Towards a Multimedia Big Data-Driven Approach for Earthquake Monitoring and Forecasting Early Warning System

Qing Ye[1], Bo Xia[1,2*], Ye Ren[1]
[1]Shanghai Seismological Bureau, Shanghai 200062, China
[2]Shanghai Sheshan National Geophysical Observatory, China
E-mail: xiabo0111@139.com, leafyale@qq.com, renyerenye@126.com

Keywords: Geo-social networks (GN), earthquake, big data, hybrid lion-optimized random forest (HLO-RF)

Received: February 29, 2024

Digital networks have fundamentally changed the way that people think by enabling them to share their location and other personal information for the benefit of their communities. The popularity of geo-social networks (GN) like Instagram, Twitter, Facebook, and Flickr has increased significantly in recent years. As a result, everyone in the world may now express their opinions, immediately report an occurrence, and interact with others. GN data therefore gives comprehensive data on individual current developments. By evaluating geo-social data in real-time, modern GN may be used as digital assets for countries and their governments. Hybrid lion-optimized random forest (HLO-RF) is the proposed technique of this research. It seamlessly integrates Lion Optimizing with random forest methods to improve complicated data-driven activities' forecasting accuracy. The novel technique promotes effective and resilient decision-making in several applications. To explore GN while gathering information and rendering decisions in real-time while monitoring and forecasting various natural occurrences, we offer an effective system and Machine Learning (ML) technique in this research. To predict the early warning system using HLO-RF. To monitor earthquake occurrences, and incidents in real-time, make potential real-time decisions, and assist in planning for the future, unique in a real-world setting, the proposed system is designed and used. We demonstrate that the proposed system has increased performance and can analyze a massive quantity of GN data in real-time, while simultaneously identifying any occurrence.

Povzetek: Članek predstavi sistem zgodnjega opozarjanja na potrese, ki temelji na multimedijskih velikih podatkih in algoritmu hibridne levje optimizacije z naključnimi gozdovi (HLO-RF). Sistem analizira podatke iz geo-socialnih omrežij za izboljšano natančnost in učinkovitost napovedi potresov.

1 Introduction

Earthquakes are considered to be the most dangerous however least expected natural disasters. The natural occurrence of catastrophes and earthquakes leads to fatalities, major infrastructure damage, the annihilation of civilizations in an instant, as well as an immediate decrease in the nation's economy [1]. A sudden discharge of resources across faults within the crust of the globe, disasters may inflict severe harm, involving the collapse of buildings, fatalities, and disruptions to the economy [2]. Using this technology, a lot of data is gathered, processed, and analyzed, including seismic data, satellite imaging, and data collected via GN.

In this field [3], multimedia big data technology-based earthquake monitoring and forecasting early warning systems can deliver real-time, precise, and in-depth information on seismic occurrences. Scientists and first responders may be better able to forecast, be ready for, and handle earthquakes with the aid of this knowledge. These devices identify movement in the soil caused by the shaking waves that travel throughout the outermost layer of Earth's surface after a tremor giving important details regarding the main system, position, and size of the tectonic occurrence [4].

Early warning systems (EWS), planning and estimation, disaster risk analysis, conversation and readiness activities, devices, and methods are all instances of hazard monitoring systems with integrated architecture that allow individuals, civilizations, officials, firms, and others to act quickly to reduce disaster risks previous to harm [5].In general, using multimedia big data technology-based early warning systems for earthquake monitoring and forecasting is essential for reducing the damage that earthquakes do to infrastructure and populations [6]. These technologies have the potential to deliver more precise and timely information to enhance earthquake preparedness and response activities as technology and data processing continue to evolve.

The present article is arranged as follows: Section 2 provides the literature survey of this article. Section 3 suggests the methodology; it includes data collection, pre-processing, service and application layer, and proposed technique. Section 4 shows the results of this experiment. Finally, the conclusion is represented in section 5.

