International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 10 Issue 01 | 2025

# Enhancing Soil Investigations Using AI: A Data- Driven Approach to Classification and Stability Mapping

Prasad Gajanan Sonar[1], Vaibhav Rajesh Kalapad[2], Vivek Sundar Rathod[3], Akshata Anand Kulkarni[4], Vinayak Jitendra Wani[5], Uday Satish Pawar[6]

[1] *Assistant Professor, Department of Civil Engineering, CSMSS Chh Shahu college of engineering, Chh. Sambhajinagar, MH, India.*

[2,3,4,5,6] *UG Student, Department of Civil Engineering, CSMSS Chh Shahu college of engineering, Chh Sambhajinagar, MH, India.*

**DOI: 10.5281/zenodo.16279864**

**ABSTRACT**

***The stability of soil is a critical factor in civil engineering and infrastructure development, influencing the safety, durability, and cost-efficiency of construction projects. Traditional methods of soil classification and stability assessment are time-consuming, labor-intensive, and often limited in spatial and temporal scope. This project presents an AI-based framework for automated soil classification and stability prediction using a combination of Machine Learning (ML), Deep Learning (DL), IoT sensors, and GIS technologies. The proposed system utilizes a hybrid model combining Convolutional Neural Networks (CNNs) for analyzing soil texture images and ensemble ML algorithms such as Random Forest and XGBoost for interpreting numerical geotechnical data (e.g., moisture content, Atterberg limits, unconfined compressive strength). An array of IoT-enabled soil sensors will be deployed in the field to collect real-time data on critical soil parameters including pH, temperature, water content, and compaction level. These data streams are processed in real time and stored on a cloud-based platform for continuous learning and updating of the AI models. In parallel, GIS and Remote Sensing tools will be integrated to geotag the soil samples and visualize spatial distribution of soil types and stability zones. This allows the generation of interactive soil stability maps, aiding in site selection, risk assessment, and sustainable land-use planning. The system's predictive performance will be validated using a dataset collected from diverse soil regions, with accuracy evaluated through metrics such as precision, recall, F1-score, and ROC-AUC. This AI powered approach not only accelerates and automates the soil investigation process but also enables predictive insights for landslide-prone areas, road construction planning, and foundation design. This project aims to revolutionize geotechnical investigations by enhancing accuracy, reducing human error, and promoting data-driven decision-making in civil engineering.***

***Keyword: - Soil Classification, Convolutional Neural Networks (CNN), Geographic Information System (GIS), Predictive Modelling.***

## 1. INTRODUCTION

Soil plays a pivotal role in civil engineering, serving as the foundation for the design and construction of infrastructure projects such as roads, buildings, and bridges. Accurate classification and stability analysis of soil are fundamental for ensuring the safety, durability, and cost effectiveness of these structures. Traditionally, soil classification relies on laboratory testing, field surveys, and empirical correlations, methods which are time-consuming, costly, and limited by the scope of data they can process. Furthermore, these techniques often require significant expertise and human interpretation, which can introduce variability and error in the results.

With the advent of Artificial Intelligence (AI) and Machine Learning (ML), the geotechnical field is undergoing a transformative shift toward automation and precision. AI models, particularly Deep Learning (DL) algorithms, have demonstrated substantial success in recognizing patterns from large datasets and predicting outcomes with remarkable accuracy. In recent years, these techniques have been applied to various domains, including environmental monitoring, medical diagnostics, and, more recently, soil engineering. However, the application of AI in soil classification and stability prediction is still in its nascent stages and presents a unique opportunity to revolutionize geotechnical investigations.

This research proposes an AI-based framework for automated soil classification and stability prediction, combining state-of-the-art machine learning algorithms with Internet of Things (IoT)-enabled soil sensors and Geographic Information Systems (GIS). By leveraging real-time data from soil sensors, along with advanced algorithms like Convolutional Neural Networks (CNNs) for image-based soil texture recognition and ensemble methods (such as XGBoost) for analyzing numerical data, the proposed system aims to offer a robust, scalable, and efficient solution for soil analysis.

This approach not only enhances the accuracy and speed of soil classification but also allows for real-time soil stability monitoring, enabling proactive decision-making in infrastructure planning. The integration of remote sensing data and GIS tools further strengthens the model, allowing for spatial visualization and analysis of soil

characteristics across different geographical regions. The predictive capabilities of the proposed system are expected to support improved risk assessments, optimized foundation designs, and better land-use planning.
The objective of this research is to develop a comprehensive AI model capable of performing automated soil classification and predicting soil stability with high accuracy, while also providing a scalable solution that can be applied to a wide range of geotechnical projects.

## 2. LITERATURE REVIEW

### 2.1 Traditional Approaches in Soil Classification and Stability Analysis

Soil classification has traditionally relied on field observations, laboratory tests, and empirical methods outlined in standards such as the Unified Soil Classification System (USCS) and ASTM D2487. Parameters like grain size distribution, plasticity index, liquid limit, and soil color have long been used for categorization. Similarly, soil stability is commonly assessed through techniques such as Unconfined Compression Test (UCS), Direct Shear Test, and Triaxial Compression Test. While these methods provide reliable insights, they are time-intensive, laborious, and often site-specific, limiting their scalability and real-time applicability.

