IoT-based Earthquake Prediction Using Fog and Cloud Computing

Kanika Saini*, Sheetal Kalra

Department of Computer Science and Engineering Guru Nanak Dev University, Regional Campus, Jalandhar, Punjab, India

*Corresponding author

doi: https://doi.org/10.21467/proceedings.114.28

Abstract

Earthquakes are severe, unexpected, life-threatening catastrophes that affect all kind of living beings. It frequently results in the loss of life and property. Predicting earthquake is the most important aspect of this field. With the golden age of the Internet of Things (IoT), an innovative new idea is to couple IoT technology with cloud and fog computing to improve the potency and accuracy of earthquake monitoring and forecasting. The embedded IoT-Fog-Cloud layered structure is adopted in this research to predict earthquakes using seismic signal data. This model transfers sensed seismic signals to fog for analysis of seismic data. At fog, Fast Walsh Hadamard transform is used to extract time and frequency domain features and PCA is employed to reduce the dimensionality of feature sets. Random Forest algorithm has been used to classify seismic signals into two different events, viz., earthquake and non-earthquake accompanied by the real-time warnings. When compared to other classification models, implementation findings indicate that the Random Forest classifier achieves high values of specificity, sensitivity, precision, and accuracy with average values of 88.50%, 90.25%, 89.50%, and 92.66%. Hence make this framework more real-time compliant for earthquake prediction.

1 Introduction

Earthquakes are the most devastating among all natural catastrophes. It has an adverse impact on millions of people across the globe, resulting in significant loss of life, massive damage to property, and infrastructure. The recent Nepal earth- quake with a magnitude of 7.8 killed over 8000 people and destroyed vast amounts of property [1]. The Mount Everest Avalanche, a result of this disastrous earthquake has induced numerous environmental changes, including a one-inch reduction in Mount Everest's height [2]. It is necessary to predict where and when large earthquake will occur in the future in order to mitigate the risk. Alleviating seismic risk seems to be a challenging task that necessitates the collaborative efforts of researchers, engineers, and administrators and must be approached at various time frames.

Despite the fact that earthquakes are completely unpredictable, it is plausible to leverage some peculiarities which lead to rapid discovery of them using a proper combination technological infrastructures and real-time services. Further with significant technological advancements in the fields of sensor networks, communications infrastructure, cloud computing, fog computing, and data analytics, it is now possible to establish an integrated earthquake monitoring and prediction framework. In this study, we proposed a robust IoT-fog- cloud centric monitoring system that aids in the real-time monitoring, seismic statistics assessment and associated events along with seismic data, ground-based data, and location- based data. The fog/cloud computing paradigm is regarded as a highly viable approach for meeting the ever-increasing demand for real-time monitoring and adequately dealing with growing amount of raw data and latency issues. The proposed framework comprises of three distinct but completely incorporated and interoperable layers: the data collection layer, the fog layer and the cloud layer. The first layer comprises of IoT-based sensors responsible for the acquisition of seismic data and transmission to fog layer for additional handling. Fog layer focuses on the analysis of seismic signals and the pre-processing of data for real-time decisions.

Proceedings DOI: 10.21467/proceedings.114; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-947843-8-8

The cloud layer is accountable for accumulating and compiling of data that cannot be interpreted by the fog layer.

The objectives of paper are: (a) a resilient iot-fog-cloud based framework for seismic monitoring, capable of meeting rigorous latency requirements and high accuracy and throughput, (b) fog computing enabled quick earthquake detection followed by real-time alerts, and (c) communication of analysis results with seismological departments and response agencies.

The remainder of the paper is structured as follows: Section 2 outlines the related work in given field. The proposed framework is described in Section 3. Section 4 summarizes the implementation outcomes. Finally, in section 5, we present our conclusions.

