www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 250-258. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2025.v16i4.1006

Classification of Soil Type in Buildings using Pseudo Spectral Acceleration Readings from Earthquake Events

Erick Rosete-Beas[1], *Laura M. Rodríguez Peralta*[1], *Eduardo Ismael Hernandez*[1]
[1] Universidad Popular Autónoma del Estado de Puebla, Engineering Department, Puebla, México
erick.rosete@upaep.edu.mx, lauramargarita.rodriguez01@upaep.mx, eduardo.ismael@upaep.mx

Abstract. Soil type is a critical factor influencing the seismic performance of buildings, as it affects the level of damage sustained during earthquakes. This paper presents a novel approach to classifying building soil types using pseudo-spectral acceleration readings recorded during seismic events. By leveraging machine learning classifiers, the study develops a model that accurately identifies soil types from pseudo-spectral acceleration data, achieving an accuracy of 89.16%. The methodology involves preprocessing the seismic data, extracting key features, and applying various classifiers to determine the most effective model. Performance is evaluated using metrics such as accuracy, precision, and recall. The findings indicate that this approach significantly improves soil classification accuracy over traditional methods, providing a practical tool for seismic hazard assessment and building design. This research further advances earthquake engineering by offering a data-driven solution to enhance building resilience.
Keywords: Pseudo Spectral Acceleration, Soil Type Classification, Earthquake Engineering, Seismic Data Analysis, Machine Learning

Article Info
Received February 25, 2025
Accepted July 6, 2025

1 Introduction

The seismic performance of buildings is intrinsically linked to the type of soil on which they are constructed (Borcherdt & Glassmoyer, 1992). Different soil types can amplify or attenuate seismic waves to varying degrees, significantly influencing a structure's response during an earthquake. Accurately assessing soil-structure interaction is therefore essential for effective seismic hazard assessment and the design of earthquake-resilient buildings.

Pseudo Spectral Acceleration (PSA) readings, representing the maximum acceleration response of a structure at a specific frequency, are widely utilized in earthquake engineering to characterize the dynamic behavior of buildings during seismic events (Baker & Cornell, 2006). These readings encapsulate crucial information about soil conditions and the seismic response of structures, making them a promising basis for soil type classification.

Traditional methods of determining soil type often rely on site-specific geotechnical investigations, which are time-consuming and costly (Gong et al., 2017). Furthermore, these methods typically provide localized, point-based assessments that may not fully capture the spatial variability of soil properties across a site. In contrast, PSA readings offer a more data-driven and potentially cost-effective approach, enabling broader and more comprehensive soil type classification across larger areas.

Recent advances in machine learning have demonstrated the potential of data-driven models in predicting seismic parameters such as Peak Ground Acceleration (PGA). For instance, Chiang et al. (2022) showed that artificial neural networks could accurately predict PGA and Peak Ground Velocity, highlighting the potential for similar approaches in soil type classification. Additionally, Khosravikia and Clayton (2021) explored the variability of ground motion predictions across different events and sites using machine learning algorithms, underscoring the effectiveness of data-driven approaches in seismic analysis.

This study proposes a methodology that utilizes classification techniques to identify the soil type of buildings based on their PSA readings during earthquakes. The approach involves preprocessing the PSA data, extracting relevant features, and evaluating

various classifiers to determine the most effective model for this task. By applying advanced machine learning techniques, a robust model is developed capable of accurately classifying soil types, even in the absence of detailed geotechnical investigations. This model could significantly enhance seismic hazard assessments and contribute to more resilient infrastructure design.

The key contributions of this research are: (1) a data-driven model that uses PSA readings and machine learning to classify soil types with an accuracy of 89.16%.; (2) the preprocessing and consolidation of soil type data from seismic stations across Mexico, addressing data imbalances and enhancing the reliability of the results; (3) extensive feature engineering, including the use of the Horizontal-to-Vertical Spectral Ratio (H/V ratio) and Principal Component Analysis (PCA), to distill valuable information from PSA data; and (4) an ablation study to evaluate multiple machine learning models and identify the most effective approach for soil type classification.

