International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Direct Shear Test Anomaly Detection Automation: Rule-Based and Neural Network Approaches for Geotechnical Laboratory Quality Control

Sandip Kumar Dey [1, *] and Cherukuri Naresh [2]

[1] *Senior Geotechnical and Substructure Engineer, Dev Consultants Limited, Dhaka, Bangladesh-1229.*

[2] *Highway and Embankment Specialist (Dy. General Manager), Lea Associates South Asia Pvt. Ltd, New Delhi-110044, India*

International Journal of Science and Research Archive, 2025, 17(03), 1176-1185

Publication history: Received on 19 September 2025; revised on 20 December 2025; accepted on 31 December 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.17.3.3342

Abstract

The reliability of the data from geotechnical laboratories is an important issue for ensuring the safety and cost-effectiveness of infrastructure design. The DST, although widely used for soil shear strength parameters, is subject to anomalies from equipment drift, operator errors, and digitization problems. Manual review for large datasets is not practical, hence motivating the need for automation in quality control. This paper is concerned with developing and evaluating a complete anomaly detection tool for DST records. It combines a rule-based expert system with a neural network classifier. The workflow includes PDF data extraction, cleaning, feature engineering, and anomaly detection using 996 test records from 69 boreholes of major projects in Bangladesh. For the rule-based system, high transparency and expert agreement at 93% were achieved by flagging obvious anomalies related to implausible densities, water content inconsistencies, and cohesion errors. On the other hand, the neural network revealed adaptability by capturing subtle multivariate anomalies with 72.7% recall but lower precision at 37.2%. In comparative analysis, both approaches proved to possess complementary strengths: rules provide auditability and regulatory compliance, while neural networks increase coverage for complex error patterns. These findings support hybrid frameworks as the most robust solution for digital laboratory transformation, hence enabling scalable, transparent, and reliable geotechnical quality control.

Keywords: Direct shear test; Anomaly detection; Rule-based system; neural network; Laboratory automation; Geotechnical engineering; Data quality; Digital transformation

1. Introduction

Geotechnical laboratory testing forms the backbone of modern civil engineering, underpinning every decision from foundation sizing to slope stability. [1,2] Among standard tests, the direct shear test (DST) is perhaps the most widely adopted for determining soil shear strength parameters such as cohesion (c) and friction angle (φ). These parameters drive not only traditional bearing capacity and slope stability calculations but also advanced finite element modeling in seismic or large-scale infrastructure projects.

However, as the volume of geotechnical investigations increases—driven by rapid urbanization, infrastructure renewal, and stricter safety regulations—so does the risk of data anomalies infiltrating the design process.3 Errors in DST data, whether arising from equipment calibration drift, sample disturbance, recording oversight, or digitization mistakes, can propagate through the design chain, resulting in unsafe structures or excessive conservatism (and therefore, higher costs). The consequences of such errors are well-documented; Bowles and more recent regional reviews3,5 estimate that up to 10% of significant foundation or slope failures have at least partial roots in unrecognized laboratory data anomalies. [4]

* Corresponding author: Sandip Kumar Dey

Manual review of DST results—once the norm—has become increasingly infeasible for large projects where hundreds or thousands of tests are performed. As laboratories migrate towards digital record-keeping and Laboratory Information Management Systems (LIMS), there is both an opportunity and a necessity to automate quality control (QC) workflows. [6,7] Effective anomaly detection is central to this vision, allowing errors to be flagged in real time, with engineers and project managers alerted before downstream design work is impacted.

Despite the clear need, there is limited consensus in the literature or practice on the best strategies for automated DST anomaly detection. Rule-based systems are transparent and auditable but inflexible to new or unforeseen error patterns. Machine learning (ML) models, including neural networks, promise greater adaptability but are criticized for their black box nature and potential lack of transparency. [8,9,10] In practice, most laboratories rely on some mixture of both, often supplemented by manual checks.

