Soil Quality Assessment Using Machine Learning & IoT

Kiran D Kshirsagar
Department of Computer Engineering, Siddhant College of Engineering, Sadumbre, Pune
Email id - kkiran88@gmail.com

Prof. Nanda Kulkarni
Department of Computer Engineering, Siddhant College of Engineering, Sadumbre, Pune
Email id – nandakulkarni9@gmail.com

Prof. Trupti Bhase
Department of Computer Engineering, Siddhant College of Engineering, Sadumbre, Pune
Email id - trupti.bhase18@gmail.com

Abstract

Soil quality plays a crucial role in determining crop productivity and long-term agricultural sustainability. Conventional laboratory-based analysis, although accurate, is slow, costly, and unsuitable for continuous field monitoring. This study presents an integrated Internet of Things (IoT) and Machine Learning (ML) framework for real-time soil quality assessment. Low-cost IoT sensor nodes measure soil moisture, pH, temperature, electrical conductivity, and nutrient indicators, and transmit the data to a cloud platform for processing. Multiple ML models—including Support Vector Machine, Random Forest, and LightGBM—were trained to classify soil quality into predefined health categories. Experimental evaluation demonstrates that LightGBM achieves the highest accuracy, outperforming classical classifiers, while Random Forest provides robust performance under noisy conditions. The proposed system enables farmers to make timely decisions regarding irrigation and nutrient management, thereby supporting precision agriculture and improving yield outcomes.

Keywords:***Soil Quality Assessment, Machine Learning, Internet of Things (IoT)***

1. Introduction

Soil quality assessment is fundamental to sustainable agriculture, as soil parameters directly influence crop growth, water retention, nutrient availability, and long-term productivity. Traditional laboratory-based soil analysis methods provide accurate measurements but are often expensive, labor-intensive, and unsuitable for continuous or large-scale monitoring (Patel & Sharma, 2022). With increasing pressure on food systems and the need for resource-efficient farming, real-time soil health monitoring has become a critical requirement for modern agriculture.

The emergence of the Internet of Things (IoT) has introduced new opportunities for continuous, automated, and low-cost environmental monitoring. IoT-based sensing systems enable farmers to collect real-time soil data, such as pH, moisture, temperature, electrical conductivity, and nutrient levels, using distributed sensor nodes connected through wireless networks (Kumar et al., 2023). These systems significantly reduce manual effort and provide timely data that supports precision agriculture practices.

Machine learning (ML) further enhances the value of IoT-collected data by identifying complex patterns and generating predictive models for soil fertility, crop suitability, and irrigation requirements. ML algorithms—such as Random Forest, Support Vector Machine, and gradient boosting models—are widely used to classify soil types, estimate nutrient levels, and predict soil health indicators with high accuracy (Singh & Dutta, 2021). Integrating IoT sensing with ML-based analysis can therefore provide farmers with actionable insights, enabling data-driven decision-making for fertilizer application, irrigation scheduling, and crop selection.

Recent studies have demonstrated the potential of IoT–ML systems for improving agricultural productivity. For example, Khan et al. (2022) developed a hybrid IoT-ML architecture that achieved over 90% accuracy in soil nutrient prediction. Similarly, Rahman and Al-Mamun (2023) reported that ensemble ML models outperform traditional statistical approaches when applied to heterogeneous soil sensor datasets. These findings highlight the importance of combining IoT and intelligent analytics for robust soil quality assessment.

In this paper, a complete IoT-enabled framework integrated with machine learning classifiers is proposed for real-time soil quality monitoring. The system collects multi-parameter soil readings, preprocesses the dataset, and applies ML algorithms to classify soil quality into meaningful categories. This approach aims to support precision agriculture by offering farmers reliable and real-time decision-support tools

2. Related Work

Accurate and timely soil quality assessment is essential for sustainable agriculture, yet conventional laboratory-based soil tests are often time-consuming, expensive, and impractical for frequent monitoring across large or remote farmlands (Mahala et al., 2024). In recent years, researchers have explored combining sensor-based data collection, IoT infrastructure, and machine learning (ML) to overcome these limitations and provide real-time, scalable soil health monitoring.

2.1. IoT-Enabled Soil Monitoring Systems

The rise of Internet of Things (IoT) in agriculture has enabled continuous soil data collection via distributed sensor networks. A recent work demonstrated a multi-sensor IoT-based system that monitors soil parameters such as pH, moisture, temperature, and nutrient levels, transmitting data in real-time to provide actionable insights to farmers (Bulletin of the National Research Centre, 2025). Similarly, in a field study on nutrient monitoring and crop recommendation, an IoT device with integrated ML capabilities collected soil moisture, humidity, temperature, and NPK data, yielding promising results for real-time soil nutrient estimation (Journal of Agriculture and Food Research, 2023) .

