Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

Smart Crop Recommendation Using Ensemble Machine Learning Models

Anam Umera[1*], Zainab Khalid[2], Sania Qamar[2], Muhammad Zaid[3], Mohsin Saeed[3], and Muhammad Yousif[3]

[1]School of System and Technology, Department of Informatics and System, Univeristy of Management and Technology, Lahore, Pakistan.
[2]School of System and Technology, Department of Software Engineering, Univeristy of Management and Technology Lahore, 54770, Pakistan.
[3]Department of Computer Science, Minhaj Univeristy,Lahore, Pakistan.
[*]Corresponding Author: Anam Umera. Email: anam.umera@umt.edu.pk

Received: April 11, 2025 **Accepted:** May 23, 2025

Abstract: Agriculture is the primary occupation for a huge portion of Pakistan's population and plays a vital role in the country's economic growth and food security. Crop growth is significantly influenced by various environmental and soil-related factors such as weather, chemical inputs, soil moisture, phosphorus levels, humidity, temperature, and rainfall. To enhance crop productivity and decision-making, this research proposes a smart crop recommendation system based on ensemble learning using supervised machine learning models. Sensor data is used to monitor key factors, which are then analyzed using an ensemble of different supervised learning techniques. By combining the strengths of multiple models through a voting-based approach, more accurate recommendations are generated. Among the evaluated models, decision trees and artificial neural networks provided the most effective results, with the artificial neural network achieving an accuracy of up to 98%. This approach supports the development of precision agriculture, which emphasizes site-specific crop management using modern agricultural technologies. Precision farming is gradually gaining traction in developing countries like Pakistan, offering improved efficiency and sustainability in agriculture.

Keywords: Ensemble Learning; Crop Recommendation; Artificial Neural Network

1. Introduction

Pakistan's economy and employment depend heavily on agriculture. Globalization has caused significant changes in agriculture in recent years. Pakistan is among the nations that biggest producers of agricultural goods, although farm productivity is still poor. Fo1r many farmers, agriculture is their primary source of income. The most frequent issue Pakistani farmers deal with is choosing the wrong crop for their soil. Productivity suffers as a result. To get more money from the same amount of land with less effort, farmers must enhance productivity. Numerous cutting-edge technologies have been created to help people regain their health [1]. There are more workers in the field, production has increased, and living conditions have improved due to digital technology's reduction of the need for manual labor in agriculture. Pakistan's agriculture has advanced significantly in recent years. Better improvements have been made to precision agriculture, which is crucial when suggesting crops. Crop recommendations are based on a number of factors. Predicting soil characteristics such as pH, humidity, temperature, nitrogen (N), phosphorus (P), and potassium (K) is the first and most important stage of farming [2]. These have a direct bearing on the climate and geography of the region being used. Global warming has caused significant changes in the climate in recent years. Inappropriate crop selection depletes all available resources, including the price of seeds, fertilizer, and other supplies, which has a significant negative influence on farmers' aspirations.

Traditional farming can be transformed with the use of machine learning (ML), a crucial technology. To help farmers grow more high-quality, productive crops with less waste, and this research attempts to present an ML-based crop recommendation system [3]. Humans learn by attentively or randomly observing the occurrences around them, gaining some exposure, followed by predicting the subsequent occurrence, which mostly takes place without human awareness. The processes of identifying pertinent groupings some experience, followed by predicting the subsequent occurrence, which mostly takes place without human awareness. Fitting prediction models to a target dataset or within data are sometimes referred to as "machine learning." Essentially, machine learning seeks to replicate or imitate human aptitude and pattern recognition through computational methods [4]. When the dataset under study is too big or complicated for humans to manage, or when we want to develop an automated platform for repeatable and time-efficient target dataset analysis, machine learning is especially helpful. Agricultural data often exhibits these characteristics [5]. There are two main goals when applying machine learning to agriculture. First, accurate predictions should be made and utilized to guide future research efforts, regardless of how adequate or inadequate the evidence gathered is. Since scientists are interested in understanding mechanism, the other goal is to apply machine learning to improve and increase understanding of crop development processes, including numerous phenotypic, genotypic, biological, agronomic, and climatic components [6]. Researcher also provide a summary of some of the drawbacks and uses of machine-learning techniques, as well as certain data-related issues for crop improvement researchers [7].

