



LightGBM-Driven Earthquake Magnitude Prediction: A Comparative Machine Learning Framework Using Global Seismic Data

Nima Khodadadi^{1,*}

¹ Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA

Email: nima.khodadadi@miami.edu

Received: January 14, 2025 Revised: March 18, 2025 Accepted: May 10, 2025 ★ Corresponding author

ABSTRACT

Earthquakes are among the most destructive natural hazards, causing severe damage to communities, loss of life, and major economic consequences. This study evaluates machine-learning models for magnitude prediction using an open-access global earthquake dataset from 2023. Five regressors are compared: Light Gradient Boosting Machine (LightGBM), Support Vector Regression (SVR), k-Nearest Neighbors (KNN), Ridge Regression, and Extra Trees. Training uses stratified cross-validation and hyperparameter optimization. The evaluation covers Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), the coefficient of determination (R^2), Nash–Sutcliffe Efficiency (NSE), Willmott Index (WI), Pearson’s correlation coefficient (r), and Relative RMSE (RRMSE). LightGBM performs best in the reported comparison, attaining $MSE = 0.0474$ and $R^2 = 0.9241$. Its leaf-wise tree construction, scalability, and regularization support generalization to unseen data while maintaining computational efficiency. The results indicate that LightGBM can capture patterns in high-dimensional seismic data and provide a reproducible analytical framework for earthquake-prone regions. In combination with other advanced ML systems, such models may contribute to early-warning research, faster emergency response, and better-informed risk planning. The use of an open dataset and documented modeling steps also supports transparency and collaborative research.

Keywords: Earthquake prediction ▪ Seismic data analysis ▪ LightGBM ▪ Machine learning ▪ Magnitude prediction

1. INTRODUCTION

Earthquakes are among the most unpredictable natural hazards, threatening human life, infrastructure, and economic activity. Their sudden release of energy can cause fatalities, extensive structural damage, and substantial financial losses. Events such as the 2004 Indian Ocean tsunami, the 2011 Japan earthquake, and the 2023 Turkey–Syria earthquake demonstrate the need for improved disaster preparedness [1, 2]. Machine-learning methods are increasingly applied to earthquake monitoring and other seismological tasks,

supporting the development of improved seismic analysis and warning systems [3].

Reliable forecasts can help emergency responders act quickly, support warnings to vulnerable communities, and guide decisions about critical infrastructure such as transport systems and nuclear facilities. Governments may also use forecasts to plan sustainable development in seismic areas and to prioritize infrastructure investments and post-disaster logistics. The global reach of seismic activity and the continued growth of cities in tectonically active regions make these capabilities increasingly important.

Traditional physical and statistical seismological methods have advanced understanding of seismic behavior, but can struggle to represent its nonlinear spatial and temporal dynamics. Machine-learning (ML) methods can learn complex patterns from data without relying on predefined physical constraints and have therefore become increasingly prominent in seismic analysis [4, 5]. Their application nevertheless presents challenges: high-dimensional data can increase computational and memory demands and encourage overfitting, while redundant features can reduce model reliability [6, 7]. Model performance also depends on selecting suitable learning rates, tree depths, regularization terms, and kernel settings; automated methods such as randomized search and Bayesian optimization can help tune these parameters [8]. Generalization across datasets, noisy measurements, and limited ground-truth labels remains an active concern [9].

Earthquake-prediction datasets may combine geospatial coordinates, seismic-wave characteristics, distances to fault lines, and environmental measurements. As the number of variables grows, the dimensionality curse can increase computational-resource and memory requirements, and make models more susceptible to overfitting. Feature selection and dimensionality reduction are therefore important for retaining informative signals without sacrificing efficiency. Redundant variables require data validation and careful handling so that repeated information does not distort model fitting. Hyperparameter choice is similarly consequential: learning rates, tree depth, regularization strength, and kernel configuration can determine whether a model underfits or overfits. Automated search strategies help compare configurations over different training folds, but performance under noise and across geographic areas remains a central generalization challenge.