2 Related Work

This section examines several significant contributions to the field of earthquake prediction. Alphonse & Ravi [3] proposed the IoT based framework for earthquake prediction using wireless sensor networks. Fischer et.al. [4] proposed self organizing wireless mesh network based earthquake early warning system. Majhi et al. [5] put forth an integrated model based on neural network with machine learning and optimization algorithms for earthquake magnitude prediction. G Reddy [6] put forth a framework for earthquake prediction using wavelet Transforms and clustering methodologies. Cao et al. [7] proposed a model to predict earthquakes premised on crowdsourcing data of abnormal animal traits from both passive and active sources. Yamamoto et al. [8] formulated Stochastic Model using the wavelet packet transform to classify intricate time variant earthquake ground motions. Asencio et al. [9] employed regression algorithms to develop a cloud- based big data infrastructure for earthquake prediction in California.

3 Proposed Model

This paper proposes an IoT-fog-cloud-based hierarchical structure for earthquake monitoring and prediction, as illustrated in Figure 1. The model composed of four layers: Seismic Data collection Layer, Fog Layer, Cloud Layer and Communication Layer. The first layer i.e. Seismic Data Collection Layer characterized by an extensive variety of IoT sensors mounted in seismic areas to acquire data on numerous seismic attributes. The amassed data is then divulged to the fog layer for extraction of features, selection of relevant features, and classification for real-time earthquake detection. The Cloud layer is accountable for storing compiled data as well as fore- casting and predicting results, thereby aiding the Seismological departments and response agencies in mitigating and managing earthquakes effectively.

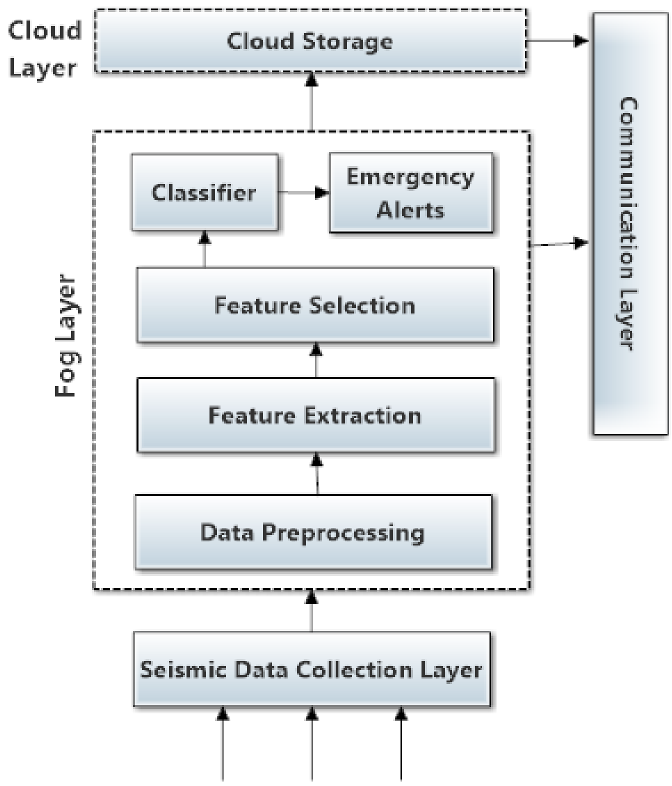

Fig. 1: The Proposed System's Layered Architecture

3.1 Seismic Data Collection Layer

The earthquake prediction system needs information about seismic signals and earthquake phenomena-related indicators that directly or indirectly trigger seismic events. This task is carried out by deploying various available sensors in the study to collect data about seismic samples and identified indicators. The earthquake-related indicators are divided into two datasets: Geological dataset and Location dataset. Geological dataset involves information on seismic signals, ground motion, ground level variations, and seismic noise. The location dataset includes indicators such as latitude, longitude, and epicenter depth.

3.2 Fog Layer

1) *Data Preprocessing:* When sensors record seismic data, the actual data is adulterated by various extrinsic influencing factors like noise and artifacts. This module's function is to filter out all noise by passing seismic signals through a high- pass filter to curtail baseline drifts and cut off frequency.
2) *Feature Extraction:* Due to the non-linear nature of signals and frequency variants with time, it is unfeasible to assess earthquake signals precisely, therefore applying an appropriate feature extraction method can optimize seismic signals monitoring. The fast Walsh Hadamard Transform (FWHT), designed mainly for signal compression, has been used in this study to extract features from earthquake signals [10]. The Walsh-Hadamard transform is a non-sinusoidal, orthogonal transformation method that converts a signal into a set of basic functions defined as Walsh coefficients, which are square or rectangular waveforms with +1 or – 1 value [11]. Each Walsh function has a unique sequence value and is used to calculate the original signal frequency. The FWHT can identify signals with sharp discontinuities more accurately and reduces computation to O(n^2) using minimal coefficients. As a result, it only requires nlogn addition and subtraction operations. The FWHT of a signal s(t) of length N is described as follows:

   $$r_n = \frac{1}{N}\sum_{i=0}^{N-1} s_t\, FWHT(n,i), n = 1,2 \ldots\ldots, N-1$$

   where r_n is the resultant coefficient and FWHT(n,i) is the applied transformation (Walsh functions) [10]. The coefficients extracted from the seismic signal in the form of features must be normalized to eliminate any potential errors caused by badly extracted features. Moreover, discriminating features of seismic signals are exemplified in the frequency and spectral segments, allowing signals to be classified more accurately and quickly.

   Seven statistical time-domain features and three frequency- domain features are included in the extracted feature. Mean, median, variance, standard deviation, kurtosis, root mean square, and skewness of the time-series signal are statistical time-domain features. The frequency-domain features are the dominant frequency, frequency amplitude, and signal energy.
3) *Feature Selection:* The extracted feature vectors might be closely associated together. Hence, the feature set must be reduced to a minimal level but sufficient one. The principal component analysis (PCA) is a popular unsupervised dimension reduction algorithm that uses a linear transformation ma- trix to preserve the most important information in feature sets [12]. The set of original variables that retain the most information from the actual data expressed as Principal Components (PCs). The eigenvectors of the extracted features covariance matrix are used to construct the transformation matrix. Let λ_e define the covariance matrix's eigenvalues and v_e the eigenvectors, then we can devise the transformation matrix by selecting the K largest λ_e's and the respective eigenvectors. We can deduce a variety of most influential features by employing the transformation matrix to the actual extracted features.
4) *Classification:* Once relevant features are extracted from the seismic signal, seismic data samples are

classified into different events such as Earthquake or Non-Earthquake. The Random Forest classification algorithm is used to classify seismic signals that have been characterized by various features obtained from feature selection component [13]. This algorithm is based on the machine learning approach, commonly used for classification problems. It is a group of several independent and unpruned decision trees that consolidate the outcomes of various trees for classification, making it a more precise classifier with better learning efficiency. Random Forest technique can address a huge set of input attributes with short training and prediction times and can acknowledge learning and classification for non-linear data sample, emerged as a recommended approach for classification [14][15]. After classifying the current event as an earthquake, emergency alerts are delivered to seismological departments and response teams in order to effectively mitigate seismic risk.

3.3 Cloud Layer

The goal of this layer is to store in- formation about seismic-oriented data and compiled results of seismic event analysis at the fog layer. This proved to be a massive benefit to the administration, response teams, seismological departments, and disaster management agencies in effectively mitigating and managing earthquakes.

3.4 Communication Layer

The proposed system's outcomes are significant inputs for different organizations and appropriate agencies that can intervene timeously to ameliorate the aftereffects. These outcomes can be prioritized by government agencies as the most important considerations while formulating short and long-term policies to manage and minimize the adverse effects of the earthquake.

4 Result Evaluation

Various datasets are used for the implementation of the proposed paradigm, including a geological dataset comprising of earthquake seismic signals acquired from the National Research Institute of Earth Science and Disaster Prevention (NIED) [16] and a location dataset accessed from the United States Geological Survey (USGS) [17]. These datasets are gathered and integrated in the Amazon Elastic Compute Cloud (Amazon EC2). To extract discriminatory features, we used the fast Walsh- Hadamard transform to decompose a signal into a frequency domain by transforming its time domain. By applying FWHT to each seismic data file, 1023 coefficients are generated using MATLAB2019B. Figure 2 depicts the original seismic signal and the FWHT coefficients computed for earthquake and non-earthquake incidents. The calculated FWHT coefficients provide a visual representation of the signal in both the time and frequency domains. Several statistical time-domain and frequency-domain features extracted based on FWHT coefficients are shown in Table 1.

