



OPEN IoT based soil moisture measurement and type prediction using advanced regression and machine learning models

Md. Mahmud Sazzad^{1,3}, Tanvir Ahmed^{1,3}✉, Golam Kibria^{1,3} & Ishmam Khan^{2,3}

Measuring soil moisture has a big impact on resource efficiency and decision-making in geotechnical engineering, agriculture, and environmental sustainability. This study presents a revolutionary Internet of Things-based method for predicting soil type and moisture in real time. Water content was verified using the traditional oven-dry method for accuracy evaluation, and capacitance measurements were gathered using a sensor to generate a custom dataset in the lab. The method outperformed linear regression in water content prediction, achieving 96.49% accuracy using an Excel logarithmic regression equation. Furthermore, a machine learning model that employed polynomial regression was able to measure the water content of the soil and predict values for fine, medium-coarse, and coarse sand types. With an R^2 score of 0.79, the model can account for almost 79% of the variation in water content and produce 1.71% of MAE (Mean Absolute of Error), indicating a strong relationship between capacitance and water content. On the expanded dataset, Random Forest classifier was chosen for classification, which correctly identified the intended soil types with an accuracy of roughly 97.77%. By combining sensor data with sophisticated algorithms, the suggested methodology makes it possible to analyze soil qualities effectively and non-destructively. This scalable method offers substantial potential for environmental management, soil monitoring, and precision agriculture and is flexible enough for the creation of mobile applications. Predictive modeling and real-time data processing combined improve resource management effectiveness while lowering the need for human intervention. In order to improve forecast accuracy and application and support more environmentally friendly farming methods and environmental monitoring, further research will surely give efforts to enrich the dataset, incorporate a variety of soil types, and take environmental elements like temperature and salinity into consideration.

Keywords Soil moisture content, Laboratory tests, Capacitive sensor, Oven dry method, Calibration, Artificial intelligence, Machine learning, Classification, Random forest classifier, Regression, Polynomial regression, Internet of things

The moisture content of soil is essential to many physical, chemical and biological processes in both natural ecosystems and man-made structures. It indicates the percentage of water in the soil, and a deep understanding of its dynamics is necessary to maintain the fragile soil-water balance. Accurate and real-time soil moisture monitoring is essential to improve agricultural productivity, ensure infrastructure durability, and promote environmental sustainability. In agriculture, it is essential to maintain soil health, increase crop yields, and protect water supplies. Monitoring soil moisture is crucial in civil engineering to prevent freeze-thaw cycles and chloride intrusion from damaging structures. Therefore, in order to protect the integrity of the environment and the health of the infrastructure, precise and efficient methods for measuring soil moisture must be developed.

Soil moisture was crucial to increase crop productivity, soil health, and agricultural water use efficiency. Although traditional gravimetric methods were believed to be accurate, their inability to provide real-time data made them inappropriate for dynamic scenarios. Capacitance-based sensors were developed as a continuous and economical means of monitoring the dielectric behavior of soils that were affected by the presence of moisture

¹Department of Civil Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh.

²Department of Computer Science and Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh. ³These authors contributed equally to this work. ✉email: tanvirahmedb37@gmail.com

levels¹. Those sensors translate moisture levels into measurable signals (e.g., frequency or pulse width modulation) for real-time monitoring. While² proposes a new cylindrical electrode design for bulk materials, optimizing sensor geometry for granular media such as soil,³ evaluates the practical deployment challenges of such sensors in IoT-based irrigation systems, including calibration drift and field variability. Together, these studies highlight the trade-offs between sensor innovation² and operational reliability³ in agricultural applications. Their long-term dependability and suitability for clay and silty soils were still restricted. To develop universal, real-time sensing systems that might be used in a variety of agricultural contexts, more study was required. CWM-four and other calibration-free models demonstrated potential⁴. For the compaction management process in geotechnical engineering, precise soil moisture measurement was essential. For fine-grained soils, capacitive and resistive sensors provide a portable, affordable, and real-time substitute for oven-drying methods. However, environmental elements such as soil type, compaction, porosity, temperature, and salt may affect their accuracy. These sensors demonstrated the ability to enhance geo-technical processes, despite problems with coarse-grained soils and calibration difficulties for various soil types⁵. Despite the dearth of existing calibration methodologies, it was demonstrated that volumetric water content accuracy was improved by soil-specific calibration, ensuring great repeatability and low variability. Small, high-frequency capacitive sensors can be used to assess the moisture content of urban soil⁶. Among gravimetrically calibrated soil moisture sensors, low-cost irrigation sensors were shown to have the lowest error, especially for sandy and sandy loam soils⁷.

The electromagnetic fields produced by electrodes, which were influenced by the dielectric characteristics of the soil, were tracked using capacitance-based sensors. Because of their adaptability, they were able to determine the moisture content of the soil at different depths⁸. Due to their accuracy, speed, and low energy consumption, these sensors were widely utilized in agricultural and grain storage⁹. However, their efficacy was occasionally hindered by salinity, soil texture, environmental diversity, and sensor-to-sensor variability^{6,10}. It has been shown that calibration techniques like quadratic regression increase the accuracy of these sensors when compared to linear models, particularly when soil-specific changes are present^{11,12}. In order to improve real-time monitoring and lower mistakes, techniques like logistic regression were used to lessen the impact of electrical conductivity (EC) on sensor accuracy, even in large agricultural settings. Despite being inexpensive, this method was ineffective in high-salinity and organic-rich soils and ignored elements such soil porosity and texture. Despite these limitations, the technique shown potential for accurate soil moisture monitoring¹³.

Recent studies (e.g.,^{3,12}) have used piecewise linear regression and temperature compensation algorithms (e.g., based on Arrhenius equation adjustments) to mitigate sensor drift caused by temperature fluctuations ($\pm 0.5\%$ accuracy loss per $^{\circ}\text{C}$ in uncontrolled environments). These refinements improved volumetric water content estimates across sand to loam ratios, though challenges persist in clay-rich soils.

For civil engineering projects to last, it was essential to keep an eye on soil moisture throughout freeze-thaw cycles and chloride intrusion. Advanced methods such nuclear magnetic resonance (NMR), electrical impedance spectroscopy (EIS), and X-ray imaging were used to generate high-resolution data; nevertheless, their field applicability was constrained by their cost, intricacy, and limited usefulness¹⁴. Recent advancements have combined low-dimensional materials, like molybdenum disulfide (MoS_2), with hybrid sensing techniques to increase sensor sensitivity and reliability across a range of environmental conditions¹⁵. In order to solve problems specific to soil, data-driven calibration methods and hybrid models that integrate machine learning and sensor data showed promise¹⁶. Water-cut meters for near-linear transfer functions¹⁷, frequency-based grain probe moisture meters for agricultural use¹⁸, and Internet of Things-enabled hydroponic drip fertigation systems¹⁹ are examples of recent advancements in capacitance-based technologies. The need for additional study was highlighted by the persistence of issues such soil-type dependence and decreasing sensitivity at high volumetric water content (VWC) in spite of these developments²⁰. In a range of environmental conditions, new materials such carbon-based composites could improve sensor performance and stability²¹. Combining IoT-enabled soil moisture monitoring with renewable energy sources increased the sustainability of agricultural practices²².