This study develops a data-driven system for predicting earthquake magnitude from an open global dataset and compares model accuracy, computational efficiency, and geographic and temporal generalization. Light Gradient Boosting Machine (LightGBM) is selected for its ability to process tabular data efficiently using histogram-based, leaf-wise decision-tree learning, including data with missing and categorical values [10]. Randomized hyperparameter search with cross-validation is used to identify a model configuration while limiting overfitting [11]. The comparative evaluation uses standardized regression metrics and considers the influence of preprocessing, feature selection, and tuning on predictive performance [12, 13]. By documenting the use of public seismic data and the modeling procedure, the work aims to provide a reproducible basis for regional prediction studies, stress-anomaly analysis, and real-time early-warning research.

The main contributions are: (i) an optimized LightGBM model using randomized hyperparameter tuning; (ii) a systematic comparison of five ML algorithms for earthquake-magnitude prediction; (iii) an analysis of dimensionality, redundant features, and model generalization in seismic data; and (iv) a reproducible framework based on open-access data and transparent modeling practices to support seismic risk research and decision-making.

2. LITERATURE REVIEW

Earthquake prediction is an important disaster-prevention problem because seismic events occur abruptly and can have

severe effects on people and infrastructure. Researchers have applied Random Forest regression and neural networks to magnitude and depth prediction using historical seismic activity, location, and event-time features. Statistical models combined with geographic-information-system data have been used to detect patterns that conventional seismological analysis may not capture [14]. Comparative investigations have also examined alternative prediction methods. In the Horn of Africa, transformer-based magnitude prediction was compared with LSTM, BiLSTM, and attention-based recurrent models; the transformer reported lower MAE, MSE, RMSE, and MAPE and was designed to capture spatial and temporal relationships [15, 16].

Hybrid work combines geological domain knowledge with machine learning. A Random Forest connected to geological models has been studied for short-term prediction along the Pacific Ring of Fire, with the integrated approach reported to provide greater predictive accuracy and reliability than either component alone. In Southern California, two hybrid models combined the Flower Pollination Algorithm (FPA) with Extreme Learning Machines (ELM) and least-squares support vector machines (LS-SVM). Seven seismic indicators were used to forecast events up to 15 days in advance; FPA-LS-SVM showed favorable RMSE, MAE, SMAPE, and PMRE results [16, 17]. For early warning, the Ensemble Earthquake Early Warning System (E3WS) uses machine-learning algorithms to detect, locate, and estimate earthquake magnitude from as little as three seconds of P-wave data recorded at a single station. Its training data include records from Peru, Chile, Japan, and the STEAD global dataset [18].

Bayesian ML with wavelet analysis has been used to identify seismic cycles in five global zones over records spanning 1900–2021. The cited study reported periodicities from 3.7 to 55 years, discussed elevated earthquake potential in Japan for 2024–2026 and South America for 2026–2031, and used the method to examine the timing of the Parkfield earthquake [19]. In India, seven supervised approaches were evaluated for earthquake-severity classification across six datasets; Simple Logistic and Logistic Model Tree classifiers were reported to perform well [20]. GPS satellite total electron content has also been studied as a possible precursor. An LSTM model trained on measurements from preceding days reportedly achieved 0.82 accuracy, exceeding SVM, LDA, and Random Forest classifiers [21].

Other studies applied artificial neural networks and LSTM to Himalayan earthquake data, reporting useful detection for medium-scale events but difficulty identifying high-magnitude earthquakes [22]. In Iran, fault density calculated using kernel-density estimation and bivariate Moran's I was incorporated with spatial features; SVM, decision-tree, shallow-neural-network, and deep-neural-network models were evaluated, with spatial variables improving prediction of larger magnitudes [23]. For earthquake-induced landslides, high-resolution satellite imagery and digital elevation models have been combined with stacked autoencoders, with reported improvements over logistic regression and Random Forest [24]. Satellite infrared and hyperspectral data for 1,371 strong earthquakes have also been analyzed using Inverse Boosting Pruning Trees and compared with six alternative ML methods [25].

A systematic review of deep learning for earthquake disaster assessment analyzed 204 articles and classified research by assessment objects, data, models, and application stages, while discussing challenges and opportunities [26]. In Indonesia, four-second raw-seismogram data were used to train regression and classification deep neural networks for earthquake early warning. The classification network was reported to have a lower final loss than Random Forest for real-time magnitude estimation [27]. This body of work points to deep architectures, ionospheric and remote-sensing observations, spatial features, hybrid models, and ensembles as continuing directions for seismic risk research.

