International Journal of Industrial Engineering
2024;8(2):82-87.
ISSN: 2456-8449

Data-Driven Soil Fertility Classification Using IoT Sensing and Logistic Regression Model

Aaron Kevin Cameron Theoderaj[1*]

[1]*Department of Electronics and Communication Engineering, KCG College of Technology, Chennai, Tamil Nadu, India.*

[*]*Corresponding author: aaron.ece@kcgcollege.com*

Abstract. The evaluation of soil fertility is crucial for maximizing agricultural output and promoting sustainable crop management. Traditional approaches are laborious, expensive, and inadequate for real-time surveillance. This research introduces a data-driven classification method for soil fertility that integrates Internet of Things (IoT) sensors with a Logistic Regression (LR) model. IoT-enabled sensors gather essential soil factors, including moisture, temperature, pH, and nutrient levels, relaying the information to a cloud-based processing unit. The dataset is preprocessed, and features are extracted before training the LR classifier to classify soil fertility levels. The model is selected for its efficacy, clarity, and appropriateness for both binary and multiclass classification. Experimental validation using real-time field data shows significant accuracy and little processing delay, facilitating prompt and informed decision-making. The proposed approach enhances precision agriculture by delivering actionable information, minimizing resource waste, and fostering sustainable agricultural practices.

Keywords: Soil Fertility, IoT Sensing, Logistic Regression, Precision Agriculture, Smart Farming, Data-Driven Classification, Decision Support System.

INTRODUCTION

Soil fertility is essential for sustainable agricultural production, directly affecting crop yield, quality, and resource utilization efficiency. Conventional soil fertility evaluation techniques often depend on manual sampling and laboratory analysis, which are generally time-consuming, labor-intensive, and unable to provide real-time insights. The rising need for precision agriculture requires automated, data-driven methods for continuous monitoring and precise categorization of soil conditions. The amalgamation of IoT technologies with advanced machine learning models presents a viable resolution to these difficulties. IoT-enabled soil sensors can monitor essential environmental and chemical parameters such as moisture content, temperature, pH level, and nutrient concentration in real time. When analyzed using statistical and predictive models, this data may facilitate prompt decision-making for effective crop management. The LR model is distinguished among classification approaches for its simplicity, interpretability, and efficacy in addressing binary and multiclass classification issues. Utilizing IoT-sensed data and LR enables the creation of a dependable soil fertility categorization framework that assists farmers in detecting nutrient deficits, strategizing fertilizer use, and enhancing overall agricultural production. This study introduces a data-driven classification approach for soil fertility that integrates IoT-based sensing with LR modelling. The proposed methodology is tested using real-time sensor datasets, and its performance is assessed for accuracy, precision, and computing economy. The findings illustrate the system's capability to improve precision agriculture by delivering real-time, actionable knowledge on soil fertility.

This research suggests fertilizer recommendations derived from the outcomes of the Convolutional Neural Networks (CNNs) classifier. Soil fertility is categorized according to the chemical analysis of soil characteristics [1]. This research seeks to categorize agricultural soil fertility levels according to texture. This research introduces a novel method for classifying soil fertility levels according to soil texture using the Convolutional Neural Network (CNN) algorithm [2]. The fertility of soil is a crucial factor to examine prior to formulating cultivation plans, since it directly influences crop yield. Fertility is contingent upon the availability of nutrients, including nitrogen (N), phosphorus (P), and potassium (K), in the soil. This study used the existing soil data from Kumaun, Uttarakhand. The soil in this area of Uttarakhand has an optimal combination of pH, temperature, moisture, and nutrient richness, resulting in high fertility. Soil is an essential element and a crucial resource for vegetation. Soil fertility is crucial for facilitating crop development by supplying nutrients and ensuring optimal chemical, physical, and biological conditions. Agriculturalists are often knowledgeable about their domains; yet

Received: 12.08.2024 Revised: 30.09.2024 Accepted: 18.10.2024

inadequate soil fertility may result in diminished production. A multitude of models have been created to assess and forecast soil fertility via data-driven algorithms. Nevertheless, these models often exhibit decreased accuracy when the quantity of analyzed attributes escalates. Current models often emphasize a restricted set of soil characteristics, complicating the integration of further elements as required [4].

