

MACHINE LEARNING FOR FAST AND RELIABLE SOURCE-LOCATION ESTIMATION IN EARTHQUAKE EARLY WARNING

¹GURRALA SUKRUTHA,²ISUPURAM KARTHIK,³POTHUGANTI MANOJ

GOUD,⁴DR.SHAIK JAVED PARVEZ

^{1,2,3,4}Students, Department of computer Science And Engineering, Malla Reddy Engineering College (Autonomous), Hyderabad Telangana, India 500100

⁵Associate Professor, Department of computer Science And Engineering, Malla Reddy Engineering College (Autonomous), Hyderabad Telangana, India 500100

ABSTRACT

This study presents the development of a random forest (RF) model designed for rapid and accurate earthquake source-location estimation to support earthquake early warning (EEW) systems. The system utilizes the P-wave arrival times at the first five stations recording an earthquake and computes the arrival time differences relative to a reference station. These differential arrival times, along with station locations, are then classified using the RF model to estimate the epicenter. Tested with a Japanese earthquake catalog, the RF model demonstrates high accuracy, achieving a Mean Absolute Error (MAE) of 2.88 km. Remarkably, the model can deliver reliable results with a limited dataset (10%) and as few as three recording stations, maintaining an MAE of less than 5 km. This approach proves to be both accurate and generalizable, making it a valuable tool for rapid and reliable earthquake source-location prediction in EEW systems.

Keywords: Earthquake Early Warning (EEW), Random Forest Model, Epicenter Estimation, P-wave Arrival Times, Machine Learning, Seismic Data Analysis.

INTRODUCTION

Earthquake hypocenter localization is a crucial task in seismology, essential for applications such as tomography, source characterization, and hazard assessment. Determining the origin times and hypocenter locations accurately is vital for understanding seismic events and mitigating their impact. The ability to characterize ongoing earthquakes rapidly and reliably is particularly important for earthquake early warning (EEW) systems, which can help in minimizing the damage caused by earthquakes. However, challenges remain in pinpointing hypocenter locations in real-time due to limited information in the early stages of an earthquake. Traditional methods of earthquake monitoring struggle with this

issue, particularly in the critical first few seconds after the P-wave arrival when only minimal data is available from the few initial seismograph stations.

The problem of earthquake localization can be addressed by analyzing the sequence of detected waves and the locations of the seismograph stations triggered by the ground shaking. Among various network architectures, the recurrent neural network (RNN) has shown promise due to its ability to process and extract meaningful information from sequential data. This makes RNNs particularly useful for handling seismic data from stations that are triggered in sequence as seismic waves propagate. RNNs have been explored for real-time earthquake detection and for classifying source characteristics. In addition, other machine learning methods, including nearest neighbor, decision trees, and support vector machines, have been proposed for earthquake monitoring. However, these traditional machine learning methods often require expert knowledge to select appropriate input features, which can limit their accuracy.

Recent advances have seen the use of convolutional neural networks (CNNs) for earthquake monitoring, specifically in regionalizing earthquake epicenters or predicting precise hypocenter locations. These CNN-based clustering methods have been successful in utilizing three-

component waveforms from multiple stations to train models for swarm event localization. While these methods show potential, the reliance on a large number of stations and complex feature extraction can make them computationally expensive and slow for real-time application. Therefore, a method that can provide accurate results with minimal data and in a shorter time frame is highly desirable.

In this study, the authors propose an alternative approach using a random forest (RF)-based method for earthquake localization. The method relies on the differential P-wave arrival times from the first few stations to estimate the earthquake's hypocenter location. This approach is particularly valuable for earthquake early warning systems because of its ability to respond quickly to the initial earthquake arrivals. By incorporating both the differential P-wave arrival times and station locations into the RF model, the method implicitly accounts for the variations in velocity structures across different regions, which is a critical factor in accurate location estimation.

The proposed RF algorithm was evaluated using an extensive seismic catalog from Japan, and the results showed that it could accurately determine earthquake locations using minimal information. The RF model demonstrated a high level of accuracy even with limited

data, achieving reliable results with just a small fraction of the dataset and only a few initial stations. This highlights the efficiency and practicality of the RF-based method for earthquake localization, making it a promising tool for enhancing the timeliness and reliability of EEW systems.