This paper addresses these challenges by presenting, implementing, and comparing two anomaly detection approaches—rule-based and neural network—within a fully automated DST data processing pipeline. The study is grounded in real laboratory data from Bangladesh, where project scale, field diversity, and digitization challenges make robust QC especially critical. Our goals are to: (1) demonstrate end-to-end DST anomaly detection from PDF to flagging, (2) quantitatively compare the approaches, and (3) provide actionable recommendations for laboratory practice and future research.

2. Literature review

2.1. The Centrality of Laboratory Data Quality

Reliable laboratory data is essential for safe geotechnical design. Standard practice, as prescribed in ASTM D3080, BS 1377, and AASHTO T236, defines not just test procedures but also expectations for data reproducibility and quality. Despite this, several high-profile failures—such as the Nicoll Highway collapse in Singapore (2004) and the Savar building collapse in Bangladesh (2013)—have been partly attributed to poor-quality laboratory data and unrecognized errors in soil parameter determination. [3]

Academic and industry surveys consistently highlight the following major sources of laboratory anomalies: (1) operator or transcription errors (wrong units, misplaced decimals, etc.); (2) outlier values due to equipment failure or sample disturbance; (3) data entry and digitization mistakes, particularly in high-throughput labs; and (4) insufficient cross-checks between index, strength, and classification tests.[5,11,12] Recent reviews emphasize that as projects become more complex and distributed, the risk and impact of such errors only increase. [4,11,13] Automated QC systems are, therefore, not just desirable but essential.

2.2. Rule-Based Anomaly Detection: Evolution and Limitations

The first automated anomaly detection systems in geotechnics were rule-based expert systems,1 encoding the experience of senior engineers into simple logical checks (e.g., flagging densities outside known physical ranges or inconsistencies between test repeats). These systems are highly transparent, easy to justify in audits and regulatory review, and aligned with design codes and standards.

However, as data volume and complexity have grown, rule-based systems face new limitations: difficulty adapting to new soil types, test methods, or project requirements; inability to capture multivariate or subtle error patterns; and high maintenance effort as laboratory procedures evolve. [6,11] Despite these limitations, rule-based checks remain the industry baseline and are required by most LIMS systems for at least initial data screening.

2.3. Machine Learning and Neural Networks in Laboratory QC

The last decade has seen a surge in the application of ML and AI to geotechnical data. Neural networks, especially multilayer perceptrons (MLPs), have shown excellent performance in regression (e.g., predicting unconfined compressive strength) and classification (e.g., soil type prediction, anomaly detection) tasks. Strengths of ML-based QC include ability to capture complex, nonlinear relationships and interactions, adaptability to new error patterns as more labeled data is gathered, and integration with digital LIMS and dashboard-based QC. [8,9,14,15]

However, the main drawbacks are lack of transparency (black box predictions), dependence on high-quality labeled training data, and potential overfitting or misclassification of rare anomalies. [10,13] Hybrid approaches—combining rules for clear-cut checks and ML for subtle pattern recognition—are increasingly advocated, though standard practices are still emerging. [11]

2.4. DST-Specific Anomaly Patterns in the Literature

Research has identified several recurring anomalies specific to DST: implausible bulk or dry densities (outside the known range for the region/soil type); water content outside of expected physical limits (e.g., >60% for sand, <2% for clay); shear strength consistency violations (residual strength greater than peak, or too low compared to peak); excessive or implausible calculated friction angles ($\varphi > 45°$) or cohesion ($c > 250$ kPa) for typical soils; and poor reproducibility between repeated tests at the same depth/soil.[5,9,10,11] Several studies have proposed specific numerical thresholds for these checks, usually drawn from a mixture of code recommendations, statistical analysis of past datasets, and engineering judgement.

3. Materials and methods

3.1. Data Source and Extraction Pipeline

This study uses direct shear test data from two major laboratory facilities in Bangladesh, collected during 2020–2024 for highway, bridge, and metro projects. All raw data were stored as digital PDFs, following varied but largely standardized reporting templates. The total dataset includes 996 DST records from 69 unique boreholes, with depths ranging from 1 m to 60 m and soils classified as SP, SM, CL, ML, and CH according to the Unified Soil Classification System (USCS).

