Proceedings of the Third International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT-2025)
IEEE Xplore Part Number: CFP25CV1-ART; ISBN: 979-8-3315-2754-9

2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) | 979-8-3315-2754-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/IDCIOT64235.2025.10915073

Optimizing Crop Yield - An IoT and ML-based Soil Testing and Crop Recommendation System

Prathosh S,
II M. Sc [CS],
Dept. of Computer Science,
Karpagam Academy of Higher Education,
Coimbatore, India
Mail id: sprathosh3@gmail.com

Veerasamy K,
Assistant professor,
Dept. of Computer Science,
Karpagam Academy of Higher Education,
Coimbatore, India
Mail id: ksveerasamy27@gmail.com

***Abstract*−Agriculture is the primary source of income for a large segment of the Indian population, making this sector essential to achieving high crop yields for food security and economic stability. There is a growing concern that the misuse of fertilizers is disrupting nutrient balance, which negatively affects soil health and reduces productivity. Moreover, inderterminate climate patterns have affected crop yield leading to extensive farming losses. To address these challenges, a system leveraging the Internet of Things (IoT) and Machine Learning (ML) was developed for improved soil testing, real-time monitoring, and crop guidance. This innovative system employed a network of sensors that evaluated critical soil characteristics such as temperature, moisture, pH, and NPK levels. The data collected by these sensors was processed through a microcontroller and analyzed using the Random Forest algorithm, resulting in precise nutrient recommendations. Additionally, the robustness of the model was further enhanced through Stacking and Adaptive Feature Fusion techniques. Additionally, the system included a built-in plant health monitoring module that utilized Integrated Convolutional Neural Networks (CNNs) to analyze images captured by integrated cameras, allowing for the early detection of plant diseases and timely interventions to prevent outbreaks. This solution refines agricultural practices through precise nutrient management and early disease identification, ultimately improving crop yields and promoting soil sustainability. In addition, this initiative has the potential to significantly transform the livelihoods of millions of farmers across India. The proposed approach outperformed existing algorithms achieving a 99.5% accuracy.**

Keywords− Identification of Soil Nutrient, NPK: Nitrogen, Phosphorus, Potassium, IoT, Machine Learning, Adaptive Feature Fusion, Stacking, Random Forest, Crop Recommendation.

I. INTRODUCTION

Agriculture is one of the main economic activities in India [1]. This sector is facing many challenges that lead to farmer suicides due to weak crop yields and enormous debts for farming investments. Unpredictable climate change has complicated the problems leading to low quality agricultural output and farmer bankruptcy. Furthermore, the abuse of fertilizers for increased crop yield has caused concerns about the negative effect on soil health and productivity. Over the past century, Indian agriculture has steadily improved with precision agriculture emerging as a potential solution to these problems. Precision farming uses site-specific parameters to observe problems and identify agricultural problems. "Site-specific" farming is at the heart of this innovation that involves tailoring techniques to enhance resource use and crop productivity. Yet, this solution also faces significant challenges that hinder efficacy.

However, advances in IoT and Machine Learning technologies can solve this problem. For instance, mathematical and statistical models used with projections of agricultural data, tend to minimize the effects of climate variability. Moreover, a data-driven approach helps farmers obtain real-time recommendations on ideal crop selection appropriate to the prevailing environmental conditions. This set of tools encompassed soil testing, IoT devices, and the ML algorithms. As a result, farmers experienced increased crop yields, maximised profits and reduced debt-bearing loads. Further, this system provided sustainable access to improved agricultural productivity and debt relief [2].

Crop recommendations can be used to tackle challenges faced in precision farming for increased yields and profits. This considers many factors, which include the environment, nutrient concentrations, and even the soil condition. Soil testing with IoT and ML technologies generated more accurate results because this enabled the collection of precise data and analysis, thus providing personalized recommendations to farmers on what crops should be planted. Thus the aim is to increase the productivity level, and thereby, improve decision-making at the farm level by

overcoming most of these concerns. Despite the advances in precision agriculture, not all outputs are correct. Consequently, a mistake in this technology can lead to massive material and financial losses for farmers. Therefore, with such high stakes, the recommendations should be accurate and specific so that farming might be sustainable and profitable. The high accuracy and efficient crop prediction models developed were thus the final product of this research. Through the utilization of IoT-based soil testing and ML algorithms, this system provided crop recommendations that reduced errors and improved overall productivity [3].

