International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Using Machine Learning to Estimate Source Location Early Earthquake Warning

Radha R.[*1], Prakash Babu Yandrapati[2], M. Abirami[3], G. Victo Sudha George[4], A. Joshi[5]

Submitted: 24/08/2023 **Revised**: 12/10/2023 **Accepted**: 25/10/2023

Abstract: Early warning systems for earthquakes can mitigate their destructive potential by spreading information about the quake's magnitude and location long before destructive waves reach populated areas. Source-location estimations in these systems need to be timely and accurate for them to be useful. This study presents a novel approach for enhancing the precision and speed of seismic early warning using machine learning techniques. Timely warnings may be delayed due to the precision but slowness of traditional seismic techniques for calculating earthquake sites. The purpose of the random forest (RF) model for fast earthquake localization is to aid in the quick decision making required by earthquake early warning (EEW) systems. This approach takes use of the P-wave arrival times recorded by the first five stations to record an earthquake and calculates the variations in these timings with regard to the first station. In order to determine the epicenter, the RF model categorizes these differences in Pwave arrival timings and station locations. The model is used to train and validate the proposed method using a Japanese earthquake dataset. The RF model is quite accurate in predicting earthquake epicenters, with a Mean Absolute Error (MAE) of just 2.88 kilometers. Additionally, the suggested RF model may learn from as little as 10% of the information and as little as three recording stations while still producing usable results (MAE5 km) in most cases. This novel algorithm provides a robust and flexible method for predicting the location of EEW sources in real-time.

Keywords: *Random Forest Model, Earthquake Early Warning, P-Wave Arrival Times, Epicentral Location Estimation, Mean Absolute Error, Rapid and Reliable Prediction*

1. Introduction

Seismologists rely heavily on earthquake hypocenter localization for tasks like tomography, source characterization, and hazard assessment, among many others. This highlights the need for reliable seismic monitoring systems that can pinpoint the exact moment an earthquake began and its epicenter. Seismic hazard reduction methods, such as earthquake early warning (EEW) systems, rely on accurate and timely characterization of active earthquakes. Despite the widespread use of classical approaches in EEW system design, identifying earthquake hypocenters in real time remains difficult, mostly owing to a lack of data available early on.

[1] *Assistant Professor, Department of Data Science and Business Systems, S R M Institute of science and technology, kattankulathur, Chennai – 603203, INDIA, radhar@srmist.edu.in, https://orcid.org/0000-0001-9822-0949*
[2] *Assistant Professor, Department of Computer Science and Engineering, GITAM University Hyderabad, Rudraram-502329, INDIA, pyandrap@gitam.edu, https://orcid.org/0000-0003-1934-9125*
[3] *Assistant Professor, Department of Computer Science and Engineering, Panimalar Engineering college , Chennai – 600123, INDIA, abiprj@gmail.com, https://orcid.org/0000-0003-4935-8938*
[4] *Professor, Department of Computer Science and Engineering, Dr.M.G.R Educational and Research Institute, Chennai – 600095, INDIA, victosudhageorge@drmgrdu.ac.in, https://orcid.org/0000-0002-8172-6721*
[5] *Professor, Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai – 600123, INDIA, joshiseeni@gmail.com,https://orcid.org/0000-0001-5141-2994*
** Corresponding Author Email: radhar@srmist.edu.in*

Timeliness is an important aspect of EEW, and more work needs to be done to improve hypocenter location estimates with minimal data from

1) The first few seconds after the P-wave arrival and

2) The first few seismograph stations that are triggered by the ground shaking.

A Support Vector Machine Regression (SVMR) approach calculates local magnitude (Ml) in five seconds following the P wave beginning of a three-component seismic station. The method was trained on 863 earthquake data, using exponential regression parameters based on the predicted waveform envelope and highest observed value for each component in a single station. The mean absolute error for a normalized polynomial kernel was calculated using ten-fold cross validation for various exponents and complexity settings. The local magnitude (Ml) may be approximated with a mean absolute error of 0.19 units [1].

Seismograph stations are activated by earth tremors, and their positions and the timing of the waves they detect may be used to solve the localization issue. When dealing with a network of seismic stations that are activated in succession as waves travel through the earth, a recurrent neural network (RNN) is the best option because of its ability to accurately extract information from a series of input data. Research into this strategy has been conducted with the goal of enhancing the effectiveness of earthquake detection and source-attribute categorization in real time. Additional machine learning-based seismic monitoring systems have been presented. The earthquake detection issue has also been used to compare and contrast some of the more classic machine learning techniques, such as the closest neighbor, decision tree, and support vector machine.