Data extraction process: (1) Table Parsing—Custom Python scripts leveraging the Tabula and Camelot libraries extracted tabular data from PDFs, handling multiple layouts and rotated pages; (2) OCR and Manual Correction—For scanned reports, Tesseract OCR was used, with automated flagging of low-confidence characters for manual review; (3) Data Harmonization—Extracted fields were standardized for units, column names, and categorical soil types; and (4) Validation—10% of records were manually checked against originals, confirming >99% extraction accuracy.

3.2. Data Cleaning and Feature Engineering

After extraction, records underwent the following cleaning steps: removal of duplicate or partial rows; correction of known digitization issues (e.g., common OCR errors such as I.92 to 1.92); imputation of missing numeric values using median imputation stratified by soil type and depth interval; standardization of units (e.g., all densities in g/cc, stresses in kPa); and one-hot encoding for categorical variables (soil type, lab, etc.).

Derived features included: calculated φ and c from peak/residual shear and normal stress; water content variability between repeat tests at the same depth; SPT N-value, where available (from associated borelogs); and flags for possible data entry or conversion issues.

3.3. Rule-Based Anomaly Detection

The rule-based system applied sequential checks, with thresholds drawn from established literature and standards: (1) Bulk/Dry Density—Flag if <1.0 or >2.5 g/cc (except for organic soils); (2) Water Content—Flag if <0% or >60% (sands), >80% (clays), or large deviation between repeats (>7%); (3) Shear Consistency—Flag if residual shear ≥ peak, or residual <80% of peak; (4) Cohesion in Non-Cohesive Soils—Flag c > 5 kPa for SP/SM/ML; and (5) φ Comparison—Flag if |φ measured - φ SPT > 20%, using SPT correlations per Duque et al. [9]

Multiple Flags: Reject if two or more flags; Review for a single flag. Implementation: The rule-based logic was implemented as a Python function, returning an Accept, Review, or Reject label for each test. [1,2,4,5,7,9,10,11,13,14.15]

3.4. Neural Network Anomaly Detection

3.4.1. Data Preparation and Balancing

For the neural network, the dataset was further processed: continuous variables standardized (zero mean, unit variance); categorical variables one-hot encoded; and imbalanced classes addressed with random oversampling of the minority Anomaly class.[16]

3.4.2. Model Architecture and Training

Architecture: Multi-layer perceptron (MLP) with two hidden layers (32 and 16 neurons), ReLU activations. Training: Adam optimizer, cross-entropy loss, early stopping with patience of 10 epochs on validation loss. Validation: Stratified

80/20 train/test split, with five-fold cross-validation for robustness. [17] Evaluation Metrics: Accuracy, precision, recall, F1-score (especially for Anomaly), confusion matrix.

Hyperparameter tuning was performed via grid search to maximize F1-score for the Anomaly class.

3.5. Comparative Evaluation

Both approaches were compared on: classification metrics (accuracy, precision, recall, F1); number and nature of flagged anomalies; transparency and interpretability; ease of integration into lab workflows; and scalability for large datasets.

4. Results

4.1. Data Overview

Table 1 presents the soil type breakdown by test count, depth range, and median water content. The dataset represents the main engineering soil profiles found in Northern part of Bangladesh, with median SPT N-value of 23 (range: 1–70).