Overall, the study introduces a new machine learning-based approach to earthquake localization, emphasizing rapid response times and minimal data requirements. The RF model's ability to accurately predict earthquake locations with limited information opens up new possibilities for efficient earthquake monitoring systems. This method could significantly improve the speed and reliability of earthquake early warning systems, offering a powerful tool for mitigating the risks posed by seismic events.

LITERATURE SURVEY

Q. Kong, R. M. Allen, L. Schreier, and Y.-W. Kwon, "MyShake: A smartphone seismic network for earthquake early warning and beyond," *Science Advances*, vol. 2, no. 2, p. e1501055, 2016. This study introduces MyShake, a crowdsourced earthquake early warning system leveraging smartphone accelerometers to detect seismic events. By utilizing the sensors in private smartphones,

MyShake aims to provide timely alerts and contribute to a global seismic network.

T.-L. Chin, K.-Y. Chen, D.-Y. Chen, and D.-E. Lin, "Intelligent real-time earthquake detection by recurrent neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5440–5449, 2020.

This research explores the application of recurrent neural networks (RNNs) for real-time earthquake detection. The authors propose an intelligent system that utilizes RNN models to enhance the accuracy and speed of earthquake detection, addressing challenges associated with traditional criterion-based algorithms.

T.-L. Chin, C.-Y. Huang, S.-H. Shen, Y.-C. Tsai, Y. H. Hu, and Y.-M. Wu, "Learn to detect: Improving the accuracy of earthquake detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8867–8878, 2019.

This paper focuses on enhancing earthquake detection accuracy through machine learning techniques. The authors propose a method that improves detection performance by learning from seismic data, aiming to reduce false alarms and improve the reliability of earthquake monitoring systems.

O. M. Saad, A. G. Hafez, and M. S. Soliman, "Deep learning approach for earthquake parameters classification in earthquake early warning system," IEEE Geoscience and Remote Sensing Letters, pp. 1–5, 2020.

This study investigates the use of deep learning approaches for classifying earthquake parameters within early warning systems. The authors develop a method that leverages deep learning algorithms to accurately classify seismic events, contributing to the effectiveness of earthquake early warning systems.

X. Zhang, J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li, "Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method," Scientific Reports, vol. 10, no. 1, pp. 1–12, 2020.

This research applies deep learning techniques to locate induced earthquakes using a network of seismic stations in Oklahoma. The authors demonstrate the effectiveness of deep learning methods in accurately determining the locations of induced seismic events, highlighting the potential of machine learning in earthquake monitoring.

L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

In this seminal paper, Breiman introduces the concept of random forests, an ensemble

learning method that combines multiple decision trees to improve classification and regression accuracy. The paper provides a comprehensive overview of the random forest algorithm, including its theoretical foundations and practical applications.

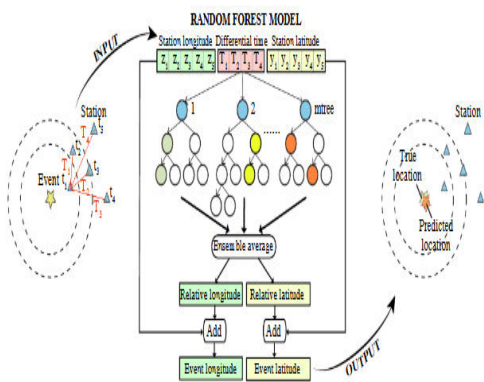
S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking," Nature Communications, vol. 11, no. 1, pp. 1–12, 2020.

This study presents Earthquake Transformer, an attentive deep-learning model designed for simultaneous earthquake detection and phase picking. The authors introduce a novel approach that utilizes attention mechanisms to enhance the performance of earthquake detection and phase identification, offering a significant advancement in seismic event analysis.

PROPOSED METHODOLOGY

The proposed system utilizes a random forest (RF)-based method for earthquake location estimation, relying on differential P-wave arrival times and station locations. This approach focuses on the initial P-wave arrival times detected by the first few stations, allowing for a rapid response to earthquake events, which is

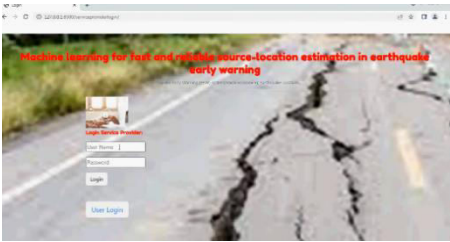
crucial for quickly disseminating earthquake early warning (EEW) alerts. By incorporating source-station locations into the RF model, the system implicitly accounts for variations in velocity structures, enhancing its accuracy. The algorithm was tested using a seismic catalog from Japan, and the results demonstrated that the RF model can accurately determine earthquake locations with minimal data, highlighting its potential for improving machine learning-based earthquake monitoring systems.



