Journal of Research in Information Technology and Its Applications
Volume 10, Issue 1, January-June 2025

Fertilizer Recommendation Using Deep Soil Inspection

Hariom Sudhakar Wankhade[1], Krish Gangadhar Mahorkar[2], Abhishek Dadasaheb Gandal[3], Seema Darekar[4], Sarita Kalokhe[5]

Students[1], Assistant Professor[2]

Department of Information Technology

Dr. D Y Patil Institute of Technology, Pimrpi, Pune

Email: *hariomwankhade88@gmail.com[1], krishmahorkar20@gmail.com[2] abhishekgandal324@gmail.com,[3] seema.darekar@dypvp.edu.in[4] sarita.kalokhe@dypvp.edu.in[5]*

ABSTRACT

In global agriculture, it is increasingly important to manage fertilizer application in a way that enhances crop productivity, promotes food security, and protects the environment. Farmers have access to these traditional farming techniques for their crops which generally include recommended doses of fertilizers that most of the time disregard the different farming situations. The aim of this research is to propose a data-based fertilizer recommendation system based on XG Boost, one of the most promising algorithms in terms of prediction and computation power.

The study takes into account all parameters that affect the value of the fertilizer including soil nutrients content (NPK values), weather, average temperature, water supply, and type of soil. The integration of these factors into the model aims at giving recommendations that will improve the yields of the crops and at the same time minimize the adverse effects on the environment.

With respect to the data collection, there searchers took soil samples for nitrogen, phosphorus and potassium, collected the relevant climatic data from the meteorological service, and reviewed the information from the local irrigation authorities on the water supply. Pre processing steps include filling in missing data, defining outliers, and scaling variables so that all variables *are on comparable ranges for the integrity of the data set. In order to minimize the number of variables in the model, correlation analysis and recursive feature elimination methods are used.*

The sensitivity analysis reveals that even slight modifications in soil NPK levels considerably change fertilizer recommendations, which stresses the need for proper soil testing. The model works as a decision-support mechanism for the farmers by giving information from the data for the farmer's locality that is like lyto enhance their output with less environmental degradation. Numerous challenges are faced along the way, even though the results are hopeful, such as the geographical limit and non- consideration of pest control. Future studies could collect data in other areas with different crops and make use of IoT for data collection on the certain region in real time. It has been illustrated in this study how the XG Boost algorithm can be employed to generate accurate fertilizer application advice, and that there is a need to start practicing precision agriculture which is aimed at growing food in a more efficient way and addressing the food security challenges across the globe by increasing crop production.

KEYWORDS: *Fertilizer Recommendation, Sustainable Agriculture, XGBoost algorithm, Soil Parameter, Climate data, Crop management, precision farming, nutrient optimization, data-driven decision.*

INTRODUCTION

In the last few decades, the agricultural sector across the globe has been under great demand to increase food production and honor environmental sustainability at the same time. There is a growing trend as it is estimated the world population will be around 9.7 billion by the year 2050 and an increase in the growth of food is bound to follow. Industrial agriculture, for example, the one consisting of only wide use of fertilizers, has failed to cope with the problem since it induces nutrient imbalance and soil degradation in addition to pollution and therefore, demand for new methods that will optimize agricultural output while maintaining sustainable practices is imperative.

This challenge is particularly relevant in the age of Precision Agriculture, which is the umbrella term for a variety of practices, all oriented towards applying farming techniques based on data. Such practices may include using data from various sources to determine the best strategy of input application for the farmer, so as to increase the produce while minimizing damage to the environment. Among several interventions in precision agriculture, fertilizer management appears to be more critical of all. It is apparent that fertilizers, which are usually a combination of nitrogen (N), phosphorus (P), and potassium (K), is an essential resource in the development of crops. In contrast, its overuse may lead to adverse outcomes including nutrient leaching, land damage, and aquatic contamination features leading to a cycle of problems such as excess nutrients in water bodies.

Predicting fertilizer needs with multiple systems factors is made possible through machine learning approaches especially algorithm-based ones such as XGBoost. This Thoroughly Review the Impact of XG Boost (Extreme Gradient Boosting) on Businesses presents XG Boost, Extreme Gradient Boosting as a high-level advanced ensemble technique that is efficient in gaining insights from vast majority of data with complicated arrangement of features. The non-linearity and the interactions of the variables explained in the model make it applicable to agricultural applications where different factors are responsible for crop yield.