One potential flaw in the aforementioned machine learning based frameworks is that they often need on expert knowledge to choose input characteristics. Epicenters of earthquakes have been regionalized and their hypocenters predicted using clustering techniques based on convolutional neural networks. To train the model for swarm event localization, the latter instance makes use of three-component waveforms from numerous stations. Using differential P-wave arrival timings and station locations, we present an RF-based approach for earthquake localization in this work (Figure 1). The first few stations' P wave arrival timings are all that are used in the proposed method. In order to quickly disseminate EEW notifications, it must react quickly to first earthquake reports. By include the source-station coordinates in the RF model; this approach implicitly takes into account the impact of the velocity structures. Using a large-scale Japanese seismic database, researchers tested the suggested technique. The test findings reveal that the RF model can pinpoint earthquake epicenters with just a little amount of data, providing novel insight into the creation of effective machine learning.

2. Related Works

Earthquake Early Warning System (EEWS) gives information on the projected arrival time of S waves, which may deliver considerable and damaging seismic energy, utilizing P wave information. Technological advancements in big data, network connectivity, and high-performance computing have made earthquake early warning difficult to process using contemporary seismological methods in the 4.0 industrial revolution. Detecting earthquakes early is crucial for efficient information transmission. Deep learning is used to identify and classify earthquake P waves and noise signals in West Sumatra's subduction zone using historical data from the 3 component BMKG single station (2014-2020). Feature selection for the waveform is limited to earthquakes around the station centroid.

Training and testing outcomes are statistically consistent. This project aims to use deep learning to classify earthquake p-wave and noise signals and predict early earthquake location utilizing three component record channels [2]. Early warning (EEW) can lessen earthquake risk. Today, EEW is used to quickly classify earthquake magnitude, with big earthquakes that need warning in the positive category and vice versa in the negative category. Magnitude quick categorization using traditional information signal processing procedures is time-consuming and data imbalance-prone. This work introduces Deep Learning (DL) techniques for EEW. Using DenseBlock with Bottleneck and Multi-Head Attention, this research presents a DL model (EEWMagNet) to extract spatial and temporal characteristics from the China Earthquake Network Center (CENC) three-component seismic waveform record of 7 s. Extensive trials using Chinese field data show that the suggested model quickly classifies magnitude. Comparison trials show that epicenter distance information is essential and that normalization hinders the model's amplitude accuracy [3]. We investigated forecasting structural drift from the first seconds of P-wave data for On-site Earthquake Early Warning (EEW) applications. This study compared the performance of linear least square regression (LSR) against four non-linear machine learning models: Random Forest, Gradient Boosting, Support Vector Machines, and K-Nearest Neighbors. Furthermore, we investigate the transferability of calibrated models from one location to another. The LSR and ML models are calibrated and validated using a dataset of ~6,000 waveforms from 34 Japanese structures (steel, reinforced concrete, and steel-reinforced concrete) and a smaller dataset from 69 US buildings (240 data points). For EEW information, we used three P-wave parameters (Pd, IV2, and ID2) across three time-windows (1, 2, and 3 s) to forecast the drift ratio as a structural response. The Japanese dataset is used to calibrate and investigate the LSR and ML models' effectiveness in predicting structural drift. Our study examined several subsets of the Japanese dataset, including one building, one construction type, and the complete dataset. Variable ground motion and building response impact drift prediction robustness. For example, the accuracy of forecasts decreases with increasing dataset complexity in terms of building and event variability. ML approaches outperform LSR models owing to intricate feature linkages and data non-linearity.

To identify the primary drivers of drift variability, we demonstrate the use of residuals analysis. Finally, Japanese dataset models are applied to the US dataset. Exported EEW models increase forecast variability, although adding adjustment terms based on magnitude may significantly reduce this issue. We found that small model adjustments can forecast drift for US structures [4].