While existing studies demonstrate the efficacy of capacitance-based sensors for soil moisture monitoring^{1–7}, several critical limitations remain unaddressed. First, most calibration models—such as linear or quadratic regression^{11,12}—fail to capture the nonlinear relationship between capacitance and moisture across diverse soil types, particularly in high-plasticity or mixed-grain soils. Second, although hybrid machine learning and sensor approaches show promise^{16,19}, they often overlook practical real-time deployment due to high computational complexity. Third, the classification of soil types using low-cost sensors is still underexplored in laboratory settings, despite its relevance to geotechnical and environmental engineering. Additionally, most IoT-based systems focus solely on moisture content and either assume the soil type is known or do not offer a mechanism for concurrent identification, limiting their adaptability in real-world field scenarios^{19,22}.

This study is distinct in bridging these gaps through three key contributions: (i) it proposes a polynomial regression model tailored to nonlinear capacitance–moisture relationships, achieving improved accuracy (96.49%) over traditional linear/logarithmic models; (ii) it uniquely integrates real-time field measurements with laboratory-based soil classification using the same low-cost capacitance sensor, a feature not present in prior IoT-enabled systems; and (iii) it enables minimal recalibration needs through AI-driven adjustments, reducing the dependency on soil-specific calibration^{5,11}. Unlike previous studies that treat moisture and soil type prediction as separate tasks, our approach offers simultaneous, dual-functionality—real-time water content estimation and soil type classification—within a unified, cost-effective framework. This makes it especially valuable for both agricultural applications and geotechnical evaluations.

The study presents a unique capacitance-based method for laboratory soil type detection and real-time, on-site soil moisture measurement. Using a custom dataset of three soil types and an AI-driven polynomial regression model, the process provides real-time water content readings on-site and predicts soil types in the lab with high accuracy. This approach, which has uses in environmental monitoring, agriculture, and geotechnical

engineering, blends complex calibration processes with machine learning algorithms to offer an economical and effective solution.

Methodology

In order to accurately detect and classify soil moisture, this study uses a methodical methodology to assess the performance of capacitive soil moisture sensors in combination with cutting-edge machine learning approaches. Key steps including sensor selection and calibration, soil sample preparation, controlled laboratory testing, and the use of both statistical and AI-driven models are all part of the technique, which is intended to guarantee reliable data collection and interpretation. Three different soil types are the specific focus of the study, and each phase is carefully planned to preserve scientific rigor and guarantee repeatability and reliability in measurements under a range of moisture circumstances.

Sensor identification and data recording process

Modern irrigation systems make considerable use of capacitive soil moisture sensors because of their remarkable dependability and longevity. These sensors' electrodes are covered in materials that resist corrosion, guaranteeing their durability and reliable operation. Compared to resistive sensors, capacitive sensors provide greater precision and accuracy while retaining dependable performance in a variety of soil types and field circumstances. They are perfect for long-term irrigation management applications because of their continuous sensitivity to changes in temperature and salinity as well as their increased longevity because of materials that are resistant to corrosion²³. An electromagnetic method called capacitive sensing uses capacitive coupling to gauge the soil's dielectric permittivity. The apparent permittivity of soil is closely related to its water content because water's relative permittivity ($\epsilon_r \sim 80$) is much higher than that of other soil components, including organic matter ($\epsilon_r \sim 4$), mineral soil ($\epsilon_r \sim 2-9$), and air ($\epsilon_r \sim 1$)⁵. Bogena et al.'s (2007)²⁴ article goes into great detail into the electronic configurations and circuit diagrams of the majority of capacitive sensors.

The capacitance sensor's (Fig. 1) integrated voltage regulator chip, which guarantees steady performance across a wide input voltage range of 3.3 to 5.5 VDC, makes it compatible with 3.3V Arduino main control boards. It offers an output voltage range of 0 to 3.0 VDC and a PH2.54-3P interface for convenient connection. The sensor is adaptable to a wide range of applications due to its small size (98 mm × 23 mm). Using less than 5mA of current, the sensor produces about 1.5V for wet soil and 3.0V for dry soil when powered at 5V²⁵.

Capacitance sensors solely provide analog output signals, and the collection of these raw signals adheres to a specified protocol. The sensors transform the physical attribute of capacitance into an electrical signal represented as voltage, which is then converted into digital data via an analog-to-digital converter (ADC)

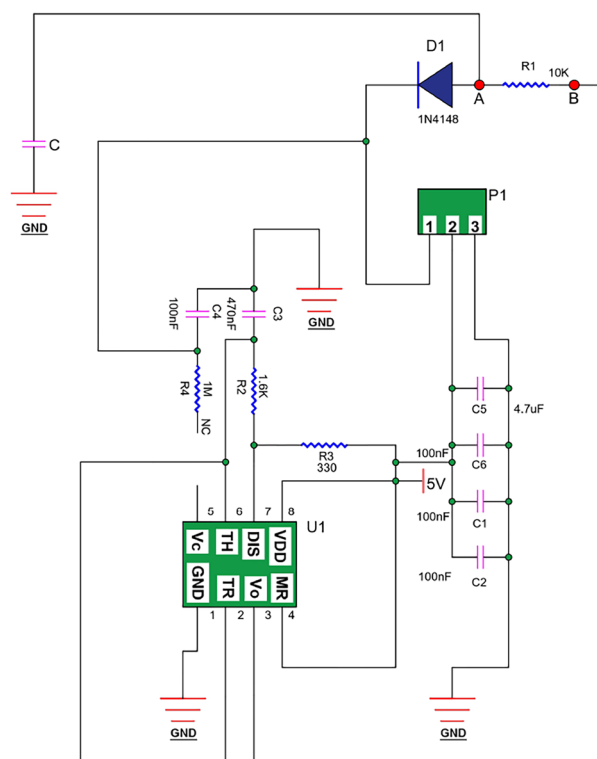


Fig. 1. The capacitive soil moisture sensor (version 1.2) circuit diagram, which includes a signal conditioning connection (P1), resistors, capacitors, and a diode (1N4148). The circuit produces voltage signals that are appropriate for microcontroller-based processing and correlate to the moisture content of the soil.

incorporated within an Arduino-compatible microcontroller board. This study employs the commonly deployed ESP32 microcontroller²⁶.

The ESP32 microcontroller is integrated into the bespoke solution for the capacitive sensor. These cost-effective, low-power 32-bit microcontrollers contain a 12-bit analog-to-digital converter (ADC), which simplifies the conversion of voltage levels from sensor output (ranging from 0 to 4.2 V) into 2^{12} (4096) discrete values, assuring precise digital representation.

The system employs an ESP32 microcontroller to interact with a Capacitive Soil Moisture Sensor v1.2 for real-time monitoring of soil moisture. The analog output of the sensor is read by the ESP32, which is powered by a 3.7V 18650 lithium-ion battery. This battery is managed using a Battery Management System (BMS) to ensure safe operation²⁷. The ESP32 processes the sensor data before sending it over Wi-Fi to the Blynk server, where it is shown on the Blynk mobile app for remote control and monitoring (Fig. 2).