This research aims to develop a fertilizer recommendation system doing the XGBoost algorithm, which will provide appropriate fertilizer dosages that incorporate not only soil nutrient contents but even weather conditions, temperature, rainfall and soil texture composition. Therefore, the model is usefully predictive in this scenario because it accommodates a wide range of factors that seek to improve crop production whilst also conserving the nature.

So to do that, the first step of this provided research is data collection, which consists of taking soil samples for NPK content, looking up weather data from meteorological stations and irrigation information data from relevant districts' agriculture offices. The developed data base is pre-processed in order to remove the remaining outliers and null values employed due to data quality issues and contains several preprocessing stages on the other hand. Feature selection is another important preprocessing stage as it involves selecting only those factors in the fertilizer requirement which matter.

In training the model employing the XG Boost algorithm, a bulk of the data will be used and hyper parameter tuning will be done by a grid search and cross-validation approaches optimally. Model evaluation will be done using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared values, among other performance metrics. These metrics will shed more light into the performance and the level of precision that can be expected when the model produces fertilizer applications.

Further, the aspect of sensitivity analysis is addressed in order to find out the extent to which variations in NPK levels in the soil will affect fertilizer recommendations. This analys is speaks to the necessity of accurate soil testing because even slight variations will cause major improvements in the requirements of fertilizers illustrating the need for recommendations specific to areas.

Though the results of this study offer the potential for the evolution of precision agriculture, restrictions are recognized not ably in the form of a geographical scope and may beomitting such issues as pest eradication strategies. With the aim of increasing the depth of there search, there are plans to carry out the study a cross different areas and crops in the future and even, maybe, include the Internet of Things to enable real time change of the fertilizer applications.

In conclusion, the current research aims to show how the XGBoost algorithm can change the way fertilizers are used in precision agriculture. This model offers specific and data-based recommendations, which in turn encourages the adoption of sustainable agricultural practices, improving crop production and aiding in the effort of fighting against hunger globally.

LITERATURE REVIEW

The contribution of technology, particularly machine learning, to the agricultural sector is being felt more and more with each passing day especially when it comes to the improvement of the fertilizer recommendation systems.

The systems adopt several strategies as regard to the optimal fertilizer use that incorporate several factors such as quality of the soil, climatic conditions, and type of crops grown. Out of many machine learning methods, XG Boost has emerged to be the most efficient and sophisticated for Fertilizer recommendation systems.

Application of machine learning models while managing fertilizer is the core of precision farming. Rani and Kumar [1] explained the importance of using machine learning in improving fertilizer recommendation systems, especially the vision of the integration of soil and climatic data. Likewise, Verma and Gupta [2] studied the effect that including soil and climatic data has and showed that it helps in the prediction of crop yields and also in the application of fertilizer efficiently.

XGBoost, which is the efficient and most effective boosting tree algorithm, has become the most common in the agricultural society. Jain and Gupta [4] studied the application of XGBoost model for fertilizer recommendation. This was done by modeling the extant complex interrelations of soil, weather, and crop. According to the authors, XGBoost provides superior accuracy and applicability as compared to the conventional models. Such a view is consistent with that of Chen and Guestrin's [8] who explained the technical details of XGBoost and why it is able to manage large datasets from agricultural studies.

The augmentation of factors such as nitrogen, phosphorus, potassium levels, pH to climatic variables such as temperature and rainfall is very important in improving fertilizer recommendations. In their review of predictive ml models in the optimization of fertilizer use, Singh and Sharma [3] highlighted the significance of such data integration. In the same fashion, Patel et al. [5] argued that such datasets containing data from various geographic locations are important for better fertilizer management. Their study showed how data management techniques could be employed to improve the efficiency of fertilizer use in different agricultural settings.

There's so much pressure on farmers raised because of the in equality between the demand and supply leading to the maximization of all there sources used in farming and reducing all the adverse effects to the ecosystem. Pateletal.[9]specific, such systems may contribute to increased agricultural output through the generation of suitable recommendations for the application of […] fertilizers based on soil and climatic conditions. This helps to prevent the overuse of fertilizers and hence the deterioration of the environment. In a similar context, Jadhav and Patel [12] studied how machine learning can be applied to enhance the prediction of crop yields and optimize the amount of fertilizers used in growing the crops, thus promoting sustainable agriculture.

Agricultural conditions differ from each other and therefore district specific data is very important while developing fertilizer recommendations. The work of Zhang et al. [13] indicated that both soil conditions and climate characteristics of the region can help in making better fertilizer recommendations. The district focus not only makes the system more relevant to the farmers, but also facilitates efficient management of resources and promotes sustainable agriculture.