2 Literature survey

An overview [7] of the IoT solutions utilized in early warning (EW) systems' designs and needs is presented in this paper. The study will, specifically, focus on the four most significant natural disasters flooding, earthquakes, tidal waves, and collapses we first certain fundamental ideas and requirements for Smart EWSs, as well as a basic IoT design. It has [8] only recently (within the last 30 years) that the techniques and tools necessary to put it into effect have been established. The fundamental objective of early warning systems is to notify users. An effective tool for tracking and forecasting earthquakes is a neuro-inspired system. These techniques have the potential to improve monitoring and alert systems and provide invaluable knowledge regarding the behavior of earthquakes [9]. Earthquakes Preliminary Attention methods may benefit from using the earthquake's magnitude predictions using earlier or later preliminary earthquake waves collected through certain geophysical locations, especially in soil motion-based techniques [10]. After learning an innovative method, the created system can find many earthquake events in real-time as well as limit the most significant features from the earthquake information. The resulting set of criteria is particularly informative because of its ability to find and categorize seismographs that might not have been included in the equation [11]. These methods for choosing features have a big effect on the way these methods' function [12]. However, the Actual signal classification algorithm is laborious and faults DL techniques; a feature map is created by automatically selecting features from a set of hidden layers. Deep learning algorithms [13] such as multiple hidden layers and graph neural networks efficiently absorb spatial data like units and their interconnections. When it comes to reliably and quickly analyzing complex and big datasets, deep learning techniques have been demonstrated to outperform a variety of tasks, including image analysis, standard ML methods [14], textual categorization, voice identification, and others. The three main architectures used in deep learning are CNN, GNN, and RNN. The multi-layer perceptron (MLP) ML algorithm [15] is the main ancestor of these models. MLP is an effective input, concealed, and output layer in a fully linked stack. Convolutional, pooling, dropout, and softmax layers have all been used in the advanced architecture of MLP to better grasp complicated data and improve performance. For data categorization, accessing seismic data, and extracting seismic characteristics, CNN-based systems have been utilized in several research [16]. It employed [17] CNN to forecast earthquakes and classify earthquake major sources using a single station's digital data. The suggested CNN simulation tool forecasts earthquake origins over a broad spectrum of locations and significance levels in tests using three-component amplitude data acquired from the US Geographical Survey. Two CNN-based models were created, as per article [18], the proposed CNN model reliably predicts the locations of earthquakes over a broad range of ranges and magnitudes. The models that have been created may rightly forecast both the amplitude and the regular intervals.

In a paper [19, 20], they showed how sensor location data mixed with graph-based systems may benefit from time-series data. Using the GNN paradigm, this was demonstrated via the usage of graph-based networks. Investigative findings from two successive databases that include earthquake waves are ambitious.

Table 1: Literature survey summary

Reference	Methodology	Dataset	Performance metrics
[8]	Transformer earthquake alerting model (TEAM)	Two nation scale data	PGA at 20%
[10]	Seismic contrastive graph neural network	Real world seismic dataset	MSE (0.5326), SD (0.6858), CC (70.07%), R^2 (49.10%)
[13]	Time Series Extrinsic Regression-Graph convolutional network (TISER-GCN)	seismic dataset	PGA in MSE (0.20), MAE (0.31), RMSE (0.44)
[16]	Convolutional neural network (CNN)	Earthquake database	0.000138 of CNN at the site of FKSH17
[17]	CNN	Raw waveform data	PGA (0.3 s), PGV (1 s), PSA (3 s)

3 Methodology

The system architecture is displayed in Figure 1. The system is composed of three fundamental levels from top to bottom: data gathering, data analysis, and services and applications. Along with the foundational levels, there are two further layers at work. These extra layers enable unstructured and structured data transmission and storage. Every high-speed internet connectivity, including WiMAX, 3G, and LTE, can be used for external connection with the geo-social network servers for data collection. As well as inner interaction among the different servers using various information techniques, such as Wi-Fi and Ethernet. In distributed storage using the Hadoop File System (HDFS), multiple regions of data are managed by the storage layer as utilities for backing up that might be organized, unstructured, or raw for use in the future or planning. The storage layer only collaborates with the last two fundamental levels, namely data processing and services, and applications, as opposed to the communication layer, which interacts with all three basic layers. The results are stored on HDFS data nodes after analysis and decision-making, and the data are recorded in the database following classification.

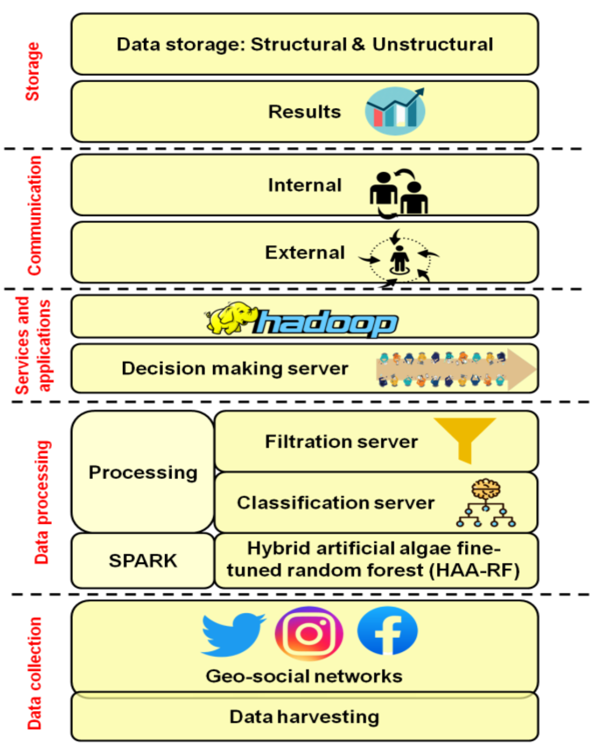

Figure 1: System architecture

3.1 Data collection layer

The datasets were gathered from Andaman & Nicobar region, in India. The period of time in the earthquake dataset is from 1900 to 2020. There are 5585 earthquakes within the Andaman & Nicobar earthquakes database. It is separated into testing and instruction parts of the same dimension. Because there are a comparable number of events in each category, this division guarantees fair coverage. In earthquake detection and prediction infrastructure, this kind of division makes it easier to create and assess accurate simulations [21].

