WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

FERTILIZER RECOMMENDATION AND SOIL HEALTH ANALYSIS USING EXPLAINABLE AI

Dr. Jai Ruby [a], M. Saranya [b]

[a]**Assistant Professor**, Department of Computer Applications and Research Centre, Sarah Tucker College(Autonomous), Tirunelveli.
[b]**Student, Reg.No: 25081205300112026,** Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Tirunelveli.

ABSTRACT :

Agriculture plays an important role in ensuring food security and supporting economic development. Maintaining soil fertility and selecting suitable fertilizers are essential for sustainable crop production. However, fertilizer requirements vary depending on soil properties, nutrient availability, crop type, growth stage, environmental conditions, and agricultural practices. Improper fertilizer application can cause nutrient imbalances, reduce crop productivity, increase cultivation costs, and affect soil quality. Therefore, an intelligent and interpretable approach is needed to support appropriate fertilizer selection and soil health assessment.

This paper presents an Explainable Artificial Intelligence (XAI)-based approach for fertilizer recommendation and soil health analysis using deep learning techniques. The study uses an agricultural dataset containing 10,000 records, 19 input features, and seven fertilizer categories: Urea, DAP, MOP, NPK, SSP, Compost, and Zinc Sulphate. Artificial Neural Network (ANN), Deep Neural Network (DNN), and TabNet algorithms are considered for fertilizer classification and performance comparison. The approach analyses soil characteristics, nutrient levels, crop information, environmental conditions, and agricultural practices to identify suitable fertilizer categories. In addition, Explainable AI techniques, including SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), are used to interpret model predictions and identify the important features influencing fertilizer recommendations. Soil health assessment is also included to provide information about soil conditions and nutrient status. By combining deep learning-based classification with explainable techniques, this research aims to improve prediction transparency, support informed fertilizer selection, and contribute to sustainable agricultural practices.

Keywords: Fertilizer Recommendation, Deep Learning, Soil Health Analysis, Explainable Artificial Intelligence, Deep Neural Network, SHAP, LIME, Precision Agriculture.

1. INTRODUCTION

Agriculture plays a significant role in food production and economic development. Crop growth and productivity depend on several factors, including soil fertility, nutrient availability, crop type, climatic conditions, and agricultural practices. Fertilizers provide essential nutrients required for plant growth. However, selecting an appropriate fertilizer requires careful consideration of soil characteristics and crop requirements. Improper fertilizer selection may result in nutrient imbalances, reduced crop productivity, and deterioration of soil quality.

Traditional fertilizer selection often depends on general recommendations and previous farming experience. Such approaches may not fully consider the relationships between soil properties, crop growth stages, and environmental conditions. Therefore, analysing multiple agricultural factors is important for making appropriate fertilizer decisions. Machine learning and deep learning techniques can analyse agricultural data, identify patterns, and support fertilizer classification. These techniques provide opportunities to examine different agricultural parameters and generate data-driven recommendations.

Although deep learning models can produce predictions, understanding their decision-making processes can be challenging. Explainable Artificial Intelligence (XAI) techniques, such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), help interpret model predictions by identifying the input features that contribute to individual outputs. These explanations improve prediction transparency and provide a better understanding of the factors influencing fertilizer recommendations.

This research focuses on an explainable approach to fertilizer recommendation and soil health analysis using deep learning techniques. It analyses soil properties, nutrient levels, crop characteristics, environmental conditions, and agricultural practices to identify suitable fertilizer categories. Artificial Neural Network (ANN), Deep Neural Network (DNN), and TabNet models are considered for fertilizer classification. Furthermore, SHAP and LIME techniques are used to explain the factors influencing model predictions. Soil health assessment is also included to provide insights into soil conditions. The main objective of this research is to support transparent, data-driven fertilizer recommendations, improve the understanding of soil conditions, and encourage informed agricultural decision-making.

2. LITERATURE REVIEW

Machine learning and deep learning techniques are increasingly used in agriculture to analyse soil properties, nutrient levels, crop characteristics, and environmental conditions. These techniques help identify patterns in agricultural data and support fertilizer- and crop-related predictions. LeCun, Bengio, and Hinton discussed the fundamentals and applications of deep learning, highlighting its ability to learn complex patterns from data [4]. Pedregosa et al. introduced Scikit-learn, a machine learning library that supports various data analysis and predictive modelling techniques [6].

Arik and Pfister introduced TabNet, a deep learning architecture specifically designed for tabular data. It uses sequential attention to identify relevant features during prediction, making it suitable for structured agricultural datasets [3]. Venkateswara and Padmanaban investigated interpretable deep learning models for independent fertilizer and crop recommendations. Their study considered soil properties, nutrient levels, and environmental factors and incorporated SHAP and TabNet-based interpretability for agricultural recommendations [5].