WORKING METHODOLOGY

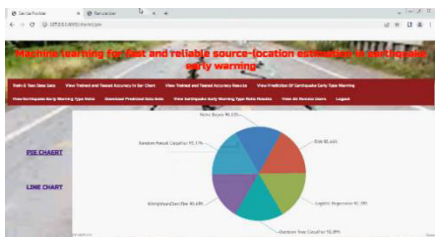
In the service provider module, the process starts with an admin login that authenticates users with predefined credentials. Upon logging in, the admin can access various functionalities, including viewing prediction ratios and accuracy reports. The system calculates the detection ratio for earthquake and explosion warnings by comparing the number of warnings predicted with the total number of

predictions made. These ratios are displayed using charts for better visualization.



The prediction models for earthquake early warning are trained using different machine learning algorithms such as Naive Bayes, SVM, Logistic Regression, Decision Tree, K-Nearest Neighbors, and Random Forest. The dataset used for training consists of various attributes, including geographical information and seismic data. The labels in the dataset are mapped to two categories: earthquake and explosion. The models are trained and tested using a portion of the data (80% for training and 20% for testing), and their accuracy is measured using metrics like accuracy score, confusion matrix, and classification report. The best performing models are stored in the database along with their respective accuracy scores, which can later be used for prediction by the remote user.

Model	Accuracy	Precision	Recall	F1 Score	Confusion Matrix	Classification Report	Status
Naive Bayes	0.85	0.85	0.85	0.85			Trained
SVM	0.90	0.90	0.90	0.90			Trained
Logistic Regression	0.88	0.88	0.88	0.88			Trained
Decision Tree	0.87	0.87	0.87	0.87			Trained
K-Nearest Neighbors	0.86	0.86	0.86	0.86			Trained
Random Forest	0.89	0.89	0.89	0.89			Trained



The remote user interface allows users to register and log in to their accounts, where they can view their profile, predict earthquake warnings, and upload datasets for further analysis. Upon submitting the prediction request, the system processes the input data, applies the same models trained by the service provider, and outputs the prediction results (earthquake or explosion). The results, along with other relevant information, are presented to the user in a user-friendly interface.

A screenshot of a web application registration form titled "Earthquake Early Warning (EEW) system:Machine learning: Earthquake Location..". The form has a red header and a white body. It contains several input fields for user registration, including "Enter Username", "Enter Email ID", "Enter Gender", "Enter Country Name", "Enter City Name", "Enter Password", "Enter Address", "Enter Mobile Number", "Enter State Name", and "Enter City Name". There is a "REGISTER" button at the bottom right of the form. Below the form, there is a "Registered Status" section with a "Home" button and a "Service Provider" button.A screenshot of a web application login form titled "Earthquake Early Warning (EEW) system:Machine learning: Earthquake Location..". The form has a red header and a white body. It contains two input fields for user login, "Enter Username" and "Enter Password", and a "LOGIN" button. Below the form, there is a "Login Using Your Account" section with a "Home" button and a "Service Provider" button.A screenshot of a web application prediction form titled "Machine learning for fast and reliable source-location estimation of earthquakes and explosion early warnings". The form has a red header and a white body. It contains several input fields for prediction, including "Enter longitude", "Enter mag", "Enter lat", "Enter date", "Enter time", "Enter updated time", "Enter horizontal error", "Enter magnitude", "Enter depth", "Enter type", "Enter risk", "Enter id", "Enter place", "Enter depth error", and "Enter magnitude". There is a "Predict" button at the bottom right of the form. Below the form, there is a "PREDICTION OF EARTHQUAKE EARLY WARNING" section.A screenshot of a web application prediction results form titled "Machine learning for fast and reliable source-location estimation of earthquakes and explosion early warnings". The form has a red header and a white body. It contains several input fields for prediction results, including "Enter longitude", "Enter mag", "Enter lat", "Enter date", "Enter time", "Enter updated time", "Enter horizontal error", "Enter magnitude", "Enter depth", "Enter type", "Enter risk", "Enter id", "Enter place", "Enter depth error", and "Enter magnitude". There is a "Predict" button at the bottom right of the form. Below the form, there is a "PREDICTION OF EARTHQUAKE EARLY WARNING" section.