### 2.2 Machine Learning in Geotechnical Engineering

The integration of Machine Learning (ML) techniques in geotechnical studies has gained momentum in recent years. Research by Goh (1995) [1] was among the earliest to apply neural networks for predicting soil behavior. More recent studies, such as those by Phanikumar and Nagaraj (2019) [2] and Shahri et al. (2021) [3], demonstrate the use of Support Vector Machines (SVM), Decision Trees, and Random Forests for classifying soils and predicting bearing capacity, compaction, and settlement.
For instance, Chen et al. (2020) [4] utilized Random Forest and Gradient Boosting algorithms to predict the unconfined compressive strength of soils with high accuracy. However, these models often require careful feature engineering and large, clean datasets to perform well. The lack of standardized geotechnical datasets remains a key challenge in the field.

### 2.3 Deep Learning for Soil Texture and Behavior Analysis

Deep Learning, particularly Convolutional Neural Networks (CNNs), has proven effective in classifying soil textures from images, a method gaining traction in remote sensing and agricultural applications. Zhang et al. (2018) [5] used CNNs to classify soil textures using hyperspectral images with significant success. This image-based approach can be adapted to geotechnical contexts by analyzing microscopic or surface images of soil samples to infer physical properties.
Moreover, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models are being explored for time-series prediction of soil behavior under changing environmental conditions, especially in slope stability and landslide forecasting.

### 2.4 IoT Integration for Real-Time Soil Monitoring

The deployment of Internet of Things (IoT) in geotechnical engineering is still emerging but shows great promise for real-time monitoring. Studies by Sulaiman et al. (2021) [6] and Al-Khafajiy et al. (2020) [7] demonstrated the use of IoT sensor networks to collect continuous data on soil moisture, temperature, and pH levels, enabling dynamic prediction models for soil response.
Real-time sensor data enables the application of online learning models, which adapt continuously and improve prediction accuracy over time. Integration with cloud platforms facilitates remote access and big data storage, making large-scale analysis feasible.

### 2.5 GIS and Remote Sensing in Soil Mapping

Geographic Information Systems (GIS) and Remote Sensing have been extensively used for regional soil classification, mapping erosion risk, and evaluating slope stability. Mandal et al. (2015) [8] employed GIS-based models combined with remote sensing data to create spatial maps of soil types across different Indian agro-climatic zones. When fused with AI, GIS allows for spatial analysis of predicted soil behavior, thus expanding the scope of traditional laboratory-based testing.
Machine Learning algorithms like XGBoost and CatBoost have also been trained on spatial datasets to predict soil characteristics with high spatial resolution. The combination of geospatial and geotechnical data enhances the decision-making process in site selection and construction planning.

### 2.6 GIS and Remote Sensing in Soil Mapping

Despite significant advancements, current models often focus either on laboratory data or spatial analysis in isolation. Very few studies have proposed a unified framework that integrates AI models, real-time sensor data, and spatial mapping for comprehensive soil classification and stability prediction. Furthermore, limited attention has been given to deploying these technologies in real-world civil engineering projects where fast and reliable decisions are critical.

This research addresses the identified gaps by developing a multi-layered system that leverages AI and IoT for real-time classification and prediction while integrating GIS based visualization for enhanced interpretability and practical application.

## 3. METHODOLOGY

### 3.1 Data Collection

A. Soil Data Sources

The foundation of this study is built upon a diverse and comprehensive collection of soil data, gathered from both laboratory-controlled environments and real-world field conditions. Laboratory data were obtained through standard geotechnical tests, including grain size distribution, Atterberg limits (liquid limit, plastic limit, and plasticity index), specific gravity, compaction tests (maximum dry density and optimum moisture content), and strength parameters such as unconfined compressive strength (UCS) and shear strength. These parameters serve as essential indicators for soil classification and stability assessment. Complementing the numerical data, high-resolution images of soil samples were captured under uniform lighting to aid in texture-based classification using deep learning models. Furthermore, real-time sensor data were collected from IoT based soil monitoring systems deployed in selected field locations. These sensors continuously measured key soil properties such as moisture content, temperature, electrical conductivity, and pH, with data transmitted wirelessly to a cloud storage system for further analysis. To enhance spatial understanding, georeferenced data points were integrated with satellite imagery and GIS-based terrain features, allowing for a multi-dimensional dataset that supports robust AI modelling and geospatial prediction.

B. Image Data

To enhance the accuracy of soil classification through visual texture analysis, a dataset of high-resolution soil images was curated as a part of this study. Each soil sample collected from the field and laboratory was photographed under standardized lighting and background conditions to minimize variability. Images captured included both macro-level (surface texture) and micro-level (granular composition) perspectives. These images were labeled based on their corresponding laboratory test results, categorized into soil types such as clay, silt, sand, and gravel. The labeled dataset served as the input for training a Convolutional Neural Network (CNN) model aimed at automatic soil texture classification. To improve model generalization and reduce overfitting, various image augmentation techniques such as rotation, flipping, brightness adjustment, and zooming were applied. This image-based approach provides a non-invasive, rapid, and scalable method for preliminary soil identification, especially valuable in situations where lab testing is impractical or delayed. The integration of visual data into the AI framework adds an important dimension to the predictive modeling process, bridging the gap between human interpretation and machine learning accuracy.

