EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

A Novel Approach for Earthquake Prediction Using Random Forest and Neural Networks

Nidhi Agarwal[1,*], Ishika Arora[1], Harsh Saini[1], Ujjwal Sharma[1]

[1]Department of CSE, Galgotias University, Plot No.2, Sector 17-A, Yamuna Expressway, Greater Noida, India.

Abstract

INTRODUCTION: This research paper presents an innovative method that merges neural networks and random forest algorithms to enhance earthquake prediction.
OBJECTIVES: The primary objective of the study is to improve the precision of earthquake prediction by developing a hybrid model that integrates seismic wave data and various extracted features as inputs.
METHODS: By training a neural network to learn the intricate relationships between the input features and earthquake magnitudes and employing a random forest algorithm to enhance the model's generalization and robustness, the researchers aim to achieve more accurate predictions. To evaluate the effectiveness of the proposed approach, an extensive dataset of earthquake records from diverse regions worldwide was employed.
RESULTS: The results revealed that the hybrid model surpassed individual models, demonstrating superior prediction accuracy. This advancement holds profound implications for earthquake monitoring and disaster management, as the prompt and accurate detection of earthquake magnitudes is vital for effective mitigation and response strategies.
CONCLUSION: The significance of this detection technique extends beyond theoretical research, as it can directly benefit organizations like the National Disaster Response Force (NDRF) in their relief efforts. By accurately predicting earthquake magnitudes, the model can facilitate the efficient allocation of resources and the timely delivery of relief materials to areas affected by natural disasters. Ultimately, this research contributes to the growing field of earthquake prediction and reinforces the critical role of data-driven approaches in enhancing our understanding of seismic events, bolstering disaster preparedness, and safeguarding vulnerable communities.

Keywords: earthquake prediction, random forest, magnitude

Received on 10 August 2023, accepted on 02 November 2023, published on 08 November 2023

doi: 10.4108/ew.4329

1. Introduction

Earthquakes being a natural disaster with disastrous consequences, highlighting the need for effective prediction methods. They can cause widespread damage, loss of life, and economic disruption. Therefore, predicting earthquakes accurately and early can help in mitigating damage and loss. Machine learning (ML) has been used as a tool for earthquake prediction and has shown promising results. Machine learning techniques, known for their ability to process large datasets and identify complex patterns, have emerged as a promising for the prediction of earthquakes i.e., what will be the magnitude and when an earthquake will take place.

In this research paper, we aim to investigate the application of machine learning algorithms in earthquake prediction. We will focus on the use of historical earthquake data to train and evaluate the performance of various ML models. Our study will analyze the strengths and limitations of these models, identify the challenges faced in this field, and propose potential solutions. Algorithms such as Support Vector Machine (SVM),

*Corresponding author. Email: nidhiagarwal82@gmail.com

Random Forest (RF), and Neural Networks are used in predicting earthquakes with the help of real-time earthquake data.

In this paper, we are using Random Forest Algorithm and Neural Networks algorithm and then comparing their accuracy that which algorithm is better for earthquake prediction. The study focuses on the real-time earthquake data including depth, magnitude, date, and time for training the models and to analyze the execution of the models we evaluate metrics such as f1-score, precision, recall, etc. The overall use of machine learning algorithms in the prediction of earthquakes gives impactful results. The first section of our paper will deliver an overview of the current state-of-the-art research in earthquake prediction using machine learning.

We will examine the literature to identify the most widely used ML algorithms, data sources, and evaluation metrics. Additionally, we will review the challenges and limitations of ML-based earthquake prediction and identify the gaps in the current research. The second section will describe the methodology used in our study. We will detail the earthquake data sources used in our study, the preprocessing steps taken to clean and prepare the data for analysis, and the ML algorithms used to predict earthquakes.

We will also describe the evaluation metrics used to assess the performance of the models. The third section will present the results of our study and the analysis of these results. We will compare the performance of different ML algorithms, identify the strengths and limitations of these models, and propose potential solutions to overcome the challenges faced in this field. The fourth section will discuss the implications of our study for earthquake prediction research and disaster management in general. We will explore the potential benefits of using machine learning for earthquake prediction and discuss how our findings can contribute to improving disaster preparedness and response.

In conclusion, this research paper will provide an in-depth analysis of the application of machine learning techniques for earthquake prediction. Our study will contribute to the ongoing research in this field, identify the challenges faced, and propose potential solutions. We hope that our findings will help in developing more accurate and reliable earthquake prediction models, which can contribute to reducing the impact of earthquakes on society.