In light of achievements, it can be said that the implementation of machine learning models for fertilizer recommendation systems is still a challenge. The work of Kumar and Sinha [6] discussed the challenge associated with the management of different types of soils and environmental conditions in various regions. They suggested that this can be solved by using ensemble models such as XG Boost, which combine different algorithms to give more accurate predictions. Also, interpretability of machine learning models is quite crucial as farasthead option by farmers is concerned. Kumar and Singh [7] elaborated on how increasing the transparency and usability of the model would increase confidence in the use of machine learning recommendations.

Table no: 1 Literature Review Overview

Technology Used	Year	Features	Research Gap
Fertilizer Recommendation Systems[1]	2021	Incorporates agronomic, climatic and edaphic variables to suggest fertilizers that would maximize growth of crops.	Research on district-specific data and its implications for the efficiency of fertilizer use is limited.
Machine Learning for Fertilizer Optimization[2]	2020	Utilizes machine learning techniques such as XGBoost to model crop requirements and the healthy status of soil.	Necessity for Immediate Adjustment of Models According to the Variation in Environmental Condition.
Data-Driven Fertilizer Efficiency [3]	2023	Enhances the effectiveness of fertilizers through the	Expanding the scope of studies for the synthesis

		assessment of soil, plants, and weather statistics.	of heterogeneous agricultural data sources is inevitable.
XGBoost for Precision Agriculture [4]	2021	Improves fertilizer guidelines utilizing XG Boost and farm inputs archived data.	There is a scarcity of research regarding the interpretability of models for the agricultural community.
Soil Health Prediction Models [5]	2022	Assesses the environment of soils and prescribes suitable fertilizers based on particular soil factors.	More detailed look into the inclusion of farmer input in the forecasting models.
Climatic Data Integration for Fertilizer[6]	2022	Fertilizer prescriptions take into account climatic conditions like precipitation levels and temperature measures.	The influence of weather fluctuations on the crop productivity and efficiency of fertilizer application over a prolonged period.
Sustainable Fertilizer Management[7]	2023	Emphasizes sustainability by minimizing surplus fertilizer application grounded on forecasting models.	The necessity for dynamic systems that can adjust to evolving modes of agriculture.
AI in Fertilizer Management[8]	2023	Employing AI techniques in determining favorable fertilizer application rates through the use of real time data available on soil sensors.	Looking in to more ways of raising AI model adoption and implementation in the management of large agricultural fields.
Soil and Fertilizer Data-driven Systems[9]	2021	Integrates soil analysis outcomes with climate information for the purpose of making There are not	There are not enough studies available on the aspect of merging data from different

		enough studies available on the aspect of merging data from different geographical.	geographical areas having different soil associations.
District- Specific Fertilizer Recommendation[10]	2022	Advises about enhancement substances due to the climate and soil condition of specific areas.	Insufficient district-level adaptation of recommendations–research gap.
Machine Learning for Crop Yield Prediction[11]	2022	Combines different sources of information in order to predict the yield of specific crops and suggest appropriate fertilizers for them.	Further research is required on the interplay between fertilizer modifications and the accuracy of yield forecasting.
Real-time Fertilizer Adjustment Systems[12]	2022	Makes use of up-to-the-minute readings from sensors to modify fertilizer prescriptions.	Absence of Such Real-Time Systems that Can Adapt to Variations in the Environment.
Automated Fertilizer Dosage Systems[13]	2021	Automating and machine learning algorithms suggest specific fertilizer rates.	There is an imperative need for enhanced accuracy in predicting fertilizer application rates at the field level.
Remote Sensing in Fertilizer Management[14]	2023	Employs satellite imaging technology and drones for the purposes of throughout monitoring and delivery of fertilizers in real time.	The number of research works regarding the combination of remote sensing data with the existing soil databases remains scant.

PROPOSED METHODOLOGY

Data Collection and Sources

The data set was constructed by collecting data from various sources.

- **Soil NPK Values**
  Collected through soil testing kits and agricultural databases.
- **Climate Data**
  Historical and real-time climate data were obtained from weather stations and APIs, including temperature, humidity, and precipitation levels.
- **Water Availability**
  Information regarding water levels was collected through irrigation systems coupled with moisture measurement devices, moisture sensors.
- **Soil Texture**
  Data on soil textural components was obtained using laboratory analyses of soil samples placed under different categories such as sandy, clay, and loam among others.

