Interpretability of XGBoost in Imbalanced Seismic Performance Classification of Reinforced Concrete Buildings Using Resampling Techniques

Ryan James S. Olivo
Center for Structural Engineering and Informatics, MSU - Iligan Institute of Technology

Joel G. Opon
Center for Structural Engineering and Informatics, MSU - Iligan Institute of Technology

https://doi.org/10.5109/7395554

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 11, pp.443-449, 2025-10-30. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

Interpretability of XGBoost in Imbalanced Seismic Performance Classification of Reinforced Concrete Buildings Using Resampling Techniques

Ryan James S. Olivo[1,2,4], Joel G. Opon[1,2,3]
[1] Center for Structural Engineering and Informatics, MSU – Iligan Institute of Technology, Iligan City
[2] Research Institute for Engineering and Innovative Technology, MSU – Iligan Institute of Technology
[3] Department of Civil Engineering and Technology, MSU – Iligan Institute of Technology
[4] Department of Civil Engineering, University of Southern Mindanao

Corresponding author email: ryanjamesolivoce@gmail.com

Abstract: *XGBoost is widely used in performance-based earthquake engineering (PBEE) because of its excellent performance, scalability, efficiency and ability to capture complex patterns. However, it often lacks interpretability, which makes it harder to trust, especially when classifying the seismic performance of reinforced concrete buildings into Immediate Occupancy (IO), Life Safety (LS), and Collapse Prevention (CP). This issue becomes more difficult when the CP class is underrepresented. This study examines how three resampling methods: ADASYN (oversampling), EditedNN (undersampling), and SMOTE-Tomek (hybrid) affect how well we can understand XGBoost predictions. SHAP and Partial Dependence Plots (PDPs) are used to explain which features the model focuses on. Results show that oversampling reveals more influential and important features beyond Sa(1.0), such as We, SR, Eh, and Sa(T1), improving model transparency. Careful resampling, therefore, is critical. It is essential for interpretable and trustworthy models in PBEE.*

Keywords: Imbalance Classification; Interpretability; Seismic Performance Levels; XGBoost

1. INTRODUCTION

Machine learning (ML) has become an important tool in performance-based earthquake engineering (PBEE), especially for predicting the seismic performance of buildings, particularly in determining, immediate occupancy (IO), life safety (LS), and collapse prevention (CP) [1–5]. Models like XGBoost are widely used because of their high accuracy, scalability, and ability to capture complex patterns and relationships in structured data. However, their lack of interpretability, particularly in explaining how structural and seismic features influence predictions, can limit trust and practical use. Interpretability helps engineers or designers understand which structural or ground motion parameters influence predictions toward IO, LS, or CP, enabling them to optimize or improve structural designs for better performance under seismic events. A common challenge in seismic performance classification is class imbalance [5–7], where rare but critical cases such as collapse prevention are underrepresented. This can result in biased predictions and poor learning of important outcomes [8–10]. Resampling techniques are often applied to handle this imbalance, but their impact on model interpretability in multiclass seismic problems is still not well studied. While tools like SHAP (SHapley Additive exPlanations) have been used in engineering and finance to explain model behavior, few studies have examined how resampling methods affect model interpretability, especially for rare seismic performance levels like CP.

This paper addresses that gap by applying Edited Nearest Neighbors (EditedNN), Adaptive Synthetic Sampling (ADASYN), and Synthetic Minority Oversampling Technique using Tomek links (SMOTE-Tomek) to a seismic performance dataset with a focus on the rare CP class. Using SHAP and Partial Dependence Plots (PDP), the study explores how resampling affects feature importance and local model explanations. The goal is to better understand how these techniques influence both prediction performance and interpretability in XGBoost models trained on imbalanced data.

2. LITERATURE REVIEW

This section clearly and briefly explains the key ideas and past studies that support the concept of this paper. It is divided into four parts: (1) the XGBoost classifier and its application in seismic classification, (2) resampling techniques used to address class imbalance in machine learning, (3) interpretability methods for understanding ML decisions in structural engineering, and (4) the emerging intersection between resampling and interpretability. Each part highlights important findings and shows what is still missing in current research.

2.1 The XGBoost Classifier

XGBoost, short for Extreme Gradient Boosting Machine, was introduced by Chen and Guestrin in 2016 [11]. It is a scalable tree boosting algorithm that incorporates regularization, sparsity-aware learning, and cache-efficient computations, making it highly effective for structured data. These enhancements enable XGBoost to train efficiently on large datasets while maintaining high predictive accuracy. In a recent study, Zhang et al. [4] applied XGBoost to classify seismic damage states in reinforced concrete (RC) frames, achieving 80% testing accuracy. Seismic design intensity and spectral accelerations were identified as the most influential features. These result highlight XGBoost's reliability and increasing use in PBEE.