To determine the US West Coast ShakeAlert earthquake early warning (EEW) system's performance and limits, we test it during temporally near earthquake pairings. Our performance criteria include source parameter correctness, ground-motion prediction accuracy, and alerting timeliness. Ground-motion time series for synthetic earthquake sequences are created by integrating signals from well-recorded earthquakes ($4.4 \leq M \leq 7.1$) with time shifts from

−60 to +180 s. The study examines fore- and aftershock sequences, near-simultaneous occurrences, and simulated offshore and out-of-network earthquakes. The ShakeAlert algorithms EPIC, FinDer, and PLUM operate mostly as intended. EPIC offers the fastest source location estimates but may underestimate magnitudes or miss large earthquakes. FinDer offers real-time line-source models and unsaturated magnitude estimates for large earthquakes, but cannot process concurrent events and may mislocate offshore earthquakes. PLUM identifies strong ground motion but may overestimate alert areas. Space and time close events are hard to distinguish, challenging scenarios with close foreshocks can lead to missed alerts for large earthquakes, and algorithms can often estimate ground motion better than source parameters. To enhance EEW, we recommend reevaluating algorithm weighting in ShakeAlert, using ground-motion data to aggregate warnings from several algorithms, and optimizing algorithm ground-motion estimations. We recommend adding 25 of our 73 scenarios to the baseline data set for ShakeAlert and other EEW system testing and certification [5]. Using 3 seconds of P waves from a single station, the Ensemble Earthquake Early Warning System (E3WS) uses Machine Learning algorithms to identify, localize, and estimate earthquake magnitude. The system has 6 Ensemble Machine Learning algorithms trained on temporal, spectral, and cepstral ground acceleration time series properties. Peru, Chile, Japan, and STEAD are in the training set. Detection, P-phase picking, and source characterisation comprise E3WS. Depth, magnitude, epicentral distance, and back-azimuth are estimated. E3WS distinguishes earthquakes from noise with 99.9% accuracy, with no false positives and few false negatives. All false negatives are $M < 4.3$ earthquakes, considered unlikely to cause damage. The Mean Absolute Error for P-phase choosing is 0.14 s, suitable for earthquake early warning. The E3WS estimates are practically unbiased for source characteristics, better for magnitude estimation than single-station methods, and marginally better for earthquake location. The method provides earthquake source-time-dependent magnitude estimations by updating estimates every second. E3WS estimates quicker than multi-station warning systems, giving you seconds for precautionary measures [6].

Real-time earthquake magnitude and location estimations are crucial for early warning and reaction. Rapid earthquake assessment techniques based on deep learning recently proposed employ seismic data from a single station or a specified group of stations. Our attention-based transformer network model for real-time magnitude and position estimation is shown here. Our method surpasses deep learning baselines in magnitude and position estimation using waveforms from dynamically shifting stations. Compared to a traditional localization approach, it outperforms a classical magnitude estimation algorithm rather well. The probabilistic inference-based uncertainty estimates in our real-time prediction model are realistic. This research also examines training data needs, training methodologies, and common failure modes. Targeted experiments and qualitative error analysis are performed on three distinct and huge data sets. Several major findings come from our investigation. In particular, a four-fold bigger training set decreases magnitude and position prediction errors by more than half and real-time assessment time by four. Second, the fundamental model systematically underestimates major events. Adding events from other locations to the training via transfer learning may lessen or address this problem. Thirdly, location estimation is accurate in regions with enough training data but poor outside the training distribution, resulting in huge outliers. We found that most deep learning models for quick evaluation have similar traits with our model. They are caused by black box models and may need physics-based neural network limitations. Practical applications must address these traits [7].

Researchers and seismic networks in Europe are exploring novel earthquake early warning (EEW) methods, building and running test systems, and sometimes giving operational EEW to end customers. We discuss recent European EEW research, the networks and locations where EEW is being tested or developed, and the two systems in Turkey and Romania that offer operational systems to a restricted number of end users [8].

3. Proposed Methodology:

3.1 Data Collection:

Seismic Stations:

Seismic data is collected using seismometers or seismographs.

These sensors constitute a seismic station network at important places. Station location depends on the seismic monitoring program's aims and geographical region. Seismic stations are commonly located near fault lines, earthquake-prone locations, or seismic activity.

Recording Data:

In reaction to earthquake seismic waves, seismometers measure vertical, north-south, and east-west ground motion. These sensors gather analog or digital indications of ground displacement over time.

Telemetry and Data Transmission:

Modern seismic monitoring systems provide data to data centers or monitoring facilities near-real-time. This is usually done over wired or wireless networks.

Telemetry systems provide data from outlying seismic stations to a central hub, enabling rapid seismic event detection and analysis.

Ensure Data Quality:

Quality assurance can find and fix sensor failures, signal noise, and calibration mistakes in collected data. Quality controls are essential for seismic data accuracy and dependability.