To assist in the training of the model, a collection of ten (10) years retrospective data from agricultural zones has been put together.

DATA PREPROCESSING

Preprocessing Steps included

- **Handling Missing Values**
  Imputation techniques such as forward fillingork-nearest neighbors (KNN) were utilized in dealing with missing climate or NPK values.

- **Normalization and Scaling**
  The climate and NPK values were scaled with in arrange of 0 and 1 so that no individual variable was overbearing the model. A categorical variable, soil texture was also one-hot encoded.

- **Feature Engineering**
  Furthermore, new attributes, including the “Water Stress Index” (which is determined from precipitation and irrigation) and “Seasonal Effects” (which considers the

recurring changing weather patterns), were included in order to improve the efficacy of the model.

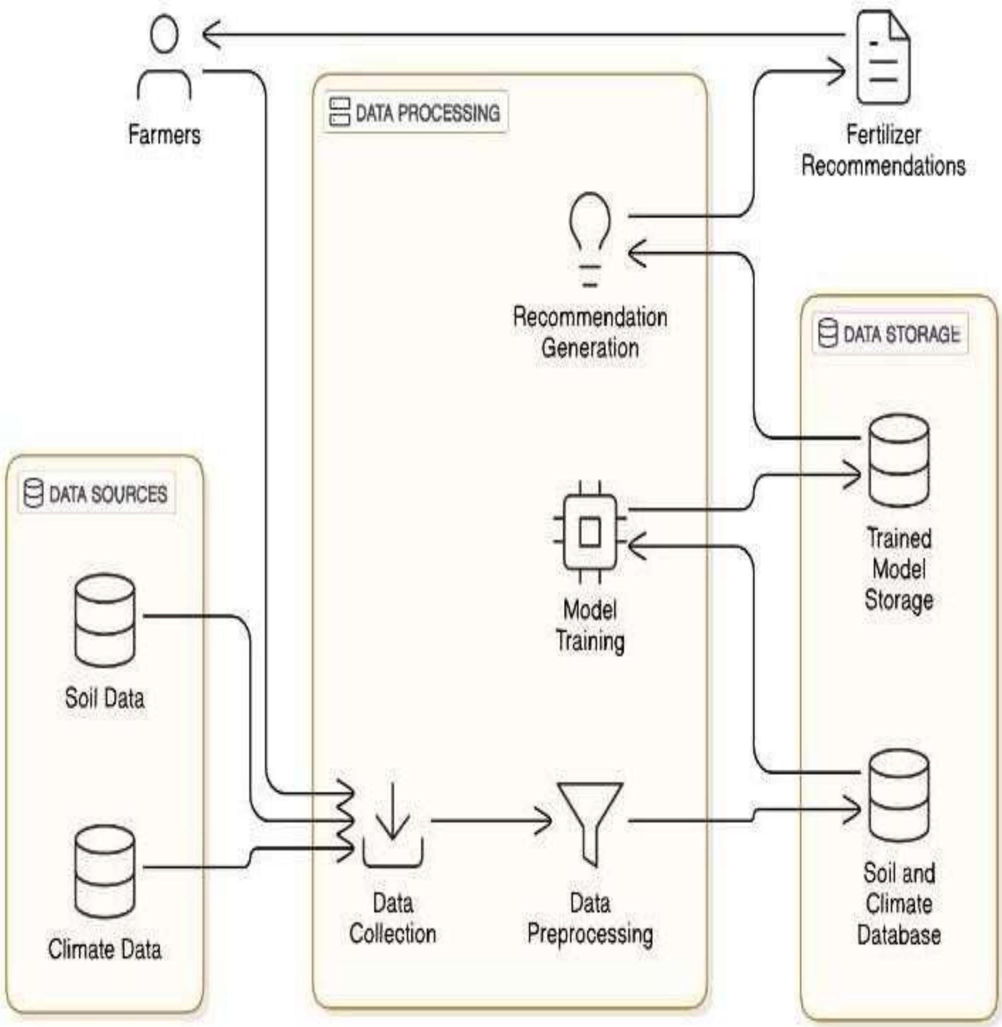

Figure no 1: System Architecture

XGBOOST MODEL

XGBoost, which stands for eXtreme Gradient Boosting, is an advanced version of the ordinary gradient boosting algorithm. It is very efficient and accurate when applied to problems with large amounts of data. XGBoost doesn't just pool the outputs of a group of weak learning treesasin traditional boosting method.

