MACHINE LEARNING-BASED GROUND MOTION PREDICTION MODELS FOR INDONESIA: A PERFORMANCE EVALUATION OF XGBOOST AND RANDOM FOREST AGAINST CONVENTIONAL GMPE

Rayhan Irfan Hielmy[1] , Edy Santoso[2], Muzli Muzli[1*], Sigit Pramono[2], Setyoajie Prayoedhie[2]
[1]Geophysics Department, State College of Meteorological, Climatological, and Geophysics, Indonesia
[2]Indonesian Agency for Meteorological, Climatological and Geophysics, Indonesia.
**E-mail: muzli@bmkg.go.id*

Received: February 18, 2026 Reviewed: June 20, 2026 Accepted: June 26, 2026

ABSTRACT

Accurate ground motion prediction is critical for seismic hazard mitigation in Indonesia, a region characterized by complex tectonic settings. Conventional Ground Motion Prediction Equations (GMPEs) often struggle to capture the non-linear attenuation characteristics of local seismicity. This study develops national-scale Machine Learning (ML) models—specifically XGBoost and Random Forest (RF)—to predict Peak Ground Acceleration (PGA) and benchmarks them against the widely used Zhao et al. [23] model. The dataset comprises 20,287 strong-motion records from 1,573 events (M 1.6–7.9) recorded by 667 stations across Indonesia between January 2023 and April 2025. Performance evaluation reveals that the XGBoost model outperforms both RF and the conventional GMPE, achieving the lowest Mean Squared Error (MSE) of 0.8885. In contrast, the conventional model showed less reliability with a significantly higher MSE of 1.5905. Feature importance analysis indicates that hypocentral distance and magnitude are the dominant predictors, while site condition (Vs30) plays a secondary role. Validation on independent test sets and recent significant seismic events confirms the model's robust generalization capability compared to conventional approaches. These findings demonstrate that ML-based approaches provide a more reliable alternative for estimating ground motion in Indonesia, offering significant potential for improving early warning systems and hazard maps.

Keywords: Ground Motion Prediction Equation (GMPE); Machine Learning Seismology, Peak Ground Acceleration (PGA), Seismic Hazard Assessment, Indonesia

1. Introduction

Indonesia is considered one of the most earthquake-prone regions in the world. This is due to the country's geographical location at the intersection of the Indo-Australian, Eurasian, and Pacific tectonic plates. Earthquakes in this region tend to damage infrastructure and threaten the safety of residents. Therefore, estimates of the ground motion caused by earthquakes are essential. One of the most widely used methods is Peak Ground Acceleration (PGA), which measures the maximum acceleration in a given area. This value is used in assessing damage caused by earthquake [1] .

To date, the use of Ground Motion Prediction Equations (GMPE) remains the standard practice for estimating the level of ground motion caused by earthquakes. This method depends on the earthquake source, wave path, and local geology. Various GMPE models, including those developed by Zhao et al. [23], have been routinely used in engineering in Indonesia [2], [3], [4]. Conventional GMPE models struggle to fully account for how seismic waves propagate locally and the complex depth profiles of earthquake sources, particularly in the context of Indonesia's highly complex geology and tectonics. In addition, these models face significant limitations: classifying earthquakes based on their type—whether intraslab, interface, or shallow crustal earthquakes—as well as their source mechanisms. These classification challenges often complicate accurate modelling efforts in certain regions, especially in geographically complex areas such as Indonesia.

Recent technological advances have positioned Machine Learning (ML) as an innovative approach to predicting ground motion with greater precision. One of the advantages of ML is its ability to capture complex interactions and non-linear relationships that traditional statistical models cannot handle. By processing large datasets, ML generates predictive models that offer superior adaptability and flexibility. This method aligns with the concept of statistically enhanced ML, which uncovers patterns in data—including context-dependent patterns, high-level interactions, and non-linear relationships—that are typically overlooked by traditional statistical models [5]. As a result, ML provides a solution to overcome the limitations of GMPE in capturing the complexity of earthquake sources and propagation paths.

Predicting ground motion due to earthquakes through an ML approach has become a major focus of global research initiatives. Historically, the development of Ground Motion Prediction Models (GMPMs) has largely relied on traditional regression analysis, in which mathematical equations are integrated with predetermined coefficients to produce a predictive framework [6]. However, this method shows limited effectiveness in capturing complex nonlinear dynamics in seismic data. In contrast, ML algorithms—particularly Artificial Neural Networks (ANN) and Extreme Gradient Boosting (XGBoost)—excel at modeling nonlinear parameter relationships. Specifically, research by Mohammadi et al. [6] using the Turkish earthquake dataset demonstrates XGBoost's remarkable precision in predicting PGA values.

Japanese researchers have identified a critical limit whereby ML models tend to systematically ignore ground motion intensity during major earthquakes. This bias is primarily due to the limited availability of strong motion data from powerful earthquakes. To address this issue, they propose a hybrid approach that synergistically combines conventional ground motion prediction equations with machine learning algorithms. Empirical evidence shows that this combined methodology significantly reduces bias while providing much better accuracy than each approach separately [7].

In Indonesia—the country with the most intense seismic activity in the world—Rachmadan et al. [26] conducted a comparative assessment of three machine learning algorithms: Random Forest (RF), Gradient Boosting (GB), and ANN. The objective of this study focused on developing GMPM tailored to the seismotectonic characteristics of West Java. The method used combined topographic data—including elevation and slope measurements from each observation station—into a dataset that was systematically classified based on the type of earthquake source. The results show that all three ML models consistently outperform the classic approach of Zhao et al. [23], with the GB model providing the best performance in predicting PGA. Further analysis also reveals that epicentral distance (Repi) and magnitude (Mw) are the most influential variables, while Vs30 has the smallest contribution.

Khosravikia and Clayton [30] conducted a comparative analysis of several ML algorithms—RF, XGBoost, ANN, and Support Vector Machine (SVM)—using seismic datasets from Oklahoma, Kansas, and Texas covering the period starting in 2005. Their research shows that these ML algorithms excel at identifying critical physical parameters such as magnitude scale and distance attenuation effects without relying on predetermined equations or coefficients. Overall, the ML approach showed superior prediction accuracy compared to traditional linear regression, with Random Forest being the most effective method. However, conventional methods still have their own advantages in scenarios with limited data availability.

In Western Part of China, a Support Vector Regression (SVR)-based ground motion model (GMM) has been developed to predict Peak Ground Acceleration (PGA) and spectral acceleration response at various periods. This model uses earthquake magnitude (Mw), hypocenter distance (Rhypo), and average shear wave velocity of the top 30 meters of soil (Vs30) as parameters. The results of the study show that the SVR-based GMM is capable of providing accurate estimates and has good generalization capabilities for earthquake characteristics similar to the data used [8] .

Beyond scalar predictions, the methodological frontier of seismology has recently advanced toward Generative Waveform Models (GWMs) using denoising diffusion techniques [9]. These advanced generative models possess the capability to synthesize full broadband seismic waveforms rather than being limited to single scalar intensity measures like PGA. Furthermore, the open-source architectures associated with these models, such as HighFEM, offer immense potential as they could hypothetically be retrained or evaluated using Indonesian seismic networks. While the present study focuses on tree-based algorithms due to their interpretability and lower computational demands, acknowledging these generative approaches highlights the rapid, ongoing evolution of seismic hazard assessment tools.

This study aims to develop a PGA prediction model using Random Forest (RF) and XGBoost algorithms trained with comprehensive Indonesian earthquake data. Unlike previous studies, which were mostly limited to specific regions such as West Java, this study covers the entire Indonesian archipelago. This national coverage is expected to provide an overview of the characteristics of ground motion in the country. Furthermore, the findings from this study can serve as a basis for future focused investigations.

The selection of RF and XGBoost is based on their consistent and robust performance, which has been proven in previous studies. Both algorithms have clear advantages over conventional linear regression methods in modelling ground motion, thanks to their ability to handle complex and nonlinear parameter relationships. By leveraging the capabilities of ML to analyse complex nonlinear dynamics, the proposed model aims to overcome the major limitations inherent in traditional GMPE approaches. In particular, the classification of earthquake types and their source mechanisms remain a challenge that often hinders the development of truly accurate predictions. In addition to model development, this study also evaluates the relative

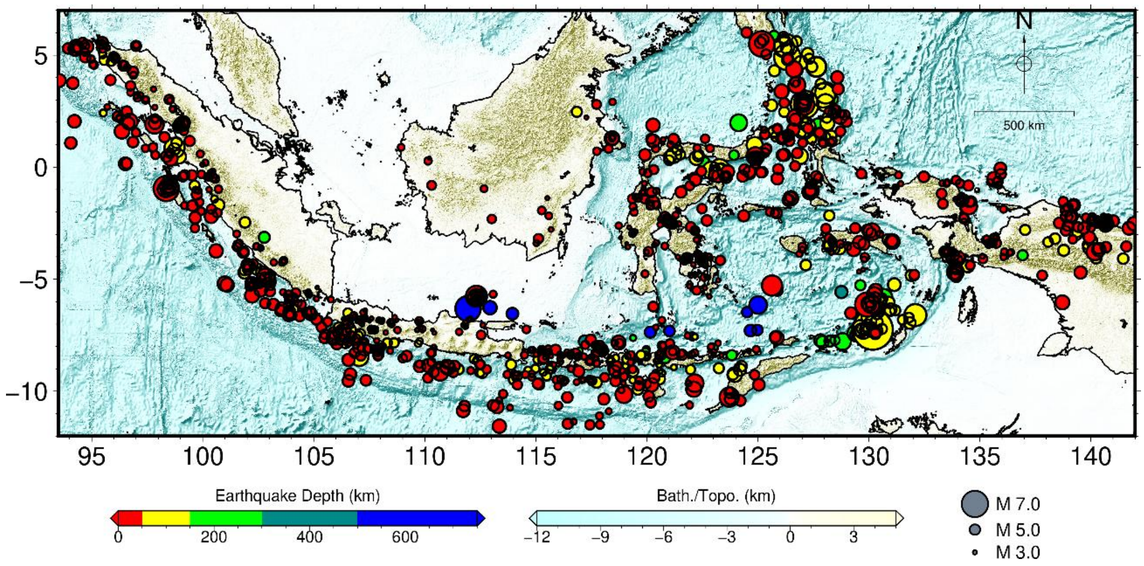

Figure 1 Earthquake distribution utilized in this study.

