Agri Crop Recommendation System Using Suitable Machine Learning Techniques

[1]Ananthi S, [2]Bauvya Shree T G, [3]Bhava Tharani R, [4]Geethanjali R, [5]Suchithra B and [6]Sathya R
[1,2,3,4,5]Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore, India.
[6] Kongu Nadu College of Engineering and Technology, Trichy, Tamil Nadu, India.
[1]ananthi68@gmail.com, [2]bauvyashree.t.g@sece.ac.in, [3]bhavatharani.r@sece.ac.in, [4]geethanjali.r2019cse@sece.ac.in, [5]suchithra.b@sece.ac.in,[6]sathyaphdkncet@gmail.com

Article Info
S. Venkatesh et al. (eds.), *Emerging Trends in Mechanical Sciences for Sustainable Technologies,* Advances in Intelligent Systems and Technologies,
Doi: https://doi.org/10.53759/aist/978-9914-9946-4-3_16

Abstract- In coastal states like Tamil Nadu, agricultural uncertainty lowers productivity due to a lack of modern technology, unstable climate changes, and a lack of good irrigation facilities. The ability to utilize modern technology in agriculture is mandatory in the current situation. Agricultural issues like crop rotation, crop forecasting, crop protection, water requirement, and fertilizer requirements can be solved using machine learning technology. It is possible to forecast and derive a precise model from the data using machine learning techniques. Agricultural crop recommendations based on productivity and season is proposed in this work. This might make it possible for prospective farmers to engage in improved agriculture. A farmer may receive a system of ideas to increase production and can move forward smoothly. Crops are advised based on quantity and weather concerns. The analysis of the crop dataset resulted in the recommendation of crops based on their productivity and growing season.

Keywords - K-Nearest Neighbors, Machine Learning, Random Forest, Support Vector Machine, Naive Bayes, Recommender System.

I. INTRODUCTION

According to the population projection report by the Census of India, Tamil Nadu is the sixth-largest by population and the top manufacturer of agricultural goods.Agricultural crop recommendation systems are gaining popularity due to their ability to provide farmers with precise information about crop yields and production. These systems utilize various machine learning algorithms to predict the most appropriate crops based on productivity and season, enabling farmers to make informed decisions about what crops to plant and when.

The dataset includes the cultivation of rice, groundnuts, cotton, coconuts, maize, and sugarcane, which are among the most grown crops in India. After acquiring the data, we utilized it to train and test various machine learning models such as naive bayes, k-nearest neighbours, random forest, and support vector machines.

The performance of the machine learning models was assessed by utilizing various evaluation metrics such as precision, recall, and k-fold cross-validation scores. These metrics are derived from the true positive, true negative, false positive, and false negative rates of each algorithm. After evaluating the performance of several algorithms, it was observed that the random forest algorithm exhibited the highest accuracy in predicting crop yields based on productivity and season.

The agricultural crop recommendation system developed in this study has the potential to transform crop selection for farmers, enhancing efficiency and effectiveness through personalized recommendations. By providing farmers with accurate and timely suggestions, the system can aid in decision-making and ultimately boost crop yields.

II. LITERATURE SURVEY

Techniques such as data pre-processing, feature selection, and model evaluation used to develop and test the system. The techniques helped in optimizing the system's accuracy and potential usefulness for farmers to increase their crop yields and income [1]. A hybrid recommendation system that adopts CBR-Case-Based Reasoning to increase the success of the system. The novelty of the model consists in the analysis of region-specific agricultural data for the prediction of future climatic [2]. Intelligent system that helps farmers make the best decision about which crop to grow based on productivity and season. Harvest data was analysedand product recommendations were made with 92 percent accuracy based on productivity and season [3]. Sonal Agarwal aims to improve the accuracy of crop yield prediction using intelligent techniques, combining

Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN) algorithms [4].The Random Forest algorithm was used to train the model, and the results showed high accuracy in predicting crop yield, indicating its potential usefulness for farmers in optimizing their crop yields and improving their income [5].An analysis method for agricultural total factor productivity using Stochastic Block Model (SBM) and machine learning. The SBM and machine learning techniques were used to analyse and predict the agricultural total factor productivity [6].A website which enables people to communicate with a machine learning model and generate predictions from given parameters[7].The collected and processed agricultural images and trained a model using the VGG-16 architecture.

