

## XGBoost for Predicting Peak Ground Acceleration Based on the NGA West2 Ground Motion Database

Asadullah Ziar<sup>1</sup> , Wahidullah Hazim<sup>2</sup> 

<sup>1</sup> Ghazni Technical University, Department of Civil Engineering, Faculty of Geology and Mine Engineering, Ghazni, Afghanistan

<sup>2</sup> Kabul University, Department of Civil Engineering, Faculty of Engineering, Kabul, Afghanistan

 E-mail: [asad.ziar2020@gmail.com](mailto:asad.ziar2020@gmail.com) (corresponding author)

---

### ABSTRACT

Earthquakes remain among the most destructive natural hazards worldwide, causing significant loss of life, economic damage, and infrastructure failures each year. Accurate prediction of peak ground acceleration (PGA), a key indicator of ground motion intensity, is fundamental to reliable seismic hazard assessment and earthquake-resistant structural design. Inaccurate or oversimplified PGA estimates can lead to unsafe structural designs or overly conservative construction costs, directly affecting public safety, infrastructure resilience, and economic sustainability. However, conventional ground motion prediction equations are typically based on predefined functional forms that may not fully capture the complex, nonlinear relationships among seismic source, path, and site parameters, leading to inaccuracies in PGA estimation. To address this limitation, this study develops an extreme gradient boosting (XGBoost) regression model to predict PGA using the NGA-West2 ground motion dataset from the Pacific Earthquake Engineering Research Center (PEER). The dataset, filtered to 16,211 records, includes moment magnitude ( $M_w$ ), rake angle (RA), hypocentral depth, closest rupture distance (ClstD), and average shear-wave velocity ( $V_{s30}$ ) as input features. At the same time, the logarithm of PGA serves as the target variable. The model achieved high accuracy on both training ( $R^2 = 0.971$ , RMSE = 0.184) and testing ( $R^2 = 0.936$ , RMSE = 0.269) datasets, demonstrating excellent learning and generalization performance. Feature importance analysis identified  $M_w$  and ClstD as the dominant predictors of PGA, accounting for approximately 64% and 29% of the total contribution, respectively. The results confirm that the XGBoost model effectively models nonlinear dependencies and complex interactions among seismic parameters.

---

### ARTICLE INFO

#### Article history:

Received: October 6, 2025

Revised: March 8, 2026

Accepted: May 18, 2026

Published: June 30, 2026

#### Keywords:

Ground motion prediction equations; Machine learning; Peak ground acceleration; Seismic hazard analysis, XGBoost

---

**To cite this article:** Ziar, A., Hazim, w. (2026). XGBoost for Predicting Peak Ground Acceleration Based on the NGA West2 Ground Motion Database. *Journal of Natural Science Review*, 4 (2), 407-420.

<https://doi.org/10.62810/jnsr.v4i2.332>

**Link to this article:** <https://kujnsr.com/JNSR/article/view/332>



Copyright © 2026 Author(s). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## INTRODUCTION

Earthquakes release elastic energy that propagates through the Earth as seismic waves, which are generally classified into body waves and surface waves. Body waves include

primary P waves, which are compressional waves that propagate through both solids and fluids, and secondary S waves, which are shear waves that propagate only through solids. Surface waves, including Love and Rayleigh waves, travel along the Earth's surface and often contribute significantly to ground shaking intensity. When these waves reach the surface, they produce ground shaking that can cause loss of life, damage to infrastructure, and long-term economic disruption. The intensity and duration of shaking at a given site depend primarily on the earthquake source (magnitude, rupture geometry, and focal mechanism), the propagation path (distance and anelastic attenuation), and local site conditions (near-surface velocity structure, layering, and amplification). Peak Ground Acceleration (PGA), the maximum instantaneous acceleration of the ground during an earthquake, is one of the most commonly used scalar measures of shaking intensity for seismic-hazard assessment, code development, and rapid post-event screening (Douglas, 2022; Javan-Emrooz et al., 2018; Kramer, 1996).

To estimate seismic ground motion intensity at a site, empirical ground motion prediction equations (GMPEs), commonly referred to as attenuation relationships, are widely used in earthquake engineering practice. These models provide predictive relationships between seismic intensity measures such as peak ground acceleration, peak ground velocity, and spectral ordinates, and a set of explanatory variables. The explanatory variables typically include moment magnitude, a distance metric such as closest distance to the rupture plane or hypocentral or epicentral distance, faulting mechanism including strike slip, normal, and reverse faulting, and site condition proxies such as  $V_{s30}$ , defined as the time averaged shear wave velocity in the upper 30 meters of the ground (Douglas, 2022; Javan-Emrooz et al., 2018).