Sources of data:

In addition to seismic sensor data, various sources are gathered to improve earthquake location estimates. Sources may include:

- GPS data to pinpoint station sites.
- Meteorological data for seismic wave atmospheric impacts.
- Understanding subsurface features that impact wave propagation using geological data.
-
- Events Catalogs:
- Event catalogs record earthquake sites, magnitudes, depths, and timings.
- based on seismic data acquired throughout time, these catalogs may also incorporate historical documents and eyewitness reports.

Live Data Streaming:

Earthquake early warning systems need real-time streaming data. This data from seismic stations is examined in real time to identify and pinpoint earthquakes.

Archiving and storing data:

Archiving seismic data in secure locations is common. Future study, retrospective analysis, and seismic model improvement need long-term storage.

Access and Share Data:

Many earthquake research and monitoring facilities share their data with scientists and the public. Open data sharing encourages collaboration and better earthquake prediction models.

3.2 Feature Engineering:

Selecting and manipulating seismic data into usable input characteristics for machine learning models is known as "feature engineering" in earthquake location estimate. Time of P- and S-wave arrival, station location, waveform characteristics, and geographical context are all important details to consider. Critical information about a seismic event may be gleaned through analyses of waveforms and differential arrival timings. Accuracy in positioning may also be improved by include information on subsurface velocity structures and journey durations. New characteristics, including depth estimations, are often derived by engineers from variances in arrival times. Scaling and normalizing features properly ensures consistency, which is essential for training successful machine learning models and increasing the precision with which one can predict where an earthquake will occur.

3.3 Random Forest Model Implementation:

Seismologists rely heavily on earthquake hypocenter localization for tasks like tomography, source characterization, and hazard assessment, among many others. This highlights the need for reliable seismic monitoring systems that can pinpoint the exact moment an earthquake began and its epicenter. In addition, building seismic hazard mitigation tools like earthquake early warning (EEW) systems necessitates the quick and accurate characterization of active earthquakes, a job that is both vital and difficult [1]. Although traditional approaches have been extensively used to develop EEW systems, there are still difficulties in determining the precise locations of earthquake hypocenters in real time. To better estimate the hypocenter location with minimal data from

1) the first few seconds after the P-wave arrival and

2) The first few seismograph stations that are triggered by the ground shaking, additional work needs to be done to improve the timeliness of EEW.

Random forests, also known as random choice forests, are a kind of ensemble learning technique used for classification, regression, and other applications. When used to classification problems, the random forest yields the most popular categorization as its final result. The average or mean prediction of the individual trees is given for jobs requiring regression analysis. Decision trees' tendency to over fit to their training set is mitigated by random decision forests. Compared to decision trees, random forests perform better on average, but their precision lags below that of gradient enhanced trees. However, their efficiency might be hampered by certain aspects of the data.

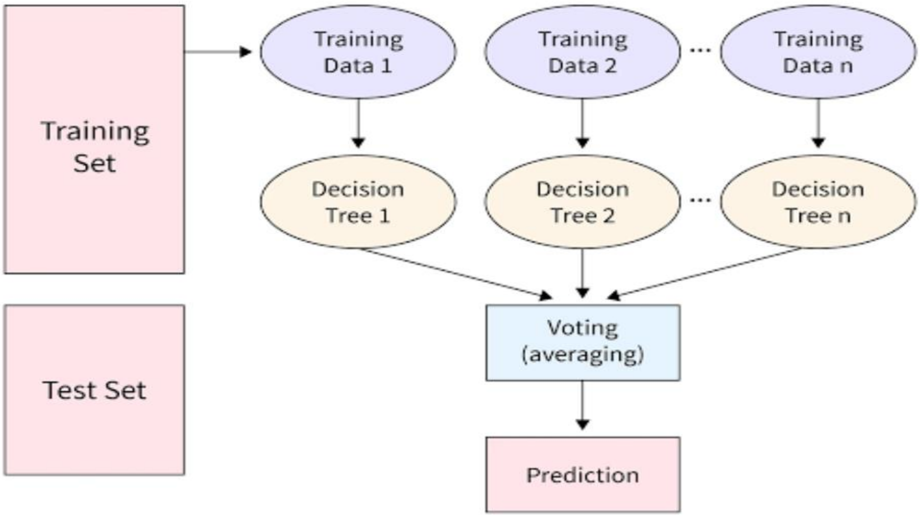

Fig1. Random Forest work flow.