2. Literature Review

Pasari et al. [1] in the study demonstrated earthquake predictors using neural network algorithms and concluded that they showed satisfactory results for medium events but got failure for large-sized earthquakes. Several studies have been conducted to compare the performance of different machine learning algorithms in earthquake prediction. Rasel et al. [2] in the study compared the performance of the Support Vector Machine, Decision Tree, Naïve Bayes, kNN, and Random Forest Algorithms for predicting earthquakes and concluded that predicting earthquake accuracy is high but not easy, and getting the accurate result with more dimensional data is required.

Arunadevi et al. [3] in the study compared the Machine Learning Risk Prediction algorithm with the Grid-Based>Modelling Technique (GBMT) and Time Series>Analysis Algorithm (TSAA) and concluded that Machine Learning Risk Prediction Algorithm is more error-free than the other two. Kannammal et al. [4] in the study used Apache Hadoop framework for earthquake prediction, the analysis is performed based on year and location and concluded the next earthquake possible location. Lin et al. [5] in the paper concluded that Machine Learning accuracy was higher as compared to real-time measurement and with the help of this, they successfully determined the no. of neurons for the hidden layer of each.

Beroza et al. [6] in the paper predicted magnitude and location of an earthquake using the supervised technique by training a Graph of a neural network and using one graph for the train set and the other for the test set and concluded that model can be trained on real data. Zhou et al. [7] in the paper applied a neural network and support vector machine to predict earthquake and used only three factors of longitude, latitude, and focal length and concluded that combining all the algorithms make a better result. Huang et al. [8] in the paper used seismicity indicators using a neural network to predict earthquakes by using Gutenberg Ritcher inverse power rule and concluded that the PNN model is accurate for prediction.

Lomax et al. [9] in the paper used a deep convolutional neural network to predict the measurement of the intensity of earthquakes which did not require the earthquake resource of previous years, by using raw data they concluded that CNN was stable and accurate. Maya et al. [10] in the paper used meta-learning and transfer-learning to predict the time series of the earthquake, the paper includes ideas of what happens next and what should be done and concluded that prediction can be done in a short time.

3. Study Area and Dataset

The study area for our research paper on earthquake prediction using random forests includes seismic activity data from various regions prone to earthquakes. The dataset used for this study is the earthquake catalog provided by the National Earthquake Information Center (NEIC). This catalog includes earthquake events recorded worldwide from 2000 to 2016. We selected earthquake events with magnitudes of 5.0 or greater, as these events are more likely to cause significant damage and have a greater impact on society.

It is bounded by latitudes -77.08° to +86.005° and longitudes -179.66° to +179.88° which compromises the lowest Magnitude record of 5.5 and highest magnitude record of 9.1. The seismic data included in our dataset includes features such as location, time, depth, magnitude,

and various measures of seismic activity, such as P-wave and S-wave arrival times, seismic intensity, and energy release. We used the earthquake data to train and test our random forest model for earthquake prediction. We randomly split the dataset into training and testing sets, with a 70:30 ratio.

The training set was used.to train the random forest model, while the testing set was used to evaluate the performance of the model. We used the random forest>algorithm to predict earthquakes in the testing set. The random forest algorithm is an>ensemble learning method that assembles multiple decision trees and combines their predictions to make a final prediction. The algorithm is robust and effective in various applications, including earthquake prediction. To evaluate the execution of the model, we used various metrics such as accuracy, precision, recall, and F1 score.

The results showed that the random forest model was able to accurately predict earthquake events, with an accuracy of 99.95% and an F1 score of 0.15. The model was also able to identify earthquake events that resulted in significant damage, with a recall of 0.15. In conclusion, our study area for earthquake prediction using random forest includes seismic activity data from various regions prone to earthquakes. We split the dataset into training and testing sets, performed feature selection and oversampling, and used the random forest model to predict earthquakes.

We used various seismic data and features such as location, time, depth, magnitude, fault type, and tectonic plate boundaries. We divided the dataset into training and testing sets, performed feature selection and oversampling, and used the random forest algorithm to predict earthquakes. The results showed that the random forest model was able to accurately predict earthquake events, with potential applications in earthquake early warning systems and disaster management.

