iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81143

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A Survey on Explainable IoT Framework for Soil Fertility Prediction using LightGBM and CatBoost

Sheela Rani C M[1], Srilakshmi Seshadri[2], Prabhudeva B N[3], Shreya Praveen Ramadurg[4], Sharath M R[5]
[1]Assistant Professor, Dept. of CSE Sapthagiri College of Engineering
2, 3, 4, 5Dept. of CSE Sapthagiri College of Engineering

Abstract: Soil fertility has a direct impact on crop growth and overall agricultural productivity. Understanding soil conditions helps farmers make better decisions regarding irrigation, fertilizer usage, and crop selection. However, traditional soil testing methods mainly depend on manual sampling and laboratory analysis, which are time-consuming and not suitable for frequent monitoring.

In recent years, IoT-based solutions have made it possible to continuously observe soil conditions using sensors. These sensors capture parameters such as pH, moisture, and temperature in real time. Along with this, machine learning techniques are used to analyze the collected data and identify patterns that help in predicting soil fertility more effectively. In this survey, attention is given to boosting models such as LightGBM and CatBoost, which are known for handling structured data efficiently.

The study also highlights the role of explainable AI techniques like SHAP and LIME, which help in understanding how predictions are generated. Various recent approaches are reviewed, and their strengths and limitations are discussed. Overall, the work focuses on developing systems that are accurate, interpretable, and suitable for real-time agricultural applications.

I. INTRODUCTION

Soil fertility plays an important role in agriculture as it directly affects plant growth and crop yield. Several factors such as nutrient content, moisture, temperature, and pH influence soil quality. Since these factors vary across regions and over time, evaluating soil conditions accurately becomes a challenging task.

Conventional methods of soil testing usually involve col-lecting samples manually and analyzing them in laboratories. Although these methods provide reliable results, they are time-consuming and not suitable for continuous monitoring. In many cases, delays in obtaining results can affect timely decision-making, especially when soil conditions change rapidly.

With the introduction of IoT, it has become possible to monitor soil conditions in real time using sensors placed in agricultural fields. These sensors collect important data, which can be used to make better decisions related to irrigation and fertilizer usage. This reduces dependency on periodic testing and improves efficiency in farming practices.

Machine learning techniques are widely used to analyze soil data and make predictions. Models such as Random Forest, Support Vector Machines, LightGBM, and CatBoost are capable of identifying complex relationships between soil parameters.

These models help in predicting soil fertility more accurately when trained with proper data.

However, many machine learning models lack transparency, as they do not explain how predictions are made. This reduces trust among users. To overcome this issue, explainable AI techniques such as SHAP and LIME are used to provide insights into model behavior. This survey focuses on combining IoT, machine learning, and explainable AI to develop efficient and practical soil fertility prediction systems.

II. BACKGROUND AND MOTIVATION

A. Need for Intelligent Soil Monitoring

Agriculture today relies more on data-driven approaches to improve productivity. Monitoring soil conditions continuously helps in taking quick actions when changes occur. Since soil properties and climate conditions keep changing, traditional periodic testing is often not sufficient. This creates a need for systems that can provide real-time insights using sensing technologies.

B. Limitations of Conventional Soil Analysis

Traditional soil testing methods require manual sampling and laboratory processing. These steps take time and increase cost, making them unsuitable for frequent use. In addition, such methods do not support real-time monitoring, which is important for effective decision-making in agriculture.

C. Role of Explainable AI in Agriculture

Although machine learning models can provide accurate predictions, they often do not explain how results are obtained. This makes it difficult for users to trust the system. Explainable AI techniques help solve this issue by showing how different factors influence the prediction, making the system easier to understand and use.

III. LITERATURE SURVEY

A. IoT and AI-based Soil Nutrient Prediction Systems

Author and Year: Uwadia O. A., Dahunsi F. M. (2026)

Methodology: This study proposes an integrated framework that combines IoT-based sensing with machine learning for soil nutrient prediction. Sensors deployed in agricultural fields continuously measure parameters such as pH, moisture, temperature, and electrical conductivity. The collected data is transmitted to a cloud platform where preprocessing and feature extraction are performed. Machine learning models, particularly ensemble techniques, are then applied to analyze the data and predict soil fertility levels. The system enables continuous monitoring and supports data-driven agricultural decisions.

Limitation: The reliance on multiple sensors increases implementation cost and maintenance requirements. In addition, the system lacks sufficient interpretability, making it difficult for users to understand the reasoning behind predictions. Sensor calibration and environmental variations may also affect accuracy.