Soil properties

Four types of soil were selected for testing with water content sensors, consisting of four coarse-grained soils: Sand 1, Sand 2, Sand 3, and Sand 4. These soils are representative of materials commonly used in the construction of roadways, foundations, and embankment dams. Sand 1 and Sand 2 were sourced from a building construction site, while Sand 4 was previously utilized in a runway foundation in Rajshahi, Bangladesh. The grain size distribution of the soil samples was determined through sieve analysis, a standard method for characterizing coarse-grained soils. In this study, sieve analysis was conducted for Sand 1, Sand 2, Sand 3, and Sand 4. These soils are classified as uniform sands with minimal fine content. Sand 1 is coarse-grained, with a grain size (D_{50}) of 0.439 mm, whereas Sand 2 is medium coarse-grained, with a grain size (D_{50}) of 0.268 mm. Sand 3 and Sand 4 are classified as fine sands, with grain sizes (D_{50}) of 0.190 mm and 0.138 mm, respectively. Among these, Sand 1 is well-graded, while the other sands are classified as poorly graded. After this for dataset purpose, Sand 3 and Sand 4 are merged. Fig. 3 presents the grain-size distribution curves of the selected soils.

Test configurations and methods

Laboratory-prepared specimens encompassing a wide range of water contents are used to test the moisture sensors for four sandy soil types. An auger or soil corer is used to gather soil samples from the field at predetermined depths and locations, guaranteeing an exact representation of the land being studied. Table 1 represents the physical properties of the selected soils used in this study, including their classification, fines content, and grain size distribution parameters.

Key terms: D_{50} = median grain size; C_u = uniformity coefficient (range of particle sizes); C_c = curvature coefficient (gradation smoothness).

Note: Soil samples are classified based on the Unified Soil Classification System (USCS). Sand 1 is well-graded, while Sand 2, Sand 3, and Sand 4 are poorly graded. Percent fines influence permeability and compaction. Bulk density and specific gravity indicate weight and mineral composition. $C_u (> 4 \text{ for sands})$ measures the range of particle sizes, and $C_c (1-3 \text{ for well-graded soils})$ assesses gradation smoothness.

Soil samples are placed on aluminum or ceramic plates, ensuring the containers are clean and dry to prevent contamination. The initial wet weight of each soil sample is determined using a precision balance, accounting for both soil particles and water content. The containers holding the soil samples are placed in an oven maintained at a constant temperature, typically $105 \pm 0.5^\circ\text{C}$ ($221 \pm 0.9^\circ\text{F}$). The samples are dried for a specified period, usually 24 to 48 hours, or until their mass remains constant. Once the drying process is complete, the containers are removed from the oven and cooled in a desiccator to prevent moisture absorption from the air. After cooling,

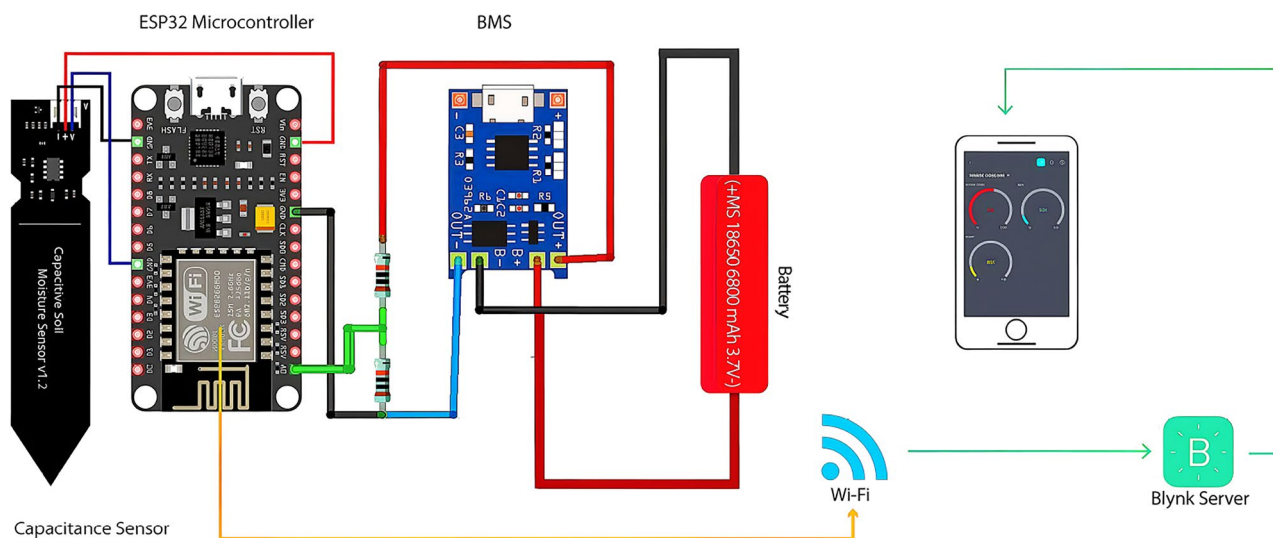


Fig. 2. Circuit diagram of the IoT-based soil moisture monitoring system using an ESP32 microcontroller, capacitive sensor, and Blynk server for wireless data transmission.

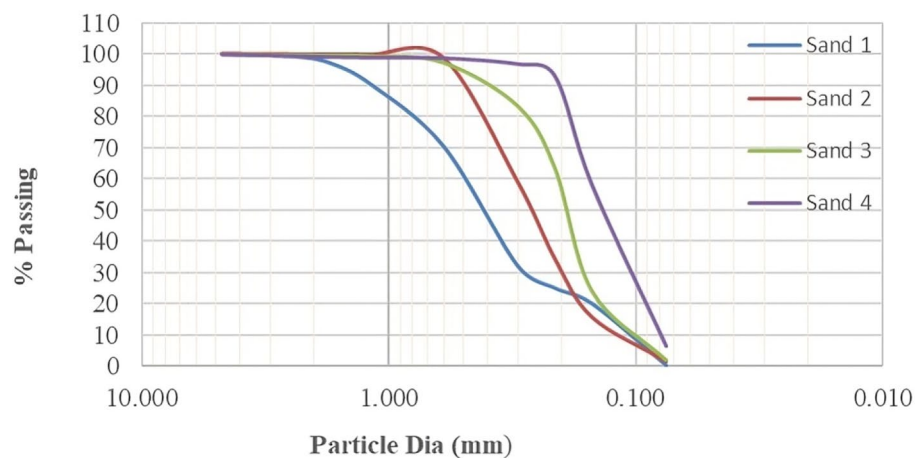


Fig. 3. Variation in particle size and gradation characteristics based on the percentage passing versus particle diameter is depicted by the grain size distribution curves for Sand 1 (Coarse Sand), Sand 2 (Medium-Coarse Sand), Sand 3 (Fine Sand), and Sand 4 (Fine Sand).

Soil type	USCS Class	% Fines	% Sand	Bulk density (g/cc)	Sp. gravity	D_{50} (mm)	Cu	Cc
Sand 1	SW-Well graded	0.42	99.58	1.88	2.67	0.439	4.65	1.29
Sand 2	SP-Poorly graded	1.51	98.50	1.75	2.64	0.268	2.58	1.05
Sand 3	SP-Poorly graded	2.06	97.95	1.74	2.63	0.190	2.03	1.20
Sand 4	SP-Poorly graded	6.30	93.71	1.70	2.63	0.138	1.91	0.97

Table 1. Physical properties of selected soil.

the dried soil samples are weighed to determine their dry weight. The soil moisture content can be found using the following formula:

$$\text{Moisture Content (\%)} = \frac{\text{Wet Weight} - \text{Dry Weight}}{\text{Dry Weight}} \times 100$$

The difference between the initial wet weight and the dry weight represents the mass of water in the soil sample. Dividing this difference by the dry weight and multiplying the result by 100 yields the soil's moisture content as a percentage. The oven-drying method is employed to calibrate the soil moisture measuring device. Achieving accurate results requires careful handling of the soil samples and precise control of the drying conditions, which can be time-intensive.