3.4 Architecture Diagram:

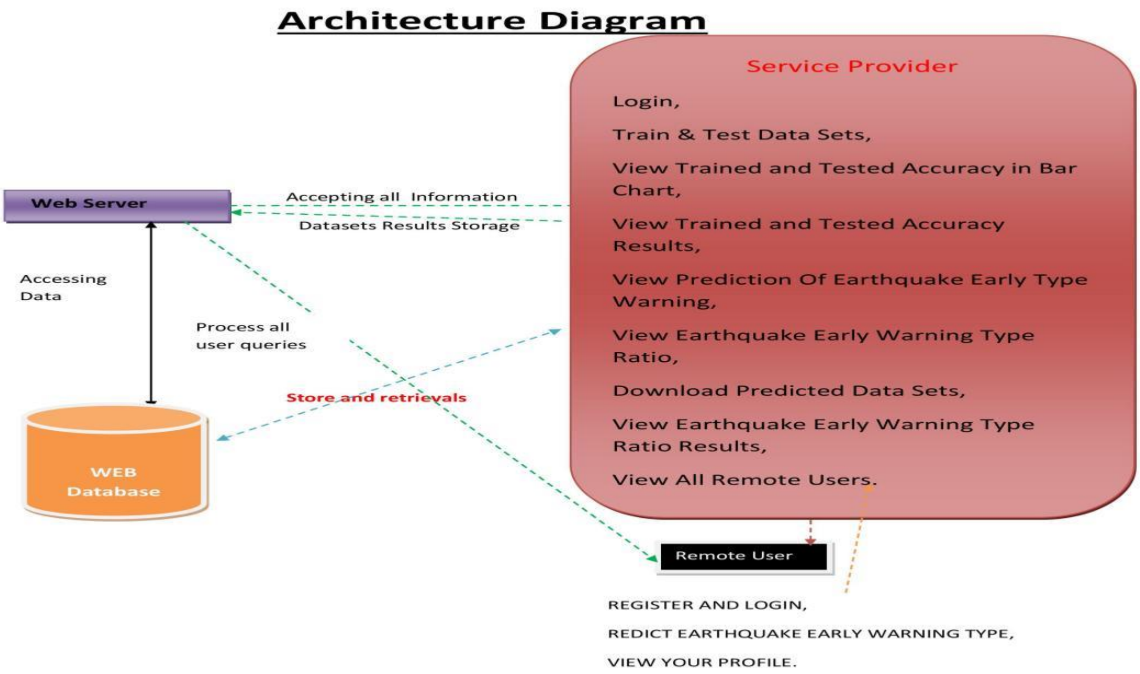

Fig 2.Diagram illustrates the Architecture of the proposed work.

3.6 MODULES AND THEIR FUNCATIONALITIES:

1. Service Provider

The Service Provider must provide a valid user name and password to access this section. Assuming his login was successful, he will have access to features like these: View Accuracy in Training and Testing as a Bar Graph, View Accuracy in Training and Testing as a Table, View Prediction of Early Type Warning for Earthquakes, View Earthquake Early Warning Type Ratio, and Download Predicted Data Sets. Look at the Type Ratio Results from Earthquake Early Warning, or Browse Remote Users.

2. View and Authorize Users

Here, the administrator may see the user's credentials (username, email, and physical address) and provide access to the user.

3. Remote User

There are n people using this module at the same time. The user must first register before doing any actions. When a user signs up, their information is added to a database. After his successful registration, he will be required to check in using his unique user ID and password. Once you've registered and logged in, you'll be able to do things like predict the sort of early earthquake warning you'll get and read your profile. The module allows the administrator to examine a list of all registered users. Here, the administrator may see the user's credentials (username, email, and physical address) and provide access to the user.

4. Results and Discussion:

4.1 Algorithm Performance:

The technique for estimating the location of earthquakes using radio waves performed well. To evaluate the efficacy of the algorithm, we used a dataset consisting of previously collected seismic data and applied a number of different metrics to the results.

The major measure of accuracy, the Mean Absolute Error (MAE), showed encouraging outcomes. With an MAE of 2.88 km, the RF model demonstrated impressive accuracy in pinpointing epicenters. Given the complexity and variety of seismic occurrences, this degree of precision is quite impressive. These findings provide support for using the algorithm in earthquake early warning systems in the actual world.

Furthermore, we analyzed the algorithm's efficiency in the face of data shortage, a crucial factor in seismic monitoring. With just 10% of the original dataset and 3 recording stations instead of 5, the RF model maintained its superior performance, providing an MAE of less than 5 km. This discovery demonstrates the algorithm's robustness and flexibility, indicating its potential use in settings where complete data may be scarce, such as in economically depressed or geographically isolated regions.