Hence, this proposal encompassed supervised, unsupervised, and reinforcement learning models which have different strengths and limitations. Supervised learning was utilized in developing a mathematical model that the generated predictions, provided the datasets had a set of inputs and the corresponding outputs. On the other hand, unsupervised learning utilized algorithms that identified relationships or patterns in datasets that lacked labels. Further, Reinforcement Learning (RL) is applied in cases whereby there exists little information about a concept, entity, or object. The RL agent acquires knowledge by interacting with the environment [4]. The IoT-based soil testing process with machine learning algorithms was focused on optimizing the crop yield prediction, reducing errors, and facilitating informed decision-making by farmers that boosted productivity and profitability.

This research inferred which crops should be cultivated through the major soil and environmental parameters such as temperature, moisture, pH value, rainfall, and NPK. The analysis compared the prediction capabilities of future yields of 22 crops, ranging from apple and papaya to coconut and cotton, jute, and coffee, among many fruits and cash crops by applying various supervised machine learning approaches suitable for Indian agriculture. Moreover, the dataset used in this system included crucial parameters and the model offered applied ML algorithms—Adaptive Feature Fusion (AFF), Stacking and Random Forest (RF) techniques—to suggest crops viable for farmers. This system enhanced crop selection and productivity through methods that employed IoT-enabled soil testing.

The rest of the paper is organized as follows: Section 1 reviews the work done on optimizing crop yield. Section 2 presents a comparative analysis of crop recommendation systems using ML approaches. Section 3 outlines the proposed system. Section 4 presents experimental results and discussion. Section 5 not only concludes the paper, but also sets out prospects for further study in this area.

II. LITERATURE REVIEW

Kumar et al (2020) discovered that a crop yield prediction system utilizing historical data can effectively account for variables such as temperature, humidity, pH levels, precipitation, and crop type, while also considering the distribution of various crops across different districts in India. This system aimed to identify the most suitable crops for local conditions. The implementation of machine learning algorithms, particularly Decision Tree (DT) and RF, significantly improved crop yield predictions, with RF achieving the highest level of accuracy for this application [1].

Suresh et al (2021) identified the most appropriate crop based on specific input data. This research utilized the Support Vector Machine (SVM) algorithm to improve both accuracy and efficiency. The study considers two datasets: one containing sample location data and the other focused on crop-specific information. By analyzing nutrient values, the study determined the optimal crop based on nitrogen (N), phosphorus (P), potassium (K), and pH levels. Additionally, the researchers assessed the current nutrient levels present in the soil and recommended the necessary fertilizer amounts for cultivating crops such as black gram, carrot, rice, radish, and maize [5].

Reddy et al (2019) highlighted three primary factors that were essential for recommending suitable crops for cultivation: soil properties, soil classifications, and crop yield data. To predict the best crop options based on particular weather conditions and data from state and district levels, various machine learning methods such as CHAID, Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbor (K-NN) were utilized. This approach aimed to assist farmers in selecting the appropriate crops according to soil requirements, ultimately enhancing agricultural productivity on a national scale [3].

Rajak et al (2017) performed a survey to identify suitable crops based on a comprehensive soil database. The study assessed nine different crops—Banana, Sorghum, Pulses, Coriander, Sugarcane, Cotton, Vegetables, Paddy, and Groundnut—considering various soil characteristics such as Soil Depth, Soil Texture, pH, Color, Water Holding Capacity, Erosion,

Drainage, and Permeability. Further, the research utilized several Machine Learning classifiers, including NB, SVM, RF, and Artificial Neural Network (ANN). This system, noted for its high accuracy and efficiency, provided crop recommendations tailored to specific site parameters. The research aimed to create an online platform for farmers, enhancing agricultural productivity, minimizing soil degradation, reducing chemical usage in farming, and optimizing water resource management [6].

Doshi et al (2018) introduced an innovative system called AgroConsultant, which consisted of two main components: i) Crop Viability Estimator and ii) Precipitation Forecasting Tool. This system was tailored to support fifteen minor crops and five major crops, taking into account a range of factors including location parameters, soil pH, temperature, soil type, soil thickness, precipitation, and aquifer thickness. AgroConsultant utilized several Machine Learning techniques, including K-NN, Neural Networks, RF, and DTs to perform multi-label classification. The rainfall prediction model demonstrated an accuracy of 71%, while the neural network-based crop suitability predictor achieved an impressive accuracy of 91%. The primary objective of this system is to improve agricultural decision-making by delivering accurate predictions for crops and rainfall [7].