Lundberg and Lee introduced SHapley Additive exPlanations (SHAP), a framework that explains model predictions by identifying the contribution of individual input features [1]. SHAP helps explain how soil properties, nutrient values, and other agricultural parameters influence prediction outcomes. Ribeiro, Singh, and Guestrin proposed Local Interpretable Model-Agnostic Explanations (LIME), a technique that explains individual predictions by locally approximating the behaviour of a complex model [2]. These techniques provide useful insights into machine learning predictions.

Abekoon et al. applied explainable machine learning techniques, including SHAP and LIME, to predict soil nitrogen, phosphorus, and potassium nutrient content in cabbage cultivation [7]. Their study highlights the importance of interpreting agricultural predictions using soil nutrient information. Yaganteeswarudu et al. examined the use of LIME for crop recommendations and demonstrated how local explanations can improve the transparency of smart farming predictions [8].

Dey, Ferdous, and Ahmed studied machine learning-based agricultural and horticultural crop recommendations using NPK values, soil pH, and climatic variables such as temperature, rainfall, and humidity [9]. Their study emphasises the importance of considering soil nutrients and environmental conditions in agricultural prediction. Xu et al. proposed an interpretable machine learning framework for soil health assessment using SHAP and causal inference, highlighting the role of important soil indicators in evaluating soil health [10].

Based on these studies, the present research considers ANN, DNN, and TabNet algorithms for fertilizer classification using soil, crop, nutrient, and environmental parameters. SHAP and LIME techniques are incorporated to explain model predictions, along with soil health assessment, to support understandable and data-driven fertilizer recommendations.

3. PROPOSED METHODOLOGY

The proposed approach follows an integrated deep learning and Explainable Artificial Intelligence framework for fertilizer recommendation and soil health analysis. The overall workflow includes soil and crop data collection, data preprocessing, feature encoding and scaling, DNN-based fertilizer prediction, confidence score calculation, soil health assessment, SHAP-and LIME-based explanations, fertilizer quantity recommendation, database storage, and dashboard-based monitoring.

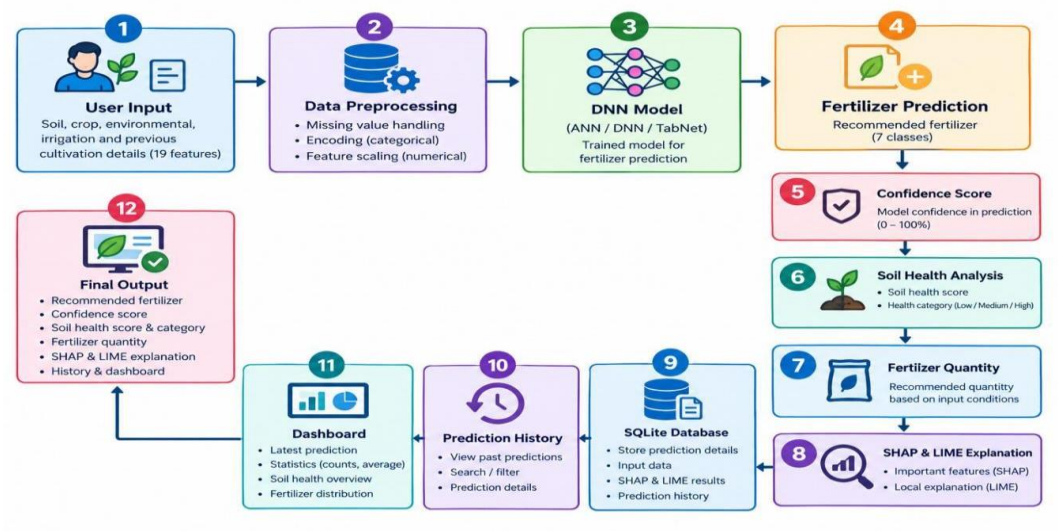


Fig. 3.1. Proposed System Architecture and Workflow

As illustrated in Fig. 3.1, the system processes the input data, predicts the recommended fertilizer, calculates the confidence score, and analyses soil health. SHAP and LIME are used to explain the prediction results. The prediction details are stored in an SQLite database and presented through the dashboard and prediction history.