The system achieves high accuracy in crop detection and recommendation, indicating the possibility of practical application in the field of precision agriculture [8][9].The decision tree-based model outperformed other techniques and could effectively assist farmers in making informed decisions about rice cultivation [10].Multispectral images of crop fields are gathered through remote sensing, and machine learning algorithms are then used to classify the crops and predict their yields.

Based on a study of the data gathered, the system suggests appropriate crop kinds and planting methods [11].The system is made up of data gathered from a variety of sources, such as satellite photography and ground-based sensors, and related locations are grouped together using clustering algorithms [12].A system for recommending crops that takes into account many elements including temperature, soil moisture, and precipitation and bases its recommendations on clustering and decision trees [13].A crop recommendation system that predicts the best crop for a given set of conditions more accurately using ensemble learning techniques including gradient boosting, random forests, and decision trees [14] [15].

III. PROPOSED SYSTEM

To assist farmers in selecting the best crops for their land and environment, crop recommendation systems can make use of the Random Forest type of machine learning algorithm. Each tree in a Random Forest ensemble is trained on a different subset of data and generates a forecast on its own. The ensemble's forecasts are combined, either by taking the average or by employing a majority vote, to produce the final prediction. This method produces a more precise and reliable prediction while assisting in reducing overfitting, a major issue in Random Forest algorithms. **Fig 1** depicts the architecture of the proposed Agricultural Crop Recommendations Based on Productivity and Season. **Fig 1** depicts the proposed model.

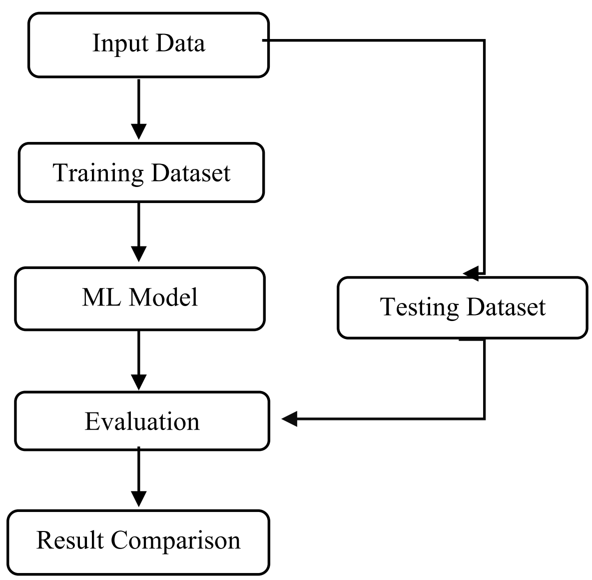

Fig1. Proposed System Architecture.

Data Collection

The Crop Production Dataset was compiled from several sources, including the Tamil Nadu Agricultural University, the Indian Meteorological Department, and Kaggle. This dataset comprises 13,000 records, with each record corresponding to a particular location in Tamil Nadu, India. It contains various attributes such as the state name, district name, crop year, season, crop name, area, and production. The dataset includes the cultivation of rice, groundnuts, cotton, coconuts, maize, and sugarcane, which are among the most grown crops in the region.