B. Real-Time Soil Fertility Monitoring

Author and Year: Priyanshu Tiwari et al. (2026)

Methodology: This work focuses on real-time soil fertility prediction using IoT sensors integrated with deep learning techniques. Soil parameters such as pH, temperature, moisture, and nutrient content are collected and used as input to an Artificial Neural Network. The model learns relationships among these features and predicts soil fertility dynamically. Continuous data updates allow the system to adapt to changing environmental conditions, enabling timely monitoring and decision-making.

Limitation: The approach requires high computational resources for training and deployment. Additionally, the model lacks interpretability, making it difficult for end users to understand predictions. Dependence on continuous data flow may also affect performance in unstable network conditions.

C. Soil Fertility Analysis using ML and DL

Author and Year: Kanimozhi Gunasekaran et al. (2025)

Methodology: In this study, a sensor-based system is developed to collect real-time soil and environmental data, including pH, NPK values, humidity, and temperature. After preprocessing, multiple machine learning and deep learning models such as Random Forest, Extra Trees, MLP, and LSTM are trained and evaluated. Performance is measured using metrics like accuracy, precision, recall, and F1-score. The system not only predicts soil fertility but also recommends suitable crops based on soil conditions.

Limitation: The system requires extensive hardware infras-tructure, which increases cost and complexity. It also lacks explainable AI techniques, reducing transparency. Scalability and deployment in large agricultural areas remain challenging.

D. IoT-enabled AI Crop Prediction

Author and Year: M.D.S. Sharafat et al. (2025)

Methodology: This study integrates IoT-based soil data with external weather information obtained through APIs. Multiple machine learning algorithms such as Random Forest and Gradient Boosting are combined using ensemble techniques to improve prediction accuracy. The system is deployed on edge devices like Raspberry Pi to enable faster processing and reduce latency. Explainable AI methods, particularly LIME, are used to interpret model predictions.

Limitation: The system partially depends on pre-existing datasets instead of fully utilizing real-time data. Integration complexity increases due to multiple data sources. Network dependency and hardware constraints may also affect system performance.

E. Explainable AI for Soil Fertility

Author and Year: Remya Praveen et al. (2025)

Methodology: This work emphasizes the importance of explainable AI in soil fertility prediction. Machine learning models such as XGBoost are used to analyze soil properties including organic carbon, nitrogen, phosphorus, potassium, and pH. Explainability techniques such as SHAP and LIME are applied to determine feature importance and provide clear insights into the prediction

process. This improves transparency and supports better decision-making.
Limitation: The study is limited to static datasets and does not incorporate real-time IoT-based data collection. It also does not address scalability or deployment challenges in practical agricultural environments.

F. Comparative ML Study

Author and Year: Walid Abdullah (2025)
Methodology: This study presents a comparative analysis of various machine learning algorithms for soil fertility prediction, including Random Forest, SVM, LightGBM, XGBoost, KNN, and MLP. Data preprocessing techniques such as normalization, feature selection, and class balancing are applied before training. Each model is evaluated using standard performance metrics. The results show that boosting algorithms perform better due to their ability to capture complex relationships in data.
Limitation: The work focuses mainly on performance comparison and does not consider real-time implementation or IoT integration. It also lacks explainability, limiting its practical usefulness.

G. STGNN Soil Prediction

Author and Year: C. P. Thamil Selvi et al. (2025)
Methodology: This study introduces a spatiotemporal graph neural network model for soil prediction. Soil data is repre-sented as a graph structure where nodes and edges capture relationships between parameters. The model processes both spatial and temporal dependencies simultaneously to improve prediction accuracy. Optimization techniques are used to enhance convergence and performance.
Limitation: The model is computationally intensive and requires high processing power. Its complexity makes it difficult to deploy in real-time agricultural applications with limited resources.

H. LightGBM vs CatBoost

Author and Year: A. Sri Lakshmi et al. (2024)
Methodology: This study checks gradient boosting algo-rithms such as LightGBM and CatBoost for soil fertility prediction. The dataset goes under preprocessing steps includ-ing normalization and handling missing values. LightGBM is selected for its speed and efficiency, while CatBoost is effective in handling categorical features. The results indicate that boosting algorithms outperform traditional machine learning approaches.
Limitation: The study does not include real-time data collection using IoT devices. It also lacks explainable AI techniques, limiting interpretability and practical applicability..

IV. METHODOLOGY

A. Soil Data Acquisition Module

[2] In this module, soil parameters are obtained directly from agricultural fields using sensing devices. Low-cost sensors such as pH, moisture, and temperature sensors are put into the soil to capture real-time environmental conditions. These sensors continuously monitor variations in soil properties caused by climatic and irrigation changes. The sensors are interfaced with a microcontroller, which reads the raw signals and converts them into digital values for further processing. This approach eliminates the dependency on traditional laboratory testing methods, which are often time-consuming and expensive. The collected data reflects the actual field conditions, making the system more practical and reliable. Proper calibration of sensors is necessary to ensure accurate readings. The sampling process is performed at regular intervals to capture dynamic changes in soil behavior. This module serves as the primary input source for the system.