2. Related Work

The majority of prediction models employ conventional modeling techniques, the application and verification of which have been the subject of several studies in the aforementioned manners. Regretfully, these models are not entirely representative of the most anticipated crop production outcome due to their low reliability. Traditional mathematical models of plant yield have both linear and nonlinear results. Linear regression (LR), logistic regression (LOR), support vector machines (SVM), decision trees (DT), K-nearest neighborhood (KNN), and current presentation of crop yield-based crop recommendations by machine learning algorithms as an inferred outcome over input parameters like soil, nutrition, and environmental condition—which may be a complex and nonlinear function—makes them remarkable.

The authors have introduced a crop prediction method based on artificial neural networks and support vector machines (ANN-SVM). Initially, the state of Maharashtra provided the crop prediction data. This included soil pH, rainfall, soil type, and humidity. The input data was then preprocessed using min-max scaling and normalization. Then, two classifiers, including ANN and SVM, were given the preprocessed data. Lastly, two classifiers (ANN-SVM) were used to carry out the crop prediction. The output takes longer to produce because of the challenging computation [8].

Crop loss brought on by climate change and other environmental stresses on farming techniques was discussed by the author of [9]. For the soil fertility prediction approach, the authors suggest a probabilistic neural network, which provides better accuracy and faster processing. Several artificial intelligence methods for increasing agricultural harvest yields are included in the publication [10]. These methods are based on crop recommender systems, which are an essential part of the precision agriculture paradigm. Among the methods that are covered in detail are neural networks, ensemble learning, K-nearest neighbor, and similarity-based classifiers. To choose the best crops, the authors offer a model that takes into account climatic variables including temperature, precipitation, and soil profile. When resources are used more effectively, higher yields are produced.

Climate change is the term used to describe long-term shifts in local or global temperatures or weather patterns. Addressing global warming and lowering greenhouse gas emissions are made more challenging by the legal and regulatory issues surrounding climate change [11]. Machine learning has found numerous applications in the agricultural sector, such as predicting crop diseases, forecasting weather patterns, optimizing irrigation systems, estimating crop yields, and setting minimum support prices. Specifically, supervised learning techniques have been adopted by researchers to enhance the accuracy of crop production forecasts [12].

Using Apriori Algorithm, propose to create a model recommender system for the agriculture sector that uses the Apriori algorithm. According to scientists, a system like this can assist farmers in making educated decisions regarding fertilization and crop selection, boosting agricultural productivity, and decreasing waste [13].

This method enables thorough mapping of soil characteristics across many fields, including pH, nutrient levels, and moisture content. Then, by adjusting their soil management techniques to each section's particular requirements, farmers may maximize inputs like water and fertilizer [14].

Table 1. Precision crop's use in a variety of fields (brief summary)

S/N	Precision agrarian application	Benefits in numbers	Local fence addressed	References
1	Using cutting-edge technology, precision crop improves resource efficiency and soil health to promote sustainable crop production.	20%–30% improvements in soil health and 15% increases in resource utilization efficiency	Low soil fertility and wasteful use of resources	[15]
2	By improving irrigation methods for sustainable agricultural growth through the use of sensor technologies and data analysis development, precision crop conserves water.	30%–50% decrease in water use and 10%–20% improvement in crop yield	Lack of water and ineffective irrigation systems	[16]
3	Drones, satellite imaging, and Internet of Things sensors are used in precision farming to monitor crops in real time, improve decision-making, and increase yields.	10%–25% increase in yield and 15% decrease in input costs	Lack of up-to-date agricultural data and costly inputs	[17]
4	Precision farming uses data-driven methods to maximize nutrient use for sustainable crop production.	20% increase in nutrient use efficiency and a 25% decrease in fertilizer expenses	excessive fertilizer use and expensive nutrient prices	[18]
5	To efficiently control pest and disease outbreaks and protect crops, precision farming uses data-driven tactics and remote sensing.	20%–40% decrease in pests and diseases; 15%–25% reduction in crop loss	Challenges in managing diseases and pests	[19]

3. Proposed Methodology

This process is divided into several steps in the suggested system, as shown in Figure 1 below: data preprocessing, data visualization, data splitting, ensemble models, and recommendation models, which are regarded as performance models.

3.1. Input Layer

In the input layer dataset which taken from kaggle[20] consisting of phosphorous, photashium, nitrogen, humadity, rainfall and ph level. This dataset consist eleven different crops. Rice, maize, chickpeas, kidney beans, pigeon peas, moth beans, mung beans, black grams, lentils, pomegranates, bananas, mangoes, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee are among the 22 crops covered in this dataset[21].

