STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–17.
Published online in International Academic Press (www.IAPress.org)

Bayesian Optimization for Enhanced Prediction of Earthquakes with Machine Learning Techniques

Muzahem Al-Hashimi*, Heyam Hayawi, Mohammed Qasim Yahya Alawjar

Department of Statistics and Informatics, University of Mosul, Iraq

Abstract Earthquakes are destructive natural hazards that strike suddenly and unexpectedly, caused by elastic strain energy released along faults, mainly resulting from the gradual accumulation and persistent movement of tectonic plates. Several new approaches have been suggested to predict earthquakes. Machine learning (ML) methodologies have recently emerged as a robust mechanism in dealing with huge, complex, nonlinear, and less dependent on stringent assumptions. Most studies are often too narrow model comparisons, systematic hyperparameter tuning, and remain geographically constrained. Comprehensive benchmarking of regression-based machine learning and Bayesian tuning remains scarce, especially for the high seismicity regions of northern and eastern Iraq. multiple Machine learning methods were employed to model the earthquake magnitude in Iraq for the period 2004-2024. Six Bayesian optimization methods were implemented across the ML methods to optimize the model hyperparameters for the models. Eight mathematical indicators of seismicity parameters are extensively used as input features for the target defined as the earthquake magnitude observed prediction. The Extra Trees Regressor method demonstrated dominant predictive capability among the evaluated metrics, after parameter optimization using the Bayesian Optimization Hyperband. Magnitude deficit, or the difference between the largest observed magnitude and the largest expected magnitude based on the Gutenberg-Richter relationship, had the key influence on earthquake magnitude, as notably by the analysis of variable importance for the eight seismic parameters. These findings provide clear evidence of the effectiveness of Extra Trees Regressor in elucidating the complex relationships underlying earthquake magnitude and provide significant insights to guide the development of data-driven strategies aimed at improving earthquake response in Iraq.

Keywords Bayesian Optimization, Machine Learning Techniques, Seismic Parameters, Earthquake forecasting, Iraq

AMS 2010 subject classifications 62F15, 68T05, 86A15

DOI: 10.19139/soic-2310-5070-3563

1. Introduction

Earthquakes are stored energy within the Earth that is released as seismic energy, causing sudden earthquakes that occur and affect the Earth's crust. Earthquakes can cause landslides, resulting in loss of life and property [1, 2]. Highlighting the need for better forecasting, construction standards, and comprehensive hazard readiness. Earthquakes do not show identifiable patterns. Therefore, predicting the magnitude of an earthquake is a challenging task and can lead to imprecise predictions [3]. No reliable prediction currently exists [4, 5, 6], even though improving forecasting accuracy and reliability in earthquake prediction is undergoing scientific growth using machine learning. The ambition remains to deliver effective predictions, as evidenced by their positive implementation of preventive measures that lower quake damage and risk [3]. Given these challenges, several new approaches have been suggested to predict earthquakes. Machine learning (ML) methodologies have emerged as a powerful mechanism in dealing with huge, complex, nonlinear, and less dependent on stringent

*Correspondence to: Muzahem Al-Hashimi (Email: muzahim_alhashime@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Iraq (41002).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

assumptions [8, 9, 10, 11, 12, 13]. Ensemble and tree-based methods have been widely applied for earthquake prediction. Eight different machine-learning algorithms, including K-nearest neighbors, Support Vector Machine, and Classification and Regression Trees, were applied to regional seismic data to predict the occurrence status of future major earthquakes [14]. Different classification algorithms have been applied, such as Naive Bayes, K-nearest neighbors, logistic regression, and support vector machines, to historical seismic data for earthquake prediction [15]. Ensemble classifier methods were applied to the datasets of three regions to predict earthquakes of larger magnitudes using 60 seismic indicators [16]. Two hybrid machine learning approaches, using seven seismic have been applied in southern California. The study found that the flower pollination algorithm - least squares support vector machine outperforms the flower pollination algorithm - extreme learning machine in prediction accuracy [17]. Four machine learning algorithms to predict the magnitudes of future earthquakes in Iran. The findings suggest that Support Vector Machine and Deep Neural Network perform best with the impact of spatial parameters, Fault Density. Including Fault Density improved prediction accuracy for high-magnitude earthquakes, with SVM and DNN performing best [1]. A probabilistic neural network was presented to predict the magnitude, demonstrating their potential effectiveness as a method to robustly predict earthquakes by using seismicity indicators for moderate-magnitude events [18]. Radial basis function neural network models have been applied for the estimation of earthquake event occurrence using seismicity rates, demonstrating that the proposed models can provide superior earthquake occurrence prediction and effectively estimate interevent times [19]. Deep Learning, Artificial Neural Network, and Backpropagation were applied for predicting earthquakes using feature-engineered seismic data, showing that it can achieve a high accuracy in forecasting earthquake events by using a normalized ANN model [20]. Six ML models were compared to predict the maximum magnitude of an earthquake in the southern part of China, indicating that the LSTM model achieved better prediction in predicting earthquakes of moderate magnitude [21]. Nine seismic parameters with three Machine Learning approaches were employed to investigate the potential for a substantial earthquake prediction, indicating that the Artificial Neural Networks model can provide reliable earthquake forecasts for the short term [22]. Machine learning has been widely explored for forecasting earthquake magnitudes, employing ensemble models, neural networks, and deep learning approaches across different seismic regions. However, most studies are limited to narrow model comparisons, neglect systematic hyperparameter tuning, and remain geographically constrained. Comprehensive benchmarking of regression-based machine learning and Bayesian tuning remains scarce, especially for the high-seismicity regions of northern and eastern Iraq. Therefore, this paper aims to evaluate the effectiveness of ten machine learning methods, based on multiple regression models and six advanced Bayesian optimization strategies, in forecasting the magnitude of earthquakes according to seismicity indices in Iraq's eastern and northern seismotectonic regions, which have experienced multiple notable earthquakes in recent years. Furthermore, this study aims to identify which regression models can best forecast earthquake magnitudes from given seismic parameters, to support seismic hazard assessment and risk mitigation. These findings are anticipated to enhance seismic hazard evaluation and aid in risk mitigation efforts and emergency preparedness in areas with significant seismic activity.

2. Materials and Methods

2.1. Data Collection

The earthquake dataset spans the bounding rectangle, West edge at longitude 38.6° (Iraq–Syria border) to the East edge at 47.0° Iranian border, and South edge at latitude 29.76° (Basrah) to the North edge at latitude 37.5° (Northern Kurdish Region). The data has been collected from the General Authority of Meteorology and Seismic Monitoring for the period 2004 to 2024. The data contains 1327 earthquakes after removing duplicate and non-locatable events. The earthquake dataset was composed of the following attributes: Date of the earthquake event, Actual time of the earthquake (in UTC), Latitude of the epicenter, Longitude of the epicenter, Depth of the earthquake in kilometers, Local magnitude of the earthquake, and Name of the region or country where the earthquake occurred.

2.1.1. Derived Earthquake Parameters In this paper, eight mathematical indicators of seismicity parameters are utilized as input features for earthquake prediction [23]. These indicators contribute to evaluating a region's seismic

potential. The indicators are calculated for every time period. The target variable is the observed earthquake magnitude (M-observed).

1. T-value: the T-value reflects the foreshock event frequency, which is called the time elapsed during the previous n events with a magnitude exceeding an established threshold value, and is identified as

$$T = t_n - t_1$$

where t_1 is the occurrence time of the first event and t_n is the occurrence time of the n-th event.
2. The Mean Magnitude: it is the mean of the latest n earthquakes (measuring the mean size magnitude of the foreshocks) exceeding a specified magnitude threshold, is identified as

$$M_{\text{mean}} = \frac{\sum M_i}{n}$$

where M_i is the i-th event magnitude.
3. Rate of the root of the seismic energy released: It represents the rate at which energy is released, identified as

$$dE^{1/2} = \frac{\sum E^{1/2}}{T}$$

where $E^{1/2}$ is the square-rooted seismic energy computed based on the following formula

$$E = 10^{(11.8+1.5M)} \text{ ergs}$$

where ergs are a centimeter–gram–second system energy unit.
4. b value: denotes the slope in the earthquake frequency–magnitude distribution (from the Gutenberg–Richter). The Gutenberg–Richter identified as The Gutenberg-Richter relation is given by:

$$\log_{10} N = a - bM$$

where N refers to the number of events of magnitude $\geq M$, and a, b are constants. The b-value is the slope of the line, calculated as:

$$b = \frac{n\sum(M_i \log_{10} N_i) - \sum M_i \sum \log_{10} N_i}{(\sum M_i)^2 - n\sum M_i^2}$$

where M_i is the magnitude of the i-th event and N_i is the cumulative frequency (count of events with magnitude at least M_i).
5. Magnitude deficit ΔM: it evaluates the largest observed magnitude against the largest predicted by the Gutenberg–Richter law, identified as

$$\Delta M = M_{\text{max, observed}} - M_{\text{max, expected}}$$

6. η value: Sum of the squared deviations of the regression fit to the inverse power law. It is defined as

$$\eta = \frac{\sum \left(\log_{10} N_i - (a - bM_i)\right)^2}{n-1}.$$

7. μ-value: It measures the regularity of a fault's earthquake pattern. the mean recurrence interval is defined as

$$\mu = \frac{\sum t_{i,\text{characteristic}}}{n_{\text{characteristic}}}$$

Where $t_{i,\text{characteristic}}$ is the time between consecutive characteristic earthquakes, and $n_{\text{characteristic}}$ is the total count of observed characteristic events.