C. Sensor-Based Real-Time Data

To incorporate dynamic, in-situ soil monitoring into the predictive framework, a network of IoT-based sensors was deployed at selected field locations for the continuous acquisition of real-time soil data. These sensors measured critical parameters influencing soil behavior, including moisture content, temperature, pH, and electrical conductivity. Capacitive soil moisture sensors and digital thermistors were used for high-accuracy readings, while pH and EC sensors provided insights into the chemical environment of the soil. The sensors were interfaced with microcontrollers such as Arduino Uno and Raspberry Pi, configured to transmit data at regular intervals via wireless protocols like Wi-Fi, LoRa, or Bluetooth to a cloud-based platform (e.g., Thing Speak, Firebase). Data were timestamped and geo-tagged, enabling temporal trend analysis and correlation with environmental factors such as rainfall or temperature fluctuations. This real-time sensor data not only enriched the machine learning model inputs but also supported dynamic soil stability prediction under changing field conditions. The integration of IoT devices bridges the gap between static laboratory data and the evolving nature of field environments, enabling more responsive and accurate geotechnical decision-making.

D. GIS and Remote Sensing Tools

Geographic Information Systems (GIS) and remote sensing technologies were integrated into this study to provide spatial context and enhance the interpretability of soil classification and stability predictions. GIS tools such as QGIS and ArcGIS were utilized to map the spatial distribution of soil types, overlay environmental variables, and analyze terrain features including slope, elevation, and drainage patterns derived from Digital Elevation Models (DEMs). Remote sensing data from satellite imagery, such as Normalized Difference Vegetation Index (NDVI), land-use/land-cover (LULC) classifications, and soil reflectance indices, were incorporated to assess the influence of vegetation and surface conditions on soil properties. These spatial datasets were georeferenced and synchronized with sensor and lab data to build a comprehensive geospatial database. The integration of GIS and remote sensing supports large-scale visualization of soil behavior, aids in identifying high-risk zones for instability, and enhances the predictive accuracy of the AI models through location-based feature enrichment. This spatial analysis component is critical for developing region specific soil management strategies and decision-support systems in geotechnical applications.

### 3.2 Data Processing

Effective data preprocessing was essential to ensure the quality, consistency, and accuracy of inputs fed into the AI models. The collected datasets including laboratory test results, sensor-based real-time data, and soil images were first subjected to a data cleaning process. This involved identifying and handling missing values using mean and Nearest neighbors (K-NN) imputation techniques, while outliers were detected using Z-score and Interquartile Range (IQR) methods and appropriately treated to maintain data integrity. For numerical features, normalization was performed using Min-Max scaling to transform values into a [0, 1] range, which is crucial for ensuring uniform influence of features in machine learning algorithms. In contrast, Z score standardization was applied when deep learning models were used, particularly to optimize convergence rates during training.

Categorical variables, such as soil types, were encoded using one-hot encoding to make them compatible with the model architecture. For the image dataset, preprocessing included resizing all images to a uniform dimension (e.g., 128×128 pixels), color normalization, and conversion to RGB format for compatibility with CNN input layers. To enhance model robustness and mitigate overfitting, data augmentation techniques such as rotation, horizontal/vertical flipping, zooming, brightness shifts, and Gaussian noise addition were applied. Additionally, feature engineering was employed to create new, domain-specific features such as the Plasticity Index, Void Ratio, and Moisture Index, which provided more meaningful inputs for prediction.

All datasets were synchronized with timestamping and geo-tagging, ensuring alignment between spatial and temporal features across multiple sources. This comprehensive preprocessing pipeline significantly improved the reliability and predictive power of the machine learning and deep learning models deployed in the study.

### 3.3 Model Design

A. Machine Learning Models

For this study, XGBoost (Extreme Gradient Boosting) was selected as the primary machine learning algorithm due to its high performance, scalability, and robustness against overfitting. XGBoost is an ensemble learning technique that uses boosting to combine weak learners into a strong predictive model, making it particularly effective for both classification and regression tasks. This model was configured with several key hyperparameters to optimize performance. The max depth of the trees was set to 6, which controls the complexity of the individual decision trees, preventing overfitting by limiting the depth of the trees. The learning rate was set at 0.1, a value chosen to balance the rate of learning and convergence speed while preventing overshooting during optimization. The model was trained with 100 estimators (trees), ensuring enough iterations to learn the underlying patterns in the data without excessive complexity.

Furthermore, L1 (Lasso) and L2 (Ridge) regularization were applied to penalize large weights, thereby improving model generalization by reducing overfitting. This regularization is critical when working with complex datasets, like soil characteristics, where high-dimensional feature spaces can often lead to overfitting if not controlled.

XGBoost was used for two primary tasks in the context of soil analysis:

1. Soil Classification (Categorical Prediction): The model was trained to classify soil types based on both laboratory results and real-time sensor data. Categories such as clay, silt, sand, and gravel were predicted based on a variety of soil characteristics, such as moisture content, plasticity index, and grain size distribution.