Traditional GMPEs are typically derived via regression using prescribed functional forms (e.g., log-linear, bilinear, or piecewise) and distributional assumptions for the residuals. These choices allow interpretable magnitude and distance scaling and transparent uncertainty quantification. However, they also impose structural constraints that can underserve complex, nonlinear interactions among predictors, such as magnitude–site coupling, near-fault directivity (which describes the amplification of ground motion in the direction of rupture propagation), and frequency-dependent site amplification. As a consequence, empirical models can be biased or overly smooth when applied outside their calibration domain or in regions with strong local heterogeneity in soil stratigraphy, such as vertical variation in soil layers or soil types, leading to under- or overestimation of PGA in engineering applications (Kohrangi et al., 2021; Sreenath et al., 2023).

Machine learning (ML) methods offer flexible alternatives because they do not require pre-specified algebraic forms and can represent high-order nonlinear interactions among predictors. Over the last two decades, ML methods, including artificial neural networks, deep neural networks, support-vector regression, ensemble tree methods such as random forest and XGBoost, and hybrid architectures, have been successfully applied to problems ranging from signal processing and event detection to ground-motion modeling, transportation and

structural damage assessment (Derakhshani & Foruzan, 2019; Güllü & Erçelebi, 2007; Joshi et al., 2024; Khosravikia & Clayton, 2021; Ziar, 2025). Comparative and review studies indicate that ML approaches can outperform conventional regression models for PGA and other intensity measures when sufficiently large and representative datasets are available, though careful validation and attention to extrapolation limits remain essential (Bhatta & Dang, 2023; Hussaini et al., 2024; Yang et al., 2024).

Machine learning techniques have been increasingly applied to predict ground-motion parameters such as peak ground acceleration (PGA), peak ground velocity (PGV), and peak ground displacement (PGD). Early studies demonstrated the potential of artificial neural networks (ANNs) for seismic ground-motion prediction. For example, Güllü and Erçelebi (2007) developed an ANN model using Turkish strong-motion records and reported higher prediction accuracy and stronger correlations than conventional regression-based approaches. Similarly, Alavi and Gandomi (2011) employed genetic programming and hybrid evolutionary algorithms to derive attenuation relationships without predefined functional forms, demonstrating the capability of symbolic machine learning methods to model complex nonlinear relationships in seismic data. With the advancement of deep learning, Derakhshani and Foruzan (2019) developed deep neural network (DNN) models based on the NGA-West2 database and successfully predicted PGA, PGV, and PGD simultaneously, achieving correlation coefficients of approximately 0.90 and lower prediction errors compared with traditional ANN models. More recently, Joshi et al. (2024) introduced the SeisEML ensemble framework, which integrates multiple machine learning algorithms, including kernel-based and tree-based methods, and reported substantial improvements in MAE and RMSE compared with conventional regional ground-motion prediction equations (GMPEs). Furthermore, Khosravikia and Clayton (2021) provided a comprehensive review of machine learning applications in ground-motion prediction, highlighting the strengths and limitations of methods such as random forests, support vector machines, and deep neural networks, while emphasizing the importance of uncertainty quantification and appropriate data-partitioning strategies. Collectively, these studies demonstrate that machine learning approaches can effectively capture the complex nonlinear relationships governing seismic ground motions and often outperform traditional empirical GMPEs in prediction accuracy.

The model is trained using the PEER NGA-West2 database (Ancheta et al., 2013). Five physically meaningful predictors are selected as input variables: moment magnitude ( $M_w$ ), closest distance to the rupture plane ( $ClstD$ ), hypocentral depth ( $D$ ), rake angle ( $RA$ ), and the average shear-wave velocity in the upper 30 m ( $V_{s30}$ ). These parameters were chosen to maintain a balance between model simplicity, physical interpretability, and predictive capability. XGBoost is particularly suitable for this task due to its strong performance on tabular datasets, robustness to outliers, ability to handle heterogeneous predictors, and built-in mechanisms for feature importance evaluation and model regularization.

The specific objectives of this study are:

- To develop a robust XGBoost model for predicting PGA using the NGA-West2 dataset.
- To quantify the relative importance of input features on PGA predictions.

The remainder of this article is organized as follows. First, the materials and methods used in this study are presented, including the description of the NGA West2 database, selected input parameters, data preprocessing procedures, and the development of the XGBoost regression model. Next, the obtained results and model performance evaluations are presented and discussed in detail. Subsequently, the findings of the proposed model are compared with previously published machine learning based PGA prediction studies. Finally, the main conclusions of the study are summarized together with recommendations for future research.

## **METHODS AND MATERIALS**

This section describes the dataset, data-processing procedures, and machine learning methodology employed in this study for peak ground acceleration (PGA) prediction. First, the source of the ground-motion records and the criteria used for data selection are presented. Subsequently, the characteristics of the final dataset and the input variables used for model development are discussed. Finally, the implementation of the XGBoost regression model is described.