3. DATASET

3.1 Dataset Description

The study uses the Earthquakes 2023 Global dataset, an open-access collection of global seismic events recorded during 2023 [28]. It contains 26,642 events and 22 descriptive fields spanning time, location, seismological measurements, and measurement uncertainty. The fields include:

- **Time, Latitude, Longitude, Depth, and Mag:** event timestamp, epicenter coordinates, depth in kilometers, and the target magnitude.
- **MagType, Nst, Gap, and Dmin:** magnitude scale, number of recording stations, azimuthal station gap, and distance to the nearest station.
- **Rms, HorizontalError, DepthError, MagError, and MagNst:** signal root-mean-square, location/depth/magnitude errors, and the number of stations used for magnitude estimation.
- **Net, Id, Updated, Place, Type, Status, LocationSource, and MagSource:** event identifiers, contextual descriptions, review status, and data-source metadata.

The temporal, spatial, geophysical, and uncertainty fields support event clustering, map construction, and predictive modeling of magnitude and depth. Figure 1 shows the reported correlations among horizontal, depth, and magnitude errors and the main magnitude and depth features; horizontal error has a moderate association with event magnitude and depth. The error distributions also provide an assessment of measurement quality. Depth errors are positively skewed toward smaller values, most magnitude errors fall within a restricted range, and horizontal errors show a wider dispersion. The latter may reflect variations in seismic-station coverage or local terrain (Figures 2–4). Together, the visual summaries and descriptive fields provide a basis for examining dataset reliability and for statistical tests of magnitude- and depth-prediction performance.

3.2 Data Preprocessing

Preprocessing addressed missing values, outliers, categorical variables, feature scales, and timestamps [29, 30]. Missing values in MagError, DepthError, HorizontalError, and MagNst were imputed with the median because it is less sensitive to outliers; categorical fields were imputed with their mode or removed when too sparse to be useful. Interquartile-range analysis and boxplot inspection were used for outlier detection. Records with impossible values, including negative depths or magnitudes below 1.0, were excluded to retain

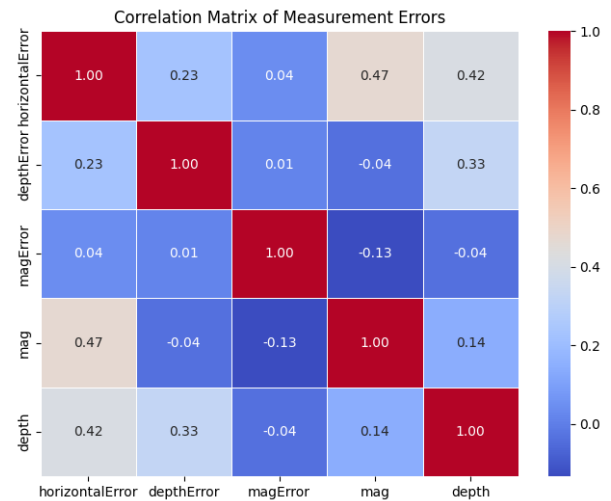


Figure 1. Correlation matrix of measurement errors.

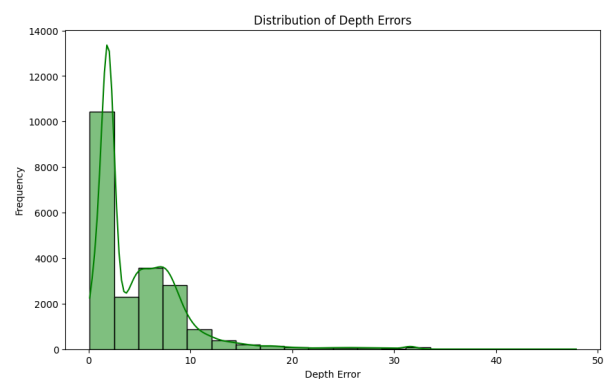


Figure 2. Distribution of depth errors.