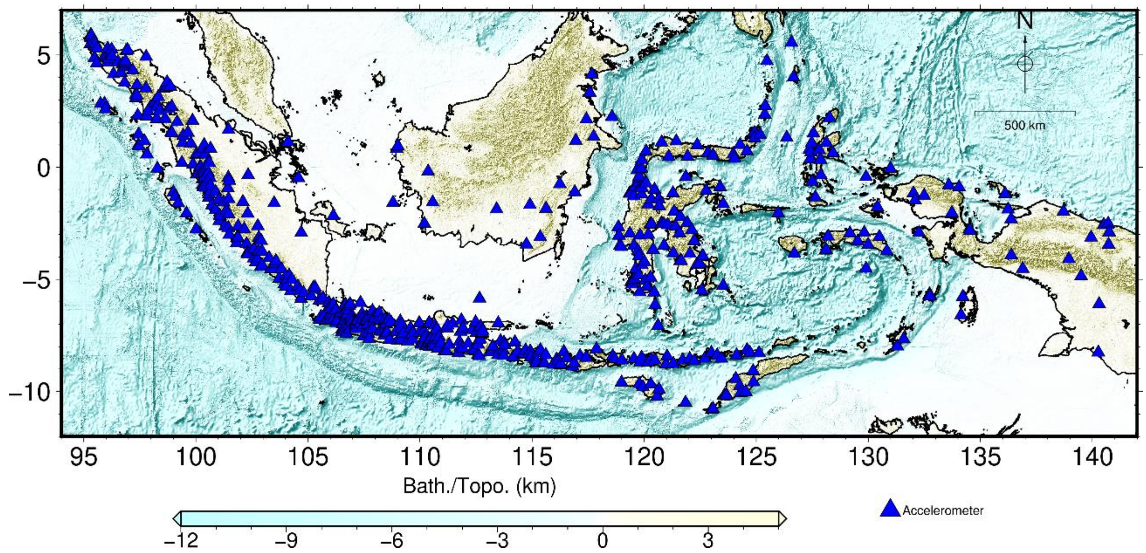

Figure 2 Station distribution utilized in this research.

performance of these two ML algorithms compared to traditional GMPE models that are widely used in Indonesia—particularly the GMPE developed by Zhao et al. [23]. This evaluation aims to provide deeper insights into the advantages and limitations of ML-based approaches in predicting ground motion in diverse seismic zones in Indonesia.

2. Methods

2.1. Dataset

In this study, the data used consists of 20,287 PGA measurements from 1,573 earthquake events recorded by 667 BMKG accelerometer stations throughout Indonesia between January 2023 and April 2025. Prior to building the dataset, a rigorous quality control process was applied based on BMKG standard procedures. This involved instrument response correction, baseline correction, applying a bandpass filter to ensure an adequate signal-to-noise ratio, and other standard data screening protocols. The PGA values utilized in this study represent the maximum absolute acceleration recorded across all three components (two horizontal and one vertical).

Earthquake data for August and September 2023 were excluded from the main dataset because they were used as random data for model testing. All data were recorded by 667 FBA sensors, with the distribution of earthquake epicentres and observation stations shown in Figures 1 and 2. The independent

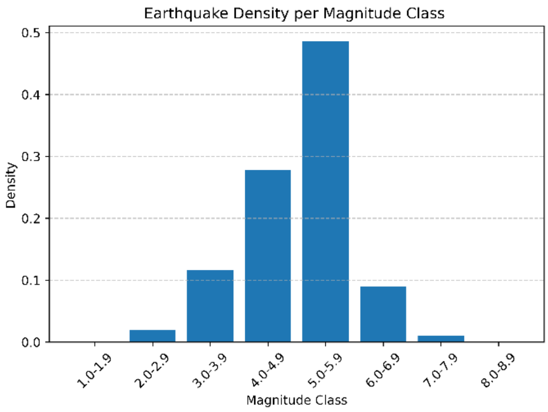

Figure 3 Earthquake density per Magnitude class.

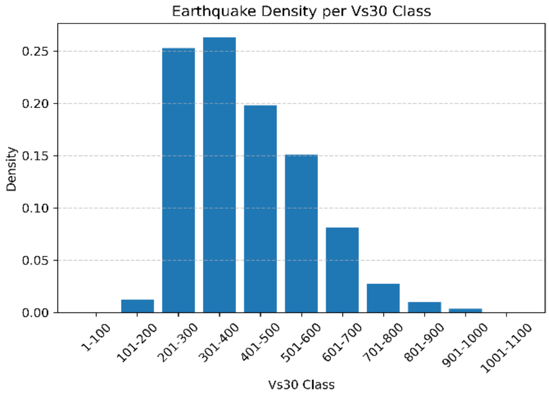

Figure 4 The histogram Density of the Vs30 (m/s).

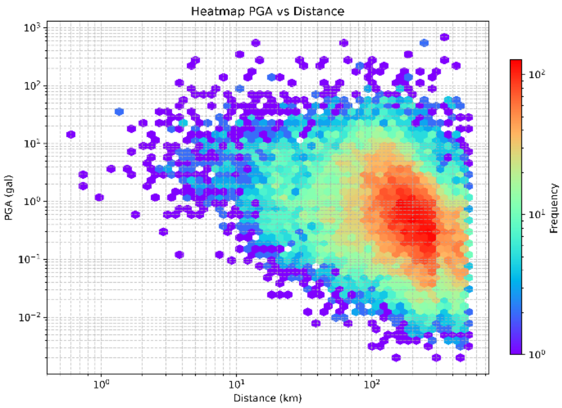

Figure 5 Heatmap PGA (gal) vs Distance (km).

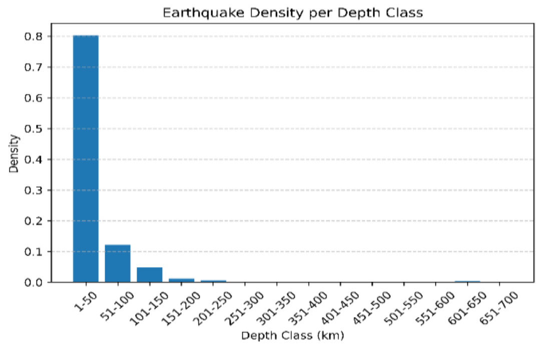

Figure 6 Earthquake density per depth class.

variables in this study include earthquake magnitude (M), which describes the earthquake source; focal depth and epicentre distance (Repi), which parameterize the earthquake source location [10]; and shear wave velocity in the soil layer to a depth of 30 meters (Vs30), which reflects local conditions. This study did not include fault depth and fault distance parameters because fault characteristic data for all earthquake events were not available.

The Vs30 parameter is used to represent site conditions. The Vs30 value is obtained based on the topographic slope proxy [11], [12], [13]. The data was obtained from the United States Geological Survey (USGS). Figure 3 shows a histogram of earthquake depth distribution, ranging from 1 to 643 km. Figure 4 shows a histogram of earthquake magnitude distribution, varying from M1.6 to M7.9. The epicentre distances used in this study range from 0.56 to 521.15 km, as shown in Figure 5, which displays a heatmap of the relationship between PGA (in gal) and epicentre distance, with the largest amount of data at distances greater than 100 km. Meanwhile, the Vs30 values used in this study range from 174.672 to 914.695 m/s, with the distribution shown in Figure 6.

2.2 Exploratory Data Analysis (EDA)

EDA is a statistical method used to identify trends and structures in data sets, thereby supporting the process of developing and refining hypotheses. EDA is important for obtaining an initial overview of the data before entering the stage of more in-depth analysis or hypothesis confirmation. This approach is exploratory and is not bound by specific model assumptions at the initial stage [14].

The main feature of EDA is its focus on understanding the substance of the data, which is answering fundamental questions about what is actually happening in the data. EDA utilizes graphical tools—including histograms, box plots, and heatmaps—to analyse data distribution, identify outliers, and reveal correlations between variables. In this framework, Pearson's correlation coefficient (r) serves as a quantitative metric for assessing the strength and direction of linear relationships between variables. These correlation values act as initial diagnostic indicators to signal potential basic associations before advanced modeling. However, we acknowledge that Pearson correlation is mathematically limited to capturing linear dependencies and cannot map complex non-linear structures. Therefore, this linear metric is used strictly for preliminary data exploration, while the actual non-linear interactions are captured fully by the tree-based machine learning algorithms.