2.2 Resampling Techniques

Class imbalance, where one or more classes in the target variable have far fewer instances than others, is a major issue in machine learning, as it often leads to bias towards dominant classes [8,10]. Although not applied to seismic performance levels, Zuhairi et al. [9] showed that SMOTE-Tomek enhances predictive performance in flood forecasting by reducing bias caused by the dominant class (non-flooding). This suggests the method may also work well for other imbalanced tasks, including PBEE classification problems. Similarly, Karampinis et al. addressed class imbalance in seismic vulnerability data by under-sampling overrepresented damage classes using Near-Miss [5]. Other studies have used different resampling techniques. For example, Edited Nearest Neighbors (EditedNN) has been used to improve performance by reducing noise in overrepresented classes [12]; While Adaptive Synthetic Sampling (ADASYN) has been applied to generate additional samples in underrepresented regions, helping models learn minority classes more effectively [13]. Synthetic Minority Oversampling Technique using Tomek links (SMOTE-Tomek), a hybrid method, has been shown to improve class separation and reduce bias in several imbalanced datasets [14]. While these techniques are not widely used in PBEE classification, they may prove useful in related fields facing similar data challenges.

2.3 Interpretability in Machine Learning (ML)

Understanding how machine learning models make decisions is important in earthquake engineering, where predictions affect public safety. A relevant example is the study by Hsiao et al. [15], which used SHAP to interpret an XGBoost model predicting liquefaction-induced lateral spreading. They found that peak ground acceleration (PGA), which is typically expected to increase the likelihood of lateral spreading, was incorrectly learned by the model due to biased training data. SHAP revealed this unexpected pattern, prompting a re-evaluation of PGA's influence and highlighting how interpretability can uncover non-obvious relationships and support better engineering decisions. SHAP, is a post-hoc interpretability method, meaning it is applied after a machine learning model has been trained, to explain how each input feature affects the model's output [16]. It helps researchers see which features increase or decrease the probability of predicting a particular class, both across the dataset and for individual predictions. Karampinis et al. [5] used SHAP to identify key building features that contribute to higher seismic vulnerability in rapid visual screening (RVS). Similarly, Kostinakis et al. [3] used feature importance plots to analyze how structural and seismic parameters affect predicted damage levels in reinforced concrete (RC) buildings. These studies show that interpretability tools can reveal patterns in structural behavior, which is very important for making informed decisions. In this study, SHAP and Partial Dependence Plots (PDP) are used to explore how resampling affects feature influence, especially for the underrepresented CP class.

2.4 Resampling and Interpretability in ML

Recent studies have shown that addressing class imbalance can improve not just predictive accuracy but also how well we understand the model's decisions. For example, Ariza-Garzón et al. [17], Alarab and Prakoonwit [18], and Moscato et al [19] used SHAP values in credit scoring and found that resampling made the explanations more reliable. Their models became better at showing how each feature affects the outcome. However, these studies focus mostly on binary problems (e.g., good vs. bad credit) and are not applied in earthquake engineering. In seismic performance classification, data is often imbalanced, and collapse-related cases like CP class are rarely seen. To the best of the authors' knowledge, no previous study has looked at how resampling affects SHAP-based explanations in this kind of application. This study addresses that gap by using three resampling methods: Edited Nearest Neighbors (EditedNN), Adaptive Synthetic Sampling (ADASYN), and SMOTE Tomek. These are used to examine how resampling changes model interpretability for the rare CP class in XGBoost.

3. METHODOLOGY

This section outlines the step-by-step procedures used in this study, including dataset preparation, the three resampling techniques, model training, and interpretability analysis focused on the CP class.