3.2 Pre-processing layer

Pre-processing and classification are each of the essential parts of the information processor stack. Analyzing every character is not practicable considering the quantity of information from social networks that it controls. Filtration reduces the quantity of knowledge by eliminating elements that are unnecessary to address a particular scenario. This entails deleting unnecessary statistics and items that have become out of date and irrelevant about execution, location, or substance. The preliminary processing places the information in an organized fashion that makes predictive procedures for categorization possible. This expedites access to information and lowers computing complexity. To provide fast warnings, the system employs specialized criteria to recognize different sorts of information, such as fires, earthquakes, and the Ebola virus. Evaluation frequently represents the next step in the knowledge-gathering process that happens following everything has been separated and categorized. By using an integrated strategy, the platform is certain to analyze data from social networks effectively, resulting in the best possible resource utilization as well as accurate knowledge about relevant incidents and subjects

The Spark contains Hybrid Artificial Algae Fine-Tuned Random Forest (HAA-RF). It incorporates sophisticated data analytics to monitor earthquakes. It improves earthquake data processing by employing a mixture of artificial algae-based sensors along with random forest techniques. The artificial algae sensor is detecting the understated environmental changes associated with earthquake activities, offering novel inputs for the modified random forest algorithm. The collaboration enables better accurate and fast earthquake predictions, as well as early warning systems. HAA-RF provides a potential solution for robust monitoring of earthquakes through the combination of biological sense with ML accuracy and emphasizes proactively risk reduction through sophisticated predictive analysis.

3.3 Proposed algorithm

Since analyzing social networks requires so much data calculation, we need a high-end hardware and software system capable of handling such a massive amount of data. We adopted the Existing technologies to store data diagonally on several nodes with data encryption, dependability, and paths. This includes the robust distributed file system (HDFS). With its parallel programming paradigm, Mapping Minimize, the Hadoop ecosystem may do processing in parallel on identical data stored on HDFS nodes. Hadoop was originally designed for batch processing; however, in our case, we want real-time processing of continually entering geo-social data. We thus installed on the extreme top of the Hadoop system is Ignite. For effective processing of real-time data to do real-time analysis and benefit from Hadoop's parallel processing capabilities.

3.4 Service and application layer

The final stage is the utility and implementation layer, which is in charge of arriving at choices and utilizing the outcomes for different apps based on requests. For instance, Hadoop provides comprehensive computing and statistical analysis of the tweets including seismic surveys at time T for earthquake prediction in various locations. Hadoop can merely deliver the results, which the decision server can then change. Based on the findings obtained, including the time, and the volume of tweets received, the decision server then chooses the location based on the information provided, including its accuracy. The probability of an earthquake or other event occurring at a specific location is calculated by the decision system. Figure 2 displays a complete stream representation of the system's execution.

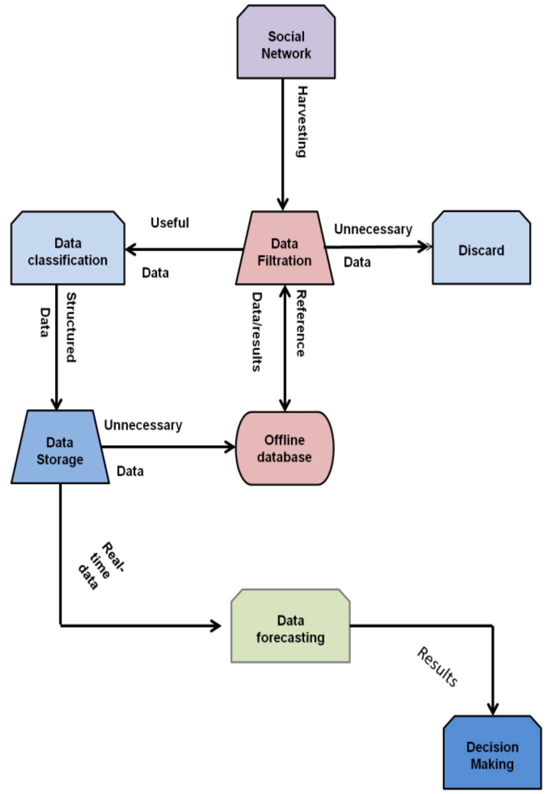

Figure 2: Proposed paradigm for systems development

3.5 K-fold cross-validations

The information is first arbitrarily assigned to several groups of K for the fivefold cross-validation. Each set is then subjected to a series of procedures as shown in Figure 3.