This research bridges the existing gap in soil information for seismic stations and offers a scalable solution for rapid soil type classification over large geographic areas, thereby enhancing current capabilities in seismic hazard assessment and building design. Despite advances in using geotechnical and ambient-vibration data for soil classification, the literature lacks models that exploit PSA signatures directly. This paper addresses that gap.

The structure of this paper is as follows: Section 2 reviews related work in soil type classification and the use of PSA in seismic analysis. Section 3 details the methodology employed, including data collection, preprocessing, and feature extraction. Section 4 presents the results and discusses the performance of the proposed classification models. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2 Related Work

Accurate soil classification is crucial for effective seismic engineering practices, as soil properties significantly influence ground motion amplification and structural response (Pitilakis et al., 2013). Traditional geotechnical investigations, while providing detailed information, are often time-consuming and costly. There has been a growing interest in utilizing seismic data and advanced computational methods to characterize soil properties more efficiently and accurately.

Machine learning techniques have shown significant potential in various seismic applications, including soil classification and ground motion prediction. Chala and Ray (2023) demonstrated the efficacy of deep learning in soil classification using cone penetration test data, achieving high accuracy in identifying soil types. Similarly, Xiao et al. (2021) proposed a coupled machine learning method to integrate borehole and piezocone penetration test data for improved soil classification, demonstrating the potential of machine learning in combining multiple data sources for more reliable site characterization.

In seismic response prediction, Derras et al. (2014) used artificial neural networks to predict ground motion parameters from various input features, including soil conditions. Their work highlighted the potential of machine learning in capturing complex relationships between soil properties and seismic responses. Building on this, Mori et al. (2022) developed a machine learning approach using Gaussian process regression to produce high-resolution ground motion prediction maps, demonstrating the capability of machine learning techniques in handling large-scale seismic data.

The application of machine learning in seismic engineering extends beyond soil classification and ground motion prediction. Researchers have used these techniques to enhance earthquake early warning systems (Kong et al., 2019), predict slope displacements in seismic events (Xu et al., 2012), and generate realistic synthetic seismic data (Kim & Kim, 2024). Kwag et al. (2020) demonstrated the effectiveness of artificial neural networks and Gaussian process regression in predicting the seismic performance of slopes, further underscoring the versatility of machine learning in seismic analysis.

Joshi et al. (2024) applied machine learning techniques to estimate shear wave velocity profiles from geotechnical and geophysical data, showcasing the potential of these methods for subsurface characterization. However, their work primarily focused on property estimation and did not delve into soil type classification.

Despite these advancements, none of the above studies used PSA data alone to directly classify site soil type at building locations. This research aims to address this gap by developing a robust machine learning model for classifying soil types at building locations using earthquake data, including PSA readings. By focusing on this specific application and utilizing a comprehensive dataset, this study seeks to advance seismic site characterization and provide a valuable tool for improving building design and safety.

3 Methodology

3.1 Data Description

The dataset used in this study was provided by the Accelerographic Network of the Institute of Engineering (RAII-UNAM, 2021), as a result of the instrumentation and processing efforts of the Seismic Instrumentation Unit. The data are distributed through the Accelerographic Database System online. It includes data from 160 seismic stations, with ground type information available for 140 of them. The analysis focused on PSA data from seismic events recorded at stations with known soil type information This subset consists of 6 220 recordings from events that had magnitudes greater than 3.0 and occurred in Mexico between 1985 and 2020. Each recording utilized three sensors, measuring the North-South (N/S), East-West (E/W), and vertical (V) components, resulting in 18660 PSA measurements.

Due to the initial diversity and uneven distribution of soil type classifications, some with very few samples or ambiguous descriptions, a preprocessing step was essential to enhance the reliability of the analysis. The original dataset contained multiple soil types with varying frequencies, as shown in Table 1.