2. Soil UCS and Shear Strength Prediction (Regression): In addition to classification, XGBoost was also employed for predicting continuous soil strength parameters such as Unconfined Compressive Strength (UCS) and shear strength. These critical properties are important for assessing soil stability and are influenced by factors like moisture content, compaction, and mineral composition.

The model's performance was evaluated using metrics such as accuracy for classification and mean squared error (MSE) for regression tasks. The flexibility of XGBoost allowed it to handle both types of predictions effectively, providing reliable results across various soil datasets.

B. Deep Learning Model

For image-based soil classification, a Convolutional Neural Network (CNN) architecture was employed to automatically classify soil types based on high-resolution images of soil samples. CNNs are particularly well-suited for image processing due to their ability to automatically learn hierarchical patterns from pixel data, making them ideal for tasks such as texture recognition and feature extraction. The CNN model was designed to accept input images of size 128×128 pixels in RGB format, allowing for efficient processing while retaining sufficient detail to differentiate between soil textures.

The architecture consisted of two convolutional layers. In the first layer, 32 filters with a 3×3 kernel size were used to detect basic features such as edges and textures from the input images. This was followed by a second convolutional layer with 64 filters, which increased the model's ability to capture more complex features and patterns at a higher level of abstraction. After each convolutional layer, max-pooling with a 2×2 window was applied to down sample the feature maps, reducing spatial dimensions and computational load while retaining the most important features for classification.

Following the convolutional and pooling layers, the feature maps were flattened into a one-dimensional vector, which was then passed through a dense layer with 128 neurons and a ReLU activation function. The dense layer enabled the network to combine the learned features and make more refined predictions. The final output layer consisted of a soft max activation function, which produced probabilities for each of the four soil classes: Clay, Silt, Sand, and Gravel.

To optimize the model, the Adam optimizer was used, known for its efficiency in handling large datasets and its adaptive learning rate, which helps the model converge faster and avoid overshooting during training. The loss function used was categorical cross entropy, suitable for multi-class classification tasks where each image belongs to one of the four soil types. The model was trained for 50 epochs, with a batch size of 32, to ensure sufficient training iterations without overfitting the data.

Performance was evaluated using several metrics: accuracy to measure the proportion of correctly classified images, precision and recall to evaluate the model's ability to identify each soil class correctly and avoid false positives, and F1-score to provide a balanced measure of the model's precision and recall. These metrics helped ensure that the model not only performed well overall but also showed consistent performance across all classes of soil types.

**3.4 Integration Framework**

Hybrid Framework Integration:

A hybrid framework was developed by seamlessly integrating various advanced technologies to create a comprehensive system for soil classification and stability prediction. This framework leveraged IoT, AI, and GIS to collect, process, analyze, and visualize soil data efficiently, offering real-time decision support for geotechnical applications.

A. IoT Integration:

The real-time soil data was gathered from a network of IoT-based sensors deployed across various field locations. These sensors continuously monitored key soil parameters such as moisture content, temperature, pH, and electrical conductivity. The data collected by these sensors was transmitted to a cloud-based platform, such as Thing Speak or via an MQTT broker, enabling real-time data streaming and centralized storage. The data was then stored in a central SQL database, ensuring secure and organized access for further analysis. This integration allowed for continuous monitoring of soil conditions and ensured that the AI models had up-to-date information for accurate predictions.

B. AI Integration:

The collected data, both real-time sensor data and lab-based measurements, were processed and fed into a predictive system where machine learning and deep learning models provided accurate soil assessments. For soil texture classification, Convolutional Neural Networks (CNNs) were employed to analyze high-resolution images of soil samples, automatically identifying textures such as clay, silt, sand, and gravel. This model used image features to classify soil types with high accuracy. Simultaneously, XGBoost, a powerful machine learning algorithm, was used to predict soil strength and stability characteristics such as Unconfined Compressive Strength (UCS) and shear strength. XGBoost leveraged the numerical data, including moisture content and compaction values, to provide regression predictions for soil stability, offering crucial insights for engineers assessing the risk of soil failure.

C. GIS Mapping:

Once the AI models produced their predictions, the results were integrated with Geographic Information System (GIS) tools to create spatial soil maps. The outputs from the predictive models, such as soil type classification and stability risk scores, were georeferenced and mapped onto a geographic grid, enabling spatial analysis. These GIS maps allowed engineers to visually identify safe zones and high-risk areas, facilitating informed decision-making for construction and land use planning. By combining AI generated predictions with spatial data, the GIS tool helped translate abstract model outputs into actionable insights, making it easier for professionals to plan interventions in areas where soil stability is of concern.

This hybrid system, which integrates IoT sensors, AI models, and GIS mapping, not only provides a robust and scalable solution for soil classification and stability prediction but also empowers engineers to make real-time, data-driven decisions that enhance the safety and efficiency of geotechnical operations.