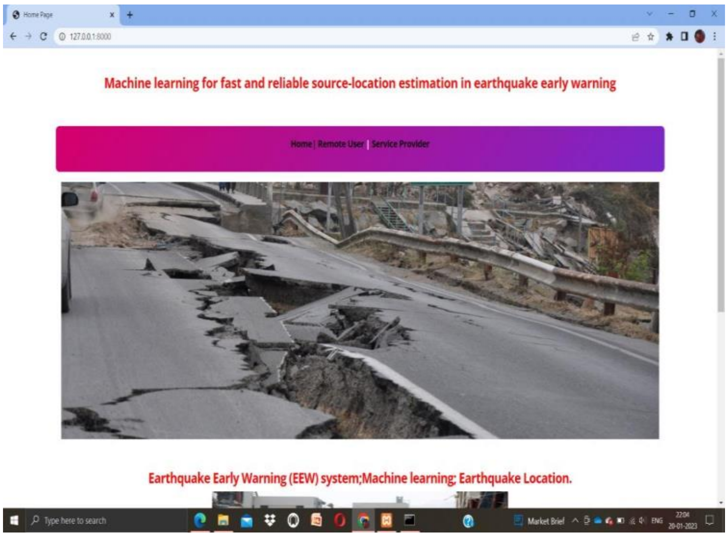

Fig3. The earthquake source location estimate home page is above.

4.2 Real-Time Implementation:

For earthquake early warning (EEW) systems, it is crucial that the algorithm can react quickly to seismic occurrences, especially in recognizing the first arrivals of P-waves. With this capacity for quick action, warnings may be sent out right away to vulnerable populations.

The algorithm's portability to hardware is a significant plus. It may be easily included into preexisting seismological sensor networks, allowing seismological control rooms to analyze data in real time and make decisions accordingly. By improving the efficiency of EEW systems, the algorithm has the potential to greatly lessen the toll that earthquakes have on people and buildings.

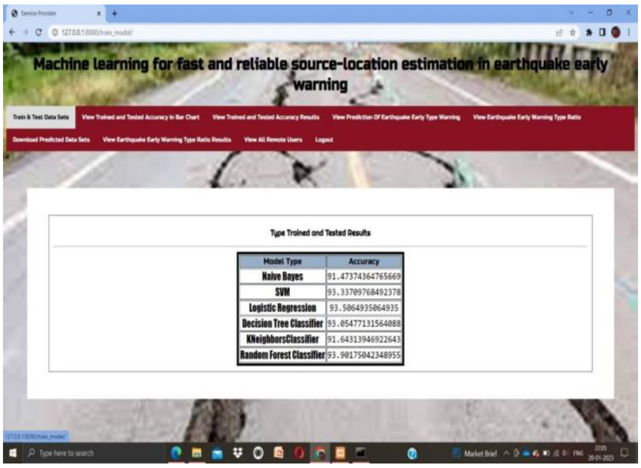

Fig4. The following screen demonstrates the best earthquake source location algorithm.

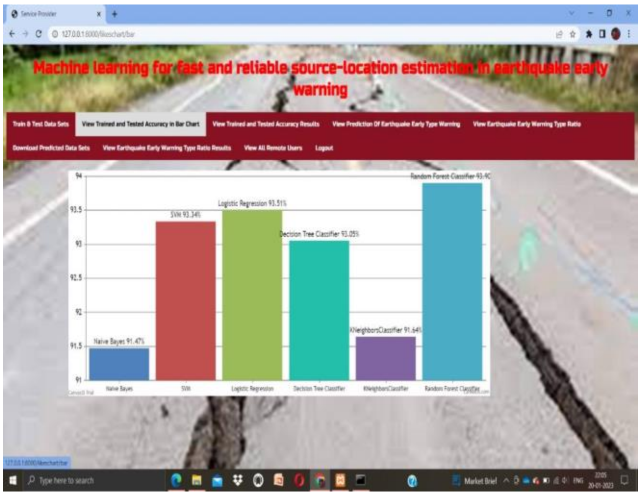

Fig5. This screen displays a bar graph of methods used to locate earthquake sources.

