Earth Science Informatics (2023) 16:1133–1149
https://doi.org/10.1007/s12145-023-00991-z

REVIEW

Machine learning for earthquake prediction: a review (2017–2021)

Nurafiqah Syahirah Md Ridzwan[1] · Siti Harwani Md. Yusoff[1]

Received: 6 December 2022 / Accepted: 4 March 2023 / Published online: 17 March 2023

Abstract
For decades, earthquake prediction has been the focus of research using various methods and techniques. It is difficult to predict the size and location of the next earthquake after one has occurred. However, machine learning (ML)-based approaches and methods have shown promising results in earthquake prediction over the past few years. Thus, we compiled 31 studies on earthquake prediction using ML algorithms published from 2017 to 2021, with the aim of providing a comprehensive review of previous research. This study covered different geographical regions globally. Most of the models analysed in this study are keen on predicting the earthquake magnitude, trend and occurrence. A comparison of different types of seismic indicators and the performance of the algorithms were summarized to identify the best seismic indicators with a high-performance ML algorithm. Towards this end, we have discussed the highest performance of the ML algorithm for earthquake magnitude prediction and suggested a potential algorithm for future studies.

Keywords Machine learning · Earthquake prediction · Review

Introduction

An earthquake is a sudden violent shaking of the ground, which occurs due to the sudden release of energy from the vibration in the Earth's crust, and typically causes great destruction and results in massive death. Generally, an earthquake occurs owing to the movement along faults, where the tensions cause bits of rock that stay together and later abruptly slip and release energy. However, human activity may also induce earthquakes. The magnitude (magnitude), focal depth (depth), epicenter location, and the distance from epicenter to an earthquake were the four seismic variables assessed during the earthquake. The first three variables were attributed to the seismic source, whereas the distance from the epicenter was primarily attributed to the impact of the observed location. Typically, an earthquake with a magnitude of five and above is capable of destroying infrastructures and causing civilians death. However, all of these catastrophic events can be minimized by predicting the occurrence of earthquakes.

Earthquake prediction typically requires the information regarding the magnitude, location, and time of occurrence. These can be categorized into short-term and long-term processes. The short-term process predicts an earthquake a few days or weeks before its occurrence and is particularly beneficial for the evacuation process. While the long-term prediction is based on the knowledge of past earthquake occurrences. Therefore, tectonic settings, geographical and historical earthquakes' records were studied to determine the occurrence of future earthquakes. Most research on earthquake prediction uses machine learning (ML) algorithms with seismic indicators based on Gutenberg Richter's law and Omori's law. Several studies have required the observation of earthquake precursors to predict future occurrences. Changes in nature, such as variations in the earth's electromagnetic field, radon gas concentrations, humidity, strange cloud formation, soil temperature, and crustal change are the potential precursors to precede earthquake occurrences (Cicerone et al. 2009).

Over the past decade, approaches and methods using ML have shown promising results in earthquake prediction. Compared to the conventional prediction techniques that have high false alarms, ML-based methods have a higher chance of predicting earthquakes accurately because of their

Communicated by: H. Babaie.

✉ Siti Harwani Md. Yusoff
aeharwani@usm.my

1 School of Aerospace Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia

high accuracy (Al Banna et al. 2020). These methods can help to prepare the disaster management teams and reduce causalities.

Hence, this study aims to provide a comprehensive review of past research on machine learning for earthquake prediction from 2017 to 2021. We also intend to study earthquake seismic indicators, because various indicators are used to predict earthquakes and observe the best seismic indicators that offer a high-performance ML algorithm.

Initially, we targeted research conducted in Southeast Asia, as it is located in the Pacific Ring of Fire, one of the regions that is frequently vulnerable to earthquakes. This region is located on active faults and an unstable tectonic geographical plate. However, we found that the studies conducted in Southeast Asia region are limited. Therefore, we have expanded our area of research globally.

In this paper, studies on earthquake prediction using ML algorithms were collected from reliable databases such as Scopus, ScienceDirect, and IEEE. After screening and review, 31 journal articles were selected. This study did not include conference papers because of peer review concerns. We have compiled and briefly summarised the work done by the 31 researchers in Table 1. This research is expected to facilitates the researchers to understand the role and importance of ML in predicting earthquakes based on the latest research.

Seismic Indicators

This section discusses the commonly and widely used seismic indicators to predict earthquake based on Gutenberg Richter's Law and Omori's Law.

Gutenberg-Richter's *a* and *b* values

These values are based on the geophysical law known as Gutenberg-Richter law. According to this law, number of earthquakes magnitude are distributed as in Eq. (1). Where, N_i is the total number of seismic events corresponding to magnitude larger or equal to M_i. While, b is the slope of the curve and a is the y-intercept.

$$\log N_i = a - \mathrm{b} M_i \tag{1}$$

Seismic energy release, *dE*

Seismic energy (dE) is the energy that keeps releasing from the ground in the form of small earthquakes. If the energy release stops, the phenomenon is known as quiescence, which may occur in the form of a major earthquake. The state of quiescence may lead to a reduction in the seismic rate of the region, and a decreasing b value.

$$dE^{\frac{1}{2}} = \frac{\Sigma\left(10^{(11.8+1.5M)}\right)^{\frac{1}{2}}}{T} \tag{2}$$

Time of *n* events, *T*

Time (T) in days, during which n number seismic events have occurred before earthquake.

$$T = t_n - t_i \tag{3}$$

Mean magnitude, M_{mean}

Mean magnitude (M_{mean}) refers to the mean value of n events

$$M_{mean} = \frac{\sum_i \mathrm{M}}{n} \tag{4}$$

Deviation from Gutenberg Richter law, *η*

It is the deviation η of actual data from the Gutenberg-Richter inverse law as

$$\eta = \frac{\sum (logN - a - bM)^2}{n-1} \tag{5}$$

Standard deviation of b value, *σb*

$$\sigma b = 2.3b^2\sqrt{\frac{\sum_{i=1}^{n}\left(M_i - mean(M)\right)^2}{n(n-1)}} \tag{6}$$

Magnitude deficit, ΔM

Magnitude deficit (ΔM) is the difference between the maximum observed earthquake magnitude and maximum expected earthquake magnitude.

$$M_{max.expected} = \frac{a}{b} \tag{7}$$

$$\Delta M = M_{max.actual} - M_{max.expected} \tag{8}$$

Model evaluation metrics

This section briefly summarizes the parameters used to assess the performance of the proposed method. True positives (TP) are the number of predicted earthquakes that match the number of recorded earthquakes. True negatives