3.1 Dataset Description

This study used 2,744 instances generated by subjecting 56 existing RC buildings in Cotabato Province, Region XII, Philippines (Fig. 1), to 49 ground motions (GMs) (22 far-field, 27 near-field) adopted from FEMA P695 [20], (response spectra is shown in Fig. 2). Each instance has 10 features selected based on the relevancy within the context of PBEE, these are: height (H), slenderness ratio (SR), seismic weight (We), stiffness (Ke), fundamental period (T1), hysteretic energy (Eh), peak ground acceleration (PGA), and spectral accelerations Sa(T1), Sa(0.2), and Sa(1.0). SeismoLEE [21] was used to get the parameters for the ground motions, and the response spectra (5% damp) were created using PRISM v2025 [22]. Seismic performance levels (IO, LS, CP) were determined via the Capacity Spectrum Method (CSM) [23] and was implemented using the Nonlinear Static Procedure (NSP) in ETABS v2022 [24].The maximum drift ratio (MDR%) was computed from Equation (1), by normalizing the performance point (PP) displacement with building height. Based on FEMA 356 [25], IO corresponds to MDR <1%, LS to 1–2%, and CP to >2%. These thresholds are the basis for the classification of each building's seismic performance. Table 1 summarizes the dataset.

$$\text{MDR\%}=\frac{\text{Max. Disp. (PP)}}{\text{Building Height}}\text{x}100\% \qquad (1)$$

Table 1. Statistics of the 10 features used in this study

Features	Mean	SD	Min	Max
H(m)	11.88	7.55	5.30	36.00
SR	0.85	0.36	0.14	2.00
We(kN)	2.05e5	3.09E4	1.32E3	1.64E5
Ke(kN/m)	1.06e5	1.28E5	1.97E3	6.68E5
T1(sec)	0.92	0.48	0.309	2.79
Eh(kNm)	1.02e3	2.11E3	1.04	1.23E4
PGA(g)	5.00	2.23	1.49	12.46
Sa(T1)(g)	0.72	0.46	0.04	3.47
Sa(0.2)(g)	0.99	0.46	0.32	2.36
Sa(1.0)(g)	0.62	0.43	0.11	2.55

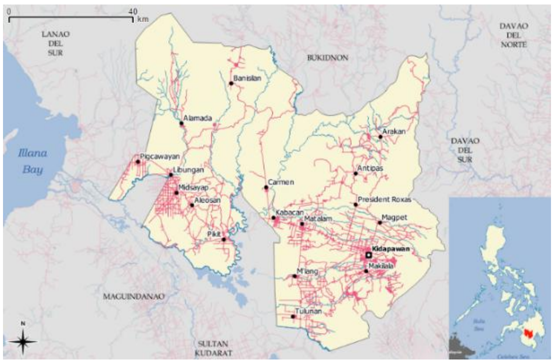

Fig. 1. Map of Cotabato Province showing locations of 56 studied buildings. *Adapted from PhilAtlas* [26].

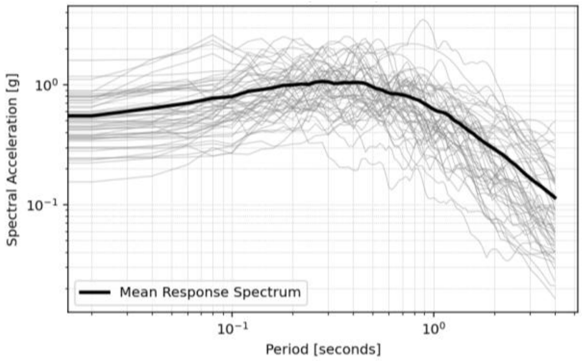

Fig. 2. Response spectra (5% damping) of the 49 GMs used in this study.

3.2 Data Preprocessing

Before training, the seismic performance levels (IO, LS, CP) were encoded as 0, 1, and 2 respectively, reflecting their ordinal nature. The dataset was divided into three parts: 70% for training, 15% for testing, and 15% for validation. Resampling was applied after this split to avoid data leakage. The test set was used to evaluate model performance, while the validation set was used during k-fold cross-validation to monitor learning progress and reduce overfitting. Statistical standardization was applied to all input features to help improve training stability and ensure consistent feature scaling.

3.3 Resampling of the Dataset

The study uses the `imblearn` library [27] and was coded in python. The original training dataset was resampled using these three techniques:

(1) EditedNN was applied using the “not minority” strategy, which resamples all classes except CP (the minority). It uses 3 nearest neighbors to check if a sample is surrounded by different classes. A sample is removed if most of its 3 neighbors belong to another class. Only samples with neighbors from the same class are kept.
Parameters: sampling_strategy='auto', n_neighbors=3, kind_sel='all', n_jobs=None

(2) ADASYN was applied using the "all classes" strategy, creating synthetic samples for all classes from the 5 nearest neighbors. A fixed random seed ensured consistent results.
Parameters: sampling_strategy='auto', random_state=42, n_neighbors=5