The sensor readings of each section of the soil sample is measured using the device (Fig. 4a and b). For each sample, the corresponding raw sensor readings (capacitance measurements) are recorded. A calibration curve is established by plotting the raw sensor readings on the x-axis and the associated moisture content values, derived from reference measurements, on the y-axis.

A calibration curve (logarithmic regression) correlates raw sensor readings (x-axis) with oven-dry moisture values (y-axis). This curve translates capacitance measurements into moisture content. The best-fit calibration curve is then used to derive an equation that relates soil moisture content to raw sensor readings (capacitance measurements). This equation serves as a tool to convert raw sensor data into meaningful soil moisture values during subsequent measurements.

The soil moisture content is determined by measuring the sensor readings (Fig. 5) of the samples and applying the calibration equation. To ensure the accuracy and reliability of the moisture readings, they are compared with reference measurements.

AI model implementation:

Linear regression

Linear regression is a machine learning method used to predict a continuous target variable based on one or more input features. It is categorized as a supervised learning algorithm, where labeled training data, consisting of input-output pairs, is utilized to train the model. The general representation of the linear regression model is given by: $y = mx + b$

In this equation, b represents the y-intercept, while m denotes the slope or coefficient of the line.

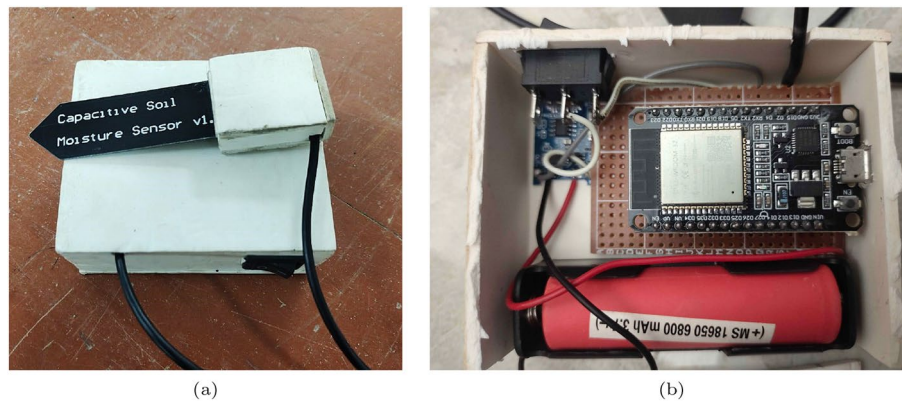


Fig. 4. (a) External view of the assembled soil moisture sensing device using a capacitive sensor. (b) Internal layout showing the ESP32 microcontroller, lithium-ion battery, and supporting circuit components for data acquisition and transmission.

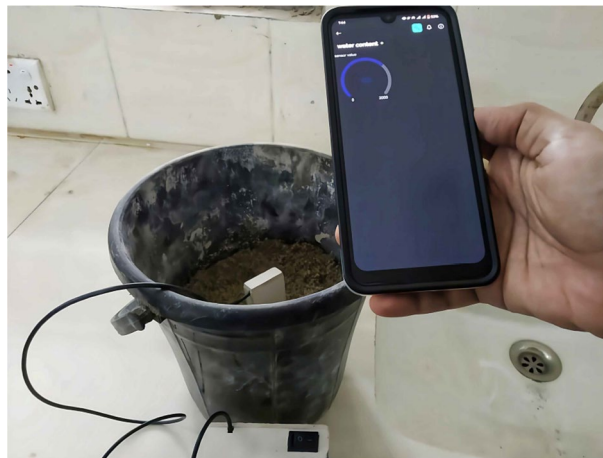


Fig. 5. Monitoring soil moisture using the Blynk app. The mobile app shows the raw sensor value indicating water content after receiving real-time moisture data from the sensor embedded in the soil.

For capacitance-based moisture devices, the purpose of linear regression is to establish a linear relationship between capacitance measurements (input features) and the corresponding soil moisture levels (target variable).

Polynomial regression

Polynomial regression is applied after generating the dataset to model the relationship between variables with a polynomial equation. The general form of the polynomial regression model is:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n + \varepsilon$$

Where:

- Y represents the soil moisture content.
- X denotes the capacitance measurement.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are the coefficients of the polynomial terms.
- ε is the error term.

The degree of the polynomial (n) is typically determined through testing or cross-validation, depending on the complexity of the relationship between capacitance and soil moisture.

In order to effectively estimate unknown data, machine learning models are used to forecast soil capacitance based on different percentages of water content. In order to examine the connection between soil capacitance and water content, a linear regression model is first implemented. A simplified polynomial regression model is then implemented to get predictive water content values.

Data points that correlate percentages of water content with the relevant values of soil capacitance are used in the training phase. The model forecasts soil capacitance for a variety of water content levels, including 2%,

4%, 6%, 8%, 10%, 12%, 14%, and 16%, when training is finished. If we want to perceive water content, we need to use a regression model.

The model uses the given training data to classify the kind of soil after predicting the soil capacitance. It designates particular classifications, such as medium-coarse sand, coarse soil, etc. The model also produces a range of the soil and parameters that help to identify the soil type. This technology supports thorough soil analysis and related applications by enabling precise predictions of soil capacitance and efficient soil type classification.

Random forest classification

To predict soil types based on capacitance and water content parameters, a Random Forest classifier was used on the supplemented dataset. The mode of the classes (for classification) is the final prediction of the Random Forest ensemble learning technique, which builds several decision trees during training.

The following is a summary of a Random Forest classifier's general methodology:

- Multiple decision trees are trained on random subsets of the training data and features.
- Each tree creates an independent prediction.
- The final classification output is determined by majority voting among the trees.

A dataset of different soil types is used to train the algorithm. Initially, we have to select whether we need to know water content or capacitance. If we need capacitance, we need to give the machine a classification model and then it will determine the soil type. This technique enhances generalization and lessens overfitting, particularly for small to medium-sized datasets. In our instance, the Random Forest model correctly and reliably categorized the appropriate soil types, achieving an accuracy of roughly 97.77% on the test set.

Method execution

A flowchart that graphically describes the process of applying machine learning models to forecast soil type and water content is shown in the Fig. 6. It describes how inputs are processed to provide either a soil categorization using a trained classifier or an estimation of the water content using regression, with a focus on capacitance data from an Internet of Things device. This framework's two main functions are soil type recognition and water content prediction.