Table 1 Summary of reviewed researches

Year	Research	Authors
2017	Presented a methodology to predict large magnitude earthquakes with a horizon of prediction of five days	(Fernández-Gómez et al. 2017)
	Applied ML to datasets from laboratory experiments that represent Earth's faulting and predict the timing of an earthquake	(Rouet-Leduc et al. 2017)
	Presented a new seismicity indicator-based Earthquake Early Warning System (EEWS) model to forecast earthquake magnitudes and locations using a combination of classification algorithm and mathematical optimization algorithm	(Rafiei & Adeli 2017)
	Used Artificial Neural Network (ANN) to predict the time and magnitude of an earthquake based on seismotectonic survey and faults	(Cheraghi & Ghanbari 2017)
	Predicted earthquake using Long Short-Term Memory (LSTM) by learning the spatio-temporal relationship of earthquakes in different locations	(Wang et al. 2017)
	Four ML techniques were applied to model the relationship between the seismic parameters and incoming earthquake occurrences	(K. M. Asim et al. 2017)
2018	Combined regression algorithms with ensemble learning in the context of big data to predict earthquakes magnitude	(Asencio-Cortés et al. 2018)
	Proposed a deep learning (DL) method for continuous earthquake prediction using historical seismic events	(Huang et al. 2018)
	Used tree-based algorithm to predict the earthquake by finding the earthquake precursor patterns	(Florido et al. 2018)
	Applied Backpropagation Neural Network (BPNN) to determine the optimal neuronal number in the hidden layer to predict earthquake magnitude	(Lin et al. 2018)
	Support Vector Regression (SVR) and Hybrid Neural Network (HNN) were applied to build an earthquake prediction model	(Khawaja M. Asim et al. 2018a)
	Proposed a prediction model using SVR with a particle filter-based algorithm to predict the magnitude and number of earthquakes	(Hajikhodaverdikhan et al. 2018)
	Presented a model to predict the aftershocks after an earthquake occurrence for the next five days using ANN	(Shodiq et al. 2018)
	Proposed an earthquake predictor system based on ensemble method by combining seismic indicator with Genetic Programming (GP) and AdaBoost	(Khawaja M. Asim et al. 2018b)
2019	Compared Feed-forward Neural Network (FFNN) and LSTM models in forecasting the trend of an earthquake	(Vardaan et al. 2019)
	Imitate seismic cycles on a laboratory scale and applied ML to predict the time and size of the earthquakes	(Corbi et al. 2019)
	Used various ML algorithms with slip distribution, locations of active faults, and Coulomb stress change to predict an earthquake's aftershock patterns	(Karimzadeh et al. 2019)
2020	Combined Functional Link ANN (FLANN) with a moth flame optimization algorithm to predict earthquake magnitudes	(Majhi et al. 2020)
	Proposed two ML algorithms to build an earthquake prediction system by predicting the b-value as a parameter that suggests the precursor to earthquakes	(Rahmat et al. 2020)
	Three ML algorithms are employed for short term prediction of earthquake occurrences	(Khawaja M. Asim et al. 2020)
	Compared Support Vector Machine (SVM), Naïve Bayes (NB) and multinomial logistic regression (MLR) performances in long term earthquake prediction	(Murwantara et al. 2020)
	Proposed a hybrid predictor for earthquake by combining ML and a conventional ground motion equation	(Kubo et al. 2020)
2021	Used six ML algorithms and estimated the accuracies of prediction for each component of ground motion for different tectonic environments	(Chanda et al. 2021)
	Proposed a Convolutional Neural Network (CNN) model with a 3D feature map to classify the earthquake magnitude using experimental data	(Bao et al. 2021)
	Investigated the effect of spatial parameters on four ML algorithms performances to predict the earthquake magnitudes	(Yousefzadeh et al. 2021)
	Developed a ML prediction model to forecast earthquake in a short term based on satellite data	(Xiong et al. 2021)
	Studied the pros and cons of various ML algorithms' performances in predicting earthquake	(Khosravikia & Clayton 2021)
	Built seven prediction models using different ML algorithm to predict earthquake magnitude	(Debnath et al. 2021)
	Compared ANN and Random Forest (RF) performances in predicting the earthquake acceleration, depth and velocity	(Essam et al. 2021)
	Proposed two hybrid ML algorithms to predict earthquake magnitude during fifteen days	(Salam et al. 2021)
	Modelled a LSTM network to predict and create a long-term earthquake catalogue	(Cao et al. 2021)

(TN) occur when there are no predicted or actual earthquakes. If the model predicts an earthquake when there is no actual earthquake, it is called a false positive (FP), and false negative (FN) indicates that the model predicts zero earthquakes when there is an earthquake. From these metrics, several well-known measures can be calculated. In particular, the sensitivity (Sn), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), Matthew's correlation coefficient (MCC), accuracy and R-score (R) were calculated. Their formulas are listed below,

$$S_n = \frac{TP}{TP + FN} \tag{9}$$

$$S_p = \frac{TN}{TN + FP} \tag{10}$$

$$PPV = \frac{TP}{TP + FP} \tag{11}$$

$$NPV = \frac{TN}{TN + FN} \tag{12}$$

$$MCC = \frac{(TP \bullet TN) - (FP \bullet FN)}{\sqrt{(TN + FN)(TP + FP)(TP + FN)(TN + FP)}} \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

$$R = \frac{(TP \bullet TN) - (FP \bullet FN)}{(TP + FN) \bullet (TN + FP)} \tag{15}$$

As for regression models, mean absolute error (MAE), mean squared error (MSE), and R-squared (R2) are the metrics commonly used to evaluate the models' performances. MAE is the average of the absolute differences between the predicted and actual values. It does not square the errors, so all errors are given equal weight regardless of their size. This makes MAE less sensitive to outliers and can be less affected by the presence of large errors in the data. MSE, on the other hand, measures the average squared difference between the predicted and actual values. It squares the errors, which means that larger errors are penalized more heavily than smaller errors. This makes MSE sensitive to outliers and can be affected by the presence of large errors in the data. It is a measure of the quality of an estimator—it is always non-negative, and values closer to zero are better. Both MAE and MSE provide a measure of the average magnitude of the errors in a set of predictions, but MSE gives a higher weight to larger errors, whereas MAE treats all errors as equal. R-squared is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. It ranges from 0 to 1, and a higher value indicates a better fit. In general, lower values for MAE and MSE and higher values for R-squared indicate a better-performing model.

Related works

In this section, the methods used in the selected 31 papers will be discussed separately based on the numbers of algorithms applied in their researches and whether they used hybrid or ensemble algorithm to predict earthquake. An example of a prediction process is shown in Fig. 1.

Earthquake prediction using a single algorithm

A shear laboratory experiment demonstrating an earthquake was set up, analysed, and RF was applied to predict the remaining time before the subsequent shear failure (Rouet-Leduc et al. 2017). The experiment was set up using a fault gouge material subjected to double direct shear in a two-fault arrangement. The shearing block displaces after a stick–slip frictional failure (laboratory earthquake), while the gouge layer dilates and intensifies simultaneously. As the materials undergo near failure, numerous tiny shear failures generate impulsive acoustic emissions. The unstable state ended when the shear stress and friction of the laboratory earthquake dropped dramatically as the gauge gouges disintegrated. To build the prediction model, a dataset consists of continuous acoustic time series data obtained from the fault was used as an input. The RF model is a weighted average of several decision trees. Each decision tree uses a series of decisions based on statistical data extracted from the time windows to predict the amount of time until the next failure. Features used in this study were selected recursively from a computed set of around 100 potentially significant statistical traits such as mean, variance, kurtosis, and autocorrelation in each time window.

A study done by Wang et al. (2017) has employed a Long Short-Term Memory (LSTM) algorithm was employed to predict earthquakes based on spatio-temporal correlation. The study was done in China where the region was divided into equal sized subregions and used a catalogue of earthquake magnitude greater than 4.5 from 1966 to 2016. To establish spatial and temporal correlation, they developed a two-dimensional matrix to represent seismic data with the same timestamp but from different regions and used it as an input to the LSTM layers. The LSTM network was built with 128 neurons in the LSTM layer, 256 and 64 for layer one and layer two of the dense network, and 9 for the output layer.

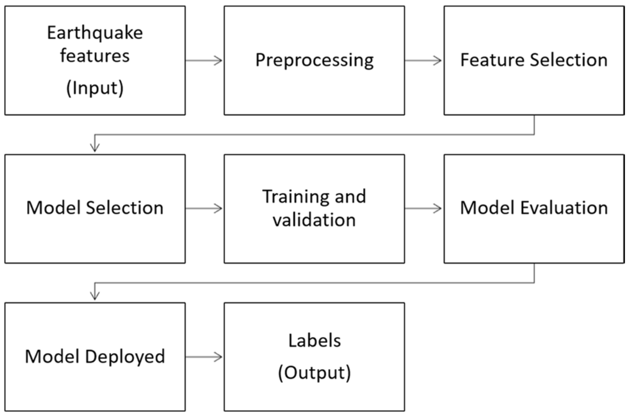

Fig. 1 A chart of earthquake prediction process using one machine learning algorithm

A dropout was applied to the output layer to avoid overfitting. The network has a SoftMax function as an activation function and RMSprop optimizer to optimize it. A decomposition method in which the subregions was group based on the same fault zone was used to improve the algorithm's performance.

