

<Conference abbreviation>

<Conference Series name>

<Volume number and Year> <DOI Number>

Earthquake Forecasting through Machine Learning: Predicting Magnitudes from USGS Seismic Data

Doaa Sami Khafaga

Department of Computer Sciences College of Computer & Information Sciences

Princess Nourah bint Abdulrahman University Riyadh, Saudi Arabia dskhafga@pnu.edu.sa

Abstract.

Estimating earthquake quantity in absolute terms such that the potential risks involved in an earthquake can be well prevented is a significant need in successful alert systems, a facet we consider instrumental in this work. This work discusses the capability of using machine learning models to predict earthquake magnitudes using data from USGS. Real-time data informing of all earthquakes in the world were used to compare model performances against the four assessment measures of MSE, RMSE, MAE, and R2. KNN and Random forest models were the most effective since they recorded low error rates with MSE of 0.11 and 0.14, respectively. By comparing KNN and Random Forest, this analysis demonstrates their ability to identify non-linear interactions of a dataset, such as the seismic data, to even aid in choosing the best model for prediction earthquake magnitude prediction. The results emphasize the sensitivity of prediction to model selection and suggest that machine learning could strengthen disaster planning.

Keywords. Earthquake Prediction, Machine Learning, USGS Seismic Data, Random Forest, K-Nearest Neighbors, Predictive Modeling

1. INTRODUCTION

Over the last decade, an enhancement in the demand for optimizing techniques to solve discrete/non-discretized engineering and computational problems has made metaheuristic optimization more appealing in areas like earthquake prediction. Therefore, training metaheuristic optimization algorithms is necessary to meet the complex challenge of multi-dimensional and nonlinear problems, which cannot be handled with existing methods in terms of efficiency and accuracy. These optimization algorithms are not only used in different data mining tasks of predictive modeling, including linear, nonlinear, and non-stationary but also in various fields such as telecommuting churn prediction and wind power forecasting. These applications show how metaheuristic algorithms are vital to address system robustness concerns in sensitive fields, including earthquake and disaster handling mechanisms.

Metaheuristic applications can be used for different mathematical problems and have been utilized in areas such as medical imaging and diagnostic correctness. For example, methods of solving the inverse boundary value problem for heat conduction have provided helpful information about thermal properties when all conditions are highly maintained [1]. Additionally, metaheuristic-driven convolutional neural network (CNN) models have been instrumental in optimizing hyperparameters for diabetic maculopathy diagnosis [2] and enhancing transfer learning in chest X-ray diagnostics [3]. Such improvements are reflected

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

in ensemble modeling methodologies, which were found to be specifically tailored for predicting bandwidth in metamaterial antennas [4]. These algorithmic flexibility underlines the ability of these algorithms to tackle complex engineering problems. By incorporating these techniques in metaheuristics, one finds these metaheuristic techniques to provide the tools crucial in anticipating and mitigating disasters in areas prone to disasters such as those caused by earthquakes.

In summary, metaheuristic optimization is a robust and applicable method in numerous fields, and those areas are progressively improving in terms of effectiveness, extensibility and adjustability of the solutions for dynamic and complex problems. Such progression speaks for their relevance in high-impact applications in environmental and industrial systems, disaster preparedness and response machinery, including earthquake-sensitive systems [5].

3. DISCUSSION AND RESULTS

This section briefly analyzes the dataset as well as compares and outcomes of methodologies utilized machine learning models for the prediction of an earthquake magnitude.

A. Dataset

The dataset used for this study was sourced from the U.S. Geological Survey (USGS), which offers scientific data to aid in understanding the Earth's geological processes and to reduce risks associated with natural disasters. USGS monitors and reports on earthquakes, offering real-time updates and comprehensive data on seismic events.

B. Data Analysis

Information which could potentially be used in earthquake prediction is available in the data set of this study, as demonstrated in the heatmap in Figure 1. It is evident from the Figure 2 that by comparing the automatic earthquake data distribution against the earthquake data passed through analyst review, there is much variation.

C. Machine Learning Results

The results of each machine learning model designed to calculate the earthquake magnitude are shown in Table I below. For all models the average values of the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), R-squared (R²) statistical and others are indicated to have a comprehensive notion of model performance and credibility.



Fig. 1. Heatmap for Dataset Features

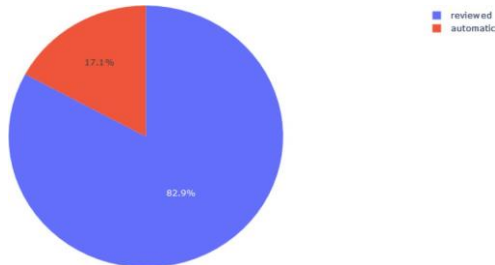


Fig. 2. Distribution of Reviewed vs. Automatic Earthquake Data

TABLE I : ML MODEL RESULTS FOR EARTHQUAKE MAGNITUDE PREDICTION

Model	MSE	RMSE	MAE	r	R ²	RRMSE	NSE
SVR	0.038	0.195	0.142	0.311	0.097	0.203	0.097
Linear Reg.	0.036	0.190	0.138	0.380	0.144	0.198	0.144
Decision Tree	0.019	0.139	0.094	0.735	0.540	0.145	0.540
Random Forest	0.014	0.119	0.077	0.813	0.661	0.125	0.661
KNN	0.011	0.107	0.065	0.854	0.730	0.111	0.730

Figure 3 shows the MSE of each model with Random Forest and KNN models having the lowest error rates. The MSE, therefore, depicts a clear picture of one model being more suitable than the other due to case with non-linear features within the data set.

4. CONCLUSION

Comparing performances and errors of several machine learning algorithm models employed in predicting the earth- quake magnitudes, the K- Nearest Neighbors and Random Forest models are more precise and less erroneous than other models in different configurations of performance metrics. It contains some sort of non-linearity within the seismic data

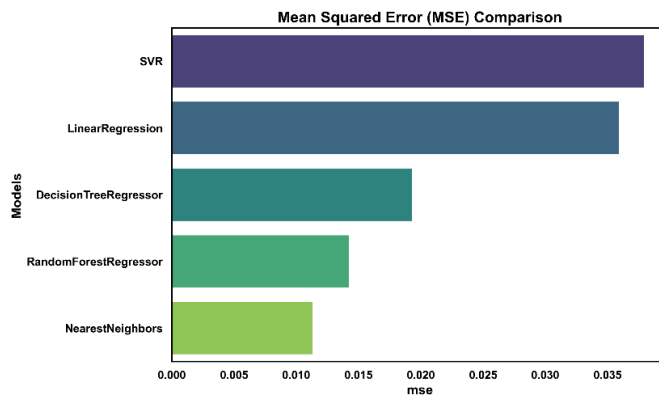


Fig. 3. Mean Squared Error (MSE) Comparison

REFERENCES

- [1] El-Sayed M El-Kenawy, Abdelhameed Ibrahim, Amel Ali Alhussan, Doaa Sami Khafaga, Ayman EM Ahmed, and Marwa M Eid. Smart city electricity load forecasting using greylag goose optimization-enhanced time series analysis. *Arabian Journal for Science and Engineering*, pages 1–19, 2025.
- [2] El-Sayed M Elkenawy, Amel Ali Alhussan, Doaa Sami Khafaga, Zahraa Tarek, and Ahmed M Elshewey. Greylag goose optimization and multilayer perceptron for enhancing lung cancer classification. *Scientific Reports*, 14(1):23784, 2024.
- [3] R. Kumar, Vevekanandam, and R. Dey, “IoT-enabled bioimpedance and thermal imaging system for non-destructive fruit maturity assessment,” *International Journal of Vegetable Science*, pp. 1–23, 2025, doi: 10.1080/19315260.2025.2574898..
- [4] El-Sayed M Elkenawy, Amel Ali Alhussan, Marwa M Eid, and Abdelhameed Ibrahim. Rainfall classification and forecasting based on a novel voting adaptive dynamic optimization algorithm. *Frontiers in Environmental Science*, 12:1417664, 2024.
- [5] Kamal, M. K. Singla, R. Kumar et al., “HGOA-based framework for multi-objective optimization and performance prediction of PEM fuel cells,” *Scientific Reports*, vol. 15, p. 35925, 2025, doi: 10.1038/s41598-025-19852-4.