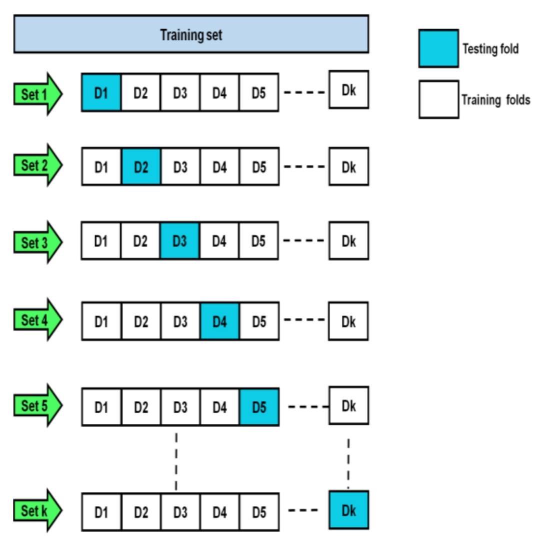

Figure 3: Fivefold cross-validation

- Choose an initial fold to serve for the experimental data set.
- The test consists of the rest K−1 groupings.
- Encourage the framework using the chosen training database, then assess it using the evaluation information.
- Develop the algorithm using the chosen learning information set, then assess it using the test information.

Typically, k in the tiny sample of this research has been set to 5 a theoretical number derived from several test attempts. When the model is used right away, the results have a small variation with little skew. These 5 sets of information were then successively entered into the HLO-RF framework. By using 5-time cross-validating, the inaccurate modeling assessment resulting from the unintentional partition of the test databases may be eliminated.

3.5 Hybrid lion-optimized random forest (HLO-RF)

Through the analysis of tweets on social media networks, the well-known ML algorithm Random Forest can be used to forecast early warnings of earthquakes. Millions of individuals use social media sites like Twitter to express their ideas and opinions about a variety of subjects, including earthquakes and other natural catastrophes. The RF algorithm is a paradigm for making decisions based on the principle of generating several decision trees to forecast an event's result. The ultimate choice in the model is determined by the majority vote of all the decision trees, each of which is constructed individually. To arrive at a final forecast, the algorithm first divides the data into various subsets, builds decision trees on each subset, and then combines them. To use RF for early earthquake warning prediction from social media tweets, the system is trained using a dataset of tweets submitted before and after an earthquake. The dataset should contain a variety of tweets, including those in which people express fear or anxiety about the earthquake and tweet details like its location and magnitude.

As a further step, the computer uses the newly acquired model to analyze recent tweets to foretell the occurrence of an earthquake. Predictions are based on several factors, including the language used in the tweet, its location, and the mood expressed in it. Because RF can handle big datasets with numerous variables and is very accurate, it is a preferred choice for early seismic warning prediction. It can be trained on both numerical and categorical data, and it can even manage missing data. In conclusion, RF is a potent ML algorithm that, through the analysis of tweets on social media sites, may be used to forecast early warnings of earthquakes. The system may be trained on a collection of tweets to identify trends and forecast earthquake probabilities with high accuracy.

There are several techniques to investigate the application of LOA and RF in earthquake prediction. Using LOA to optimize RF's hyperparameters, such as the total number of trees, the maximal depth for every tree, as well as the number of characteristics to consider at every splitting, is one strategy. Determining the ideal arrangement of these factors might enhance the effectiveness of RF in earthquake prediction. An alternative strategy would be to train an RF model on the subset of characteristics that HLO determines to be most important for monitoring earthquakes. As a result of removing superfluous or unnecessary characteristics, the RF model may become more accurate and efficient. All things considered, the hybrid technique may improve the precision and effectiveness of earthquake prediction models. It is crucial to remember that earthquake prediction is a complicated topic, and further study and testing would be required to properly assess this method's efficacy.

"Lion optimization algorithm (LOA)" is the name of a stochastic metaheuristic optimization approach. By mimicking how lion pride behaves, in which individual lions cooperate and compete to accomplish shared objectives, this method may be used to efficiently search for optimum solutions inside a search space. Each time the issue is solved, the improved metaheuristic optimization approach yields a successful outcome. In the solution domain, the population is first randomly generated to begin the LO process. The LOA procedure refers to each answer as a lion, as seen in Equation (1)

$$Lion = \ x1, x2, x3, \dots . x_{N_{var}} \tag{1}$$

Where $\% N$ the fraction of is randomly produced nomad lions, and the remaining populations are the resident lions, the lion's locations are represented byx_1; x_2; x_3, and the dimension space is denoted by$x_{N_{var}}$. To represent the network's users, an encoding scheme is selected using a genetic process. The location of a lion is represented by x, which has n components, including$(x1, x2, \dots, xn)$. The lion's location within a searching region reflects the solutions and determines each lion's fitness value (FV). The fitness value constitutes a function that accepts possible solutions for the problem in input along with outputs on how efficient the solution's in regards to issue under the consideration. The next sections describe the processes used by the LOA for community identification:

Algorithm process:

Stage 1: Initialization: During a solution space, the populations are produced randomly, and the lion's positions are maintained as a matrix. The objective functions, which will be sorted and stored in the form of a matrix and presented within Eq. 2, are utilized to compute the FV for every lion.

$$f(lion) = f(x1, x2, \dots, x_{N_{var}}) \tag{2}$$

Where xNvar is the space of dimensions, $(the\ lion)$ represents the lion's FV and the dimension space is expressed by$x_{N_{var}}$.