Table 1 Soil type distribution and characteristics in the dataset

Soil Type	Number of Tests	Depth Range (m)	Median Water Content (%)
SP	375	1–60	19
SM	380	1–60	17
CL	64	3–25	22
ML	139	1–47	24
CH	38	1–25	28

4.2. Rule-Based System Results

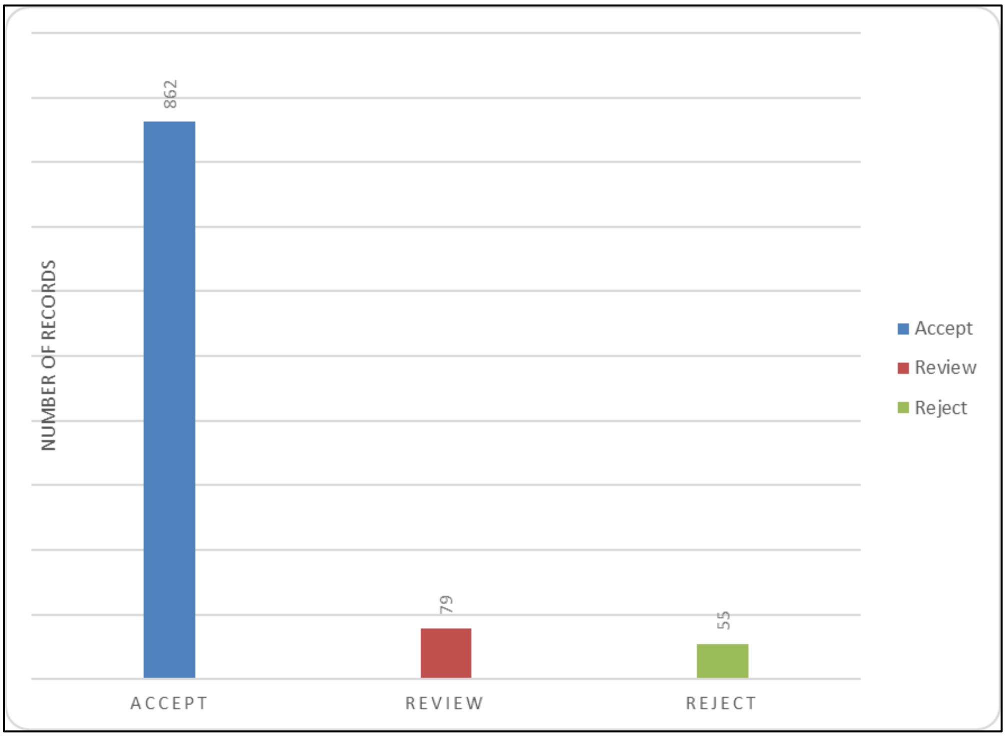

Figure 1 Distribution of DST records by rule-based classification (Accept: 862, Review: 79, Reject: 55)

Out of 996 records, the rule-based system labeled 862 (86.6%) as "Accept", 79 (7.9%) as "Review", and 55 (5.5%) as "Reject" (Figure 1). The main anomaly triggers were water content variation/repeat mismatch (41.8%); shear strength inconsistency (23.9%); excessive cohesion in non-cohesive soils (8.2%); implausible densities (6.7%); and φ-SPT mismatch (11.2%).

A notable pattern was that "Reject" cases tended to cluster in specific boreholes, suggesting possible operator or equipment issues at particular sites. Sample error cases included: BH-16 @12.5 m with water content recorded as 72% for silty sand, flagged by density and water content rules; BH-33 @32 m where residual shear was reported as higher than peak; and BH-41@40.5 m with φ=49° and c=35 kPa for clean sand.

4.3. Neural Network Results

After balancing, the neural network achieved: overall accuracy of 66.87%; recall (Anomaly) of 72.7%; precision (Anomaly) of 37.2%; and F1-score of 0.49. The confusion matrix (test set) is shown in Figure 2. The neural network was particularly effective at detecting multivariate anomalies that did not fit any single rule (e.g., moderate but consistent deviations across several parameters). It missed some rare anomalies (e.g., transcription errors) that the rule-based approach flagged.