An Artificial Neural Network (ANN) algorithm was proposed to predict the time and magnitude of an earthquake. The earthquake data was based in Iran from the year 1900 to 2008 (Cheraghi & Ghanbari 2017). As input, they employed earthquake time, depth, magnitude, and distance from the epicentre to the fault line. Initially, all of the data was normalised. The learning algorithm for the NN was the BP algorithm, while the error computation was mean squared error (MSE). The NN had two hidden layers, one with three neurons and the other with two neurons. The activation function was a sigmoid function, which added nonlinearity to the network. They used the estimated magnitude for energy prediction. The first, second, and output layers' momentum was set to 0.02, 0.01, and 0 accordingly, with the initial learning rate set to 0.001.

The work in Hajikhodaverdikhan et al. (2018) has proposed a model to predict earthquake magnitude and occurrences for the next month in Tabriz, Iran, using a Support Vector Regressor (SVR) enhanced with a particle filter-based algorithm. They used two types of data, which are meteorological data from Iran Meteorological Organization (IRIMO) and seismic data from Iranian Seismological Center (IRSC). The meteorological data include mean temperature, maximum temperature, average wind speed and precipitation. The model in SVR looks for a hyperplane that can divide a dataset into distinct parts based on classes. With the presence of noise, particle filters transform the state of a linear system into a random behaviour system. The parameters C, ϵ, and kernel scale are crucial in SVR. A large value of C will drop the model's generalisation, but its error performance will improve. The ϵ denotes a loss function whose lower value is desired; however, there is a possibility of overfitting if it is zero. Gaussian RBF was utilised as a kernel filter in this model. While a particle filter is used to select the main parameters in the particle filter-based support vector regression, in this study, parameter selection is based on particle weight in this filer, which is done by calculating the probability density function.

An ANN model was presented by Shodiq et al. (2018) to predict the aftershock of an earthquake five days after the occurrence based on automatic clustering. The data on Indonesia from 1910 to 2017 was obtained from the USGS and BMKG, with a total of 82,850 earthquake events. Seven seismic indicators were used as input parameters. The clustering process means the large dataset will be divided into smaller data groups. Three steps were taken in this process where firstly finding the best number of clusters, in which the authors implemented the valley tracing and hill-climbing algorithms. The number of clusters that worked best was six. The data were then clustered into six clusters using a Hierarchical K-means algorithm. Aftershocks were predicted using a neural network. Trial and error were used to determine the number of hidden layers and neurons. With a learning rate of 0.1, the optimal combination comprised two hidden layers, each with thirty-two neurons.

Research done by Lin et al. (2018) presented a Back-propagation Neural Network (BPNN) model to predict earthquakes with $M > 5$ with depths of 300 km and determined the optimum neuronal number in each hidden layer of the model. They used Philippine real-time series of data from 2000 to 2011 and tectonic plate slip rate data. The data was retrieved from Central Weather Bureau (CWB). The initial weights and biases of BPNN were set randomly with values of 0 to 1. An initial BPNN model was created, having two hidden layers and data from 2000 to 2010 as an input. The best predictions were made with 10 neurons for each hidden layer and a learning rate of 0.83. BPNN is embedded with the same initial weights as the prior model, as well as plate's slide rate, and the model is then used to forecast future earthquakes. The following data used is from 1990 to 1999 and 2011 to 2014 to train the model. The initial BPNN was then trained with data from 1990 to 2014 with a learning rate of 0.33, which is the best accuracy. To evaluate the model, the correlation coefficient, standard deviation, and MSE were calculated.

A Convolutional Neural Network (CNN) was used to predict the magnitude of an earthquake for the next 30 days in Taiwan using historical seismic events (Huang et al. 2018). Data was obtained from the China Seismic Information net between 1st January 1970 and 25th May 2016. The magnitude threshold was set at $M = 6$. Two dropout layers was applied to prevent overfitting and the activation function used was ReLu. The model produced the best result with a learning rate of 0.001 and a momentum value of 0.9 was used to accelerate convergence of the model.

A tree-based algorithm was employed to find precursors pattern in the historical data involving large magnitude earthquake events (Florido et al. 2018). They applied this model to seven different datasets, from different regions, namely the Alboran Sea and Western Azores-Gibraltar fault in the Iberian Peninsula, Santiago, Talca, Valparaiso and Pichilemu in Chile and Tokyo, Japan. Seismic features derived from the Gutenberg-Richter's b-value are used to study the seismic patterns. The datasets were taken from the authors' previous research. The model goes through a training process that involved the act of clustering, grouping, building a precursor tree, pattern extraction, and pattern selection. Clustering and grouping processes are both dependent on parameters K and A, therefore, a thorough search has been performed around the core. Parameter K stands for the number of clusters used in the clustering phase and A is the group size used in the grouping phase. K-means algorithm was applied during the clustering process. After the grouping was created, a tree-based model called the precursor tree was constructed and the best K precursors were extracted from this tree.

An ANN algorithm was employed as an earthquake prediction model in terms of earthquake acceleration, depth, and velocity, in Terengganu, Malaysia (Essam et al. 2021). Data collected from six seismological stations in Terengganu for two channels, provided from Malaysian Meteorological Department, were used to develop the model. For each ground motion parameter in each channel, the data were divided into six sets, with each set using data from separate groups of five stations for training and one station for testing. Earthquake depth can be described accurately univariate, that is, using only the appropriate output parameter, earthquake depth, as the input parameter. Earthquake acceleration and velocity cannot be univariately modelled accurately, and therefore earthquake depth is included as input parameters.

In order to predict an earthquake magnitude based on electromagnetic signals, Bao et al. (2021) proposed a CNN algorithm to build the prediction model. This study used electromagnetic signals data collected from 1st January 2017 to 1st January 2021 with a total of 6936 samples. They analyse data collected every 27 days to predict if there will be an earthquake on the 28th day. For the model structure, a High-Dimensional-Feature-Extraction (HDFE) block with a 3D feature matrix as its input feature and a Temporal-Correlation block consists of four convolution unit were created to obtain a deep relationship between sensed data and seismicity. In addition, noise modelling and SMOTE oversampling technology were used to enhance samples due to the imbalanced sampling. Finally, the model was evaluated using a large amount of data collected from the suggested inductive electromagnetic sensor.

In Cao et al. (2021), LSTM was used to predict earthquake pattern based on a three-dimensional strike-slip fault model. A three-dimensional visco-elastic–plastic finite element model was constructed. The long-distance earthquake cycle of a single fault was simulated and then calculated the stress evolution of the fault. The fault model was constructed with the depth of the mid-lower crust and upper mantle model of 100 kms. The fault element has a width of 2 km to model a 90-degree-dip fault with the fault depth being 20 km, and the element was a particular type of element with strain-softening elastoplastic. The fault simulation elastoplastic material had a crust on the outside, and the lower crust and higher mantle were approximated by Maxwell's rheology viscosity. After being integrated for about 100,000 years, the model was running in a quasi-steady state until the regional stress patterns stabilised, and the stress fluctuated around the background stress field due to the earthquakes. Although the LSTM layer has a high ability to study time series, its ability to adapt may be insufficient. To enhance learning capabilities, fully-connected layer was built on top of a single LSTM layer. In addition, the dropout of the LSTM layer was tuned to avoid overfitting and the layers were constructed with 100 neurons placed in the hidden layers, and one neuron was placed in the output layer.

Table 2 summarises the data sources, input features and prediction variable used in the studies mentioned above.

Earthquake prediction by comparing two or more algorithms

Several ML algorithms such as Naïve Bayes (NB), K-Nearest Neighbors (KNN), SVM and RF together with the slip distribution, Coulomb stress change, and locations of neighbouring active faults was applied to predict the aftershock patterns of the Kermanshah earthquake (Karimzadeh et al. 2019). The aftershocks data was retrieved from Iran Seismological Network from 12 November 2017 to the end of September 2018. This study was approached as a binary classification problem, and the location of aftershocks with a magnitude larger than 2.5 was predicted. A nonlinear and linear inversion approach was used to calculate the slip distribution. The slip distribution, friction coefficient, and Skemton's coefficient were used to determine the Coulomb stress change, which described the faults and neighbouring faults triggered by the earthquake. The geographical coordinates of the aftershocks were used to create a binary grid map, then NB, KNN, SVM, and RF classifiers were used to predict whether any aftershocks occurred on each grid. When Coulomb stress change and slip distribution are represented, these models simply predicted zero or one.