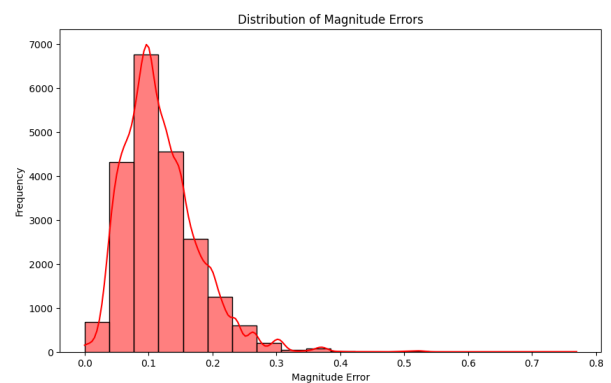


Figure 3. Distribution of magnitude errors.

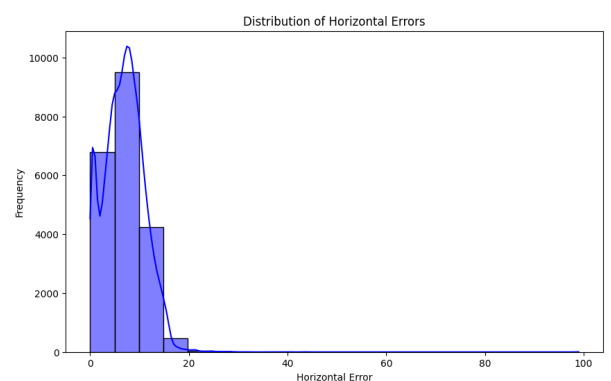


Figure 4. Distribution of horizontal errors.

seismically meaningful observations.

MagType, Type, and Status were one-hot encoded so that categorical information could be used by the learning algorithms. Id, Updated, and Net were removed as non-predictive identifiers. Depth, Mag, Gap, and Dmin were standardized for scale-sensitive models, and latitude and longitude were normalized to a uniform scale. The Time field was transformed into cyclic sine and cosine features to represent daily and seasonal patterns. Pearson correlation analysis identified a strong relationship between Rms and MagNst; both fields were removed to reduce multicollinearity. The data were split into training and test sets at an 80:20 ratio before preprocessing, and k-fold cross-validation was used during model development to assess generalization while limiting train-test bias.

4. RESULTS AND DISCUSSION

4.1 Machine Learning Models

Five regression models were compared because they provide

different approaches to linear and nonlinear relationships in structured, high-dimensional seismic data. LightGBM builds trees leaf-wise rather than level-wise, selecting the leaf that produces the greatest loss reduction. This can yield deeper trees and faster convergence. Parallel computation, histogram-based learning, and GPU support are intended to make the method efficient on large datasets; its regularization controls overfitting in data with irregular distributions. Light-GBM adds weak learners sequentially to reduce the loss; its prediction at iteration t is represented as:

$$X_t(f) = X_{t-1}(f) + \eta y_t(f), \quad (1)$$

where η is the learning rate and $y_t(f)$ is the decision tree added at iteration t . Randomized search and five-fold cross-validation were used to tune LightGBM. The final hyperparameters are listed in Table 1.

Table 1. Optimized hyperparameters for the LightGBM model.

Parameter	Value	Description
num_leaves	31	Controls tree complexity; more leaves allow more complex models.
learning_rate	0.1	Step size at each iteration while minimizing the loss function.
n_estimators	100	Number of boosting iterations (trees) used for fitting.
subsample	0.8	Fraction of observations sampled for each tree to help prevent overfitting.
colsample_bytree	0.8	Fraction of features used to construct each tree, improving generalization.
min_child_samples	20	Minimum samples required in a leaf node to control overfitting.
random_state	42	Random seed for reproducible results.