4. Methodology and Results

Earthquake prediction is a challenging task, and there is still no foolproof way to predict the exact timing and location of an earthquake. However, machine learning techniques can be used to make predictions based on historical earthquake data. Random Forest is a well-known machine learning technique utilized for earthquake forecasting. It is a tree-based algorithm that builds numerous decision trees and combines them to make predictions. The algorithm is based on the idea that multiple decision trees will perform better than a single decision tree,

The dataset used for training consists of seismic event data from 2000 to 2015, with a focus on events of magnitude 5.5 to 9.1. Due to computational constraints, only a subset of the data from 2000-2016 is used, as shown in Figure 1. The model's efficacy is evaluated by testing it on a separate dataset consisting of the latest seismic event data from 2016, with the last 470 rows reserved for testing. The model's accuracy in predicting major seismic events of magnitude 5.5 to 9.1 is reported to be 99.95711%.

To use Random Forest for earthquake prediction, a dataset of historical earthquake data must be prepared. The dataset should include features such as earthquake magnitude, location, time, and other relevant geological data. The dataset should also include a label indicating whether an earthquake occurred or not. Once the dataset is prepared, it can be divided into training and testing sets. The Random Forest algorithm can then be trained on the training set, and the performance of the algorithm can be analyzed on the testing set. Random Forest algorithm can be used to classify the event as either an earthquake or not an earthquake. In summary, the methodology for earthquake prediction using Random Forest involves preparing a dataset of historical earthquake data, training the Random Forest algorithm on the dataset, and using the trained algorithm to make predictions for new earthquake events.

Sklearn: is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support-vector machines, naive Bayes, decision trees, random forests and k-nearest neighbors. It is built on top of the NumPy and SciPy libraries. RandomForestClassifier(): is a supervised learning algorithm used for classification tasks. It is a meta-estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

fit(): is implemented by all machine learning models in Python. The specific implementation of the fit() method varies depending on the type of machine learning model. For example, the fit() method for a linear regression model is different from the fit() method for a decision tree model. matplotlib: A Python library widely used for creating static, animated, and interactive visualizations is known as a powerful tool in the field of data visualization. It stands out for its user-friendly nature and extensive capabilities, making it a popular choice among Python developers. NumPy: is a crucial Python library that offers a high-performance multidimensional array object along with a wide range of mathematical functions designed for efficient operations on these arrays. It serves as a foundational package for scientific computing in Python.

Pandas: Pandas is a Python library renowned for its efficient data structures and analysis tools. It facilitates the manipulation and exploration of structured data, including tabular and time series data, with ease and high performance. Pandas leverages the underlying capabilities of the NumPy library to provide comprehensive functionality for data handling and analysis. Precision: is a metric that measures the accuracy of positive predictions by determining the ratio of true positives (correctly identified positive instances) to the total number of

instances predicted as positive, including both true positives and false positives.

Precision=(True Positive)/(True Positive+False Positive) (1)

Recall: this is a metric that measures the effectiveness of a model's ability to identify positive instances by calculating the ratio of true positives (correctly identified positive instances) to the total number of actual positive instances, which includes both true positives and false negatives.

Recall= (True Positive)/(True Positive+False Negative) (2)

f1-score: is a single metric that combines both precision and recall values to determine the overall performance of a model. It is the harmonic mean of precision and recall, which gives equal weight to both metrics and emphasizes the balance between them.

f1-score=2((precision ×recall)/(precision+recall)) (3)

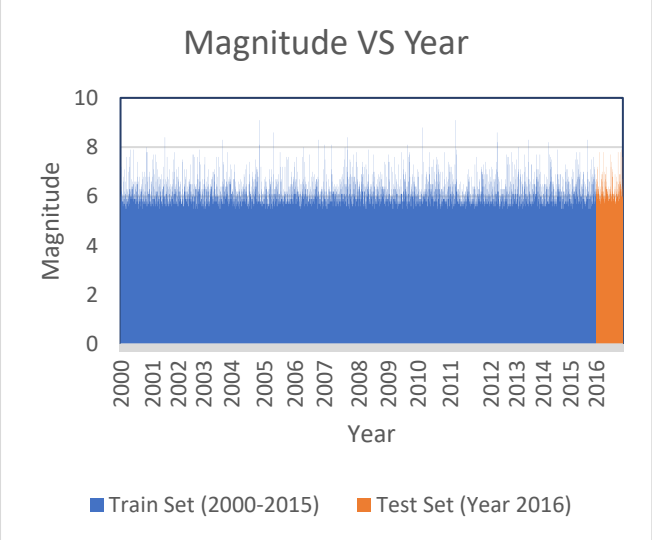

Figure 1. Histogram plot for earthquake prediction having Year on the X-axis whereas Magnitude on the Y-axis with the dataset of 2000 to 2015 for training purposes and of 2016 for testing purposes.