Table 1. Data Extraction for Coimbatore District

District	Year	Season	Crop	Area	Production
COIMBATORE	1997	Whole Year	Arhar/Tur	1393	700
COIMBATORE	1997	Whole Year	Bajra	518	910
COIMBATORE	1997	Whole Year	Banana	7269	285930
COIMBATORE	1997	Kharif	Banana	5619	183740
COIMBATORE	1997	Whole Year	Black pepper	131	30
COIMBATORE	1997	Whole Year	Cardamom	834	70
COIMBATORE	1997	Whole Year	Cashewnut	63	40
COIMBATORE	1997	Whole Year	Castor seed	199	70
. . .	. . .	. . .	. . .	. . .	. . .
COIMBATORE	2013	Whole Year	Coconut	8453 1	1212000000
COIMBATORE	2013	Whole Year	Coriander	138	48
COIMBATORE	2013	Kharif	Cotton(lint)	336	768
COIMBATORE	2013	Rabi	Cotton(lint)	12	29
COIMBATORE	2013	Whole Year	Dry chilies	481	131
COIMBATORE	2013	Whole Year	Gram	1162	811

With information on several districts, Coimbatore district data are displayed in **Table 1**. The number of Coimbatore crop varieties was large.

Data RRE-Processing

The data pre-processing techniques are essential for developing accurate and reliable agricultural crop recommendation systems. In the proposed work, various data pre-processing steps were performed, including data cleaning, data transformation, and data normalization. The data cleaning step involved the removal of missing and irrelevant data from the dataset. The data transformation step was performed to convert the data into a format suitable for analysis. Finally, data normalization was used to standardize the data range and improve the accuracy of the model. Overall, the data pre-processing techniques applied in this study are crucial for developing a robust and effective agricultural crop recommendation system.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step in the development of an agricultural crop recommendation system. It involves understanding the properties of the data and uncovering patterns and relationships between variables. In this phase, various data visualization techniques, statistical methods, and machine learning algorithms are applied to identify key factors affecting crop productivity and seasonality. EDA allows for the identification of potential outliers, missing values, and other data quality issues that need to be addressed before moving on to the modeling phase.

Model Selection

Random Forest

A Random Forest is trained for crop recommendation using historical data on crop yields as well as other pertinent elements, such as weather patterns, soil type, and water availability. The specific crop for a given set of input can then be predicted using the existing data and the Random Forest method. This can assist farmers in selecting the crops that are most suited to their land and environment, boost yields, reduce risk, and result in a more sustainable and successful agricultural enterprise. Random Forest Process is depicted in **Fig2.**

Random Forest model is shown in equation 1

$$f(x) = \frac{1}{N} * \text{sum } i = 1^{N} T_{i(x)} \tag{1}$$

Were,

f(x) - Prediction function
N - Number of trees in the forest
T_i(x) - Prediction of i(x)

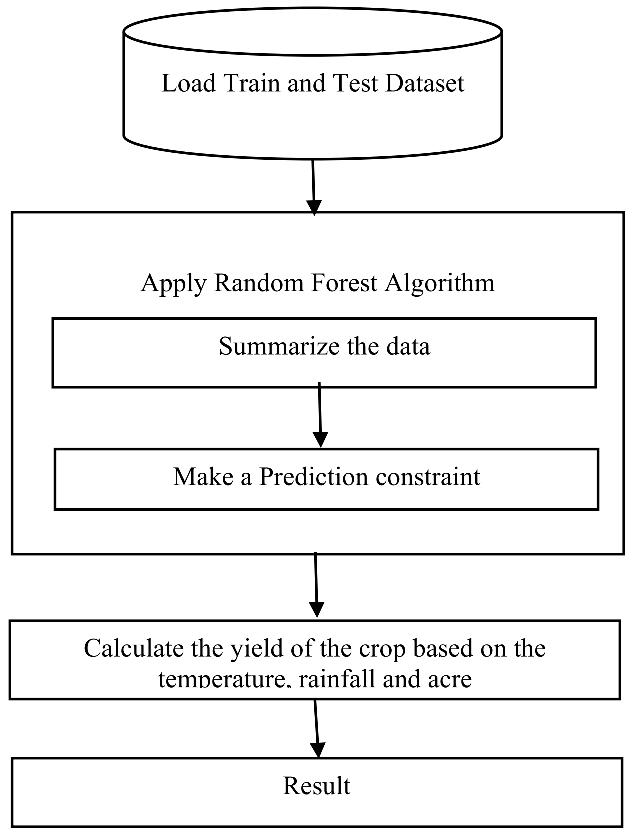

Fig 2. Random Forest Process.