B. Data Transmission Module

[3] After taking the soil data, it is sent to the processing system using wireless communication. A Wi-Fi-enabled mi-crocontroller such as ESP32 is used to establish connectivity with the network. The collected sensor readings are formatted into structured data packets and sent to the application through API-based communication. This enables smooth integration between the hardware and software components of the system. Real-time transmission ensures that the system processes current soil conditions without delay. The use of wireless communication improves flexibility and allows deployment in remote agricultural areas. This module supports scalability, making it possible to extend the system across multiple locations.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

C. Data Preprocessing Module

[1] The transmitted data is processed to improve its quality before being used for prediction. Sensor readings may contain noise, inconsistencies, or missing values due to environmental disturbances. In this stage, such issues are addressed using appropriate preprocessing techniques. Outliers are identified and removed to avoid incorrect predictions. Missing values are handled using estimation or interpolation methods. Normaliza-tion is applied to bring all features into a comparable range, preventing bias toward any specific parameter. Data formatting is performed to ensure compatibility with machine learning models.

D. Feature Selection Module

Feature selection is done to identify the most significant parameters influencing soil fertility[4]. Among the taken data, certain features such as pH, moisture, and temperature play a important role in determining soil quality. Selecting only relevant features reduces computational complexity and improves system efficiency. This step helps in eliminating duplicate unnecessary information. It also reduces the risk of putting two many values in machine learning models. By focusing on main attributes, the model can learn meaningful relationships more effectively.

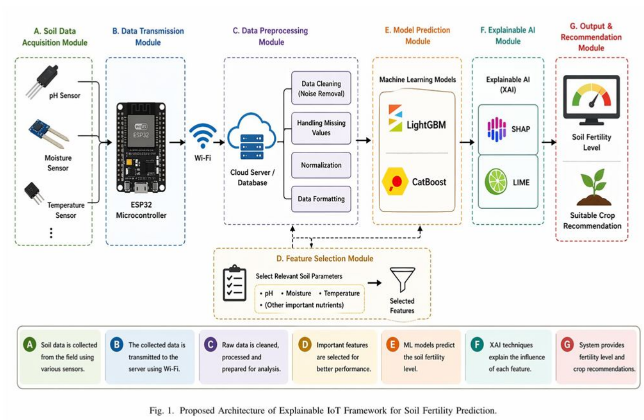


Fig. 1. Proposed System Architecture for Soil Fertility Prediction

E. Model Prediction Module

In this module, the processed data is used for predicting soil fertility using machine learning algorithms. LightGBM and CatBoost are employed due to their ability to handle structured data efficiently. These boosting algorithms combine multiple weak learners to improve prediction accuracy. They are capable of capturing complex and nonlinear relationships between soil parameters. The models are trained using prepared datasets and then applied to real-time inputs. Their computational efficiency makes them suitable for real-time applications. The prediction output is generated based on learned patterns from the data.

F. Explainable AI Module

[5] To ensure transparency, explainable AI techniques are integrated into the system. Methods such as SHAP and LIME are used to interpret the predictions generated by the models. These techniques provide insights into how each input feature contributes to the final result. This helps in understanding the importance of different soil parameters. Instead of acting as a black-box system, the model provides clear reasoning behind its predictions. This improves user trust and makes the system more practical for real-world applications.

G. Output and Recommendation Module

[8] The last module gives the output based on the predicted results. The system displays the soil fertility level in an understanding format. Along with the prediction, important influencing factors are also given to the user. on basis of the soil condition, suitable crop recommendations are provided to support better decision-making.which shows us which crops we can grow,the output can be accessed through a user interface or an API. This module completes the workflow by converting predictions into meaningful insights.