3.2. Data Preprocessing

Given dataset is un-normalize which taken form Kaggle in csv file exploratory data analysis is implemented for normalize in which df.head() is used for displaying header data record and df.footer() is used for footer files. df.decribe() this show statistical vales.

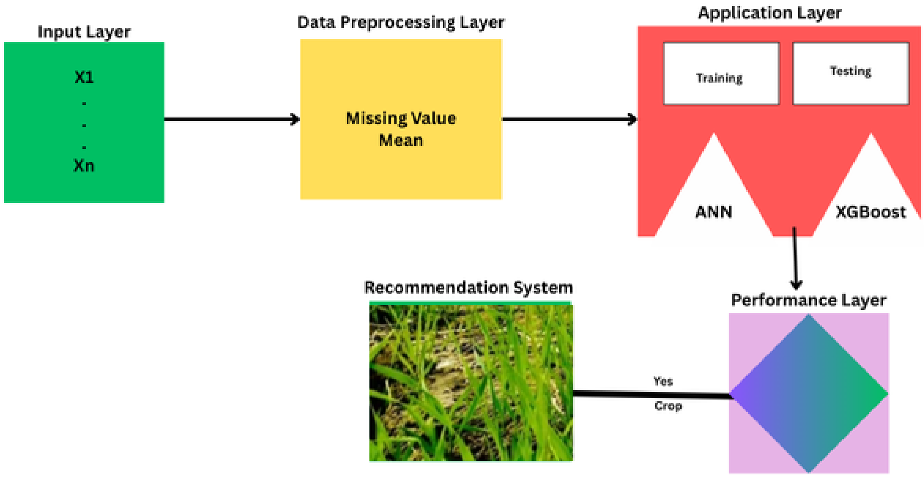

Figure 1. Proposed System

3.3. Data Visualization

The production quantity of various cops has been displayed against the elements influencing their creation. This will give us information on how different weather conditions and soil constituents impact production amount. "Figure.2" displays the bar graphs that represent the data that we plotted.

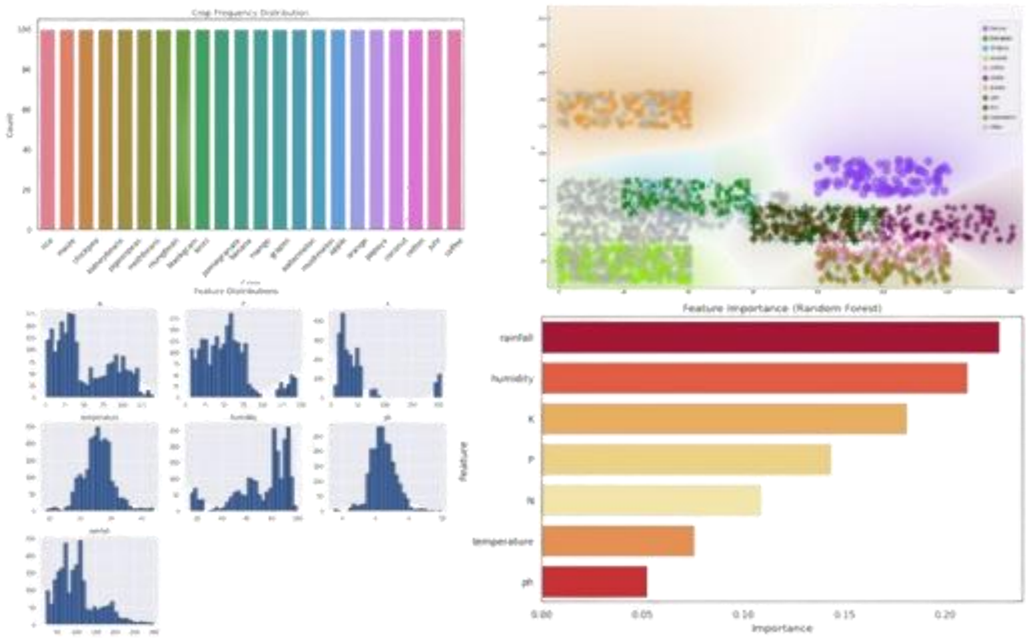

Figure 2. Data Visualization

3.4. Data Splitting

Table 2 respresent the implementation of crop dataset. Splitting of data which shows training , testing and validation with ratio of 70:15:15. For all the data we achieved accuracy 98% when propoer setup is used. The configuration is achived with optimal term performance and accuracy of Table 2.