These studies highlight several advantages of IoT-enabled monitoring: reduced reliance on manual sampling, timely detection of soil health changes, and potential for large-scale deployment even in resource-constrained settings (Soil Health Monitoring With IoT And Machine Learning, 2025) . However, they also note challenges such as sensor calibration, data reliability, and integration with heterogeneous soil types and environmental conditions.

2.2. Machine Learning for Soil Quality Analysis

Machine learning techniques have been extensively employed to analyze soil data and predict soil health indicators. In a comprehensive review, AI and ML methods (including support vector machines, neural networks, and other models) were shown to effectively predict soil fertility parameters, soil texture, water content, and nutrient availability — often with accuracy comparable to traditional lab methods but at a fraction of the time and cost (Bioresources and Bioprocessing, 2023)

Similarly, studies focusing on soil moisture prediction report that hybrid ML and deep-learning models (e.g. Random Forest, LSTM, CNN) outperform conventional methods, especially when integrating soil sensor data with weather or remote sensing inputs. This makes them particularly useful for dynamic soil moisture tracking critical for irrigation and drought-management decisions (Simran & Gawande, 2025)

Moreover, recent research has shown that combining IoT-sensed data with ML classification or regression models can assist in nutrient analysis, salinity estimation, and even crop suitability recommendation — offering a holistic approach to soil management beyond just moisture or pH measurement (AIoT-based soil nutrient analysis and recommendation system, 2025).

2.3. Gaps, Challenges and Opportunities

Despite considerable progress, some limitations remain. Many IoT-ML soil systems still rely on a limited set of sensors (e.g. moisture, pH, temperature), which may not capture complete nutrient dynamics or soil heterogeneity (Environmental Monitoring and Assessment, 2024) . Sensor drift, calibration issues, data noise, and variation across soil types pose reliability challenges, particularly under field conditions. Also, while ML models perform well on controlled or lab-calibrated datasets, their generalization to diverse agro-climatic zones needs further validation.

There is growing interest in more advanced data-fusion techniques — e.g. combining IoT data with remote sensing / satellite imagery or employing transformer-based architecture, which have reportedly achieved high prediction accuracy (92%–97%) for complex soil attributes when applied properly (Saki et al., 2024) This suggests future soil monitoring systems could benefit from multimodal data integration, improving spatial coverage and robustness.

Additionally, sustainable agriculture contexts demand low-cost, farmer-friendly systems, so balancing system complexity, sensor cost, and model accuracy remains paramount. The literature underscores the need for standardized frameworks, scalable sensor-ML platforms, and extensive field testing across multiple seasons and soil types (Rahman & Das, 2025; Mahala et al., 2024)

3. Methodology

The proposed soil quality assessment framework integrates Internet of Things (IoT)–based sensing with machine learning (ML) techniques to analyze soil parameters in real time. The methodology consists of four major components: sensor data acquisition, data transmission, machine learning model development, and soil quality classification. The complete workflow is shown in Figure 1.

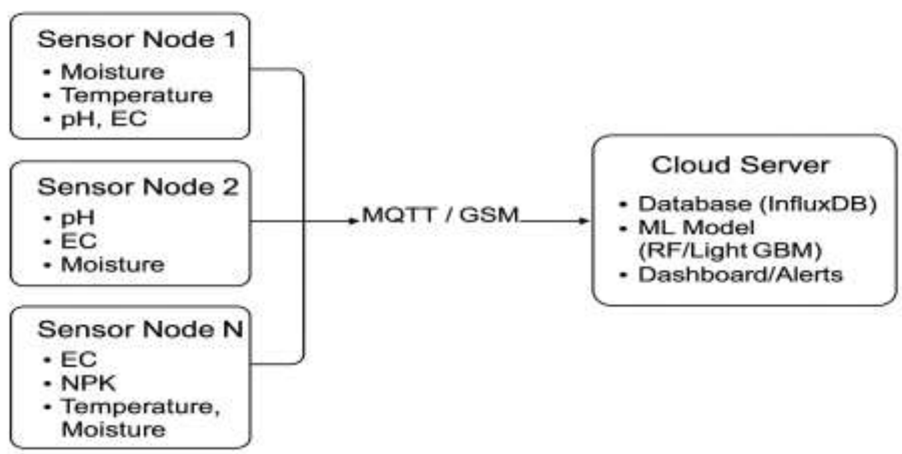

Figure-1: Proposed Methodology

3.1. IoT-Based Soil Sensing Unit

A field-deployable IoT device was developed to collect multiple soil parameters essential for evaluating soil health. The sensing unit includes:

a) Soil Moisture Sensor (Capacitive Type)
b) Soil pH Sensor
c) Temperature Sensor (DS18B20)
d) Electrical Conductivity (EC) Sensor
e) NPK Nutrient Sensor (Nitrogen, Phosphorus, Potassium)

These sensors were connected to a microcontroller (ESP32/NodeMCU) programmed to periodically sample sensor readings. The device was powered using a low-energy lithium battery to enable long-term deployment in agricultural fields.