Support Vector Machine

Support Vector Machine (SVM) is a well-known algorithm for classification and regression tasks, and can be used in crop recommendation to predict the most suitable crop based on productivity and season. SVM is particularly useful when the data is high-dimensional and when the different classes of crops can be separated by a clear hyperplane.SVM is depicted in **Fig3**.

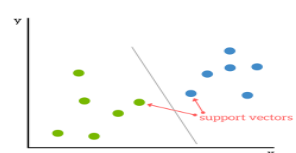

Fig 3. Support Vector Machine.

SVM algorithm is illustrated in equation (2)

$$f(x) = w^t x + b \quad (2)$$

such that g (x) ≥ 0 for y = +1 and g (x) < 0 for y = −1.
Were,
f(x) - Decision function

x - Vector of input features
w -Vector of weights
b - Bias term

K-Nearest Neighbors
K-Nearest Neighbors (KNN), is an algorithm utilized for classification and regression analysis in machine learning. k-NN can be used to identify crops that have performed well under similar conditions (i.e., similar productivity and season). This algorithm is particularly useful when the decision boundary is non-linear, but may be computationally expensive for large datasets. KNN is depicted in **Fig4**.

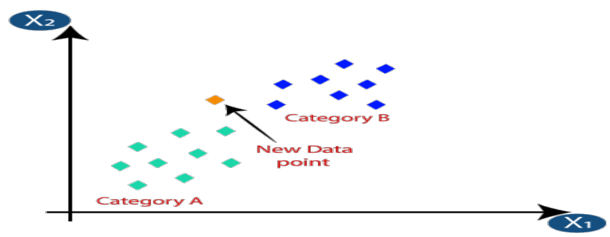

Fig 4. K-Nearest Neighbors.

k-NN algorithm is illustrated in Equation (3) and (4).

$$f(x) = \text{mode}(y_i) \quad \text{for } i \text{ in } N_k(x) \tag{3}$$

Regression:

$$f(x) = \text{mean}(y_i) \quad \text{for } i \text{ in } N_k(x) \tag{4}$$

Were,
$f(x)$ - Prediction function
y_i - value of the i-th data point in the training set.
$N_k(x)$ - set of k closest data points to input x

Naïve-Bayes
Naïve-Bayes is a probabilistic algorithm that can also be used for crop recommendation based on productivity and season. This algorithm calculates the probability of each crop given the input data, and is particularly useful when the data is high-dimensional and when the features are independent of each other. Naïve-Bayes is depicted in **Fig 5**.

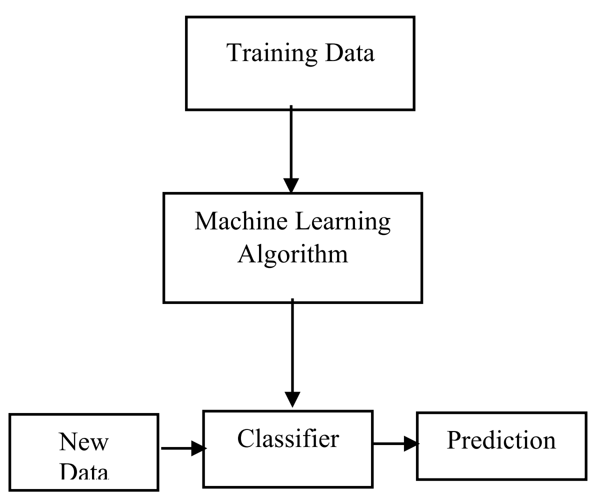

Fig 5. Naïve- Bayes Model

The posterior probability of class- y for input features is shown in Eqn. (5).