(3) SMOTE-Tomek was applied using the "not majority" strategy, resampling only LS and CP classes, excluding the majority class (IO). A fixed random number generator was used for consistent and reproducible results. The resampling also used both the default parameters for SMOTE and the ‘resample all classes’ strategy for the TomekLinks.
Parameters: sampling_strategy='auto', random_state=42, smote=None, tomek=None

3.4 XGBoost Training and Evaluation

For model training, the XGBoost classifier used a fixed hyperparameters to keep the pipeline process fast, efficient, and consistent for all resampling methods. This helps ensure that any changes in interpretability are due to the resampling techniques and not from tuning the model. The model was fixed and set to softmax multiclassification [multi:softprob], with the number of trees set to 100 [n_estimators = 100], a maximum depth of 6 [max_depth = 6], and a learning rate of 0.1 [learning_rate = 0.1]. Regularization was applied using L1 [reg_alpha=1] and L2 [reg_lambda=10] penalties. These hyperparameters were both applied to both the original and resampled training datasets, to focus on how resampling affects interpretability.

For model performance, this study used a One-versus-Rest (OvR) strategy with averaging (macro) to evaluate the model's performance for predicting three seismic performance levels: IO, LS, and CP. OvR transforms the problem into three binary classifications (Table 2), each class compared against the rest. Standard binary metrics were computed per class and averaged to ensure equal weight. The five metrics used were: Accuracy (Eq. 2), Precision (Eq. 3), Recall (Eq. 4), and F1-score (Eq. 5) as adapted from Kyaw et al. [28]. Model training and evaluation was coded in Python using the libraries `sklearn` [29] and `xgboost` [11]. TP – True Positive, FP – False Positive, TN – True Negative, FN – False Negative.

Table 2. OvR strategy by averaging the 3 binaries

	Positive Class	Negative Class
Binary 1	IO	LS & CP
Binary 2	LS	IO & CP
Binary 3	CP	IO & LS

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{2}$$

$$\text{Precision} = \text{TP/(TP+FP)} \tag{3}$$

$$\text{Recall} = \text{TP/(TP+FN)} \tag{4}$$

$$\text{F1 score} = \frac{2\times\text{Recall}\times\text{Precision}}{\text{Recall + Precision}} \tag{5}$$

3.5 XGBoost Interpretability Methodology: CP Class

This study used SHAP (SHapley Additive exPlanations) to explain how the XGBoost model makes predictions, with focus on the underrepresented CP class (positive class). The `TreeExplainer` tool from the SHAP library [16] was used because it works well with tree-based models like XGBoost. The three interpretability techniques are: (1) SHAP feature importance to find which features have influence for predicting the CP class, (2) SHAP waterfall plots are used to explain the contributions of features towards a logit [softmax] of a single CP class instance and (3) using Partial Dependence Plots (PDP) to see the global marginal effects of features towards the model's CP output under different resampling methods. These tools helped show both overall (global) and individual (local) feature effects, making it easier to understand how resampling changes the model's interpretability.

4. RESULTS AND DISCUSSION

4.1 Class Distribution After Resampling

The result shows the class distributions under different resampling techniques, revealing severe imbalance in the original training set (Fig. 3a) with only 6.7% in the CP class. This imbalance creates problems because the model may struggle to learn the rare CP class, making its predictions less accurate and harder to understand. EditedNN (Fig. 3b) improves CP representation to 14.4% by removing noisy majority class data, while ADASYN (Fig. 3c) and SMOTE-Tomek (Fig. 3d) achieve near-equal class distributions (about 33% per class). This supports the findings of Chi et al. [7], who showed that combining resampling with machine learning improves classification in seismic evaluation of school buildings using imbalanced data. Their study also highlighted that improving minority class learning helps avoid unnecessary retrofitting decisions, saving time and resources. Balancing the dataset helps the model learn each performance level better, especially the rare CP class, leading to fairer predictions and improved understanding of how decisions are made.

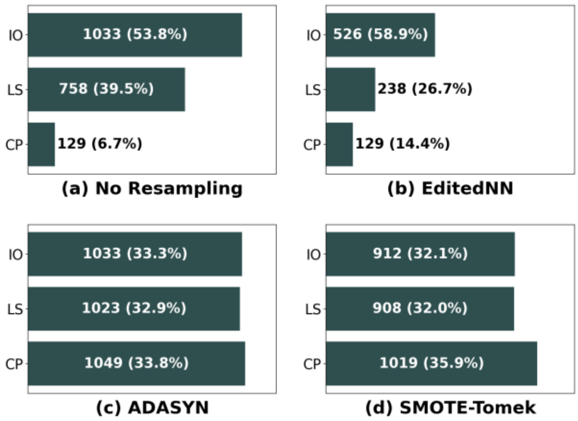

Fig. 3. Histograms of resampled training sets compared to the original (no resampling) training sets.