Gosai et al (2021) introduced novel methods for determining the most suitable crops by analyzing particular environmental and soil factors, such as soil composition, precipitation, and temperature. The reseearchers' system utilized various algorithms, including DT, SVM, RF, NB, and Logistic Regression (LR), to categorize crops into groups like Kharif and Rabi. This methodology guaranteed a high level of accuracy, leading to precise crop recommendations that enhanced agricultural productivity [8].

III. PROPOSED METHODOLOGY

Phases

1) Data Collection
2) Data Pre-processing and Noise Elimination
3) Feature Extraction
4) Application of Various ML Algorithms
5) Creation of a Recommendation System
6) Crop Recommendation

Proposed system was divided into several stages, as illustrated in Fig. 1.

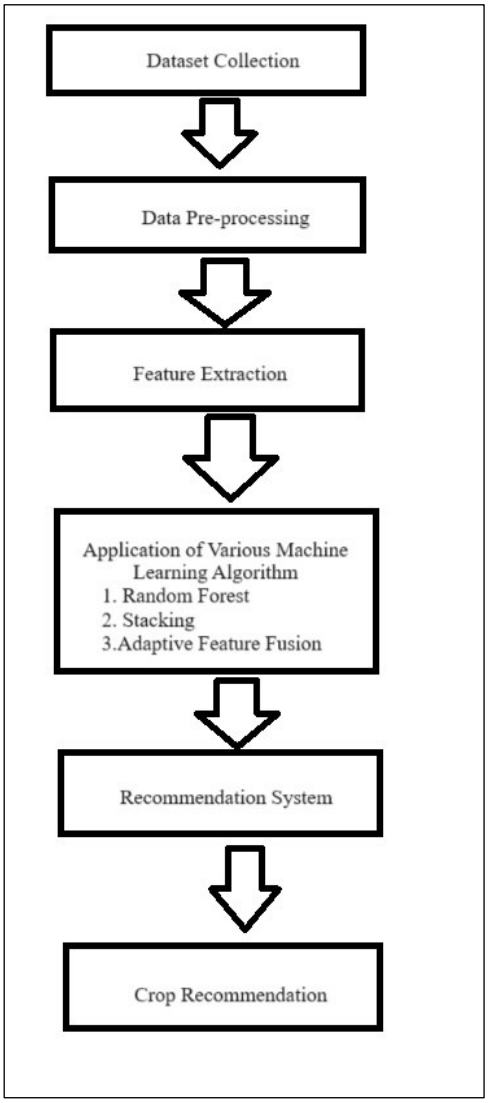

Fig. 1. System Design for the Crop Recommendation Platform

Flow of the Proposed System

The sentiment extraction process, illustrated in Fig. 1, was executed as follows:

(1). Dataset Collection:

This included rainfall, temperature, humidity, N, P, and K as well as the soil's pH value. The datasets were obtained from the GitHub website. There were 2,200 websites in the collection. Historical data was used to build these instances. Black Gramme, Lentil, Mung Beans, and Moth Beans were among the eleven crops present. Fig. 2 is part of the dataset used in this research.

1	N	P	K	temperatu	humidity	ph	rainfall	lal
2	90	42	43	20.87974	82.00274	6.502985	202.9355	ric
3	85	58	41	21.77046	80.31964	7.038096	226.6555	ric
4	60	55	44	23.00446	82.32076	7.840207	263.9642	ric
5	74	35	40	26.4911	80.15836	6.980401	242.864	ric
6	78	42	42	20.13017	81.60487	7.628473	262.7173	ric
7	69	37	42	23.05805	83.37012	7.073454	251.055	ric
8	69	55	38	22.70884	82.63941	5.700806	271.3249	ric
9	94	53	40	20.27774	82.89409	5.718627	241.9742	ric
10	89	54	38	24.51588	83.53522	6.685346	230.4462	ric

Fig. 2. Dataset

(2). Data Pre-processing and Noise Elimination:

This task enabled the data to be used effectively in creating high-performing models. The majority of data gathered from various sources is usually unprocessed and frequently contains redundant, inconsistent, or missing information. As a result, the dataset had to be normalised and redundant data removed [4]. Fig. 3 shows the result checking for missing values in the dataset.

```
Columns in dataset: Index(['N', 'P', 'K', 'temperature
'humidity', 'ph', 'rainfall', 'label'], dtype='object'
Check for missing values in dataset:
 N              0
P               0
K               0
temperature     0
humidity        0
ph              0
rainfall        0
label           0
```

Fig. 3. Data Pre-processing

(3). Feature Extraction:

In order to maximise the usage of classifiers, this involved identifying only the most pertinent qualities in a particular dataset and then applying them, along with eliminating redundant and irrelevant information [9].