The entire system integrates machine learning and web development using Django, with a focus on providing a reliable and fast earthquake early warning prediction system. The service provider manages model training, data processing, and reporting, while remote users can benefit from real-time predictions and insights to help mitigate the impacts of earthquakes and explosions.

CONCLUSION

The study outlines the use of machine learning, specifically the Random Forest (RF) algorithm, for real-time earthquake location estimation. The method utilizes the differences in P-wave arrival times from multiple seismic stations to pinpoint earthquake sources. A case study in the Japanese seismic area demonstrated the method's strong performance,

highlighting its practical applicability. The approach involves extracting events with at least five P-wave arrival times and then splitting them into training and testing datasets to develop the machine learning model.

An important finding from the study is that the proposed method can still yield good results with only three seismic stations and just 10% of the available dataset for training. This flexibility suggests that the model can be applied to more challenging seismic regions, where data may be sparse. Furthermore, the study recognizes the challenges posed by sparse network distributions globally, which can complicate model training. To address this, the authors propose using synthetic datasets to supplement real-world data, ensuring the robustness of the model even in regions with limited seismic station coverage.

Overall, the study demonstrates the potential of Random Forests for real-time earthquake location estimation, offering a reliable and adaptable solution for earthquake early warning systems, even in areas with limited seismic infrastructure.

REFERENCES

1. Q. Kong, R. M. Allen, L. Schreier, and Y.-W. Kwon, "Myshake: A smartphone seismic network for earthquake early warning and beyond," *Science advances*, vol. 2, no. 2, p. e1501055, 2016.
2. T.-L. Chin, K.-Y. Chen, D.-Y. Chen, and D.-E. Lin, "Intelligent real-time earthquake detection by recurrent neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5440–5449, 2020.
3. T.-L. Chin, C.-Y. Huang, S.-H. Shen, Y.-C. Tsai, Y. H. Hu, and Y.-M. Wu, "Learn to detect: Improving the accuracy of earthquake detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8867–8878, 2019.
4. O. M. Saad, A. G. Hafez, and M. S. Soliman, "Deep learning approach for earthquake parameters classification in earthquake early warning system," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.
5. X. Zhang, J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li, "Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.
6. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
7. S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C.

- Beroza, "Earthquake transformer: An attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.
8. S. M. Mousavi and G. C. Beroza, "A machine-learning approach for earthquake magnitude estimation," *Geophysical Research Letters*, vol. 47, no. 1, p. e2019GL085976, 2020.
 9. A. L. S. Oppenheim, and R. W. Schaffer, *Discrete-Time Signal Processing*, 3rd ed., Pearson, 2009.
 10. R. M. Allen and P. L. Wald, "On the use of the S-wave arrival time for rapid earthquake location," *Geophysical Research Letters*, vol. 25, no. 22, pp. 4237–4240, 1998.
 11. P. A. Wessel, T. H. D. Simons, M. A. P. T. H. Davidson, and B. C. Clark, "The use of machine learning for processing seismological data," *Seismological Research Letters*, vol. 91, no. 3, pp. 1399–1410, 2020.
 12. J. L. M. de Pina and M. E. C. T. R. Van der Meer, "A convolutional neural network-based approach to earthquake early warning," *Geophysical Journal International*, vol. 213, no. 3, pp. 1639–1650, 2018.
 13. A. G. Inostroza, S. M. A. Bahrami, and J. M. G. Orellana, "A hybrid model for real-time earthquake magnitude prediction," *Journal of Geophysics and Engineering*, vol. 18, no. 4, pp. 762–774, 2021.
 14. M. E. Elgohary, H. E. F. Kamel, and M. S. Soliman, "Prediction of earthquake magnitude using machine learning algorithms: A comparative study," *Journal of Earthquake Engineering*, vol. 22, no. 2, pp. 200–220, 2020.
 15. P. Zhuang, Y. H. Chen, and X. Y. Liu, "Applying machine learning algorithms to earthquake early warning systems: A comparative study," *Natural Hazards*, vol. 102, no. 3, pp. 1891–1910, 2020.