TABLE 1: Features derived from seismic signal

No	Feature	Domain
1.	Mean	Time
2.	Median	Time
3.	Standard Deviation	Time
4.	Variance	Time
5.	Kurtosis	Time
6.	Skewness	Time
7.	Root mean square	Time
8.	Dominant frequency	Frequency
9.	Frequency amplitude	Frequency
10.	Signal energy	Frequency

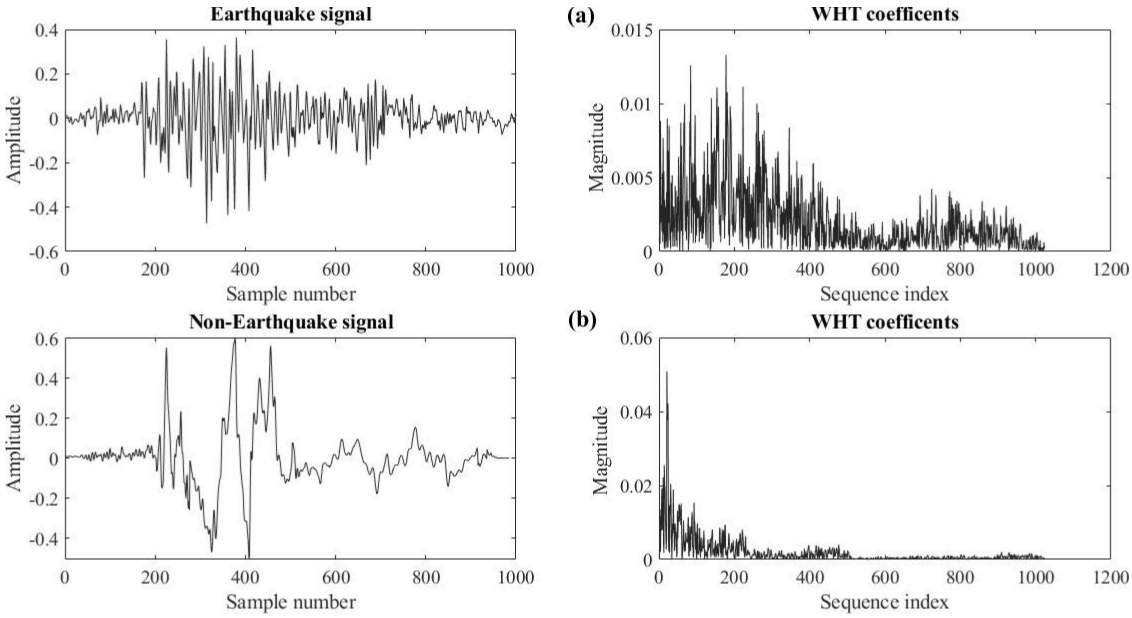

Fig. 2: Actual seismic signal and FWHT coefficients: (a) Earthquake (b) Non-Earthquake

PCA creates new components that store the most valuable information of the features by capturing a high variance. The eigenvalue-one metric is used in the analysis to assess various significant components to be preserved. Thus, we retained all of the components having eigenvalues above 0.5. Each component accounts for single unit of variance as an individual variable. Accordingly, components having eigen- values above 0.5 represented a greater variance over their contribution as independent variables and were chosen for further classification. Table 2 depicts the resultant principal components (PC1 to PC6) with eigenvalues above 0.5 and variance corresponding to individual extracted features. The PCA has downsized 10 features to 6 meaningful features.

TABLE 2: Result of features reduction using PCA

Features	**Eighenvalues**	**Variance**
Mean	0.927	6.58
Variance	0.856	6.02
Kurtosis	0.712	5.44
Skewness	0.689	4.86
Dominant frequency	0.906	6.49

Based on the identified features, the Random Forest classifier is implemented at the FC Layer to classify an event as earthquake or non-earthquake. The accurate classification of seismic data for earthquake detection is a major focus in our proposed model, the performance of the Random Forest classifier has been compared to the performance of other familiar classification algorithms, namely Multilayer Perceptron (MLP) [18], Naive Bayes [19], and C4.5 decision tree (DT) algorithm [20], using the WEKA 3.8 tool. The performance of various classifiers is assessed using performance metrics such as sensitivity, specificity, precision, and accuracy, as shown in Table 3. These performance indicators are inferred from the confusion matrix [21], which contains true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. The confusion matrix of Random Forest classifier is presented in Table 4.

