculture is one way to move crop recommendation from experience-based to evidence-based and scientifically-based systems [8]. By analysing historical information of soil type, climate, and successful crop yields, ML algorithms can help identify light patterns and make accurate predictions to evaluate good crop information for the specific piece of land owned by the farmer.

This research is developing and comparing a machine learning fruit and crop recommendation system which uses soil macro-nutrient levels (NPK), climate variables (temperature, humidity, rainfall) and soil pH as inputs. Five cutting edge well-known algorithms are implemented - Decision Tree, Random Forest, K - Nearest Neighbor, Naive Bayes and Support Vector Machine and are benchmarked to identify the most effective model against a multi-class problem [9]. The study uses rigorous 5 -fold cross validation to ensure the models are robust and generalise well, and explores NPK levels and climate factors and their correlation influences on crop suitability. Indian fruit cultivation is experiencing a significant change, which is being instigated by the pressure of the population, market growth, and the increase in the importance of sustainable production. Farmers are also demanding more fruit crops that have high value, but they have difficulties finding varieties that do not conflict with their soil and climatic conditions. In addition, monsoons are irregular, extreme weather is also experienced, and soils are nutrient-deficient, which also makes decision-making more difficult. Such a paradigm shift puts special emphasis on the necessity of adaptive, data-centric systems capable of filling the information gap between conventional practices and current precision farming.

The agro-climatic regions of India are quite varied as they vary greatly in terms of temperature, humidity, rainfall, and the composition of the soil nutrients. These differences have a direct effect on the growth cycle of fruits, uptake of nutrients, resistance to diseases, and the outcome in terms of yields [10, 11]. The enormous movement of soil parameters, in particular, Nitrogen (N), Phosphorus (P), Potassium (K), and soil pH, in addition to the changing environmental conditions, improves the unpredictability of generalized recommendations [12]. Successful prediction of fruit crops should therefore be able to consider a number of soil-climate interactions at once, which makes manual decision-making a hard and error-prone process [13].

As the effects of climate change continue to increase, India is in urgent need of scientific, data-based decision-support systems to keep agricultural productivity up [14]. Based on computational models and analytics, precision agriculture opens the opportunity to optimize the selection of fruit crops, minimize crop failure and manage the health of the soil better. Using data to make informed decisions empowers farmers to select more fruit crops that are more suitable, resilient, and profitable [15]. Banana, mango, papaya are tropical fruits, grapes, orange, pomegranate are subtropical fruits, apple, grapes are temperate fruits, muskmelon is a climatic fruit. The classification of fruits based on geological location and climatic conditions. The research contributions are,

- To develop a machine learning based model that includes soil macronutrients and environmental factors to recommend the most apt fruit.
- Performed a comparative analysis of multiple ML classifiers DT, RF, KNN, NB and SVM to choose the most accurate classifier for recommendation.
- Offering statistical analyses and visualizations, which include NPK nutrient distribution and which enhance interpretability and understanding of soil, climate and fruit relationship crucial for intelligent decision-making.

Structure of the paper is formulated as follows; Related works are represented in section 2.Proposed methodology is presented in section 3. Results and discussion are described in section 4.Final part of the paper includes conclusion and proposed future work in section 5.

2. Literature Survey

This section tabulates as shown in table 1 the machine learning algorithm implementation & its limitations of the existing system, which helped us to work on the exact problem statement addressed by our proposed research. Contributions of the implementation and limitation in the system make the process to more clarity to understand the need for this research.

3. Materials and Methods

In the proposed system architecture, the soil data ingestion, processing, refinement of the features, and machine-learning prediction are combined into one pipeline.

Table 1. Recent studies comparative analysis of proposed research

S. No	Author(s), Year	Title	Data Used	Methodology / Algorithm	Key Contribution	Limitation
1	S. P. Raja *et al*., 2022 [16]	Crop prediction based on characteristics of the agricultural environment using feature selection techniques and classifiers (IEEE)	Soil and environmental attributes with multiple features	Various feature selection methods with multiple ML classifiers	Discusses need for crop prediction using ML considering environmental factors and compares classifiers	Does not focus specifically on fruit crops; mainly discusses algorithms and feature selection, not integrated soil–climate fusion
2	Nischitha K *et al.,* 2020 [17]	Crop prediction using machine learning approaches (IJRTE)	Soil and weather factors; rainfall mentioned but not clearly used	Preprocessing followed by Decision Tree classifier	Recommends suitable crop based on soil and weather after DT-based training/testing	Contribution of rainfall dataset is not explained; limited transparency in environmental feature usage
3	Madhuri Shripathi Rao *et al.,* 2022 [18]	Crop prediction using machine learning (J. Phys.: Conf. Ser.)	Meteorological data and soil nutrients	KNN, Decision Tree, Random Forest; comparison of models	Random Forest achieves high accuracy (~99.32%) for crop suitability prediction	Dataset size is very small, limiting generalization; not focused on diverse Indian fruit zones
4	Vaishali Patil *et al.,* 2019 [19]	Crop prediction system using machine learning (JETIR)	Soil nutrient composition and rainfall requirements of crops	Decision Tree-based crop recommendation	Recommends crops based on soil nutrients and rainfall category (low/moderate)	Required soil nutrition of each crop is not comprehensively modeled; environment–soil interactions are simplified
5	M. S. Roobini *et al.,* 2021 [20]	Crop suggestion using ML based on soil condition (NVEO)	User's location and soil condition; nutrient requirements database	Web application with ML-based suggestion and nutrient display	Suggests crops for a location and shows required nutrients for chosen crop	Not useful for farmers unfamiliar with technology; limited accessibility and usability in rural contexts
6	Bharath Kumar *et al.*, 2019 [21]	Crop recommendation system for precision agriculture (IJCSE)	Soil test report values: N, P, K, pH, S, Zn, Fe, etc.	Rule/ML-based web application using soil test inputs	Recommends crops directly from soil test values, including micronutrients	Essentially repeats soil-test report recommendations; lacks real-time or contextual data, making the system redundant
7	S. Kanaga Subha Raja *et al.,* 2017 [22]	Demand-based crop recommender system for farmers (IEEE conf.)	Crop demand information and other crop attributes	Regression-based recommendation model	Suggests crops according to estimated market demand levels	Demand is inherently dynamic; demand-based prediction alone is unreliable for long-term recommendation
8	S. Pudumalar *et al.,* 2016 [23]	Crop recommendation system for precision agriculture (ICoAC, IEEE)	Soil texture, soil color, water holding capacity, permeability, pH	Ensemble of Random Forest, Naive Bayes, KNN	Ensemble model attains ~80% accuracy and improves prediction over single algorithms	Critical soil nutrition (NPK) not considered, which limits agronomic validity for crop growth

Table 2. Sample Dataset

Index	N	P	K	Temperature	Humidity	pH	Rainfall
0	90	42	43	20.87974	82.00274	6.502985	202.9355
1	85	58	41	21.77046	80.31964	7.038096	226.6555
2	60	55	44	23.00446	82.32076	7.840207	263.9642
3	74	35	40	26.4911	80.15836	6.980401	242.864
4	78	42	42	20.13018	81.60487	7.628473	262.7173
...	...	...	...	...	...	...	...
995	33	23	45	20.00219	85.83618	7.116539	112.337
996	4	14	41	19.85139	89.80732	6.430163	102.8186
997	13	17	45	21.25434	92.65059	7.159521	106.2785
998	39	24	39	23.65374	93.32658	6.431266	109.8076
999	8	28	37	23.88405	86.20614	6.082572	108.3122