In this model data is spitted into 70:15:15 ratio.

3.5. Ensemble Models

3.5.1. Logistic Regression

This one is a statistical approach with binary dependent variable and parameters of logistic models.

3.5.2. Support vector Machine

To overcome regression challeneges this model utilized as supervised training with N-corrdinate.

3.5.3. Decision Tree

This one is a greedy approach with hierarchical method using supervised models.

3.5.4. KNN

K-Nearest-Neighbors this one is both supervised and regression supervised machine model.This one is predicted new class from previous one.

3.5.5. XGBoost

This one is a gradient boosting framework which is most suitable for supervised predictions utilized with training data to predict target variable with optimal.

3.5.6. ANN

Artificial neural network resembling of human brain decision strategy. This one work as input layer, hidden layer and output layer.

Three layers in total—an input layer, a hidden layer, and an output layer—were used in the suggested smart crop recommendation. The back propagation method consists of several processes, such as initializing weight, feedforward, back propagation of error, and updating weight and bias. The activation function of each neuron in the buried layer is f(x)=Sigmoid(x). The suggested smart crop recommendation's input and hidden layer sigmoid functions can be expressed as

$$¥_x = \mathfrak{h}_1 + \sum_{\mathfrak{i}=1}^{\mathfrak{M}} (\omega_{ix} * \mathfrak{r}_{\mathfrak{i}}) \tag{1}$$

$$\oint_j = \frac{1}{1 + e^{-¥_x}} \ where \ \varkappa = 1,2,3 \dots n \tag{2}$$

From the output layer, the input is

$$¥_k = \mathfrak{h}_2 + \sum_{\varkappa=1}^{n} \left(\mathfrak{v}_{x k} * \oint_x\right) \tag{3}$$

Below is the output layer activation function

$$\oint_k = \frac{1}{1 + e^{-\gamma_{\mathfrak{K}}}} \quad where \ \mathfrak{K} = 1,2,3 \dots \mathfrak{r} \tag{4}$$

$$E = \frac{1}{2} \sum_{k} \left(t_{\mathfrak{K}} - \oint_{\mathfrak{K}}\right)^2 \tag{5}$$

The back propagation error is represented by the equation above, where $t_{\mathfrak{K}}$&$out_{\mathfrak{K}}$ stand for the intended and estimated outputs, respectively.

The layer is expressed as the rate of change in weight for the output in equation (7).

$$\Delta\omega \propto -\frac{\partial E}{\partial \omega}$$

$$\Delta v_{x,k} = -\varepsilon \frac{\partial E}{\partial \mathfrak{v}_{x,\mathfrak{K}}} \tag{7}$$

The above equation can be expressed as follows after using the Chain Rule method:

$$\Delta \mathfrak{v}_{x,k} = -\varepsilon \frac{\partial E}{\partial \oint_k} \times \frac{\partial \oint_{\mathfrak{K}}}{\partial \gamma_{\mathfrak{K}}} \times \frac{\partial \gamma_{\mathfrak{K}}}{\partial v_{\varkappa,\mathfrak{K}}} \tag{8}$$

Equation (7) displays the value of the weight change that results from changing the variables in equation (8).

$$\Delta v_{x,\mathfrak{K}} = \varepsilon\left(\tau_{\mathfrak{K}} - \oint_{\mathfrak{K}}\right) \times \oint_{\mathfrak{K}}\left(1 - \oint_{\mathfrak{K}}\right) \times \left(\oint_x\right)$$

$$\Delta \boldsymbol{v}_{x,\mathfrak{R}} = \boldsymbol{\varepsilon} \boldsymbol{\xi}_{\mathfrak{R}} \oint_{x} \tag{9}$$

Where,

$$\boldsymbol{\xi}_{\mathfrak{R}} = \left(\boldsymbol{\tau}_{\mathfrak{R}} - \oint_{\mathfrak{R}}\right) \times \oint_{\mathfrak{k}}\left(\mathbf{1} - \oint_{\mathfrak{R}}\right)$$

Use the chain rule to update the input and hidden layers' weights.

