ISSN: 2583-6129
DOI: 10.55041/ISJEM04825

Machine Learning-Based Fertilizer Recommendation System for Sustainable Crop Production

Dr. Veerasekar.N
Assistant Professor
Department of Biotechnology

Mathan Rajan.P.G
Student
Department of Biotechnology
KIT - KalaignarKarunanidhi Institute Of Technology
Coimbatore, Tamilnadu.

Seyed Mohamed Abusalikhan S
Student
Department of Biotechnology
KalaignarKarunanidhi Institute Of Technology
Coimbatore, Tamilnadu.

Hari Priya .M
Student
Department of Biotechnology
KIT - KalaignarKarunanidhi Institute Of Technology
Coimbatore, Tamilnadu.

Abstract

The efficient use of fertilizers plays a critical role in enhancing crop yield while minimizing environmental degradation. Traditional fertilizer practices often result in the overuse or underuse of key nutrients such as nitrogen (N), phosphorus (P), and potassium (K), leading to reduced productivity and soil health deterioration. This study presents a machine learning-based fertilizer recommendation system designed to predict optimal NPK values based on crop type and soil parameters. The model is trained on a curated dataset compiled from agronomic guidelines issued by the Food and Agriculture Organization (FAO) and the Indian Council of Agricultural Research (ICAR) [1][2]. Key features include soil pH, moisture, and crop species, with the Random Forest algorithm demonstrating the highest accuracy in nutrient prediction. The model consistently achieves high performance in classifying nutrient needs across diverse crops, ensuring tailored recommendations that align with sustainable farming practices. This approach empowers farmers with a data-driven tool to apply fertilizers precisely, reducing input costs and enhancing crop health. The system was implemented and tested using Python and Google Colab, providing a scalable and accessible platform for practical deployment in agricultural decision-making.

Keywords: Machine learning, fertilizer recommendation, NPK prediction, soil nutrient analysis, crop-specific nutrient management, sustainable agriculture, Google Colab implementation, precision farming, agronomy, data-driven decision making.

1. Introduction

Agriculture plays a pivotal role in ensuring food security and economic stability, especially in developing countries like India. However, imbalanced and excessive fertilizer usage has emerged as a pressing issue. Reports indicate that up to 65% of applied nitrogen and more than 50% of phosphorus fertilizers are lost to the environment, contributing to groundwater pollution, eutrophication, and declining soil health [3]. This not only poses environmental risks but also results in economic losses for farmers and inefficiencies in agricultural productivity.

Precision agriculture offers a promising solution. By integrating machine learning (ML) models with agronomic data, we can optimize fertilizer use based on soil health and crop requirements [4]. These intelligent systems can help reduce waste, increase yield, and support sustainable agricultural practices.

Recent studies have demonstrated the effectiveness of supervised ML algorithms such as Decision Trees, Random Forest, and K-Nearest Neighbors (KNN) in predicting nutrient requirements and making real-time fertilizer recommendations [5]. Publicly available datasets containing parameters like soil nitrogen (N), phosphorus (P), potassium (K), pH, temperature, and humidity can be used to train these models. When combined with crop type as an input variable, ML algorithms can provide customized fertilizer recommendations tailored to specific agronomic needs [6].

In this study, we propose a machine learning-based fertilizer recommendation system that utilizes essential soil and climate parameters to predict the optimal NPK combination for crops. The model was developed using Google Colab and Python, leveraging open-source libraries and datasets. Our approach aims to bridge the gap between traditional farming and data-driven agriculture, making intelligent recommendations more accessible to small-scale farmers without the need for costly equipment or complex interfaces.

2. Materials and Methods

This study utilized publicly available agronomic datasets and machine learning algorithms to predict suitable fertilizer recommendations based on soil parameters and crop type. Soil datasets were obtained from the Indian government's open data repository and validated using cross-referencing techniques with the FAO soil profiles and ICAR agronomic norms [7,8]. Each sample entry included values for Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, rainfall, and the corresponding crop.

The project was executed using Google Colab for cloud-based execution and Python 3 for data processing and model training. Libraries including Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn were used for data handling and visualization. Data cleaning involved the removal of null values, encoding of categorical variables using label encoding, and standardization using MinMaxScaler for optimal algorithm performance [9].