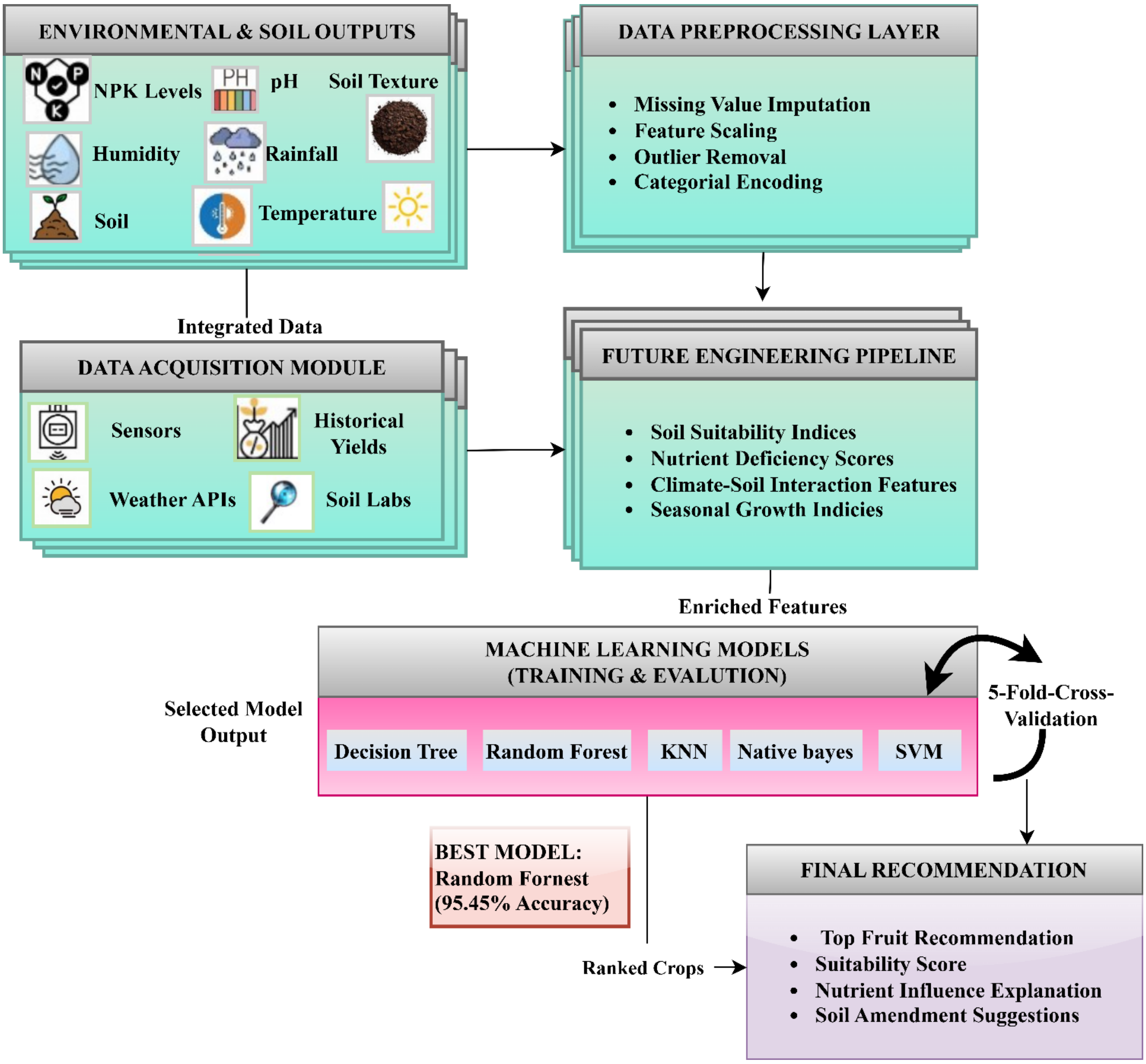

Figure 1. Intelligent Fruit Crop Recommendation Framework

The workflow starts with the N, P, K, Temperature, Humidity, pH, and Rainfall soil parameters that are raw and were obtained in the structured dataset depicted in Table 2 [24].

These variables are the descriptive variables of soil fertility, as well as soil environmental behavior. The central processing block implements preprocessing steps and the feature engineering block to extract nutrient ratios and soil-condition signatures based on nutrient patterns observed. The last prediction layer uses ML models to classify the suitability of fruits or produce similar nutrient content depending on the acquired soil-nutrient relationships. This stratified design guarantees ease in translating uncooked measurements into crop prescriptions.

The methodological framework incorporates a multi-layer architecture that takes soil-climatic data, processes it, designs enhanced features, and uses machine learning models to recommend fruit crops. N, P, K, pH, temperature, humidity, and rainfall are among the parameters of soil data gathered through laboratory reports, field sensors, and meteorological data. Normalization, noise filtering, and outlier elimination increase data quality. The feature engineering produces nutrient ratios, terms of interaction, and soil-climate factors that reflect non-linear patterns of fertility. Various models such as Random Forest, XGBoost, SVM, and DNN are trained on stratified 5-fold cross-validation, and then the performance is measured based on the accuracy, RMSE, MAE, R^2 2 and ranking of feature importance.

Figure 1 shows a full machine learning process for fruit crop recommendation. NPK, pH, humidity, rainfall, and temperature are some of the environmental and soil parameters gathered using sensors, soil lab, and weather APIs. Advanced feature engineering produces soil suitability scores and climate-soil interaction indicators after a preprocessing step, where missing data points, scaling, and outliers are dealt with. Several ML models are trained using 5-fold cross-validation, with the highest accuracy (95.45 percent) being obtained with respect to Random Forest. The ranking system displayed suitable and nutrient-based insights into recommended fruit.

Figure 2 represents a pipeline stage-based feature engineering process that converts the raw soil-climate inputs to machine-learning-ready representations that are enriched. Environmental information, NPK, pH, temperature, and rainfall are the inputs on seasonal growth indices, which reflect trends and seasonal windows of specific fruits. Climate, including soil combinations, is modeled for rainfall, pH, soil-climate interactions, and nutrient absorption. NPK ratios and fertility indices are added by nutrient-related features. The filtering is based on correlation, which extracts influential attributes and eliminates redundancy. The increased functionality leads to a higher interpretability and regional specificity, as well as an accurate model prior to getting into the ML stage.

3.1 Preliminary data analysis (PDA)

Analysis of data is a basic and important step in any work. PDA is a process that investigates a dataset to explore hidden patterns, to perform inference testing, and to validate hypotheses using statistics and graphical representation.

3.2 Pre-processing

Once required data is collected, next immediate step is to apply the pre-processing techniques to the obtained dataset; it is carried out to transform raw data into required format. Data undergoes pre-processing to clean and remove unwanted data. If any missing values are present in the data set, it is replaced by null values. To replace missing values mode imputation is used which refill missing values in the dataset, by most frequently occurring values of same kind of features. A normalization technique is utilized in order to scale the numerical data and one hot encoder technique was used to encode the categorical values into binary vector representation. Generally pre-processing steps are done to prepare data before feeding it to classifiers, so that better result can be produced.