4.2 Model Performance Evaluation (Test Set)

Figure 4 shows that EditedNN achieved the best performance across all metrics, while SMOTE-Tomek and ADASYN also outperformed the baseline (No Resampling). These findings support Zuhairi et al. [9], who found SMOTE-Tomek improves model performance compared to no resampling, and López-García et al. [30], who showed ADASYN enhances minority class prediction. EditedNN's strong performance, is consistent with the previous study of Alarab and Prakoonwit [18], which the outstanding increase of model performance is attributed to the EditedNN's ability to remove noisy data. These results show that performance gains are linked to class distribution strategies, highlighting the role of resampling in boosting the performance in seismic classification tasks.

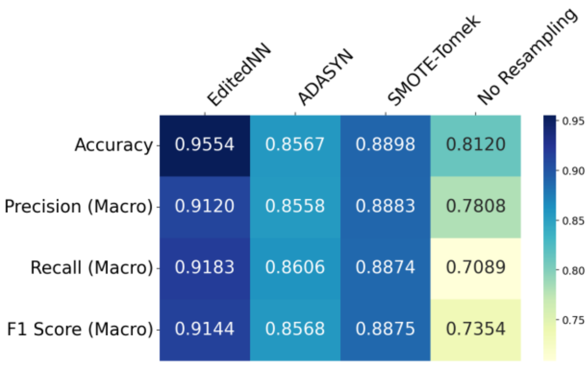

Fig. 4. XGBoost performance of test dataset across resampling techniques.

4.3 Feature Importance for CP Class Prediction

SHAP analysis (Fig. 5) shows that resampling, especially oversampling (ADASYN) and hybrid (SMOTE-Tomek), increases the influence of features in predicting the underrepresented CP class. ADASYN boosts the contribution of secondary features such as hysteretic energy (Eh) and seismic weight (We), indicating higher sensitivity to collapse related behavior. SMOTE-Tomek shows similar pattern but with improved performance. EditedNN (under-sampling), while greatly improving model performance, has little effect on feature importance, suggesting that it may help with accuracy but not with understanding how the model predicts the underrepresented class. This is in contrast with the study of Alarab and Prakoonwit [18] who observed otherwise in binary classification, implicating these multiclass problems could behave differently. Without resampling, the model's predictions relies heavily on a single feature, Sa(1.0), potentially overlooking important minority class (CP) indicators, such as We and Eh. These findings point out the value of resampling in enhancing both the performance and interpretability of the model for seismic risk assessment and PBEE.

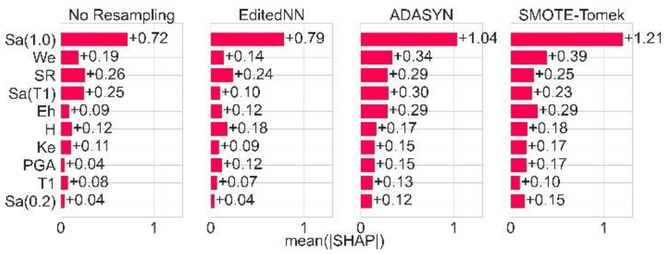

Fig. 5. SHAP feature importance of XGBoost with different resampling techniques for predicting CP class.

4.4 Local Interpretability for CP Class Prediction

To explore the influence of resampling techniques on XGBoost's local interpretability, SHAP waterfall plots (Fig. 6) and the corresponding logit–softmax summary (Table 3) were analyzed for a single instance that was both predicted and labeled as CP. The SHAP plots illustrate the additive contributions of input features to the CP logit prediction, f(x), while the table summarizes the logits and softmax probabilities for all seismic performance levels (SPLs), highlighting how resampling shifts the model's class confidence.