Stage 2: Mating: This method involves giving people the potential to generate the finest novel solution utilizing existing ones that include "mutation and crossover". The best solutions include killing weak cubs.

Stage 3: territorial defense: It is a comparison of the fitness levels that reside along with migrant lions. The resident lions will be replaced with nomadic lions if the FV for the nomadic lions is more powerful compared to the resident lions.

Stage 4: Territory takeover: The group selects every lion (cub and male) depending on fitness. Weaker male lions are ejected from the pride and become nomads. Metaheuristic lion optimization finds important features and improves classification generalization.

This hybridization aims to leverage the strengths of both methods to enhance the accuracy along efficiency of the early warning system. The HLO-RF approach can effectively identify potential risks or events before they occur. The primary application of the HLO-RF approach is to develop an early warning system that can effectively predict and alert about potential risks, anomalies, or critical events. By training the HLO-RF model on historical data that includes patterns or indicators leading up to specific events, the system can learn to recognize these patterns and issue warnings when similar patterns begin to emerge in new data. This proactive approach can be invaluable in various fields, such as finance, healthcare, environmental monitoring, and more, where early detection of anomalies or risks is crucial for effective decision-making and risk management.

HLO-RF is to improve the efficacy and precision of the evaluation of earthquake information. It is more adept at navigating the changing and intricate terrain of earthquake information. It seeks to advance notification capacities across areas vulnerable to earthquakes by improving earthquake incident findings, categorization, and predictions. HLO-RF are particularly good at detecting the complex connections and trends present in data from seismic surveys because they make use of a collection of tree models that have been educated using various portions within the information. Using this knowledge, they can precisely identify seismic occurrences, estimate their amplitudes, and anticipate subsequent quakes. Furthermore, HLO-RF helps ensure reliability under all difficult circumstances by providing resistance to outliers and other types of noise frequently present in everyday life earthquake information. This has to be done to use cutting-edge, information-driven tactics to improve the capacity to reduce earthquake risk and safeguard populations that are exposed to danger.

4 Result and discussion

The fivefold cross-validation method was utilized to to choose the test sets for the input information in an orderly way, followed by the HLO–RF and can ensure that the simulation performs properly in the absence of information as well as obtaining an additional reliable assessment of its efficacy as shown in the Table 2.

Table 2: Numerical value in fivefold cross-validation

Evaluation metric	**1**	**2**	**3**	**4**	**5**	
Accuracy	87	73	85	90	96	
MAE	0.3576	0.4389	0.3627	0.5017	0.3200	
RMSE	0.4501	0.3872	0.4079	0.3423	0.3107	

By applying the established approaches Bayesian ridge, Decision tree (DT), and Extreme gradient boosting (XG Boost) we verified our suggested approach using "Output, Accuracy, MAE and RMSE" The proposed methodology was HLO-RF. Figure 4 and Table 3 depict the Count of tweets about earthquakes. When the period that has passed since the tweet grows, an increased number of earthquake tweets can be found.

Table 3: Count of tweets about earthquakes

No.of Earthquake Tweets Detected	Time after Earthquake (Min)
3	1
5	1.5
7	6
9	6.5
11	2.5
13	5
15	5.5

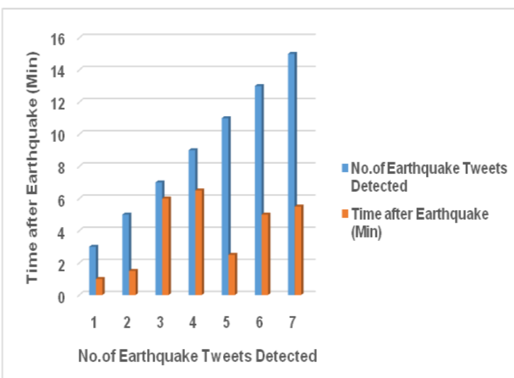

Figure 4: Graphical representation for a count of tweets about earthquakes

Accuracy is demonstrated in Figure 5 and Table 4. How closely a measurement or forecast matches the actual or anticipated value is referred to as its accuracy. It is a statistical notion that is frequently used to assess how well a method or system performs when monitoring or categorizing events. A graph showing the accuracy of an earthquake monitoring and forecasting early warning system over time. The accuracy and dependability of forecasts depend upon the date, position, magnitude, and appearance of earthquakes. To adequately warn and mitigate the possible effects of disasters on populations and facilities great precision must be attained. Real strong earthquakes must be recognized by interfering factors and ambient sound for a precise identification to take place, which promises that every single pertinent incident is tracked down as soon as possible. It might contain measures like earthquake forecast accuracy, false reports, and warning time between earthquake detection and real events. Bayesian ridge [22] (84%), XGB [22] (86.5%), and DT [22] (90.1%) are the existing methods. The proposed method HLO-RF attains 98%. Comparing our suggested method to the present one, it is superior.