The above equation can be expressed as follows after simplification:

$$\Delta \boldsymbol{\omega}_{i,j} = \boldsymbol{\varepsilon} \boldsymbol{\xi}_{x} \boldsymbol{\alpha}_{\mathbb{i}}$$

Where,

$$\boldsymbol{\xi}_{x} = \left[\sum_{\mathfrak{k}} \boldsymbol{\xi}_{\mathfrak{R}}\left(\mathfrak{v}_{x,\mathfrak{k}}\right)\right] \times \oint_{x}\left(\mathbf{1} - \oint_{x}\right)$$

$$\mathfrak{v}_{x,\mathfrak{k}}^{+} = \mathfrak{v}_{x,\mathfrak{R}} + \lambda_{F} \Delta \mathfrak{v}_{x,\mathfrak{R}} \tag{10}$$

The weights between the output and hidden layers are updated using the equation above.

$$\boldsymbol{\omega}_{\mathbb{i},x}^{+} = \boldsymbol{\omega}_{\mathbb{i},x} + \lambda_{F} \Delta \boldsymbol{\omega}_{\mathbb{i},x} \tag{11}$$

Additionally, the weights between the input and hidden layers are updated using the equation above.

4. Simulation & Results

Based on important performance parameters including accuracy, precision, recall, F1 score, and ROC score, the table.2 compares different machine learning models.With the greatest accuracy and F1 score of 0.98 among the models assessed, the Artificial Neural Network (ANN) performs exceptionally well overall in terms of making accurate predictions and striking a balance between recall and precision. Additionally, XGBoost performs remarkably well, matching ANN in the majority of metrics and even significantly outperforming it in recall, indicating that it is especially good at spotting favorable examples. K-Nearest Neighbors (KNN) and Decision Tree come in second and third, respectively, with constant scores of 0.96 to 0.97 across all metrics. While still effective in terms of accuracy and F1 scores of approximately 0.95, Support Vector Machine (SVM) and Logistic Regression fall somewhat short in terms of precision and recall when compared to the more sophisticated models. In conclusion, this comparison shows that ANN and XGBoost are the most reliable models, performing better in terms of prediction across almost all evaluation metrics.

$$\text{ACCURACY} = \frac{\text{TN+TP}}{\text{TN+FP+TP+FN}} \tag{12}$$

$$\text{PRECISION} = \frac{\text{TP}}{\text{TP+FP}} \tag{13}$$

$$\text{RECALL} = \frac{\text{TP}}{\text{TP+FN}} \tag{14}$$

Recall shows 1 if both numerator and denominator are equall.

$$\text{F1SCORE} = 2 * \frac{\text{PRECISION}*\text{RECALL}}{\text{PRECISION+RECALL}} \tag{15}$$

F1SCORE become 1 if precion and recall becomes 1 and high of both respectivel shows high.

Table 2. Over all performances Metrices for crop recommendation

Models	Accuracy	Precision	Recall	F1score
Logistic Regression	0.94	0.95	0.94	0.94
Decision Tree	0.96	0.96	0.96	0.96
SVM	0.98	0.97	0.97	0.97
KNN	0.97	0.98	0.97	0.97

XGBoost	0.97	0.97	0.98	0.98
ANN	0.98	0.98	0.98	0.98

In figure 3 shows the correlation between -1 to 1 values close to 1 shows posstive correlation and values close to -1 shows negative correlation a value around 0 show no relationship.

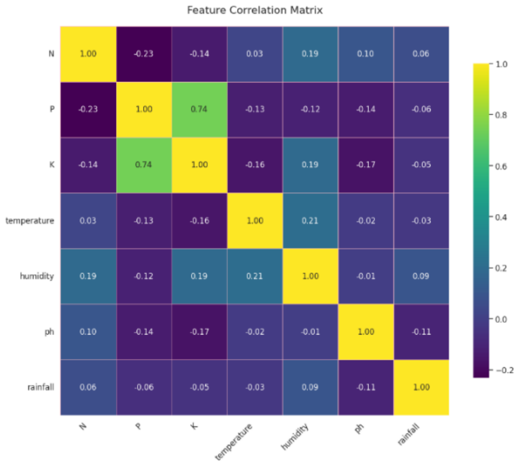

Figure 3. Feature Correlation of Crop

5. Conclusion

In this paper, the researchers present a Smart Crop Recommendation system utilizing Ensemble Learning of Supervised Models, designed for easy application by farmers across Pakistan. This model facilitates informed decision-making for crop cultivation by considering multiple parameters. As a result, it enhances agricultural productivity on a countrywide scale, providing improved decision-making for crop production. In comparing the different models, the results indicate that all listed models exhibit strong performance across various metrics, with the Artificial Neural Network (ANN) achieving the highest levels of accuracy and precision.