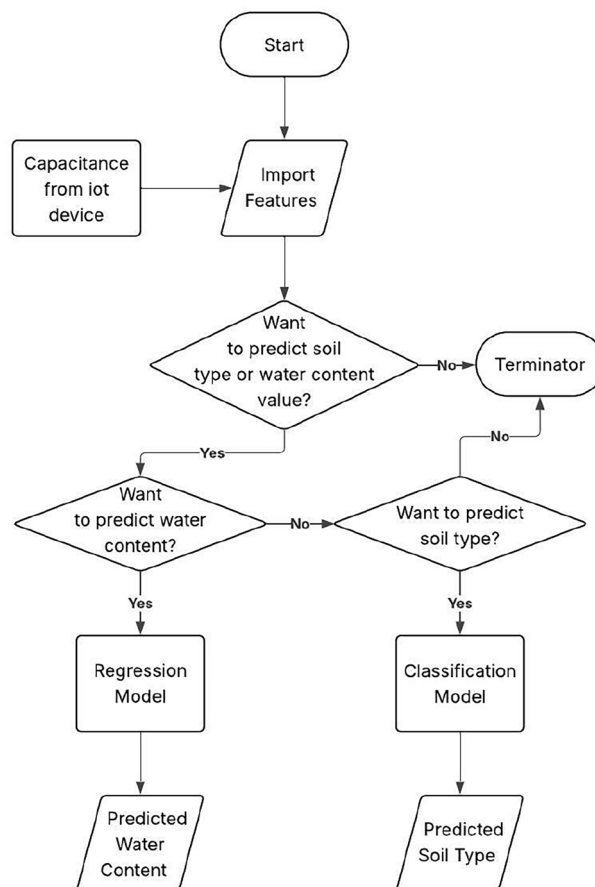


Fig. 6. Flowchart showing the decision process for predicting either soil type or water content based on sensor capacitance readings and input parameters.

Dataset preparation and augmentation for generalization

Three different soil types—fine sand, medium coarse sand, and coarse sand—were experimentally measured for moisture and capacitance in order to produce a specific dataset for this study. There were 120 samples of fine sand, 119 samples of medium coarse sand, and 218 samples of coarse sand in the original dataset. We used the Synthetic Minority Over-sampling Technique (SMOTE) to address the class imbalance problem and enhance the machine learning models' capacity for generalization. Using their feature space, this technique creates new samples for the underrepresented classes in a synthetic manner. Following augmentation, each of the three classes received 218 samples, for a final dataset of 654 samples. In addition to avoiding overfitting during classification²⁸, this augmentation step made sure the model was equally exposed to each type of soil, which improved the prediction robustness and dependability of the model in practical applications.

Selection of input:

The user starts the process by entering necessary data, such as capacitance readings from the Internet of Things device and other pertinent characteristics. The user's goal—whether it's to identify the type of soil or forecast the amount of water—determines the next stage. The flowchart's course is determined by this choice.

Prediction for water content:

The user must enter the capacitance value and the necessary features if they decide to predict the water content. These inputs are processed by the regression model, which was trained on the relationship between capacitance and moisture, to determine the soil's water content. To improve the prediction accuracy, both linear and polynomial regression models have been applied.

Soil type prediction:

When the goal is to estimate the kind of soil, the user inputs the capacitance values that correspond to known percentages of water content (e.g., 2%, 4%, 6%, up to 16%) together with other required information. The pre-recorded capacitance patterns of various soil types under identical water content conditions are then compared with these input values. Using a machine learning-based classification process, the trained Random Forest Classifier determines the most likely soil type based on the closest match between the input pattern and the stored data.

Training and prediction of the model

The model was trained using datasets that showed the percentage of water content correlated with the corresponding capacitance values for various types of soils. The model determines the most likely soil category for soil type prediction by comparing the input capacitance to the training data. Additionally, we used SMOTE (Synthetic Minority Over-sampling Technique) to improve generalization, reduce the danger of overfitting, and augment the dataset after gathering all the data. We used label encoding to translate categorical soil categories into numerical labels for the classification job of predicting soil type based on capacitance or water content. We divided the dataset into training and testing sets using an 80-20 stratified split to guarantee that the class distribution held steady across subsets. Furthermore, to lessen bias and improve the robustness of our findings, we used Stratified K-Fold Cross-Validation (K=5). Given the short and perhaps unbalanced size of our dataset, our approach ensures that every fold preserves the original class distribution.

Generated insights:

To improve the accuracy of soil categorization, the model additionally produces a variety of related metrics in addition to the main outputs. These metrics offer more information about the properties and behaviors of soil.

This process is visually represented by the flowchart, which guarantees logical progression and clarity. This methodical technique allows for precise soil categorization and enables reliable estimates of soil capacitance and water content, making it appropriate for a variety of applications in geo-technical engineering, agriculture, and environmental monitoring.

Test results and analysis

The forecasts of soil capacitance and water content are compared in this part using two different approaches: machine learning (ML)-based predictions and manual computation using Excel-based datasets.

The first section provides a detailed analysis of the Excel results, emphasizing the direct correlations between capacitance values and percentages of water content in various soil types. By using established datasets, this method provides a baseline for evaluating prediction accuracy. The second section evaluates machine learning (ML) models trained on these datasets. Key metrics—including a confusion matrix and classification report—assess their performance in predicting soil water content and type. These metrics offer a thorough evaluation of the model's overall efficacy, recall, precision, and classification accuracy. To confirm the model's effectiveness and spot any inconsistencies, the ML-based predictions are contrasted with Excel-based outcomes, highlighting the benefits and drawbacks of both methods for soil analysis and classification.

A linear regression analysis was conducted using capacitance sensor values (analog signals converted into digital numbers) as the independent variable (x-axis) and the corresponding water content values, expressed as percentages, as the dependent variable (y-axis). Furthermore, a computational algorithm was developed to predict water content based on a given capacitance value. The regression analysis was carried out using Microsoft Excel.

Figure 7 illustrate the relationships between sensor values and moisture content for Sand 1, Sand 2, Sand 3, and Sand 4, respectively. For each sand type, two types of regression equations were developed: a linear equation and a logarithmic equation.