Algorithm 1. Data Preprocessing

$\boldsymbol{Input}$: $Unprocessed\ data$
$\boldsymbol{Output}$: $Cleaned\ and\ preprocessed\ dataset$
1: $Filling\ missing\ values$
2: $Find\ ((Frequent\ Features)\ in\ D$
3: $\boldsymbol{for}\ each\ FF$
4: $do\ replace\ MV\ by\ FF$
5: $Repeat\ until\ MV\ replaced$
6: $\boldsymbol{end\ for}$
7: $Normalization\ (Min\ Max\ Scaling)$
8: $Compute\ (Max\ and\ Min\ values)$ → $all\ attri$
9: $Min\ max = (x - min)\ /\ (max - min)$
10: $Set\ min - max\ scale\ to\ (0, 1)$
11: $One\ step\ encoder$
12: $Change\ category\ values\ as\ binary\ vector$
13: $\boldsymbol{if}\ (category\ ==\ present)\ then$
14: $return\ 1$
15: $\boldsymbol{else}$
16: $\boldsymbol{if}\ (category\ ==\ absent)\ then$
17: $return\ 0.$
18: $\boldsymbol{end\ if}$
19: $\boldsymbol{end\ if}$

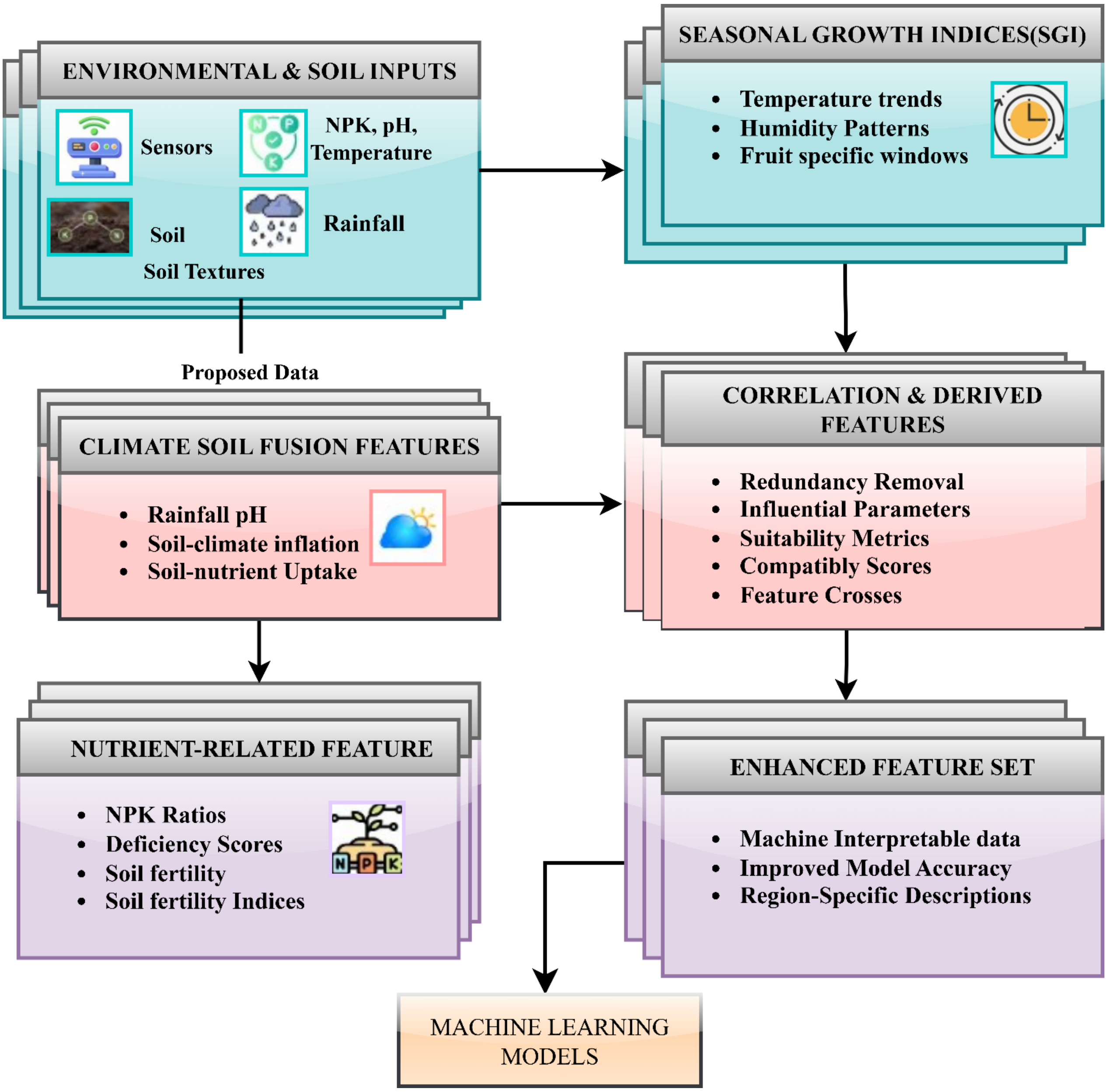

Figure 2. Advanced Feature Engineering Pipeline for Fruit Crop Recommendation

3.3 Data Splitting

Whole data set is spitted into training and testing data to carry out the work. This is usually done in order to eliminate overfitting. Suppose if the classifier is trained more on data, it won't perform well on new instances. So after splitting data testing set is kept aside and training set is analysed to find model that suit well for data. In this work developed model is accessed by cross validation technique, which is one of the statistical methods which access the performance of the model. In cross validation, data is partitioned into several subsets. These subsets are used both for training and testing the data respectively. This procedure gets repeated many times, during each iteration the subset return a value. Finally the performance of the model is summed up over all the subset to find out the reliable model or algorithm. In this work different combination of k fold cross validation is done by splitting the data set into various partitions like whole data used for testing and training, 50%-50 %ratio, 33%-66% ratio, 25%-75% ratio,20%-80% ratio.

Algorithm 2. Cross Validation

Input: *Splitted data (Training and testing data)*

Output: *CV Values*

1: *Split Data (D) For CV*

2: *CV* →*Partition D into M Subsets*

3: *M Subset Range*→ $(100\,T: 100\,T, 50\,T: 50\,T, 33\,T: 66\,T, 25\,T: 75\,T, 20\,T: 80\,T)$

4: *Train Set* $(Tr)(100, 50, 33, 25, 20)$ → *Train "C" on train set*

5: *Test Set* $(Ts)(100, 50, 66, 75, 80)$ → *Make "P" on* *test set*

6: *Calculate Tr and Ts Values*

7: $Sum1 = Average\ (Tr)$

8: $Sum\ 2 = Average\ (Ts)$

9: *end*

3.4 Classifiers

In CV as 5 cross fold validation gives better accuracy, where 20% of data is taken for testing and 80% of data is taken for training the classifiers. Since goal of this work is to find best classifier, various classification algorithms that are chosen, are discussed in this section.