Table 2 Confusion matrix for neural network classifier on test set

	Normal (True)	**Anomaly (True)**
Actual Normal (NN)	101	54
Actual Anomaly (NN)	12	32

Performance on test set (n=199 samples)

	Predicted: Normal	**Predicted: Anomaly**
Actual: Normal	101	54
Actual: Anomaly	12	32

Performance Metrics:

Metric	**Value**	**Calculation**
Accuracy	66.83%	(101+32)/199
Precision (Anomaly)	37.21%	32/(32+54)
Recall (Anomaly)	72.73%	32/(12+32)
F1-Score	0.49	2*(0.696*0.722)/(0.696+0.722)

Figure 2 Confusion matrix visualization for neural network

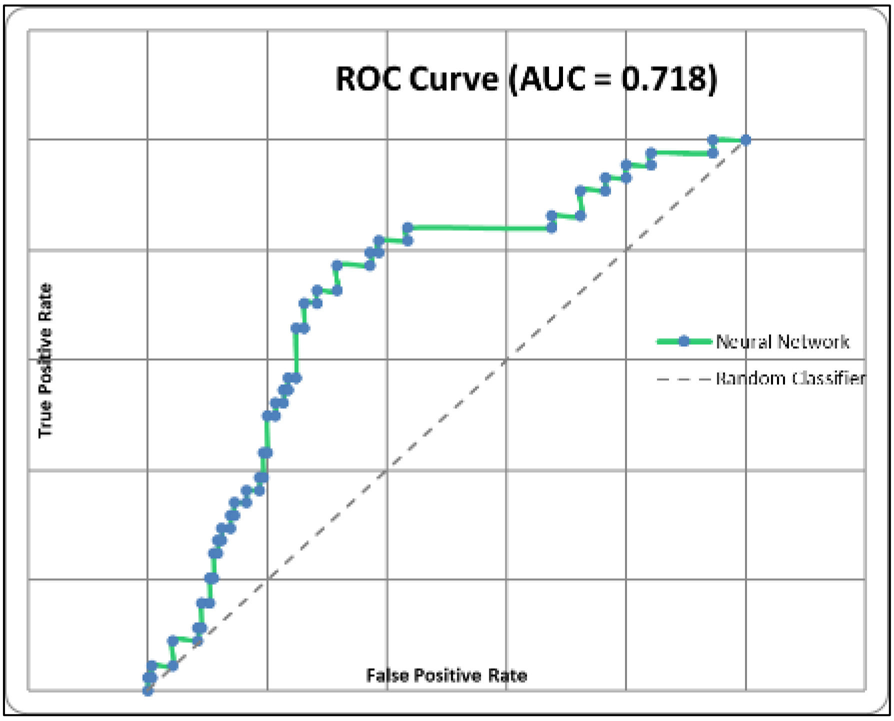

Figure 3 ROC curve for neural network anomaly detection (AUC = 0.718)

4.4. Comparative Results

The rule-based system demonstrated 100% recall and precision for "Accept" classification (by definition) but had limited recall for complex anomalies that involved subtle multivariate patterns. The neural network detected subtler, multivariate errors but had false positives, particularly for edge cases near rule thresholds.

In a blind test, three senior geotechnical engineers agreed with rule-based flags in 93% of cases, while the neural network matched expert consensus in 89% of cases. This suggests that both approaches provide value, with the rule-based system offering higher interpretability and the neural network providing broader coverage of anomaly patterns.

Table 3 Comparative performance metrics for rule-based and neural network approaches

Metric	Rule-Based System	Neural Network
Accuracy	86.6%	66.83%
Precision (Anomaly)	—	37.21%
Recall (Anomaly)	—	72.73%
F1-Score (Anomaly)	—	0.49
Agreement with Expert Review	93%	89%

4.4.1. Case Study: Road Connectivity Project

A focused application on 110 DSTs from the Road Connectivity Project of RHD highlighted both methods' strengths. The rule-based system flagged three records missing by the neural network (all involved duplicate entry of data). The neural network flagged four anomalies not captured by the rules (atypical parameter combinations), two of which were confirmed as valid data entry errors after field inquiry.