8. The c value (aperiodicity): assesses the consistency of the earthquake recurrence cycle relative to the mean μ, defined as

$$c = \frac{\text{standard deviation of the observed times}}{\mu}$$

We calculated the seismic indicators using a rolling 15-day window, with the target value specified as the observed magnitude in the subsequent period. This resulted in 259 input–target pairs, which were used to train and test the regression models.

2.2. Bayesian Optimization Algorithm

Bayesian Optimization is based on the principles of Bayesian inference. It offers a robust and effective strategy for optimizing the model hyperparameters for machine learning models by balancing exploration and exploitation. These facilitate an effective search for the hyperparameter that enhances forecasting accuracy substantially and the reliability of the model. Bayesian Optimization iteratively selects hyperparameters, incorporating information from each trial [24]. It develops a probabilistic model to estimate model performance for different hyperparameters. Optimizing hyperparameters efficiently is essential to the performance of seismic prediction models, which often involve complex nonlinear relationships between seismic parameters and earthquake magnitude. Various types of Bayesian Optimization Algorithms. Six Bayesian Optimization Algorithms were applied in this study to enhance predictive accuracy and are suitable for different data types and problem settings. The following algorithms are examined:

2.2.1. Gaussian Process Bayesian Optimization (GP-BO) It estimates the objective function across hyperparameter sets while quantifying uncertainty. The next hyperparameter set is identified for testing by using the expected improvement acquisition function to manage the trade-off between exploration and exploitation. Iterations continue until convergence.

2.2.2. Covariance Matrix Adaptation Evolution Strategy (CMA-ES) This algorithm employed a multivariate Gaussian distribution defined by the mean vector. It proceeds via successive iterations of the distribution. The algorithm allows to learn and utilizing the features of the local solution space throughout the optimization process by refining the covariance matrix of the distribution in response to the observed performance of the sampled solutions [25].

2.2.3. Bayesian Neural Network-Based Bayesian Optimization (BNN-BO) Neural networks are widely used models because of their expressive power and generalization capabilities [26]. Under the Bayesian paradigm, a Bayesian Neural Network is defined by a Bayesian treatment of weights, whose weights are modeled probabilistically [27]. The essence of Bayesian Neural Networks lies in viewing neural network weights as random variables instead of predetermined constants. The learner estimates a posterior distribution over the model's parameters conditioned on the observed data. It differs from classical learning, which aims to estimate a single optimal parameterization of the model using maximum-likelihood optimization [28].

2.2.4. Optuna Tree-Structured Parzen Estimator (TPE-BO) Optuna is a modern, it is autonomously operating hyperparameter optimization framework that accelerates the process of optimizing the performance of ML models [29]. Three main benefits in hyperparameter tuning that optuna provides: the API defines the hyperparameter search space dynamically (define-by-run paradigm), a robust pruning and sampling scheme, and it is simple to configure [30]. Optuna integrates Bayesian Optimization, where a surrogate model, as a Tree-structured Parzen Estimator, is used to determine the optimal set of hyperparameters tailored to a given model. The Tree-structured Parzen Estimator predicts which hyperparameters are promising and then employs the model to steer the optimization of parameters instead of testing every possible configuration.

2.2.5. Sequential Model-Based Algorithm Configuration (SMAC-BO) (Random Forest-based Bayesian Optimization) Sequential Model-based Algorithm Configuration (SMAC) (Random Forest BO). SMAC has been

proposed as an enhancement to the sequential model-based optimization. During cross-validation, SMAC limits its evaluation to the minimum number of folds needed to show that a configuration is not better than the current best-performing solution [31, 32]. It evaluates multiple configurations across instances, but its primary advantage is in modeling performance with Random Forests and utilizing an acquisition function. It typically uses Random Forests as the surrogate model, rendering it suitable for both numerical and categorical feature spaces [**?**]. SMAC has the functionality for identifying optimal configurations within large configuration spaces, and it can estimate the runtime or evaluation cost of a configuration in advance.

2.2.6. Bayesian Optimization with Hyperband (BOHB) BOHB integrates Bayesian optimization (BO) with Hyperband (HB) [33]. BOHB depends on HB to decide how many configurations to evaluate and the budget to allocate according to model-based search. The standard successive halving procedure is executed using these configurations after reaching the target number of configurations. The validation performance of each configuration hyperparameter is monitored during every training budget evaluation, and this information forms the foundation of our subsequent iteration models [34]. We adopt the HB's strategy for selecting budgets and continue utilizing successive halving, while introducing a BO component in place of random sampling to steer the search.

2.3. Model Selection and Hyperparameter Tuning for Earthquake Magnitude forecasting

We applied ten regression models by using Bayesian optimization exclusively on the training data for each model for hyperparameter tuning, and evaluated the performance of the models to predict earthquake magnitudes based on seismic parameters. The dataset is split into 80% for hyperparameter tuning and internal validation, and the remaining 20% test set is strictly held out and not used during hyperparameter optimization. The split is chronological to respect temporal order. The methods used are NuSVR (kernel-based support vector regression model), Random Forest Regressor (bagging of decision trees with randomness), Bagging Regressor (bootstrap aggregating ensemble for variance reduction), ExtraTreeRegressor (a single randomized tree), Extra Trees Regressor (ensemble of randomized trees), Hist Gradient Boosting Regressor (boosting using histograms, efficient on large data), LGBMRegressor (LightGBM, gradient boosting with histogram-based splitting), Decision Tree Regressor (tree-based regression model), XGBRegressor (XGBoost, gradient boosting with advanced regularization), and Gaussian Process Regressor (probabilistic kernel-based regression model). To measure forecasting performance, the metrics used are mean squared error (MSE), root mean squared error (RMSE), and R^2 score. For Bayesian optimization, we used six methods of Bayesian optimization: Gaussian Process (GP-BO), Covariance Matrix Adaptation Evolution Strategy (CMA-ES), Bayesian Neural Network (BNN-BO), Optuna (TPE), Sequential Model-based Algorithm Configuration (SMAC) (Random Forest BO), and Bayesian Optimization Hyperband (BOHB), then selected the best parameters for hyperparameter selection. For each model, we fitted it on the training data and predicted earthquake magnitudes on the test data. We computed MSE, RMSE, and R^2. Additionally, we performed a separate evaluation for large-magnitude earthquakes. The thresholds ($M \geq 4.5$ and $M \geq 5.0$) were used to avoid inflated accuracy driven by small-magnitude events, because the earthquake catalogue is dominated by small-magnitude events. The performance for larger-magnitude events is analyzed separately to address class imbalance caused by the dominance of small-magnitude earthquakes. All models were executed in the Python 3.8 environment.

3. Results and Discussion

Figure 1(a) shows a map of Iraq. The earthquakes were most concentrated along the eastern side of Iraq, especially near the Iran–Iraq border. The northeastern region suggests that this region experiences the highest earthquake activity, while Southern Western, and central regions have lower earthquake activity. Figure 1(b) reveals the bar chart of the number of the top 10 earthquakes recorded in different Iraqi governorates during the period 2004-2024. Diyala Governorate stands out with the highest number of earthquakes (229), which is nearly double that of Sulaymaniyah Governorate, which reports 122 earthquakes, tying for second place.