Table 2 Summary of used data sources, input features, output variables for one algorithm

Data type/sources	Input (Features)	Output (Prediction)	References
Laboratory	Continuous acoustic time series data	Time to failure	(Rouet-Leduc et al. 2017)
USGS	Time, latitude, longitude, magnitude	Earthquake occurrence	(Wang et al. 2017)
Not stated	Earthquake timing, focal distance, fault distance, magnitude	Time and magnitude of earthquake	(Cheraghi & Ghanbari 2017)
Meteorological data: IRIMO, IRSC	Mean temperature, maximum temperature, average wind speed and precipitation	Earthquake magnitude and occurrences for next month	(Hajikhodaverdikhan et al. 2018)
USGS & BMKG	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, $M_{expected}$, μ,5 increments of b-value, Maximum magnitude, Dynamic GR's law	Earthquake aftershock five days after the occurrence	(Shodiq et al. 2018)
CWB	Occurred time, epicentre, depth, slip rate	Earthquake magnitude	(Lin et al. 2018)
China Seismic Information	Time, latitude, longitude, magnitude,	Earthquake magnitude for the next 30 days	(Huang et al. 2018)
Not stated	T, M_{mean}, $dE^{\frac{1}{2}}$, η, μ, c, M_{def}, a and b value, maximum magnitude, probability of earthquake occurrence, 5 increments of b-value,	Next day earthquake occurrence	(Florido et al. 2018)
Seismograph	Earthquake acceleration, depth, and velocity	Earthquake acceleration, depth, and velocity	(Essam et al. 2021)
Electromagnetic sensor	51 features of electromagnetic disturbances	Earthquake magnitude	(Bao et al. 2021)
Geodynamic Simulation	Quasi-periodic earthquake magnitude	Earthquake pattern	(Cao et al. 2021)

An earthquake trend was predicted by comparing the Feed-forward NN (FFNN) and LSTM (Vardaan et al. 2019). Data for this study was collected from the Indian subcontinent region, Afghanistan, Tajikistan, Thailand, Laos, Vietnam, Malay Peninsula, and several provinces in China. In this study, the feed-forward NN used 2 hidden layers with 20 and 60 neurons, respectively, and all nodes having a sigmoid activation function. RMSprop is used as the learning role. For the LSTM, the model has 2 hidden layers with 40 hidden units in each LSTM cell and a 15 steps BP through time. To avoid overfitting, a drop out layer was inserted between the 2 hidden layers. The Agadrad algorithm is used to reduce the RMS loss.

A study done by Rahmat et al. (2020) has proposed an earthquake prediction model by using Extreme Learning Machine (ELM) and Deep Learning (DL), by predicting the b-value as a seismic indicator. This study used earthquake catalogue retrieved from the International Seismological Centre (ISC) Sumatra–Andaman region, in the region of 92°-106° East Longitude (EL) and 6.5° South Latitude (SL)—8° North Latitude (NL), from January 1973 to November 2014. The data is divided into 12 months or 1-year, where each of these is used to predict the next year's earthquakes. A total of 31 input nodes, 1024 hidden nodes, and 1 output node are designed to build the ELM structure, while DL have three hidden layers with 31 nodes each. ReLU and Sigmoid activation is used on DL structure. Data were processed with weights and biases using the appropriate activation function during the training process. Therefore, the output of the ELM and DL network are compared with the desired target output until the error is acceptable. During the testing process, the best weight is used, and new data is provided during this phase, which is then analysed at each layer until a network output called prediction is achieved. The results of this prediction will be utilised to predict the b-value, which is considered to be an earthquake precursor.

A comparison of machine learning performance in medium to long-term prediction using machine learning algorithms, such as MLR, SVM and NB has been presented in Murwantara et al. (2020). The data used in this study was collected from the USGS, Incorporated Research Institution for Seismology (IRIS), National Oceanic and Atmospheric Administration (NOAA), European-Mediterranean Seismological Centre (EMSC), International Seismological Centre (ISC), Istituto Nazionale di Geofisica e Vulcanologia (INGV), GeoForschungZentrum (GFZ), Indonesia Tsunami Early Warning System (InaTEWS), Global Historical Earthquake Archive (GHEA), and Badan Meteorologi, Klimatology dan Geofisika (BMKG) Indonesia from 1900 to 2018. The retrieved dataset consists of time, date, latitude, longitude, magnitude and depth. Earthquakes are predicted in two groups of 10 and 30 years. To predict the magnitude of the earthquake, the prediction was factorized into two factors. Earthquake power is derived from the latitude and longitude first, and then the magnitude is predicted based on a combination of location and depth. The depth of earthquake was predicted by factorizing the opposite of the magnitude prediction.

Three ML algorithms which are SVM, ANN and RF has been employed to predict earthquake magnitude in Khawaja M. Asim et al. (2020). The earthquake catalogue used in this study was collected from the Euro-Mediterranean Seismological Centre (EMSC) with the duration from 1924 to 2018. A de-clustering method was used on the data to eliminate the aftershocks and all earthquakes below magnitude 3.0. A set of 13 seismic attributes was used to develop the prediction model. In this study, the ANN had three layers with 20 neurons in the hidden layer. As for the SVM, the dataset is projected into a higher dimension to produce non-linear separability using a radial basis kernel. The number of trees selected for RF was 20. The models were constructed to predict earthquakes for various magnitude threshold of 3.0, 3.5, 4.0, and 4.5, and with prediction period of 5 days, 7 days, 10 days, and 15 days.

A study by Yousefzadeh et al. (2021) presented SVM, Decision Tress (DT), Shallow Neural Network (SNN) and Deep Neural Network (DNN) to predict the earthquake's magnitude for the next seven days in Iran. They used the catalogue from January 1973 to July 2019 with magnitude 3 and above, based in Iran, and collected from USGS and IIEES. The input data used to construct the models consisted of 16 seismic parameters, latitude, longitude, depth and a fault density parameter calculated based on Kernel Density Estimation Function. Both SNN and DNN were calibrated during model construction to provide a high level of generalization while reducing overfitting. To reduce the effect of overfitting, Weight Decay parameter were applied for SNN and Dropout for the DNN. The number of layers and nodes, dropout rate, activation function, and weight decay are adjusted for both models. They are constantly adjusted, trained, and validated to ensure high performances. For SVM, the calibration is best when using RBF kernel. Iterating over different value ranges generated the C parameter and the kernel width. In the calibration procedure for DT, the Trials parameter were tuned, which controlled the number of boosting rounds.

The study in Debnath et al. (2021) has employed several ML algorithms which are BN, RT, SL, RF, LMT, ZeroR, and LR to predict earthquake magnitudes in India. In this study, the data was obtained from USGS from the period of 1900 to 2000. The study focused on six states in India that are considered high risk cities namely the Andaman & Nikobar, Gujarat, North India, North East India, UP Bihar and Nepal. The prediction is made based on three earthquake magnitude range namely Mild Earthquake, Moderate Earthquake and Fatal Earthquake. A Bayesian network (BN) is classified as a probabilistic graphical model, which means it has nodes and directed edges. This model includes both relationships conditionally dependent and conditionally independent variables. Three components are needed to create the Bayes Network: random variables, a conditional relationship, and probability distributions. Simple logistic (SL) regression is extremely similar to linear regression, except the dependant attribute in simple logistic regression should be nominal rather than measurable. One purpose of simple logistic regression is to determine the probability of a certain nominal attribute value associated with a measurement attribute. The Logistic Model Tree (LMT) is a combination of decision tree and logistic regression. This algorithm is built based on tree information of previous model. It is type of decision tree in which the leaves provide a piecewise form of linear regression using linear regression models. The LogitBoost technique generates each node of a tree using the LR model. ZeroR focused on the relevant target and ignores all other factors. The majority class is predictable, although the ZeroR algorithm lacks predictive capabilities. It constructs a frequency table for the target variable and select the most common value from it. The algorithm is helpful in creating a baseline performance that can be used as a reference point for other classification algorithms. Logistic regression (LR) is applied when the target variable corresponds to a particular category type. It is a probability model-based forecasting algorithm and a form of linear regression that employs the sigmoid function, which is a sophisticated cost function. Between 0 and 1, the logistic regression hypothesis is most common. It is a linear regression-like equation. As a result, regression is a part of this method. A logistic function is used to carry out the logistic regression. It computes a logistic model's parameters.