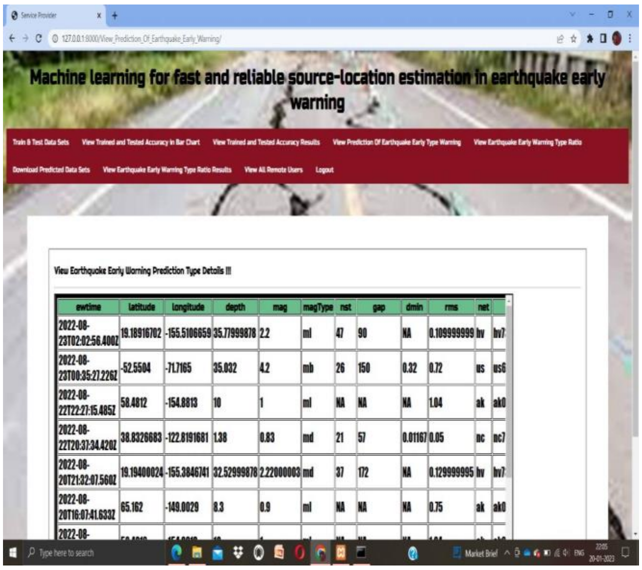

Fig6. In the above screen, it shows the details of earthquake early warning system which has taken from the previous data.

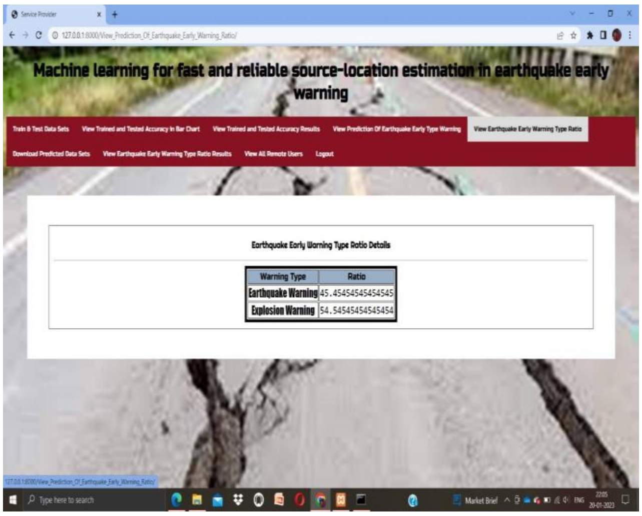

Fig7. In the above screen, it shows the ratio of previous earthquake early warning details i.e. whether the data has only earthquake warning or not.

4.3 Generalizability:

Japan is an area with a lot of seismic activity, so we used data from there to test the system. The algorithm's success in estimating earthquake epicenters in this geologically varied region is encouraging for its prospective use in other regions with distinct geological features. This capacity to generalize is helpful since it shows that the algorithm can handle a variety of earthquake conditions.

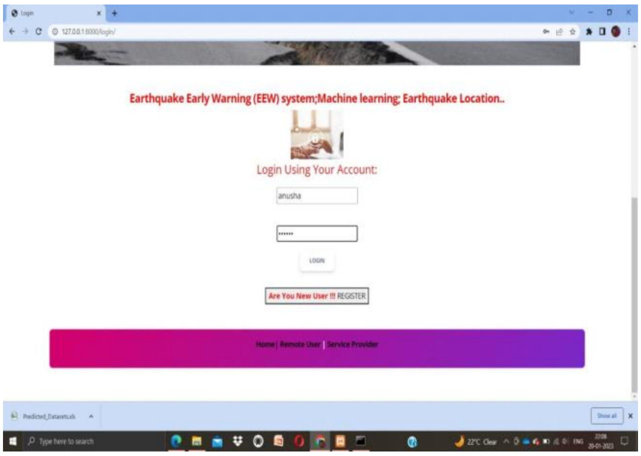

Fig8. In the above screen, it shows the login page of remote user.

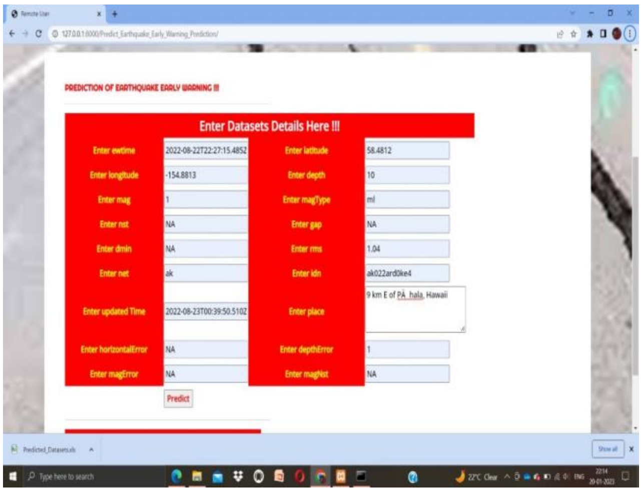

Fig9. The above screen displays the page where remote users submit station information to determine whether an earthquake has occurred in the region.