In this study, EDA was systematically applied to describe the nature of the dataset and determine the optimal input parameters for Peak Ground Acceleration (PGA) prediction. The analytical approach used EDA techniques to identify underlying patterns, detect anomalous data points, and explain the relationships between variables. Specifically, we evaluated different input parameter transformations—mainly natural logarithm (ln) and base 10 logarithm (log10) conversions. The analysis focused on examining the correlation between PGA values and seismic parameters: magnitude (M), focal depth (Depth), epicentre distance (Repi), and shear wave velocity at a depth of 30m (Vs30), with logarithmic transformations applied subsequently.

In addition, we calculated the correlation coefficient (r) between the measured PGA values and each input parameter in its original version, which had been transformed using ln, and which had been transformed using log10. The optimal input parameters for ML modeling are determined through analysis of the average correlation values across all transformation combinations. In addition to parameter selection, the EDA methodology provides critical insights into data distribution patterns, identifies anomalous observations, and reveals relationships between parameters that may affect the performance of the PGA prediction model. The relationship between two variables, x and y, is shown by the formula used to calculate the correlation coefficient (r):

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} \tag{1}$$

2.3 Algorithms and Hyperparameter Settings

This study uses two ML algorithms—RF and XGBoost—to develop GMPM. This model integrates four main explanatory variables: earthquake magnitude (M), focal depth (Depth), epicentre distance (Repi), and average shear wave velocity at 30 meters above (Vs30). Meanwhile, the expected response variable is the PGA value. This study focuses on developing a GMPM to predict PGA values. Here is one way to formulate a model for predicting PGA values:

$$\hat{Y} = f(\mathrm{M}, \mathrm{Repi}, \mathrm{Depth}, \mathrm{Vs30}) \tag{2}$$

where $\hat{Y}$ represents the predicted PGA value in cm/s^2, and f(·) represents the nonlinear function formed by the ML algorithm used, either RF or XGBoost.

The data used to train and test the models in this study was taken from a uniform source. Before the training process was carried out, the dataset was randomly divided with 70% used as training data consisting of 14,200 samples and 30% as test data consisting of 6,087 samples. In developing GMPM, the hyperparameters in the RF model were set to n_estimators = 1000 and random_state = 42. For the XGBoost model, the initial settings were the same as RF, but max_depth = 6 and learning_rate = 0.1 were added to control model complexity and learning speed.

2.4 Random Forest (RF)

Individual decision trees generally exhibit significant variability and tend to overfit training data, so that although the model may capture training data very accurately, its performance often declines when it is used on new or unfamiliar data [15]. To overcome this shortcoming, the RF algorithm builds a set of decision trees using the bootstrap sampling approach. RF applies bootstrap aggregation to form an ensemble of decision trees, thereby effectively overcoming the limitations found in previous approaches. As a supervised learning algorithm, RF has gained widespread adoption in various applications—particularly in classification and predictive modelling—thanks to its robustness and flexibility [16], [17] .

When applied to regression with n observations, RF begins bootstrap aggregation by randomly selecting n samples with replacement to generate N resampled datasets. This process causes some samples to appear more than once, while others are not selected at all. Each resampled dataset then trains an unpruned decision tree, which is allowed to grow to its maximum depth.

At each split point, the algorithm does not consider all features, but only f features randomly selected from the total F features available in the dataset (f<F). In other words, F is the total number of explanatory variables (M, Repi, Depth, Vs30), while f is a small subset of these variables used to evaluate the split at a particular node. The selection of this subset encourages diversity between trees and reduces the correlation between predictors, thereby improving generalization ability and reducing the risk of overfitting. For unseen samples, predictions are obtained by aggregating the outputs of all decision trees through averaging. The prediction equation is formulated as:

$$Ypred = \frac{1}{N}\sum_{n=1}^{N} Tn(x) \tag{3}$$

where $T_n(x)$ represents the prediction of the nth tree for input x, and N represents the total number of trees in the forest. Figure 7 illustrates the workflow of the RF algorithm where individual decision trees systematically partition the seismic dataset based on localized feature thresholds. Each split within the ensemble evaluates specific combinations of magnitude, focal depth, epicentral distance, and proxy Vs30 values to minimize prediction variance. The final ground motion intensity prediction is obtained by aggregating the unweighted individual outputs from all trained decision trees through averaging. This operational framework is implemented via the Scikit-Learn library to ensure

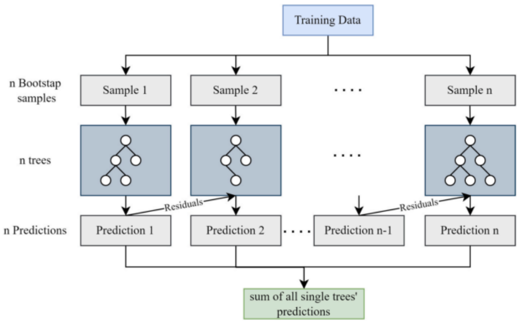

Figure 7 Extreme Gradient Boosting method framework [6].

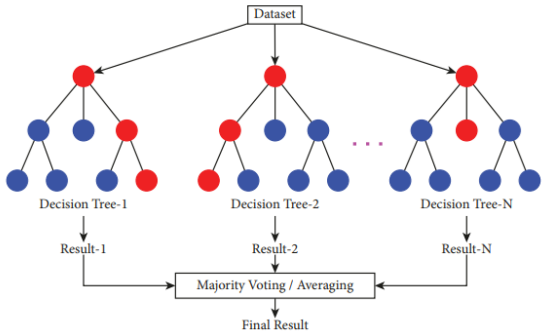

Figure 8 Random forest (RF) method framework [20].

high computational stability during national-scale processing [18],[19].

2.5 Extreme Gradient Boosting (XGBoost)

XGBoost is a highly effective implementation of the gradient boosting method. Its architecture integrates a regularization framework that reduces the risk of overfitting while increasing model simplicity (Chen and Guestrin, 2016). Thanks to its exceptional computational efficiency in processing large-scale datasets, XGBoost has become one of the most widely used ML approaches. However, this algorithm has a higher complexity and requires a deeper technical understanding compared to alternative boosting techniques [22].

Figure 8 illustrates the XGBoost algorithm framework. The prediction workflow begins with building a base decision tree from a randomly selected training subset. Subsequent trees are then trained on residuals—defined as the difference between actual values and the current model's predictions. Because this process is iterative, the model is continuously updated so that prediction accuracy improves each time new residuals are calculated. The loss function (L), which assesses the extent to which predictions differ from observed results, is represented by the first component of the objective function, which is used in conjunction with a penalty for model complexity to avoid overfitting. The objective function can be expressed as follows:

$$\sum_{i=1}^{N} L(yi, yi) + \sum_{i=1}^{t} \Omega(fi) \tag{4}$$

$$\Omega(f) = \gamma K + \frac{1}{2}\lambda \sum_{j=1}^{K} w_j^2 \tag{5}$$

where γ represents the complexity level of each leaf node in the tree, K refers to the total number of leaves, λ serves as a regularization factor, and w_j corresponds to the value assigned to leaf j.

2.6 GMPE Model Zhao et al. [23]

In this study, the GMPE used refers to the model developed by Zhao et al. [23]. This model aims to estimate PGA in cm/s² units by considering factors such as earthquake magnitude, epicentre distance, focus depth, and location conditions. In this study, there are several modifications to the original model, particularly related to the use of site class constants and the replacement of the moment magnitude (Mw) parameter with magnitude (M). The basic equation of the model used is [23] :

$$\ln(y) = \mathrm{aM} - \mathrm{bx} - \ln(\mathrm{r}) + \mathrm{e(h} - 15)\delta\mathrm{h} + \mathrm{FR} + \mathrm{SI} + \mathrm{SS} + \mathrm{SSL}\ln(x) + Ck \tag{6}$$

$$r = x + c\exp(dM) \tag{7}$$

In this equation, y represents the ground acceleration measured in cm/s², M refers to the magnitude of the earthquake, x is the horizontal distance to the earthquake source, h is the focal depth, and hc is the critical depth used in the calculation (in this study, it is taken as 15 km because it provides the

Table 1 NEHRP site classes [24]

Site Class	Soil Profile Name	VS30 (m/s)
A	Hard rock	> 1500
B	Rock	760 – 1500
C	Very dense soil and soft rock	360 – 760
D	Stiff soil	180 – 360
E	Soft soil	< 180