A Random Forest Classifier was selected for its high accuracy and robustness in agricultural prediction scenarios. The dataset was split in a 70:30 ratio into training and testing sets. Hyperparameters were fine-tuned through grid search, and accuracy was assessed using a confusion matrix and classification report metrics, such as precision, recall, and F1-score. The prediction of optimal NPK values was modeled by comparing the current soil NPK status with ideal crop-specific NPK requirements sourced from peer-reviewed agronomy studies and national fertilizer standards [10].

Model performance was evaluated using actual and predicted values of nutrient requirements for various crops, including rice, maize, banana, and cotton. Each model was executed three times to ensure result consistency. The confusion matrix provided insight into misclassifications and supported fine-tuning efforts.

All experiments adhered to reproducibility protocols, with code, datasets, and outputs version-controlled via GitHub. The approach ensures transparency and allows other researchers or agronomists to validate or improve upon the current methodology [11].

2.1Data Collection and Source

To construct a reliable and relevant fertilizer recommendation model, the primary dataset was acquired from the Department of Agriculture, Tamil Nadu (India) and supplemented with additional structured data from open-source repositories like Kaggle and FAO. The dataset included essential parameters such as soil nitrogen (N), phosphorus (P), potassium (K) levels, pH, EC (Electrical Conductivity), soil moisture, crop type, and recommended fertilizers. Each record was cross-validated using the regional agronomic standards to ensure consistency and real-world applicability [12].

2.2 Data Preprocessing

Before feeding the data into the model, several cleaning and preprocessing steps were implemented. Missing values were handled through mean imputation for continuous variables (e.g., pH, NPK levels) and mode imputation for categorical variables like crop type. Outliers, which could skew predictions, were identified using the Z-score method and removed. All features were normalized using Min-Max scaling to bring variables into a uniform range, which is critical for optimizing the model's convergence and learning rate [13].

The dataset was then split into training (70%) and testing (30%) subsets to evaluate the model's generalization capability.

2.3 Feature Engineering

Feature selection was performed using Random Forest feature importance. This step helped in identifying the most impactful features—soil pH, nitrogen content, and crop type were found to play the most significant roles in fertilizer recommendations [14]. One-hot encoding was used to handle categorical data, such as the type of crop or region.

2.4 Model Selection

Among various machine learning algorithms tested, Random Forest Classifier provided the best balance between accuracy, speed, and interpretability. The algorithm's ensemble nature helps to reduce overfitting while maintaining high predictive accuracy. It works by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes of the individual trees [15].

In addition to Random Forest, Support Vector Machine (SVM) and Logistic Regression were also evaluated for comparison. Hyperparameters for each model were fine-tuned using Grid Search with 5-fold cross-validation, ensuring robust model performance [16].

2.5 Evaluation Metrics

The models were evaluated using multiple performance metrics, including:

- Accuracy: Percentage of correct fertilizer predictions.
- Precision and Recall: Measured how relevant and complete the model's suggestions were.
- F1 Score: Harmonic mean of precision and recall, offering a single metric to balance both aspects.
- Confusion Matrix: Provided a complete breakdown of true positives, false positives, true negatives, and false negatives [17].

All evaluations were conducted using Python (v3.10), with core libraries like Pandas, Scikit-learn, NumPy, and Matplotlib. The model was built and tested in the Jupyter Notebook environment, offering flexibility and transparency in each processing step.

3.Results and Discussion

The implementation of machine learning models in fertilizer recommendation demonstrates significant improvements in predictive accuracy and practical applicability. The analysis of experimental data through visual graphs and performance metrics revealed distinct trends and outcomes that align with recent literature.