Table 4: Values for Accuracy

Algorithm	Accuracy (%)
Bayesian Ridge [22]	84
XGB [22]	86.5
DT [22]	90.1
HLO -RF [Proposed]	98

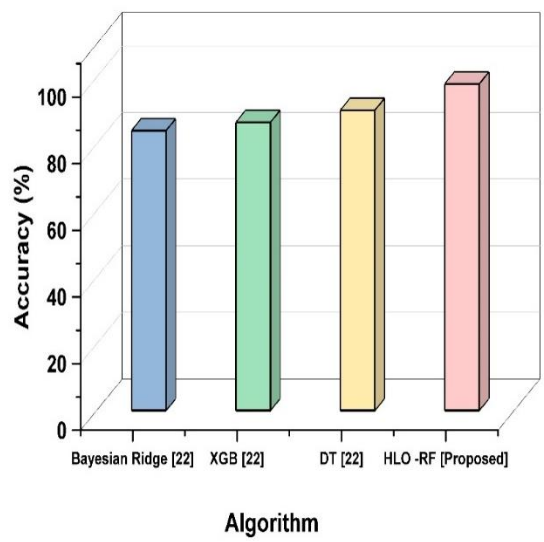

Figure 5: Graphical representation for accuracy

Figure 6 and Table 5 show MAE results. An MAE graph shows the Mean Absolute Error (MAE) values for different ML or simulation approaches or parameterization. When it comes to seismic characteristics like earthquake size, dimension, or recurrence duration, the MAE measures the mean amount of deviations among the actual and projected quantities. The coefficient, which is independent of the spatial location of shortcomings, offers a simple way to assess the simulation that predicts earthquake occurrences. Regression models are frequently evaluated using the MAE, which quantifies the average absolute difference between the expected and actual outcomes. Bayesian ridge [22] (0.2367), XGB [22] (0.2776), and DT [22] (0.3537) are the existing methods. The proposed method HLO-RF attains 0.1852. As opposed to the present method, ours is superior.

Table 5: Values for MAE

Algorithm	MAE
Bayesian Ridge [22]	0.2367
XGB [22]	0.2776
DT [22]	0.3537
HLO -RF [Proposed]	0.1852

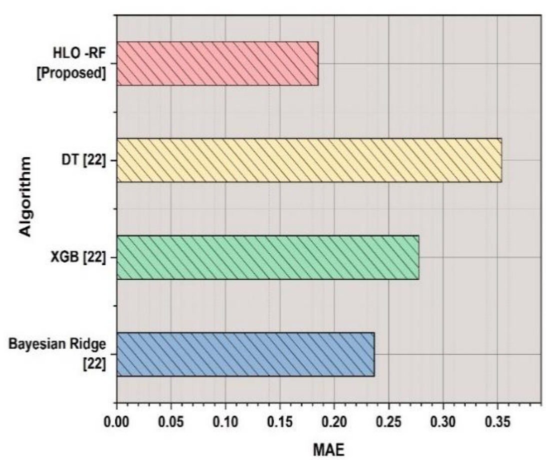

Figure 6: Graphical representation for MAE

Root Mean Square Error (RMSE) is another widely used metric for evaluating the performance of a predictive model, including the Hybrid Lion Optimized Random Forest (HLO-RF) model for the early warning system. MAE is identical to RMSE, however, it focuses on greater mistakes by calculating the squared roots for the average for the squared variations between predicted along actual values. A critical indicator for evaluating the precision of forecasting techniques is the potency of tracking mechanisms. The relative level of variance between actual and expected earthquakes can be determined by RMSE, which offers information about the system and predicts future earthquake activity. A single amount which indicates the average size of prediction mistakes gets generated by this technique. It is particularly useful for assessing the variance between the predicted and actual values. Figure 7 and Table 6 show RMSE results and they show that our proposed approach performs better than another existing approach. Bayesian ridge [22] (0.3607), XGB [22] (0.4018), and DT [22] (0.5350) are the existing methods. The proposed method HLO-RF attains 0.1947.

Table 6: Values for RMSE

Algorithm	RMSE
Bayesian Ridge [22]	0.3607
XGB [22]	0.4018
DT [22]	0.5350
HLO -RF [Proposed]	0.1947

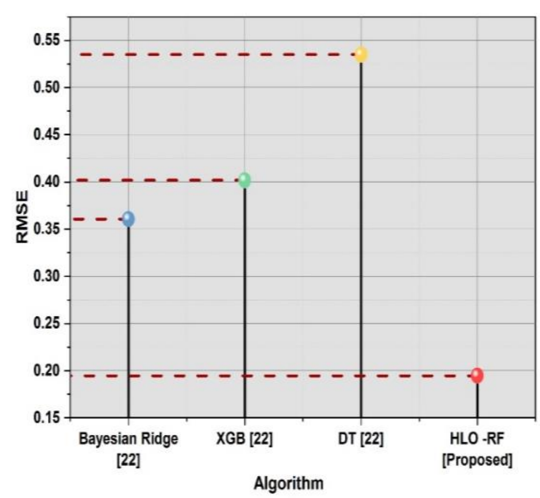

Figure 7: Graphical representation for RMSE

4.1 Earthquake warning time

The predicted warning time using our suggested model is covered in this section. Large cities and other places that need notice are far from the station and the event, thus the warning time depends on that distance. Since the result time has a significant impact on the overall procedure time, our approach seeks to limit the amount of data required. The following may be used to indicate the EEW network warning time:

$$T_w = t_s - t_{epd} - t_{proc} \quad (3)$$

The period $t_s$reflects the lapse between when the earthquake started and when the S wave finally reached its destination. In contrast, the t_{epd} value shows the earliest possible time (including communication time) at which P wave detection is declared by the fewest number of stations. Finally, the sum of t_{proc} is the time it took for several stations to verify the event and pinpoint its size. Only three stations are required to make an early P phase arrival announcement. In our tests, we found that the time it took for both communication and triggering was around one second. Additionally, the HLO-RF approach needs a waveform waiting time of three seconds to assess an earthquake's magnitude and location. For locations located behind the station, the warning time depends on how far the station is from the site and how close the station is to the event source. Therefore, the time S wave arrives is not the same as the time an alarm is received on a local level.

4.2 Discussion

Assessing a CNN's [9] potential performance in detecting tremors and characterizing its position and size since it operates immediately on brief periods of single-station seismograph waves with little to no preparation was a fundamental problem in seismic surveillance. If the emphasis focuses on component representations rather than charts, then using point fixes was inappropriate since the depiction's range of values therein is going to be considerably more uniform than and not as useful for telling us apart nodes as individuals [14]. The site FKSH17 has occurred 0.000138 in mean square error (MSE) of 0.001734 SHAKE 2000, the site FKSH18 obtained 0.011128 in MSE of 0.123401 in SHAKE 2000 [16]. The source waves' highest amplitudes must be preserved given that they are essential towards precisely projecting the desired peaks instant messages. They normalize the waveform being used times its highest point so as to get around these constraints, following which enter the normalization values through the final completely linked layers [17]. The primary kinds collected platforms endurance-capable multicolor drones will face challenges from the collection of data across wider regions. In overall, greater sorties must be performed to provide the intended region covering regardless of the drone's durability. Additionally legal constraints like limit operative elevations and being close to controllers' and witnesses' visible path of vision make it difficult to collect information across wider regions [18]. The proposed method of HLO-RF involves highly accurate for involving the regression as well as classification. It also means that they can be highly precise in accurately classifying earthquakes and forecasting magnitudes or other relevant variables within the setting of earthquakes surveillance and it performed 98% at accuracy, 0.1852 at MAE, and 0.1947 at RMSE.

4.3 Implication of the study

The recognition of earthquake-generated waves of vibration in small tremors or incidents that happen away from monitoring facilities, though, could not register on the network's detecting thresholds, missing possibly hazardous occurrences. Incorrect alerts from warning systems for earthquakes may still happen, particularly in areas that have complicated geography or a great deal of backdrop tremors. False alarms may sap public confidence in the network and cause people to become complacent or ignore subsequent alerts. A significant amount of cash and continuous maintenance is needed to construct and keep up extensive seismic surveillance and beforehand signal infrastructure. Costly technology acquisition, setup, upkeep, and employee instruction might prove problematic, especially in areas having limited resources. Bayesian Ridge [22] finds it difficult to convey the complex chronological and geographic trends found in seismic readings. The framework's basic parameters may not adequately capture the effect of several elements, including faults, continental plate actions, and geographical characteristics, on earthquake recurrence.

4.4 Limitations of existing methods

XGB [22] overestimation was a common problem with XGBoost designs, particularly when these models undergo training on noisy or extremely dimensional information. The accuracy of earthquake forecasts may be jeopardized if excessive fitting results in low generalization efficiency, whereby the system works on initial information and nonetheless struggles to generalize to new information. DT [22] gives predictable forecasts despite explicitly measuring the likelihood of class affiliation, according to the overwhelming vote within leaf branches. Using decision trees may not consistently generate stochastic results that are why stochastic outcomes have become popular in earthquake forecasting and surveillance to measure variability and evaluate predicted dependability.

5 Conclusion

The earthquake monitoring while predicting early alerting system, which depends on multi-media big data technologies along with the hybrid lion optimization random forest (HLO-RF) technique, indicates substantial progress in accuracy and effectiveness. The conceptual framework of the system incorporates data collecting, pre-processing, as well as service layers, having an emphasis on real-time analytics along with making decisions. The HLO-RF technique improves the accuracy of earthquake forecasts by analyzing social media information. The findings indicate the superiority of the suggested technique over existing methods, as demonstrated by increased efficient for earthquake monitoring. This model's estimated warning time also highlights its potential for proactively risk reduction. Further study and testing are needed to provide a full evaluation of the system's effectiveness in earthquake predictions.

References

[1] Tehseen R, Farooq MS, Abid A, 2020 Mar 19; Earthquake prediction using expert systems: a systematic mapping study. *Sustainability.* 12(6):2420.

[2] Liu, Y., Zhao, Q. and Wang, Y., 2024. Peak ground acceleration prediction for on-site earthquake early warning with deep learning. *Scientific reports*, *14*(1), p.5485.