Table 1. Original Station Soil Type Distribution

Station Soil Type	Station Count
rock	84
clay	12
basaltic rock	8
soft	7
soil	5
alluvial	4
lake zone	3
soft soil	3
compacted clay	3
sandy silt	3
transition zone	2
sedimentary rock	2
granite	2
Others	22
Total	160

To address the issue of uneven distribution and improve the statistical significance of the analysis, the various soil types were consolidated into three primary categories: Hard, Transitional, and Soft soil. This grouping was based on geological characteristics and seismic response properties.

The hard category includes stable geological formations known for lower seismic wave amplification. Soil types grouped under this category are rock, basaltic rock, sedimentary rock, granite, travertine, metamorphic rock, basalt, altered granite, fractured rock, and similar formations. The transitional category is characterized by materials like compacted clay, sandy silt, and sand-silt-clay mixtures, which can amplify seismic motions due to lower shear strength. The soft category encompasses soils with high compressibility that significantly influence seismic wave propagation, such as clay, lake zone preconsolidated soil, and alluvial soils.

To ensure data integrity, soil types with ambiguous descriptions or insufficient data, such as "soil" and "unknown," were excluded from the analysis. Non-soil entries, including open ground, structures, and archaeological monuments, were also omitted.

After preprocessing and reclassification, the final dataset distribution is summarized in Table 2. Fig. 1 displays the geographical distribution of the seismic stations across Mexico, color-coded by the newly defined soil categories.

Table 2. Final Soil Type Distribution After Preprocessing

Soil Category	Number of Stations	Number of Records
hard	102	4798
soft	29	1056
transitional	9	466
Total	140	6220

The consolidation into three categories improved class balance and enhanced the statistical significance of subsequent analyses. The PSA data associated with each soil category revealed distinctive patterns, reinforcing the rationale for using PSA measurements in soil type classification.

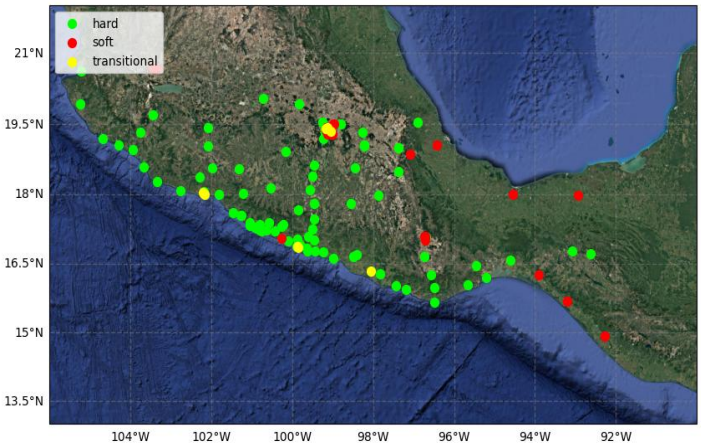

Fig. 1. Geographical distribution of seismic stations across Mexico, color-coded by soil type.

Given the limited and often imprecise soil information available for seismic stations in Mexico, this study explores the use of PSA data as a proxy to address this information gap. PSA is instrumental in understanding how seismic waves interact with local soil conditions. By leveraging machine learning techniques, soil types are predicted based on the distinctive PSA patterns observed across different seismic events. Fig. 2 illustrates the variability and consistency of PSA signals across stations, supporting the potential of PSA data in soil type prediction.