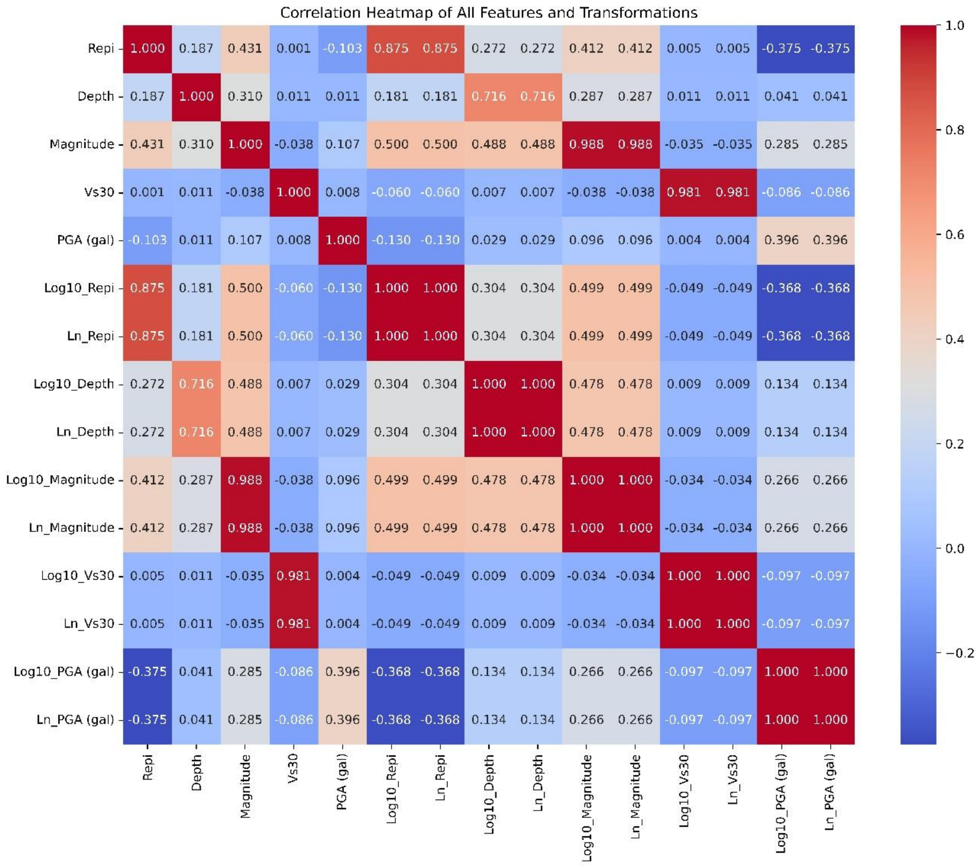

Figure 9 Correlation Heatmap of all features and transformation by EDA.

best depth effect for shallow crustal events), and δh is 1 if $h \geq hc$ and 0 otherwise. The parameters FR, SI, SSS, and SSL represent the source mechanism effect and source classification, but in this study these parameters are not used and are therefore ignored. In this study, modifications were made to the use of site class constants in the Zhao model, where Ck was calculated as $Ck = f \times SC$. The SC value was determined based on site classification (Table 1), with SC = 1 for hard rock, SC = 2 for rock, SC = 3 for very dense soil and soft rock, SC = 4 for stiff soil, and SC = 5 for soft soil. This modification aims to simplify the site class function in the model, while adjusting it to local geological conditions and available data limitations.

Although this approach ignores several detailed parameters found in Zhao's original formulation, the modification was justified in order to maintain a balance between model accuracy and data availability. Thus, the results obtained are still representative of regional conditions, but should be noted as a limitation of the study. Furthermore, constants a, b, c, d, e, and f were obtained through nonlinear regression analysis of observed ground acceleration records. This nonlinear approach is essential because the model incorporates exponential and logarithmic parameters—mathematical forms that cannot be addressed by standard linear regression alone. Through iterative optimization, this method systematically minimizes the discrepancy between field observations and model predictions in successive iterations. Convergence is achieved when the constant values show optimal predictive power relative to the underlying data distribution.

2.7 Evaluation metrics

This study uses three main metrics to evaluate GMPM performance: the coefficient of determination

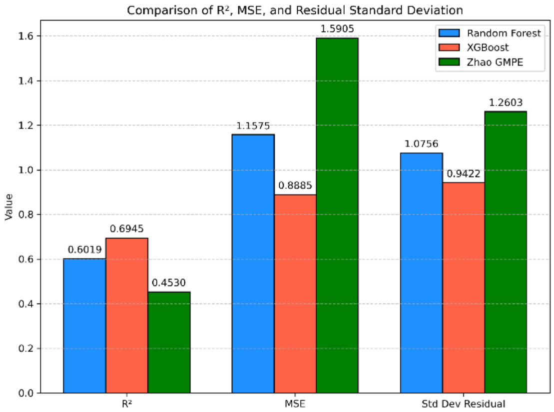

Figure 10 Evaluation of the performance of the three models using the R², MSE, and residual standard deviation metrics.

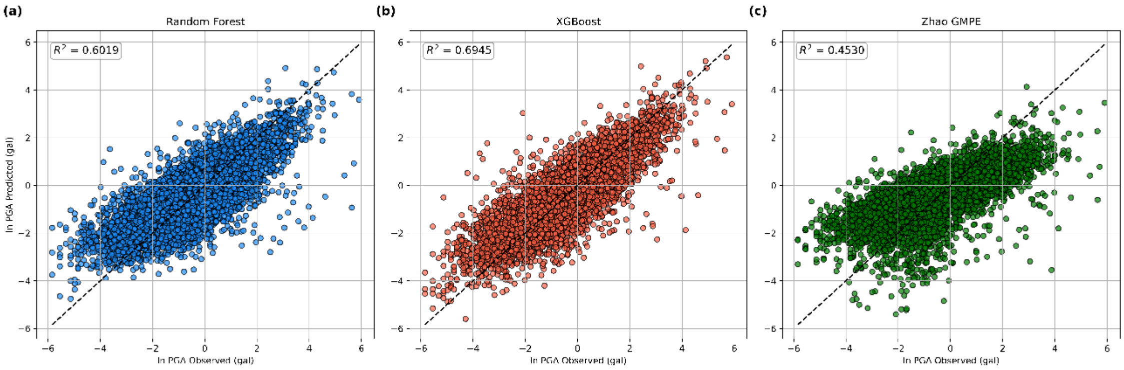

Figure 11 Comparison of observed and predicted PGA from three models using a natural logarithm scale.

(R^2), Mean Squared Error (MSE), and residual standard deviation. The R^2 metric measures the model's ability to explain the observed data variation. Conceptually, this metric represents the proportion of variation in the dependent variable explained by the independent variables in the model [25]. The R^2 value ranges from 0 to 1, and a value close to 1 indicates a better ability to explain the observed data variation. The mathematical formula for R^2 is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \bar{y})^2} \qquad (8)$$

The true value of the target in observation i is represented by yi in this formula. Meanwhile, $\bar{y}$ represents the average value of all yi in the complete dataset, and ŷi is the value predicted by the model for observation i. MSE measures the degree of discrepancy between the recorded PGA value and the value generated by the model. This metric is excellent for identifying outliers. The quadratic component of this function means that large errors have a much greater influence, so when the model produces highly inaccurate predictions, the contribution of these errors increases significantly [25]. Models with lower MSE values have better prediction accuracy.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2 \qquad (9)$$

In addition, the residual standard deviation assesses the spread of prediction errors and measures the extent to which the residuals deviate from their mean [26]. A lower value indicates a higher degree of agreement between the model's predictions and the observed data. This metric is calculated through:

$$\sigma_{res} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ei - \hat{e}i)^2} \qquad (10)$$

where:

- ei = yi - ŷi = residual at observation i
- ê = mean residual (usually close to zero)
- n = total number of data points

3. Result and Discussion

Two ML models—developed using RF and XGBoost—were compared with the conventional GMPE of Zhao et al. [23]. In the GMPE, parameters were recalibrated through nonlinear regression. The selection of these models was based on their widespread application in Probabilistic Seismic Hazard Assessment (PSHA) studies in the region and their operational use by the Indonesian Agency for Meteorological, Climatological and Geophysics (BMKG) for shakemaps [27], [28]. Model performance is evaluated using several metrics: residual standard deviation (σ_{res}), mean squared error (MSE), and coefficient of determination (R^2). In addition, probability density function analysis

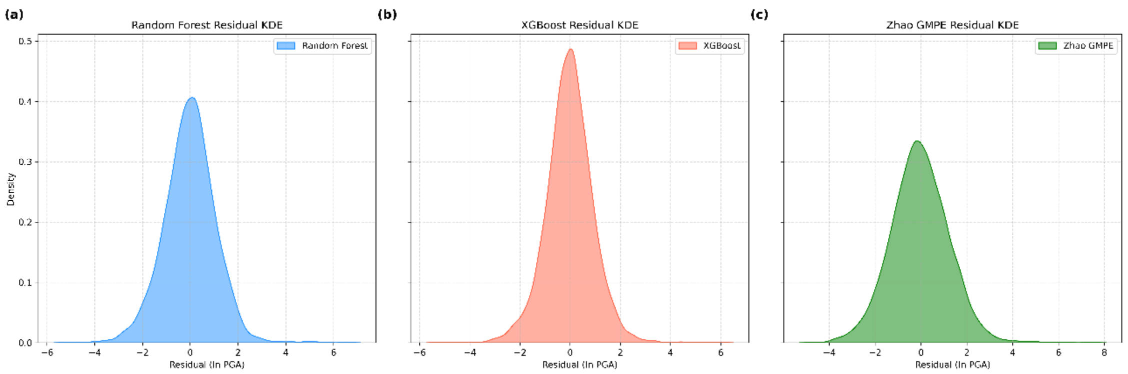

Figure 12 Distribution of predictions from three models with kernel density estimation on test data

examines residual patterns across various magnitude and distance ranges while describing their statistical distribution.

Based on Figure 9, a correlation was observed between all parameters analysed. The correlation results show that the PGA parameter has a negative relationship with Repi and Vs30. This means that as Repi and Vs30 increase, the PGA value tends to decrease. Conversely, the Magnitude parameter shows a positive correlation with PGA, which means that the greater the magnitude of the earthquake, the higher the PGA value.