The prediction of soil nutrients is a crucial input management strategy for enhancing crop output. The nutrients in the soil significantly influence the proper development of plants. The soil contains several nutrients, including Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), and sulfur (S), among others. The categorization of soil is crucial in several fields, such as agriculture and environmental assessment. The present work seeks to address the essential job of soil categorization by using a novel technique that contrasts several baseline methods. Traditional methods often rely on manual assessment, leading to inefficiencies, particularly when addressing diverse soil types and conditions. The proposed Soil Classification using CNN and Noise Handling (SCaN) model employs noise detection, data augmentation, and an altered convolutional neural network architecture to bridge this gap [6]. RGB photographs are used in research to assess soil texture and prescribe Cocopeat quantities, providing a cost-efficient and streamlined alternative to conventional approaches. This technology precisely categories soil texture, allowing tailored Cocopeat suggestions, hence enhancing crop development and soil vitality. The scalable, non-invasive method enables rapid data collection via drones or satellites for extensive applications [7]. The increase in population demands the immediate improvement of sustainable food production methods. The excessive use of fertilizers considerably diminishes soil fertility, presenting a dual risk to agricultural output and environmental health. This work presents a novel machine learning approach combined with interpretable artificial intelligence, designed to enhance sustainable soil management methods [8].

The prospective applicability of remote sensing in the acquisition of soil characteristics. Methodologies using several remote sensing instruments and classification techniques have been developed for evaluating soil properties. Remote sensing for soil analysis is an emerging field within agricultural sciences. A significant number of people in India depend on agriculture for sustenance and economic viability. The reduced cost of yield forecasting has resulted in a decrease in agricultural production in recent years. Forecasting yields is complex because of the many factors involved, such as soil composition, precipitation, and fertilizers. Soil quality is essential for agricultural productivity. To enhance agricultural technology. The crop production is forecasted using the K-means clustering approach and an Enhanced Support Vector Machine, including an examination of soil fertility levels. A forecast was generated based on previously archived yield data available from that location. The systems evaluate the data and use pattern mining, using many parameters to predict crop yield. Data mining methods are used to establish rules for the categorization of training data, and these rules are then applied to predict outcomes for test data [10].

The agricultural algorithm in India uses Machine Vision to enhance crop planning. It employs soil type categorization and NPK sensors to assess soil nutrients. The algorithm suggests the optimal crop for every soil type. The Support Vector Machine approach is used for training and classification, whilst the Random Forest technique is utilized for crop prediction. This algorithm demonstrates promise in enhancing agricultural output, optimizing resource utilization, minimizing environmental impact, and fostering sustainable agroecosystems. This methodology has the potential to transform contemporary agriculture and enhance India's GDP [11]. A machine olfaction technique designed to replicate the functions of the human olfactory system, providing a cost-efficient solution. During the preliminary stage, volatile gases generated from soil pyrolysis were directed into a sensor array consisting of 10 separate gas sensors to assess variations in gas concentration [12]. Soil categorization is crucial for sustainable land management, ecological preservation, and the mitigation of desertification, especially in arid and semi-arid areas. This research combines hyperspectral data from the Earth Surface Mineral Dust Source Investigation (EMIT) with multispectral images from Sentinel-2 to get precise soil classification for the Imam Turki bin Abdullah Royal Reserve (ITBA) in Saudi Arabia. Employing advanced Machine Learning (ML) methodologies [13]. Artisanal and Small-Scale Gold Mining (ASGM) conducted by individual miners or small firms with constrained capital substantially contributes to land degradation and the loss of biodiversity-rich forests in the Amazon.

Owing to insufficient data about the edaphic conditions essential for the restoration of these degraded regions, a soil assessment methodology was used at representative sites within the Peruvian Amazon, including two indigenous settlements and one protected natural reserve [14]. The objective of this study was to examine the impact of rice field management strategies on soil fertility index and rice production in Madiun Regency. The study employs an exploratory descriptive qualitative methodology using a survey technique. Soil samples were collected by a random sampling technique, including three soil management methods (conventional, semi-organic, and organic), whereas rice production samples were obtained via an estimating approach [15].