The results clearly show that among several classifiers, the Random Forest classifier yields the highest classification accu- racy of 92.66%, the highest sensitivity of 90.25%, the highest specificity of 88.50%, and the highest precision of 89.50%, makes it an effective classifier for seismic event classification. Further, figure 3 depicts the latency time, response time, and execution time of our proposed model for various numbers of seismic samples. Because of the limited computation load, the model has a lower latency, response, and execution time for smaller samples. Also, Table 5 compares the performance of different classifiers with respect to possible error values.

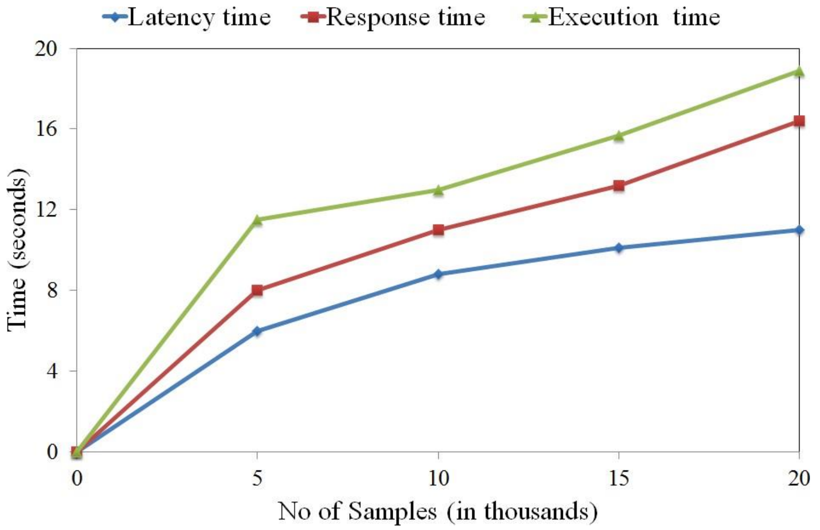

Fig. 3: Performance evaluation in terms of latency time, response time and execution time

TABLE 3: Performance analysis of different classifier

Classifier	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
Random Forest	90.25	88.50	89.50	92.66
Multilayer Perceptron	82.05	80.22	85.13	86.82
Naïve Bayes	72.08	69.66	75.80	77.38
C4.5 Decision Tree	67.43	65.71	68.24	70.24

TABLE 4: Description of confusion matrix for the classification result

Classified as	Earthquake	Non-Earthquake
Earthquake	92.90%	7.10%
Non-Earthquake	16.44%	83.56%

TABLE 5: Error performance of various classifiers

Error values	MLP	C3.5 DT	Naïve Bayes	Random Forest
Mean square error	0.1701	0.24	0.3108	0.2185
Root mean square error	0.3616	0.4899	0.3921	0.3582
Relative absolute error (%)	37.7682	53.2819	68.9937	30.8292
Root relative absolute error (%)	76.2302	103.2882	82.6609	70.5279

5 Conclusion

The current study proposed an iot-fog-cloud-based frame- work for earthquake prediction using seismic signals. The framework used the Fast Walsh-Hadmard Transform (FWHT) to extract information about the signal's nonlinearity. PCA- based strategies make it easier to retrieve discriminatory features associated with seismic events. The Random Forest classifier is used to classify a set of features as earthquake or non-earthquake event. The model's performance evaluation shows that Random Forest classifier outshines other classification algorithms with a maximum accuracy of 92.66%, a maximum specificity of 88.50%, a maximum sensitivity of 90.25%, and a maximum precision of 89.50%. This study has enormous potential for swift monitoring and detection of earthquakes, thereby mitigating their negative consequences. Furthermore, the proposed system yielded acceptable results, paving the way for a new knowledge domain in this area and assisting seismological departments as well as other disaster management authorities in developing informed interventions to prevent human life loss.