However, there is an interesting finding related to the depth parameter. Theoretically, earthquake depth should have a negative correlation with PGA (the deeper the earthquake, the weaker the surface vibrations). However, the correlation results show that Depth actually has a positive relationship with PGA. This may be due to the low correlation coefficient between Depth and PGA, indicating that its influence on the model is insignificant, or it may be influenced by the distribution of the data used.

The best parameter used in ML modelling is Ln(PGA), with a correlation value of -0.375 with Repi, 0.285 with Magnitude, 0.134 with $\log_{10}$(Depth), and -0.097 with $\log_{10}$(Vs30). The average absolute correlation value is 0.223, which is considered low. This is likely due to the large number of parameters involved and the possibility that other parameters that have not been analysed also affect the PGA value.

Figures 10a–c shows the coefficient of determination graphs for the three prediction models tested. A comparative analysis shows that the XGBoost and Random Forest algorithms perform much better. These two models show the highest coefficient of determination among the models tested. The XGBoost model proved to be the most superior with an R^2 value of 0.6945, followed by the Random Forest model with an R^2 value of 0.6019. Meanwhile, the Zhao GMPE model produced the lowest R^2 value, which was 0.4545. These results indicate that the XGBoost model is capable of providing more accurate predictions than the other two models.

Figure 11 shows that ML-based models perform better than conventional methods. In addition to showing higher R^2 values, both ML models (XGBoost and Random Forest) also show better performance in MSE. The MSE values for XGBoost and Random Forest are recorded at 0.8885 and 1.1575, respectively, which are lower than the MSE value of Zhao's GMPE model at 1.5905. A similar trend was also seen in the residual standard deviation values, where the XGBoost and Random Forest models produced values of 0.9422 and 1.0576, respectively, which were lower than the residual standard deviation of the Zhao GMPE model, which was 1.2603.

In addition to applying the key evaluation indicators mentioned earlier, additional residual checks were performed to measure how well the three models performed. Residuals describe the extent to which the model results differ from actual observations. Kernel Density Estimation (KDE) graphs of the residuals generated by each method are shown in Figures 12a–c. By visualizing the distribution of residuals, these graphs provide important information about the accuracy of predictions and the possibility of bias in each model.

Overall, the residual distribution pattern between the estimated results of the three models and the observed data is close to normal and centered at zero, indicating that the model predictions do not show any recognizable systematic bias. Among all models, XGBoost produces a KDE graph with the tightest spread and highest peak, indicating excellent prediction accuracy and consistency. The Random Forest model shows a slightly lower distribution peak and wider residual spread compared to XGBoost. Meanwhile, the Zhao GMPE model displays a lower peak and more scattered residual distribution compared to Random Forest, indicating relatively less consistent prediction performance.

Figures 13a-c, 14a-c, 15a-c, and 16a-c illustrate the residual patterns with respect to epicentre distance, magnitude, Vs30, and depth for all analysed models. The purpose of these graphs is to evaluate whether the residual distribution is consistent with the characteristics of stochastic error—that is, randomly scattered without a systematic pattern.

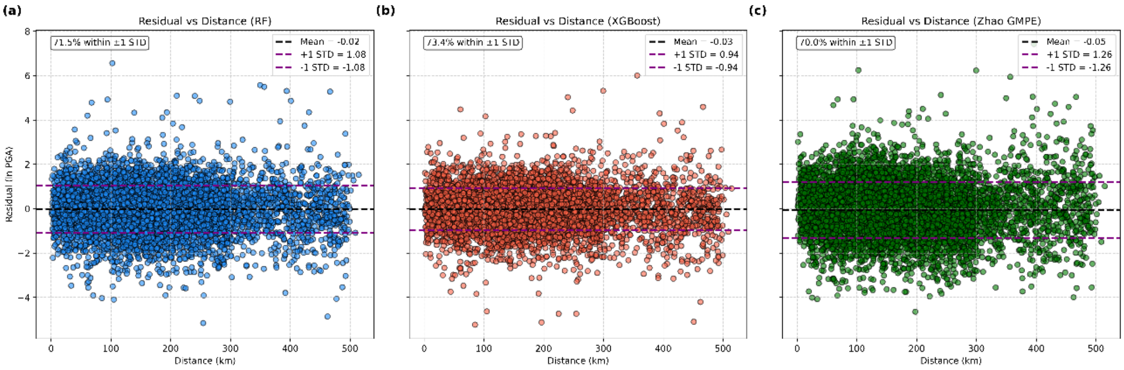

Figure 13 Relationship between residual change and distance (km) for the three prediction models analysed.

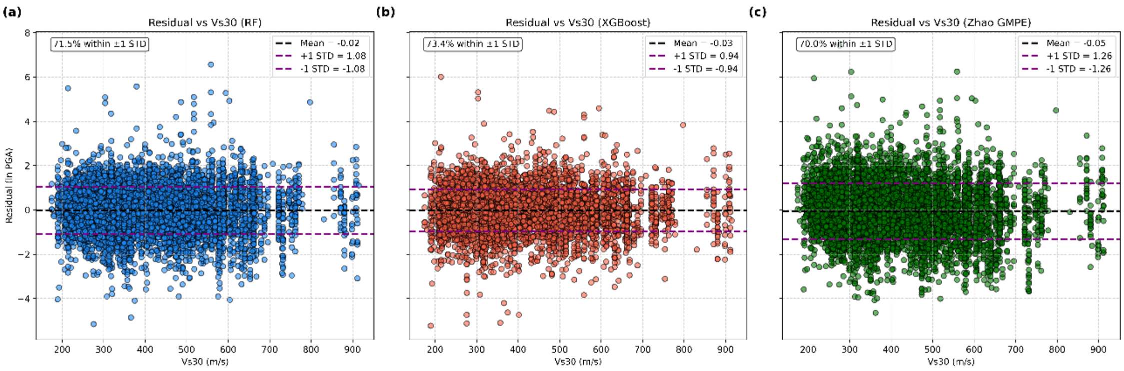

Figure 14 Relationship between residual change and Vs30 (m/s) for the three prediction models analysed.

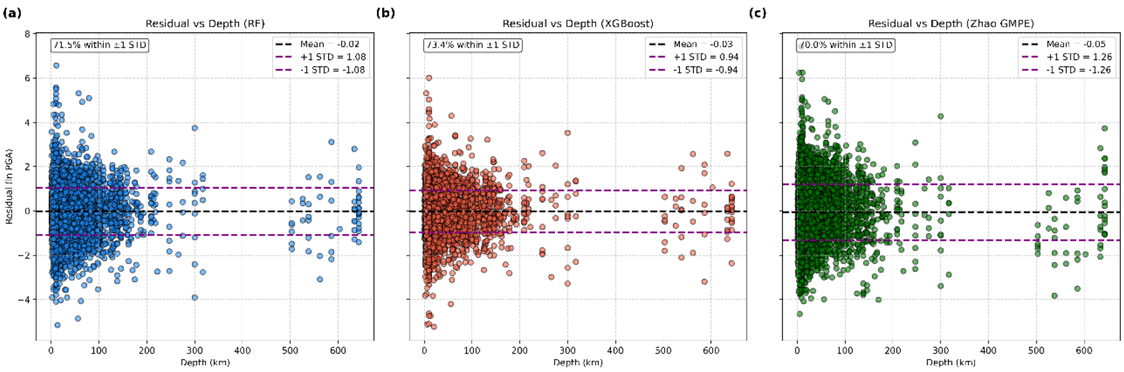

Figure 15 Relationship between residual change and depth (km) for the three prediction models analysed.

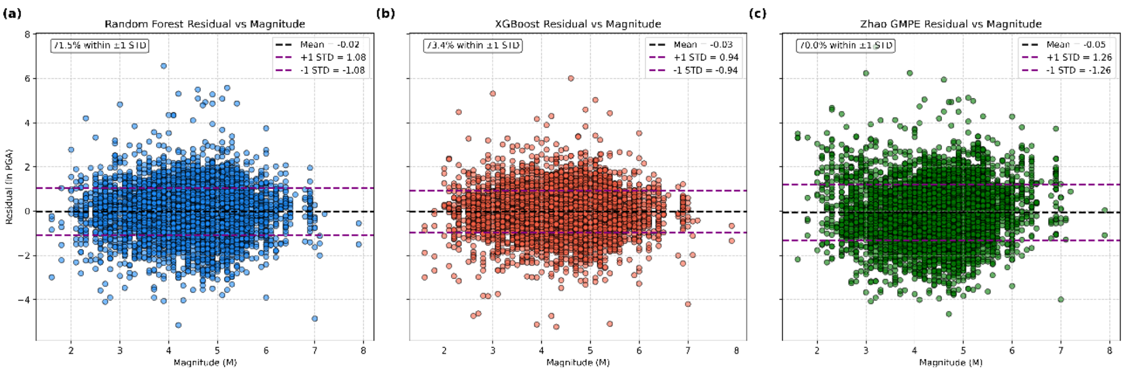

Figure 16 Relationship between residual change and magnitude (M) for the three prediction models analysed.