(4). Approach:

Some machine learning techniques, such as RF, Stacking, and AFF were used in the proposed system.

1. Adaptive Feature Fusion (AFF):

During training, one measures the feature importance, most of the time using an algorithm like RF. Only the best features are retained to train the model in order to generalize better and remove the irrelevant features. Factors like temperature, perception, rainfall, and the production were used for ML modelling. The major parameters of the AFF technique include: (i) Threshold for Feature importance which was the cut-off score in terms of the value with which the feature could be differentiated as important to be selected, and further features not up to this threshold were excluded from consideration. (ii) Feature selection process whereby the most most relevant features were selected after the importance scores were calculated. This just made use of only those values above the threshold while creating the final feature set for the model for training. (iii) Fused Feature Set was the final outcome of the fusion process that had the set of features chosen. Filtered features were used in training the model and the model focused only on the most important variables.

Adaptive Feature Fusion was applied in the following steps:

(i) The accuracy_score method from sklearn.linear module was imported
(ii) Then, the accuracy_score object was generated
(iii) Lastly, the model was fitted on the data

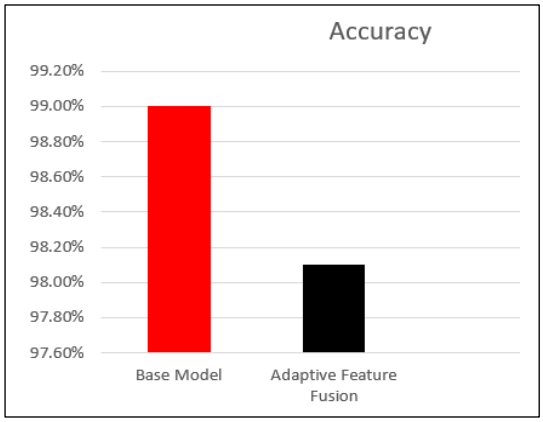

Fig. 4. Adaptive Feature Fusion Accuracy Bar Chart

In Fig. 4, the accuracy of the base and AFF models is compared. The base model had a 99% accuracy result while the AFF model had a 97.5% accuracy score.

2. Stacking:

Base models are stacked together to improve prediction accuracy. In this process, multiple base models, either classifiers or regressors were trained and the predictions from these models were combined using a meta-learner to produce the final output [10]. Input data processed using the meta-learner was used

to generate the final output using the forecasts of the base models. The meta-learner was utilized since the model could capture patterns that individual models missed, resulting in higher accuracy scores [11]. Variables considered for the dataset were rainfall, perception, temperature, and production. Generally, a portion of the total dataset is kept as the testing set while the rest of the dataset is divided into two: a training set and a validation set. The Stacking algorithm took into account the following parameters: (i) Base models: These are individual models such as RF, LR, etc, whose predictions were combined to produce the stacked model. (ii) Meta-learner: The model that aggregated the predictions from the base models to generate the final outcome [12]. (iii) Cross-validation: In this technique, the dataset was split several times in order to ensure that each of the base models and the meta-learner had distinct training and validation samples.

Stacking was implemented as follows:

(i) The StackingClassifier class was imported from sklearn
(ii) Then a StackingClassifier object was created
(iii) Finally, the model was fitted to the data

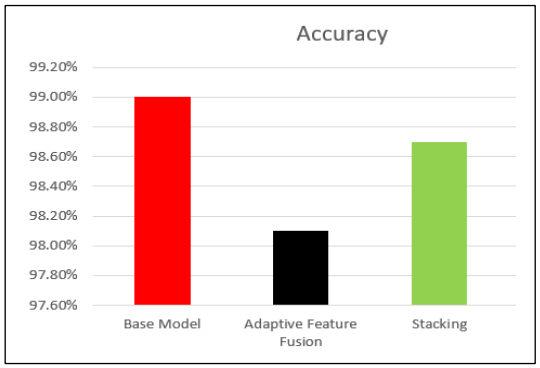

Fig. 5. Stacking Accuracy Bar Chart

Fig. 5 shows the accuracy results of three models.