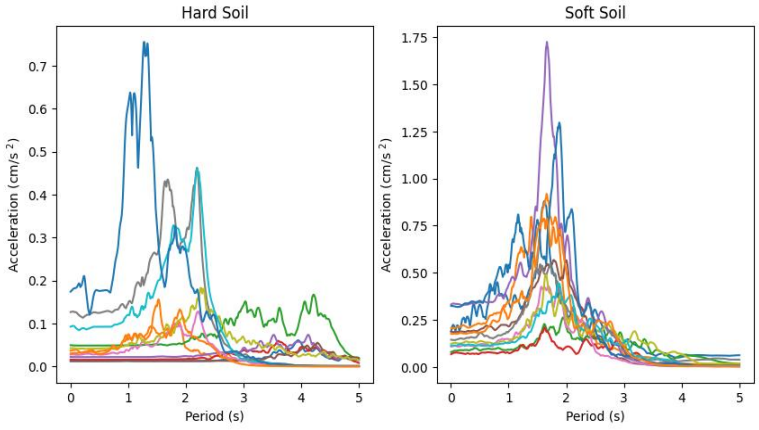

Fig. 2. PSA spectra for individual recordings at Hard (left) and Soft (right) soil stations. Each colored curve is one event's PSA (in cm/s²) plotted against period (0 - 5 s); note that Soft-soil spectra peak much higher (around 1 - 2 s) than Hard soil spectra, indicating stronger site amplification. These consistent, intra-category patterns support using PSA signatures to distinguish soil types.

3.2 Feature Engineering

The feature selection for this study was driven by the need to identify the most impactful seismic parameters for soil type classification. Features were derived from PSA data using the Spectral Ratio method. The Spectral Ratio was calculated using the Horizontal-to-Vertical (H/V) Spectral Ratio technique (Macau et al., 2015). Specifically, the horizontal component was computed as the Euclidean norm of the north-south and east-west PSA components, which was then divided by the vertical PSA to obtain the Spectral Ratio. This method captures the relative amplification between horizontal and vertical motions and is particularly valuable for distinguishing differences in seismic wave behavior across various soil types.

In addition to the Spectral Ratio (SR) features, six Conventional Earthquake Measurement (CEM) features commonly used in seismic analysis were included: magnitude, maximum recorded acceleration in the vertical (V), east-west (E/W), and north-south (N/S) orientations, depth, and duration.

Each PSA recording provides 1,000 frequency components ranging from 0.1 Hz to 100 Hz. To manage the high dimensionality of the Spectral Ratio data and extract meaningful information, two strategies were employed. The first involved aggregating the Spectral Ratio measurements by calculating statistical summaries: maximum, mean, minimum, median, and standard deviation across the frequency components, resulting in five features that capture the overall characteristics of the Spectral Ratio. The second strategy applied Principal Component Analysis (PCA) to the Spectral Ratio data, retaining the first 4 principal components that captured over 99% of the variance.

Depending on the feature extraction strategy, the total number of input features ranged from 6 to 1,006 (see Table 3). When only the six conventional earthquake measurements were used, the feature vector comprised six dimensions. Augmenting these six CEMs with five SR summary statistics produced eleven features (6 CEM + 5 SR stats). Alternatively, retaining the first four principal components from PCA on the SR data and combining them with the six CEMs resulted in ten features (6 CEM + 4 PCA). Finally, using the entire 1,000-component SR vector alongside the six CEMs yielded a high-dimensional feature space of 1,006 features.

Table 3. Summary of feature extraction strategies and resulting feature dimensionalities

Feature Set	No. of Features	Description
CEM Only	6	Contains only the six Conventional Earthquake Measureme (CEM) features.
SR Only	1000	Consists exclusively of the raw 1,000-component Spectral (SR) vector.
CEM + SR	1006	Appends the full 1,000-component SR vector to the six CE
CEM + SR Statistics	11	Combines the six CEMs with five SR summary statistics (r mean, min, median, standard deviation).
CEM + PCA Reduced SR	10	Merges the six CEMs with the four principal components r from PCA on the SR data.

These feature sets correspond to the tests presented in our results (Table 4), allowing the assessment of the impact of different combinations of features on soil type classification performance.

3.3 Ablation Study and Feature Evaluation

To evaluate the effectiveness of different feature sets in predicting soil types, an ablation study was conducted, systematically analyzing the impact of specific features on the overall performance of the model. This approach helps identify the most influential features for accurate classification. The feature sets evaluated in the ablation study are those described in the Feature Engineering section: CEM Only, SR Only, CEM + SR, CEM + SR Statistics, and CEM + PCA Reduced SR.