Table 3 shows that ADASYN produced the highest CP logit (2.307) and softmax probability (0.904), followed by SMOTE-Tomek (1.973, 0.884) and EditedNN (1.090, 0.707). Without resampling, the CP logit (1.602) has the lowest probability (0.579), indicating class confusion and weak model certainty as expected with this under represented CP data. Figure 6 shows how different resampling techniques influence the SHAP-based feature contributions towards predicting the CP class for the single instance. Without resampling (Fig. 6a), the model relied heavily on only two features: We and Sa(1.0), with limited input from others, resulting in a weak and narrowly focused prediction. EditedNN (Fig. 6b) provided a more distributed influence across features like SR, H, and PGA, though the contributions were modest and mixed in direction. ADASYN (Fig. 6c) produced the highest logit (2.307) and demonstrated the most balanced feature contributions, significantly enhancing the roles of SR, Ke, and H, which indicates a broader sensitivity to collapse-related behavior. SMOTE-Tomek (Fig. 6d) also achieved good interpretability by leveraging multiple features, including SR, Ke, We, and Sa(1.0). Overall, oversampling methods, particularly ADASYN and SMOTE-Tomek, not only improved model certainty in predicting the underrepresented CP class but also revealed a richer, more transparent feature interaction. These results show that using resampling techniques helps the model become more explainable, which is important for making better decisions in seismic risk and building performance assessments.

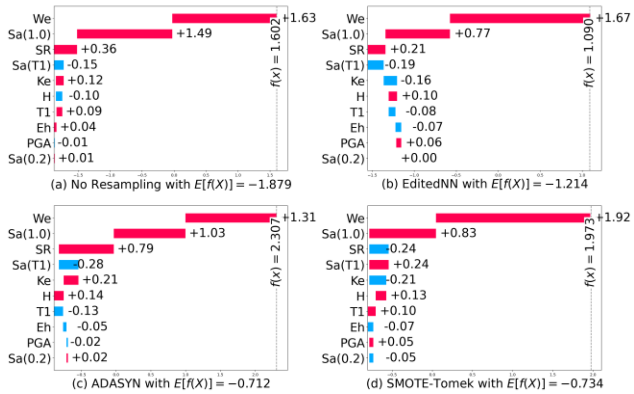

Fig. 6. SHAP waterfall plots for a single CP-class instance under different resampling techniques.

Table 3. Logit and softmax outputs formatted as logit [probability], with the highest value per row in bold.

	No Resample	Edited NN	ADASYN	SMOTE-Tomek
IO	-0.273 [0.089]	-0.537 [0.139]	-0.366 [0.062]	-0.838 [0.053]
LS	1.046 [0.332]	-0.436 [0.154]	-0.975 [0.034]	-0.669 [0.063]
CP	**1.602 [0.579]**	**1.090 [0.707]**	**2.307 [0.904]**	**1.973 [0.884]**

4.5 Global Interpretability for CP Class Prediction

To examine how resampling techniques affect the model's global interpretability of key features, this study uses Partial Dependence Plots (PDP), a global interpretability tool that shows the average effect of a feature on the model's output. PDP helps reveal whether a feature increases or decreases the predicted risk. The plots illustrate the marginal effects of each standardized input: Sa(1.0), We, SR, and Sa(T1), on the model's average predicted logit output E[f(x)]. Based on SHAP rankings (Fig. 2), these top four features were selected for discussion due to their strong influence and consistent trends across resampling techniques, as shown in Fig. 7.

The plot for Sa(1.0) reveals that higher spectral acceleration increases the model's predicted likelihood for the CP class. This effect is most pronounced in models trained with ADASYN and SMOTE-Tomek, which show steep logit increases beyond standardized Sa(1.0) values of zero, followed by a constant plateau. This trend aligns with structural engineering expectations, where more intense shaking typically correlates with greater collapse potential.

For Seismic Weight (We), the PDP curves show an early spike in predicted logit values at low weight levels, which then stabilize across the range. Models using ADASYN and SMOTE-Tomek maintain higher sensitivity throughout, suggesting these techniques help the model better account for the influence of structural mass on collapse potential. By contrast, EditedNN and especially the No Resampling model demonstrate flatter responses, indicating limited learning of mass-related effects, particularly in underrepresented collapse scenarios.

The plot for Slenderness Ratio (SR) shows that taller, narrower structures are generally associated with a higher predicted likelihood of collapse. This trend is again more evident in ADASYN and SMOTE-Tomek models, which capture the geometric instability risks linked with high SR values. The No Resampling model, however, shows minimal variation across the SR range, further emphasizing its reduced sensitivity to key collapse indicators. Finally, for Sa(T1), the PDP curves resemble similar to Sa(1.0), with oversampling techniques showing a stronger and more consistent rise in logits as Sa(T1) increases. This suggests that ADASYN and SMOTE-Tomek enable the model to better learn the role of resonance and dynamic amplification in a potential collapse behavior. Generally, the results show that resampling methods not only help the model make better predictions but also make it easier to understand how the model works. Models trained with oversampling (like ADASYN and SMOTE-Tomek) respond more clearly to important factors like earthquake intensity, building shape, and weight, making their behavior more in line with real engineering knowledge. In contrast, models without resampling might miss important indicators of collapse, especially in rare but risky building cases. These findings highlight the importance of using tools like PDP to check if the model is learning the right patterns and to help make better decisions in earthquake risk assessments and PBEE.