Three ML models; ANN, RF and SVM was proposed to predict ground-motion intensity in terms of PGA and PSA at three different periods (Khosravikia & Clayton 2021). This study used a database of 4528 ground motion recordings from 374 earthquakes recorded at seismic stations in Texas, Oklahoma, and Kansas since 2005. The main features used in this study were moment magnitude, hypocentral distance and averaged shear wave velocity up to 30 m depth. For the ANN model, a hidden layer size of 4 with log sigmoid activation function was selected based on the sensitivity analysis, and the predictive and output parameters are normalised to their minimum and maximum values. The RF model used k-fold cross-validation analysis to solve the overfitting error of 180 trained models. The dataset was divided into k subsets using k-fold cross-validation. RF with 120 trees, and in-bag-fraction of 3/4, was selected to predicted the output based on the results from the 180 models. This method balances between model accuracy and model simplicity. The third model, SVM, trained 1500 different models for each

Table 3 Summary of used data sources, input features, output variables for methods using two or more algorithms

Data type/sources	Input (Features)	Output (Prediction)	References
Iran Seismological Network	Slip distribution, Coulomb stress change, and locations of neighbouring active faults	Aftershock pattern	(Karimzadeh et al. 2019)
Not stated	Historical earthquake occurrences	Earthquake trend	(Vardaan et al. 2019)
ISC	B-value	B-value as a precursor for next year earthquake occurrence	(Rahmat et al. 2020)
USGS	Time, date, latitude, longitude, magnitude and depth	Earthquake location, depth, magnitude	(Murwantara et al. 2020)
EMSC	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, σb, M_{def}, T_r, z, δm, a and b value, probability of earthquake occurrence	Earthquake magnitude	(Khawaja M. Asim et al. 2020)
USGS & IIEES	T, M_{mean}, $dE^{\frac{1}{2}}$, η, μ, c, ΔM, maximum magnitude, probability of earthquake occurrence, 5 increments of b-value, a and b value, latitude, longitude, depth,fault density	Earthquake magnitudes	(Yousefzadeh et al. 2021)
USGS	20 attributes including time, latitude, longitude, depth	Earthquake magnitude with different ranges	(Debnath et al. 2021)
Seismograph	Moment magnitude, hypocentral distance, averaged shear wave velocity over the top 30 m of soil	PGA and PSA at different period	(Khosravikia & Clayton 2021)

output variable. Given that all the SVM models are trained, it is observed that εl of 0.15, Cl of 5, and γl of 0.5 results in higher accuracy without overfitting problems, which ensures that the model has generalization capabilities for future data.

A summary of the data sources, input features and prediction variable used in the studies of earthquake prediction by comparing several algorithms is displayed in Table 3.

Earthquake prediction with hybrid/ensemble algorithms

A new methodology was presented by Fernández-Gómez et al. (2017) to predict large-magnitude earthquakes for a period of five days by applying imbalanced classification techniques and ensemble learning. The study was conducted in four Chilean cities, Santiago, Valparaiso, Talca, and Pichilemu, and used imbalanced datasets consists of the seismic activities in the cities. In the first stage, the dataset was rebalanced using a pre-processing algorithm and then processed by a classification algorithm to ensure an accurate result. All the classifiers created were called simple classifiers. High performance algorithms were selected and combined with various pre-processing methods to improve the prediction results. In the second stage, the ensembles were generated based on the best simple classifiers to help boost their performances. The ensembles were generated in iterations until all simple classifiers were selected or a perfect ensemble was found.

A neural Earthquake Early Warning System (NEEWS) model, which uses a combination of four classification algorithms (CA) and a mathematical optimization algorithm (OA) was introduced by Rafiei & Adeli (2017) to forecast the earthquake magnitude and location weeks before occurrence. They compiled an earthquake catalogue for the southern region of Californian from 1932 to 2016. The earthquake catalogue was then used to calculate eight indicators of earthquake seismicity. The four CAs that were used to develop the NEEWS model consisted of Neural Dynamic Classification (NDC), Probabilistic Neural Network (PNN), Enhanced Probabilistic Neural Network (EPNN), and Support Vector Machine (SVM), while the OA was the Neural Dynamic Optimization of Adeli and Park (NDAP). CAs are used to forecast the range of maximum earthquake magnitude, and OA is to estimate the longitude and latitude of the earthquake location in a particular region and time lag number. The time lag number is a multiple of a predetermined time resolution representing the time difference between the present and a predetermined period in the future for earthquake prediction. A period of month, two-week, or a week can be used as a time resolution. In this study, the time lag number was derived using a half-month time resolution and a magnitude threshold of 4.5, 5.0, 5.5, 6.0, and 6.5.

The performance of several algorithms, Piecewise Recurrent Neural Network (PRNN), Random Forest (RF), Recurrent Neural Network (RNN) and Linear Programming Boosting (LPBoost) were evaluated when predicting earthquakes with magnitudes larger than 5.4 based on the calculated seismic parameters (K. M. Asim et al. 2017). The earthquake catalogue was taken from the United States Geological Survey (USGS) and the Centre for Earthquake Studies, focusing on the Hindukush region. A total number of 11,137 events ranging between January 1976 and December 2013, were used in this study. Eight seismic indicators representing the seismic state and ground potential were used to predict the earthquakes. For the PRNN, the Levenberg–Marquardt backpropagation (LMBP) algorithm was used to train the network as it is faster than the standard backpropagation (BP). Two hidden layers with 12 neurons each and a tan-sigmoid transfer function for layer one and log-sigmoid for layer two were used to build the network. The RNN was built similar to the PRNN except that it has six and seven neurons in each layer and the ability to store the internal state as a directed cycle existed between the units. The RF was created with 50 numbers of decision trees. LPBoost is a combination of many tree classifiers, where each classifier was linearly added.

A method to predict earthquake magnitude within seven days by using a regression algorithm with ensemble learning in terms of big data catalogue has been proposed (Asencio-Cortés et al. 2018). Amazon Web Services (AWS) was used as the cloud-based platform to run the methodology proposed in this paper. The 1 GB big data catalogue consists of earthquake data in California from 1970 to 2017 and was divided into a matrix with a cell size of 0.5×0.5 (latitude and longitude). The cells generated 16 seismic features and 27 regression datasets. The selected datasets are the cell with at least 500 events and one or more earthquakes with $M > 5$. The 27 cells and 16 seismic features have been stored in Amazon Redshift in a new table. Four ML algorithms; generalized linear model (GLM), gradient boosting machine (GBM), deep learning (DL) and RF, and a RF-based stacking ensemble were used to create the regressor model and were executed in H2O library.

In Khawaja M. Asim et al. (2018a), SVR and Hybrid Neural Network (HNN) based classification system was built to predict earthquake with $M > 5$. The prediction was made on Hindukush, South California and Chile using data from USGS between January 1980 to December 2016. Sixty seismic features relative to every earthquake occurrence such as Gutenberg-Richter Law, foreshock frequency, total recurrence time, seismic energy release and seismic rate changes were computed. A two-step feature selection approach is used in the suggested methodology. The Maximum Relevance and Minimum Redundancy (MRMR) technique was employed to select the most useful features. The selected features are then fed into the SVR. The trend predicted by SVR is then sent into the next step of the prediction model, ANN, along with seismic features. SVR-output has been added to ANN as a feature to pass on the knowledge learned through SVR. Following SVR, the dataset is subjected to three layers of neural networks and combined with Enhanced Particle Swarm Optimization (EPSO). In place of SVR-output and feature set, the output of each ANN is used as an input to the next ANN. Each ANN layer's weight modifications are also passed on to the next ANN, so that the next ANN does not have to start learning from scratch. The inclusion of EPSO serves to optimise the weights of ANN, which tends to get stuck in the local minima.