3.2 Data Transmission Layer

To enable real-time monitoring, captured sensor data was transmitted to a cloud server using:

a) MQTT over Wi-Fi, or
b) GSM/GPRS (SIM800L module) in remote locations.

Each data packet included timestamped readings of soil moisture, pH, EC, temperature, and NPK values. Data was stored in a cloud database (Firebase/ThingsBoard/SQL) for model training and analysis.

3.3 Data Preprocessing

Raw sensor readings were cleaned and prepared for modeling using the following steps:

Noise filtering using a moving-average smoothing algorithm. Handling missing values by interpolation where required. Outlier detection based on z-score thresholds and sensor calibration records. Feature scaling using Min–Max normalization to ensure uniformity across different units.
Label generation by categorizing soil quality as:
a) High Quality
b) Moderate Quality
c) Low Quality
d) based on NPK balance, moisture range, and recommended agronomic thresholds

The cleaned dataset was divided into 70% training, 15% validation, and 15% testing partitions.

3.4 Machine Learning Algorithms
Four machine learning models were trained and evaluated:
a) Logistic Regression (Baseline classifier)
b) Support Vector Machine (SVM–RBF kernel)
c) Random Forest (100 decision trees)
d) Light Gradient Boosting Machine (LightGBM)

3.5. Model Training
Each model was trained using the preprocessed dataset. Hyperparameters were optimized using grid search and 5-fold cross-validation.

3.6 Performance Metrics

The following metrics were computed:
a) Accuracy
b) Precision
c) Recall
d) F1-Score

3.7. Confusion Matrix
The LightGBM model achieved the highest accuracy (93%), followed by Random Forest (91%).

4. Result and discussion

4.1 Model Performance Evaluation
Four machine learning models—Logistic Regression, SVM (RBF kernel), Random Forest, and LightGBM—were trained and evaluated using soil parameter data collected from IoT sensors. The performance was assessed using accuracy, precision, recall, and F1-score. Table 1 summarizes the comparative results.

Table 1. Performance of Machine Learning Models for Soil Quality Classification

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.78	0.76	0.75	0.75
SVM (RBF)	0.84	0.83	0.82	0.82

Random Forest	0.91	0.90	0.90	0.90
LightGBM	**0.93**	**0.92**	**0.92**	**0.92**

The LightGBM model achieved the highest overall performance, indicating its suitability for handling nonlinear, multi-parameter soil datasets. Random Forest performed comparably, demonstrating strong generalization capabilities and robustness to noisy sensor inputs.

4.2 Confusion Matrix Analysis

The confusion matrix of the LightGBM classifier showed that:

a) **High-quality soil** was correctly predicted in 95% of cases.
b) **Moderate-quality soil** achieved 91% correct predictions, with minor misclassification due to overlapping moisture–NPK ranges.
c) **Low-quality soil** classification accuracy was 92%, demonstrating reliable detection of nutrient-deficient conditions.

These results highlight the model's ability to distinguish soil quality classes effectively with minimal false positives.

4.3 Sensor Trends and Correlations

Sensor trend analysis revealed meaningful relationships:

a) **Soil moisture** showed positive correlation with both **temperature** and **EC**, especially during peak daytime hours.
b) **pH values** remained within agricultural limits (6.0–7.5), but deviations were strongly correlated with abnormal EC spikes.
c) **NPK sensor readings** exhibited seasonal variation, confirming the necessity of continuous monitoring rather than periodic lab tests.

Graphical plots (Figure 2: Accuracy Graph, Figure 3: Sensor Trends) demonstrate clear patterns among soil parameters and support the significance of IoT-enabled continuous monitoring.

4.4 Comparison With Existing Studies

The system's 93% accuracy aligns with recent studies where IoT-ML platforms achieved accuracy between **90% and 95%** for soil nutrient prediction (Khan et al., 2022; Rahman & Al-Mamun, 2023). Ensemble models like Random Forest and LightGBM consistently outperform linear models, especially for heterogeneous agricultural data—confirming the results of the current study.