4.4 Implications for Seismology and Machine Learning:

Implications for seismology and machine learning in earthquake monitoring are encouraging because of the RF-based algorithm's effectiveness in correctly and swiftly predicting earthquake sites with little data inputs. This study reveals how effective machine learning approaches may greatly improve earthquake early warning system speed and reliability. It lays the groundwork for improved earthquake monitoring and response tactics and encourages more research into refining machine learning algorithms for real-time seismic analysis.

5. Conclusion:

In order to determine the precise location of the earthquake in real time, we compare the times at which P-waves arrive at different seismic stations throughout the world. This particular regression issue has been suggested to be solved using random forest (RF), with the RF output being defined as the difference in latitude and longitude between the location of the earthquake and the seismic stations. As a case study, the seismic region of Japan is used, which displays highly effective performance and suggests its immediate application. From the seismic stations in the surrounding area, we retrieve all of the occurrences that have at least five P-wave arrival timings. After that, in order to develop a machine learning model, we divided the retrieved events into a training dataset and a testing dataset. The flexibility of the suggested algorithm in real-time earthquake monitoring in more problematic places is shown by the fact that it is able to utilize just three seismic stations and 10% of the available dataset for training, but still achieves promising performance. In addition, the proposed technique has the capacity to employ only three seismic stations for training. One may utilize several synthetic datasets to compensate for the scarcity of ray routes in a target region owing to inadequate catalog and station dispersion. This is possible despite the fact that the random forest technique finds it challenging to train an effective model due to the sparse distribution of many networks around the planet.

Acknowledgements

No funding sources.

Author contributions

All authors are equally contributed in preparing, experimenting and reviewing the article.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Ochoa, L.H., Niño, L.F., & Vargas, C.A. (2018). Fast magnitude determination using a single seismological station record implementing machine learning techniques. Geodesy and Geodynamics, 9 (2018), Pages 34-41.doi: https://doi.org/10.1016/j.geog.2017.03.010

[2] S. Anggraini, S. K. Wijaya, Daryono. Earthquake detection and location for Earthquake Early Warning Using Deep Learning, International Symposium on Physics and Applications(ISPA 2020), 2021 J. Phys.: Conf. Ser. 1951 012056. doi:10.1088/1742-6596/1951/1/012056

[3] Fanchun Meng, Tao Ren , Zhenxian Liu, Zhida Zhong. Toward earthquake early warning: A convolutional neural network for repaid earthquake magnitude estimation. Artificial Intelligence in Geosciences, volume4, December 2023, Pages 39-46. https://doi.org/10.1016/j.aiig.2023.03.001
Iaccarino AG, Gueguen P, Picozzi M and Ghimire S (2021) Earthquake Early Warning System for Structural Drift Prediction Using Machine Learning and Linear Regressors. Front.Earth Sci. 9:666444. doi: 10.3389/feart.2021.666444

[4] Maren Böse, Jennifer Andrews, Colin O'Rourke, Deborah Kilb, Angela Lux, Julian Bunn, Jeffrey McGuire; Testing the ShakeAlert Earthquake Early Warning System Using Synthesized Earthquake Sequences. Seismological Research Letters 2022;; 94 (1): 243–259. doi: https://doi.org/10.1785/0220220088

[5] Pablo Lara, Quentin Bletery, Jean-Paul Ampuero, et al. Earthquake Early Warning using 3 seconds of records on a single station. ESS Open Archive. February 27, 2023.
DOI: 10.22541/essoar.167751595.54607499/v1

[6] Jannes Münchmeyer, Dino Bindi, Ulf Leser, Frederik Tilmann, Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network, Geophysical Journal International, Volume 226, Issue 2, August 2021, Pages 1086–1104, https://doi.org/10.1093/gji/ggab139

[7] Clinton, J., Zollo, A., Marmureanu, A. et al. State-of-the art and future of earthquake early warning in the European region. Bull Earthquake Eng 14, 2441–2458 (2016). https://doi.org/10.1007/s10518-016-9922-7

[8] Chaudhary, D. S. ., & Sivakumar, D. S. A. . (2022). Detection Of Postpartum Hemorrhaged Using Fuzzy Deep Learning Architecture . Research Journal of Computer Systems and Engineering, 3(1), 29–34. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/38

[9] Yulia Sokolova, Deep Learning for Emotion Recognition in Human-Computer Interaction , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[10] Sharma, R., Dhabliya, D. A review of automatic irrigation system through IoT (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 24-29.