Support Vector Regression (SVR). SVR models nonlinear relationships by fitting a function within a tolerance ε :

$$q(\mathbf{x}) = \langle \mathbf{h}, \mathbf{x} \rangle + a. \quad (2)$$

Its optimization objective and constraints are:

$$\min_{\mathbf{h}, a, \xi, \xi^*} \frac{1}{2} \|\mathbf{h}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

$$\text{subject to } y_i - \langle \mathbf{h}, \mathbf{x}_i \rangle - a \leq \varepsilon + \xi_i, \quad (4)$$

$$\langle \mathbf{h}, \mathbf{x}_i \rangle + a - y_i \leq \varepsilon + \xi_i^*, \quad (5)$$

$$\xi_i, \xi_i^* \geq 0. \quad (6)$$

k-Nearest Neighbors (KNN). KNN predicts an output from the average of the k nearest observations:

$$\hat{E}(y) = \frac{1}{k} \sum_{i \in N_k(y)} E_i. \quad (7)$$

Ridge Regression. L2 regularization is added to the least-squares objective to reduce coefficient magnitudes and address multicollinearity:

$$\hat{\beta} = \arg \min_{\beta} \left[\sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2 + \lambda \|\beta\|^2 \right]. \quad (8)$$

Extra Trees Regressor. The ensemble averages predictions

from T randomized trees:

$$\hat{v}(i) = \frac{1}{T} \sum_{t=1}^T b_t(i). \quad (9)$$

The algorithms were evaluated on the same dataset split to enable direct comparison of predictive performance. Their different assumptions and ensemble structures provide a basis for assessing both regression accuracy and the potential use of these approaches in seismic monitoring.

4.2 Evaluation Metrics

Reliable assessment requires multiple measures because a single score cannot describe error size, bias, correlation, and agreement with observed values. MSE penalizes larger deviations more heavily; RMSE returns the error to the target's scale, while MAE gives an intuitive average absolute deviation. MBE indicates the direction of systematic error. Correlation and coefficient-of-determination measures characterize association and explained variance, while NSE compares residual variation with a mean-value baseline. WI measures agreement on a normalized scale, and RRMSE facilitates comparisons across different value ranges. Table 2 gives the metrics and formal definitions used in the study. Here, y_i is the observed value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ and $\hat{\bar{y}}$ are their respective means.

Table 2. Machine-learning prediction metrics.

Metric	Description and definition
Mean Squared Error (MSE)	Average squared difference between predicted and actual values; larger errors receive greater weight. $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	Square root of MSE, expressed in the target variable's units. $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
Mean Absolute Error (MAE)	Mean absolute difference between predicted and actual values. $MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
Mean Bias Error (MBE)	Mean signed difference; with the definition shown, positive values indicate underprediction. $MBE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$
Pearson Correlation (r)	Linear association between predicted and observed values. $r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}$
Coefficient of Determination (R^2)	Proportion of target variance explained by the model. $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Relative RMSE (RRMSE)	RMSE normalized by the range of observed values. $RRMSE = \frac{RMSE}{\max(y) - \min(y)}$
Nash–Sutcliffe Efficiency (NSE)	Predictive efficiency relative to the mean-observation baseline. $NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Willmott Index (WI)	Agreement between predicted and observed values; 1 indicates perfect agreement. $WI = 1 - \frac{\sum_{i=1}^n y_i - \hat{y}_i }{\sum_{i=1}^n (y_i - \bar{y} + \hat{y}_i - \bar{\hat{y}})}$

4.3 Results

The five models were evaluated using MSE, RMSE, MAE, MBE, Pearson correlation, R^2 , RRMSE, NSE, and WI. These nine measures provide complementary views of predictive error, correlation with observations, and the capacity to retain

patterns in the seismic data. Table 3 summarizes their reported values. LightGBM (Tuned) achieved the lowest errors and the strongest overall agreement with observed magnitudes.

Table 3. Evaluation of machine-learning models for earthquake-magnitude prediction.

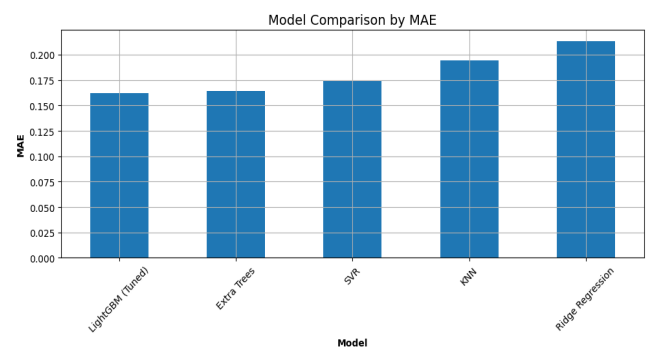
Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
LightGBM (Tuned)	0.0474	0.2176	0.1619	0.0018	0.9613	0.9241	0.0538	0.9241	0.9799
SVR	0.0562	0.2371	0.1743	-0.0160	0.9542	0.9100	0.0586	0.9100	0.9761
KNN	0.0687	0.2621	0.1944	-0.0021	0.9434	0.8900	0.0648	0.8900	0.9703
Ridge Regression	0.0799	0.2827	0.2136	0.0083	0.9339	0.8720	0.0699	0.8720	0.9647
Extra Trees	0.0490	0.2213	0.1644	0.0076	0.9600	0.9216	0.0547	0.9216	0.9791