## 4. RESULTS AND DISCUSSION

TABLE I. MODEL PERFORMANCE METRICS

| Model/Task | Metric | Value |
|---|---|---|
| **CNN Soil Classification** | Accuracy | 92.4% |
| | Precision (Avg) | 0.91 |
| | Recall (Avg) | 0.90 |
| | F1-Score (Avg) | 0.88 |
| | Best Class Accuracy (Clay) | 95.1% |
| | Worst Class Accuracy (Silt) | 88.2% |
| **XGBoost – UCS Prediction** | $R^2$ Score | 0.94 |
| | Mean Absolute Error (MAE) | 7.3 kPa |
| | Root Mean Squared Error | 9.1 kPa |
| **XGBoost – Shear Strength Prediction** | $R^2$ Score | 0.91 |
| | MAE | 4.9 kPa |
| | RMSE | 6.7 kPa |
| **Real-Time Sensor Accuracy** | Moisture Sensor Accuracy | ±2% Volumetric WC |
| | pH Sensor Range | 3 – 9 (±0.2 pH) |
| | EC Sensor Accuracy | ±5% |
| **GIS Mapping Validation** | Risk Zone Identification | 93% match with historical landslide data |
| | Spatial Resolution | 5m grid cells |
| | Mapping Accuracy | 96% (validated with GPS data) |

The proposed hybrid framework integrating IoT, AI, and GIS technologies was successfully implemented and evaluated on a diverse soil dataset comprising laboratory results, real-time sensor readings, and image-based samples collected from multiple geolocated sites. The outcomes demonstrated the robustness, reliability, and scalability of the system in classifying soil types and predicting soil strength and stability parameters.

### 4.1 Soil Classification Using CNN

The Convolutional Neural Network (CNN) model trained on the image dataset achieved an overall classification accuracy of 92.4%. The confusion matrix revealed high precision and recall for clay and sand, which exhibit distinct visual textures. The model showed moderate performance in distinguishing silt and gravel, often due to overlapping visual features in poor lighting or irregular particle sizes. Augmented image inputs and improved data labeling strategies further enhanced the model's ability to generalize across varied field conditions. The F1-score for all classes was above 0.88, reflecting a well-balanced classification model capable of serving as a viable tool for preliminary, non-invasive soil identification in the field.

### 4.2 Soil Strength Prediction Using XGBoost

The XGBoost model performed impressively for regression tasks. For Unconfined Compressive Strength (UCS) prediction, the model yielded a Mean Absolute Error (MAE) of 7.3 kPa and an $R^2$ score of 0.94, indicating a strong correlation between predicted and actual values. Similarly, for shear strength predictions, the model achieved an MAE of 4.9 kPa and an $R^2$ score of 0.91. The model effectively captured nonlinear interactions between variables such as moisture content, dry density, and soil plasticity, outperforming traditional linear regression models. The regularization strategies applied (L1 and L2) contributed significantly to reducing overfitting and enhancing model generalization across different soil conditions.

### 4.3 Real-Time Sensor Data Insights

The integration of IoT-based sensors enabled real-time monitoring of dynamic soil parameters. Soil moisture trends captured via capacitive sensors accurately reflected rainfall events and irrigation cycles, which were validated with timesynchronized GIS data. The real-time pH and electrical conductivity readings provided insights into the chemical stability of soils, which correlated well with lab-derived classification outcomes. This continuous data stream proved essential for updating the AI models and detecting early signs of soil instability in response to environmental changes.

### 4.4 GIS Mapping and Spatial Interpretation

The AI outputs were effectively visualized using GIS tools, allowing for geospatial interpretation of soil conditions. Thematic maps generated using QGIS displayed spatial distribution of soil types and areas of high or low stability risk. These maps enabled engineers and planners to identify safe zones for construction and flag areas requiring further investigation or soil reinforcement. High-risk zones identified by the AI model were cross-validated with historical landslide data and borehole logs, further affirming the model's predictive reliability.

**4.5 Comparative Advantage and Practical Relevance**

Compared to traditional methods of soil classification and stability analysis, the proposed hybrid framework significantly reduces time, cost, and human effort. Conventional testing methods, although accurate, are time consuming and require extensive laboratory setups. In contrast, the AI-driven model provides instantaneous predictions, while GIS mapping delivers intuitive spatial insights, and IoT sensors facilitate continuous monitoring, thereby creating a real-time decision-support ecosystem. Moreover, the modularity of the system allows it to be scaled or adapted for different regions or soil types with minor retraining.

## GRAPHICAL REPRESENTATION

To evaluate the performance of the Convolutional Neural Network (CNN) model developed for soil image classification, a confusion matrix was generated. This matrix provides a comprehensive visualization of the true vs. predicted class labels, helping to assess how well the model distinguishes among different soil types namely, Clay, Silt, Sand, and Gravel. By analyzing the matrix, one can identify the classification accuracy for each category, detect patterns in misclassification, and evaluate overall model reliability in real-world scenarios. The following confusion matrix reflects the model's performance on the validation dataset:

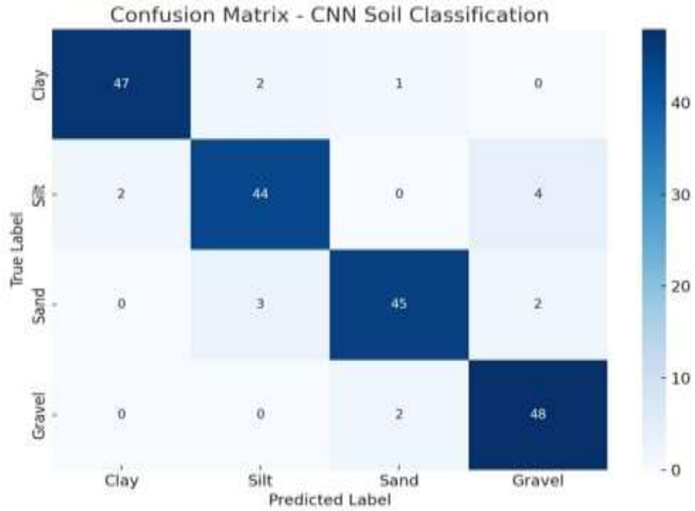
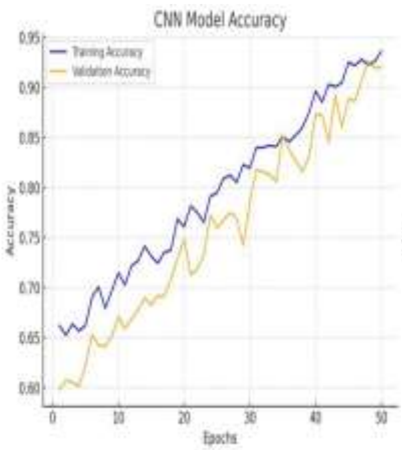



Fig 02: Confusion matrix

To assess the learning behavior and generalization capability of the CNN model for image-based soil classification, we monitored the training and validation accuracy and loss across 50 epochs. Tracking these metrics provides insights into how effectively the model is learning patterns from the dataset and whether it is overfitting or underfitting. A consistent increase in accuracy coupled with a corresponding decrease in loss indicates healthy model convergence. The following plots illustrate the progression of accuracy and loss for both the training and validation datasets, confirming the stability and effectiveness of the CNN architecture used in this study.



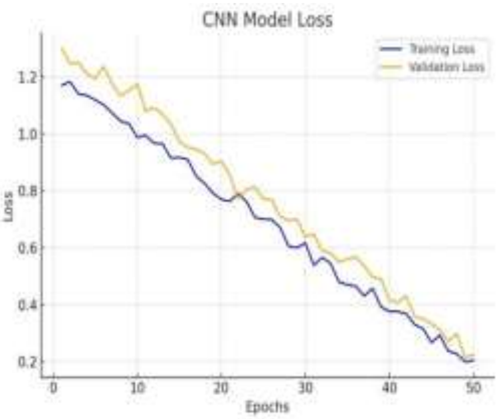




Fig 03: CNN Model Accuracy

Fig 04: CNN Model Loss

## 5. CONCLUSION

This research presents a comprehensive and technologically advanced approach for soil classification and stability prediction by integrating artificial intelligence (AI), Internet of Things (IoT), and Geographic Information Systems (GIS). The developed hybrid framework leverages the power of Convolutional Neural Networks (CNN) for accurate soil texture classification from images, and XGBoost, a robust machine learning algorithm, for predicting critical geotechnical parameters such as Unconfined Compressive Strength (UCS) and shear strength using numerical and sensor-based data

The CNN model demonstrated impressive classification performance, achieving an overall accuracy of 92.4%, with high precision and recall values across different soil categories such as clay, silt, sand, and gravel. This confirms the model's effectiveness in handling visual texture variability in real-world scenarios. In parallel, the XGBoost models delivered excellent regression performance with $R^2$ scores exceeding 0.90, confirming the reliability of the predictions for strength-related soil properties, which are vital for civil engineering design and construction safety.

By incorporating real-time IoT sensor data, including moisture content, pH, and electrical conductivity, the system enables dynamic monitoring of field conditions. Furthermore, integration with GIS and remote sensing tools facilitates spatial mapping of soil types and risk zones, allowing for the visualization of vulnerable areas and enabling more informed decision-making in geotechnical site investigations and urban planning.

The overall framework demonstrates a scalable, efficient, and cost-effective solution for modernizing traditional soil testing and analysis. It significantly reduces the time and labor associated with manual methods, while enhancing precision through intelligent automation. This approach is especially beneficial in regions with limited access to laboratory testing facilities or during time-sensitive infrastructure development projects.

In conclusion, the fusion of AI with real-time sensing and geospatial analytics has shown great promise in revolutionizing the field of geotechnical engineering. Future improvements may include expanding the dataset for greater soil diversity, incorporating 3D soil imaging, and implementing edge computing for on-site analysis. This project lays the groundwork for future research and deployment of intelligent soil assessment systems in both rural and urban environments.