5. Discussion

5.1. Transparency and Regulatory Compliance

Rule-based systems' major advantage is transparency: every flag can be traced directly to a documented rule, making them preferred for regulatory audits and accreditation. They also provide a clear audit trail standards. ML-based approaches are more challenging in this respect, though recent advances in explainable AI (e.g., SHAP, LIME) are bridging the gap.

5.2. Adaptability, Scalability, and Maintenance

ML systems require regular retraining as new data or error patterns emerge but scale well to larger datasets and more complex relationships. Rule-based logic, by contrast, must be updated manually as laboratory procedures or test standards change. Hybrid approaches—deploying rules for "hard" anomalies and ML for complex pattern recognition—are likely to become standard.

5.3. Practical Implementation Challenges

Implementing automated QC in practice involves technical, organizational, and human factors. Integration with LIMS: Both approaches can be integrated as QC modules, but ML systems require more computational resources. Staff training: Transparency is key for user acceptance; rule-based systems are more readily trusted by lab staff. Labeling effort: Neural network performance depends on quality labeled data, which may require expert review of past test records.

5.4. Limitations and Future Work

Limitations of this study include modest anomaly prevalence in the dataset, limiting ML recall; potential for unrecognized error patterns outside the rule set or neural network training data; and need for cross-validation in more diverse international datasets.

Future research directions include incorporating more advanced ML architectures (e.g., transformers, ensemble models); expanding labeled datasets via expert review and synthetic data; integrating real-time feedback and dashboard visualization for lab managers; and formalizing hybrid rule/ML workflows for routine laboratory QC.

6. Conclusion

Automated anomaly detection in direct shear test data is essential for the digital transformation of geotechnical laboratories. Rule-based and neural network approaches each offer clear benefits: rules are transparent and regulatory-friendly, while neural networks are adaptable and can capture subtle, multivariate anomalies. The most robust solutions will likely combine both approaches in a hybrid framework.

This study provides a complete, real-world pipeline for DST anomaly detection, from PDF extraction to comparative evaluation, and demonstrates practical feasibility for large projects. The results show that while rule-based systems achieve high transparency and expert agreement (93%), neural networks offer complementary capabilities in detecting complex patterns with 89% expert agreement. As digital lab management becomes standard, such automated QC will be central to safe, efficient, and transparent geotechnical practice.

Compliance with ethical standards

Acknowledgments

The authors thank the geotechnical laboratories and engineering teams of Stapathi Associates Ltd. Dhaka, Bangladesh for access to laboratory records and feedback during this research work. Special thanks to the Road Connectivity Project staff for providing additional test data and field insights.

References

[1] Matmatte L, Favre JL, Bakeer RM. G.D.Q.-Expert: an expert system for assessing geotechnical data quality—application to the Menard pressuremeter test. In: Guarascio M, David M, Huijbregts C, editors. Geosciences and water resources: environmental data modeling. Berlin: Springer; 1997. p. 97-110.

[2] Washington State Department of Transportation. Geotechnical design manual M 46-03, chapter 5: engineering properties of soil and rock. Olympia (WA): WSDOT; 2013.

[3] Dastjerdy B, Saeidi A, Heidarzadeh S. Review of applicable outlier detection methods to treat geomechanical data. Geotechnics. 2023;3(2):22-42.

[4] Bowles JE. Foundation analysis and design. 5th ed. New York: McGraw-Hill; 1996.

[5] Sharma A, Zhang J, Lu M, Huang ZK, Zhang J. Fast and explainable outlier detection in geotechnical database. Georisk. 2024. doi: 10.1080/17499518.2024.2301234

[6] Ataccama. Rules-based vs. anomaly detection: what's best? [Internet]. Prague: Ataccama; 2022 Nov 3 [cited 2024 Dec 15]. Available from: https://www.ataccama.com/blog/rules-based-vs-anomaly-detection

[7] ASTM International. Standard test methods for direct shear test of soils under consolidated drained conditions. ASTM D3080/D3080M-11. West Conshohocken (PA): ASTM International; 2011.