LightGBM obtained an MSE of 0.0474, RMSE of 0.2176, and MAE of 0.1619. Its MBE was 0.0018, with Pearson correlation $r = 0.9613$, $R^2 = 0.9241$, NSE = 0.9241, RRMSE = 0.0538, and WI = 0.9799. Extra Trees performed comparably (MSE = 0.0490, $R^2 = 0.9216$, WI = 0.9791), while Ridge Regression and KNN had larger errors. The negative MBE reported for SVR indicates a positive signed difference under the definition in Table 2; the result is retained as reported.

The high correlation and R^2 scores show that LightGBM captured a large share of the variation in observed magnitudes. Its NSE and WI values also indicate close agreement with the observations, and the low RRMSE suggests that its errors remain small relative to the observed range. The leaf-wise construction and regularization of LightGBM may help explain its results on nonlinear, high-dimensional data. Extra Trees was the closest alternative on several measures. Ridge Regression and KNN showed larger RMSE and MAE, while SVR's MBE was negative as reported.

Figures 5–8 compare models by MAE, RMSE, R^2 , and NSE. LightGBM has the lowest MAE (0.1619), indicating the smallest average absolute difference between predictions and observations; Extra Trees is close, and Ridge Regression has the largest MAE. The RMSE comparison likewise places

LightGBM first at 0.2176, followed by Extra Trees and SVR, which is relevant because RMSE gives greater weight to larger residuals. For R^2 , LightGBM (0.9241) and Extra Trees (0.9216) lead the alternatives. NSE shows a similar pattern, with the ensemble models ahead of Ridge Regression and KNN. The WI values likewise favor LightGBM, although the source study did not include a WI plot. Taken together, the metrics indicate that the tuned LightGBM model provided the strongest results in this comparison and merits further assessment for seismic risk analysis and early-warning applications.

**Figure 5.** Model comparison by mean absolute error (MAE); LightGBM (Tuned) has the lowest average prediction error.

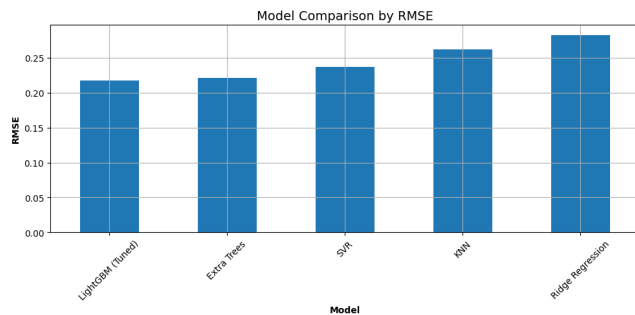


Figure 6. Model comparison by RMSE; LightGBM (Tuned) minimizes large prediction errors.

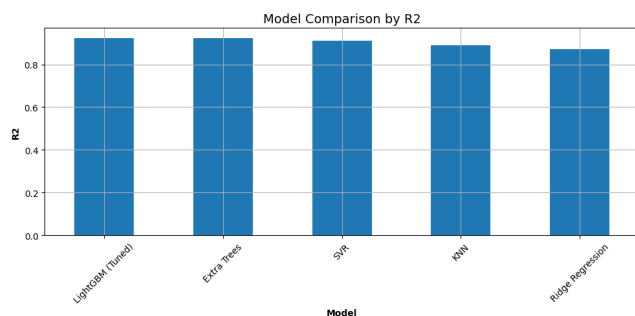


Figure 7. Model comparison by R^2 ; LightGBM (Tuned) and Extra Trees show high explanatory power.