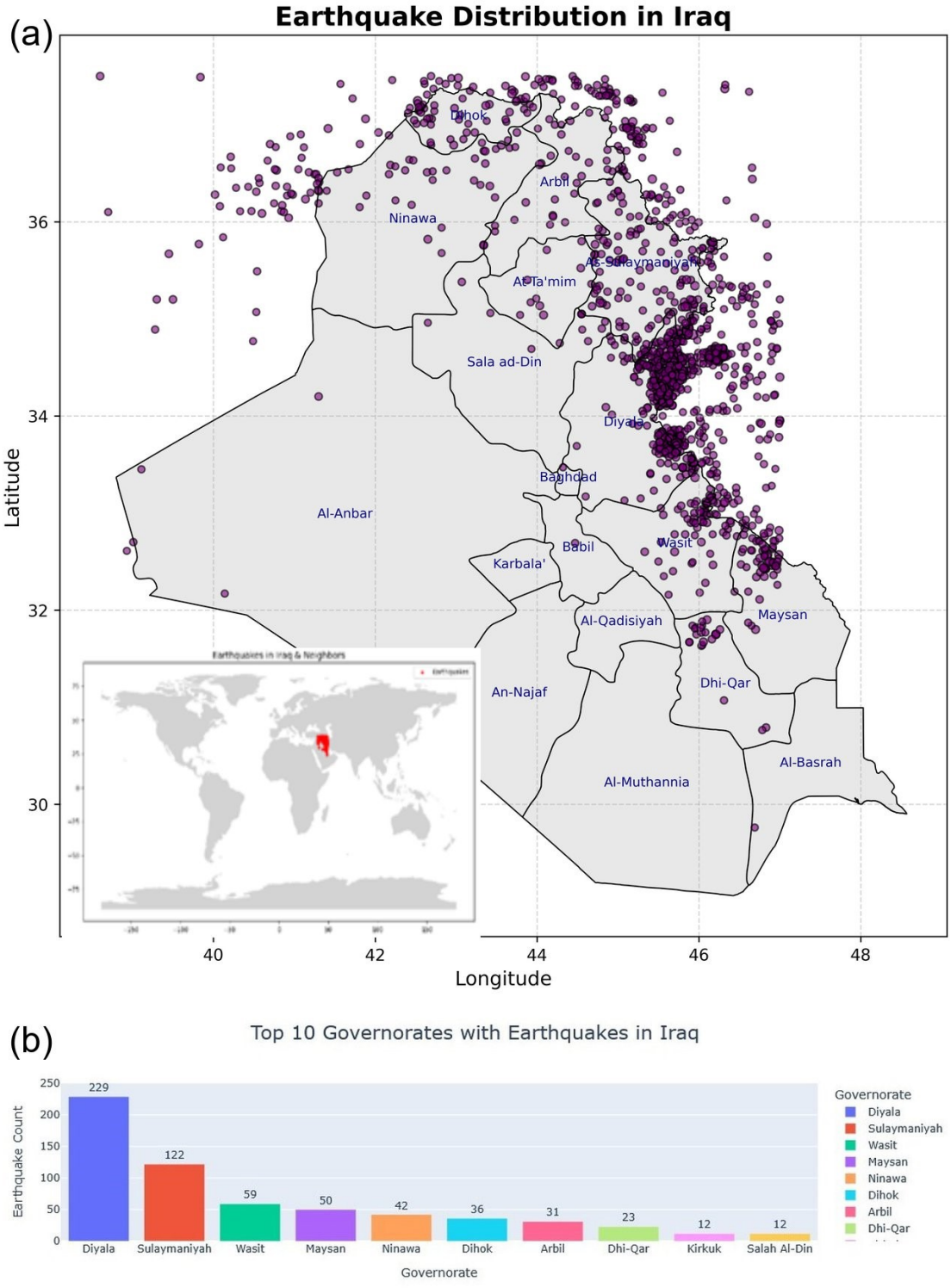

Figure 1. Geospatial distribution of earthquakes in Iraq (2004–2024): (a) Map of Iraq showing the locations of earthquakes in Iraq; (b) Bar chart of the top 10 earthquakes recorded in different Iraqi governorates (2004–2024).

Figure 2(a) illustrates the annual number of earthquakes for the period 2004-2024. The values fluctuated, ranging from 7 in 2004 to the first notable peak of 81 earthquakes in 2008. From 2017, the trend began rising again, with the highest in the entire series of 254 earthquakes in 2019, while the earthquakes stabilized between 46 and 63 annually between 2021 and 2024. Figure 2(b) depicts the distribution of earthquakes in Iraq and the border areas near Iraq. The distribution is right-skewed; most of the earthquake-observed values are grouped between 3.5 and 4.0, with frequencies exceeding 400. With the magnitude increase beyond 4.0, the frequency decreases sharply, and very few earthquake magnitudes extend beyond 5.5–7.2.

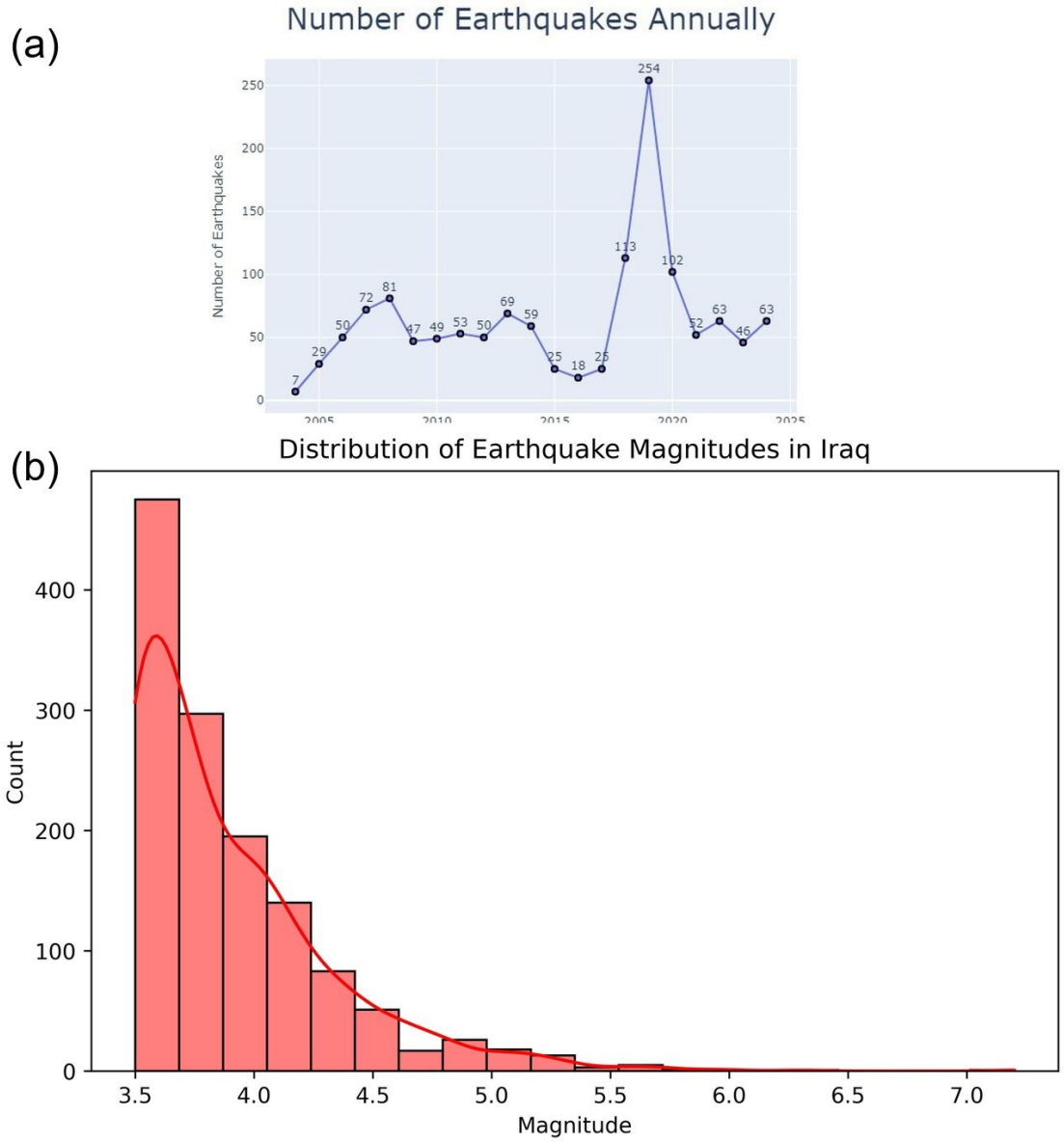

Figure 2. Statistical properties of earthquakes in Iraq (2004–2024): (a) Annual number of earthquakes; (b) Distribution of earthquake magnitudes.

Figure 3(a) highlights the depth distribution of the earthquakes in Iraq during 2004-2024. The distribution is positively skewed, with most earthquake depths concentrated among shallow depths. The highest frequencies occur

between depths 0 and 20, with noticeable peaks around 5-15. After depth 20, the frequency decreases steadily, with a few smaller spikes up to around 40 depth. Beyond 40 depth, earthquakes become relatively rare, but isolated cases appear up to a depth of 80. Figure 3(b) represents a scatter plot showing the relationship between depth and magnitude. Most earthquakes, across varying depths, occur at magnitudes 3.0–4.5. The distribution shows that most earthquakes occur at lower magnitudes (3–4.5), and they involve earthquakes of all depths, though small and middle-depth earthquakes are more frequent. As magnitude increases, earthquakes become less common, with very few events beyond magnitude 6.

Table 1 shows the performance of all Bayesian optimization (BO) methods combined with ten regression models. To quantify uncertainty, test R^2 values with their 95% confidence intervals were reported. Chronological Cross-validation R^2 mean and standard deviation (CV R^2) offer additional insight into model stability. Root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were included to assess predictive accuracy. This table allows for comparison across models and optimization strategies. Figures 4–6 show the high-performing earthquake magnitude forecasting models. The plots compare the predicted earthquake magnitudes versus the actual observed values (M-observed) for the ML methods with six BO methods. Figure 6(b) shows that ExtraTreesRegressor achieves a high predictive performance in all methods used. The model has the highest predictive accuracy (R^2), RMSE, MAPE, MAE), and it is the closest to the perfect forecasting using Bayesian Optimization Methods. The ensemble tree-based models (RandomForestRegressor, BaggingRegressor, XGBRegressor, HistGradientBoostingRegressor, and LightGBM), which combine multiple decision trees to improve predictive accuracy and stability, outperform a single tree (DecisionTree, ExtraTree). Although single-tree models perform poorly because they are less robust to noise and small sample sizes, this is critical for seismic datasets. The R^2 for ExtraTreesRegressor achieved 0.948 using BOHB optimization, indicating that the model structure has a stronger influence. The MSE is 0.0094, and it corresponds to an RMSE of about 0.097 magnitude units. showing, on average, the predicted earthquake magnitudes show only a minor deviation by ±0.1 from the observed values.

Table 1. Performance metrics for BO methods and regressors with Test R^2 95% confidence intervals.