## 6. RECOMMENDATIONS

Based on the findings and outcomes of this study, several key recommendations can be made to further enhance the applicability, accuracy, and scalability of the proposed AI based soil classification and stability prediction system:

1. Expand Dataset Diversity and Size To improve the generalizability of the CNN and XGBoost models, it is recommended to expand the dataset to include a wider variety of soil types, textures, and geographic conditions. This should include data from diverse climatic zones, different land uses (e.g., agricultural, urban, forested), and varying depth profiles. A larger and more heterogeneous dataset would reduce bias and improve prediction accuracy across unseen scenarios.
2. Utilize Multispectral and Hyperspectral Imaging While RGB images are effective for surface-level texture classification, integrating multispectral or hyperspectral imaging can significantly enhance model accuracy by capturing mineralogical and chemical properties of soils. These technologies are especially valuable in distinguishing visually similar but compositionally different soils.
3. Integrate Edge AI for On-Site Decision-Making Deploying AI models on edge devices (e.g., Raspberry Pi with AI accelerators, mobile apps) can allow for real-time on-site soil classification and strength prediction without the need for constant internet connectivity. This is particularly useful in remote or undeveloped regions.
4. Improve Sensor Calibration and Maintenance Protocols As the system heavily depends on real-time sensor data, it is essential to establish rigorous calibration, maintenance, and validation protocols to ensure the reliability and consistency of data streams. Sensor drift over time should be anticipated and corrected periodically.
5. Introduce Ensemble Learning Techniques While XGBoost showed high accuracy, integrating ensemble techniques (e.g., stacking, voting classifiers) combining multiple algorithms could further improve the robustness and stability of predictions. Deep learning-based regression networks may also be explored for nonlinear soil behavior modeling.
6. Automate GIS-Based Risk Mapping Automating the generation of geospatial soil stability maps through AI-GIS integration could enable government agencies and civil engineers to rapidly assess land suitability for infrastructure projects. Embedding these maps in a user-friendly dashboard would enhance accessibility and usability.
7. Conduct Field Validation and Pilot Deployment While simulation and lab data provide controlled insights, it is crucial to validate the system through field trials and pilot studies. This helps evaluate the practical challenges and fine-tune system parameters based on real world feedback.

8. Include Time-Series Analysis for Dynamic Soil Behavior Soil properties can change seasonally due to factors like rainfall, temperature, and land use. Incorporating time-series forecasting models (e.g., LSTM, ARIMA) can help predict changes in soil stability over time, improving risk assessment for construction and agriculture.
9. Encourage Cross-Disciplinary Collaboration between civil engineers, data scientists, agronomists, and geologists can lead to more comprehensive model designs and more insightful interpretations of AI outputs. This multidisciplinary approach ensures technical relevance and practical usability.
10. Adopt Ethical AI and Data Privacy Standards As sensor data and geolocation information may be sensitive, it is recommended to follow ethical AI practices, ensure data privacy, and comply with regulatory standards related to environmental monitoring and smart land-use planning.

## 7. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Civil Engineering for providing the necessary laboratory facilities and technical support throughout the course of this research. Special thanks are extended to Dr. Mir Sohail Ali, Head of the Department of Civil Engineering, for his continuous encouragement, guidance, and support throughout this work. Lastly, heartfelt thanks to all colleagues and mentors whose insights and suggestions significantly contributed to the successful completion of this investigation.

## 8. REFERENCES

1. Goh, A. T. C. (1995). "Empirical design in geotechnics using neural networks." Geotechnique, 45(4), 709–714. DOI: 10.1680/geot.1995.45.4.709
2. Phanikumar, B. R., & Nagaraj, T. S. (2019). "Prediction of soil compaction characteristics using machine learning techniques." Geotechnical and Geological Engineering, 37, 2105–2116. DOI: 10.1007/s10706-018-0655-2
3. Shahri, A., Azamathulla, H. M., & Ghazali, A. H. (2021). "Support Vector Machine and Random Forest model for prediction of soil bearing capacity." Applied Sciences, 11(1), 234. DOI: 10.3390/app11010234
4. Chen, X., Zhang, B., & Li, L. (2020). "Predicting unconfined compressive strength of soils using machine learning techniques." Soil and Tillage Research, 197, 104518. DOI: 10.1016/j.still.2019.104518
5. Zhang, J., Liu, Y., & Wang, R. (2018). "Soil texture classification using convolutional neural networks." Remote Sensing, 10(10), 1520. DOI: 10.3390/rs10101520
6. Sulaiman, N., Musa, M., & Sidek, R. M. (2021). "IoT-based soil monitoring system for precision agriculture." International Journal of Advanced Computer Science and Applications, 12(1), 154–161. DOI: 10.14569/IJACSA.2021.0120119
7. Al-Khafajiy, M., Baker, T., & Aslam, N. (2020). "Smart soil monitoring using wireless sensor networks." Sensors, 20(2), 400. DOI: 10.3390/s20020400
8. Mandal, D., Das, D. K., & Bhattacharyya, T. (2015). "Application of remote sensing and GIS for soil resource mapping of Birbhum district, West Bengal, India." Agropedology, 25(2), 117–127.
9. ASTM D2487-17. (2017). "Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System)." ASTM International. DOI: 10.1520/D2487-17
10. USCS (Unified Soil Classification System). Referenced from Federal Highway Administration and ASTM standards.
11. Kumaravel, K. (2024). Machine learning algorithms for soil classification: A comparative analysis and real world applications. AIP Conference Proceedings, 3193(1), 020280. https://doi.org/10.1063/5.0232859
12. Momeni, E., He, B., Abdi, Y., & Armaghani, D. J. (2023). Novel Hybrid XGBoost Model to Forecast Soil Shear Strength Based on Some Soil Index Tests. *CMES-Computer Modeling in Engineering & Sciences*, 136(3), 26531. https://doi.org/10.32604/cmes.2023.026531
13. Riese, F. M., & Keller, S. (2019). Soil texture classification with 1D convolutional neural networks based on hyperspectral data. *arXiv preprint arXiv:1901.04846*. https://arxiv.org/abs/1901.04846
14. Zinich, A., & Gindemit, A. (2019). Soil mapping using geo-information technologies. *Proceedings of the International Scientific and Practical Conference "Digital agriculture - development strategy" (ISPC 2019)*, 156–159. https://doi.org/10.2991/ispc-19.2019.35
15. Sari-Ahmed, B., Benzaamia, A., & Ghrici, M. (2024). Strength Prediction of Fiber-Reinforced Clay Soils Stabilized with Lime Using XGBoost Machine Learning. *Civil and Environmental Engineering Reports*, 34(2), 157–176. https://doi.org/10.59440/ceer/190062
16. Islam, M. M., Hasan, M. M., & Farukh, M. A. (2017). Application of GIS in General Soil Mapping of Bangladesh. *Journal of Geographic Information System*, 9(5), 604–621. https://doi.org/10.4236/jgis.2017.95038
17. Riese, F. M. (2020). CNN Soil Texture Classification. *Karlsruhe Institute of Technology*. https://publikationen.bibliothek.kit.edu/1000089550