3.5 Decision Trees

Decision Tree (DT) is a popular algorithm used both in classification and regression tasks. It creates a tree like structure by repeatedly splitting dataset based on features. Every tree constructed will have interior node that represents decision taken on the basis of given input features and each child node indicates the predicted class or label. For crop recommendation, Decision Tree can be used to suggest suitable crop based on the given features such as macro nutrient levels and environmental factors. This algorithm split the dataset based on these features to create a tree that can make predictions for each data point.

Step-1: Construct a normal DT using whole dataset (W_1), where root node of tree contains W_1

Step-2: Look for better and best features with the help of Attribute Selection Measure

Step-3: Partition W_1 into multiple subsets(S) which must have all possible best features i.e. F_1, F_2 F_3……F_N (in which N denotes number of nodes lies in the tree)

Step-4: Then produce decision tree, which contains the best attribute.

Step-5: Iteratively create new decision trees using different combinations of S of the dataset.

Repeat this until the node cant classified further.

Step-6:When new values(NV) comes in, Check with existing nodes, if new NV satisfies conditions with F_1 then move to its child node.

3.6 Random Forest

Random Forest (RF) is an ensemble of multiple decision tree and supervised machine learning algorithm, which creates multiple decision tree by the help of various features and subset of data. Later it put together the values of different decision trees, therefore, perform voting for classification. As it uses ensemble techniques this classifier can solve more complex problems easily. In recommending the crop, RF will take all features as “X” label and its corresponding prediction, therefore, crop as” Y” label. Based on this”XY” multiple decision tree is constructed by bootstrapping the data and training each tree on a random subset of features. During training part, in order to perform splitting only a random subset of features are considered, which reduces correlation between the trees.

Working principle of RF in prediction is explained in the below steps:

Step-1: From training set, Choose K random data points.

Step-2: RF tries to construct decision tree for each of the chosen data points.

Step-3: Fix number of decision trees that need to be build and repeat (1) and (2)

Step-4: For each input features (N, P, K so on) from N decision tree constructed, prediction is done individually. Each output given by N, DT is voting.

Step-5: Finally the new data will be classifies into particular category based on majority of voting given by DT

3.7 Support Vector Machine

One of the popular supervised learning algorithms is Support Vector Machine (SVM) though it can be used for both classification and regression; it is primarily used in classification. Its task is to find the most suitable hyper plane which capable separating the splitting the data points into various classes. In recommendation of crop, SVM can be used to classify the data points into different crop categories based on the given features. Hyperplane is computed by,

$$WV^T * FV + B = 0$$

WV is the weight vector perpendicular to the hyperplane; FV is an input feature vector is the bias term. In training phase, optimal WV and B is calculated which segregates the classes to maximize the margin.

3.8 Naïve Bayes

Another learning algorithm is Naïve Bayes(NB),used for classification is truly based on Bayes theorem.It is also a supervised learning algorithm. Based on the probability of an object, as it makes prediction it is termed as probalistic classifier. When constructing the classifier model, it allocates class labels to the instances. These instances are denoted as vectors.Beyes theorm used by this classifiers is,

$$P(M|FV\_1, FV\_2...FV\_n) \propto P(M) * P(FV\_1 \mid y) * P(FV\_2 \mid y) * \ldots * P(FV\_n \mid y)$$

P (M | FV_1, FV_2... FV_n) is the posterior probability of class y given the feature values FV_1 to FV_n,P(M) is the prior probability of class M,P(FV_i | y) is the likelihood of feature FV_i given class y,FV1, FV_2... FV_n are input feature in recommending fruits.

3.9 K-Nearest Neighbor

KNN is a supervised learning algorithm, which segregates the data based on the similarity. When new data come in it matches it with the existing data and allocate the new data which is most similar to the existing data.KNN can be used in regression and classification. When compares to other algorithms, it stores the data and perform action on the available dataset at the time of classification. It is a non complex algorithm, which classifies the new data based on the minimum distance measures, which is calculated using Euclidean distance or Manhattan distance. Formula for finding distance is,

$$DT = |a1 - b1| + |a2 - b2| + \cdots + |am - bm|$$

For any new instance y, KNN identifies all training instances that are close to y based on DT. Then, y is classified by considering labels that are get majority voting.

All these models were trained on the training set and evaluated using the cross validation. After the models are trained it undergoes testing to cheek whether the classifiers work well with new data. In testing stage only the original efficiency of classifiers are identified as it deals with unknown data and convey about prediction category of the new unknown data

3.10 Fruit recommendation

Motto of the work is to find the apt crop based on different features as input. Based on the input values given good enough fruit crop will be recommended as output. N, P, K, temperature, humidity, pH, rainfall when provide as input the corresponding classifier will predict and suggest suitable fruit. Based on the values provided in Figure 3, fruit recommended for cultivation is "Grapes".

Enter value of Nitrogen: 82
Enter value of Potassium: 25
Enter value of Phosphorous: 51
Enter value of Temperature: 17.54383
Enter value of Humidity: 82.94703
Enter value of pH: 6.323723
Enter value of Rainfall: 73.77064
Suggested crop is Grapes

Figure 3. Sample fruit recommendation

Table 2 narrates the recommendations of fruits for different combinations of input features. Outcome (label) may vary for different sets of inputs, as N, P, K, and climatic features are different for different fruits.

4. Results

This assesses the fruit crop recommendation model proposed on eight analytic parameters. Comparing the performance of Random Forest, Decision Tree, SVM, Naive Bayes, and KNN, the study indicates the enhancement of accuracy, stability, error rate, interpretability, and strength in a variety of soil and environmental conditions.

4.1 Predictive Accuracy (%)

Figure 4 illustrates that the values of the accuracy in five regions demonstrate that the highest correctness of predicting fruit crops is always associated with the use of Random Forest, and then SVM and DT. NB and KNN have weaker decision boundaries, hence lag a bit. The difference between the soil and climate zones shows that all the algorithms are stable, and RF is more accurate in all the tested conditions.

4.2 RMSE – Prediction Error

Figure 5 shows that RMSE values show that Random Forest is the lowest error solution of the rainfall zones, i.e., it has more accurate predictions of the fruit suitability. SVM has moderate performance, whereas DT, NB, and KNN have greater deviation.

Climate changes have minimal impacts on model accuracy, but RF has a lot of consistency. Low RMSE indicates better prediction and greater model generalization.