A. Dataset Collection and Understanding

The fertilizer recommendation dataset is used to train and evaluate the proposed deep learning model. It contains information about soil characteristics, crop types, environmental conditions, and fertilizer usage. These attributes help the model identify patterns associated with different fertilizer categories. The input features include soil type, soil pH, soil moisture, organic carbon, electrical conductivity, nitrogen, phosphorus, potassium, temperature, humidity, rainfall, crop type, growth stage, season, irrigation method, previous crop, region, fertilizer used, and previous yield. The target attribute is Recommended_Fertilizer, which represents the fertilizer category to be predicted.

The study considers seven fertilizer categories: Urea, DAP, MOP, NPK, SSP, Compost, and Zinc Sulphate. Before model development, the dataset is analysed to understand its structure, feature types, and target distribution.

B. Data Preprocessing

Data preprocessing is an essential step in preparing the dataset for deep learning. The dataset is examined for missing values, duplicate records, and inconsistent data. Categorical and numerical attributes are handled according to their respective data types.

Categorical variables, such as soil type, crop type, season, and irrigation method, are converted into numerical representations before being provided to the model. Numerical features are checked for valid values and prepared for model training. These preprocessing steps help maintain data consistency and improve the reliability of the training process.

C. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is performed to understand the characteristics and relationships within the dataset. The distribution of fertilizer classes is examined to identify the frequency of each recommendation category. Numerical features, such as nitrogen, phosphorus, potassium, soil pH, temperature, humidity, and rainfall, are analysed using descriptive statistics and visualisations.

Graphs and correlation heatmaps are used to identify patterns and relationships among the input variables. The findings provide a better understanding of the dataset and support the model training process.

D. Feature Encoding and Scaling

The dataset contains both categorical and numerical features. Categorical attributes are converted into numerical values using an appropriate encoding method. Numerical features may be standardised using a scaling technique to bring their values into comparable ranges.

The processed features are then arranged in the format required by the deep learning model. The same preprocessing transformations are applied during both training and prediction to maintain consistency between the training data and user-provided inputs.

E. Model Training and Comparison

Different deep learning models are evaluated to identify a suitable model for fertilizer classification. The models considered in this study are Artificial Neural Network (ANN), Deep Neural Network (DNN), and TabNet.

The dataset is divided into training and testing sets. The training data is used to learn the relationships between the input features and fertilizer categories, while the testing data is used to evaluate the models on unseen records.

The models are compared using evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics provide a broader understanding of classification performance. The final model is selected based on the evaluation results and its suitability for integration into the web application.

F. DNN Model Architecture

The Deep Neural Network is used as the primary fertilizer classification model. It receives the processed input features and learns the relationships between soil conditions, crop characteristics, environmental factors, and fertilizer categories through multiple layers.

The proposed architecture consists of an input layer with 19 features, followed by hidden layers containing 128, 64, and 32 neurons. ReLU activation is used in the hidden layers, while dropout is applied where configured to reduce overfitting. The output layer contains seven neurons with Softmax activation, representing the seven fertilizer categories.

G. Fertilizer Prediction and Confidence Score

The trained DNN model predicts a suitable fertilizer based on the input values provided by the user. These inputs include soil nutrient levels, soil properties, crop information, and environmental conditions. After preprocessing, the input data is passed to the trained model to obtain prediction probabilities for the fertilizer categories.

The fertilizer category with the highest predicted probability is selected as the model's recommendation. The corresponding probability can be displayed as a confidence score, indicating the model's relative certainty in its prediction.

The confidence score represents the model's output probability and should not be considered a guarantee that the recommendation is agronomically correct.

H. Soil Health Analysis

The study includes a soil health analysis component that provides an assessment based on the soil-related information supplied by the user. Parameters such as soil pH, moisture, organic carbon, electrical conductivity, nitrogen, phosphorus, and potassium provide information about soil conditions.

A soil health score is calculated, and a soil health category is assigned according to the implemented scoring method. This result is displayed along with the fertilizer recommendation, allowing users to examine soil conditions and fertilizer predictions within a single interface.

The scoring rules and category thresholds should be documented according to the actual implementation. The score should not be described as a laboratory-validated soil health measurement unless such validation has been performed.

I. SHAP and LIME Explainability

Explainable Artificial Intelligence is included to help users understand the factors influencing fertilizer predictions. SHAP and LIME are used to provide explanations for the model outputs.

SHAP (SHapley Additive exPlanations) estimates the contribution of individual features to a prediction. Depending on the explanation method and output being examined, it can help identify which input variables have a positive or negative influence on the predicted class.

LIME (Local Interpretable Model-Agnostic Explanations) explains an individual prediction by approximating the model locally using a simpler, interpretable model. It highlights the features that contribute to the prediction for a particular input record.