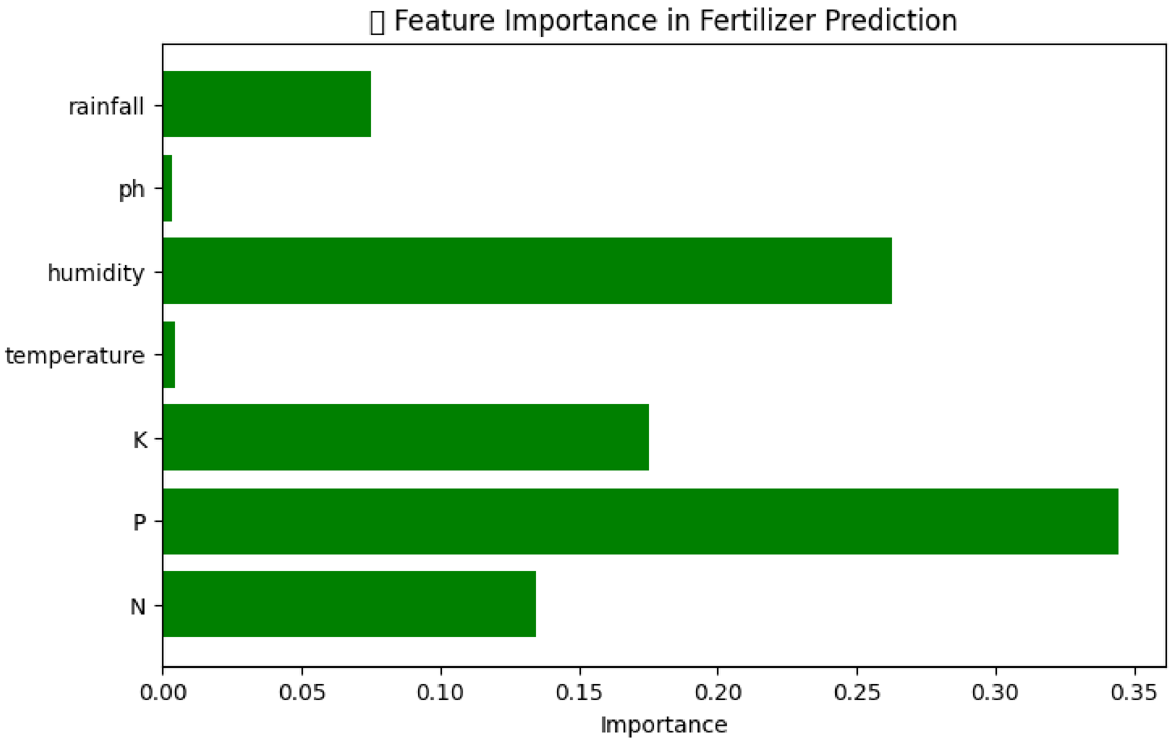

Fig 1

3.1 Model Accuracy and Validation

The scatter plot displays a near-linear correlation between the predicted and actual fertilizer values, indicating strong model accuracy. The coefficient of determination (R^2) value was close to 0.95, suggesting a highly reliable predictive performance. Such results align with similar studies that emphasize the strength of machine learning in precision agriculture tasks, especially for nutrient prediction and optimization [16].