References

1. Modi, D., Sutagundar, A. V., Yalavigi, V., & Aravatagimath, A. (2021, October). Crop recommendation using machine learning algorithm. In 2021 5th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-5). IEEE.
2. Garg, D., & Alam, M. (2023). An effective crop recommendation method using machine learning techniques. International journal of advanced technology and engineering exploration, 10(102), 498.
3. Khatibi, S. M. H., & Ali, J. (2024). Harnessing the power of machine learning for crop improvement and sustainable production. Frontiers in plant science, 15, 1417912. https://doi.org/10.3389/fpls.2024.1417912
4. Wang, J., Wang, Y., Li, G., & Qi, Z. (2024). Integration of remote sensing and machine learning for precision agriculture: a comprehensive perspective on applications. Agronomy, 14(9), 1975.
5. Nagesh, O. S., Budaraju, R. R., Kulkarni, S. S., Vinay, M., Ajibade, S. S. M., Chopra, M., ... & Kaliyaperumal, K. (2024). Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. Discover Sustainability, 5(1), 78.
6. Upadhyay, S. K., Kushwaha, P., & Garg, P. (2024, March). A Machine Learning Approach for Predicting Crop Yield in Precision Agriculture. In 2024 International Conference on Automation and Computation (AUTOCOM) (pp. 289-294). IEEE.
7. Bhatti, M. A., Zeeshan, Z., Syam, M. S., Bhatti, U. A., Khan, A., Ghadi, Y. Y., ... & Afzal, T. (2024). Advanced plant disease segmentation in precision agriculture using optimal dimensionality reduction with fuzzy c-means clustering and deep learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
8. Fegade, T.K., Pawar, B.V.: Crop prediction using artificial neural network and support vector machine. Data Manage. Anal. Innov.: Proc. of ICDMAI 2019(2), 311–324 (2020).
9. Rajamanickam, Jayalakshmi, and Savitha Devi Mani. "Kullback chi square and Gustafson probabilistic neural network based soil fertility prediction." Concurrency and Computation: Practice and Experience 33.24 (2021): e6460.
10. Katarya, Rahul, et al. "Impact of machine learning techniques in precision agriculture." 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE). IEEE, 2020.
11. McEldowney JF (2021) Climate change and the law. In: the impacts of climate change, Elsevier. 503–519.
12. Hasan, M., Marjan, M. A., Uddin, M. P., Afjal, M. I., Kardy, S., Ma, S., & Nam, Y. (2023). Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation. Frontiers in Plant Science, 14, 1234555.
13. Konaté J, Diarra AG, Diarra SO, Diallo A (2020) SyrAgri: a rec ommender system for agriculture in mali'. Information 11(12):561
14. Kala K., Nandhini M., Chakkravarthi M. K., Thangadarshini M., and Verma S. M., Deep Learning Techniques for Crop Nutrient Deficiency Detection—A Comprehensive Survey, Precision Agriculture for Sustainability. (2024) 319–326.
15. Liang C. and Shah T., IoT in Agriculture: The Future of Precision Monitoring and Data-Driven Farming, Eigenpub Review of Science and Technology. (2023) 7, 85–104.
16. Panda C. K., Advances in Application of ICT in Crop Pest and Disease Management, Natural Remedies for Pest, Disease and Weed Control, 2020, Elsevier, Amsterdam, Netherlands, 235–242.
17. Holt N., Sishodia R. P., Shukla S., and Hansen K. M., Improved Water and Economic Sustainability With Low-Input Compact Bed Plasticulture and Precision Irrigation, Journal of Irrigation and Drainage Engineering. (2019) 145.
18. Roy T. and George K J., Precision Farming: A Step towards Sustainable, Climate-Smart Agriculture, Global Climate Change: Resilient and Smart Agriculture. (2020) Springer, Berlin, Germany, 199–220.
19. Parihar C., Jat H., Jat S., Kakraliya S., and Nayak H., Precision Nutrient Management for Higher Nutrient Use Efficiency and Farm Profitability in Irrigated Cereal-Based Cropping Systems, Indian Journal of Fertilisers. (2020) 16, 1000–1014.
20. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset
21. Chauhan, G., & Chaudhary, A. (2021, December). Crop recommendation system using machine learning algorithms. In 2021 10th international conference on system modeling & advancement in research trends (SMART) (pp. 109-112). IEEE.