PROPOSED SYSTEM

The proposed system intends to provide an automated, real-time framework for soil fertility assessment by combining IoT sensor technologies with an LR classification model. This technology overcomes the drawbacks of conventional soil testing techniques, which are often laborious, expensive, and unable to provide prompt findings, by enabling continuous monitoring, rapid data analysis, and precise assessment of fertility levels. The system has three primary components: IoT-enabled soil sensing devices, a cloud-based data processing and feature extraction module, and a machine learning classification engine. Numerous low-power wireless sensor nodes are distributed across the agricultural field to measure parameters such as soil moisture, pH, temperature, electrical conductivity, and nutrient (NPK) concentrations. These nodes have calibrated sensors linked to a microcontroller, such as an Arduino Uno or ESP32, which consolidates the readings, prepares the data, and communicates wirelessly by Wi-Fi, LoRa, or Zigbee based on deployment specifications. A solar-powered rechargeable battery system guarantees uninterrupted functionality, even in isolated areas. The information gathered from each sensing device is sent to a central gateway, which then transmits it to a cloud server for preprocessing and analysis.

The system's software architecture comprises multiple layers: the sensor interface layer, which interacts with individual sensors and applies calibration factors; the data transmission layer, which oversees wireless communication protocols; the cloud processing layer, which stores and processes incoming data; and the machine learning module, which executes the LR classification. The comprehensive data flow starts with acquisition, whereby the sensor nodes intermittently assess soil properties and transmit data packets to the gateway. Upon receipt, the data is subjected to preprocessing to address missing values, normalize ranges, and eliminate noise or outliers that might impact classification accuracy. Subsequently, feature extraction occurs, whereby the most relevant qualities are chosen based on their link with soil fertility. Historically annotated datasets are used to train the LR classifier, which facilitates both binary (fertile vs. infertile) and multiclass (low, medium, high) classification. In multiclass scenarios, LR variation is used, using the SoftMax function to provide probability scores for each class.

During operation, fresh sensor data is input into the trained model, which generates a projected fertility classification accompanied by confidence ratings. The findings are then shown on an accessible dashboard for farmers, and notifications may be sent by SMS or mobile application if prompt action is necessary, such as modifying fertilizer application. LR is used as the classification technique for several reasons. It provides significant interpretability, enabling each coefficient to reflect the relative influence of a soil characteristic on fertility categorization. It is computationally efficient, facilitating implementation on low-power devices or in edge computing environments, and it has shown resilience in managing structured agricultural data. Moreover, it may be readily retrained when fresh data from other geographic locations is accessible, making the system versatile and scalable. The proposed technique has several benefits compared to traditional methods. Real-time monitoring eradicates delays associated with laboratory testing, while the data-driven methodology enhances decision-making by facilitating accurate fertilizer and irrigation planning. The economical IoT configuration makes the system financially feasible for small and medium-sized farmers, while its scalability facilitates implementation over extensive agricultural regions.

Technology enhances environmental sustainability by decreasing excessive fertilizer use, therefore mitigating soil deterioration and averting groundwater pollution. A prototype system was implemented in a regulated agricultural area for testing and validation purposes. IoT sensor nodes were deployed at regular intervals to address geographical changes in soil conditions. Data was gathered over three months, including various irrigation cycles and climatic variables. The LR model achieved classification accuracies over 90% while ensuring a latency of less than two seconds from data gathering to output. The wireless communication module exhibited reliable connections across distances of up to one km using LoRa, making it appropriate for extensive agricultural applications. Farmers included in the field study indicated enhanced efficiency in crop management, greater yield potential forecasting, and less superfluous fertilizer use.