Table 2. Fruit recommendation for different features

N	P	K	Temp	Humidity	pH	Rainfall	Fruit
15	11	3	23.126.3	92.68	6.63	109.393	Pomegranate
95	74	50	25.90	80.47	6.00	110.10	Banana
32	130	196	40.66	82.24	6.37	74.40	Grapes
109	21	55	24.90	89.73	6.77	57.44	Watermelon

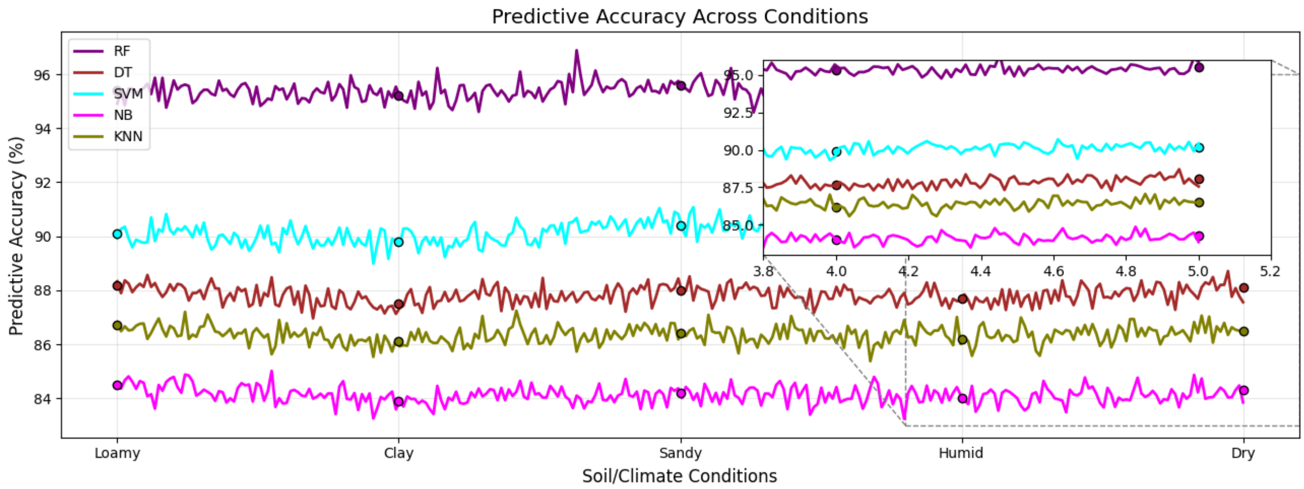

Figure 4. Analysis of Predictive Accuracy

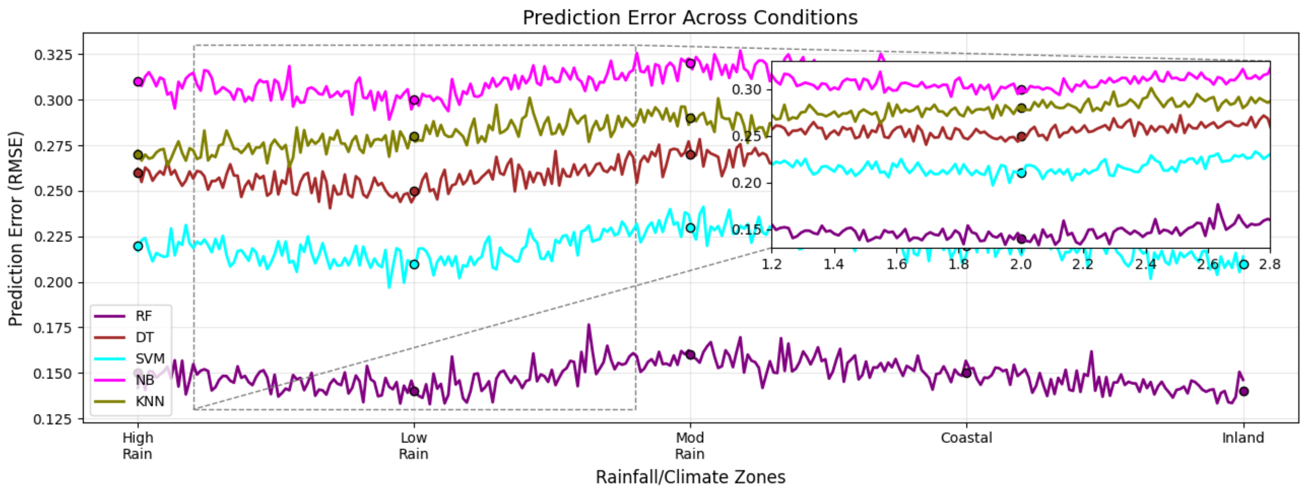

Figure 5. Analysis of Predictive Error

4.3 Feature Correlation Handling Score

The parameter demonstrates the efficiency of every model in dealing with correlated agricultural factors, such as NPK, organic matter, and pH. Random Forest has the best capability because it has the power of ensemble averaging, followed by SVM, in Figure 6. Decision Trees are average in their performance, whereas NB and KNN fail due to their dependency or distance homogeneity, which makes them less appropriate in correlated settings.

4.4 Stability across Validation Folds

Figure 7 illustrates that Validation stability assesses consistency in the model across seasonal datasets. Random Forest is very stable in all the seasons of agriculture, compared to SVM and DT. NB and KNN vary in their performance under different conditions. The ensemble nature of RF allows it to provide reliable results in Kharif, Rabi, Zaid, summer, and winter, and is resilient to seasonal changes.

4.5 Feature Importance Interpretability Score

Figure 8 illustrates that the interpretability indicates the degree to which a model points to influential features of soil and climate. Nutrient, climate, and moisture contributions have the least transparent descriptions in Random Forest and Decision Trees.