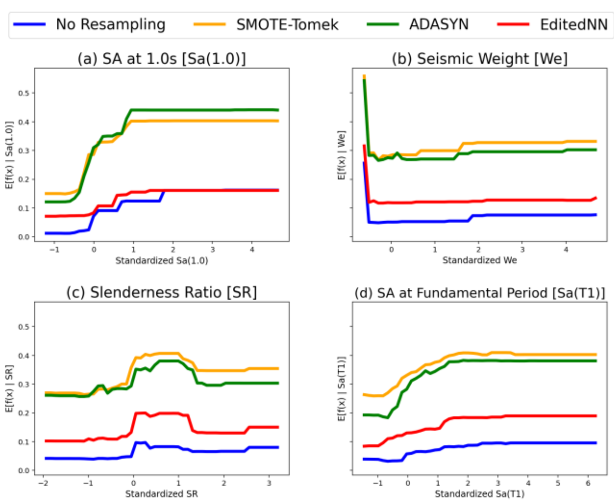

Fig. 7. PDP showing how the top four features influence CP prediction across resampling techniques.

5. CONCLUSION

This study examined how resampling techniques affect the interpretability and performance of XGBoost in predicting seismic performance levels, with focus on the underrepresented Collapse Prevention (CP) class. Resampling not only addressed class imbalance but also improved the model's ability to highlight meaningful features using SHAP and Partial Dependence Plots (PDP). EditedNN yielded the highest accuracy, while ADASYN and SMOTE-Tomek enhanced interpretability and model confidence for CP predictions. Without resampling, the model relied heavily on a single feature, Sa(1.0), while failing to capture the influence of other important predictors. In contrast, oversampling techniques like ADASYN and SMOTE-Tomek revealed a broader range of influential features, including Seismic Weight (We), Slenderness Ratio (SR), Hysteretic Energy (Eh), and Sa(T1). These features are essential for practical engineering insights and should not be ignored. These findings emphasize two key points. First, class balancing is critical not only for improving accuracy but also for ensuring fair and reliable predictions, especially for rare but safety-critical cases like CP. Second, model interpretability depends on the proper representation of rare classes during training. Oversampling enables the model to recognize multiple collapse-related factors, making its decisions more transparent and aligned with structural behavior. Resampling, therefore, is not optional, it is essential for building interpretable and trustworthy models in performance-based earthquake engineering (PBEE).

6. REFERENCES

[1] O. Coskun, R. Aktepe, A. Aldemir, A. E. Yilmaz, M. Durmaz, B. G. Erkal, and E. Tunali, Seismic risk prioritization of masonry building stocks using machine learning, Earthquake Engineering Struct. Dyn. 53 (2024) 4432-4450.

[2] N. Ravichandran, B. Bidorn, O. Mercan, and B. Paneerselvam, Data-Driven Machine-Learning-Based Seismic Response Prediction and Damage Classification for an Unreinforced Masonry Building, Appl. Sci.15 (2025) 1686.

[3] K. Kostinakis, K. Morfidis, K. Demertzis, and L. Iliadis, Classification of buildings' potential for seismic damage using a machine learning model with auto hyperparameter tuning, Eng. Struct. 290 (2023) 116359.

[4] H. Zhang, X. Cheng, Y. Li, D. He, and X. Du, Rapid seismic damage state assessment of RC frames using machine learning methods, J. Building Engg. 65 (2023) 105797.

[5] I. Karampinis, L. Iliadis, and A. Karabinis, Rapid Visual Screening Feature Importance for Seismic Vulnerability Ranking via Machine Learning and SHAP Values, Appl. Sci. 14 (2024) 2609.

[6] L. Iturburu, J. Kwannandar, S. J. Dyke, X. Liu, X. Zhang, and J. Ramírez, Towards rapid and automated vulnerability classification of concrete buildings, Earthq. Eng. Eng. Vib. 22 (2023) 309–

[7] N.-W. Chi, J.-P. Wang, J.-H. Liao, W.-C. Cheng, and C.-S. Chen, Machine learning-based seismic capability evaluation for school buildings, Autom. Constr., 118 (2020) 103274.