3. Random Fores (RF)t:

This is an ML algorithm that builds many DTs that are used to make predictions during training [13]. The generation of predictions was done by taking a majority vote of classification problems and the average forecasts of regression problems [14]. The addition of more trees generally improved the accuracy of the model [15]. Variables used for training include: rainfall, precipitation, temperature, and RF algorithm. This research assigned approximately 66% of data for training and kept 34% for testing purposes. The RF algorithm has three key parameters: n_tree that decided how many trees were generated; m_try that showed how many variables should be tried in the process of splitting a node; and Node_size, which defined the minimum number of observations required at the terminal nodes [16].

Random Forest (RF) implementation:

(i) The RandomForestClassifier class was imported from sklearn.ensemble
(ii) Then, the RandomForestClassifier object was created
(iii) Finally, the model was fitted to the data

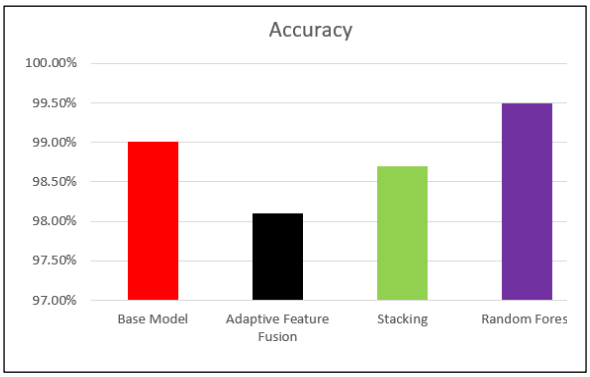

Fig. 6. Random Forest Accuracy Bar Chart

Fig. 6 presents the accuracy scores of four models.

Key observation:

The RF model achieved the highest accuracy, reaching approximately 99.50%. The Base Model closely followed with an accuracy of nearly 99.00%. The Stacking model demonstrated strong performance as well, with an accuracy around 98.80%. In contrast, AFF recorded the lowest accuracy among the four models, scoring 98.20%. This comparative analysis indicated that the Random Forest model was the most effective, surpassing the Base, Stacking, and AFF models. This research which combined multiple models, also showed that AFF underperformed compared to both RF and the Base Model in this evaluation despite AFF's ability to dynamically select features.

IV. EXPERIMENTAL RESULT

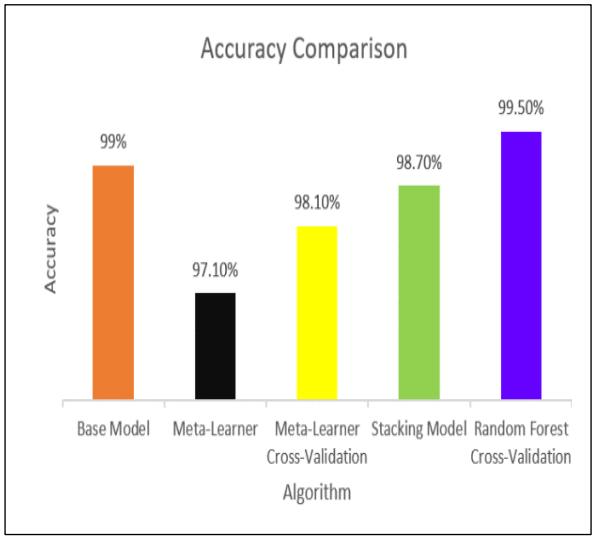

Fig. 7. Accuracy Comparison

Fig. 7 compares ML models according to individual accuracy scores. The x-axis lists the many techniques, such as Base Model, Meta-Learner, ML Cross Validation, Stacking Model, and RF Cross Validation, while the y-axis shows the % correctness, which ranges from 0 to 100. Fig. 8 indicates the machine learning model accuracy result while Fig. 9 shows the prediction output.

```
Base Model Accuracy: 0.990909090909091
Meta-Learner Accuracy after Tuning: 0.9712121212121212
Random Forest Cross-Validation Accuracy: 0.9949999999999995
Meta-Learner Cross-Validation Accuracy: 0.9818181818181818
Stacking Model Accuracy: 0.9878787878787879
```

Fig. 8. Machine Learning Vice Accuracy Result

```
# Make prediction using the trained RandomForestClassifier
prediction_numeric = base_model.predict(new_data)
predicted_crop = crop_names[prediction_numeric[0]]
print("Predicted Crop:", predicted_crop)

Predicted Crop: coffee
```

Fig. 9. Prediction Snapshot

V. RESULT ANALYSIS

TABLE I: ALGORITHM-WISE ACCURACY RESULTS (%)

Algorithm	**Accuracy**
Base Model	99%
Meta-learner	97.1%
Stacking Model	98.7%
Meta-learner Cross-Validation	98.1%
Random Forest Cross- Validation	**99.5%**

Based on soil test results and other pertinent information gathered by IoT sensors, these high accuracy results imply that the machine learning models are highly successful at forecasting appropriate crops or making suggestions. The use of cross-validation strongly supports the above analysis. Further, ensemble techniques like stacking, significantly improve model performance across various approaches. These results are quite encouraging for a recommendation system in agriculture. Based on soil conditions and other pertinent data gathered by IoT devices, this solution can offer farmers in India extremely accurate guidance for maximizing agricultural production. Table I above shows the results of the models with the proposed model achieving a 99.5% accuracy.