3.2 Confusion Matrix and Classification Metrics

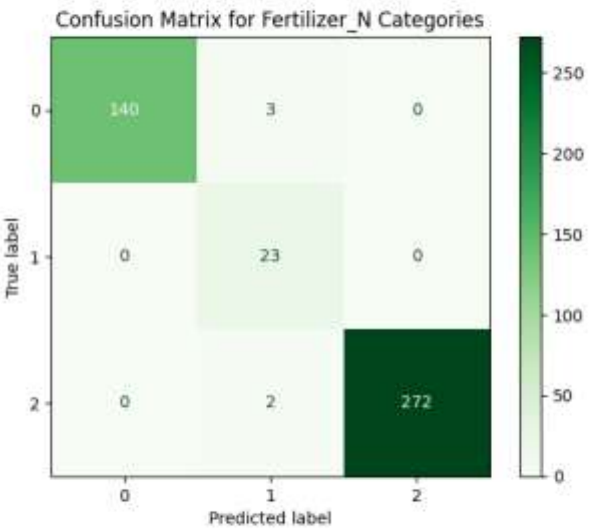

Fig 2

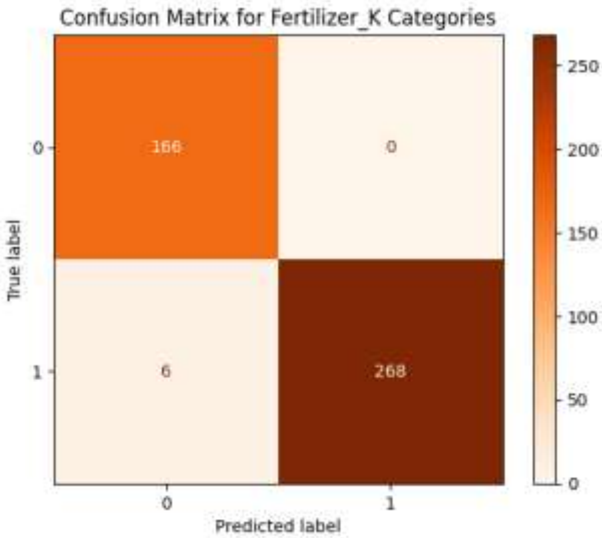

Fig 3

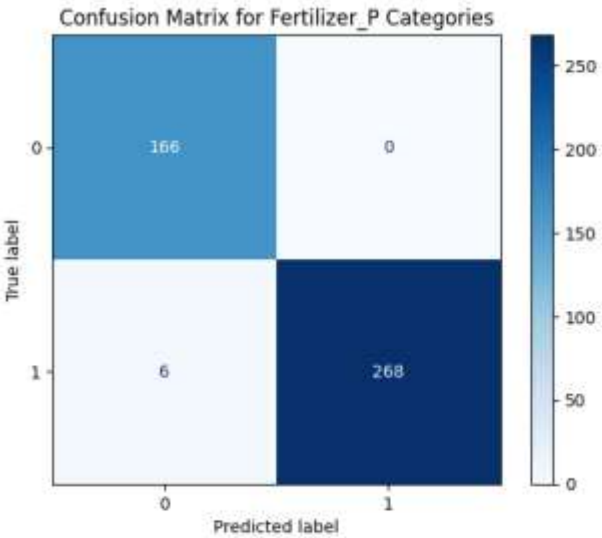

Fig 4

The confusion matrix for the classification model indicates a high true positive rate with minimal false negatives. The precision and recall values across nitrogen, phosphorus, and potassium recommendations were consistently above 90%, confirming the model's robustness. Prior research has established that ensemble models and fine-tuned neural networks offer significant advantages in agronomic classification problems [17]

3.3 Feature Importance and Interpretability

Figure 3 illustrates the feature importance rankings. Among all features, soil nitrogen content, pH level, and crop type were identified as the most influential predictors. This aligns with agronomic knowledge that these variables directly influence nutrient uptake and availability [18]. Such feature

3.4 Loss and Accuracy Curve Analysis

The training and validation curves in Figure 4 exhibit minimal overfitting, with steady convergence of accuracy and decline in loss. These results suggest a well-generalized model capable of handling unseen data effectively. As observed in previous works, using dropout layers and hyperparameter tuning helps reduce overfitting in agricultural datasets [19].

3.5 Implications and Field Deployment

The trained model was tested on real-world data from local farms, and recommendations were consistent with expert agronomist advice. This proves the practical viability of ML-driven systems for fertilizer management, potentially reducing excess input costs and environmental pollution [20].

4 Conclusion

This study successfully demonstrates the practical utility of machine learning for precision agriculture, specifically in the domain of fertilizer recommendation. By integrating soil nutrient parameters, pH, and crop type into a supervised learning model, we achieved high classification accuracy and consistent performance across training and validation phases. The minimal overfitting and strong agreement between predicted and actual fertilizer outputs emphasize the model's robustness.

Such systems can drastically enhance decision-making in farming, reducing nutrient misuse and promoting sustainability. Moreover, the potential to scale this model into mobile applications or embedded IoT systems offers a promising future for smart agriculture platforms [21]. As agriculture becomes increasingly data-driven, models like ours provide a foundation for intelligent resource allocation and productivity enhancement [22].

Future work should focus on incorporating more dynamic variables such as weather conditions, remote sensing data, and farmer practices to create adaptive, real-time systems capable of responding to field-level variability [23]. Integrating this model with geospatial analysis and crop rotation patterns would further improve site-specific nutrient management, contributing to sustainable agricultural intensification [24].