The XGBoost classification algorithm was utilized for this study. XGBoost, standing for Extreme Gradient Boosting, is a scalable and efficient implementation of gradient-boosted decision trees. It is renowned for its high performance and speed in handling classification and regression tasks, making it suitable for processing large-scale datasets.

To ensure a robust and unbiased evaluation of the models, 5-fold stratified cross-validation was employed on the record-level during the training and testing phases. Stratified cross-validation maintains the original class distribution within each fold, which

is especially important given the class imbalance in the dataset (as shown in Table 2). By partitioning the dataset into five folds with proportional representation of each soil category, each fold served as a representative sample of the overall data. The models were trained on four folds and validated on the remaining fold, with this process repeated five times to ensure that each fold served as the validation set once. Performance metrics were then averaged across the five folds, providing a reliable assessment of model performance.

3.4 Model Evaluation

To identify the optimal model for the task, several well-established machine learning algorithms were experimented with: XGBoost, Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). All models were implemented using the scikit-learn library (version 1.3.2) in Python, with default hyperparameters to ensure replicability. Using default settings provides a baseline performance for each algorithm without the influence of hyperparameter tuning. For models sensitive to feature scaling, such as SVM and MLP, standardization was applied to the feature sets using scikit-learn's StandardScaler.

4 Results

All evaluations in this study were conducted using five-fold stratified cross-validation to preserve the original class proportions of soil types in each fold. First, the impact of different feature sets on classification performance was assessed using XGBoost. Table 4 presents the performance metrics accuracy, precision, recall, and F1-score corresponding to five feature-set configurations: Conventional Earthquake Measurements only (CEM only), Spectral Ratio only (SR only), CEM combined with the full 1,000-component SR vector (CEM + SR), CEM combined with a PCA-reduced SR vector (CEM + PCA-reduced SR), and CEM combined with summary statistics computed from the SR vector (CEM + SR statistics).

Table 4. Performance Metrics for Different Feature Sets (ranked by accuracy; all values obtained using XGBoost)

Feature set	Accuracy	Precision	Recall	F1-Score
CEM + SR	**89.16%**	**85.04%**	**70.62%**	**76.04%**
SR only	87.15%	79.75%	65.42%	70.51%
CEM + PCA reduced SR	85.14%	74.24%	62.49%	66.65%
CEM + SR statistics	82.73%	68.44%	56.74%	60.65%
CEM only	76.33%	50.08%	40.28%	41.44%

The combination of CEM with full SR features (CEM + SR) yields the highest accuracy (89.16%) as well as the best balance between precision (85.04%) and recall (70.62%), resulting in an F1-score of 76.04%. This performance improvement over the "SR only" configuration (87.15% accuracy, 79.75% precision, 65.42% recall, F1-score 70.51%) suggests that conventional measurements (e.g., peak ground acceleration, PGA, and other time-domain descriptors) provide complementary information to the spectral features. Notably, when only CEM features are used, accuracy drops to 76.33% with a much lower precision of 50.08% and recall of 40.28%, indicating that conventional measurements alone lack sufficient discriminative power to distinguish among soil types. The configurations in which SR features were summarized, either via PCA reduction or by computing summary statistics, still outperform CEM alone, but both fall short of the full SR vector's representational capacity. Specifically, CEM + PCA-reduced SR achieves 85.14% accuracy (F1-score 66.65%), while CEM + SR statistics obtains 82.73% accuracy (F1-score 60.65%). Because the dataset is imbalanced (see Table 2), it is essential to consider precision, recall, and F1-score alongside accuracy: the high precision of CEM + SR (85.04%) indicates relatively few false positives, while the recall of 70.62% shows that some samples, primarily in underrepresented soil categories, remain misclassified. Overall, CEM + SR offers the strongest performance across all reported metrics.

Next, using CEM + SR (the best feature set), six classification algorithms were evaluated to identify the most suitable model for this task. Table 5 compares the accuracy, precision, recall, and F1-score of XGBoost, Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Multilayer Perceptron (MLP), all trained with default hyperparameters on the same cross-validation splits.