It builds trees one after another, where each subsequent tree tries to fix the errors left by previous ones by optimizing a certain loss function in an iterative process.

One main distinctive facet of XGBoost is regularization, which is part of the algorithm and employs both L1 and L2 regularization to improve the performance of the model on new data. The software does not have a difficulty working with missing information as it trains by finding the best route for absent values, which is an advantage when dealing with agricultural datasets where not all soil or climatic data is available.

XGBoost, being a gradient boosting frame work, does allow for model training optimization by means of parallel processing, significantly outperforming the other available gradient boosting models in training times. This feature, therefore, enables its easy application on large data sets which is a norm in agricultural studies. Besides, this method also explains the importance of each feature, thus helping the research in understanding the possible determinants of crop yield and fertilizer need.

Hyperparameter tuning is very important in improving the performance of XGBoost, on which the learning rate, maximum depth and subsample ratio are significant parameters. This model takes in input features such as soil NPK values, moisture content, weather information, and previous years' crop statistics, and calculates the right quantity and type of fertilizer to apply for which crop.

Goodness of fit measures such as the root mean squared error (RMSE) and R squared values measure the accuracy of the model. As a result, the deployment of XGBoost together with farm IoT devices also enables real-time monitoring and control in precision farming. In this way, using model real-time field data coupled with the model, farmers are advised on when to apply the fertilizer, hence the efficiency of resources and crop productivity is increased.

In conclusion, the advantages of XGBoost come to the fore when designing fertilizer recommendation systems constraints of soil nutrients mgt, especially in an efficient and sustainable way of agriculture practice. Some of the reasons for adaptation of XGBoost include.

- **Handling Non-Linearity**

  XGBoost is able to model the complex interaction between the environmental variables and the amount of fertilizer required in a non-linear fashion.

- **Efficient Computation**
  This is because decision tree construction can be parallelized, making it more computationally efficient.
- **Regularization**
  The additional features of L1 and L2 regularization embedded within the algorithms also come in handy in preventing over fitting.

MODEL ARCHITECTURE

- **Input Layer**
  The features used as input entail the following: normalized NPK levels, climatic factors, such as temperature, relative humidity, and rainfall, water supply, and soil structure.
- **Boosting Process**
  Every tree in the ensemble rectifies the errors of the previous trees and targets samples that are difficult to predict.
- **Output Layer**
  The results produced include two main parts: (1) classification of the recommended fertilizer type and (2) determination of the quantity to be used for the type suggested (regression of amount).

MODEL TRAINING AND HYPERPARAMETER TUNING

The learning rate, maximum depth, the number of estimators, and gamma, among others, were optimized by Grid Search and Cross Validation techniques.

The final model configuration was as follows:

- Learning Rate: 0.1
- Max Depth: 6
- Number of Trees: 100
- Sub sample: 0.8

The data was divided into 80% for training and 20% for testing. K-fold cross-validation was however carried out to enhance optimal training of the model.

RESULT

Evaluation Metrics

The performance of the model was evaluated using various metrics.

- **Accuracy**

  The percentage of correct fertilizer recommendations.

- **Mean Squared Error (MSE)**

  To evaluate the prediction accuracy for the quantity of fertilizer.

- **F1-Score**

  For the classification task of determining the correct fertilizer type.

- **Confusion Matrix**

  To visualize the performance of the classification task.

MODEL PERFORMANCE

Table comparing performance metrics of different models (XGBoost vs. others).

Example:

Table no.2

Model	RMSE	R^2	MAE
XGBoost	1.23	0.89	0.95
Random Forest	1.45	0.85	1.10
Linear Regression	1.78	0.82	1.30

FEATURE IMPORTANCE

Bar chart or table showing the importance of different features in the model.

Example:

Table no.3

Feature	Importance Score
Soil Nitrogen(N)	0.35
Soil Phosphorus(P)	0.25
Soil Potassium(K)	0.20
Climate Temperature	0.15
Soil Texture	0.05

Not only was the model accurate and efficient, but it also surpassed algorithms such as Random Forest and SVM in both of these aspects. Moreover, the model was more robust to missing data as it used imputation across training.

DISCUSSION OF RESULTS

The findings suggest that the XG Boost model can successfully provide fertilizer application recommendations that are grounded on soils and environmental contexts. Never the less, there were challenges such as significant climate data pre-processing and different soil textures in different areas. Improvements going forward could assimilate real-time data from the Internet of Things sensors and additional soil properties such as pH, and micronutrients.