$$P\frac{y}{x_1,x_2,\ldots,x_n} = \frac{P(y)*P\frac{x_1}{y} * P\frac{x_2}{y} * \ldots * P\frac{x_n}{y}}{P(x_1,x_2,\ldots,x_n)} \tag{5}$$

Where,
$x_1, x_2, \ldots, x_n$ – input feature
P(y) – prior probability of class y
$P\frac{x_i}{y}$ – conditional probability of feature $\frac{x_i}{y}$

Model Evaluation
Knowing that the production it delivers maintains a suitable level throughout the "whole year," it is clear from Fig. that farmers do not have to fear more about season for the "Coconut" crop. Using this information, farmers may choose when to start cultivating crops. These guidelines can be taken out and used to teach the farmers. A greater understanding of the crops to choose for cultivation is revealed to the farmer through pictorial depiction.
Once the dataset was partitioned, the machine learning models were trained and their performance was assessed based on Accuracy, Precision, Recall, and F1 Score for validation. Four potential outcomes were considered to arrive at these scores: Real Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) for each of the machine learning techniques employed.

Accuracy is determined using Equation (6).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

Equation (7) measures the precision score,

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

Equation (8) determines the recall score,

$$\text{Recall} = TP/(TP + FN) \tag{8}$$

FI core determines the overall accuracy metric that joins precision and recall is shown in equation (9)

$$F1 = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{9}$$

The random forest algorithm also achieved the highest F1-score of 0.90, 0.84 for the SVM.Our results demonstrate the effectiveness of the random forest algorithm in crop recommendation systems.

Table 2. Performance Of Various Techniques Used

Technique	Accuracy
KNN	63
NB	67
SVM	81.8
RF	89.5

Table 2 describes the performance of various techniques used for agricultural crop recommendations. From the experimental evaluation it is observed that RF provides better results than other techniques. Accuracy comparison is illustrated in **Fig 6**.**Table 3** shows Comparison Table.

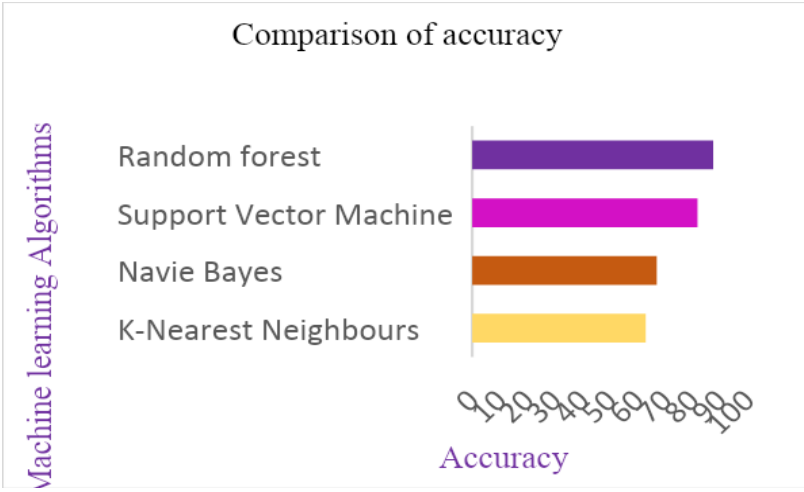

Fig 6. Comparison of ML algorithm accuracy.

Table 3. Comparison Table

Title	Year	Techniques	Accuracy(%)
Proposed approach	2023	SVM, Random Forest, Naïve-Bayes, KNN	89.5 for Random forest
[3]	2022	Linear Regression, Decision Tree, Random Forest	88.9
[5]	2021	Random Forest	85.71
[7]	2020	Linear Support Vector Machine	87.2
[9]	2020	Artificial Bee Colony Algorithm	86.5
[11]	2021	Machine Learning Techniques	83.7

IV. EXPERIMENTAL RESULTS

Data at the district level were examined, and suggestions were made. **Table 4** represents coconut crop facts.