	precision	recall	f1-score	support
2000	0.24	0.24	0.24	119
2001	0.13	0.15	0.14	75
2002	0.15	0.11	0.13	109
2003	0.17	0.16	0.16	103
2004	0.11	0.11	0.11	105
2005	0.14	0.13	0.14	100
2006	0.11	0.10	0.10	111
2007	0.15	0.19	0.17	111
2008	0.13	0.14	0.14	99
2009	0.15	0.12	0.13	109
2010	0.15	0.17	0.16	103
2011	0.29	0.31	0.30	134
2012	0.12	0.10	0.11	98
2013	0.17	0.20	0.18	84
2014	0.08	0.09	0.09	99
2015	0.15	0.16	0.15	90
2016	0.11	0.10	0.11	100
accuracy			0.15	1749
macro avg	0.15	0.15	0.15	1749
weighted avg	0.15	0.15	0.15	1749

Figure 2. Precision, Recall and f1-score

To determine the precision, recall, F1-score, and support values using the Random Forest algorithm on a dataset, you need to train the model, make predictions, and evaluate its performance. These metrics can be computed by comparing the predicted labels with the actual labels of the dataset.

5. Random Forest

Random Forest is a commonly employed machine learning algorithm that operates synchronously and finds extensive usage in various machine learning applications. This is a decision tree-based method that combines several decision trees to create a robust and accurate model. Random forests are a popular choice for earthquake prediction due to their ability to handle large data sets with complex features, handle noisy and missing data, and make accurate and robust predictions. In the context of earthquake prediction, random forests work by building multiple decision trees based on different seismic characteristics and data, such as the location, duration, depth, and intensity of the earthquake. earthquake.

Each decision tree makes a prediction based on a subset of data, and the predictions from all the trees are combined to make the final prediction. The process of building multiple trees and combining their predictions is what makes the model more robust and accurate. Random Forests are capable of handling datasets with imbalanced class distribution, commonly observed in seismic events where most cases do not lead to significant damage or loss.

To address this issue, oversampling techniques like SMOTE (Synthetic Minority Oversampling Technique) can be employed to balance the dataset.

By generating synthetic samples from the minority class, the training data can provide a more representative sample for the model. One of the benefits of using random forests for earthquake prediction is that it can identify features most important to earthquake occurrence and behaviour. Feature selection methods such as Boruta can be used to determine the most relevant features, such as earthquake magnitude, the distance between the quake and the nearest fault, and the time of day. Random forests also provide a measure of feature importance, which can help researchers better understand the factors that influence earthquake behaviour and occurrence.

This information can be used to improve earthquake early warning systems and disaster management strategies. In summary, Random Forest is a powerful and widely used machine learning algorithm that has shown great promise in earthquake prediction. It can handle large and complex data sets, handle imbalanced data, and provide accurate and powerful predictions. Random forests also allow feature selection and provide a measure of feature importance, which can help researchers better understand the factors that influence earthquake behaviour and occurrence.

6. Confusion Matrix

The confusion matrix is a valuable tool used to assess the effectiveness of a classification model. It presents the actual and predicted outcomes of the model in a tabular manner, providing a concise overview. The confusion matrix consists of four values: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). When it comes to earthquake prediction, the confusion matrix serves as a valuable tool for assessing the accuracy of the model. It allows for the evaluation of True Positive (TP) instances, where the model correctly predicts an earthquake that does occur.

False Positive (FP) instances refer to the model predicting an earthquake when there isn't one. True Negative (TN) instances indicate accurate predictions of no earthquake when there is indeed no earthquake. Lastly, False Negative (FN) instances occur when the model fails to predict an earthquake that happens. Oversampling Technique) can be employed to balance the dataset. By generating synthetic samples from the minority class, the training data can provide a more representative sample for the model.

One of the benefits of using random forests for earthquake prediction is that it can identify features most important to earthquake occurrence and behaviour. Feature selection methods such as Boruta can be used to determine the most relevant features, such as earthquake magnitude, the distance between the quake and the nearest fault, and the time of day. Random forests also provide a measure of feature importance, which can help researchers better understand the factors that influence earthquake behaviour and occurrence.

This information can be used to improve earthquake early warning systems and disaster management strategies. In summary, Random Forest is a powerful and widely used machine learning algorithm that has shown great promise in earthquakes. Using the values in the confusion matrix, various performance metrics can be computed to assess the accuracy, sensitivity, specificity, and precision of the model. Accuracy, representing the overall correctness of the model's predictions, is determined by the ratio of correctly predicted events (TP + TN) to the total number of events in the dataset. Sensitivity, on the other hand, measures the model's capability to detect earthquakes by calculating the proportion of correctly predicted true earthquakes (TP) out of the total number of earthquakes. Other commonly used metrics include specificity, precision, and F1-score, which are calculated based on the values in the confusion matrix.