### **Study Data Source**

This study utilized ground motion records from the Updated NGA-WEST2 Flatfile Vertical 5% Damping database, developed by the Pacific Earthquake Engineering Research Center (PEER) (Ancheta et al., 2013). The NGA-West2 dataset is one of the most comprehensive and widely used ground-motion databases in earthquake engineering, containing high-quality strong-motion records from numerous earthquakes worldwide, along with detailed metadata describing source, path, and site characteristics.

### **Data Selection and Sample**

To ensure data consistency and reliability for model development, several filtering criteria were applied to the NGA-West2 dataset. The closest distance to the rupture plane (ClstD) was limited to 500 km, and the moment magnitude (Mw) ranged between 3.40 and 7.90. The hypocentral depths ranged from 1.8 km to 81 km. The faulting mechanism was represented using the rake angle parameter.

After applying these selection criteria, the final dataset consisted of 16,211 ground motion records, which were used for machine learning model development and evaluation.

### **Model Development and Feature Analysis**

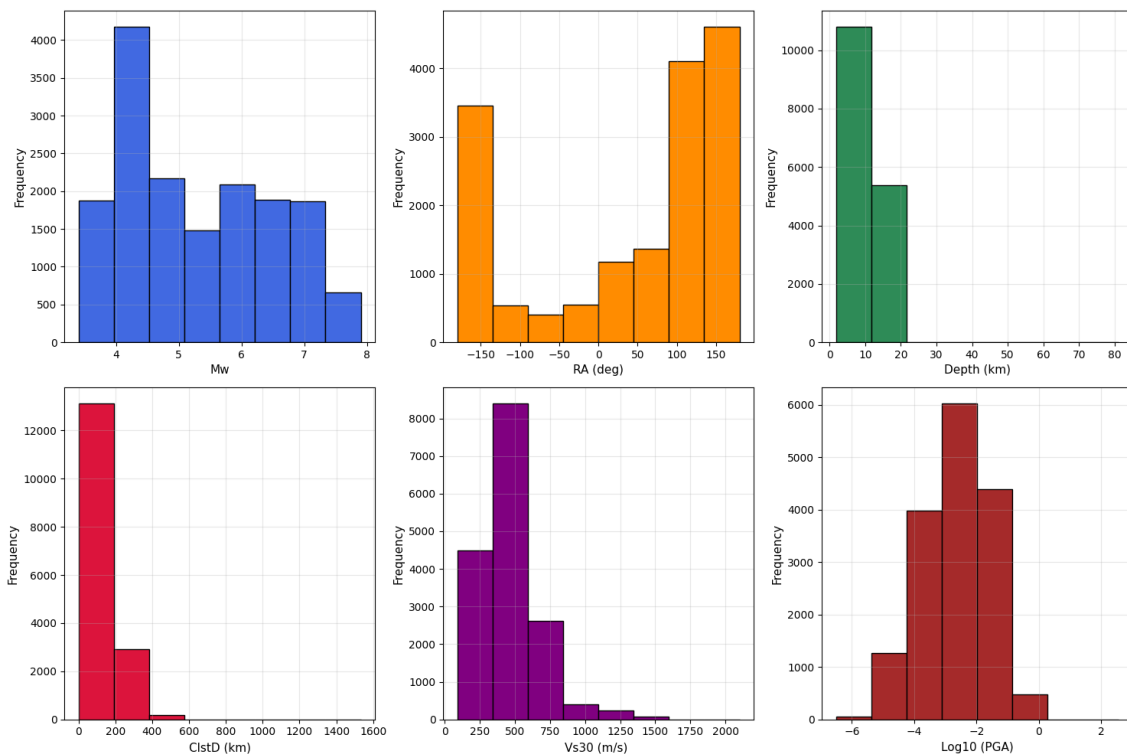
For model development, the XGBoost machine learning regression model was implemented using moment magnitude, rake angle, hypocentral depth, and average shear wave velocity in the upper 30 meters ( $V_{s30}$ , m/s) as input parameters. The target variable was the logarithm

of the peak ground acceleration ( $\log(\text{PGA})$ ). The statistical distributions of these variables are presented in Figure 1, and their descriptive statistics are summarized in Table 1.

Figure 2 presents the correlation heatmap among the input parameters and the logarithm of peak ground acceleration ( $\text{Log}_{10}(\text{PGA})$ ). Pearson's correlation values range between  $-1$  and  $+1$ , where positive coefficients indicate direct relationships and negative ones indicate inverse trends. The color scale highlights these dependencies, with warmer shades representing stronger positive associations and cooler shades showing negative ones.

$M_w$  displays the strongest positive correlation with  $\text{Log}_{10}(\text{PGA})$  ( $r = 0.66$ ), confirming that larger events generally generate stronger shaking.  $RA$  and event depth show weaker positive correlations ( $r = 0.15$  and  $r = 0.13$ ), suggesting a limited but non-negligible influence. In contrast,  $ClstD$  and  $V_{s30}$  show negative correlations with  $\text{PGA}$  ( $r = -0.55$  and  $-0.33$ ), consistent with attenuation of shaking with distance and stiffer site conditions.