Here, XGBoost again ranks highest with 89.16% accuracy, 85.04% precision, 70.62% recall, and an F1-score of 76.04%. Random Forest closely follows, achieving 87.28% accuracy, 81.76% precision, 63.02% recall, and an F1-score of 68.61%. The KNN classifier achieves 85.48% accuracy (F1-score 61.33%), while Logistic Regression and SVM fall to below 80% accuracy, with SVM showing particularly low precision (25.71%) and recall (33.33%). The MLP model performs worst overall (48.94%

accuracy), likely due to the limited sample size (6,220 recordings) and the challenge of training a neural network with insufficient data. The robust performance of tree-based models, especially XGBoost, is consistent with their known ability to capture non-linear relationships and to handle class imbalance more effectively (Chen & Guestrin, 2016).

Table 5. Comparison of classification performance for different machine learning models ranked by accuracy

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	**89.16%**	**85.04%**	**70.62%**	**76.04%**
Random Forest	87.28%	81.76%	63.02%	68.61%
KNN	85.48%	80.66%	56.20%	61.33%
Logistic regression	78.95%	58.90%	44.04%	46.49%
SVM	77.14%	25.71%	33.33%	29.03%
MLP	48.94%	38.56%	38.71%	33.92%

Because class imbalance persists even under stratified splits, random oversampling was applied to the training folds to increase the representation of minority classes. Table 6 reports five-fold weighted-average metrics for XGBoost and Random Forest trained on three feature sets: CEM only, CEM + SR, and SR only.

Table 6. Classification performance (accuracy, precision, recall, F1-score) for XGBoost and Random Forest on three feature sets after applying random oversampling to address class imbalance

Model	Feature Set	Accuracy	Precision	Recall	F1-Score
XGBoost	CEM only	68.46%	71.06%	68.46%	69.46%
XGBoost	CEM + SR	**88.99%**	**88.56%**	**88.99%**	**88.63%**
XGBoost	SR only	87.62%	86.90%	87.62%	86.99%
Random Forest	CEM only	73.87%	69.61%	73.87%	71.31%
Random Forest	CEM + SR	87.15%	86.32%	87.15%	86.39%
Random Forest	SR only	87.15%	86.28%	87.15%	86.38%

As expected, oversampling has the most pronounced effect on the less-informative feature set (CEM only), where XGBoost's accuracy remains low (68.46%) despite improved precision (71.06%) and recall (68.46%). Compared to CEM only (68.46% accuracy), including SR features improves performance substantially. In particular, CEM + SR achieves 88.99% accuracy, 88.56% precision, 88.99% recall, and an F1-score of 88.63% for XGBoost. Random Forest trained on CEM + SR reaches 87.15% accuracy (F1-score 86.39%), while using SR only yields 87.62% accuracy (F1-score 86.99%) for XGBoost and 87.15% accuracy (F1-score 86.38%) for Random Forest. These results confirm that full SR features provide substantial discriminative power even under class-balanced training, and that XGBoost consistently outperforms Random Forest by a small margin.

Finally, Fig. 3 displays the confusion matrix obtained from the XGBoost model trained on CEM + SR with random oversampling. The three soil classes, hard, transitional, and soft, are shown along both axes. Hard-soil recordings are correctly classified in 4 586 out of 4 798 cases (95.6%). Of the 4 798 true hard-soil recordings, 162 (3.4%) are misclassified as soft and 50 (1.0%) are misclassified as transitional. Soft-soil recordings (1 056 total) are correctly identified in 748 cases (70.8%), while 291 (27.6%) are mislabeled as hard and 17 (1.6%) are mislabeled as transitional. Transitional-soil recordings (366 total) are correctly classified in 202 cases (55.2%), but 122 (33.3%) are predicted as hard and 42 (11.5%) are predicted as soft. These misclassifications suggest that transitional- and soft-soil spectral patterns overlap with those of hard soils, especially in boundary cases, so some transitional recordings appear "too stiff" (and are predicted as hard), and a few soft recordings resemble transitional spectra. Despite these errors, overall accuracy remains high. To reduce the remaining mistakes, future work could employ more balanced sampling, refine the feature set further, or incorporate additional site information (for example, local geologic maps).