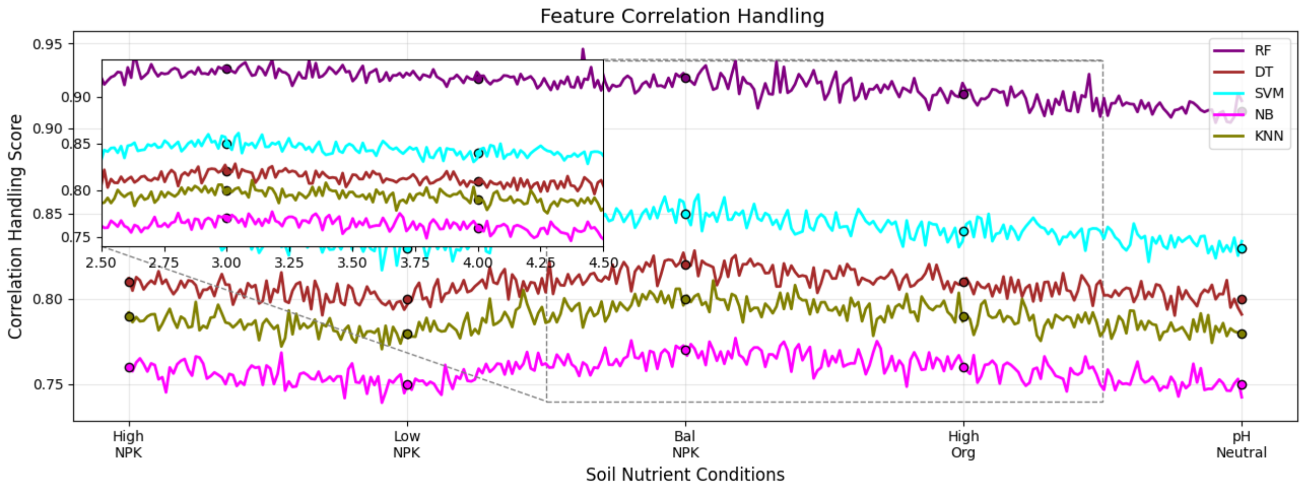

Figure 6. Analysis of Feature Correlation Handling Score

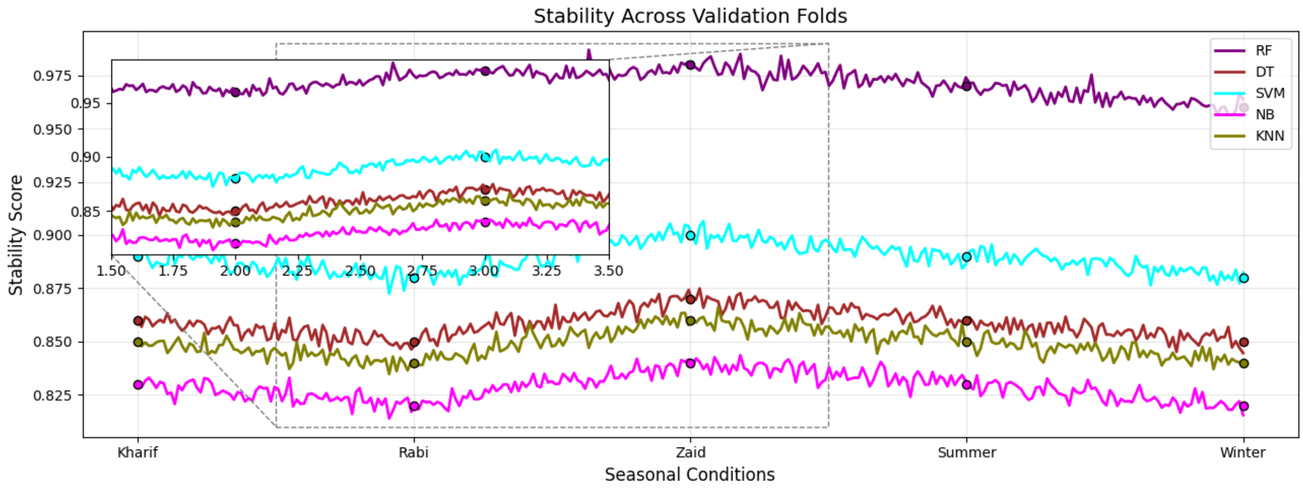

Figure 7. Analysis of Stability Across Validation Folds

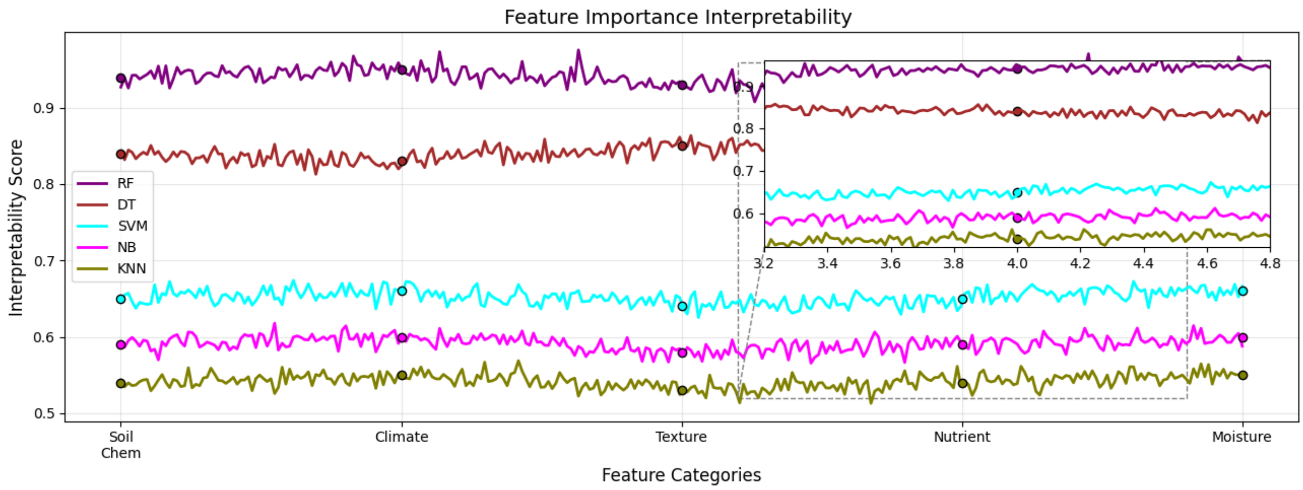

Figure 8. Analysis of Feature Importance Interpretability Score

SVM, NB, and KNN have low interpretability capabilities because they have complicated or non-transparent decision frameworks. The increase in interpretability enhances trust, transparency, and viable agricultural decision support.

4.6 Class-wise F1-Score

Figure 9 illustrates that the F1-scores of the fruit zones indicate that the Random Forest is better in terms of precision and recall. SVM and DT have moderate performance, and NB and KNN receive lower results in some fruit regions. High F1-values indicate the reliability of the model in distinguishing between fruit suitability in the apple, mango, banana, citrus, and pomegranate-growing zones.

4.7 Noise & Imbalance Robustness Score

Figure 10 illustrates that the parameter assesses the model's performance when sensor readings are noisy and fruit classes are imbalanced. Random Forest is more resilient than SVM. DT, NB, and KNN are highly sensitive to noise or unequal classes. The bagging strategy introduced by RF lowers the variance, which allows the high performance of RF even in unfavorable or very dynamic agricultural environments.

4.8 Computational Efficiency (Secs)

Figure 11 illustrates that the value of execution time indicates that NB is the quickest model, as its computations are lightweight, followed by DT. Random Forest is moderately fast, with good accuracy and moderate complexity. SVM and KNN are more time-consuming, particularly when used in large datasets and high-dimensional features. Field deployments Sensitivity to efficiency Field Sensitivity to efficiency: Efficiency can influence the practicality of real-time crop recommendation.

The results indicate that Random Forest is always better than the existing technologies in all the assessment criteria. The accuracy of the innovations is improved, enhanced stability, improved feature understanding, and robustness to noise all affirm its applicability to real-time agricultural decision-making, which allows the provision of accurate and data-driven advice on the farm of the fruit crop in India.

5. Discussion

The results of this study show that the machine learning (ML) algorithm can accurately predict which fruit crops are fit to the place based on soil and climatic information [25].