Another study done by Khawaja M. Asim et al. (2018b) has proposed an earthquake predictor system named EP-GPBoost to predict earthquake prior to 15 days before occurrence by combining Genetic Programming (GP) and Adaptive Boosting (AdaBoost), building an ensemble algorithm. Seismic dataset was obtained from USGS between January 1980 to December 2016 covering Hindukush, Chile and South California. The searching potential of GP with the boosting capabilities of AdaBoost created a robust classifier. The development of the GP is aided by boosting, which involves the evolution of many GP strings per class that operate as a single class classifier. The evolution of the GP is carried out by employing boosting for weight update. For each class, boosting is used to produce P number of GP programmes. The generated results are summed together, and the larger value from a weighted total of the outputs of GP strings generated for each class is used to determine a class label for a test instance. For all three regions, the EP-GPBoost performs wonderfully, especially in reducing false alarm production.

In the study done by Corbi et al. (2019) the timing and size of a laboratory earthquake were predicted using the Gradient Boosted Regression Trees (GBRT) algorithm. They simulated several seismic cycles in a laboratory-scale subduction zone. The model simulated earthquakes with magnitude, $6.2 < M < 8.3$ by creating both partial and full margin ruptures, similar to real subduction zones with a coefficient of variation in recurrence intervals of 0.5. The laboratory scale earthquake was set up using a gelatine wedge as the overriding plate, under thrusted at constant rate by a 10° dipping, flat rigid plate as the subducting plate. The analogue megathrust integrates two velocity weakening patches with the same size and friction separated by a velocity strengthening patch to represent asperities. The gelatine on sandpaper produced velocity-weakening and gelatine on plastic contacts produced the strengthening behaviour. The model exhibits stick–slip behaviour after an initial stress build-up phase, with periods of stress build-up broken by the spontaneous nucleation of frictional instabilities propagating at the gelatine–plate contact. These instabilities represent an earthquake. The model generates ruptures across single or twin asperities, the number of which is proportional to the barrier to asperity length ratio. The GBRT model predict the time to failure (TTF) based on 94 features characterizing surface deformation measured in a laboratory mode known as geodetic signals data. The model was trained individually for each shifting training window of N seismic and single subsequent cycles for TTF at nine target points parallel to the trench.

In Majhi et al. (2020), a Functional Link ANN (FLANN) with a moth flame optimizer (MFO) was presented to predict earthquake magnitudes. The dataset used in this study was obtained from USGS and covers global earthquake events with six seismic indicators as input features. There are no hidden layers in the FLANN where a nonlinear function gains nonlinearity. They tested standard BP, least-square optimization, gradient descent, LMBP, and MFO as learning algorithms to evaluate the one performed the best. These algorithms were used to determine the model's optimal weight. Earthquakes with magnitude greater than or equal to 5.5 has been selected from the dataset. The time and date properties were then combined into one attribute. All of the attributes were normalized and subsequently expanded.

A hybrid method of machine learning and conventional ground-motion prediction equation (GMPE) has been proposed to predict the intensity of ground motions caused by earthquakes in Japan (Kubo et al. 2020). The dataset used in this study was constructed by combining data provided by National Research Institute for Earth Science and Disaster Resilience (NIED): ground-motion records observed by K-NET and KiK-net, site information from Japan Seismic Hazard Information Station (J-SHIS) and earthquake source information provided by F-net from 1997 to 2017. Since the regression equations are studied based on the geophysical background of ground motions, conventional GMPEs are stable for extrapolation or low data-density sections. Nevertheless, because the ground-motion model is strongly connected by a particular functional form of the regression equations, these GMPEs are inflexible. Nonparametric ML approaches, such as Random Forest or ERT, on the other hand, are very easily adaptable, although its prediction capabilities are inconsistent and cannot be proven for extrapolation or low data-density sections. A hybrid predictor that includes basic GMPE prediction followed by ERT was constructed to combine the benefit of both approaches. ERT is a Random Forest-derived tree-based ensemble ML method that fits multiple decision trees on different data subsamples and combines them to determine the output to improve predicted accuracy and control overfitting. In terms of tree splitting, ERT differ from RF in which RF determines the optimal split among a random group of variables, while ERT selects a node split at random. Regardless of the minor increase in bias, ERT can reduce variance. The number of trees for ERT was set at 1000.

In the work done by Xiong et al. (2021), an Inverse Boosting Pruning Trees (IBPT) model was developed for a short-term earthquake prediction. This work used 1371 earthquakes data with magnitude 6 and above, collected from 1st January 2006 to 25th December 2013, obtained from the Atmospheric Infrared Sounder (AIRS) on NASA's spacecraft Aqua and NOAA. The parameters generated from the data are surface skin temperature, atmospheric temperature at the Earth's surface, water vapour mass mixing ratio at the surface, total integrated column ozone burden, retrieved total column co, retrieved total column CH_4, ARIS outgoing longwave radiation flux, clear-sky outgoing longwave radiation flux, land surface temperatures and Outgoing Longwave Radiation (OLR). Standard features and time series-based features were generated for comparison. To generate the standard features, the data were classified using the scalable K-means technique and subjected to the Elbow method to identify the number of groups of the original continuous dataset that should be divided into. Next, k-means clustering were applied for various values of k, such as altering k from 5 to 20 clusters and the total within-cluster sum (wss) of squares for each k was calculated. For time series-based data, a sliding time window was performed, and Dynamic Time Warping (DTW) was used for clustering analysis and to determine the demarcation point of each segmentation period. From the pre-processing, four set of datasets were generated. The IBPT is an ensemble model, which combines an Adaboost variation with pruning decision trees for classification. Decision tree was used as the boosting base estimator. A decision dump often poses an imbalance problem when an entire tree has a high variance. As a result, trimming the tree was performed to increase the generalization capability of the model. While finding the best-pruned tree, all the training samples were used, let the decision tree grow to its full potential, and then trimmed part of the tree's branches using the cost-complexity pruning method. An inverse boosting structure was constructed with the pruned trees and updated weights. Then the procedures were repeated until the maximum number of trees is reached.

Two hybrid machine learning models namely FPA-ELM and FPA-LS-SVM were proposed to predict the earthquake magnitude (Salam et al. 2021). The data used in this study was retrieved from SCEC covering Southern California from 1st January 1950 to 31st May 1978. FPA-ELM is a hybrid of the flower pollination algorithm (FPA) and the extreme learning machine (ELM). FPA has been used to solve various non-linear problems with excellent results. Since ELM is based on a feed-forward neural network, it is a simple algorithm. With only one hidden layer, the data goes in one direction. In the classification and regression fields, ELM is a well-known and well-accepted technique. ELM solves the overfitting problems and long-running time concerns as well as achieving high accuracy and speed results. FPA-ELM was built when the FPA was used to optimise the ELM in order to increase the accuracy. The LS-SVM (Least Square Support Vector Machine) algorithm is a modified version of the

Table 4 Summary of used data sources, input features, output variables for methods with hybrid/ensemble algorithms

Data type/sources	Input (Features)	Output (Prediction)	References
Not stated	Imbalanced datasets of seismic activities	Earthquake magnitude for 5 days period	(Fernández-Gómez et al. 2017)
SCEDC	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, $M_{expected}$, μ, c	Earthquake magnitude and location	(Rafiei & Adeli 2017)
USGS	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, $M_{expected}$, μ, c	Earthquakes with magnitude larger than 5.4	(K. M. Asim et al. 2017)
NCEDC	Latitude, longitude and magnitude of earthquake	Earthquake magnitude within 7 days	(Asencio-Cortés et al. 2018)
USGS	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, σb, ΔM, T_r, z maximum magnitude, probability of earthquake occurrence, a and b value	Earthquake magnitude higher than 5.0	(Khawaja M. Asim et al. 2018a)
USGS	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, σb, ΔM, T_r, z maximum magnitude, probability of earthquake occurrence, a and b value	Earthquake prior to 15 days before occurrence	(Khawaja M. Asim et al. 2018b)
Laboratory	Geodetic signals data	Timing and size of a laboratory earthquake	(Corbi et al. 2019)
USGS	Date, time. Latitude, longitude, depth, magnitude	Earthquake magnitudes	(Majhi et al. 2020)
Seismograph	The epicentral distance D, moment magnitude Mw, event depth H, top depth to the layer whose S-wave velocity is 1,400 m/s, average S-wave velocity	Intensity of ground motion	(Kubo et al. 2020)
Satellite data: AIRS NASA NOAA	Ten different infrared and hyperspectral measurements	Earthquake occurrences	(Xiong et al. 2021)
Southern California Earthquake Center	T, M_{mean}, $dE^{\frac{1}{2}}$, β, η, $M_{expected}$, μ	Earthquake magnitude during 15 days	(Salam et al. 2021)
Seismograph	Historical data of ground motion duration	Ground motion duration	(Chanda et al. 2021)

SVM. Although LS-SVM simplifies the SVM approach, it still requires kernel parameters, which are critical in regression situations. FPA is used to optimise LS-SVM in selecting the best LS-SVM parameters.