Table 4. "Coconut" Crop Facts

Crop_Year	Season	Crop	Area	Productior
2011	Whole Yea	Coconut	82704	1.25E+09
2013	Whole Yea	Coconut	56484	3.61E+08
2013	Whole Yea	Coconut	56484	3.61E+08
2008	Whole Yea	Coconut	76902	12406000
2011	Whole Yea	Coconut	3435	25900000
2017	Whole Yea	Coconut	323	2600000
2011	Whole Yea	Coconut	33742	8.24E+08
2019	Whole Yea	Coconut	11305	1.58E+08
2009	Whole Yea	Coconut	15548	205800
2004	Whole Yea	Coconut	100160	924200
2002	Whole Yea	Coconut	99250	630000
2003	Whole Yea	Coconut	24240	258500
2006	Whole Yea	Coconut	15538	2736
2004	Whole Yea	Coconut	100160	924200

The "coconut" crop production is taken into account for the experimental outcomes.

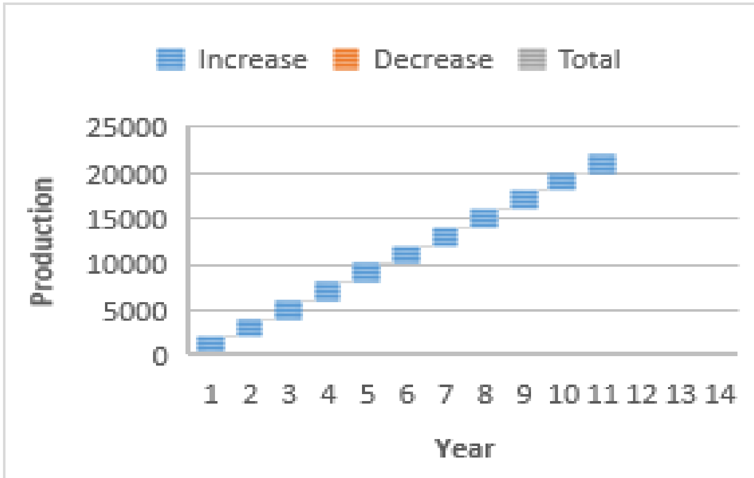

Fig 7. Production of Coconut.

It is obvious from **Fig7**, that growing coconut won't result in any losses for farmers. In 2013 and 2011, there was a lot of output. When compared to 2003, 2004 has somewhat decreased, although it is still higher than the 1990s. It is obvious that coconuts will provide cultivators with a good yield.

The agricultural season, which is depicted in **Fig 8**, should be taken into account while thinking about output.

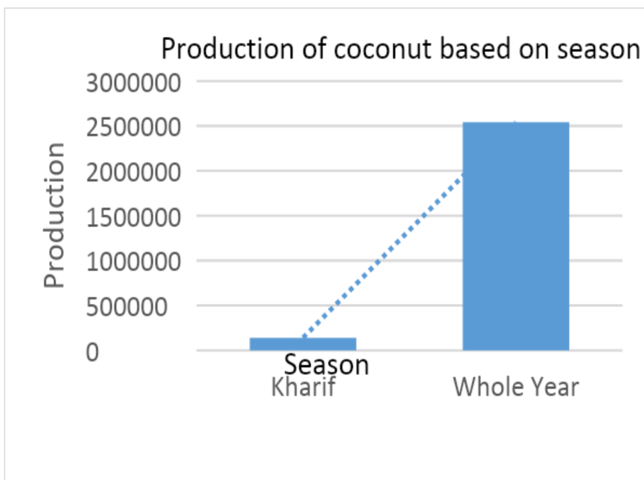

Fig 8. Production of Coconut Based on Season.

V. CONCLUSION

The importance of crop management was thoroughly examined in this paper. To raise the crops, farmers need help from modern technologies. Agriculturists can be informed on a timely basis information about crop forecasts. The agricultural parameters have been analysed using machine learning approaches. This literature review examines some of the methods used in many facets of agriculture. Machine Learning techniques were implemented for better prediction. Farmers may receive more individualized and pertinent recommendations based on factors like productivity and season, causing them to produce an adequate amount of goods.