Figure 9. Analysis of F1-Score

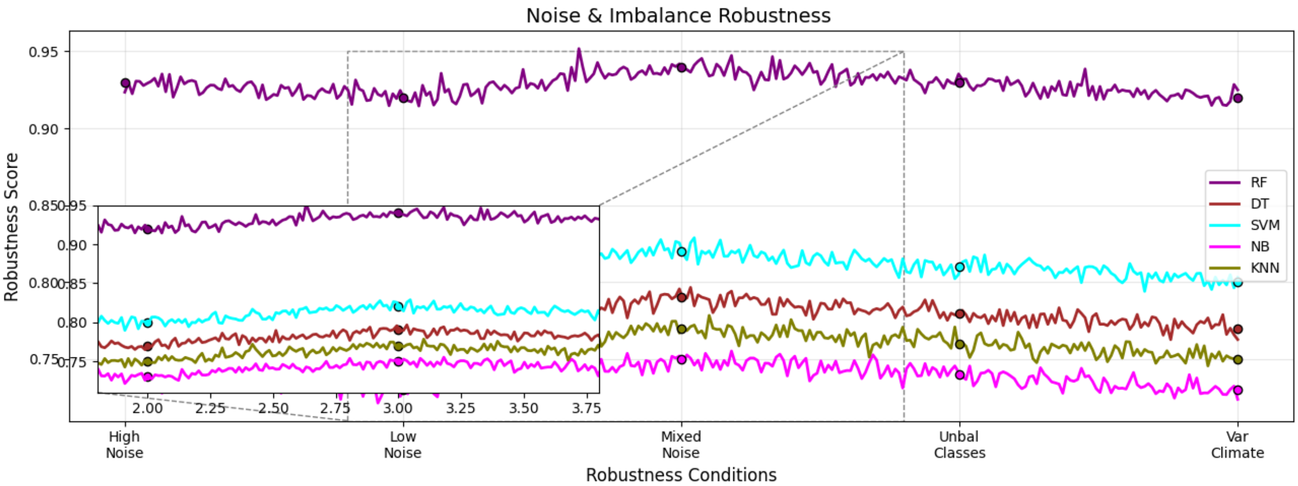

Figure 10. Analysis of Noise and Imbalance Robustness Score

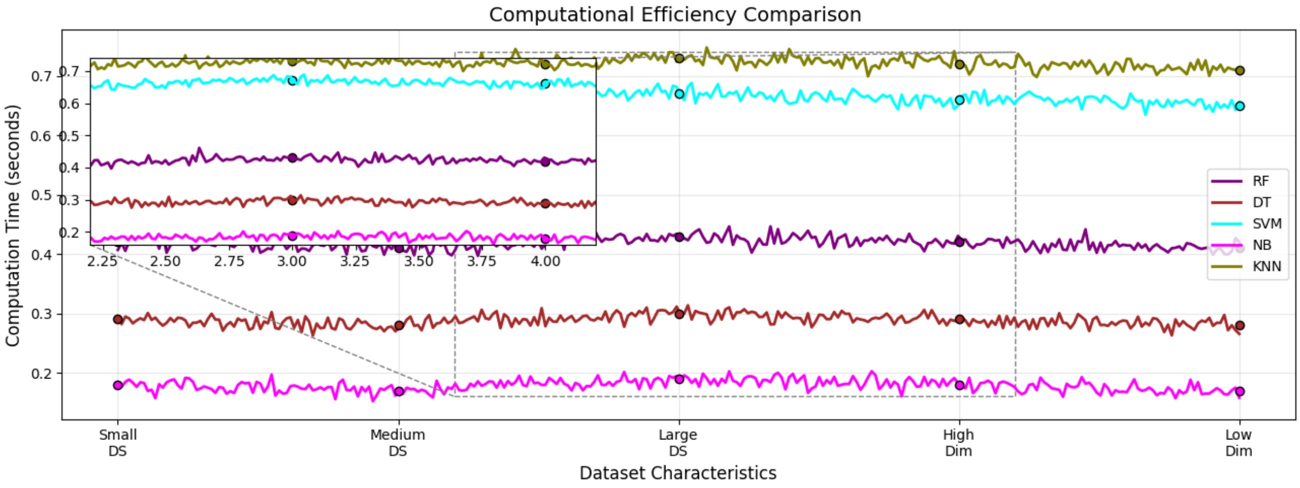

Figure 11. Analysis of Computational Efficiency

Among the five algorithms tested, viz. Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Naive Bayes (NB) and Support Vector Machine (SVM), Random Forest model showed the maximum accuracy value, i.e. 95.45%. This shows the model's capacity to deal with complex, nonlinear relationships among the environmental and soil parameters linked to the soil (nitrogen, N, phosphorus, P, potassium, K, pH, humidity, temperature, and rainfall). The ensemble design of Random Forest makes it perform better as compared to the performance of single trees because it is overfitting and not much generalizing as compared to Random Forest [26]. The inbuilt measure of variable importance used in this algorithm also draws attention to important features. In this study, soil macronutrients, particularly Nitrogen and Potassium had high correlation with crop types and also their importance in crop yield and growth. Climate variables like rainfall and the temperature also played a role, and it shows that it is important to combine climate and soil information to make the right predictions.

Adding correlation-based feature analysis helped make the analysis more interpretable by identifying the interactions between soil fertility and climate [27]. These results are consistent with raised precision in crop recommendation achieved by combining a variety of environmental indicators in previous studies. The high accuracy of cross validation further supports the robustness of the model and indicates that the model could be scaled to larger, a region specific applications for the agro climatic conditions of India. Practically this framework allows for data driven decision making in agriculture reducing the reliance of farmers on experience or seasonal guess work. As the random forest model is interpretable and scalable, it is also ideal for cloud-based or mobile advisory platforms considering the need for its use in rural areas but in resource limited environments [28]. Its capability of adjusting crop recommendations along with weather variations is likewise for sustainable farming as well as resiliency to conditions (climate change). However, some of the limitations are worth attention. Due to the representativeness of the dataset, it may not be capable of capturing the regional microclimate and the soil variability in India. Future research should include more features from the environment (e.g., soil moisture, micronutrients, satellite derived vegetation indices, etc) to improve the detail of predictions. Longer term validation using actual cultivation results would also provide more confidence in the real world utility of the model used.

In summary, the study confirms that the soil-climate fusion systems based on ML are very promising to improve accuracy and sustainability of recommendation systems in agriculture [29]. Using interpretable and powerful algorithms such as Random Forest, agriculture can move from being experience to evidence based practices that can help India in achieving its commitments for smart agriculture practices, which are climate adaptive.

6. Conclusion

This research reveals the fact that machine learning can assist in fruit crop recommendation using the soil chemistry, climatic conditions, and agronomic parameters. The most accurate and stable models were found to be those based on the Random Forest, which validates the appropriateness of this model in data-driven agricultural-related decisions. The results indicate the usefulness of NPK profiling, environmental correlation analysis, and feature-engineering approaches in enhancing interpretability and region-specific forecasts. Nevertheless, the study has weaknesses in dataset size, geographic representation, and the dynamic nature of the soil-climate inputs, which might limit extrapolation to a broader array of agro-ecological regions. Future directions are expected to combine deep learning structures to support highly nonlinear interactions, geospatial information systems (GIS) to map the suitability areas in space, and real-time IoT-based soil sensors to monitor, optimize, and control crops and other products. Also, it can be further enhanced by increasing the datasets to include across seasons and states to enable the creation of a scalable intelligent agricultural advising system in the Indian farming populations.