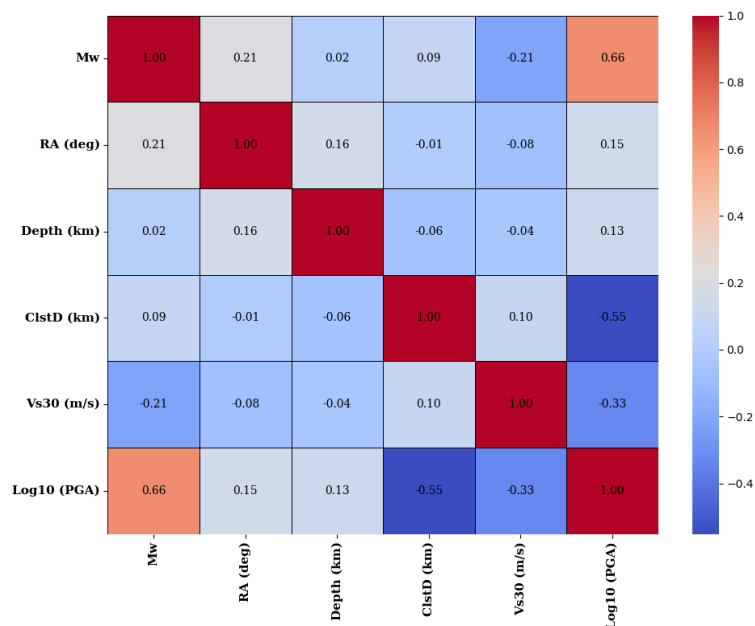
Although these correlations provide useful first-order insights, many predictor-response relationships appear weak or nonlinear (e.g.,  $RA$  and  $ClstD$ ,  $r \approx 0$ ). This highlights the limitations of linear correlation analysis alone. To capture the more complex interactions among seismic source, path, and site parameters, an ensemble learning approach such as XGBoost is more suitable. XGBoost can model nonlinear dependencies, automatically handle feature interactions, and assign relative importance to variables beyond what linear metrics suggest. In this context,  $M_w$  and  $ClstD$  are expected to emerge as the most influential features. In contrast, others, such as  $V_{s30}$  or depth, may contribute through interaction effects that correlation alone cannot capture.



**Figure 1.** Histograms showing the statistical distributions of the key parameters used in the study dataset, including moment magnitude ( $M_w$ ), rupture angle ( $RA$ ), focal depth ( $Depth$ ), closest distance ( $ClstD$ ), shear-wave velocity ( $V_{s30}$ ), and logarithmic peak ground acceleration ( $\text{Log}_{10}(\text{PGA})$ ).

**Table 1.** Summary of descriptive statistics for input and output variables

Parameters	Mw	RA (deg)	Depth (km)	ClstD (km)	Vs30 (m/s)	Log10 (PGA)
Minimum	3.40	-180.00	1.80	0.05	89.32	-6.51
Maximum	7.90	180.00	81.00	1532.66	2100.00	2.55
Range	4.50	360.00	79.20	1532.61	2010.68	9.06
Mean	5.31	35.27	10.53	117.74	460.19	-2.63
Median	5.03	90.00	10.00	97.93	405.00	-2.50
Standard deviation	1.19	126.54	3.77	87.94	204.48	1.07



**Figure 2.** Correlation heatmap of input parameters and Log<sub>10</sub> (PGA), showing positive and negative relationships. Mw has the strongest positive correlation, while ClstD and Vs30 show negative trends

### Data Analysis: Model Performance Metrics

As discussed earlier, the main PEER database was first filtered in Excel using the previously described criteria. Before training the model, any records with missing or incomplete data were removed. Outliers were identified by visually inspecting the distributions of all input and output variables and were excluded to minimize their impact on the dataset. Continuous variables were standardized using the StandardScaler() function from the Scikit-learn library to ensure consistent scaling across all parameters.

After model training, the predictive capability of the developed XGBoost model was evaluated using several statistical indicators: the Root Mean Square Error (RMSE) [Eq. (3)], the Mean Absolute Error (MAE) [Eq. (4)], and the Coefficient of Determination (R<sup>2</sup>) [Eq. (5)].

The RMSE is a widely used metric for evaluating the accuracy of regression models and is the square root of the average squared difference between predicted and observed values. Because it squares the errors before averaging, RMSE is particularly sensitive to large deviations, making it useful when large errors are especially undesirable. It has the advantage of being in the same units as the observed variable, providing an intuitive sense of the typical

magnitude of prediction errors (Chai & Draxler, 2014; Hodson, 2022). The MAE measures the average magnitude of errors in a set of predictions, without considering their direction. Unlike RMSE, it treats all deviations equally, making it more robust to outliers. MAE is straightforward to interpret and widely used in model evaluation to represent the typical error magnitude in the same units as the observed data (Willmott & Matsuura, 2005; Hodson, 2022).