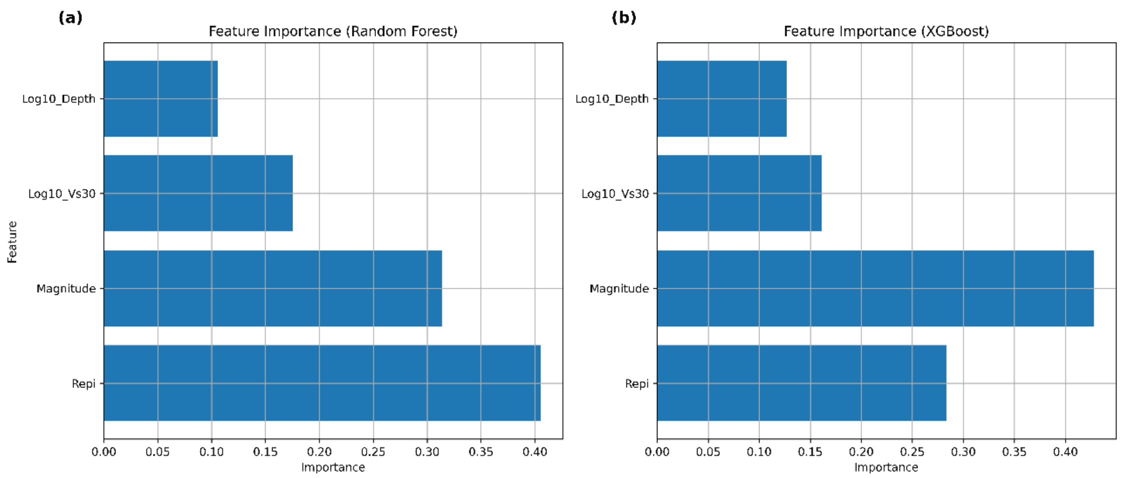

Figure 17 Feature Importance plot of Random Forest and XGBoost models.

Observations show that the residuals from the three models are randomly scattered around the mean line, which is close to zero. This indicates that the models do not show significant bias toward specific input variables and are able to represent the data well. The proportion of residuals within the standard deviation range was recorded at 73.4% for the XGBoost model, 71.5% for the Random Forest model, and 70.0% for the Zhao GMPE model. These figures indicate that most predictions are within an acceptable error range, with XGBoost showing the best performance in maintaining prediction consistency with the observed data.

In terms of epicentre distance, all three models show consistent performance at both short and long distances, without showing any systematic residual patterns. However, for magnitude, the data tends to fall within the range of 3 to 6. Within this range, many residuals are classified as outliers. Outside this range, data is very limited, making model evaluation at extreme magnitudes less representative. For the Vs30 parameter, most data falls within the range of 200 to 600 m/s, and the residual distribution is relatively even. Meanwhile, data is limited in the range of 700 to 900 m/s, where residuals tend to fall into the outlier category.

As for earthquake depth, most of the data is concentrated at depths of less than 200 km. There is a gap in the data between 200 and 500 km, and only a small amount of data is available above 500 km. Interestingly, many outliers are found at shallow depths, while at depths above 500 km, the residuals tend to remain within normal limits (non-outliers).

Overall, the residual distribution patterns of the three models (Random Forest, XGBoost, and Zhao GMPE) are relatively similar. However, the Zhao GMPE model shows a more uneven distribution compared to the other two ML models. This unevenness is in line with the lower R^2 value and higher MSE in the Zhao model. The evaluation results through performance metrics and residual analysis show that the model using XGBoost provides the most optimal results, with more stable and accurate predictions compared to the other models tested.

3.1 Model Explanation

The GMPE model proposed by Zhao has been evaluated using a nonlinear regression approach, resulting in the following equation:

$$\begin{aligned} \ln(PGA) = 1.076\mathrm{M} - 0.00313\mathrm{Repi} \\ - \ln(\mathrm{x} \\ + 53.7\exp(-0.7\mathrm{M})) \\ - 0.0004(\mathrm{h} - 15) \\ + 0.06\mathrm{SC} \end{aligned} \qquad (11)$$

In this equation, the coefficients for the magnitude (M) and site class (SC) parameters are positive. This indicates that higher magnitudes and softer soil conditions at the observation site tend to strengthen the PGA value. Conversely, the epicentre distance (Repi) has a negative coefficient, indicating that the PGA value tends to decrease as the distance between the epicentre and the station increases. For depth (h), the coefficient is negative as theoretically expected, although previous analyses found a positive correlation with PGA. This difference is likely related to the low correlation strength of depth with PGA and the distribution of the data used.

One of the main obstacles in ML-based models is their hidden nature, like a “black box,” which makes it difficult for researchers to understand the decision-making mechanisms within the model. To improve readability, we applied the feature importance approach provided by the Scikit-learn library. This method was chosen for its computational efficiency, ease of implementation, and widespread use in various scientific studies. However, we acknowledge that impurity-based importance in tree ensembles can be inherently biased toward continuous variables with high cardinality [29]. Consequently, future studies

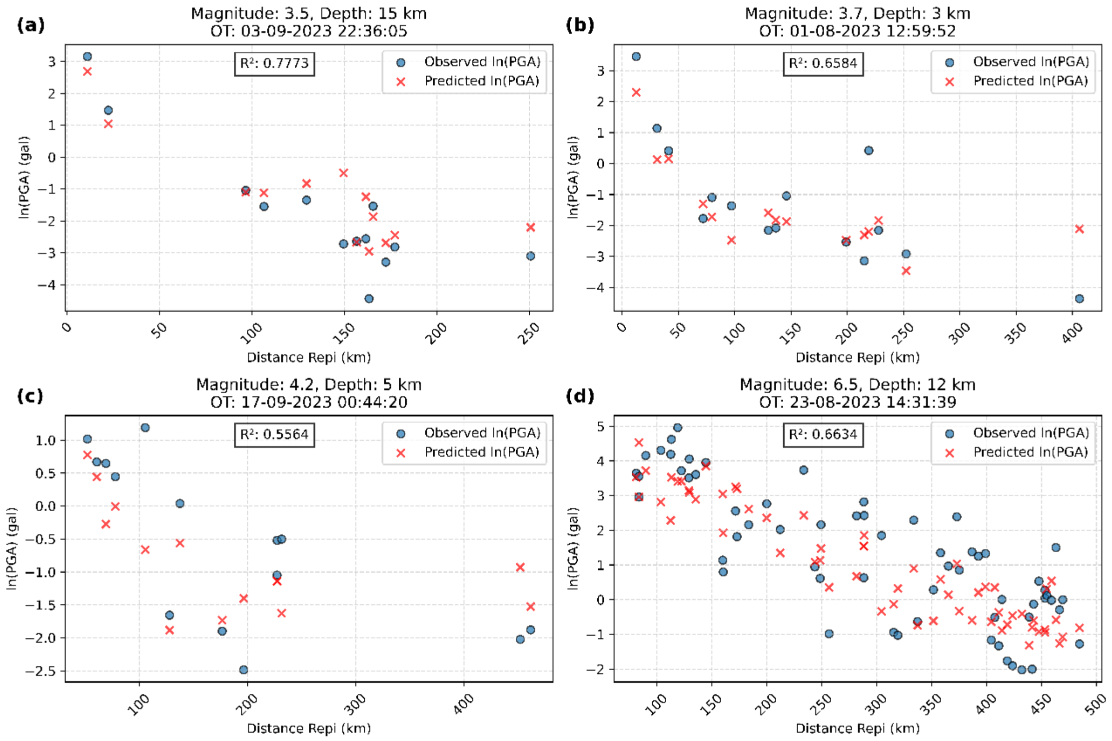

Figure 18 Comparison of observed and predicted values for four random earthquakes using the XGBoost model.

will implement SHAP (SHapley Additive exPlanations) values to provide a more robust and unbiased evaluation of feature contributions.

The results of the variable contribution analysis are shown in Figures 17a-b, which show feature importance in the Random Forest and XGBoost models. Based on these results, epicentre distance (Repi) and magnitude (M) are the variables with the greatest contribution in predicting PGA values. On the other hand, focal depth shows the lowest impact, with Vs30 being slightly higher.

The low importance of the depth variable in this study is likely due to the uneven distribution of data. Most data points are concentrated at depths of less than 200 km, while data for depths greater than 200 km is very limited. This unevenness may reduce the influence of the depth variable on the shape of the prediction model.

The distance between ground motion monitoring stations and the earthquake source is generally represented by the hypocentre distance (Rhypo) in previous studies. This parameter measures the hypocentre distance—the measurement between the monitoring station and the location of the earthquake source on the fault plane. By combining the epicentre coordinates and depth, Rhypo provides a more accurate source-to-location distance measurement than the epicentre distance alone [30]. Rhypo effectively overcomes the limitations arising from uneven depth distribution by functioning as a good distance parameter. Meanwhile, Vs30 shows a low level of importance because its value is obtained indirectly through a global model based on topographic slope proxies, rather than from MASW or borehole measurements. As a result, the actual site conditions are not fully represented.

Many studies highlight that Vs30 plays a critical role when obtained through direct measurement methods. For example, Seo et al. [31] showed that Vs30 values obtained from MASW ranked second after earthquake magnitude and epicentre distance in predicting ground motion intensity—especially in the seismic region of South Korea. Similarly, Japanese research consistently emphasizes the integral role of Vs30 in developing ground motion prediction models for shallow crustal earthquakes, with findings largely based on borehole datasets [32]. The results show that Vs30 emerges as the main parameter when local site conditions are analysed using data obtained from MASW/borehole [33]. These findings emphasize the importance of expanding research in Indonesia to obtain more accurate Vs30 values through direct field measurements. High-quality Vs30 data will not only improve the accuracy of PGA predictions but also enable the development of predictive models for additional parameters—such as Peak Ground Velocity (PGV)—in future research.