References

[1] CNN news about Nepal Earthquake, http://edition.cnn.com/2015/05/10/asia/nepal-earthquake-death-toll/

[2] http://www.weather.com/news/news/nepal-earthquake-everest-shrinks- kathmandu-lifted

[3] Alphonsa, A., and G. Ravi. ”Earthquake early warning system by IOT using Wireless sensor networks.” In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSP- NET), pp. 1201-1205. IEEE, 2016.

[4] Fischer, Joachim, Frank Ku¨hnlenz, Klaus Ahrens, and Ingmar Eveslage. ”Model-based Development of Self-organizing Earthquake Early Warning Systems.” Simul. Notes Eur. 19, no. 3-4 (2009): 9-20.

[5] Majhi, Santosh Kumar, Sk Sajeed Hossain, and Trilok Padhi. ”MFOFLANN: moth flame optimized functional link artificial neural network for prediction of earthquake magnitude.” Evolving Systems 11, no. 1 (2020): 45-63.

[6] Reddy, R. Gangadhar, M. Srinivasa Reddy, and P. R. Anisha. ”Identification of earthquakes using wavelet transform and clustering technologies.” International Journal of Civil En- gineering and Technology (IJCIET)(2017): 666–676.

[7] Cao, Kai, and Qunying Huang. ”Geo-sensor (s) for potential prediction of earthquakes: can earthquake be predicted by abnormal animal phenom- ena?.” Annals of GIS 24, no. 2 (2018): 125-138.

[8] Yamamoto, Yoshifumi, and Jack W. Baker. ”Stochastic model for earth- quake ground motion using wavelet packets.” Bulletin of the Seismolog- ical Society of America 103, no. 6 (2013): 3044-3056.

[9] Asencio–Corte´s, Gualberto, Antonio Morales–Esteban, Xueyi Shang, and Francisco Mart´ınez–A´ lvarez. ”Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure.” Computers & Geosciences 115 (2018): 198-210.

[10] Harmuth, Henning F. ”Applications of Walsh functions in communica- tions.” IEEE spectrum 6, no. 11 (1969): 82-91.

[11] Abbasi, Shuja A., and A. R. M. Alamoud. ”A novel complete set of walsh and inverse walsh transforms for signal processing.” In 2011 International Conference on Communication Systems and Network Tech- nologies, pp. 504-509. IEEE, 2011.

[12] Jimenez-Rodriguez, Luis O., Emmanuel Arzuaga-Cruz, and Miguel Ve´lez-Reyes. ”Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data.” IEEE Transactions on geoscience and remote sensing 45, no. 2 (2007): 469-483.

[13] Cutler, Adele, D. Richard Cutler, and John R. Stevens. ”Random forests.” In Ensemble machine learning, pp. 157-175. Springer, Boston, MA, 2012.

[14] Jia, Shaochun, Xiuzhen Hu, and Lixia Sun. ”The comparison between random forest and support vector machine algorithm for predicting - hairpin motifs in proteins.” Engineering 5, no. 10 (2013): 391.

[15] Subasi, Abdulhamit, Aysha Ahmed, Emina Alicˇkovic´, and Ahnaf Rashik Hassan. ”Effect of photic stimulation for migraine detection using random forest and discrete wavelet transform.” Biomedical signal processing and control 49 (2019): 231-239.

[16] National Research Institute for Earth Science and Disaster Prevention. Available online: http://www. kyoshin.bosai.go.jp (accessed on 1 February 2021).

[17] United States Geological Survey. Available at: https://www.usgs.gov/natural-hazards/earthquake-hazards/earthquakes (accessed on 1 February 2021).

[18] Singh, Kuldeep, and Sunil Agrawal. ”Comparative analysis of five machine learning algorithms for IP traffic classification.” In 2011 In- ternational conference on emerging trends in networks and computer communications (ETNCC), pp. 33-38. IEEE, 2011.

[19] Pop, Ioan. ”An approach of the Naive Bayes classifier for the document classification.” General Mathematics 14, no. 4 (2006): 135-138.

[20] Korting, Thales Sehn. ”C4. 5 algorithm and multivariate decision trees.” Image Processing Division, National Institute for Space Research–INPE Sao Jose dos Campos–SP, Brazil (2006).

[21] Powers, David MW. ”Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.” arXiv preprint arXiv:2010.16061 (2020)