[8] M. S. Z. Bin Ahmad, N. A. Ab. Aziz, and L.H. Siong, Classification of Spine Abnormalities using Deep Learning, Proc. Int. Exch. Innov. Conf. Eng. Sci. IEICES 10 (2025) 998–1004.

[9] A. H. Zuhairi, F. Yakub, M. Omar, M. Sharifuddin, K. A. Razak, and A. Faruq, Imbalanced Flood Forecast Dataset Resampling Using SMOTE-Tomek Link, Proc. Int. Exch. Innov. Conf. Eng. Sci. IEICES 10 (2024) 845–850.

[10] V. Jayaramu, Z. Zulkafli, and F. Rahmat, Spatio-Temporal Prediction of Leptospirosis in Kuantan Using Hydrometeorological Variables and Random Forest Machine Learning, Proc. Int. Exch. Innov. Conf. Eng. Sci. IEICES 10 (2024) 916–922.

[11] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA (2016) 785–794.

[12] D. L. Wilson, Asymptotic Properties of Nearest Neighbor Rules Using Edited Data, IEEE Trans. Syst. Man Cybern. 2 (1972) pp.408–421.

[13] H. He, Y. Bai, E. A. Garcia, and S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, 2008 IEEE International Joint Conference on Neural Networks, Hong Kong, China, (2008) 1322–1328.

[14] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, A study of the behavior of several methods for balancing machine learning training data, ACM SIGKDD Explor. Newsl. 6 (2004) 20–29.

[15] C.-H. Hsiao, K. Kumar, and E. Rathje, Explainable AI models for predicting liquefaction-induced lateral spreading, arXiv (2024).

[16] S. Lundberg and S.-I. Lee, A Unified Approach to Interpreting Model Predictions, arXiv (2017).

[17] M. J. Ariza-Garzon, J. Arroyo, A. Caparrini, and M.-J. Segovia-Vargas, Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending, IEEE Access 8 (2020) 64873–64890.

[18] I. Alarab and S. Prakoonwit, Effect of data resampling on feature importance in imbalanced blockchain data: comparison studies of resampling techniques, Data Sci. Manag. 5 (2022) 66–76.

[19] V. Moscato, A. Picariello, and G. Sperlí, A benchmark of machine learning approaches for credit score prediction, Expert Syst. Appl. 165 (2021) 113986.

[20] C. B. Haselton and G. G. Deierlein, Assessing Seismic Collapse Safety of Modern Reinforced Concrete Frame Buildings, Blume Earthquake Engineering Research Center, Stanford University, Stanford, California, Technical Report No. 156, 2007.

[21] M. Fragiadakis, M. Georgioudakis, and P. Georgakis, SeismoLEE: An online tool for ground motion record processing, Laboratory of Earthquake Engineering, School of Civil Engineering, National Technical University of Athens (NTUA) (2021).

[22] S.-H. Jeong, K.-H. Lee, and W.-S. Jang, PRISM 2025: Platform for Integrated Seismic

Monitoring, Earthquake Engineering Research Group, Department of Architectural Engineering, INHA University, Incheon, South Korea, (2025).
[23] FEMA, Improvement of Nonlinear Static Seismic Analysis Procedures, Federal Emergency Management Agency, 2005.
[24] CSI, ETABS: Integrated Software for Structural Analysis and Design, (2022).
[25] FEMA, Prestandard and Commentary for the Seismic Rehabilitation of Buildings (FEMA 356), Federal Emergency Management Agency, Washington, DC, 2000, pp. 110-117
[26] PhilAtlas website: https://www.philatlas.com/mindanao/r12/cotabato.html/ (accessed 24.10.2024).
[27] G. Lemaître, F. Nogueira, and C. K. Aridas, Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning, J. Mach. Learn. Res. 18 (2017) 1–5.
[28] M. T. Kyaw, Y. N. Soe, and N. S. M. Kham, ADetect: Hybrid Analysis Feature Extraction for Android malware Detection, Proc. Int. Exch. Innov. Conf. Eng. Sci. IEICES 4 (2018) 27–31.
[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[30] A. López-García, O. Blasco-Blasco, M. Liern-García, and S. E. Parada-Rico, Early detection of students' failure using Machine Learning techniques, Oper. Res. Perspect. 11 (2023) 100292.