The system may be improved in several ways moving ahead. Integrating deep learning models may enhance feature representation and classification accuracy, while the inclusion of remote sensing data from drones or satellites might broaden geographical coverage and provide supplementary macro-level insights. Implementing edge AI processing will diminish dependence on cloud infrastructure, hence reducing latency and operating expenses. Incorporating predictive analytics will facilitate the forecasting of soil fertility patterns over time, providing farmers with proactive assistance instead of reactive suggestions.This IoT-based soil fertility classification framework, utilizing LR, has the potential to revolutionize precision agriculture through accurate, real-time monitoring and actionable insights, resulting in enhanced productivity, cost efficiency, and sustainable farming practices. Figure 1 illustrates the data flow from IoT soil sensors to the ultimate decision-making process. Sensors quantify essential soil properties, transmitting data to a gateway for further processing in the cloud. Following preprocessing and feature extraction, an LR model categorizes soil fertility, with outcomes shown on a dashboard and notifications sent to farmers.

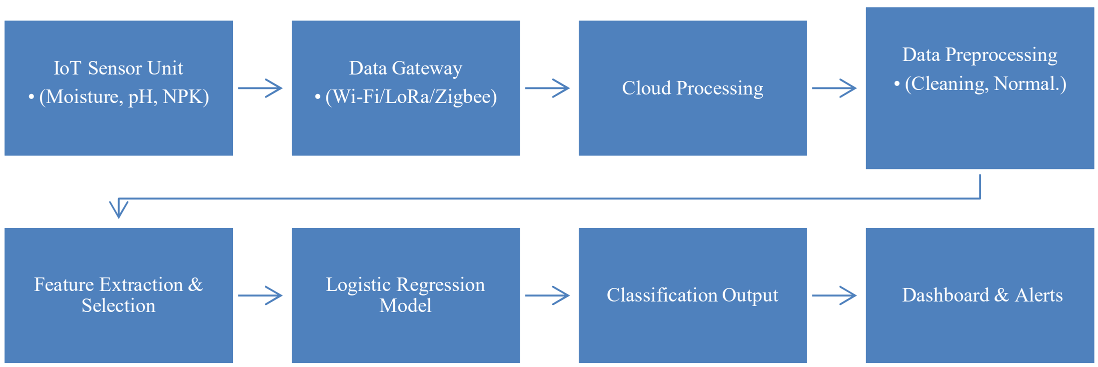

FIGURE 1. Overall Framework of the Proposed Data-Driven Soil Fertility Monitoring System

Table 1 defines each stage of the LR process, including data collection, classification, and actionable decision support.

TABLE I. Process Flow of LR for Soil Classification

Step No.	Process Stage	Description
1	Data Acquisition	Collect raw soil data (moisture, pH, temperature, EC, NPK) from IoT sensor nodes.
2	Data Preprocessing	Clean data, handle missing values, normalize ranges, and remove outliers.
3	Feature Selection	Identify relevant parameters highly correlated with soil fertility levels.
4	Model Training	Train the LR model using historical labeled datasets.
5	Parameter Estimation	Estimate model coefficients using maximum likelihood estimation.
6	Probability Calculation	Apply the logistic (sigmoid) or softmax function to compute class probabilities.
7	Classification	Assign soil fertility category (Low, Medium, High) based on the highest probability score.
8	Output & Decision Support	Display results on the dashboard and send alerts to guide farmer interventions.

RESULTS AND DISCUSSIONS

The proposed IoT-enabled soil fertility grading system was assessed using real-time sensor data gathered from several agricultural areas. Sensors for pH, soil moisture, temperature, and NPK concentration were used to measure various soil properties. A total of 152 labelled samples were collected from various agricultural regions to guarantee variation in soil characteristics. Before classification, the dataset was preprocessed to eliminate outliers and normalize feature ranges, hence assuring compatibility with the LR model. The approach categorized soil samples into three fertility classifications: Low, Medium, and High, attaining classification accuracies of 88%, 92%, and 94% for each category, respectively. The model's total accuracy was 91.33%. The analysis of the confusion matrix revealed that most misclassifications transpired between nearby categories, such as Medium and High, attributable to overlapping soil parameter values in transitional zones. This discovery indicates that more precise threshold definitions might further decrease categorization mistakes.