This PSA-based method has several clear benefits. Rather than relying on costly borehole or CPT fieldwork and lab tests, it uses existing seismic recordings to classify soil over large areas quickly and at lower cost. Compared to H/V ambient-vibration techniques (Macau et al., 2015), PSA directly measures how soils amplify real earthquake motions, making it more reliable under different noise conditions. There are some drawbacks. PSA-based classification is indirect and relies on good-quality recordings stations. Finally, while deep learning on raw waveform data could yield richer features, it requires much larger labeled datasets. Given current data availability, tree-based PSA models like XGBoost remain the most practical choice.

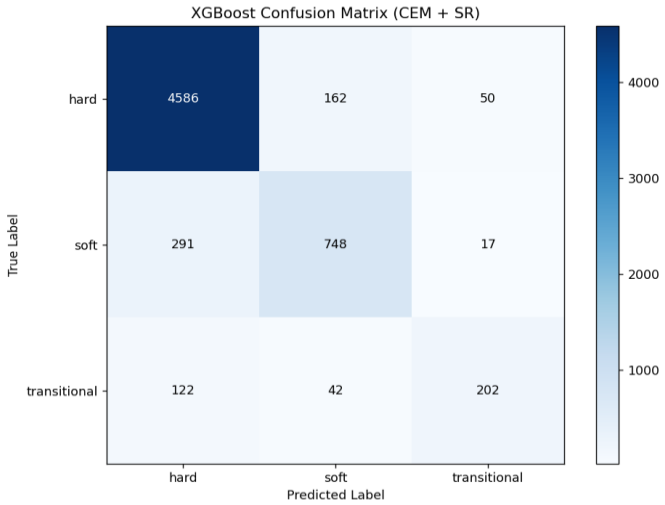

Fig. 3. Confusion Matrix for XGBoost (CEM + SR, oversampled training). The rows represent true classes; the columns represent predicted classes. Diagonal entries correspond to correct classifications; off-diagonal entries indicate misclassifications.

5 Conclusions

This study successfully demonstrates the use of Pseudo Spectral Acceleration (PSA) readings for soil type classification through machine learning. By integrating Spectral Ratio features with conventional seismic measurements, models, particularly XGBoost and Random Forest, achieved superior accuracy compared to other data-driven methods. This approach offers a cost-effective tool for seismic hazard assessment and soil-structure interaction analysis, representing a shift towards more data-driven, efficient building design strategies.

By accurately predicting soil type using our model, this study opens new avenues for improved seismic risk analysis and planning of earthquake resistant building designs. Future work should explore integrating ambient vibration analysis to predict soil type without requiring an actual earthquake event. This capability would support preventive measures and the development of comprehensive seismic risk indicators that incorporate building characteristics, ultimately guiding the design of structures that are more compatible with local soil conditions.

Overall, the findings confirm that PSA readings can be effectively utilized for predicting soil type, with each method contributing unique strengths to the classification task. The insights gained from this study have important implications for seismic hazard analysis and risk assessment, providing a robust framework for leveraging PSA data in soil type prediction. Among these preprocessed features, the Spectral Ratio was particularly valuable for distinguishing differences in seismic wave behavior across various soil types.

Acknowledgements

The authors would like to acknowledge Dr. Galdir Reges from the Universidade Federal da Bahia (UFBA) for his invaluable contributions. His expertise in applying machine learning to earthquake engineering and seismic data analysis was crucial to this research.