[3] McBride, S.K., Bostrom, A., Sutton, J., de Groot, R.M., Baltay, A.S., Terbush, B., Bodin, P., Dixon, M., Holland, E., Arba, R. and Laustsen, P., 2020. Developing post-alert messaging for ShakeAlert, the earthquake early warning system for the West Coast of the United States of America. International Journal of Disaster Risk Reduction, 50, p.101713.

[4] Pwavodi, J., Ibrahim, A.U., Pwavodi, P.C., Al-Turjman, F. and Mohand-Said, A., 2024. The role of artificial intelligence and IoT in the prediction of earthquakes. *Artificial Intelligence in Geosciences*, p.100075.

[5] Early Warning System Definition by United Nations Office for Disaster Reduction. Available online: https://www.undrr.org/terminology/early-warning-system (accessed on 1 March 2022)

[6] Pribadi, K.S., Abduh, M., Wirahadikusumah, R.D., Hanifa, N.R., Irsyam, M., Kusumaningrum, P. and Puri, E., 2021. Learning from past earthquake disasters: The need for a knowledge management system to enhance infrastructure resilience in Indonesia. International Journal of Disaster Risk Reduction, 64, p.102424.

[7] Esposito, M., Palma, L., Belli, A., Sabbatini, L. and Pierleoni, P., 2022. Recent advances in internet of Things solutions for early warning systems: A review. *Sensors*, 22(6), p.2124.

[8] Münchmeyer, J.; Bindi, D.; Leser, U.; Tilmann, F. The Transformer Earthquake Alerting Model: A New Versatile Approach to Earthquake Early Warning. Geophys. J. Int. 2021, 225, 646–656.

[9] Murugaraj, S.S., Kumar, S. and Balaji, S.R., 2024. Neuro-Inspired Algorithms for Earthquake Monitoring and Prediction. In *Novel AI Applications for Advancing Earth Sciences* (pp. 315-329). IGI Global.

[10] Murshed, R.U., Noshin, K., Zakaria, M.A., Uddin, M.F., Amin, A.S. and Ali, M.E., 2024. Real-time seismic intensity prediction using self-supervised contrastive gnn for earthquake early warning. *IEEE Transactions on Geoscience and Remote Sensing*.

[11] Rekha, M., Chilambuchelvan, A., Babu, D. and Vasudhevan, V., 2024. Development of Neuro-Inspired Algorithms for Monitoring and Prediction of Earthquakes. In *Novel AI Applications for Advancing Earth Sciences* (pp. 86-100). IGI Global.

[12] Mousavi, S.M.; Beroza, G.C. A Machine-Learning Approach for Earthquake Magnitude Estimation. Geophys. Res. Lett. 2020, 47, e2019GL085976.

[13] Kim, G.; Ku, B.; Ahn, J.K.; Ko, H. Graph Convolution Networks for Seismic Events Classification Using Raw Waveform Data from Multiple Stations. IEEE Geosci. Remote Sens. Lett. 2022, 19, 3004805.

[14] Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph Neural Networks: A Review of Methods and Applications. AI Open 2020, 1, 57–81.

[15] Audretsch, J. Earthquake Detection Using Deep Learning Based Approaches. Master’s Thesis, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, 2020.

[16] Hong, S.; Nguyen, H.T.; Jung, J.; Ahn, J, 2021. Seismic Ground Response Estimation Based on Convolutional Neural Networks (Cnn). *Appl. Sci.* 11, 10760.

[17] Jozinovi´c, D.; Lomax, A.; Štajduhar, I.; Michelini, A, 2021. Rapid Prediction of Earthquake Ground Shaking Intensity Using Raw Waveform Data and a Convolutional Neural Network. *Geophys. J. Int.*, 222, 1379–1389.

[18] Kucharczyk, M.; Hugenholtz, C.H, 2021. Remote Sensing of Natural Hazard-Related Disasters with Small Drones: Global Trends, Biases, and Research Opportunities. *Remote Sens. Environ.* 264, 112577.

[19] Amin, J.; Sharif, M.; Anjum, M.A.; Raza, M.; Bukhari, S.A.C. 2020. Convolutional Neural Network with Batch Normalization for Glioma and Stroke Lesion Detection Using MRI. *Cogn. Syst. Res*. 59, 304–311.

[20] Bloemheuvel, S.; van den Hoogen, J.; Jozinovi´c, D.; Michelini, A.; Atzmueller, M. Multivariate Time Series Regression with Graph Neural Networks. arXiv 2022, arXiv:2201.00818.

[21] Debnath P, Chittora P, Chakrabarti T, Chakrabarti P, Leonowicz Z, Jasinski M, Gono R, Jasińska E. 2021 Jan 19;Analysis of earthquake forecasting in India using supervised machine learning classifiers. *Sustainabilit*y. 13(2):971.

[22] Biswas S, KUMAR D, Bera UK. Prediction of earthquake magnitude and seismic vulnerability mapping using artificial intelligence techniques: a case study of Turkey. 2023.