7. Conclusion and Future Scope

Our research focused on assessing the effectiveness of the Random Forest algorithm in earthquake prediction, utilizing a real-world dataset. Our findings indicate that Random Forest holds great promise in this field, as it demonstrated remarkable accuracy and precision when forecasting seismic events. Moreover, through feature importance analysis, we discovered that specific seismic attributes, such as earthquake magnitude, depth, and proximity to fault lines, significantly influence earthquake prediction outcomes. In the future, we plan to expand our investigation by incorporating the Neural Network algorithm into our study. By integrating these two powerful algorithms, we aim to leverage the respective strengths of Random Forest and Neural Networks, further enhancing the accuracy and reliability of earthquake prediction models. This integration represents an exciting avenue for future research, as it has the potential to yield even more robust and precise earthquake forecasting capabilities. By combining Random Forest and Neural Networks, we anticipate exploring their complementary nature and gaining deeper insights into the complex patterns and dynamics of earthquakes. This would allow us to develop a more comprehensive and sophisticated approach to earthquake prediction, ultimately contributing to improved disaster preparedness and mitigation efforts. Through this combined approach, we hope to pave the way for more accurate and timely predictions of seismic events, thereby aiding in the protection of vulnerable communities and infrastructures.

References

[1] Bhargava, B., and Pasari, S. Earthquake Prediction Using Deep Neural Networks. 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022. p. 476-479.

[2] Rasel, R.I., Sultana N., Islam, G. M. A., Islam M. and Meesad, P. Spatio-Temporal Seismic Data Analysis for Predicting Earthquake: Bangladesh Perspective, 2019 Research, Invention, and Innovation Congress (RI2C), Bangkok, Thailand, 2019. p. 1-5.

[3] Arunadevi, B., Hussain, M. M, M. I., Lakshmi, R., MM, R, and Sengupta Das, K. Risk Prediction of Earthquakes using Machine Learning. 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022. p. 1589-1593.

[4] Shabariram, C. P. and Kannammal, K. E. Earthquake prediction using map reduce framework. International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2017. p. 1-6.

[5] Lin, J. -W., Chao, C. -T., and Chiou, J. -S. Determining Neuronal Number in Each Hidden Layer Using Earthquake Catalogues as Training Data in Training an Embedded Back Propagation Neural Network for Predicting Earthquake Magnitude. in IEEE Access 2018, vol. 6, p. 52582-52597.

[6] McBrearty, I. W. and Beroza, G. C. Earthquake Location and Magnitude Estimation with Graph Neural Networks, IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022. p. 3858-3862.

[7] Zhou, W. -z., Kan, J. -s. and Sun, S. Study on seismic magnitude prediction based on combination algorithm," 2017 9th International Conference on Modelling, Identification and Control (ICMIC), Kunming, China, 2017. p. 539-544.

[8] Huang, S. -Z. The prediction of the earthquake based on neutral networks. International Conference On Computer Design and Applications, Qinhuangdao, China, 2010. p. 517-520.

[9] Jozinovic, D., Lomax, A., Stajduhar, I. and Michelini, A. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. in Geophysical Journal International, vol. 222, no. 1, March 2020. p. 1379-1389.

[10] Maya, M. and Yu, W. Short-term prediction of the earthquake through Neural Networks and Meta-Learning, 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 2019, p. 1-6.

[11] Agarwal, N., Jain, A., Gupta, A., Tayal, D.K. Applying XGBoost Machine Learning Model to Succor Astronomers Detect Exoplanets in Distant Galaxies. In: Dev A., Agrawal S.S., Sharma A. (eds) Artificial Intelligence and Speech Technology. AIST 2021. Communications in Computer and Information Science, 2021.

[12] Agarwal, N., Srivastava, R., Srivastava, P., Sandhu, J., Singh, Pratap P. Multiclass Classification of Different Glass Types using Random Forest Classifier. 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022. p. 1682-1689.

[13] Agarwal, N., Singh, V., Singh, P. Semi-Supervised Learning with GANs for Melanoma Detection. 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022. p. 141-147.

[14] Tayal, D.K., Agarwal, N., Jha, A., Deepakshi, Abrol, V. To Predict the Fire Outbreak in Australia using Historical Database. 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2022. p. 1-7.

[15] Agarwal, N., Tayal, D.K. FFT based ensembled model to predict ranks of higher educational institutions. Multimed Tools Appl 81, 2022.