VI. CONCLUSION AND FUTURE WORK

The proposed research aimed to create an advanced crop recommendation system specifically tailored for farmers in India. This system assisted farmers in determining the most appropriate crops for planting by analyzing essential soil and environmental factors, including NPK levels, humidity, rainfall, pH value, and temperature. By leveraging these parameters, the system sought to improve agricultural productivity and profitability. The proposed methodology integrated AFF, Stacking, and RF ML techniques to provide precise crop recommendations. Among these algorithms, RF demonstrated the highest accuracy. Consequently, this model was chosen as the preferred model for this study. Ultimately, this system has the potential to significantly boost crop yields, enhance

decision-making for farmers, and contribute to the economic growth of India. However, this research had some limitations. Firstly, the focus group was Indian farmers only. Future research can focus on other farmers from other parts of the world. Secondly, the study used one dataset for research. A comprehensive dataset is needed to validate the model's robustness across different datasets.

REFERENCES

[1] Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., Neha, K., & Devi, V. G. R. R. (2020, June). Supervised machine learning approach for crop yield prediction in agriculture sector. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 736-741). IEEE.

[2] Garanayak, M., Sahu, G., Mohanty, S. N., & Jagadev, A. K. (2021). Agricultural recommendation system for crops using different machine learning regression methods. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, *12*(1), 1-20.

[3] Reddy, D. A., Dadore, B., & Watekar, A. (2019). Crop recommendation system to maximize crop yield in ramtek region using machine learning. *International Journal of Scientific Research in Science and Technology*, *6*(1), 485-489.

[4] Gandge, Y. (2017, December). A study on various data mining techniques for crop yield prediction. In *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)* (pp. 420-423). IEEE.

[5] Suresh, A., Kumar, P. G., & Ramalatha, M. (2018, October). Prediction of major crop yields of Tamilnadu using K-means and Modified KNN. In *2018 3rd International conference on communication and electronics systems (ICCES)* (pp. 88-93). IEEE.

[6] Rajak, R. K., Pawar, A., Pendke, M., Shinde, P., Rathod, S., & Devare, A. (2017). Crop recommendation system to maximize crop yield using machine learning technique. *International Research Journal of Engineering and Technology*, *4*(12), 950-953.

[7] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In *2018 fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-6). IEEE.

[8] Gosai, D., Raval, C., Nayak, R., Jayswal, H., & Patel, A. (2021). Crop recommendation system using machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *7*(3), 558-569.

[9] Gautron, R., Maillard, O. A., Preux, P., Corbeels, M., & Sabbadin, R. (2022). Reinforcement learning for crop management support: Review, prospects and challenges. Computers and Electronics in Agriculture, 200, 107182.

[10] Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019, November). Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.

[11] Medar, R., Rajpurohit, V. S., & Shweta, S. (2019, March). Crop yield prediction using machine learning techniques. In *2019 IEEE 5th international conference for convergence in technology (I2CT)* (pp. 1-5). IEEE.

[12] Suresh, G., Kumar, A. S., Lekashri, S., Manikandan, R., & Head, C. O. (2021). Efficient crop yield recommendation system using machine learning for digital farming. *International Journal of Modern Agriculture*, *10*(1), 906-914.

[13] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In *2016 eighth international conference on advanced computing (ICoAC)* (pp. 32-36). IEEE.

[14] Jain, S., & Ramesh, D. (2020, February). Machine Learning convergence for weather-based crop selection. In *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (pp. 1-6). IEEE.

[15] Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). Improving crop productivity through a crop recommendation system using ensembling technique. In *2018 3rd international conference on computational systems and information technology for sustainable solutions (CSITSS)* (pp. 114-119). IEEE.

[16] Kumar, A., Sarkar, S., & Pradhan, C. (2019, April). Recommendation system for crop identification and pest control technique in agriculture. In *2019 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0185-0189). IEEE.