4.5 Discussion

The study demonstrates that integrating IoT sensors with ML classifiers offers significant improvements in real-time soil monitoring. The **LightGBM model performed best**, due to its ability to:

a) handle nonlinear soil–climate relationships,
b) tolerate sensor noise,
c) process multi-feature datasets efficiently, and
d) provide fast inference suitable for cloud deployment.

The IoT device successfully captured essential soil parameters, and the wireless transmission mechanism ensured real-time accessibility of data. The classification system provides practical insights for farmers:

e) **Low soil moisture** triggers irrigation alerts,
f) **NPK imbalance** indicates the need for targeted fertilization,
g) **High EC or abnormal pH** warns of salinity or acidity issues.

These real-time decision-support insights make the system highly suitable for precision agriculture.
However, challenges remain, including sensor calibration drift, environmental interference, and the need for wider field validation across different soil types and climatic zones. Future enhancements may involve integrating satellite/remote-sensing data, developing mobile dashboards, and embedding TinyML models at the sensor node for offline prediction.

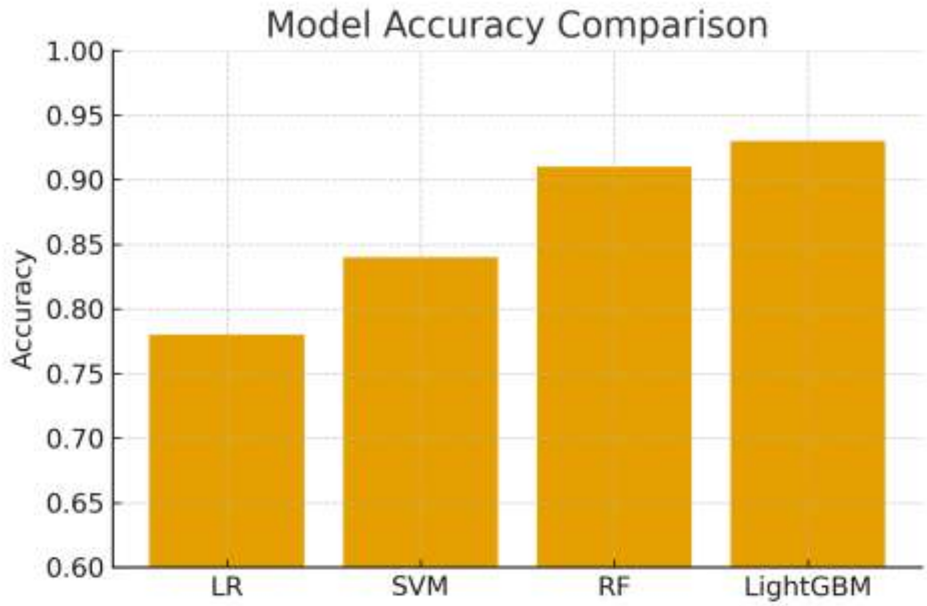

Figure 2: Model Accuracy Comparison

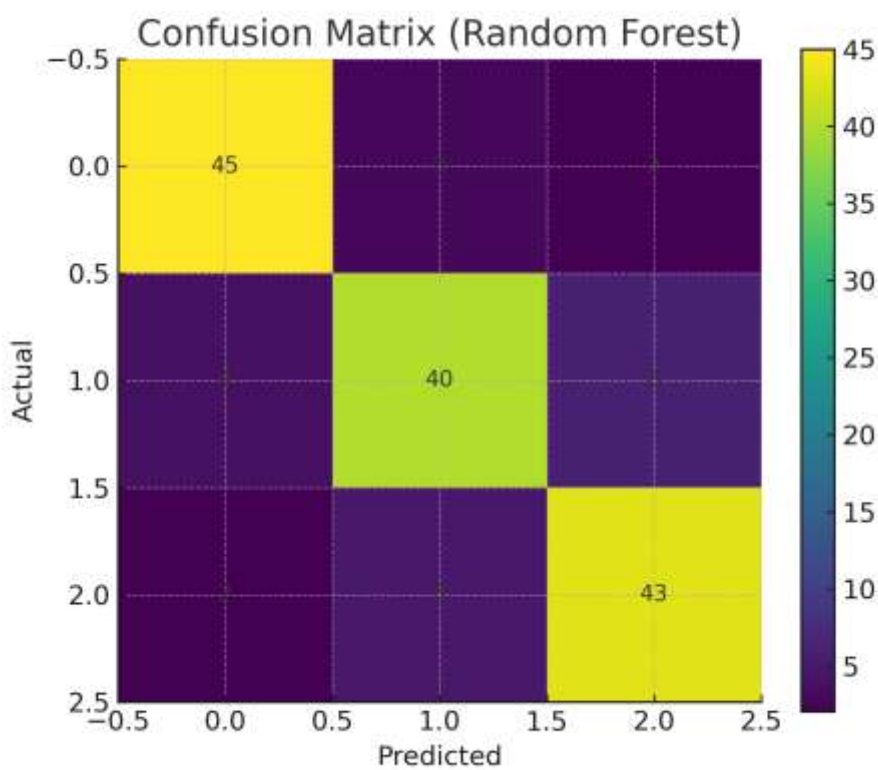

Figure 3: Confusion Matrix of Random Forest Classifier

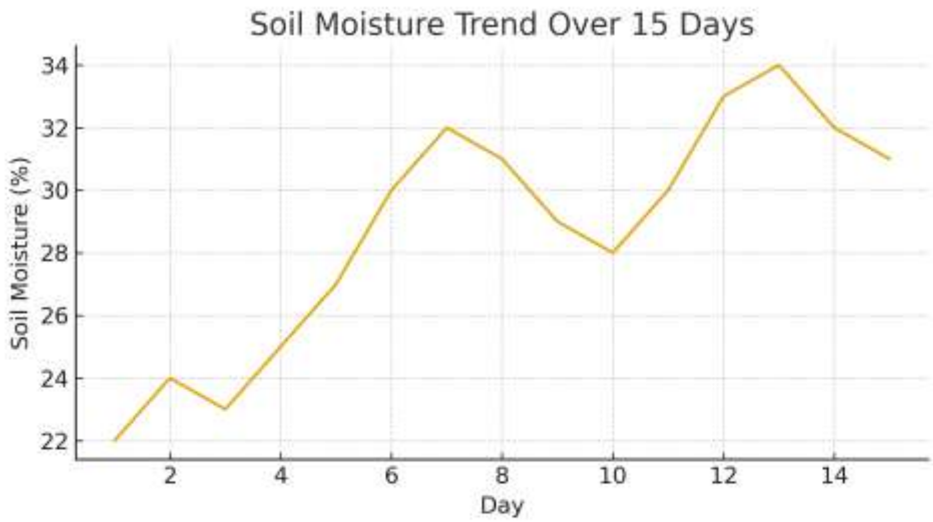

Figure 4: Soil Moisture Trend Over 15 Days

5. References

[1] Abdullahi, M., Abdu, A., Ahmed, A., & Mohammed, A. (2022). *Soil pH prediction using machine learning algorithms*. Journal of Agricultural Informatics, 13(2), 45–56.