The  $R^2$  is a statistical measure indicating how well the independent variables explain the variance in the dependent variable. It ranges from 0 to 1, where higher values suggest a better fit of the model to the observed data. While  $R^2$  is commonly used to evaluate regression models, it does not provide information about the scale of prediction errors and can sometimes be misleading in the presence of overfitting (Chai & Draxler, 2014; Li et al., 2023).

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (y_{Actual\_PGA} - y_{Predicted\_PGA})^2} \quad (3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{Actual\_PGA} - y_{Predicted\_PGA})^2 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{Actual\_PGA} - y_{predicted\_PGA})^2}{\sum_{i=1}^n (y_{Actual\_PGA} - \bar{y})^2} \quad (5)$$

In these formulas,  $n$  represents the number of samples, while  $\bar{y}$  denotes the mean of the actual values in the dataset.

The complete workflow of the study, from data selection and preprocessing to model training, feature analysis, and performance evaluation, is summarized in the methodological flow chart shown in Figure 3. This chart provides a visual overview of all steps involved in developing and validating the XGBoost model for predicting peak ground acceleration.

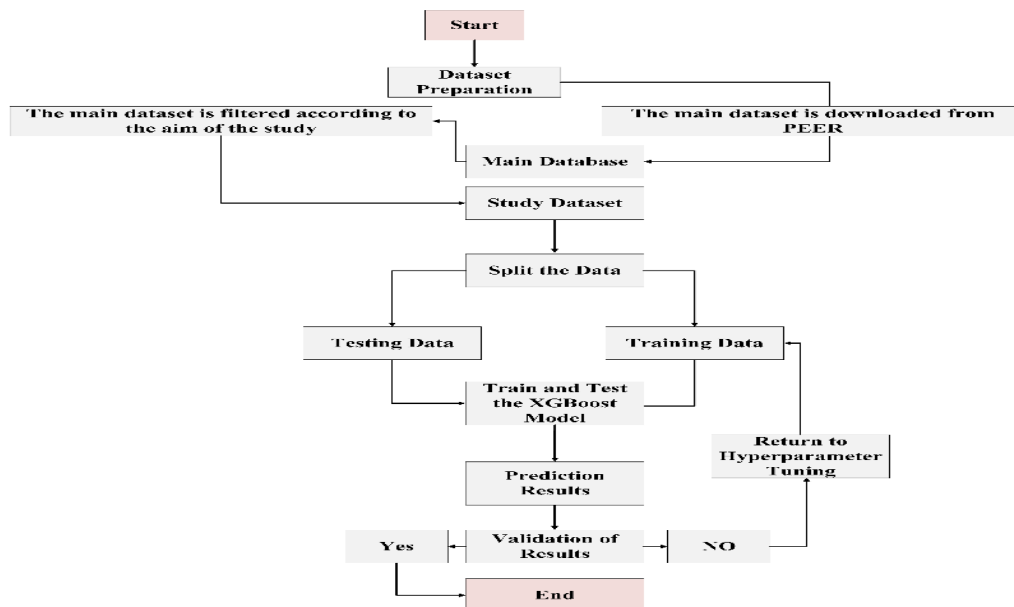


Figure 3. Flowchart of research methodology

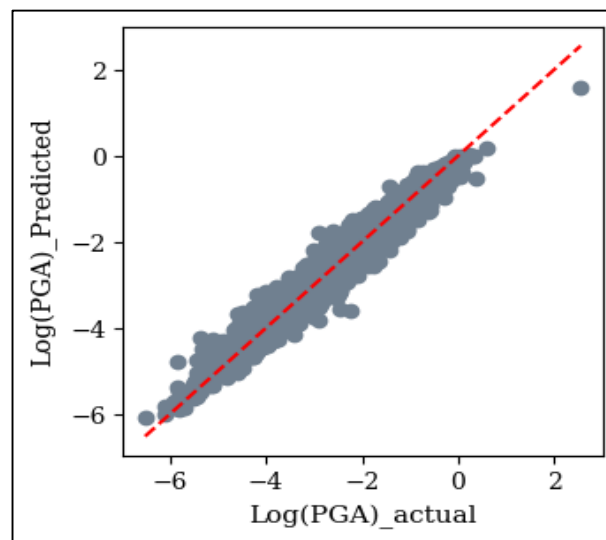
## FINDINGS

The extreme gradient boosting regression model was developed to predict peak ground acceleration using different input parameters described earlier in this study. Model performance was evaluated on both the training and test datasets, and the results are illustrated in Figures 4 and 5.