References

[1]. Neerkathalingam.V; "Agricultural crop recommendation based on productivity and season", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/ Issue:05/May-2022.

[2]. R.Parvathi, "Improvement of Crop Production Using Recommender System by Weather Forecasts," International Conference on recent Trends in Advanced Computing by Weather Forecasts , 2019,ICRTAC-2019.

[3]. Vanitha. D "Productivity and Season Based Crop Yield Prediction Using Machine Learning Algorithm," International Journal of Research Publications and Reviews, Vol 3, no8,pp 204-207,August 2022

[4]. S. Agarwal and S. Tarar, “A Hybrid Approach For Crop Yield Prediction Using Machine Learning And Deep Learning Algorithms,” Journal of Physics: Conference Series, vol. 1714, no. 1, p. 012012, Jan. 2021, doi: 10.1088/1742-6596/1714/1/012012.

[5]. K. Moraye et al., “Crop Yield Prediction Using Random Forest Algorithm for Major Cities in Maharashtra State,” International Journal of Innovative Research in Computer Science & Technology, vol. 9, no. 2, pp. 40–44, Mar. 2021, doi: 10.21276/ijircst.2021.9.2.7.

[6]. S. Sampathkumar and R. Rajeswari, “An Automated Crop and Plant Disease Identification Scheme Using Cognitive Fuzzy C-Means Algorithm,” IETE Journal of Research, vol. 68, no. 5, pp. 3786–3797, Jun. 2020, doi: 10.1080/03772063.2020.1780163.

[7]. S. Ramasamy, A. Selvarajan, V. Kaliyaperumal, and P. Aruchamy, “A hybrid location-dependent ultra convolutional neural network-based vehicle number plate recognition approach for intelligent transportation systems,” Concurrency and Computation: Practice and Experience, vol. 35, no. 8, Jan. 2023, doi: 10.1002/cpe.7615.

[8]. Manjunathan.,"Crop Yield Prediction Using Linear Support Vector Machine," European Journal of Molecular and Clinical Medicine Volume 7, Issue 06,2020.

[9]. R. Sathya, S. Ananthi, R. Rupika, N. Santhiya, and K. Lavanya, “Average Intensity Sign (AIS) Feature based Offline Signature Verification for Forgery Detection using Machine Learning,” 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Nov. 2022, doi: 10.1109/icaiss55157.2022.10010812.

[10]. W. Li et al, “Crop Recommendation System Based on Convolutional Neural Networks and Transfer Learning”,International Conference on Industrial Technology and Management (2021).

[11]. S. Ahmed et al.,” Agricultural Crop Recommendation Using Artificial Bee Colony Algorithm”, International Conference on Machine Learning and Data Engineering (2020).

[12]. S. Singh and R. Kumar "Crop Recommendation System for Rice Cultivation in India Using Decision Trees,". Conference: International Conference on Sustainable Computing and Communication Systems (2020).

[13]. D. Zhou et al. "Crop Recommendation Based on Multispectral Imagery and Machine Learning Techniques," Conference: International Conference on Information Technology and Multimedia (2021).

[14]. M. D. Dibie et al: "An Intelligent Crop Recommendation System for Smallholder Farmers in Africa," Conference: IEEE International Conference on Industrial Engineering and Engineering Management (2021).

[15]. J. Kim et al. "Crop Recommendation System Based on Clustering and Decision Trees,” Conference: International Conference on Advanced Communication Technologies and Networking (2021).

[16]. M. A. Ali and S. U. Amin. "Agricultural Crop Recommendation Using Ensemble Learning Techniques," Conference: International Conference on Information Science and Systems (2020).

[17]. Y. Li, C. Chen, F. Liu, and J. Wang, “Analysis Method of Agricultural Total Factor Productivity Based on Stochastic Block Model (SBM) and Machine Learning,” Journal of Food Quality, vol. 2022, pp. 1–11, Mar. 2022, doi: 10.1155/2022/9297205.