Six ML algorithms; KNN, SVM, RF, NN, DT and AdaBoost were compared for a ground motion prediction of every component of different tectonic environments in Chile (Chanda et al. 2021). There were 3899 instances in-slab earthquakes, 6676 instances interface earthquakes and 138 instances crustal earthquakes dataset collected from the seismic station used in this study. The dataset was divided into ten sets of 10% each for testing in which tenfold validation was used. For KNN, the weighted value of K used in this study is set to 30.

Table 4 summarises the data sources, input features and prediction variable used in the studies mention above.

ML performances

In this section, the performances of the algorithms mentioned in the previous section will be discussed. A summary of the used algorithms and their performances, along with the input features is shown in Table 5.

Rouet-Leduc et al. (2017) applied RF to predict the time remaining before the subsequent shear failure based on shear laboratory experiment. The RF model consisted of 1000 decision trees and successfully predicted the time to failure with high accuracy and a R-score of 0.89. The model also accurately predicts failure during the entire laboratory earthquake cycle.

Wang et al. (2017) used the LSTM algorithm to predict earthquake based on spatio-temporal correlations. The model achieved an accuracy of 63.50% for one-dimensional input (temporal correlations) and 87.59% for two-dimensional input (spatio-temporal correlations) with decomposing method, for 5 × 5 subregions.

Cheraghi & Ghanbari (2017) predicted the time and magnitude of an earthquake by using ANN. Overall, the maximum error of the model was 3.5%, with average error of 0.5% for magnitude prediction. The earthquake timing was predicted with 10 days of error.

Hajikhodaverdikhan et al. (2018) predicted magnitude and occurrences of earthquake for the following month using SVR enhancement with a particle filter-based algorithm. The results predict the mean magnitude with an R-score of 0.96, and the number of earthquakes has the R-score of 0.78.

Table 5 Summary of used data sources, ML algorithms and their performances

Data type/sources	ML Algorithms	Performances	References
Single Algorithm			
Laboratory	RF	R-score 0.89	(Rouet-Leduc et al. 2017)
USGS	LSTM	Accuracy 87.59% (two-dimensional input)	(Wang et al. 2017)
Not stated	ANN	Maximum error- 3.5%	(Cheraghi & Ghanbari 2017)
Meteorological data: IRIMO	SVR	R-score 0.96 (Mean magnitude)	(Hajikhodaverdikhan et al. 2018)
USGS & BMKG	ANN	Accuracy between 56 and 72%	(Shodiq et al. 2018)
CWB	BPNN	MSE between 0.01 to 0.09	(Lin et al. 2018)
China Seismic Information	CNN	R-score 0.303	(Huang et al. 2018)
Not stated	Tree based algorithm	Accuracy 93.59% (Santiago)	(Florido et al. 2018)
Seismograph	ANN	NSE value: acceleration—0.8815 depth -0.9985 velocity -0.9933	(Essam et al. 2021)
Electromagnetic sensor	CNN	Accuracy 97.88%	(Bao et al. 2021)
Geodynamic Simulation	LSTM	MSE 0.08	(Cao et al. 2021)
Comparing two or more algorithms			
Iran Seismological Network	NB, KNN, SVM and RF	Accuracy 75% (RF)	(Karimzadeh et al. 2019)
Not stated	FFNN, LSTM	R-score of LSTM is higher by 59%	(Vardaan et al. 2019)
ISC	DL, EML	Accuracy 85.92% (EML)	(Rahmat et al. 2020)
USGS	MLR, SVM, NB	MAE 0.598473 (SVM)	(Murwantara et al. 2020)
EMSC	ANN, SVM, RF	MCC 0.861 (SVM)	(Khawaja M. Asim et al. 2020)
USGS & IIEES	DNN, SNN, SVM, DT	Accuracy 100% (DT) Sensitivity 97.7% (SVM)	(Yousefzadeh et al. 2021)
USGS	BN, RT, SL, RF, LMT, ZeroR, LR	Accuracy 99.94% (SL)	(Debnath et al. 2021)
Seismograph	ANN, RF, SVM	R-score 0.739 (SVM)	(Khosravikia & Clayton 2021)
With hybrid/ensemble algorithms			
Not stated	Ensemble learning	MCC 0.96	(Fernández-Gómez et al. 2017)
SCEDC	NDC, PNN, EPNN, SVM	Accuracy 99.4% (NDC-NDAP for earthquake magnitude)	(Rafiei & Adeli 2017)
USGS	PRNN, RF. RNN and LPBoost	Accuracy 65% (LPBoost)	(K. M. Asim et al. 2017)
NCEDC	GLM, GBM, DL, RF, and RF-based stacking ensemble	MAE 0.74 (RF)	(Asencio-Cortés et al. 2018)
USGS	SVR-HNN	Accuracy 90.6% (South California)	(Khawaja M. Asim et al. 2018a)
USGS	GP-Adaboost	Accuracy 86.6% (South California)	(Khawaja M. Asim et al. 2018b)
Laboratory	GBRT	R-score between 0.7 and 0.8	(Corbi et al. 2019)
USGS	MFOFLANN	RMSE of 0.0565	(Majhi et al. 2020)
Seismograph	ERT-GMPE	R-score 0.619	(Kubo et al. 2020)

Table 5 (continued)

Data type/sources	ML Algorithms	Performances	References
Satellite data: AIRS NASA NOAA	Inverse Boosting Pruning Trees (IBPT)	MCC 0.6581 R-score 0.6429	(Xiong et al. 2021)
Southern California Earthquake Center	FPA-ELM, FPA-LS-SVM	RMSE 0.565476 (FPA-LS-SVM)	(Salam et al. 2021)
Seismograph	KNN, SVM, RF, NN, DT and AdaBoost	R-score 1.00 (AdaBoost)	(Chanda et al. 2021)

Shodiq et al. (2018) proposed an ANN model to predict the aftershock of an earthquake five days after the occurrence. The result shows that for earthquakes with a magnitude greater than 6, the model shows a better performance with accuracy between 56 and 72%.

(Lin et al. 2018) used a BPNN model to predict earthquake with M > 5 with depths of 300 km and determined the optimum neuronal number in each hidden layer of the model. As a result, the optimal neuronal number in each hidden layer is 10, the MSE varied from 0.01 to 0.09, and the standard deviation is 0.21.

Huang et al. (2018) conducted a study in Taiwan using CNN to predict earthquake magnitude for the next 30 days. After using the past 120 days of earthquakes to predict future occurrence, the model achieved an R-score of 0.303.

Florido et al. (2018) searched for precursor patterns in the historical data that involve large magnitude earthquake events by employing a tree-based algorithm. The model successfully predicted an earthquake the next day and achieved the highest accuracy of 93.59% for the dataset from Santiago.

Essam et al. (2021) proposed an artificial neural network (ANN) algorithm as an earthquake prediction model for earthquake acceleration, depth, and velocity. Compared with RF and ANN models, the proposed model exhibit good performance in forecasting earthquake acceleration with NSE of 0.8815, depth with NSE of 0.9985, and velocity with NSE of 0.9933 based on the analysis and evaluation of the findings using four predefined performance criteria.