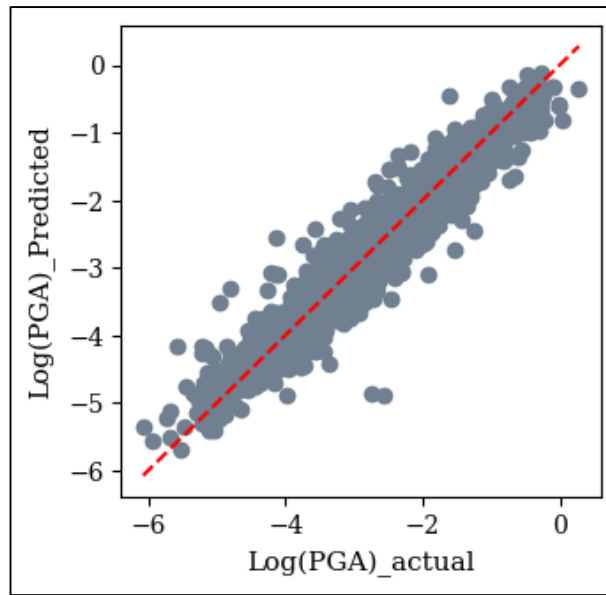
Figure 4 presents the regression performance for the training dataset. The model achieved a coefficient of determination ( $R^2$ ) of 0.971, a root mean square error (RMSE) of 0.184, and a mean square error (MSE) of 0.034. The correlation coefficient between the predicted and observed PGA values was 0.99, indicating a very strong agreement between the model predictions and the recorded data.

Figure 5 shows the regression results for the testing dataset. The model produced an  $R^2$  of 0.936, an RMSE of 0.269, an MSE of 0.072, and a correlation coefficient of 0.97. Although the predictive accuracy on the test dataset is slightly lower than on the training dataset, the results still demonstrate a strong relationship between predicted and observed PGA values. This indicates that the developed XGBoost model has good generalization capability when applied to unseen data.

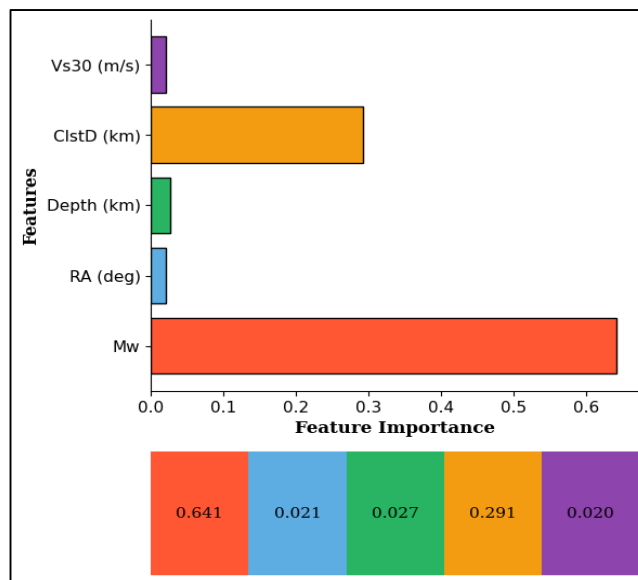
Feature importance analysis of the developed model is presented in Figure 6. Among the input parameters, the moment magnitude ( $M_w$ ) and the closest distance to the fault rupture (ClstD) were identified as the most influential variables.  $M_w$  contributed approximately 64.1% to the model predictions, while ClstD contributed about 29.1%. The remaining parameters, including RA, hypocentral depth, and  $V_{s30}$ , contributed relatively less to the prediction performance, with contributions of 2.7%, 2.0%, and 2.1%, respectively. These results highlight the dominant role of earthquake magnitude and source-to-site distance in determining the PGA values.



**Figure 2.** Regression performance of the XGBoost model on the training dataset, showing excellent agreement between predicted and observed PGA values ( $R^2 = 0.971$ ).



**Figure 3.** Regression performance of the XGBoost model on the testing dataset, confirming strong predictive accuracy and generalization ( $R^2 = 0.936$ ).



**Figure 4.** Feature importance of the XGBoost model, highlighting Mw and ClstD as the most influential parameters in predicting PGA.

## DISCUSSION

The results of this study demonstrate that the XGBoost regression model is highly effective at predicting peak ground acceleration from key input parameters. The high  $R^2$  values obtained for both the training (0.971) and testing datasets (0.936) indicate that the model successfully captures the complex nonlinear relationships between earthquake characteristics and ground motion intensity.

The relatively small difference between the training and testing performance metrics suggests that the model does not suffer from significant overfitting and maintains good predictive capability when applied to new data. This confirms that machine learning

techniques, such as XGBoost, can reliably predict seismic hazard parameters when trained on high-quality ground-motion datasets.

The feature importance results further reveal that moment magnitude and distance to the fault rupture are the most influential predictors of PGA. This finding is consistent with the physical understanding of earthquake ground motion generation: larger earthquakes release more energy and produce stronger ground shaking, while increasing distance from the seismic source generally reduces ground motion intensity.