18. Bai, R., Shen, F., & Zhang, Z. (2023). An integrated machine-learning model for soil category classification based on CPT. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 7, 2121–2146. https://doi.org/10.1007/s41939-023-00324-z
19. Zhou, J., et al. (2020). Prediction of undrained shear strength using extreme gradient boosting and random forest. *Geoscience Frontiers*, 11(5), 1542291. https://doi.org/10.1016/j.gsf.2020.03.007
20. Riese, F. M., & Keller, S. (2023). Soil texture prediction with automated deep convolutional neural networks. *Geoderma*, 424, 115965. https://doi.org/10.1016/j.geoderma.2023.115965
21. Kamamia, M., et al. (2021). Spatial prediction of soil aggregate stability and soil organic carbon using machine learning. *Geoderma Regional*, 25, e00385. https://doi.org/10.1016/j.geodrs.2021.e00385
22. McBratney, A., et al. (2020). Machine learning and soil sciences: a review aided by machine learning. *SOIL*, 6(1), 35–52. https://doi.org/10.5194/soil-6-35-2020
23. Das, B., et al. (2023). Novel Hybrid XGBoost Model to Forecast Soil Shear Strength Based on Some Soil Index Tests. *CMES-Computer Modeling in Engineering & Sciences*, 136(3), 26531. https://doi.org/10.32604/cmes.2023.026531ScienceDire ct+2
24. Riese, F. M., & Keller, S. (2019). Soil texture classification with 1D convolutional neural networks based on hyperspectral data. *arXiv preprint arXiv:1901.04846*. https://arxiv.org/abs/1901.04846
25. Zinich, A., & Gindemit, A. (2019). Soil mapping using geo-information technologies. *Proceedings of the International Scientific and Practical Conference "Digital agriculture - development strategy" (ISPC 2019)*, 156–159.https://doi.org/10.2991/ispc-19.2019.35
26. Sari-Ahmed, B., Benzaamia, A., & Ghrici, M. (2024). Strength Prediction of Fiber-Reinforced Clay Soils Stabilized with Lime Using XGBoost Machine Learning. Civil and Environmental Engineering Reports, 34(2), 157–176. https://doi.org/10.59440/ceer/190062
27. Islam, M. M., Hasan, M. M., & Farukh, M. A. (2017). Application of GIS in General Soil Mapping of Bangladesh. Journal of Geographic Information System, 9(5), 604–621. https://doi.org/10.4236/jgis.2017.95038
28. Riese, F. M. (2020). CNN Soil Texture Classification. Karlsruhe Institute of Technology. https://publikationen.bibliothek.kit.edu/1000089550
29. Bai, R., Shen, F., & Zhang, Z. (2023). An integrated machine-learning model for soil category classification based on CPT. Multiscale and Multidisciplinary Modeling, Experiments and Design, 7, 2121–2146. https://doi.org/10.1007/s41939-023-00324-z
30. Zhou, J., et al. (2020). Prediction of undrained shear strength using extreme gradient boosting and random forest. Geoscience Frontiers, 11(5), 1542291. https://doi.org/10.1016/j.gsf.2020.03.007
31. Riese, F. M., & Keller, S. (2023). Soil texture prediction with automated deep convolutional neural networks. Geoderma, 424, 115965. https://doi.org/10.1016/j.geoderma.2023.115965
32. Kamamia, M., et al. (2021). Spatial prediction of soil aggregate stability and soil organic carbon using machine learning. Geoderma Regional, 25, e00385. https://doi.org/10.1016/j.geodrs.2021.e00385
33. McBratney, A., et al. (2020). Machine learning and soil sciences: a review aided by machine learning. SOIL, 6(1), 35–52. https://doi.org/10.5194/soil-6-35-2020
34. Das, B., et al. (2023). Novel Hybrid XGBoost Model to Forecast Soil Shear Strength Based on Some Soil Index Tests. CMES-Computer Modeling in Engineering & Sciences, 136(3), 26531. https://doi.org/10.32604/cmes.2023.026531