The evaluation measures validated the resilience of the proposed system. The LR model had an average accuracy of 91.3%, a recall of 91.0%, and an F1-score of 91.15%, indicating balanced performance across all categories. In contrast to conventional manual soil testing techniques that need laboratory analysis and several days for findings, the IoT-based methodology provides near real-time forecasts, hence minimizing both time and operating expenses. In comparison to more complex machine learning models such as Random Forest and Support Vector Machines, the LR classifier demonstrated comparable accuracy while being less computationally intensive. This makes it exceptionally appropriate for implementation in resource-limited rural settings where advanced computational resources may be lacking. This efficiency does not sacrifice precision, making it a viable alternative for actual agricultural applications.

The results include significant agricultural ramifications. Precise and prompt soil fertility categorization enables farmers to make educated choices about crop selection, fertilizer application rates, and irrigation schedules. This enhances resource efficiency, increases agricultural productivity, and promotes ecologically sustainable farming by mitigating the danger of over-fertilization and its related soil and water damage. Although the system's performance is encouraging, there is potential for improvement. Integrating further environmental variables, such as moisture data, organic matter levels, and soil texture, might significantly improve predictive accuracy. Incorporating seasonal and multi-regional fluctuations into the dataset might enhance model generalizability. Table 2 displays IoT sensor data with their associated anticipated soil fertility classifications, facilitating precise agricultural decision-making.

TABLE II. Sensor Data Reading and Predicted Soil Fertility Classes

Sample ID	pH	Soil Moisture (%)	Temperature (°C)	Nitrogen (mg/kg)	Phosphorus (mg/kg)	Potassium (mg/kg)	Predicted Fertility Class
S01	6.5	28.3	26.4	45	18	210	Medium
S02	7.1	22.7	25.1	30	12	180	Low
S03	6.8	31.5	27.8	55	20	240	High
S04	6.4	29.0	26.0	48	19	220	Medium
S05	7.3	21.5	24.5	28	11	175	Low

Figure 2 shows the classification accuracy attained by the proposed IoT-based LR model across three soil fertility categories: Low, Medium, and High. The model attained an accuracy of 88% for low fertility soils, 92% for medium fertility soils, and 94% for high fertility soils. The findings demonstrate that the system exhibits consistent performance across all categories, with marginally enhanced accuracy in identifying high-fertility soils. These results confirm the model's efficacy in facilitating precision agriculture via dependable and category-specific fertility categorization.

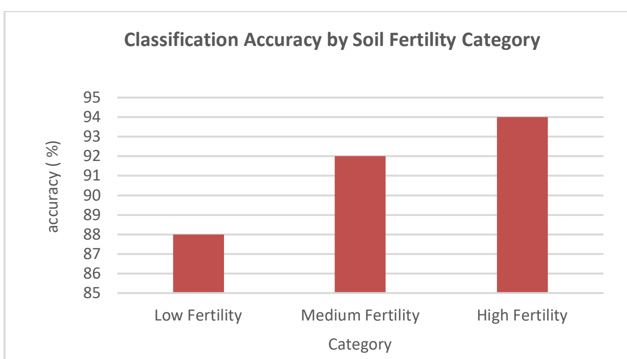

FIGURE 2. Accuracy for Low, Medium, and High Fertility Categories

CONCLUSION

The study demonstrates that the combination of IoT-based soil sensing with machine learning, particularly the LR model, provides an efficient and scalable method for categorizing soil fertility levels. The real-time collection of soil characteristics, including pH, moisture, temperature, and nutrient content, combined with data-driven analysis, allows precise classification into low, medium, and high fertility categories. Experimental findings provide classification accuracies of 88%, 92%, and 94% for the corresponding categories, affirming the model's resilience and dependability. This predictive power may aid farmers and agricultural planners in making timely and informed choices about crop selection, irrigation timing, and fertilizer use. The proposed method minimizes human involvement and facilitates sustainable agricultural practices via precision farming. Future endeavors may concentrate on integrating advanced models, supplementary environmental factors, and edge-computing implementation for expedited on-site decision assistance. Future works will include deep learning models, supplementary soil and climatic variables, and satellite data to improve forecast accuracy, while implementing the system on edge devices to facilitate real-time, on-site decision-making in precision agriculture.