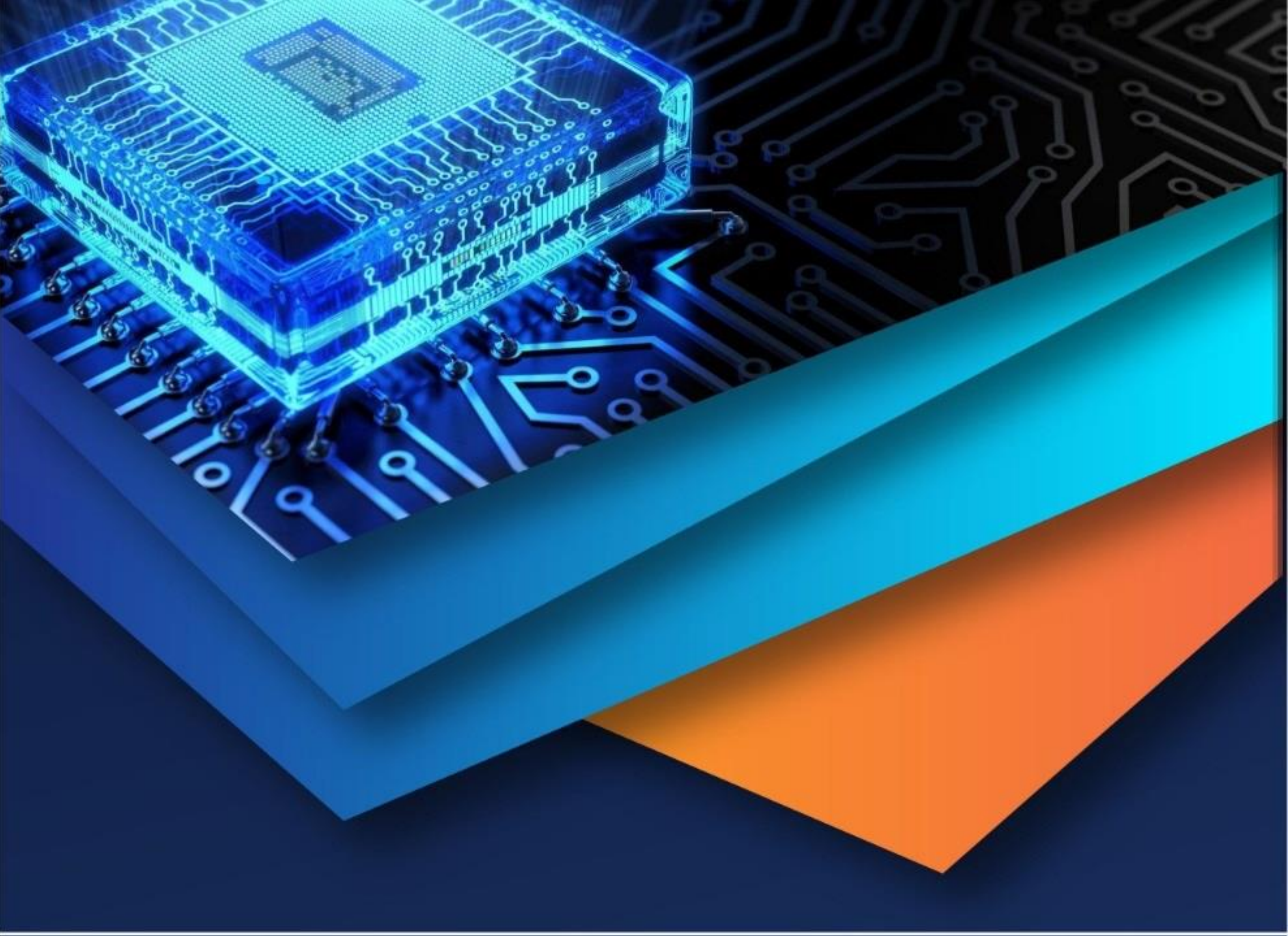
V. CONCLUSION

This survey presents a sensor-based and explainable approach for soil fertility prediction by integrating IoT, machine learning, and interpretability techniques. The system utilizes real-time soil parameters such as pH, moisture, and temperature to generate meaningful insights without relying on traditional laboratory methods. Machine learning models like LightGBM and CatBoost are used to capture relationships between soil attributes and fertility levels, enabling accurate predictions. Explainable AI techniques such as SHAP and LIME further en-hance transparency by showing how each parameter influences the results.

The proposed framework improves the speed and accessibil-ity of soil analysis while supporting better decision-making in agriculture. By combining prediction with interpretability, the system provides a practical and scalable solution for sustainable farming applications.

REFERENCES

[1] Uwadia O. A., Dahunsi F. M. (2026): IoT and Machine Learning-Based Soil Nutrient Prediction System
[2] Priyanshu Tiwari et al. (2026): Real-Time Soil Fertility Monitoring Using IoT and Deep Learning Techniques
[3] Kanimozhi Gunasekaran et al. (2025): Soil Fertility Analysis Using Machine Learning and Deep Learning Models
[4] M. D. S. Sharafat et al. (2025): IoT-Enabled AI-Based Crop Prediction System with Explainability
[5] Remya Praveen et al. (2025): Explainable AI-Based Soil Fertility Prediction Using XGBoost
[6] Walid Abdullah (2025): Comparative Analysis of Machine Learning Algorithms for Soil Fertility Prediction
[7] C. P. Thamil Selvi et al. (2025): Spatiotemporal Graph Neural Network for Soil Prediction
[8] A. Sri Lakshmi et al. (2024): Performance Comparison of LightGBM and CatBoost for Soil Fertility Prediction



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY