Earthquake Risk-Level Classification Using Machine Learning and Somersaulting Spider Optimizer

El-Sayed M. El-kenawy[1,2]

[1]*Delta Higher Institute of Engineering and Technology Department for Communications and Electronic, Mansoura 35511, Egypt*

[2]*Applied Science Research Center Applied Science Private University Amman, Jordan, skenawy@ieee.org*

Abstract.

This study presents an innovative integrated framework that combines advanced machine learning classifiers with meta- heuristic optimization algorithms to enhance earthquake prediction capabilities. A comprehensive evaluation was conducted using five state-of-the-art models, including Random Forest, XGBoost, Gradient Boosting, Extra Trees, and Light GBM, on preprocessed seismic datasets. Random Forest emerged as the superior baseline classifier, achieving exceptional performance with 92.3% accuracy, 92.31% sensitivity, and 97.44% specificity, demonstrating balanced predictive capability across critical eval- uation metrics. To further optimize classification performance, three bio-inspired metaheuristic algorithms—Somersaulting Spi- der Optimizer (SSO), Harris Hawks Optimization (HHO), and Whale Optimization Algorithm (WOA)—were systematically in- tegrated with the Random Forest model. A comparative analysis revealed that SSO-Random Forest achieved the highest overall performance, with an accuracy of 93.46% and a specificity of 97.82%, significantly outperforming both the baseline and other optimized configurations. Statistical validation, as deter- mined by one-way ANOVA, confirmed the significance of these improvements.

Keywords. Earthquake risk-level classification, Machine learning, Somersaulting Spider Optimizer (SSO), Random Forest, seismic analysis, metaheuristic optimization

1. INTRODUCTION

Earthquakes constitute catastrophic natural disasters causing significant loss of life and infrastructure damage. Traditional seismic monitoring systems face substantial limitations in speed, generalization, and predictive accuracy, while early warning delivery remains constrained by earthquake com- plexity and limited seismic data. Integrating advanced computational techniques has become urgent for strengthening early warning systems. Machine learning (ML) and Deep learning (DL) approaches have emerged as powerful tools for improving earthquake prediction performance [1].

Unlike conventional threshold-based methods, these data- driven techniques are capable of automatically extracting relevant features, modeling complex nonlinear relationships, and making rapid predictions from raw seismic inputs [2]. Numerous studies have explored convolutional neural net- works, recurrent architectures, and hybrid models for tasks such as earthquake detection, magnitude estimation, and peak ground acceleration prediction [3]. However, despite promis- ing advances, these approaches face critical challenges related to generalization across diverse seismic regions, sensitivity to noise, and the scarcity of labeled training data in low- seismicity zones [4]. Addressing these limitations requires frameworks that combine the strengths of machine learning with novel strategies for optimization and statistical validation to ensure reliability and adaptability in real-world scenarios [5].

2. LITERATURE REVIEW

Deep learning is increasingly applied to seismic waveform analysis for earthquake monitoring, but models often fail to generalize across diverse regions due to geographically limited training. As shown in [6], a data recombination method generates synthetic seismic data from various locations and stations, enabling neural networks with broader applicability for detection and parameter estimation.

Optical mesh networks provide another approach for early earthquake detection by using existing fiber infrastructure. According to [7], integrating real displacement data into waveplate models allows strain simulations in fiber cables, improving machine learning models for P-wave detection and showing the potential of fiber optic grids as seismic sensors. Accurate peak ground acceleration (PGA) prediction is vital for early warning. In [8], a CNN-based end-to-end model pre- dicts PGA directly from initial seismic waves, achieving better accuracy and lower error than peak displacement methods. Data scarcity remains a challenge in low-seismicity areas. Research in [9] applies an auxiliary classifier GAN to gen- erate synthetic earthquake data, with results showing realistic seismic characteristics validated against borehole sensors.

3. EXPERIMENTAL RESULTS

A. Machine Learning Performance Evaluation

Table I presents the comprehensive performance comparison of five classification models evaluated on the earthquake pre- diction dataset. Among the evaluated models, Random Forest demonstrates superior performance across all key metrics, achieving the highest accuracy of 0.9231, sensitivity of 0.9231, and specificity of 0.9744. Additionally, it recorded a positive predictive value of 0.9292 and negative predictive value of 0.9746. Figure 1 illustrates the average error distribution for several models through comprehensive boxplot analysis with swarm overlay visualization. Random Forest exhibits exceptionally strong performance characteristics, achieving a relatively low median error value that suggests consistently accurate predic- tions.

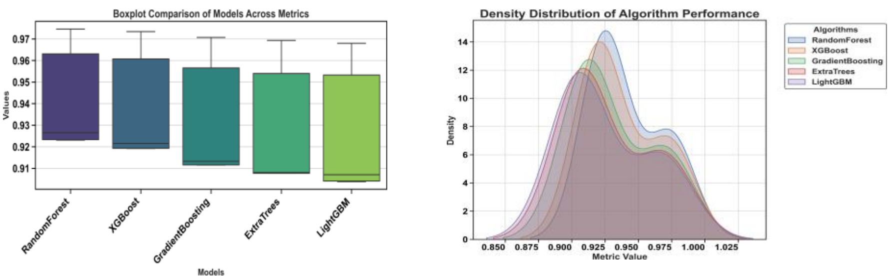

Fig. 1. Model Performance Comparison Using Boxplots with Swarm Overlay and Machine Learning Algorithm