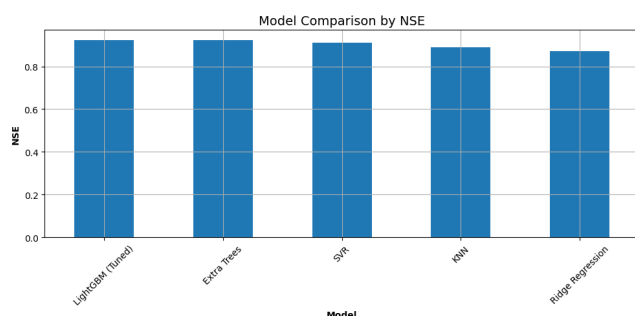


Figure 8. Model comparison by Nash–Sutcliffe efficiency (NSE).

5. CONCLUSION

This study analyzed ML models for earthquake-magnitude prediction using globally collected data from 2023. LightGBM, SVR, KNN, Ridge Regression, and Extra Trees were compared using geographic coordinates, seismic depth, and magnitude-related features. LightGBM yielded the strongest overall reported performance, with $R^2 = 0.9241$ and $MSE = 0.0474$, together with favorable results across the other statistical measures. Its regularization and scalability make it a promising approach for modeling nonlinear patterns in seismic data. The results support further investigation of ensemble ML methods for seismic risk analysis and early-warning applications, while the open dataset and documented evaluation framework provide a reproducible and extensible basis for research.

Future work will incorporate additional earthquake records to improve temporal coverage and long-term prediction, combine seismic observations with satellite ionospheric measurements, tectonic-plate velocities, and historical fault data, and examine explainable AI (XAI) methods to make model decisions more interpretable and trustworthy. Deployment of the

models in a real-time earthquake early-warning framework should also be evaluated with operational data and seismic events. Continued collaboration among seismologists, AI researchers, and geoinformatics specialists will be important for assessing these systems in practice.

REFERENCES

- [1] Artem Vesnin, Yury Yasyukevich, Natalia Perevalova, and Erman Senturk. Ionospheric response to the 6 february 2023 turkey–syria earthquake. *Remote Sensing*, 15(9):2336, 2023.
- [2] Hua-Lun Huang. Natural hazards: Earthquakes. In *Encyclopedia of Security and Emergency Management*, pages 662–669. Springer, 2021.
- [3] S. Mostafa Mousavi and Gregory C. Beroza. Machine learning in earthquake seismology. *Annual Review of Earth and Planetary Sciences*, 51:105–129, 2023.
- [4] Munish Bhatia, Tariq Ahamed Ahanger, and Ankush Manocha. Artificial intelligence based real-time earthquake prediction. *Engineering Applications of Artificial Intelligence*, 120:105856, 2023.
- [5] Assem Turarbek, Maktagali Bektemesov, Aliya Ongarbayeva, Assel Orazbayeva, Aizhan Koishybekova, and Yeldos Adetbekov. Deep convolutional neural network for accurate prediction of seismic events. *International Journal of Advanced Computer Science and Applications*, 14(10), 2023.
- [6] Alexander N. Safronov. Astronomical triggers as a cause of strong earthquakes. *International Journal of Geosciences*, 13(9):793–829, 2022.
- [7] David Teh and Tehmina Khan. Types, definition and classification of natural disasters and threat level. In *Handbook of Disaster Risk Reduction for Resilience: New Frameworks for Building Resilience to Disasters*, pages 27–56. Springer, 2021.
- [8] Mei-Yun Lin and Raluca Ilie. A review of observations of molecular ions in the earth’s magnetosphere-ionosphere system. *Frontiers in Astronomy and Space Sciences*, 8:745357, 2022.
- [9] Denis Anikiev, Claire Birnie, Umair bin Waheed, Tariq Alkhalifah, Chen Gu, Dirk J. Verschuur, and Leo Eisner. Machine learning in microseismic monitoring. *Earth-Science Reviews*, 239:104371, 2023.
- [10] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [11] Bernd Bischl, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, Difan Deng, and Marius Lindauer. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, 13(2):e1484, 2023.