The explanations are presented alongside the fertilizer recommendation, providing additional insight into the model's decision-making process. These techniques improve interpretability but do not independently establish that a recommendation is scientifically correct.

J. SQLite Database Integration

SQLite is used to store fertilizer prediction records. The database contains information such as the date and time of prediction, recommended fertilizer, confidence score, soil health score, input data, and explainability results.

The input data and explanation results are stored in JSON format where appropriate. Parameterised SQL queries are used to insert and retrieve records. Database integration ensures that prediction results remain available even after a prediction request has been completed.

K. Prediction History

The prediction history module allows users to view previously generated fertilizer recommendations. Each record contains the prediction date, recommended fertilizer, confidence information, and soil health details.

Users can select an individual record to view its input data and stored explanation results. The application also provides options to delete individual prediction records or clear all prediction records, wherever these actions are enabled.

L. Dashboard

The dashboard provides an overview of prediction activity and the latest available results. It displays the total number of predictions, average confidence, average soil health score, most frequently recommended fertilizer, and fertilizer distribution.

It also presents information from the latest prediction, including crop details, soil parameters, recommended fertilizer, and available SHAP and LIME explanations. Charts and summary tables help users understand the distribution of fertilizer recommendations and review prediction activity.

M. Complete System Integration

All components are integrated into a single workflow. The dataset is prepared and used to train the deep learning model. The trained model is connected to the Flask backend, which accepts user inputs and generates fertilizer predictions. The soil health module calculates the configured score, while SHAP and LIME provide explanations for the model outputs.

SQLite stores the prediction records, and the dashboard presents summary statistics and prediction information. This integration brings together model development, prediction, explainability, data storage, and visualisation within a single web-based system.

4. EVALUATION METRICS

The proposed approach uses three deep learning models to recommend suitable fertilizers based on soil properties, crop details, and environmental conditions. The selected models are Artificial Neural Network (ANN), Deep Neural Network (DNN), and TabNet. Their performance is evaluated using Accuracy, Precision, Recall, and F1-score. Among the evaluated models, DNN achieved the highest reported accuracy and was selected as the final fertilizer classification model.

4.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a deep learning technique inspired by the structure of the human brain. It consists of interconnected neurons that learn patterns from input data. In this study, ANN is trained using 19 input features to classify the appropriate fertilizer category.

4.2 Deep Neural Network (DNN)

Deep Neural Network (DNN) is a neural network that contains multiple hidden layers to learn complex relationships between input features. In this study, DNN uses soil properties, nutrient values, crop information, and environmental conditions to predict the recommended fertilizer. It achieved the highest reported accuracy of 86.27% and was selected as the final model.

4.3 TabNet

TabNet is a deep learning model designed for tabular data. It uses attention-based feature selection to identify important input features during prediction. In this study, TabNet is trained using the same 19 input features to classify fertilizer categories. Its performance is compared with ANN and DNN using the selected evaluation metrics.

4.4 Evaluation Measures

Accuracy: Accuracy measures the proportion of correctly classified samples among all test samples.

Accuracy = (Correct Predictions / Total Predictions) × 100

Precision: Precision measures the proportion of correct predictions among all predictions made for a particular class.

Precision = TP / (TP + FP)

Recall: Recall measures the proportion of actual samples correctly identified by the model.

Recall = TP / (TP + FN)

F1-score: F1-score represents the harmonic mean of precision and recall.

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

Here, TP represents True Positive, FP represents False Positive, and FN represents False Negative. For multiclass classification, these measures are calculated using the averaging method specified in the implementation.

5. RESULTS AND DISCUSSION

The ANN, DNN, and TabNet models were evaluated for fertilizer recommendation using 19 soil, crop, and environmental input features. Their performance was compared using Accuracy, Precision, Recall, and F1-score. The reported results are presented in Table 5.1

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ANN	85.30%	85.00%	85.30%	85.10%
DNN	**86.27%**	86.40%	**86.27%**	86.26%
TabNet	86.20%	**91.05%**	86.20%	**87.48%**

Table 5.1. Comparative Performance Analysis of Deep Learning Models

The results indicate that DNN achieved the highest accuracy of 86.27%, followed by TabNet at 86.20% and ANN at 85.30%. Based on overall accuracy, DNN was selected as the final model. However, TabNet achieved the highest reported precision of 91.05% and F1-score of 87.48%. These findings indicate that the three models show different levels of performance across the evaluation metrics.