References

Baker, J. W., & Cornell, C. A. (2006). Which spectral acceleration are you using? Earthquake Spectra, 22(2), 293-312. https://doi.org/10.1193/1.2191540

Borcherdt, R. D., & Glassmoyer, G. (1992). On the characteristics of local geology and their influence on ground motions generated by the Loma Prieta earthquake in the San Francisco Bay region, California. Bulletin of the Seismological Society of America, 82(2), 603-641.

Chala, A. T., & Ray, R. P. (2023). Machine learning techniques for soil characterization using cone penetration test data. Applied Sciences, 13(14), 8286. https://doi.org/10.3390/app13148286

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Chiang, Y. J., Chin, T. L., & Chen, D. Y. (2022). Neural network-based strong motion prediction for on-site earthquake early warning. Sensors, 22(3). https://doi.org/10.3390/s22030704

Derras, B., Bard, P. Y., & Cotton, F. (2014). Towards fully data-driven ground-motion prediction models for Europe. Bulletin of Earthquake Engineering, 12(1), 495-516. https://doi.org/10.1007/s10518-013-9481-0

Gong, W., Tien, Y.-M., Juang, C. H., Martin, J. R., II, & Luo, Z. (2017). Optimization of site investigation program for improved statistical characterization of geotechnical property based on random field theory. Bulletin of Engineering Geology and the Environment, 76, 1021-1035.

Joshi, A., Raman, B., Mohan, C. K., & Cenkeramaddi, L. R. (2024). A new machine learning approach for estimating shear wave velocity profile using borelog data. Soil Dynamics and Earthquake Engineering, 177, 108424. https://doi.org/10.1016/j.soildyn.2023.108424

Khosravikia, F., & Clayton, P. (2021). Machine learning in ground motion prediction. Computers & Geosciences, 148. https://doi.org/10.1016/j.cageo.2021.104700

Kim, J., & Kim, B. (2024). Generative adversarial network to produce numerous artificial accelerograms with pseudo-spectral acceleration as conditional input. Computers and Geotechnics, 160, Article 106566. https://doi.org/10.1016/j.compgeo.2024.106566

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. Seismological Research Letters, 90(1), 3-14. https://doi.org/10.1785/0220180259

Kwag, S., Hahm, D., Kim, M., & Eem, S. (2020). Development of a probabilistic seismic performance assessment model of slope using machine learning methods. Applied Sciences, 10(8), 2831. https://doi.org/10.3390/su12083269

Macau, A., Benjumea, B., Gabàs, A., Figueras, S., & Vilà, M. (2015). The effect of shallow Quaternary deposits on the shape of the H/V spectral ratio. Surveys in Geophysics, 36(1), 185-208. https://doi.org/10.1007/s10712-014-9305-z

Mori, F., Mendicelli, A., Falcone, G., Acunzo, G., Spacagna, R. L., Naso, G., & Moscatelli, M. (2022). Ground motion prediction maps using seismic-microzonation data and machine learning. Natural Hazards and Earth System Sciences, 22(3), 947-966. https://doi.org/10.5194/nhess-22-947-2022

Pitilakis, A., Riga, K., & Anastasiadis, K. (2013). New code site classification, amplification factors and normalized response spectra based on a worldwide ground-motion database. Bulletin of Earthquake Engineering, 11(4), 925-966. https://doi.org/10.1007/s10518-013-9429-4

RAII-UNAM. (2021). Red Acelerográfica del Instituto de Ingeniería de la UNAM [Data set]. Instituto de Ingeniería, Universidad Nacional Autónoma de México. https://aplicaciones.iingen.unam.mx/AcelerogramasRSM/Registro.aspx

Xiao, T., Zou, H.-F., Yin, K.-S., Du, Y., & Zhang, L.-M. (2021). Machine learning-enhanced soil classification by integrating borehole and CPTU data with noise filtering. Bulletin of Engineering Geology and the Environment, 80, 9157-9171. https://doi.org/10.1007/s10064-021-02478-x

Xu, C., Xu, X., Dai, F., & Saraf, A. K. (2012). Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. Computers & Geosciences, 46, 317-329. https://doi.org/10.1016/j.cageo.2012.01.002