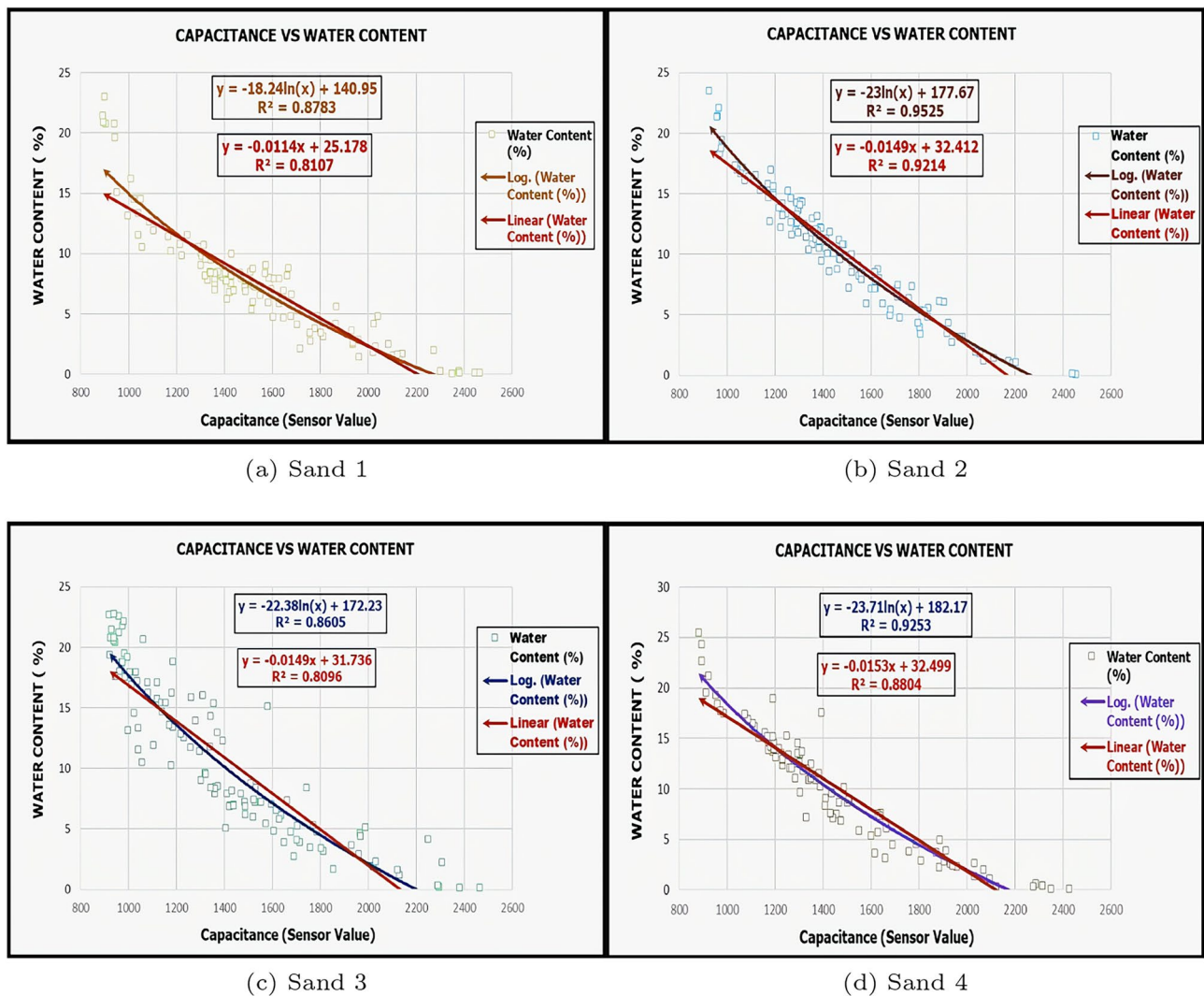


Fig. 7. Capacitance vs. Water Content in Sands. For (a) Sand 1 (Coarse Sand), (b) Sand 2 (Medium-Coarse Sand), (c) Sand 3 (Fine Sand), and (d) Sand 4 (Fine Sand), graphs illustrate the correlation between capacitance (sensor value) and water content (%). For every type of sand, linear and logarithmic trendlines are provided together with their equations and R^2 values.

For Sand 1 (Fig. 7a), the logarithmic equation $y = -18.24\ln(x) + 140.95$ ($R^2 = 0.8783$) outperforms the linear equation $y = -0.0114x + 25.178$ ($R^2 = 0.8107$).

Similarly, for Sand 2 (Fig. 7b), the logarithmic equation $y = -23\ln(x) + 177.67$ ($R^2 = 0.9525$) demonstrates a stronger correlation than the linear equation $y = -0.0149x + 32.412$ ($R^2 = 0.9214$).

In Sand 3 (Fig. 7c), the logarithmic equation $y = -22.38\ln(x) + 172.23$ ($R^2 = 0.8605$) again shows a better fit compared to the linear equation $y = -0.0149x + 31.736$ ($R^2 = 0.8096$).

Lastly, for Sand 4 (Fig. 7d), the logarithmic equation $y = -23.71\ln(x) + 182.17$ ($R^2 = 0.9253$) outperforms the linear equation $y = -0.0153x + 32.499$ ($R^2 = 0.8804$).

The negative coefficients in the equations indicate that a decrease in water content corresponds to an increase in sensor values. Meanwhile, the intercept term represents the estimated water content when the sensor capacitance is zero. Although this intercept may lack direct physical significance, it is included for mathematical completeness. The coefficient of determination (R^2) obtained from the regression analysis demonstrates a strong correlation between the sensor values and the corresponding water content for all sands. For all four sand types, the logarithmic equations consistently exhibit higher R^2 values than their linear counterparts. This suggests that logarithmic models are more effective in capturing the relationship between sensor values and moisture content, making them the preferred choice for analysis. Therefore, the logarithmic equation is used as the calibration model for this device.

Testing and validation

Using a spreadsheet program, trendline fitting of the experimental data was utilized to develop the linear and logarithmic equations used for soil moisture prediction. Because the logarithmic equation was better able to

Sample No	Device reading (%)	Lab testing value (%)	Accuracy(%)	Average accuracy(%)
1	2.85	3.24	88.00	94.47
2	5.22	5.20	99.61	
3	6.40	6.16	96.10	
4	7.95	7.41	92.71	
5	8.74	8.40	95.95	

Table 2. Device accuracy assessment via field and lab comparisons using a linear equation.

Sample NO	Device reading (%)	Lab testing value (%)	Accuracy(%)	Average accuracy(%)
1	3.03	3.24	93.52	96.49
2	4.96	5.20	95.38	
3	5.97	6.16	96.92	
4	7.39	7.41	99.73	
5	8.14	8.40	96.90	

Table 3. Device accuracy assessment via field and lab comparisons using a logarithmic equation.

capture nonlinear correlations between device readings and lab testing values, it attained a little greater accuracy (96.49%) than the linear equation (94.47%). The logarithmic model showed more consistency, with accuracies between 93.52% and 99.73%, while the linear model performed well, with accuracies ranging from 88.00% to 99.61%, as indicated in Tables 2 and 3. These findings validate the dependability of the suggested IoT system for practical uses by demonstrating the logarithmic model’s stability and applicability for accurate soil moisture prediction.

Tables 2 and 3 show field data collected from the Rajshahi Airport runaway, where device readings were systematically compared with laboratory test results to evaluate accuracy and reliability.

Note: The accuracy of soil moisture prediction using a linear regression model is evaluated by comparing IoT device readings with lab-tested values. The linear model provides a baseline for assessing prediction performance but may not fully capture nonlinear soil moisture variations.

Note: The logarithmic regression model enhances prediction accuracy by capturing nonlinear relationships in soil moisture variations. This approach reduces error compared to linear regression, making it more suitable for real-world applications where soil moisture behavior is complex.

To create reliable predictive models for the classification of soil types and water content, the ML-based approach first trained data on soil datasets. Using test data, soil type and water content predictions were developed following the training phase. A confusion matrix (Fig. 8) and a classification report (Table 4) that emphasized important metrics including accuracy, precision, recall, and F1-score were used to assess the ML models’ performance²⁹. Based on the input capacitance values, these findings shed light on the model’s capacity to accurately forecast soil type.

In order to show the potential benefits of machine learning for accurate and effective soil analysis, this section evaluates the consistency and dependability of the predictions by contrasting the manual and ML-based results.

A comparison of the expected and actual water content using linear (blue) and polynomial (red) regression models is shown in Fig 9. The ideal 1:1 relationship, where model predictions and actual values match exactly, is shown by the diagonal dashed black line. Although there are noticeable variations, particularly at lower and higher water content levels, both models generally follow the optimum trend. In comparison to the linear model, the polynomial model shows a comparatively tighter grouping of points along the 1:1 line, indicating better prediction ability and less bias. This distribution demonstrates how well the polynomial regression model can reflect the dataset’s underlying nonlinear relationship.