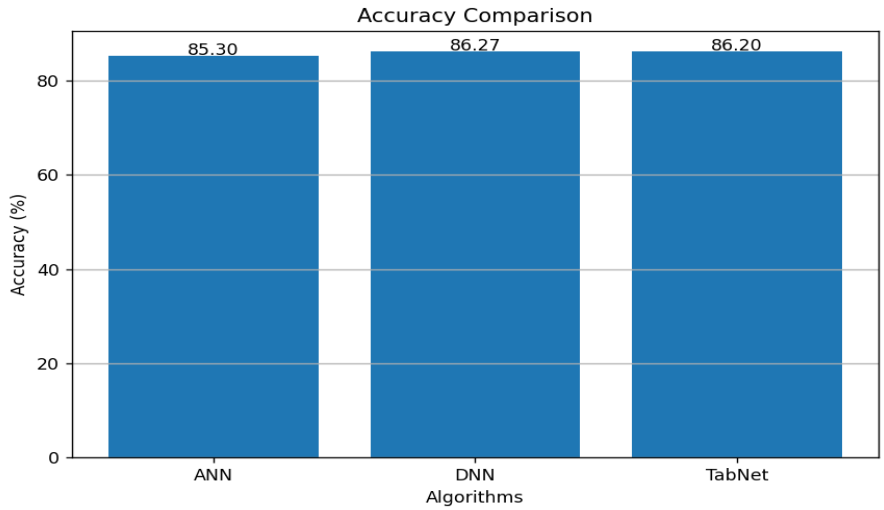


Fig. 5.2. Comparison of Accuracy of ANN, DNN, and TabNet Models.

Figure 5.2 presents the accuracy comparison of the three models, with DNN achieved the highest accuracy, followed by TabNet and ANN.

6. CONCLUSION

This paper presents a deep learning-based approach for fertilizer recommendation and soil health analysis using Explainable Artificial Intelligence. The study uses 19 agricultural input features to predict suitable fertilizer categories. ANN, DNN, and TabNet models are compared, with DNN achieving the highest accuracy of 86.27%. SHAP and LIME techniques are used to explain the factors influencing fertilizer predictions. The study also includes soil health assessment and confidence score calculation. Overall, this research supports informed fertilizer selection and soil management through deep learning and explainable AI techniques. In the future, larger datasets, real-time soil sensor data, and field experiments can be considered to improve prediction reliability and evaluate practical performance.

7. REFERENCES

1. Lundberg, S. M., and Lee, S. I. (2017). "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, 30.
2. Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "'Why Should I Trust You?' Explaining the Predictions of Any Classifier." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144.
3. Arik, S. Ö., and Pfister, T. (2021). "TabNet: Attentive Interpretable Tabular Learning." *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), pp. 6679–6687.
4. LeCun, Y., Bengio, Y., and Hinton, G. (2015). "Deep Learning." *Nature*, 521, pp. 436–444.
5. Venkateswara, S. M., and Padmanaban, J. (2025). "Interpretable Deep Learning Models for Independent Fertilizer and Crop Recommendation." *Scientific Reports*, 15, Article 41721. https://doi.org/10.1038/s41598-025-26910-4
6. Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, 12, pp. 2825–2830.
7. Abekoon, T., Sajindra, H., Rathnayake, N., Ekanayake, I. U., Jayakody, A., and Rathnayake, U. (2025). "A Novel Application with Explainable Machine Learning (SHAP and LIME) to Predict Soil N, P, and K Nutrient Content in Cabbage Cultivation." *Smart Agricultural Technology*, 11, Article 100879. https://doi.org/10.1016/j.atech.2025.100879
8. Yaganteeswarudu, A., Biswas, S. K., Aruna, V., and Tripathi, D. (2025). "Enhancing Transparency in Smart Farming: Local Explanations for Crop Recommendations Using LIME." *Procedia Computer Science*, 258, pp. 1993–2005. https://doi.org/10.1016/j.procs.2025.04.450
9. Dey, B., Ferdous, J., and Ahmed, R. (2024). "Machine Learning Based Recommendation of Agricultural and Horticultural Crop Farming in India Under the Regime of NPK, Soil pH and Three Climatic Variables." *Heliyon*, 10(3), Article e25112. https://doi.org/10.1016/j.heliyon.2024.e25112
10. Xu, X., Liu, Q., Liu, Y., Li, Y., Chen, Y., Lei, T., Kuzyakov, Y., Zhang, W., Chen, J., and Ge, T. (2025). "Novel Soil Health Assessment Framework for Legume-Based Rotation Farmland by Interpretable Machine Learning with Causal Inference." *Computers and Electronics in Agriculture*, 239, Part B, Article 111011. https://doi.org/10.1016/j.compag.2025.111011