3.2 Performance Validation Using Random Seismic Data

As explained earlier, the XGBoost model showed superior performance compared to other algorithms, so it was chosen as the GMPM for this field of research. In addition to using test data, this study also conducted additional validation by utilizing a dataset that was not included in the training or testing process, with the aim of evaluating the model's performance more objectively. This dataset consists of four randomly selected earthquake events that occurred in August and September 2023, namely: M 3.5 at a depth of 15 km on September 3, 2023; M 3.7 at a depth of 3 km on August 1, 2023; M 4.2 at a depth of 5 km on September 17, 2023; and M 6.5 at a depth of 12 km on August 23, 2023.

Figures 18a-d show graphs comparing the observed values with the predicted values from the XGBoost model at each recording station for the four earthquake events. The R^2 values obtained for each event—0.7773, 0.6584, 0.5564, and 0.6634—indicate a strong correlation and are close to the overall R^2 value of the XGBoost model (0.6945). The analysis results show that this model is capable of estimating ground movement levels with high accuracy, including when applied to new earthquake data that has never been used in the model training and testing stages.

3.3 Limitations and Future Works

While the XGBoost model demonstrates promising preliminary performance for predicting PGA in Indonesia, this national-scale study possesses several methodological limitations that will be addressed in future developments. First, the current ML models and the modified Zhao (2006) GMPE were developed without source classification terms such as interface, intraslab, and shallow crustal settings. This specific design was intentionally chosen to test predictive capabilities using only rapidly available parameters for potential rapid-assessment applications where focal mechanisms are not immediately known. Consequently, the modified GMPE served solely as an input-equivalent baseline rather than a fully realized physical model.

It is fully acknowledged that ignoring these complex tectonic source types compromises the overall physical validity of the predictions. Furthermore, the utilization of a purely random train/test split introduces a potential risk of data leakage across the events, a well-documented issue in GMPE-ML [34], [35]. The choice of epicentral distance as a primary feature is also geometrically suboptimal for deep intraslab earthquake events within the region. Finally, the present evaluation relies heavily on aggregate error metrics without incorporating a standard residual decomposition protocol as expected by the seismic community [36].

Future studies will prioritize addressing these structural and methodological challenges to improve prediction reliability. The upcoming model iterations will incorporate Slab2.0 geometry to achieve a rigorous and physically sound source classification. Transitioning to hypocentral distance will provide a more accurate parameterization of wave propagation paths across varying depths. Additionally, implementing event-based cross-validation and full between-event and within-event residual decomposition will align the research with current seismological best practices.

4. Discussion

XGBoost consistently outperforms Random Forest thanks to its systematic and adaptive learning framework. Unlike Random Forest's bagging approach—which builds multiple independent decision trees that are combined through averaging or voting—XGBoost's boosting methodology sequentially builds decision trees, with each new tree specifically targeting the error from the previous iteration. During optimization, XGBoost improves model accuracy by utilizing the first and second derivatives of the loss function. Its integrated regularization mechanism simultaneously reduces the risk of overfitting. Additional advantages—such as adaptive learning rate control, automatic handling of missing data, and high computational efficiency—collectively support XGBoost's superior performance on large and complex datasets.

Residual analysis provides critical diagnostic insights into model accuracy and potential bias. Kernel Density Estimation (KDE) visualization shows that most models exhibit minimal bias, as evidenced by residual distributions centred around zero. However, the presence of outliers and secondary peaks in the Random Forest and Zhao GMPE predictions indicates systematic bias. This inaccuracy is likely due to data limitations—particularly within certain magnitude ranges—or unrecorded local geological conditions. In contrast, the XGBoost residual distribution shows better symmetry and simplicity, indicating higher resistance to outliers and better, more consistent predictions.

Because it provides more uniform and consistent coverage across the region without relying on limited field measurement data, the use of Vs30 data from the USGS global model on a national scale, such as in Indonesia, is considered relevant. This model is based on a topographic slope proxy that has been shown to have a strong correlation with Vs30 values worldwide[12], [13]. The evaluation study also showed that the performance of proxy models, including slope-based approaches, did not differ significantly from geological or topographical estimates when compared to measured VS30 data [37]. Therefore, the use of the full USGS global dataset in this study is considered sufficiently representative as a basis for preliminary analysis of PGA values in Indonesia, especially on a national scale, prior to further validation with field measurement data.

Based on Figure 18, it can be seen that the predictions from XGBoost have a consistent pattern, namely that the further the distance from the epicentre, the lower the PGA value. However, at some points there are small fluctuations that are not significant. This is due to other factors that affect each location, one of which is the Vs30 value. This finding is consistent with Figure 17, which shows that the earthquake epicentre distance variable has a high level of importance, while Vs30 has a smaller influence.

In Figure 18b, the prediction for the M 3.7 earthquake shows that there are stations at a distance of 200–250 km and around 400 km whose PGA values deviate from the XGBoost model prediction path. However, the main prediction pattern remains consistent and does not deviate too far. This indicates that the spike in PGA values at several stations is likely influenced by other local factors that are not represented in the model.

Furthermore, in Figure 18d for the M 6.5 earthquake, it can be seen that the PGA prediction pattern follows a downward trend as the distance from the epicentre increases. As in the previous case, there are small fluctuations at several points that are influenced by the Vs30 value. Overall, these results show that the XGBoost ML method is capable of providing PGA value predictions with fairly good and consistent performance.

However, it should be noted that the prediction results obtained from statistical methods, including ML algorithms, can only be considered valid to the extent that the data used to train the model [30]. GMPM is built using specific historical data as the basis for training, so its ability to predict events outside the scope of the training data tends to be limited. Therefore, this model needs to be updated regularly by adding more data and variations, especially for magnitude ranges that currently have limited representation. In addition, incorporating additional geophysical variables—such as location elevation, slope, fault mechanism, earthquake source classification, radiation patterns, and directivity effects—as well as information about local subsurface structures, will improve the model's accuracy. Overall, the use of the latest earthquake data and real-time model adjustments are essential to maintain the reliability and accuracy of predictions in the long term.

In line with this, the use of the latest earthquake data and real-time model adjustments are important aspects for maintaining the reliability of long-term predictions. This is in line with the view that earthquake risk mitigation practices in various countries begin with a reliable risk assessment before determining building retrofit targets and strategies. Thus, the development of ML models based on more representative earthquake data has the potential to strengthen the risk mitigation process in Indonesia. These models not only provide more accurate predictions but can also serve as a stronger basis for the formulation of future seismic resilience policies [38].

5. Conclusion

This study developed a Ground Motion Prediction Model (GMPM) specifically for Indonesia as preliminary research to improve the accuracy of ground motion predictions. The data used consisted of 20,287 PGA records from 1,573 earthquake events recorded by 667 BMKG accelerometer stations throughout Indonesia (January 2023–April 2025). Data from August–September 2023 was intentionally excluded as a validation set, resulting in a magnitude range of 1.6–7.9. The data was divided into training and testing partitions with a 70:30 ratio.

We implemented prediction models using Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms, comparing their performance with the conventional GMPE approach of Zhao et al. [23]—which was recalibrated through nonlinear regression. ML inputs included magnitude (M), epicentre distance (Repi), base-10 logarithm of depth, and base-10 logarithm of Vs30. Exploratory Data Analysis (EDA) identified the natural logarithm of PGA as the target variable due to its strongest correlation with the input parameters.

Model performance was evaluated using several metrics: residual standard deviation, Mean Square Error (MSE), coefficient of determination (R^2), and residual distribution analysis. The results consistently showed the superiority of the ML model over Zhao et al. [23] GMPE, with XGBoost achieving optimal accuracy. Residual analysis confirmed these findings, showing random dispersion of XGBoost around zero without systematic bias. Kernel Density Estimation (KDE) further confirmed the near-normal distribution of residuals with a mean close to 0.

The selected XGBoost model underwent additional validation using four earthquakes from August–September 2023 that were excluded from the training process. The predictions showed strong agreement with the observed data, confirming the model's ability to estimate ground acceleration well. Feature importance analysis within the current model framework indicates magnitude and epicentral distance as the dominant predictors for ground motion. However, the apparent low influence of Vs30 is primarily attributed to the reliance on a topographic slope proxy rather than direct borehole or MASW measurements. It is highly expected that incorporating high-quality, direct Vs30 field data in future regional studies would significantly elevate its predictive importance. Therefore, the current low feature importance of site conditions must be interpreted strictly as a data limitation rather than a physical reality. With its good performance and validated results, the XGBoost-based GMPM emerges as a reliable PGA prediction tool for Indonesia. This model has the potential to improve the accuracy of ground motion estimates, strengthen disaster mitigation strategies, and reduce the risk of casualties in future major earthquakes [39], [40].

Data availability statement

The raw strong-motion data utilized in this study are subject to BMKG institutional data protection policies. Consequently, these datasets cannot be made publicly available, but access may be granted upon reasonable request to the corresponding author for specific research purposes. Additionally, the custom machine learning scripts developed for this analysis are part of an ongoing internal research framework and are not currently open-source. However, detailed descriptions of the algorithms, hyperparameter configurations, and preprocessing steps have been thoroughly documented within the methodology section to facilitate reproducibility.

Acknowledgement

We would like to thank BMKG and STMKG for providing the data and facilities used in this study.