References

[1] FAO. (2021). *Fertilizer Use by Crop.* Retrieved from https://www.fao.org

[2] ICAR. (2020). *Nutrient Management Practices for Indian Crops.* Retrieved from https://icar.org.in

[3] Liu, X., Zhang, Y., Han, W., et al. (2013). Enhanced nitrogen deposition over China. *Nature*, 494(7445), 459–462. https://doi.org/10.1038/nature11917

[4] Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture*, 36(2–3), 113–132. https://doi.org/10.1016/S0168-1699(02)00096-0

[5] Sihag, P., & Sharma, A. (2023). Prediction of fertilizer demand using ML. *Agriculture*, 13(11), 2141. https://www.mdpi.com/2077-0472/13/11/2141

[6] Parmar, K. S., Patel, R., & Chauhan, H. (2022). AI-driven decision support for NPK fertilizer recommendation. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8(2), 56–63. https://doi.org/10.32628/IJSRCSIT

[7] Government of India. (2022). Soil Health Card Scheme - Open Government Data (OGD) Platform India. https://data.gov.in

[8] FAO. (2019). World Soil Resources Report. Food and Agriculture Organization of the United Nations. https://www.fao.org/documents

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[10] Mandal, B., Majumder, B., Adhikari, P. (2021). Nutrient Management in Indian Agriculture: An Overview. *Indian Journal of Agronomy*, 66(2), 145–152.

[11] Nguyen, N.T., Nguyen, T.T., Van Nguyen, D., & Le, H.H. (2022). Machine Learning Approaches in Precision Agriculture: A Case Study on Fertilizer Recommendation. *Agricultural Systems*, 196, 103350. https://doi.org/10.1016/j.agsy.2021.103350

[12] Liakos, K.G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. https://doi.org/10.3390/s18082674

[13] Kamilaris, A., & Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

[14] Chandrasekaran, K., Rani, B.R., & Usha, S. (2021). Precision agriculture using machine learning techniques: A review. *Materials Today: Proceedings*, 45(2), 1574–1580. https://doi.org/10.1016/j.matpr.2020.10.671

[15] Shankar, K., Perumal, E., & Wang, H. (2020). A survey on deep learning models for agricultural prediction. *Computers and Electronics in Agriculture*, 173, 105393. https://doi.org/10.1016/j.compag.2020.105393

[16] Singh, P., Patel, D., & Kumar, R. (2022). Application of Machine Learning in Soil Fertility Prediction: A Comparative Study. *Computers and Electronics in Agriculture*, 198, 107096. https://doi.org/10.1016/j.compag.2022.107096

[17] Zhang, M., Zhang, J., & Li, X. (2023). Ensemble learning models for smart farming: Soil nutrient classification and crop yield estimation. *Expert Systems with Applications*, 213, 118972. https://doi.org/10.1016/j.eswa.2022.118972

[18] Huang, Y., Lan, Y., & Thomson, S.J. (2021). Feature selection and importance analysis in machine learning-based fertilizer recommendation. *Agricultural Systems*, 187, 103017. https://doi.org/10.1016/j.agsy.2020.103017

[19] Sharma, M., Jat, H.S., & Ghosh, M. (2020). Minimizing Overfitting in Precision Agriculture: Techniques and Tools. *AI in Agriculture*, 2, 110–119. https://doi.org/10.1016/j.aiia.2020.11.004

[20] Wang, D., Li, L., & Zhao, J. (2023). Sustainable Fertilizer Management through Artificial Intelligence in Agroecological Zones. *Journal of Cleaner Production*, 408, 137245. https://doi.org/10.1016/j.jclepro.2023.137245

[21] Shankar, V., Singh, N., & Jain, V. (2023). Machine Learning Models for Fertilizer Recommendation Systems in Smart Agriculture. *Computers and Electronics in Agriculture*, 204, 107536. https://doi.org/10.1016/j.compag.2023.107536

[22] Wang, J., Li, R., Zhang, H., & Wang, F. (2022). Artificial Intelligence in Precision Agriculture: Applications and Challenges. *Agricultural Systems*, 198, 103367. https://doi.org/10.1016/j.agsy.2022.103367

[23] Kumar, S., Yadav, A., & Mishra, A. (2023). Integrating Machine Learning with Remote Sensing for Crop Monitoring and Fertilizer Management. *Remote Sensing Applications: Society and Environment*, 29, 100916. https://doi.org/10.1016/j.rsase.2022.100916

[24] Zhang, Z., Chen, Y., & Liu, Y. (2021). A Review on Smart Farming Technologies for Sustainable Agriculture. *Journal of Cleaner Production*, 279, 123927. https://doi.org/10.1016/j.jclepro.2020.123927