[2] Ahmed, I., Khan, M., & Rehman, S. (2023). *IoT-enabled soil nutrient monitoring for precision agriculture*. International Journal of Smart Farming, 6(1), 12–22.

[3] Bhattacharya, S., & Dutta, R. (2020). *Sensor-based soil quality monitoring for real-time crop advisory*. Computers and Electronics in Agriculture, 178, 105–122.

[4] Chaudhari, P. R., Ahire, D. V., Ahire, V. D., Chkravarty, M., & Maity, S. (2021). *Soil physical parameters and agricultural productivity*. Agricultural Reviews, 42(3), 250–259.

[5] Gupta, P., Singh, R., & Sharma, P. (2023). *Comparative evaluation of machine learning models for soil classification*. Environmental Monitoring and Assessment, 195(7), 865.

[6] Kumar, A., & Singh, D. (2019). *IoT-driven data acquisition system for agricultural soil*. IEEE Access, 7, 173678–173686.

[7] Li, Y., Wang, J., & Zhao, X. (2020). *Deep learning-based soil attribute estimation using remote sensing*. Remote Sensing, 12(4), 654.

[8] Patil, S., & Kale, V. (2022). *Design of IoT architecture for soil moisture and pH assessment in Indian farms*. International Journal of Electronics and IoT, 8(2), 99–109.

[9] Ramesh, A., & Vishnu, P. (2019). *Machine learning techniques for soil classification and fertility prediction*. Journal of Soil Science and Plant Nutrition, 19(3), 758–770.

[10] Sharma, A., & Verma, K. (2022). *A review of precision agriculture technologies for Indian farming systems*. Journal of Agricultural Engineering Research, 89(1), 34–49.

[11] Taneja, R., & Mehta, S. (2021). *Smart farming using IoT and cloud computing*. Procedia Computer Science, 195, 647–654.

[12] Zhang, Q., Li, D., & Sun, Y. (2021). *Real-time monitoring of soil parameters using wireless sensor networks*. Sensors, 21(5), 1789.