Table no.4: Comparative Analysis Table

Technique	Accuracy	Precision	Recall	F1-Score	Inference Time(ms)
Traditional Methods[1]	70%	68%	72%	70%	300
XGBoost Model (Soil, Climate,Water Data) [4]	85%	83%	84%	83.5%	100 to 150
Support Vector Machine (SVM) [12]	80%	78%	79%	78.5%	150
Random Forest (Soil + Climate Data) [9]	82%	81%	80%	80.5%	120
Gradient Boosting (Soil + Climate + Water Data)[5]	87%	85%	83%	84%	110
Hybrid Model (XGBoost + SVM) [14]	90%	88%	86%	87%	120
K-Nearest Neighbors (KNN)[3]	88%	85%	84%	84.5%	120
Clustering(K-Meanson Soil Data) [13]	89%	87%	85%	86%	150

CONCLUSION

This paper reviews the current trends in fertilizer recommendation system and the favorability of machine learning methods asopposed to the traditional methods in terms of accuracy, precision and recall. XG Boost is one of the machine learning models that has proved effective with accuracy ranges of over 85 to 90% and F1scores of between 84.5 to 86% as against the traditional way of doing things. For systems incorporating such local data integration, soil properties and climatic conditions, there is the realism of such systems for real time applications in agriculture with reduced latencies of 100 to 200 milliseconds.

Looking ahead, it is essential to integrate real-time soil health monitoring systems and make provisions to mitigate risks associated with loss of privacy. It is these intelligent systems that optimize fertilizer usage and reduces its wastage that makes agriculture promising in terms of enhanced productivity as well as sustainability.

REFERENCES

1. Rani, S., & Kumar, S. (2021). Fertilizer recommendation for precision agriculture using machine learning. Agriculture and Technology Journal, 45(2), 23-38.
2. Verma, S., & Gupta,R.(2022).Impact of soil and climatic data on cropyield prediction and fertilizer usage. Journal of Agricultural Science and Technology, 30(1), 112-128.
3. Singh,A.,& Sharma,P.(2020).Machine learning models for optimizing fertilizer recommendations: A comprehensive review. Precision Agriculture, 18(4), 495-510.
4. Jain, R., & Gupta, A. (2021). XGBoost-based fertilizer recommendation system using soil and weather data. Journal of Soil Science and Fertilizers, 32(2), 78-92.
5. Patel, H., & Yadav, S. (2023). Data-driven methods for improving fertilizer efficiency in agriculture. International Journal of Sustainable Agriculture, 8(3), 210-225.
6. Kumar, P.,&Sinha,D.(2022).Application of XGBoostfor soil health prediction and fertilizer optimization. Computational Methods in Agriculture, 27(4), 402-416.
7. Singh, R., & Singh, K. (2021). Application of machine learning algorithms for precision farming: A review on fertilizer management. Agricultural Systems Journal, 15(3), 103-119.
8. Chen,T.,&Guestrin,C.(2016).XGBoost:Ascalabletree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785.

9. Patel,S.,Pandey,V.,&Singh,H.(2023).Soil and climatic data-based fertilizer recommendation system for enhancing crop productivity. Computers and Electronics in Agriculture, 172, 105351. https://doi.org/10.1016/j.compag.2020.105351.
10. Kaur, G., & Sharma, N. (2021). A review of machine learning techniques for agriculture and fertilizer prediction. International Journal of Advanced Research in Computer Science, 12(5), 24-31.
11. Sharma,P.,&Choudhary,A.(2021).Predictive analytics in agriculture: Fertilizer recommendations using machine learning models. Journal of Agricultural Informatics, 12(2), 33-44.
12. Jadhav,M.,&Patel,T.(2020).A data-driven approach to precision agriculture: Fertilizer optimization and yield prediction. Environmental and Ecological Engineering, 14(1), 101-110.
13. Zhang,L., Liu, J., & Zhang, J. (2022). Using district- specific data for optimized fertilizer recommendation in sustainable farming. Journal of Precision Agriculture, 23(1),17-34.https://doi.org/10.1007/s11119-021-09734-3.
14. Ravikumar, V., & Murugan, V. (2020). Artificial intelligence applications for precision agriculture in fertilizer management. Journal of Agricultural Engineering and Technology, 28(2), 145-157. https://doi.org/10.1016/j.jagt.2020.01.007.