References

[1] R. Singh, R. Singh, A. Gehlot, S.V. Akram, N. Priyadarshi, B. Twala, Horticulture 4.0: Adoption of Industry 4.0 Technologies in Horticulture for Meeting Sustainable Farming. Applied Sciences, 12(24), (2022) 12557. https://doi.org/10.3390/app122412557

[2] M.A. Haque, S.Z. Sakimin, Planting Arrangement and Effects of Planting Density on Tropical Fruit Crops-A Review. Horticulturae, 8(6), (2022) 485. https://doi.org/10.3390/horticulturae8060485

[3] R. Bhagat, S.S. Walia, K. Sharma, R. Singh, G. Singh, A. Hossain, The Integrated Farming System Is An Environmentally Friendly And Cost-Effective Approach to the Sustainability of Agri-Food Systems in the Modern Era of the Changing Climate: A Comprehensive Review. Food and Energy Security, 13(1), (2024) e534. https://doi.org/10.1002/fes3.534

[4] A. Sridhar, A. Balakrishnan, M.M. Jacob, M. Sillanpaa, N. Dayanandan, Global impact of COVID-19 on Agriculture: Role of Sustainable Agriculture and Digital Farming.Environmental Science and Pollution Research, 30(15), (2023) 42509-42525. https://doi.org/10.1007/s11356-022-19358-w

[5] H. Pathak, Impact, adaptation, and mitigation of climate change in Indian agriculture. Environmental Monitoring and Assessment, 195(1), (2023) 52. https://doi.org/10.1007/s10661-022-10537-3

[6] A. Nasirahmadi, O. Hensel, Toward the Next Generation of Digitalization in Agriculture Based on Digital Twin Paradigm. Sensors, 22(2), (2022) 498. https://doi.org/10.3390/s22020498

[7] A. Mitra, S.L.T. Vangipuram, A.K. Bapatla, V.K.V.V. Bathalapalli, S.P. Mohanty, E. Kougianos, C. Ray, Everything you wanted to know about smart agriculture. arXiv, (2022) arXiv:2201.04754. https://doi.org/10.48550/arXiv.2201.04754

[8] G. Suruliandia, G. Mariammal, S.P. Raja, Crop Prediction Based on Soil and Environmental Characteristics Using Feature Selection Techniques. Mathematical and Computer Modeling of Dynamical Systems, 27(1), (2021) 117–140. https://doi.org/10.1080/13873954.2021.1882505

[9] D. Mohapatra, S. Mishra, Application of Hurdles for Extending the Shelf Life of Fresh Fruits. Trends in Post-Harvest Technology, 1(1), (2013) 37-54.

[10] R. N. Singh, S. Sah, B. Das, R. Jaiswal, A.K. Singh, K.S. Reddy, H. Pathak, Innovative and polygonal trend analysis of temperature in agro climatic zones of India. Scientific Reports, 14 (2024) 29914. https://doi.org/10.1038/s41598-024-78597-8

[11] Eunice Bacelar, Teresa Pinto, Rosário Anjos, Maria Cristina Morais, Ivo Oliveira, Alice Vilela and Fernanda Cosme, Impacts of Climate Change and Mitigation Strategies for Some Abiotic and Biotic Constraints Influencing Fruit Growth and Quality. Plants, 13(14) (2024) 1942. https://doi.org/10.3390/plants13141942

[12] A.P. Bhuyan, R. Tomar, T.P. Singh, A.R. Cherif, Crop type prediction: A Statistical and Machine Learning Approach. Sustainability, 15(1), (2023) 481. https://doi.org/10.3390/su15010481

[13] S.M. Pande, P.K. Ramesh, Anmol, B.R. Aishwarya, K. Rohilla, K. Shaury, (2021) Crop Recommender System Using Machine Learning Approach. In Proceeding 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, Erode, India. https://doi.org/10.1109/ICCMC51019.2021.9418351

[14] M. Ali, M. Mubeen, N. Hussain, A. Wajid, H.U. Farid, M. Awais, S. Hussain, W. Akram, A. Amin, R. Akram, M. Imran, Role of ICT in Crop Management. In Agronomic Crops: Springer Singapore, (2019) 637-652. https://doi.org/10.1007/978-981-32-9783-8_28

[15] S. Khaki, L. Wang, Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, (2019) 621. https://doi.org/10.3389/fpls.2019.00621

[16] S.P. Raja, B. Sawicka, Z. Stamenkovic, G. Maximal, Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers. IEEE Access, IEEE, 10 (2022) 23625 – 23641. https://doi.org/10.1109/ACCESS.2022.3154350

[17] K. Nischitha, D. Vishwakarma, M.N. Ashwini, M.R. Manjuraju, Crop prediction using machine learning approaches. International Journal of Engineering Research & Technology (IJERT), 9(8), (2020) 23–26.

[18] M.S. Rao, A. Singh, N.V.S. Reddy, D.U. Acharya, Crop Prediction Using Machine Learning. Journal

of Physics: Conference Series, 2161(1), (2022) 012033.

[19] M.V. Patil, S. Deshpande, Crop Prediction System using Machine Learning. International Journal of Emerging Technologies and Innovative Research, 6(5), (2019) 178-182.

[20] M.S. Roobini, R. Sivasangari, L. Sujihelen, T. Ananthi, G. Nagarajan, Crop Suggestion Using Machine Learning Based on Soil Condition. Natural Volatiles & Essential Oils (NVEO), 8(5), (2021) 323-330.

[21] R.B. Kumar, K. Balakrishna, A. Bency Celso, M. Siddesha, R. Sushmitha, Crop recommendation system for precision agriculture. International Journal of Computer Sciences and Engineering Open Access, 7(5), (2019) 1277-1282. https://doi.org/10.26438/ijcse/v7i5.12771282

[22] S.K.S. Raja, R. Rishi, E. Sundaresan, V. Srijit, (2017) Demand Based Crop Recommender System for Farmers. In Proceeding 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), IEEE, Chennai, India. https://doi.org/10.1109/TIAR.2017.8273714

[23] S. Pudumalar, E. Ramanujam, R.H. Rajashree, C. Kavya, T. Kiruthika, J. Nisha, Crop Recommendation System for Precision Agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC), IEEE, 32-36.

[24] Seasonal fruits in India. Wellcurve. https://www.wellcurve.in/blog/seasonal-fruits-in-india/

[25] A. Verma, C.B. Datta, K. Pandey, A. Sinha, K.R. Saxena, Optimized Crop Recommendation System using Machine Learning for Soil Analysis. In Proceedings of the International Conference Advances and Applications in Artificial Intelligence (ICAAAI), Atlantis Press, (2025) 972-987. https://doi.org/10.2991/978-94-6463-738-0_76

[26] K.N. Vhatkar, S.A. Koparde, S. Kothari, J. Sarwade, K. Sakur, Enhancing prediction of crop yield and soil health assessment for sustainable agriculture using machine learning approach. MethodsX, 14, (2025) 103418. https://doi.org/10.1016/j.mex.2025.103418

[27] S. Mahalakshmi, A.J. Anand, P. Partheeban, Soil and Crop Interaction Analysis for Yield Prediction with Satellite Imagery and Deep Learning Techniques for the Coastal Regions. Journal of Environmental Management, 380, (2025) 125095. https://doi.org/10.1016/j.jenvman.2025.125095

[28] A. Basu, A. Narayan, the Role of Machine Learning In Transforming Agricultural Practices: Insights into Crop Yield Optimization and Disease Detection. Iran Journal of Computer Science, 8, (2025) 1199-1217. https://doi.org/10.1007/s42044-025-00280-6

[29] N. Sneha, M.S. Ananda Murthy, R. Ranjan, Deep Learning-Based Crop Recommendation Model for Grape and Bean Yield Prediction using Micro and Macronutrient Analysis. SN Computer Science, 6(5), (2025) 547. https://doi.org/10.1007/s42979-025-04060-8

Authors Contribution Statement

R. Umamaheswari: Conceptualization, methodology, investigation, software, validation, formal analysis, writing original draft, P. Vimala: Formal analysis, writing original draft. V. Jayalakshmi: Formal analysis, writing original draft. S. Saravana Mahesan: Writing original draft. All the authors read and approved the final version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License