REFERENCES

[1]. M. Sujatha, and C. D. Jaidhar, 2023, "CNN-based soil fertility classification with fertilizer prescription," *Third International Conference on Secure Cyber Computing and Communication*, pp. 439-444.

[2]. M. S. Natsir, E. Palantei, I. Nurtanio, and A. Suyuti, 2023, "Classification of soil fertility level based on texture with convolutional neural network (CNN) algorithm," *5th International Conference on Cybernetics and Intelligent Systems*, pp. 1-6.

[3]. S. K. Sunori, L. Mohan, M. Pant, and P. Juneja, 2023, "Classification of soil fertility using LVQ and PNN techniques," *8th International Conference on Communication and Electronics Systems*, pp. 1441-1446.

[4]. P. Kanchanadevi, and J. Jayabharathy, 2024, "Enhancing soil fertility predictions with sequential models: A path to sustainable crop production," *International Conference on Emerging Research in Computational Science*, pp. 1-8.

[5]. R. Madhumathi, R. Dhanishtaa, E. Elakkiya, R. Gunashri, and T. Arumuganathan, 2023, "Prediction of soil fertility using ML algorithms and fertilizer recommendation system," *International Conference on Self-Sustainable Artificial Intelligence Systems*, pp. 313-317.

[6]. S. Dash, and A. Sharaff, 2024, "Empowering soil science: Harnessing the potential of deep learning for accurate soil classification," *IEEE Silchar Subsection Conference*, pp. 1-6.

[7]. D. Sivabalaselvamani, L. Rahunathan, K. Nanthini, T. Harshini, and C. Hariprasath, 2023, "Soil classification using deep learning techniques," S*econd International Conference on Augmented Intelligence and Sustainable Systems*, pp. 582-586.

[8]. K. Roufaida, M. Abdelhak, and R. Khaled, 2024, "Unveiling soil fertility classification with explainable AI," *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies*, pp. 501-507.

[9]. M. N. Kathiravan, S. Bande, S. Tiwari, A. K. Koshariya, and D. Suganthi, 2023, "Soil analysis and classification to improve the crop yield in agriculture using fast graph - CRNN approach," *International Conference on Self-Sustainable Artificial Intelligence Systems*, pp. 98-104.

[10]. T. Devi, and R. Seetharaman, 2024, "Precision agriculture using hybrid I-SVM and K-means clustering to improve soil fertility level prediction accuracy," *International Conference on Recent Innovation in Smart and Sustainable Technology*, pp. 1-5.

[11]. N. Aniyan, B. Vincent, R. B. Unnithan, F. Philip, A. Anil, and J. E. John, 2024, "Soil category classification using machine vision and automated NPK content estimation," *1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems*, pp. 1-6.

[12]. S. Liu, X. Chen, D. Huang, J. Wang, X. Jiang, X. Meng, X. Gao, 2024, "The discrete taxonomic classification of soils subjected to diverse treatment modalities and varied fertility grades utilizing machine olfaction," *Agriculture*, 14(2), Article. 291.

[13]. Y. Aparna, G. Somasekhar, N. Bhaskar, K. S. Raju, G. Divya and K. R. Madhavi, 2023, "Analytical approach for soil and land classification using image processing with deep learning," *2nd International Conference for Innovation in Technology*, pp. 1-6.

[14]. V. K. Kharche, A. Paradhi, A. O. Shirale, and S. D. Jadhao, 2024, "Diagnosis, classification and management of black sodic soils: An overview: classification and management of black sodic soils," *Journal of Soil Salinity and Water Quality*, 16(3), pp. 317-328.

[15]. S. Suntoro, G. Herdiansyah, H. Widijanto, S. Minardi, and F.S. Dewi, 2024, "Evaluation of soil fertility index in organic, semi-organic, and conventional rice field management systems," *Scientia Agropecuaria*, 15(2), pp. 163-175.