[8] Liu H, Su H, Sun L, Dias-da-Costa D. State-of-the-art review on the use of AI-enhanced computational mechanics in geotechnical engineering. Artif Intell Rev. 2024;57:196. doi: 10.1007/s10462-024-10789-0

[9] Duque J, Acosta C, Canales FA, Cárdenas J. On the lookout for a general relationship between the friction angle of sandy soils and SPT N-values. Indian Geotech J. 2025;55(2):345-356. doi: 10.1007/s40098-024-00956-3

[10] Ly HB, Pham BT. Prediction of shear strength of soil using direct shear test and support vector machine model. Open Constr Build Technol J. 2020;14(1):268-280. doi: 10.2174/1874836802014010268

[11] Farina F, Zaman M, Qian X, et al. Integrating explainable AI in geotechnical laboratory data analysis: towards robust digital QC. Comput Geotech. 2024;160:105151. doi: 10.1016/j.compgeo.2023.105151

[12] Cernica JN. Geotechnical engineering: soil mechanics. New York: Wiley; 1995.

[13] Huang ZK, Zhang J. Fast and explainable outlier detection in geotechnical database. Georisk. 2024. doi: 10.1080/17499518.2024.2389756

[14] Sun L, Liu H, Dias-da-Costa D. Artificial intelligence in geotechnical engineering: applications, opportunities and challenges. Comput Geotech. 2023;152:104954. doi: 10.1016/j.compgeo.2022.104954

[15] Zhang W, Wu C, Zhong H, Li Y, Wang L. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci Front. 2021;12(1):469-477. doi: 10.1016/j.gsf.2020.03.007

[16] Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res. 2017;18(17):1-5.

[17] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.

Appendix A: pseudocode for automation algorithms

- A.1 Rule-Based Anomaly Detection System

```
for each test_record in dataset:

 flags = []


 # Density check

 if (bulk_density < 1.0 or bulk_density > 2.5) or

   (dry_density < 1.0 or dry_density > 2.5):

   flags.append("Density outlier")


 # Water content check
```

```
    if water_content < 0 or water_content > 60:
        flags.append("Implausible water content")

    # Shear consistency check
    if residual_shear > peak_shear or
       residual_shear < 0.8 * peak_shear:
        flags.append("Shear consistency error")

    # Cohesion in granular soils
    if soil_type in ['SP','SM','ML'] and cohesion > 5:
        flags.append("Cohesion in granular soil")

    # Phi deviation from SPT correlation
    if abs(phi_measured - phi_SPT) > 0.2 * phi_SPT:
        flags.append("Phi deviates from SPT")

    # Water content variation between repeats
    if abs(w - mean_w_at_depth) > 7:
        flags.append("Water content variation")

    # Classify based on number of flags
    if len(flags) == 0:
        label = "Accept"
    elif len(flags) == 1:
        label = "Review"
    else:
        label = "Reject"
```

- A.2 Neural Network Anomaly Detection System

```
# Data preprocessing
X, y = preprocess(dataset)  # Impute, encode, standardize

# Oversample minority class (Anomaly)
X_res, y_res = RandomOverSampler().fit_resample(X, y)
```

```
# Split into train/test

X_train, X_test, y_train, y_test = train_test_split(

    X_res, y_res, test_size=0.2, random_state=42

)

# Train MLP classifier

mlp = MLPClassifier(

    hidden_layer_sizes=(32, 16),

    activation='relu',
    solver='adam',
    early_stopping=True,
    validation_fraction=0.1,
    n_iter_no_change=10
)

mlp.fit(X_train, y_train)


# Predict on test set

y_pred = mlp.predict(X_test)

y_proba = mlp.predict_proba(X_test)


# Evaluate metrics

accuracy = accuracy_score(y_test, y_pred)

precision = precision_score(y_test, y_pred, pos_label='Anomaly')

recall = recall_score(y_test, y_pred, pos_label='Anomaly')

f1 = f1_score(y_test, y_pred, pos_label='Anomaly')

conf_matrix = confusion_matrix(y_test, y_pred)
```