TABLE I: CLASSIFIER PERFORMANCE COMPARISON TABLE

Models	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F1 Score
Random Forest	0.9231	0.9231	0.9744	0.9292	0.9746	0.9239
XGBoost	0.9192	0.9192	0.9731	0.9237	0.9734	0.9195
Gradient Boosting	0.9115	0.9115	0.9705	0.9147	0.9707	0.9118
Extra Trees	0.9077	0.9077	0.9692	0.9084	0.9693	0.9077
Light GBM	0.9038	0.9038	0.9679	0.9092	0.9680	0.9049

4. Optimization Algorithm Enhancement Results

Table I presents a comprehensive comparative analy- sis of three different metaheuristic optimization algorithms when combined with the Random Forest classification model. Among the three evaluated optimization strategies, SSO- Random Forest achieves the highest overall performance with an accuracy of 0.9346, representing a significant improve- ment over the baseline Random Forest model. The model also demonstrates exceptional specificity of 0.9782, indicating superior capability in correctly identifying negative earthquake instances.

Figure 2 displays comprehensive comparative model eval- uation through detailed boxplot analysis with swarm overlay, specifically focusing on average error distribution across the optimized models. The SSO-Random Forest model demon- strates exceptionally clustered error distribution characteristics, signifying remarkably consistent and reliable performance across multiple evaluation runs.

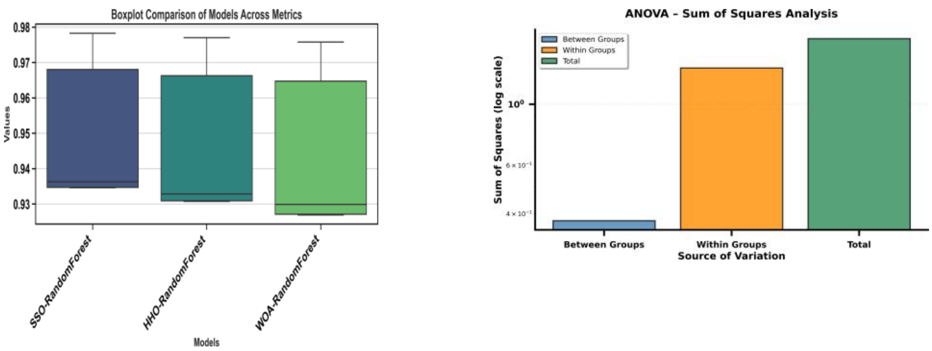

Fig. 2. Comparative Model Performance: Boxplot Analysis with Swarm Overlay for Optimized Models and ANOVA Sum of Squares Contribution Analysis

C. Statistical Significance Analysis

Table III displays comprehensive results of the one-way Analysis of Variance (ANOVA) conducted to evaluate sta- tistical significance of performance differences among the evaluated models. The Between Groups analysis reveals an F- statistic of 68.2295, which represents a substantial statistical measure indicating significant performance variations among the evaluated algorithms. The p-value of less than 0.0001 indicates extremely high statistical significance. Figure 2 illustrates comprehensive ANOVA results visualization on logarithmic scale, effectively highlighting the relative contributions of different variance sources to the overall statistical analysis.

5. CONCLUSION

The experimental results presented in this study demon strate the significant potential of integrating machine learning models with advanced metaheuristic optimization algorithms for earthquake prediction. This optimized model achieved remarkable improvements in accuracy and specificity, significantly reducing false positives while preserving strong detection capability. Statistical valida- tion through one-way ANOVA confirmed that these improve- ments were not incidental but rather statistically significant, reinforcing the reliability of the optimization-enhanced frame- work.

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] Y. K. Shejwal, A. J. Hati, J. U. Kidav, A. K. Singh, and R. Kumar, “Neural network-based classification of beamforming matrices,” International Journal of Satellite Communications and Networking, 2025, doi: 10.1002/sat.70002..

[3] Ahmed El-Sayed Saqr, Mohamed S Saraya, and El-Sayed M El- Kenawy. Enhancing co2 emissions prediction for electric vehicles using greylag goose optimization and machine learning. Scientific Reports, 15(1):16612, 2025.

[4] El-Sayed M Elkenawy, Amel Ali Alhussan, Marwa M Eid, and Ab- delhameed Ibrahim. Rainfall classification and forecasting based on a novel voting adaptive dynamic optimization algorithm. Frontiers in Environmental Science, 12:1417664, 2024.

[5] R. Kumar, D. Singh, M. K. Singla, E. S. M. El-Kenawy, and A. H. Alharbi, “Optimizing the placement of distributed renewable energy resources in large-scale distribution systems,” Journal of Electrical Engineering & Technology, pp. 1–25, 2025, doi: 10.1007/s42835-025-02279-2.

[6] 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.

[7] Hasan Awad, Fehmida Usmani, Emanuele Virgillito, Rudi Bratovich, Roberto Proietti, Stefano Straullu, Francesco Aquilino, Rosanna Pas- torelli, and Vittorio Curri. Environmental surveillance through machine learning-empowered utilization of optical networks. Sensors (Basel, Switzerland), 2024.

[8] R. Kumar, A. B. Queyam, M. K. Singla, M. Louzazni, and D. K. Mishra, “Optimizing efficiency and performance in a Rankine cycle power plant analysis,” Energy Engineering, Mar. 2025, doi: 10.32604/ee.2025.058058.

[9] Jae-Kwang Ahn, Byeonghak Kim, Bonhwa Ku, and Eui-Hong Hwang. Virtual scenarios of earthquake early warning to disaster management in smart cities based on auxiliary classifier generative adversarial networks. Sensors (Basel, Switzerland), 2023.