The primary objectives of the study were comprehensively achieved. As discussed earlier, the high predictive accuracy and strong correlation between predicted and observed PGA values confirm that the model can effectively estimate ground motion intensity using the selected input parameters. Therefore, the study provides evidence that machine learning approaches can serve as useful tools for seismic hazard assessment and ground motion prediction.

To further evaluate the reliability and effectiveness of the proposed model, the obtained results were compared with previously published machine learning studies related to PGA prediction. Recent studies have shown that machine learning algorithms such as Random Forest (RF), Artificial Neural Networks (ANN), Deep Neural Networks (DNN), and Gradient Boosting techniques are capable of successfully modeling the nonlinear characteristics of seismic ground motion data. A comparative summary of these studies together with the proposed model is presented in Table 2.

As shown in Table 2, the XGBoost model developed in this study achieved a testing  $R^2$  value of 0.936 using the NGA-West2 dataset. This performance is comparable to or slightly higher than several previously reported machine learning models for PGA prediction. For instance, the ANN model developed by (Güllü & Erçelebi, 2007) achieved an R value of 0.964 using the Turkish Strong Motion dataset, while the DNN model proposed by (Derakhshani & Foruzan, 2019) obtained an  $R^2$  value of 0.902 using the NGA-West2 dataset. Similarly, the XGBoost model developed by (Mandal & Mandal, 2024) reported an  $R^2$  value of 0.844 using the CSIR-NGRI Hyderabad dataset.

In addition, Ziar and Basari, (2025) applied the Random Forest machine learning algorithm using the same NGA-West2 dataset and obtained an  $R^2$  value of 0.928. Compared with that study, the XGBoost model proposed in this research achieved slightly higher predictive performance, with an  $R^2$  value of 0.936. This improvement indicates that the XGBoost algorithm is highly effective in learning complex nonlinear patterns in seismic datasets.

In conclusion, the comparative evaluation demonstrates that the proposed XGBoost model provides competitive, robust, and reliable predictive capability for peak ground acceleration estimation and can serve as a practical tool for future seismic hazard and ground motion prediction studies.

**Table 2.** Performance comparison of the proposed XGBoost model with previously published ANN, DNN, RF, and XGBoost models for predicting PGA

Study	Employed Model	Data Source	R or R <sub>2</sub> (Test)
(Güllü & Erçelebi, 2007)	Artificial neural network	Turkish Strong Motion	0.964
(Derakhshani & Foruzan, 2019)	Deep neural network	NGA-West2	0.902
(Mandal & Mandal, 2024)	XGBoost	CSIR-NGRI, Hyderabad	0.844
(Ziar & Basari, 2025)	Random Forest	NGA-West2	0.928
<b>This Study</b>	<b>XGBoost</b>	<b>NGA-West2</b>	<b>0.936</b>

## CONCLUSION

The primary aim of this study was to predict peak ground acceleration of ground motions using a large global ground motion database provided by the Pacific Earthquake Engineering Research Center (PEER) and filtered within the scope of the study. A highly recognized and powerful regression model, namely Extreme Gradient Boosting (XGBoost), was employed for this purpose. In the developed model, PGA was defined as the output variable. At the same time, closest distance to the rupture plane (ClstD), shear wave velocity ( $V_{s30}$ ), moment magnitude (Mw), rake angle (RA), and hypocentral depth were selected as input variables.

The developed XGBoost regression model demonstrated strong performance in predicting PGA values for ground motion records. The model achieved high accuracy on both the training ( $R^2 = 0.971$ ,  $MSE = 0.034$ ) and test ( $R^2 = 0.936$ ,  $MSE = 0.072$ ) datasets, indicating excellent learning and generalization. Feature importance analysis revealed that Mw and ClstD are the dominant factors influencing PGA, while RA,  $V_{s30}$ , and Depth have smaller effects.

Compared to conventional regression-based ground motion prediction equations (GMPEs), the proposed approach offers greater flexibility in modeling nonlinear dependencies without imposing predefined functional forms. This highlights the potential of machine learning techniques as complementary tools to traditional empirical attenuation relationships, particularly when large and high-quality datasets are available.

From an engineering perspective, the developed model may serve as a reliable tool for preliminary seismic hazard assessment, rapid ground motion estimation, and data-driven regional model development. However, the applicability of the model is limited to the parameter ranges represented in the training dataset, and extrapolation beyond these bounds should be avoided.

Despite promising results, this study has some limitations. The model was trained on the NGA-West2 vertical 5 percent damping dataset, filtered to include only records with distance to the rupture plane up to 500 km, magnitudes from 3.40 to 7.90, and depths between 1.8 km and 81.0 km. This limits its applicability to earthquakes outside these ranges.