Model	BO Method	Test R^2	(95% CI)	CV R^2 (mean)	CV R^2 (std)	RMSE	MAE	MAPE (%)
ExtraTreesRegr.	BOHB	0.948	[0.905,0.968]	0.879	0.011	0.097	0.095	1.382
RandomForestRegr.	BOHB	0.913	[0.840,0.949]	0.860	0.014	0.106	0.102	1.480
BaggingRegr.	BOHB	0.923	[0.842,0.969]	0.852	0.016	0.110	0.104	1.530
XGBRegr.	BOHB	0.921	[0.863,0.943]	0.858	0.013	0.108	0.103	1.500
HistGradientBoostingRegr.	BOHB	0.903	[0.827,0.930]	0.841	0.018	0.114	0.108	1.582
LGBMRegr.	BOHB	0.932	[0.854,0.965]	0.824	0.020	0.123	0.115	1.820
DecisionTreeRegr.	BOHB	0.838	[0.787,0.883]	0.802	0.026	0.129	0.121	1.920
ExtraTreeRegr.	BOHB	0.898	[0.832,0.921]	0.832	0.020	0.134	0.126	1.980
NuSVR	BOHB	0.733	[0.642,0.752]	0.694	0.038	0.132	0.118	2.000
GaussianProcessRegr.	BOHB	0.641	[0.604,0.682]	0.598	0.052	0.165	0.152	3.050
ExtraTreesRegr.	BNN-BO	0.943	[0.852,0.956]	0.855	0.017	0.106	0.103	1.520
RandomForestRegr.	BNN-BO	0.921	[0.855,0.950]	0.838	0.020	0.114	0.107	1.610
BaggingRegr.	BNN-BO	0.927	[0.880,0.947]	0.830	0.022	0.118	0.110	1.660
XGBRegr.	BNN-BO	0.936	[0.893,0.955]	0.835	0.019	0.116	0.108	1.630
HistGradientBoostingRegr.	BNN-BO	0.910	[0.873,0.941]	0.820	0.024	0.123	0.115	1.720
LGBMRegr.	BNN-BO	0.922	[0.865,0.960]	0.805	0.026	0.131	0.122	1.950
DecisionTreeRegr.	BNN-BO	0.780	[0.692,0.820]	0.798	0.029	0.135	0.127	2.050
ExtraTreeRegr.	BNN-BO	0.879	[0.791,0.915]	0.780	0.033	0.142	0.132	2.120
NuSVR	BNN-BO	0.734	[0.658,0.765]	0.720	0.036	0.139	0.126	2.180

Continued on next page

Table 1 (continued)

Model	BO Method	Test R^2	(95% CI)	CV R^2 (mean)	CV R^2 (std)	RMSE	MAE	MAPE (%)
GaussianProcessRegr.	BNN-BO	0.647	[0.566,0.683]	0.570	0.068	0.173	0.160	3.320
ExtraTreesRegr.	CMA-ES	0.937	[0.860,0.979]	0.848	0.020	0.109	0.104	1.570
RandomForestRegr.	CMA-ES	0.916	[0.835,0.945]	0.828	0.024	0.119	0.110	1.680
BaggingRegr.	CMA-ES	0.896	[0.827,0.917]	0.820	0.026	0.123	0.114	1.730
XGBRegr.	CMA-ES	0.927	[0.852,0.964]	0.825	0.023	0.121	0.112	1.700
HistGradientBoostingRegr.	CMA-ES	0.910	[0.813,0.929]	0.812	0.028	0.126	0.116	1.790
LGBMRegr.	CMA-ES	0.918	[0.859,0.951]	0.800	0.030	0.134	0.122	2.010
DecisionTreeRegr.	CMA-ES	0.840	[0.792,0.873]	0.790	0.034	0.140	0.126	2.180
ExtraTreeRegr.	CMA-ES	0.880	[0.810,0.922]	0.772	0.038	0.146	0.128	2.260
NuSVR	CMA-ES	0.725	[0.666,0.764]	0.705	0.040	0.143	0.128	2.270
GaussianProcessRegr.	CMA-ES	0.647	[0.573,0.674]	0.560	0.072	0.177	0.162	3.410
ExtraTreesRegr.	SMAC-BO	0.934	[0.891,0.964]	0.865	0.015	0.102	0.100	1.460
RandomForestRegr.	SMAC-BO	0.920	[0.837,0.938]	0.848	0.018	0.111	0.105	1.560
BaggingRegr.	SMAC-BO	0.915	[0.823,0.949]	0.840	0.020	0.114	0.108	1.600
XGBRegr.	SMAC-BO	0.928	[0.894,0.947]	0.845	0.017	0.112	0.106	1.580
HistGradientBoostingRegr.	SMAC-BO	0.905	[0.826,0.928]	0.832	0.021	0.118	0.112	1.650
LGBMRegr.	SMAC-BO	0.924	[0.843,0.945]	0.815	0.024	0.128	0.120	1.880
DecisionTreeRegr.	SMAC-BO	0.838	[0.756,0.856]	0.802	0.026	0.134	0.125	2.020
ExtraTreeRegr.	SMAC-BO	0.841	[0.775,0.865]	0.785	0.030	0.140	0.128	2.090
NuSVR	SMAC-BO	0.722	[0.642,0.763]	0.730	0.032	0.136	0.125	2.050
GaussianProcessRegr.	SMAC-BO	0.642	[0.600,0.676]	0.575	0.062	0.169	0.158	3.210
ExtraTreesRegr.	GP-BO	0.940	[0.892,0.974]	0.840	0.025	0.115	0.110	1.680
RandomForestRegr.	GP-BO	0.910	[0.862,0.924]	0.825	0.028	0.123	0.115	1.790
BaggingRegr.	GP-BO	0.914	[0.878,0.957]	0.815	0.030	0.126	0.115	1.840
XGBRegr.	GP-BO	0.927	[0.889,0.949]	0.820	0.029	0.125	0.112	1.820
HistGradientBoostingRegr.	GP-BO	0.912	[0.857,0.931]	0.805	0.032	0.130	0.116	1.900
LGBMRegr.	GP-BO	0.916	[0.864,0.944]	0.795	0.034	0.136	0.122	2.050
DecisionTreeRegr.	GP-BO	0.820	[0.761,0.857]	0.785	0.036	0.142	0.126	2.250
ExtraTreeRegr.	GP-BO	0.884	[0.809,0.926]	0.765	0.040	0.148	0.130	2.330
NuSVR	GP-BO	0.727	[0.648,0.773]	0.690	0.048	0.149	0.135	2.380
GaussianProcessRegr.	GP-BO	0.650	[0.567,0.686]	0.550	0.082	0.180	0.165	3.550
ExtraTreesRegr.	TPE-BO	0.928	[0.875,0.944]	0.875	0.018	0.104	0.102	1.520
RandomForestRegr.	TPE-BO	0.911	[0.870,0.955]	0.860	0.020	0.112	0.106	1.630
BaggingRegr.	TPE-BO	0.893	[0.847,0.926]	0.852	0.021	0.115	0.107	1.680
XGBRegr.	TPE-BO	0.918	[0.884,0.954]	0.865	0.019	0.109	0.103	1.600

Continued on next page

Table 1 (continued)

Model	BO Method	Test R^2	(95% CI)	CV R^2 (mean)	CV R^2 (std)	RMSE	MAE	MAPE (%)
HistGradientBoostingRegr.	TPE-BO	0.896	[0.829,0.928]	0.848	0.022	0.117	0.110	1.700
LGBMRegr.	TPE-BO	0.913	[0.876,0.954]	0.835	0.024	0.122	0.112	1.820
DecisionTreeRegr.	TPE-BO	0.857	[0.809,0.883]	0.805	0.028	0.136	0.128	2.150
ExtraTreeRegr.	TPE-BO	0.798	[0.762,0.839]	0.790	0.031	0.141	0.130	2.220
NuSVR	TPE-BO	0.724	[0.687,0.741]	0.730	0.040	0.143	0.130	2.250
GaussianProcessRegr.	TPE-BO	0.632	[0.570,0.680]	0.585	0.065	0.172	0.160	3.400

To evaluate models' performance for moderate-to-large earthquakes ($M \geq 4.5$ and ≥ 5.0), six Bayesian optimization methods were used. Model performance across magnitude thresholds was evaluated using error metrics and classification-based metrics, summarized and visualized using a heatmap plot to highlight trade-offs and tail-event sensitivity. Figure 7(a) shows the results for $M \geq 4.5$, where the best-performing configuration per model was selected and visualized using heatmaps. XGBRegressor with TPE-BO achieved the lowest MAE (0.114), RMSE (0.143), and high precision (0.938), recall (0.938), and F1-score (0.938), indicating strong overall predictive performance. The XGBRegressor consistently outperformed the other models due to its gradient boosting framework, which iteratively corrects forecasting errors and effectively prioritizes informative samples, leading to enhanced learning performance of nonlinear relationships and rare, high-magnitude events. The very competitive method is ExtraTreesRegressor, which reduces variance via averaging multiple independent trees, producing stable learning with limited responsiveness to extreme events. HistGradientBoostingRegressor is based on boosting principles and histogram-based tree construction, which explains its comparable accuracy and tail sensitivity. NuSVR and GaussianProcessRegressor, as kernel-based methods, assume a smoother functional relationship. Their ability is limited to capturing the complex, skewed patterns characteristic of large-magnitude earthquakes. Figure7(b) shows the results for $M \geq 5.0$. XGBRegressor with SMAC-BO achieved the lowest MAE (0.145), RMSE (0.190), and high precision (1.000), recall (0.625), and F1-score (0.769), surpassing other models in predicting rare events, though recall reaches 0.625, some extreme events continue to go undetected. This indicates that XGBRegressor provides reliable forecasts for moderate to strong earthquakes. The figure shows that for $M \geq 5.0$, XGBRegressor (SMAC-BO) remains best, with minimal errors while maximizing precision and recall. ExtraTreeRegressor maintains comparable errors while demonstrating lower precision. ExtraTrees and RandomForest, which are ensemble tree methods, are precise but under-detect extremes. The Kernel-based models (NuSVR and GaussianProcess) yield high errors and zero recall, which fail to capture rare large events. Results show that the XGBRegressor delivers the strongest performance for high-magnitude events, which represent the highest hazard risk. While the performance metrics for most models generally decrease, indicating the challenge of modeling large-magnitude events. This suggests models tend to under-detect events, commonly observed in skewed seismic datasets. XGBRegressor performs well due to its high flexibility, as it is capable of capturing complex patterns and recurrent low-magnitude. ExtraTreesRegressor is predominant in the training set across all magnitudes. XGBRegressor outperforms for learn rare cases and the high-magnitude events ($M \geq 4.5$ and ≥ 5.0) as a result of its boosting framework's concentration on hard-to-predict cases.

Figure8(a) shows the performance of the Extra Trees Regressor model on the training set (80%) for a magnitude dataset in Iraq. The figure shows that the model's forecasts match the observed values, which reflects strong training accuracy, while Figure8(b) presents the performance on the testing dataset. The two curves are very similar pattern. This suggests that the Extra Trees Regressor model accurately reflects the variations and trends in the data.

Figure 9 visualizes the relative contribution of the features applied in the Extra Trees Regressor model for forecasting earthquake magnitude. The highest importance feature was ΔM. The magnitude deficit is the numerical discrepancy between expected and observed magnitudes in a given region. The deficit relative to Gutenberg–Richter expectations indicates elevated levels of unreleased tectonic stress in these regions. These zones in northern and eastern Iraq lie along seismically active fault zones, particularly the Zagros fold-and-thrust belt and adjacent seismotectonic structures, where the concentration of unreleased stress increases the likelihood of larger

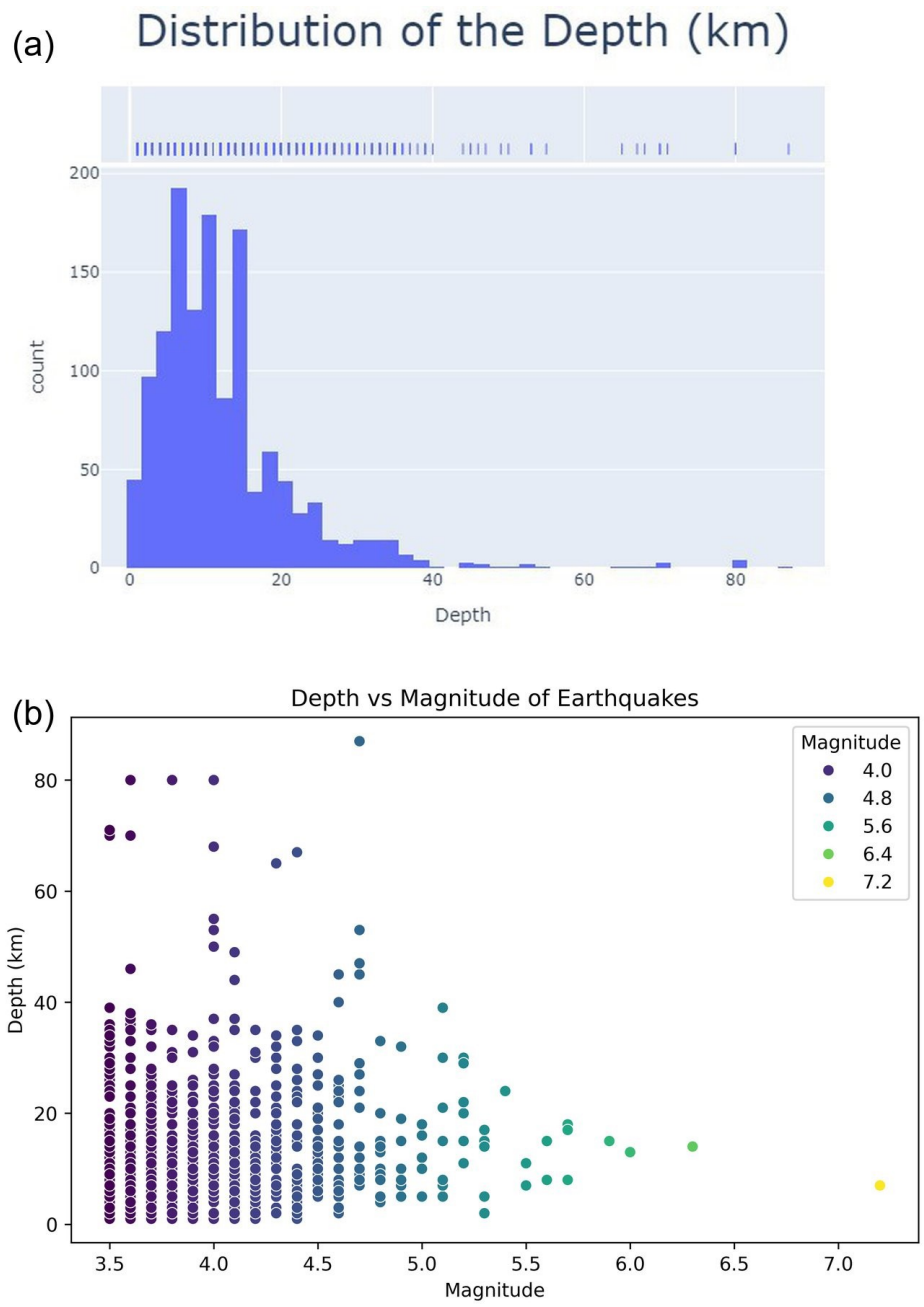

Figure 3. Depth properties of earthquakes in Iraq (2004–2024): (a) Depth distribution of earthquakes; (b) Scatter plot of the relationship between earthquake depth and magnitude.

earthquakes. This underscores the importance of assessing stress accumulation for forecasting future earthquake activity. Rate root, c-value, and μ-value also exhibit high importance, indicating that cycles of stress buildup and previous rupture events are key factors in estimating future seismic magnitudes.

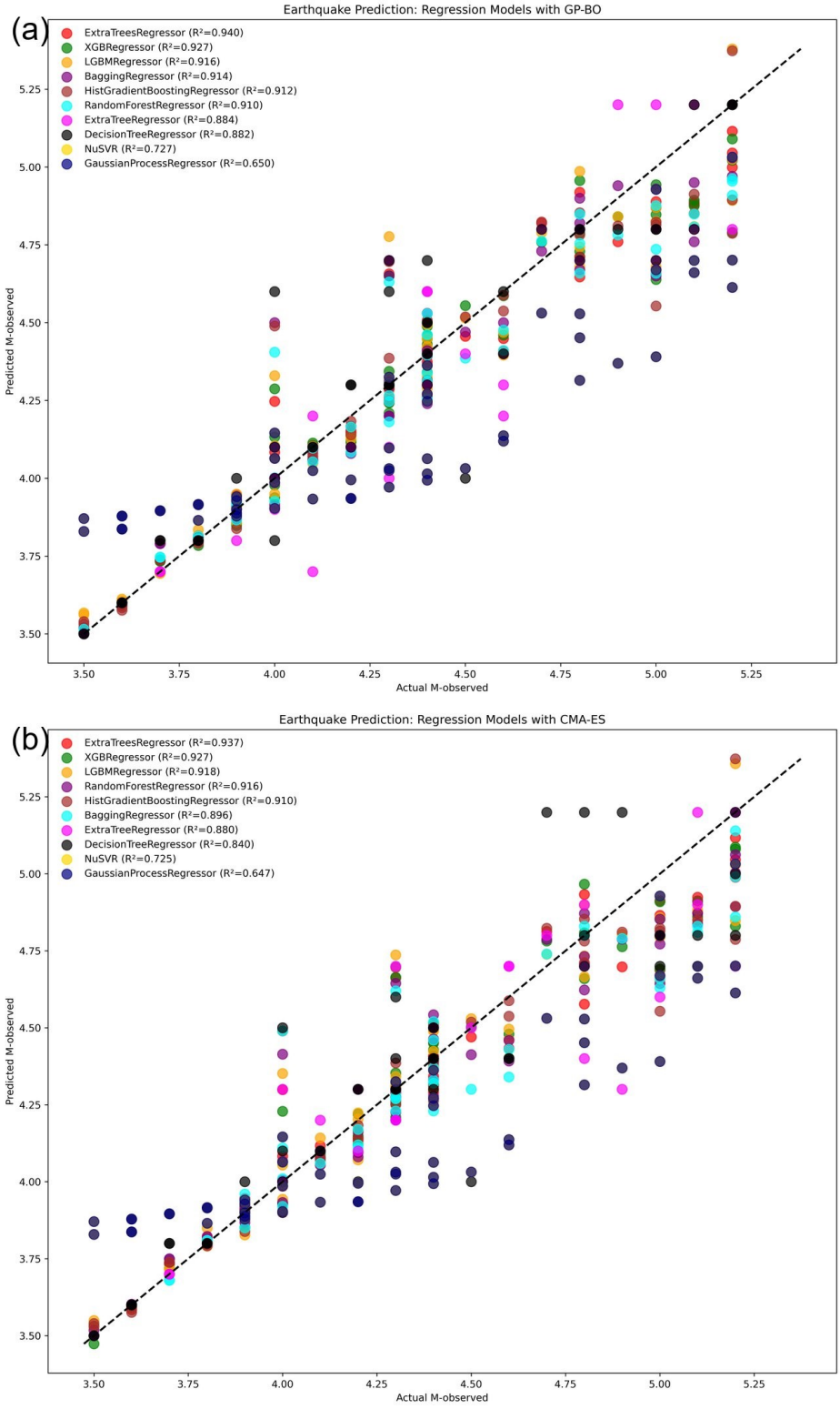

Figure 4. Comparison of Earthquake magnitude forecasting using regression models with optimization strategies: (a) GP-BO optimization results; (b) CMA-ES optimization results.

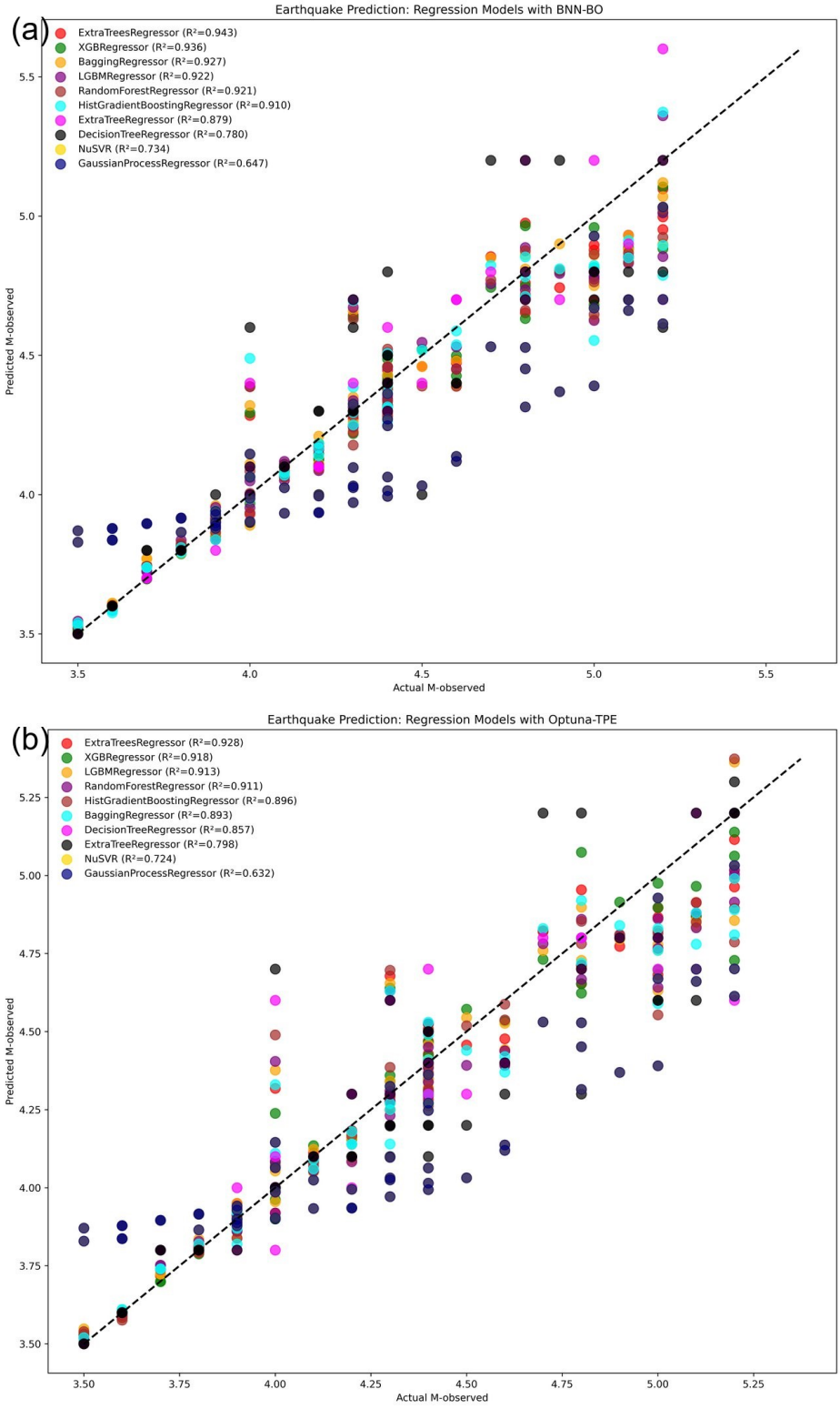

Figure 5. Comparison of Earthquake magnitude forecasting using regression models with optimization strategies: (a) BNN-BO optimization results; (b) Optuna-TPE optimization results.

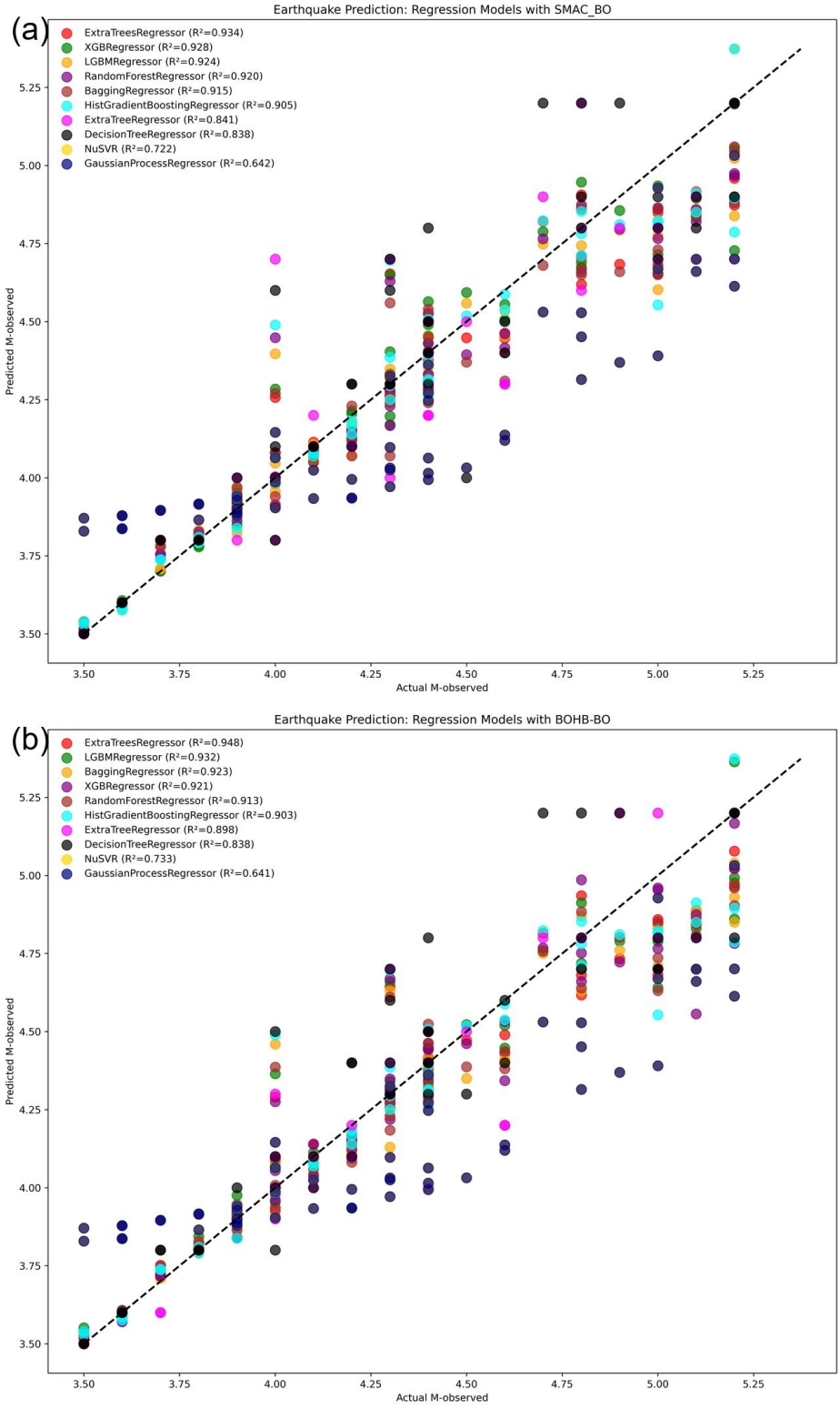

Figure 6. Comparison of Earthquake magnitude forecasting using regression models with optimization strategies: (a) SMAC-BO optimization results; (b) BOHB-BO optimization results.

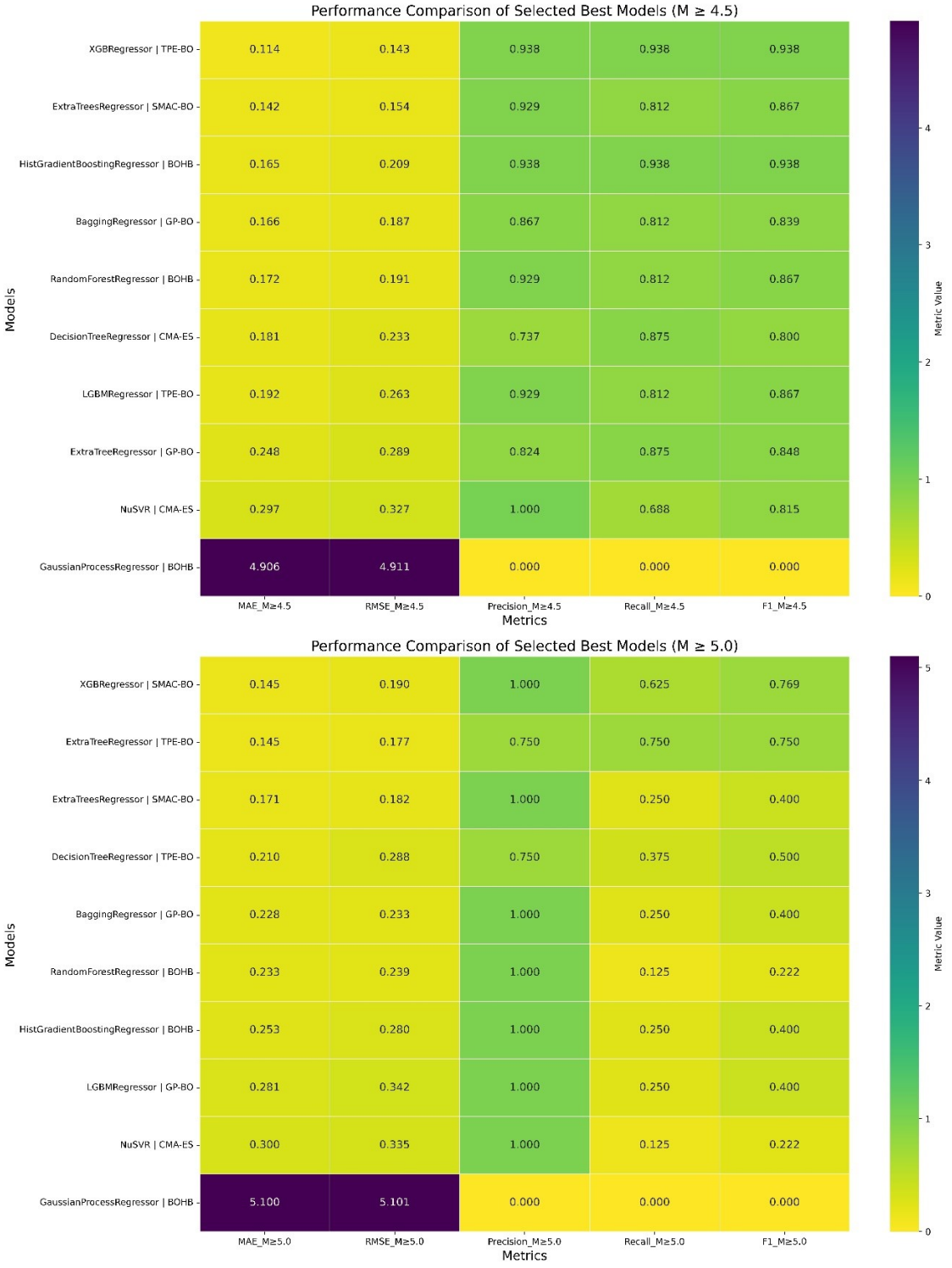

Figure 7. Heatmap for comparison of Earthquake forecasting using regression models with optimization strategies at different magnitude thresholds: (a) $M \geq 4.5$; (b) $M \geq 5.0$.

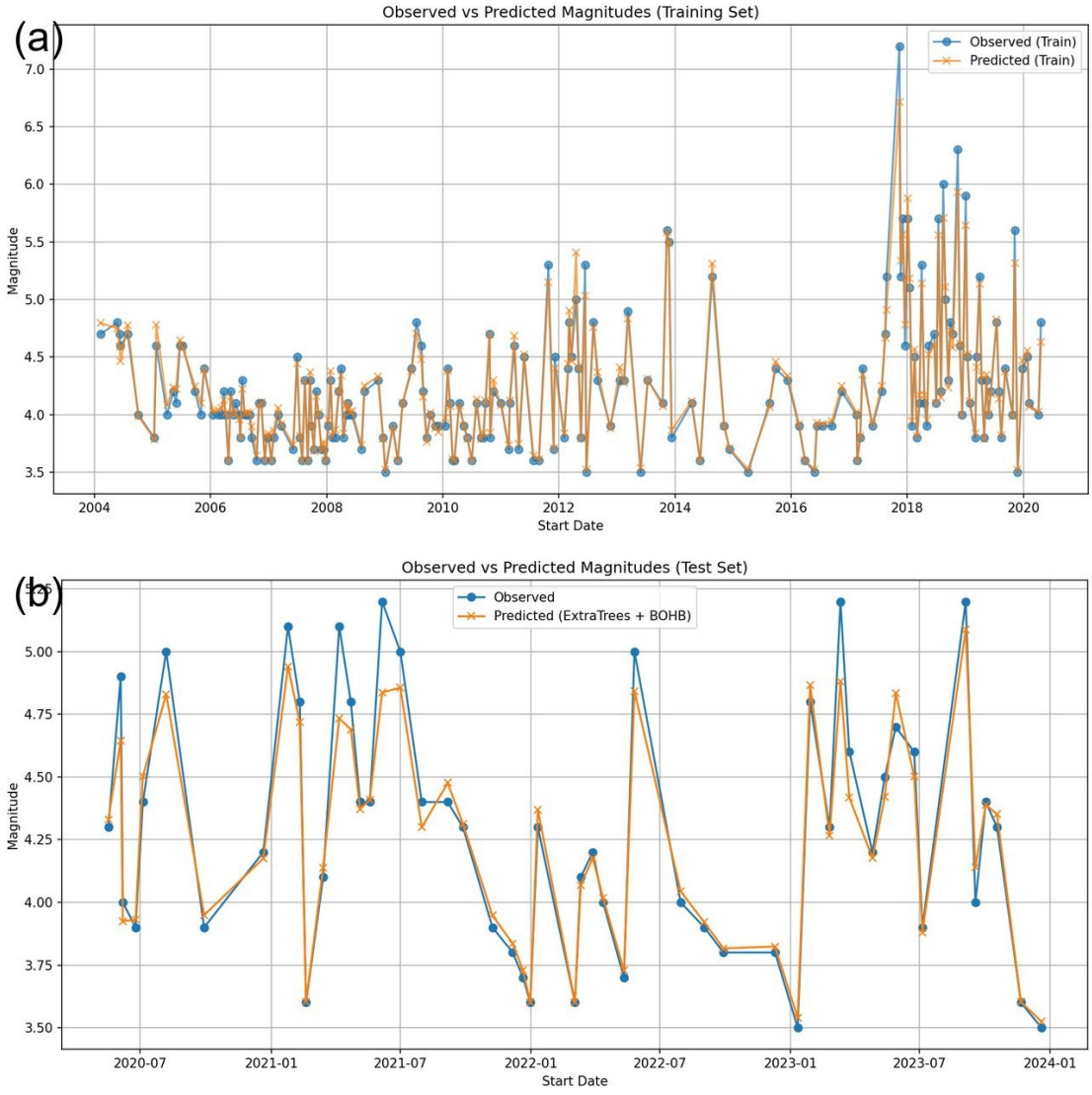

Figure 8. Performance of the Extra Trees model for earthquake forecasting: (a) Training results; (b) Testing results.

Conclusion

Ten ML methods were employed to model and characterize earthquakes in Iraq during the period 2004-2024. Eight seismic parameters as predictors with earthquake magnitude as the dependent variable. Six Bayesian optimization methods were implemented across the ML methods to ensure optimal model performance. Among the evaluated models, the Extra Trees Regressor method demonstrated superior predictive capability using the BOHB Bayesian optimization for all magnitude data. For$M \geq 4.5$ and $M \geq 5.0$ analysis, the performance metrics generally decrease for most models, reflecting the difficulty in capturing large-magnitude events. XGBRegressor delivers the strongest performance for high-magnitude events, which are of greatest hazard significance. ExtraTreesRegressor is precise but under-detects extremes. Analysis of variable importance revealed that seismic parameter ΔM had the strongest influence on earthquake magnitude. These findings highlight the effectiveness of Extra Trees Regressor in

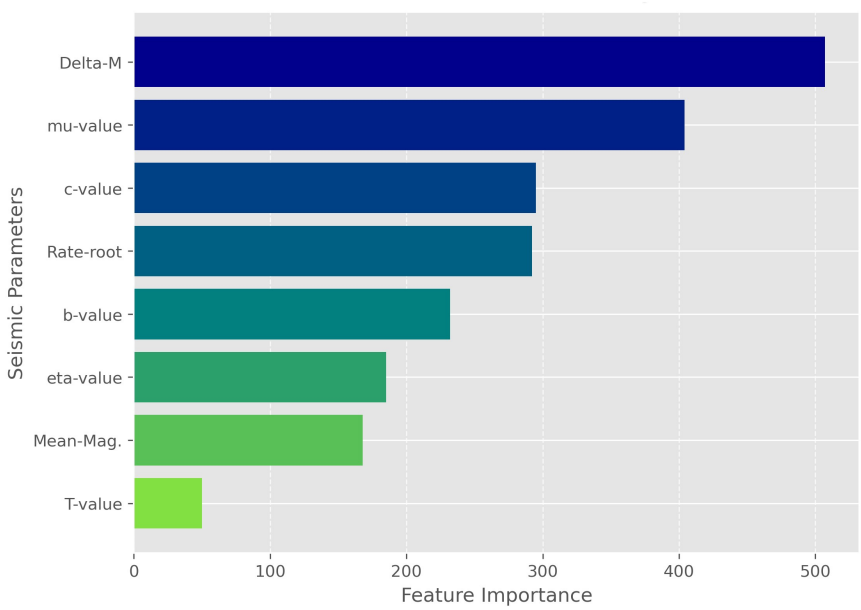

Figure 9. Feature importance of the input parameters.

capturing the complex relationships underlying earthquake magnitude and provide valuable insights for developing data-driven strategies aimed at improving earthquake response in Iraq.

4. Contributions

Muzahem, Al-Hashimi — Conceptualization, Data Analysis, Methodology, Software, Drafted the Initial Manuscript; Heyam, Hayawi— Conceptualization, Data curation, Methodology, Project Administration, Validated the Results, Writing–Review and Editing; Mohammed, Alawjar — Investigation, Supervision, Resources, Writing – review & editing.

REFERENCES

1. M. Yousefzadeh, S. A. Hosseini, and M. Farnaghi, *Spatiotemporally explicit earthquake prediction using deep neural network*, Soil Dynamics and Earthquake Engineering, vol. 144, p. 106663, 2021.
2. A. Kaushal, A. K. Gupta, and V. K. Sehgal, *Earthquake prediction optimization using deep learning hybrid RNN–LSTM model for seismicity analysis*, Soil Dynamics and Earthquake Engineering, vol. 195, p. 109432, 2025.
3. M. H. Al Banna, K. A. Taher, M. S. Kaiser, M. Mahmud, M. S. Rahman, A. S. Hosen, and G. H. Cho, *Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges*, IEEE Access, vol. 8, pp. 192880–192923, 2020.
4. M. Al-Hashimi, H. H. Alawjar, and M. Alawjar, *Ensemble Method for Intervention Analysis to Predict the Water Resources of the Tigris River*, Statistics, Optimization & Information Computing, [to appear], 2025.
5. G. V. Otari and R. V. Kulkarni, *A review of application of data mining in earthquake prediction*, International Journal of Computer Science and Information Technologies, vol. 3, no. 2, pp. 3570–3574, 2012.
6. W. XiangJun and M. M. Al-Hashimi, *The comparison of adaptive neuro-fuzzy inference system (ANFIS) with nonlinear regression for estimation and prediction*, in 2012 International Conference on Information Technology and e-Services, pp. 1–7, IEEE, 2012.
7. N. S. Ibrahim and H. A. Hayawi, *Employment the state space and Kalman filter using ARMA models*, International Journal on Advanced Science Engineering Information Technology, vol. 11, no. 1, pp. 145–149, 2021.
8. M. M. Al-Hashimi, H. A. Hayawi, and M. Al-Kassab, *A comparative study of traditional methods and hybridization for predicting non-stationary sunspot time series*, Computer Science, vol. 19, no. 1, pp. 195–203, 2024.
9. H. Hayawi, M. Al-Hashimi, and M. Alawjar, *Machine learning methods for modelling and predicting dust storms in Iraq*, Statistics, Optimization & Information Computing, vol. 13, no. 3, pp. 1063–1075, 2025.

10. I. U. Sikder and T. Munakata, *Application of rough set and decision tree for characterization of premonitory factors of low seismic activity*, Expert Systems with Applications, vol. 36, no. 1, pp. 102–110, 2009.
11. T. H. Ali, H. H. Taha, B. S. Sedeeq, and H. A. A. Hayawi, *Novel Hampel filtering and robust regression in control chart applications for regression process monitoring*, Journal of Statistical Theory and Applications, 2025.
12. T. H. Ali, H. H. Taha, B. S. Sedeeq, and H. A. A. Hayawi, *An innovative hybrid control chart combining wavelet decomposition and support vector machine for effective outlier detection*, Frontiers in Applied Mathematics and Statistics, vol. 11, Oct. 2025.
13. H. H. Taha, H. A. A. Hayawi, T. H. Ali, and S. R. Ahmed, *Wavelet Daubechies enhanced average chart incorporating classical Shewhart and Bayesian techniques*, Statistics, Optimization and Information Computing, vol. 14, Nov. 2025.
14. R. Mallouhy, C. Abou Jaoude, C. Guyeux, and A. Makhoul, *Major earthquake event prediction using various machine learning algorithms*, in 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), pp. 1–7, IEEE, 2019.
15. W. Han, Y. Gan, S. Chen, and X. Wang, *Study on earthquake prediction model based on traffic disaster data*, in 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), pp. 331–334, IEEE, 2020.
16. Y. Zhao and D. Gorse, *Earthquake prediction from seismic indicators using tree-based ensemble learning*, Natural Hazards, vol. 120, no. 3, pp. 2283–2309, 2024.
17. M. A. Salam, L. Ibrahim, and D. S. Abdelminaam, *Earthquake prediction using hybrid machine learning techniques*, International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, pp. 654–665, 2021.
18. H. Adeli and A. Panakkat, *A probabilistic neural network for earthquake magnitude prediction*, Neural Networks, vol. 22, no. 7, pp. 1018–1024, 2009.
19. A. Alexandridis, E. Chondrodima, E. Efthimiou, G. Papadakis, F. Vallianatos, and D. Triantis, *Large earthquake occurrence estimation based on radial basis function neural networks*, IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 9, pp. 5443–5453, 2013.
20. D. Mehta, P. P. Das, S. Ghosh, S. Mishra, A. Alkhayyat, and V. Sharma, *A normalized ANN model for earthquake estimation*, in 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 151–155, IEEE, 2023.
21. J. Zhang, K. Sun, X. Han, and N. Mao, *Application of machine learning models to multi-parameter maximum magnitude prediction*, Applied Sciences, vol. 14, no. 24, p. 11854, 2024.
22. S. Ommi and M. Hashemi, *Machine learning technique in the north Zagros earthquake prediction*, Applied Computing and Geosciences, vol. 22, p. 100163, 2024.
23. A. Panakkat and H. Adeli, *Neural network models for earthquake magnitude prediction using multiple seismicity indicators*, International Journal of Neural Systems, vol. 17, no. 1, pp. 13–33, 2007.
24. M. Santos, *Bayesian optimization for hyperparameter tuning*, Journal of Bioinformatics and Artificial Intelligence, vol. 2, no. 2, pp. 1–13, 2022.
25. A. Thomaser, M. E. Vogt, T. Bäck, and A. V. Kononova, *Optimizing CMA-ES with CMA-ES*, in IJCCI, pp. 214–221, 2023.
26. J. Arbel, K. Pitas, M. Vladimirova, and V. Fortuin, *A primer on Bayesian neural networks: Review and debates*, arXiv preprint arXiv:2309.16314, 2023.
27. L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, *Hands-on Bayesian neural networks—A tutorial for deep learning users*, IEEE Computational Intelligence Magazine, vol. 17, no. 2, pp. 29–48, 2022.
28. P. Izmailov, S. Vikram, M. D. Hoffman, and A. G. G. Wilson, *What are Bayesian neural network posteriors really like?*, in International Conference on Machine Learning, pp. 4629–4640, PMLR, 2021.
29. B. A. Dada, N. I. Nwulu, and S. O. Olukanmi, *Bayesian optimization with Optuna for enhanced soil nutrient prediction: A comparative study with genetic algorithm and particle swarm optimization*, Smart Agricultural Technology, p. 101136, 2025.
30. J. P. Lai, Y. L. Lin, H. C. Lin, C. Y. Shih, Y. P. Wang, and P. F. Pai, *Tree-based machine learning models with optuna in predicting impedance values for circuit analysis*, Micromachines, vol. 14, no. 2, p. 265, 2023.
31. K. Eggensperger, F. Hutter, H. Hoos, and K. Leyton-Brown, *Efficient benchmarking of hyperparameter optimizers via surrogates*, in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, no. 1, 2015.
32. A. M. Vincent and P. Jidesh, *An improved hyperparameter optimization framework for AutoML systems using evolutionary algorithms*, Scientific Reports, vol. 13, p. 4737, 2023.
33. S. Falkner, A. Klein, and F. Hutter, *BOHB: Robust and efficient hyperparameter optimization at scale*, in International Conference on Machine Learning, pp. 1437–1446, 2018.
34. T. A. A.-R. S. Al-Hasso, H. A. A. Hayawi, and T. H. Ali, *Using linear wavelets in analyzing the GARCH model with the simulation*, Pakistan Journal of Statistics, vol. 42, no. 1, pp. 27–44, 2026.