The matrix indicates that the model showed high prediction skills for both Medium Coarse Sand and Fine Sand, despite the small number of misclassifications for Coarse Sand. In addition to the metrics shown in the classification report, this visualization offers a more thorough assessment of the model’s performance and identifies possible areas for development.

Table 4 presents the classification report for soil sample prediction, detailing precision, recall, and F1-score for different sand types, along with overall accuracy and averaged performance metrics.

Note: The classification report evaluates model performance in predicting soil types. Precision measures the proportion of correctly identified positive cases among predicted positives, recall assesses how many actual positive cases were correctly identified, and the F1-score balances precision and recall. The support column indicates the number of instances per class, providing insight into dataset distribution and model reliability.

The model did remarkably well for Fine Sand, attaining flawless precision, recall, and F1-score, according to the classification report. With a high F1-score and flawless recall, Medium Coarse Sand also showed excellent performance. The recall was high for Coarse Sand, though, suggesting that the model only occasionally misclassified a few occurrences that belonged to this group. The model’s overall accuracy was 97.73%, and its weighted and macro averages showed balanced performance for all soil types. These findings highlight how well

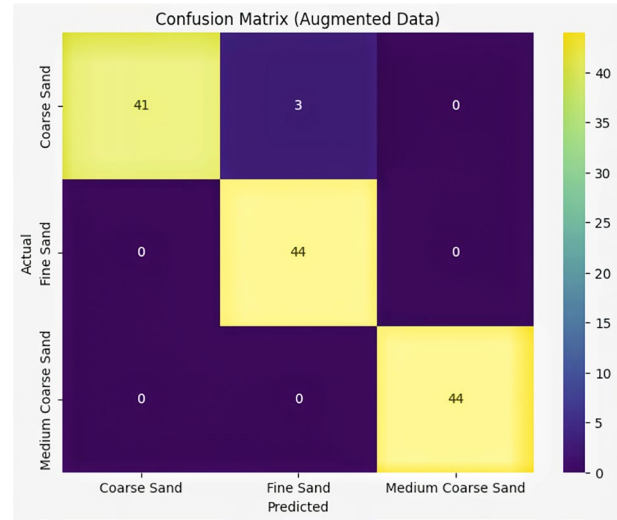


Fig. 8. Confusion matrix showing how well the machine learning model classified three different types of sand—coarse, fine, and medium-coarse—using supplemented data. Rows reflect actual labels, whereas columns represent anticipated labels. While incorrect classifications are displayed in the off-diagonal cells, correct classifications are displayed along the diagonal. With only a few misclassifications (three cases of coarse sand being forecasted as fine sand, for example), the model demonstrated great accuracy. The number of forecasts in each category is shown by the color intensity.

Class	Precision	Recall	F1-score	Support
Coarse sand	1.00	0.93	0.96	44
Fine sand	0.94	1.00	0.97	44
Medium coarse sand	1.00	1.00	1.00	44
Accuracy			0.98	132
Macro-avg	0.98	0.98	0.98	132
Weighted avg.	0.98	0.98	0.98	132

Table 4. Classification report of soil sample prediction using a polynomial regression model in machine learning.

the model classifies soil types; however, there is still an opportunity for improvement in some areas, such as coarse sand.

Conclusion

An AI-driven polynomial regression model and a capacitance-based soil moisture sensor are combined in this study to provide a reliable and affordable method for real-time soil moisture assessment and soil type categorization. Utilizing the sensor’s non-destructive, in-situ measurement capabilities, the system showed great suitability for geotechnical engineering applications like foundation design and slope stability studies. The nonlinear relationship between capacitance values and moisture content was well captured by the polynomial regression model, which performed better than the linear model with an accuracy of 96.49% as opposed to 94.47%. To further confirm this nonlinear pattern, a logarithmic regression model was employed.

With an overall accuracy of 98%, the Random Forest classifier demonstrated remarkable accuracy in classifying soil types. All three tested classes—Coarse Sand (Precision: 1.00, Recall: 0.93, F1-score: 0.96), Fine Sand (Precision: 0.94, Recall: 1.00, F1-score: 0.97), and Medium Coarse Sand (Precision: 1.00, Recall: 1.00, F1-score: 1.00)—showed high consistency with the classifier. All evaluation indicators had weighted and macro averages of 0.98, which further supported the model’s generalizability and dependability. These findings demonstrate the model’s ability for precise and effective soil categorization and water content estimation under real-world field circumstances.

The findings highlight the importance of integrating capacitance-based sensors with machine learning algorithms to increase soil monitoring systems’ accuracy. To improve prediction performance and increase real-world applicability in civil infrastructure, agriculture, and environmental monitoring, it is essential to choose regression models that take nonlinear behavior into consideration.

It is still to be confirmed how well the model performs on more complex soils like clay, silt, and high-organic-content soils, even though it showed excellent accuracy across the investigated sand types—fine sand, medium coarse sand, coarse sand, and sand-silt mixture. In order to enable wider generalization, future work

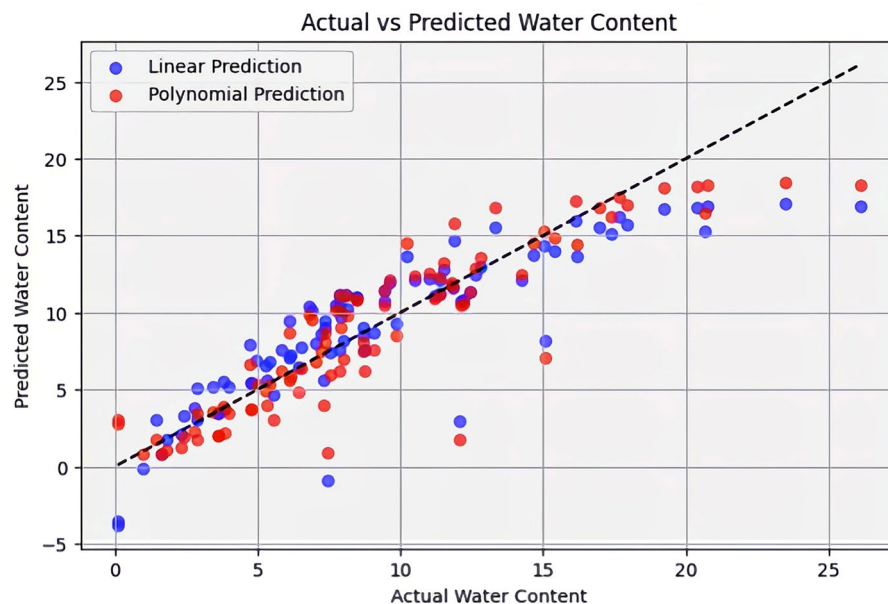


Fig. 9. Using linear (blue) and polynomial (red) regression models, the actual and predicted water content values are compared. The ideal 1:1 connection, where predicted and actual values are exactly equal, is represented by the dashed black line. The models' prediction performance is shown by the distribution of points around the line.

will concentrate on enlarging the dataset to cover various soil types. The effects of additional variables, including mineralogy, clay concentration, water chemistry, temperature, and salinity, will be investigated further. These improvements will further improve system accessibility and unlock more value for geotechnical, agricultural, and environmental applications when incorporated into clear, user-friendly applications.