References

[1] A. C. Windiyanti, karyanto, Rustadi, and Rudianto, "Analisis Zona Rawan Gempabumi Daerah Lampung Berdasarkan Nilai Percepatan Tanah Maksimum (Pga) Dan Data Accelererograph Tahun 2008-2017," *JGE (Jurnal Geofisika Eksplorasi)* , vol. 3, no. 2, pp. 17–27, 2019.

[2] S. H. Pandadaran, G. F. A. Muabuay, S. E. Kurniawan, and M. Fadhilah, "Ground motion prediction equation for West Sumatra due to distant shallow crustal, interface, andiIntraslab warthquakes," in *E3S Web of Conferences*, 2020. doi: 10.1051/e3sconf/202015603001.

[3] R. I. Hielmy, "Perbandingan Metode Davenport, Milne, dan McGuirre dalam Penentuan Nilai Percepatan Tanah Maksimum di Pulau Bali," *Buletin Meteorologi, Klimatologi, dan Geofisika*, vol. 5, no. 5, pp. 9–16, 2024.

[4] R. I. Hielmy and M. L. Rajagukguk, "Analisis probabilistik bahaya seismik di denpasar dan sekitarnya berdasarkan pendekatan psha," *Jurnal Online Physics*, vol. 11, no. 1, pp. 21–28, 2025.

[5] M. Ryo and M. C. Rillig, "Statistically reinforced machine learning for nonlinear patterns and variable interactions," *Ecosphere*, vol. 8, no. 11, pp. 1–16, 2017, doi: 10.1002/ecs2.1976.

[6] A. Mohammadi, S. Karimzadeh, S. A. Banimahd, V. Ozsarac, and P. B. Lourenço, "The potential of region-specific machine-learning-based ground motion models: Application to Turkey," *Soil Dynamics and Earthquake Engineering,* vol. 172, p. 108008, 2023, doi: 10.1016/j.soildyn.2023.108008.

[7] H. Kubo, T. Kunugi, W. Suzuki, S. Suzuki, and S. Aoi, "Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-68630-x.

[8] J. Hu, H. Zhang, and J. Hu, "Support Vector Machine Method for Developing Ground Motion Models for Earthquakes in Western Part of China Support Vector Machine Method for Developing Ground Motion," *Journal of Earthquake Engineering*, vol. 26, no. 11, pp. 5679–5694, 2022, doi: 10.1080/13632469.2021.1884146.

[9] K. H. Palgunadi *et al.*, "High resolution seismic waveform generation using denoising diffusion," *Journal of Geophysical Research: Machine Learning and Computation*, vol. 2, p. e2025JH000862, 2025, doi: 10.1029/2025JH000862.

[10] R. I. Hielmy, B. Pranata, Wijayanto, and Daryono, "Performance Evaluation of Automated and Manual Seismic Phase Picking for Rapid Earthquake Parameter," *Jurnal Meteorologi dan Geofisika*, vol. 26, no. 2, pp. 89–107, 2025.

[11] D. C. Heath, D. J. Wald, C. B. Worden, E. M. Thompson, and G. M. Smoczyk, "A global hybrid VS30 map with a topographic slope–based default and regional map insets," *Earthquake Spectra*, vol. 36, no. 3, pp. 1570–1584, 2020.

[12] T. I. Allen and D. J. Wald, "On the Use of High-Resolution Topographic Data as a Proxy for Seismic Site Conditions (VS30)," *Bulletin of the Seismological Society of America*, vol. 99, no. 2A, pp. 935–943, 2009.

[13] D. J. Wald and T. I. Allen, "Topographic slope as a proxy for seismic site conditions and amplification," *Bulletin of the*

Seismological Society of America, vol. 97, no. 5, pp. 1379–1395, 2007, doi: 10.1785/0120060267.

[14] J. T. Behrens, "Principles and Procedures of Exploratory Data Analysis," *Psychological Methods*, vol. 2, no. 2, pp. 131–160, 1997, doi: 10.1037/1082-989X.2.2.131.

[15] H. Seo, J. Kim, and B. Kim, "Machine-learning-based surface ground-motion prediction models for South Korea with low-to-moderate seismicity," *Bulletin of the Seismological Society of America*, vol. 112, no. 3, pp. 1549–1564, 2022.

[16] L. Breimann, "Random Forest," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," in *International Conference on Web Information Systems Engineering*, Cham: Springer International Publishing, 2020, pp. 503–515. doi: 10.1007/978-3-030-62008-0_35.

[18] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[19] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive text mining and database fusion approach," *Environmental Health Perspectives*, vol. 127, no. 9, pp. 2825–2830, 2019, doi: 10.1289/EHP4713.

[20] M. Y. Khan, A. Qayoom, M. S. Nizami, M. S. Siddiqui, S. Wasi, and S. M. K. U. R. Raazi, "Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques," *Complexity*, vol. 2021, no. 1, pp. 1–18, 2021, doi: 10.1155/2021/2553199.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[22] A. S. Bakouregui, H. M. Mohamed, A. Yahia, and B. Benmokrane, "Explainable extreme gradient boosting tree-based prediction of load-carrying capacity of FRP-RC column," *Engineering Structures*, vol. 245, p. 112836, 2021.

[23] J. X. Zhao *et al.*, "Attenuation relations of strong ground motion in Japan using site classification based on predominant period," *Bulletin of the Seismological Society of America*, vol. 96, no. 3, pp. 898–913, 2006, doi: 10.1785/0120050122.

[24] R. Dobry *et al.*, "New site coefficient and site classification system used in recent building code provisions.," *Earthquake Spectra*, vol. 16, no. 1, pp. 41–67, 2000.

[25] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[26] A. Rachmadan, A. Koeshidayatullah, and S. L. I. Kaka, "Developing ground motion prediction models for West Java: A machine learning approach to support Indonesia's earthquake early warning system," *Applied Computing and Geosciences*, vol. 25, p. 100212, 2025, doi: 10.1016/j.acags.2024.100212.

[27] A. L. Ashadi, U. Harmoko, G. Yuliyanto, and S. I. Kaka, "Probabilistic seismic-hazard analysis for central java province, indonesia," *Bulletin of the Seismological Society of America*, vol. 105, no. 3, pp. 1711–1720, 2015, doi: 10.1785/0120140277.

[28] R. Damanik *et al.*, "New assessment of the probabilistic seismic hazard analysis for the greater Jakarta area, Indonesia," *Geomatics, Natural Hazards and Risk*, vol. 14, no. 1, p., 2023, doi: 10.1080/19475705.2023.2202805.

[29] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures," *BMC Bioinformatics*, vol. 8, p. 25, 2007.

[30] F. Khosravikia and P. Clayton, "Machine learning in ground motion prediction," *Computers & Geosciences*, vol. 148, no. January, p. 104700, 2021, doi: 10.1016/j.cageo.2021.104700.

[31] H. Seo, J. Kim, and B. Kim, "Machine-learning-based surface ground-motion prediction models for South Korea with low-to-moderate seismicity," *Bulletin of the Seismological Society of America*, vol. 112, no. 3, pp. 1549–1564, 2022, doi: 10.1785/0120210244.

[32] H. Dang *et al.*, “Ground motion prediction model for shallow crustal earthquakes in Japan based on XGBoost with Bayesian optimization,” *Soil Dynamics and Earthquake Engineering*, vol. 177, p. 108391, 2024.

[33] X. Wang, Z. Wang, J. Wang, P. Miao, H. Dang, and Z. Li, “Machine learning based ground motion site amplification prediction,” *Frontiers in Earth Science*, vol. 11, pp. 1–14, 2023, doi: 10.3389/feart.2023.1053085.

[34] K. B. Withers, M. P. Moschetti, and E. M. Thompson, “A machine learning approach to developing ground motion models from simulated ground motions,” *Geophysical Research Letters*, vol. 47, no. 6, 2020.

[35] D. T. Trugman and P. M. Shearer, “Strong correlation between stress drop and peak ground acceleration for recent M 1--4 earthquakes in the San Francisco Bay area,” *Bulletin of the Seismological Society of America*, vol. 108, no. 2, pp. 929–945, 2018.

[36] L. Al Atik, N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn, “The variability of ground motion prediction equations and its components,” *Seismological Research Letters*, vol. 81, no. 5, pp. 794–801, 2010.

[37] E. M. Thompson, D. J. Wald, and C. B. Worden, “A VS30 Map for California with Geologic and Topographic Constraints,” *Bulletin of the Seismological Society of America*, vol. 104, no. 5, pp. 2313–2321, 2014.

[38] Y. Zhang, J. F. Fung, K. J. Johnson, and S. Sattar, “Review of Seismic Risk Mitigation Policies in Earthquake-Prone Countries: Lessons for Earthquake Resilience in the United States,” *Journal of Earthquake Engineering*, vol. 26, no. 12, pp. 6208–6235, 2022, doi: 10.1080/13632469.2021.1911889.

[39] R. I. Hielmy *et al.*, “Sosialisasi Mitigasi Bencana Gempa Bumi , Tsunami , dan Hidrometeorologi di SMAN 6,” *To Maega | Jurnal Pengabdian Masyarakat*, vol. 8, no. 3, pp. 607–616, 2025.

[40] R. I. Hielmy *et al.*, “Evaluasi Efektivitas Simulasi Mitigasi Bencana Gempa Bumi,” *To Maega | Jurnal Pengabdian Masyarakat*, vol. 9, no. 1, pp. 1–12, 2026.