Future research could explore the integration of larger and more diverse datasets, additional seismic and geological parameters, and hybrid machine learning approaches. Comparative studies between different machine learning algorithms could also help identify the most effective techniques for ground motion prediction.

### **Authors Contributions**

Asadullah Ziar: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing – Original Draft.

Wahidullah Hazim: Conceptualization, Supervision, Writing – Review & Editing.

All authors reviewed and approved the final version.

### **Funding Information**

No funding is available for the manuscript.

### **Conflict of Interest Statement**

The authors declare that they have no conflicts of interest related to this work.

### **Data Availability Statement**

Data are available on request from the corresponding author.

### **REFERENCES**

- Alavi, A. H., & Gandomi, A. H. (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *Computers and Structures*, 89(23–24), 2176–2194. <https://doi.org/10.1016/j.compstruc.2011.08.019>
- Bhatta, S., & Dang, J. (2023). Seismic damage prediction of RC buildings using machine learning. *Earthquake Engineering & Structural Dynamics*, 52(11), 3504–3527. <https://doi.org/10.1002/eqe.3907>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Derakhshani, A., & Foruzan, A. H. (2019). Predicting the principal strong ground motion parameters: A deep learning approach. *Applied Soft Computing*, 80, 192–201. <https://doi.org/10.1016/j.asoc.2019.03.029>
- Douglas, J. (2022). Ground motion prediction equations 1964–2021. [Link](#)
- Güllü, H., & Erçelebi, E. (2007). A neural network approach for attenuation relationships: An application using strong ground motion data from Turkey. *Engineering Geology*, 93(3–4), 65–81. <https://doi.org/10.1016/j.enggeo.2007.05.004>

- Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(12), 5481–5490. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hussaini, S. M. S., Caicedo, D., Mohammadi, A., Karimzadeh, S., & Lourenço, P. B. (2024). Nonparametric ground motion models of Arias intensity and significant duration for the Italian dataset. *Journal of Physics: Conference Series*, 2647(6). <https://doi.org/10.1088/1742-6596/2647/6/062001>
- Javan-Emrooz, H., Eskandari-Ghadi, M., & Mirzaei, N. (2018). Prediction equations for horizontal and vertical PGA, PGV, and PGD in northern Iran using prefix gene expression programming. *Bulletin of the Seismological Society of America*, 108(4), 2305–2332. <https://doi.org/10.1785/0120170155>
- Joshi, A., Raman, B., Mohan, C. K., & Cenkeramaddi, L. R. (2024). Application of a new machine learning model to improve earthquake ground motion predictions. *Natural Hazards*, 120(1), 729–753. <https://doi.org/10.1007/s11069-023-06230-4>
- Khosravikia, F., & Clayton, P. (2021). Machine learning in ground motion prediction. *Computers and Geosciences*, 148, 104700. <https://doi.org/10.1016/j.cageo.2021.104700>
- Kohrangji, M., Kotha, S. R., & Bazzurro, P. (2021). Impact of partially non-ergodic site-specific probabilistic seismic hazard on risk assessment of single buildings. *Earthquake Spectra*, 37(1), 409–427. <https://doi.org/10.1177/8755293020938813>
- Kramer, S. L. (1996). *Geotechnical earthquake engineering*. Prentice-Hall.
- Li, X., Chen, Y., & Zhang, L. (2023). Comparative evaluation of regression metrics in predictive modeling for environmental data. *Environmental Modelling & Software*, 164, 105678. <https://doi.org/10.1016/j.envsoft.2023.105678>
- Mandal, P., & Mandal, P. (2024). Peak ground acceleration prediction using supervised machine learning algorithm for the seismically hazardous Kachchh rift zone, Gujarat, India. *Natural Hazards*, 120(2), 1821–1840. <https://doi.org/10.1007/s11069-023-06257-7>
- Sreenath, V., Podili, B., & Raghukanth, S. T. G. (2023). A hybrid non-parametric ground motion model for shallow crustal earthquakes in Europe. *Earthquake Engineering and Structural Dynamics*, 52(8), 2303–2322. <https://doi.org/10.1002/eqe.3845>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82. <https://doi.org/10.3354/cro30079>
- Yang, C., Pan, Y., Zhang, K., Yue, M., Wen, H., & Wang, F. (2024). Machine learning-based ground peak acceleration attenuation prediction model. *Journal of Earthquake Engineering*. <https://doi.org/10.1080/13632469.2024.2443638>

Ziar, A. (2025). Machine learning approaches for predicting the structural number of flexible pavements based on subgrade soil properties. *Scientific Reports*, 15(1).  
<https://doi.org/10.1038/s41598-025-13852-0>

Ziar, A., & Basari, E. (2025). Accurate prediction of peak ground acceleration using random forests: A data-driven approach with PEER updated NGA-West2 ground motion records. *Archives of Advanced Engineering Science*, 1–9.  
<https://doi.org/10.47852/bonviewaaes52026930>