Data availability

The dataset generated during this study is not publicly available but can be obtained from the corresponding author upon reasonable request.

Received: 13 February 2025; Accepted: 9 September 2025

Published online: 13 October 2025

References

- Sheng, Z. et al. A portable pull-out soil profile moisture sensor based on high-frequency capacitance. *Sensors* **23**(8), 3806 (2023).
- Nasirov, T. & Jabborov, K. Microprocessor moisture measuring device of bulk materials based on the capacitance transformer with cylindrical electrodes. In: *E3S Web of Conferences*, 386, 03004 (2023). EDP Sciences
- Pramanik, M. et al. Evaluation of capacitance-based soil moisture sensors in iot based automatic basin irrigation system (2023)
- Deng, X. et al. A calibration-free capacitive moisture detection method for multiple soil environments. *Measurement* **173**, 108599 (2021).
- Teixeira, J. & Santos, R. Exploring the applicability of low-cost capacitive and resistive water content sensors on compacted soils. *Geotech. Geol. Eng.* **39**(4), 2969–2983 (2021).
- Kanso, T., Gromaire, M.-C., Ramier, D., Dubois, P. & Chebbo, G. An investigation of the accuracy of ec5 and 5te capacitance sensors for soil moisture monitoring in urban soils-laboratory and field calibration. *Sensors* **20**(22), 6510 (2020).
- Al, J. & Smith, J. Low-cost soil moisture sensors' assessment for their accuracy after calibration through the gravimetric method. *Journal of Soil Research*. **56**(1), 353–369 (2024).
- Wu, S., Zhang, B., Tian, Y., Zhou, S. & Ma, H. A grain moisture model based on capacitive sensor. In: *Journal of Physics: Conference Series*, 1074, 012120 (2018). IOP Publishing
- Nathk, D. & Ramanathan, P. A simplified and portable capacitance sensor to measure moisture content of paddy. *Gazi Univ. J. Sci.* **31**(3), 821–829 (2018).
- Domínguez-Niño, J. M., Oliver-Manera, J., Arbat, G., Girona, J. & Casadesús, J. Analysis of the variability in soil moisture measurements by capacitance sensors in a drip-irrigated orchard. *Sensors* **20**(18), 5100 (2020).
- Silva, C. R. et al. Calibração de uma sonda de capacitância em um argissolo. *Sci. Agricola* **64**(6), 636–640 (2007).
- Mittelbach, H., Casini, F., Lehner, I., Teuling, A.J. & Seneviratne, S.I. (2011) Soil moisture monitoring for climate research: Evaluation of a low-cost sensor in the framework of the swiss soil moisture experiment (swissmex) campaign. *Journal of Geophysical Research: Atmospheres* **116**(D5)
- Deng, X. et al. A method of electrical conductivity compensation in a low-cost soil moisture sensing measurement based on capacitance. *Measurement* **150**, 107052 (2020).
- Wang, W., Zhao, K., Zhang, P., Bao, J. & Xue, S. Application of three self-developed ect sensors for monitoring the moisture content in sand and mortar. *Construct. Build. Mater.* **267**, 121008 (2021).
- Surya, S. G. et al. An in-field integrated capacitive sensor for rapid detection and quantification of soil moisture. *Sensors Actuators B: Chem.* **321**, 128542 (2020).

16. Archer, N., Rawlins, B., Machant, B., Mackay, J. & Meldrum, P. Approaches to calibrate in-situ capacitance soil moisture sensors and some of their implications. *SOIL Discussions* **2016**, 1–29 (2016).
17. Zabolotnyi, O., Zabolotnyi, V. & Koshevoy, N. Capacitive water-cut meter with robust near-linear transfer function. *Computation* **10**(7), 115 (2022).
18. Joaquin, A. C., Ramos, M. & Martinez, R. C. Development and performance evaluation of a grain probe moisture meter for paddy. *Int. J. Environ. Sci. Dev.* **10**(10), 350–355 (2019).
19. Suranata, I. W. A. et al. Arsitektur moisture meter dengan capacitive sensing dan serverless iot untuk hidroponik fertisasi. *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)* **5**(2), 292–300 (2021).
20. Smith, W.B., Jung, H.-S. & Association, W.D.K., et al. Effect of high temperature drying on moisture content determination with electronic meters (1993).
21. Kostadinović, D. M., Dimitrijević-Jovanović, D., Vučičević, B. S., Jovanović, M. P. & Konjikušić, S. M. Smart capacitive moisture sensor calibration in mineral wool and green roof soil substrate. *Therm. Sci.* **25**, 1827–1836 (2021).
22. Patil, K.R., Eickhoff, S.B. & Langner, R. Predictive data calibration for linear correlation significance testing. arXiv preprint [arXiv:2208.07081](https://arxiv.org/abs/2208.07081) (2022).
23. Adla, S. et al. Laboratory calibration and performance evaluation of low-cost capacitive and very low-cost resistive soil moisture sensors. *Sensors* **20**(2), 363 (2020).
24. Bogaena, H. R., Huisman, J. A., Oberdörster, C. & Vereecken, H. Evaluation of a low-cost soil water content sensor for wireless network applications. *J. Hydrol.* **344**(1–2), 32–42 (2007).
25. Engineers, L.M. Interfacing Capacitive Soil Moisture Sensor with Arduino. <https://tinyurl.com/5bb2bydd>. Accessed: 2024-12-30 (2024).
26. Espressif. ESP32 Technical Reference Manual. https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf. Accessed: 2024-12-30 (2024).
27. Haldar, S., Gol, S., Mondal, A. & Banerjee, R. Iot-enabled advanced monitoring system for tubular batteries: Enhancing efficiency and reliability. *E-Prime-Adv. Elect. Eng., Elect. Energy* **9**, 100709 (2024).
28. Khatun, M.S. & Khan, I. Bangla counterfeit news identification: Using the power of bert. In: 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON), pp. 518–522 (2024). <https://doi.org/10.1109/PEEIACON63629.2024.10800650>
29. Khan, I. et al. Leveraging the robust capability of modified multi-layer gated recurrent units for fake news detection. In: 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), pp. 986–991 (2024). <https://doi.org/10.1109/ICEEICT62016.2024.10534407>

Acknowledgements

The authors acknowledge the support of the Department of Civil Engineering, Rajshahi University of Engineering & Technology, Bangladesh, for providing laboratory facilities.

Author contributions

Md. Mahmud Sazzad (1st author): Conceptualization, supervision, review. Tanvir Ahmed (2nd author): Data curation, IoT device development, methodology, original draft. Golam Kibria (3rd author): Visualization, validation. Ishmam Khan (4th author): Machine learning model development, editing. All authors reviewed the manuscript.

Funding

The authors did not receive any financial support for the research, authorship, or publication of this article.

Declarations

Competing interests

The authors declare no conflicts of interest relevant to this study.

Ethical approval

This study does not involve human participants or animals, and ethical approval is not applicable.

Consent to participate

Not applicable, as no human participants were involved.

Consent for publication

All authors consent to the publication of this manuscript.

Additional information

Correspondence and requests for materials should be addressed to T.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025