- [12] Gautam Kunapuli. *Ensemble Methods for Machine Learning*. Simon and Schuster, 2023.
- [13] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.
- [14] Gaurav Singh Manral and Alka Chaudhary. Prediction of earthquake using machine learning algorithms. In *Proceedings of the 2023 4th International Conference on Intelligent Engineering and Management*, pages 1–5. IEEE, 2023.
- [15] Nada Badr Jarah, Abbas H. Hassin Alasadi, and Kadhim Mahdi Hashim. Earthquake prediction technique: A comparative study. *IAES International Journal of Artificial Intelligence*, page 1027, 2023.
- [16] Ewnetu Abebe, Hailemichael Kebede, Mickus Kevin, and Zelalem Demissie. Earthquakes magnitude prediction using deep learning for the horn of africa. *Soil Dynamics and Earthquake Engineering*, 170:107913, 2023.
- [17] Mustafa Abdul Salam, Lobna Ibrahim, and Diaa Salama Abdelminaam. Earthquake prediction using hybrid machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 12(5):654–665, 2021.
- [18] Pablo Lara, Quentin Bletery, Jean-Paul Ampuero, Adolfo Inza, and Hernando Tavera. Earthquake early warning starting from 3 s of records on a single station with machine learning. *Journal of Geophysical Research: Solid Earth*, 128(11):e2023JB026575, 2023.
- [19] Victor Manuel Velasco Herrera, Eduardo Antonio Rossello, Maria Julia Orgeira, Lucas Arioni, Willie Soon, Graciela Velasco, Laura Rosique-de La Cruz, Emmanuel Zuniga, and Carlos Vera. Long-term forecasting of strong earthquakes in north america, south america, japan, southern china and northern india with machine learning. *Frontiers in Earth Science*, 10:905792, 2022.
- [20] Papiya Debnath, Pankaj Chittora, Tulika Chakrabarti, Prasun Chakrabarti, Zbigniew Leonowicz, Michal Jasin-ski, Radomir Gono, and Elzbieta Jasinska. Analysis of earthquake forecasting in india using supervised machine learning classifiers. *Sustainability*, 13(2):971, 2021.
- [21] Rayan Abri and Harun Artuner. LSTM-based deep learning methods for prediction of earthquakes using ionospheric data. *Gazi University Journal of Science*, 35(4):1417–1431, 2022.
- [22] Bharat Bhargava and Sumanta Pasari. Earthquake prediction using deep neural networks. In *Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems*, volume 1, pages 476–479. IEEE, 2022.
- [23] Mohsen Yousefzadeh, Seyyed Ahmad Hosseini, and Mahdi Farnaghi. Spatiotemporally explicit earthquake prediction using deep neural network. *Soil Dynamics and Earthquake Engineering*, 144:106663, 2021.
- [24] Yao Li, Peng Cui, Chengming Ye, José Marcato Junior, Zhengtao Zhang, Jian Guo, and Jonathan Li. Accurate prediction of earthquake-induced landslides based on deep learning considering landslide source area. *Remote Sensing*, 13(17):3436, 2021.
- [25] Pan Xiong, Lei Tong, Kun Zhang, Xuhui Shen, Roberto Battiston, Dimitar Ouzounov, Roberto Iuppa, Danny Crookes, Cheng Long, and Huiyu Zhou. Towards advancing the earthquake forecasting by machine learning of satellite data. *Science of the Total Environment*, 771:145256, 2021.
- [26] Jing Jia and Wenjie Ye. Deep learning for earthquake disaster assessment: Objects, data, models, stages, challenges, and opportunities. *Remote Sensing*, 15(16):4098, 2023.
- [27] M. Apriani, S. K. Wijaya, et al. Earthquake magnitude estimation based on machine learning: Application to earthquake early warning system. In *Journal of Physics: Conference Series*, volume 1951, page 012057. IOP Publishing, 2021.
- [28] Earthquakes 2023 Global. Earthquakes 2023 global. Kaggle dataset, 2023.
- [29] Cheng Fan, Meiling Chen, Xinghua Wang, Jiayuan Wang, and Bufu Huang. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9:652801, 2021.
- [30] Andrea Barucci, Stefano Diciotti, Marco Giannelli, and Chiara Marzi. Data preparation for ai analysis. In *Introduction to Artificial Intelligence*, pages 133–150. Springer, 2023.