Bao et al. (2021) applied CNN to predict earthquake magnitude based on electromagnetic signals experimental sensor. The findings of the experiment reveal that the CNN model performs well in earthquake magnitude prediction, with an accuracy of 97.88%.

Cao et al. (2021) applied LSTM to predict earthquake pattern based on a simulation of three-dimensional strike-slip fault model. The LSTM achieved a performance with MSE of 0.08.

Karimzadeh et al. (2019) predicted the aftershock patterns of the Kermanshah earthquake using several ML algorithms such as NB, KNN, SVM and RF. Of all the algorithms used in this study, RF achieved the highest accuracy with a percentage of 75%.

Vardaan et al. (2019) compared the feed-forward NN and LSTM to predict the trend of an earthquake. The LSTM achieved R-score of -0.252, which is 59% higher than FFNN.

Rahmat et al. (2020) proposed an earthquake prediction model by using Extreme Learning Machine and Deep Learning. Extreme Learning Machine performed with a success percentage of 85.92% compared to Deep Learning with 85.82%.

Murwantara et al. (2020) compared the performance of MLR, SVM and NB in medium- to-long term earthquake prediction. As a result, SVM outperforms other method for 30 years dataset with grouping for predicting earthquake magnitude. It shows the prediction accuracy as shown by MAE of 0.598473, in which explain that the prediction results of earthquake are precise compared to the other algorithm.

Khawaja M. Asim et al. (2020) used SVM, ANN and RF to predict earthquake magnitude in Cyprus for different magnitude threshold. For the magnitude 3.0, 3.5, and 4.5, RF perform the best for all prediction period with the highest MCC of 0.810 when predicting for within 15 days. Meanwhile, SVM showed the best for performance for predicting magnitude 4.0 with the highest MCC of 0.861 within 15 days.

Yousefzadeh et al. (2021) predicted earthquake magnitude for the next seven days in Iran using SVM, DT, SNN and DNN. The results showed satisfactory performances of DNN and SVM in predicting the classes of high magnitudes with sensitivity percentage of 95.5% and 97.7% respectively. However, the performance of DT was more promising in coping with events of both high and low magnitudes with an accuracy of 100%.

Debnath et al. (2021)employed several ML algorithms, namely BN, RT, SL, RF, LMT, ZeroR, and LR, to predict earthquake magnitudes in Andaman & Nikobar, Gujarat, North India, North East India, UP Bihar and Nepal. The result shows that SL achieved the highest accuracy of 99.94% in Andaman & Nikobar.

Khosravikia & Clayton (2021) predicted ground-motion intensity in terms of PGA and PSA at three different periods using ANN, RF and SVM. SVM achieved the highest R-score of 0.739.

Fernández-Gómez et al. (2017) used ensemble learning to predict the occurrence of large magnitude earthquakes for the next five days in four Chile cities (Santiago, Valpraiso, Talca and Pichilemu). The model showed a good performance by achieving an MCC of 0.96 in Talca.

Rafiei & Adeli (2017) used a combination of four classification algorithms (CA); NDC, PNN, EPNN and SVM and a mathematical optimization algorithm (OA); NDAP for predicting the magnitude and location of earthquake. The results showed that the combination of the NDC-NDAP model is the best with an accuracy of 99.4% and R-score of 0.94, followed by EPNN-NDAP with 99.3% accuracy and R-score of 0.93.

K. M. Asim et al. (2017) employed of PRNN, RF. RNN and LPBoost in predicting earthquakes with magnitude larger than 5.4. The algorithm's performances show that LPBoost has the highest accuracy percentage of 65%, followed by RNN with 64%. Although PRNN has the lowest accuracy with 58%, it produced the fewest false alarms.

Asencio-Cortés et al. (2018) proposed a method for predicting earthquake magnitude using four ML algorithms; GLM, GBM, DL and RF, and a RF-based stacking ensemble. All the algorithms were compared, and RF showed the best performance on average of all 27 datasets with a MAE of 0.74. GBM achieved a MAE of 0.77 on average and produced the highest accuracy in three datasets. GLM and DL had the lowest performances and MAE over 1.0. All RF-based algorithms had the same MAE of 0.76.

Khawaja M. Asim et al. (2018a) built an SVR and HNN based classification system to predict earthquake with $M > 5$. The SVR-HNN model was trained and tested, producing good results with a percentage accuracy of 82.7% for Hindukush, 84.9% for Chile and 90.6% for South California.

Khawaja M. Asim et al. (2018b) combined GP and AdaBoost to predict earthquake prior to 15 days before occurrence in three regions. The model achieved positive predictive value of 74%, 80%, and 84%, respectively, for Hindukush, Chile, and Southern California, implying that the ratio of false alarms is low. The model also achieved an accuracy of 78.7% for Hindukush, 84.5% for Chile and 86.6% for South California.

Corbi et al. (2019) simulated several seismic cycles in a laboratory-scale subduction zone and predicted the timing and size of a laboratory earthquake using GBRT. The results were promising, with R score between 0.7 and 0.8 for each target points.

Majhi et al. (2020) predicted earthquake magnitude by building a functional link ANN (FLANN) with a moth flame optimizer (MFO) prediction model. The MFOFLANN model achieved a RMSE of 0.0565.

Kubo et al. (2020) used a hybrid of ERT and conventional ground-motion prediction equation (GMPE) model to predict the intensity of ground motions in Japan. The proposed model achieved an R score of 0.619.

Xiong et al. (2021) predicted earthquake by employing an Inverse Boosting Pruning Trees (IBPT) model. The IBPT was the performed with the highest R-score value of 0.6429 and MCC of 0.6581.

Salam et al. (2021) built two hybrid machine learning models; FPA-ELM and FPA-LS-SVM, to predict the earthquake magnitude. FPA-LS-SVM outperformed the FPA-ELM model according to all compared criteria. Moreover, FPA-LS-SVM is the best in reducing the false alarm ratio in earthquake prediction and has achieved RMSE of 0.565476.

Chanda et al. (2021) employed KNN, SVM, RF, NN, DT and AdaBoost for ground motion prediction. Overall, Adaboost has the highest performance with R score of 1.00.

Discussion

Overall, majority of the algorithms used in the studies can be considered as high-performance algorithm as they have high accuracy (75%-100%), high R-score (closer to 1.00) and low error. For earthquake magnitude prediction, Decision Trees algorithm achieved the highest prediction accuracy with a percentage of 100%, based on the data and input features from Iran (Yousefzadeh et al. 2021). Others high performance algorithms with percentage accuracy higher than 90% are CNN, NDC-NDAP, SL, SVR-HNN, and Tree-based algorithm.

When predicting earthquake, the type of input features, the desired output variables, and the algorithm used for prediction should be taken into consideration. Although the same type algorithm was used to predict earthquake, the result will be different with varieties of input features and prediction variables. For example, five studies (Essam et al. 2021), (Cheraghi & Ghanbari 2017), (Shodiq et al. 2018), (Khosravikia & Clayton 2021) and (Khawaja M. Asim et al. 2020) employed ANN to predict earthquake and each of them achieved different types of result. Dataset size also played an important role because a small training dataset could lead to an underfitting.

Conclusion

Because earthquakes can result in large numbers of deaths, studies have been conducted to prevent these devastating consequences. One of the studies is using ML to predict earthquake as ML has higher prediction accuracy compared to the other methods. By applying ML algorithms to predict earthquake, preventative measures and precautions can be taken to reduce the negative impacts. Various ML algorithms were used to predict earthquakes, however, none of them are suitable for all types of prediction problems as there are many variables that need to be considered. Hence, this paper has reviewed the ML-based methodologies that were used to predict earthquake. A total of 31 papers were selected from academic databases from 2017 to 2021. The used features, prediction variables, and the performances of the algorithm have been discussed and summarize in tables. This study aims to help the researchers to understand the role and importance of ML in predicting earthquake based on the latest researches. Understanding the algorithms and how they work can help to build a model that could potentially solve all the earthquake prediction problems. ANN is one of the earliest algorithms used to predict earthquake and have shown good results. However, with the advancement in artificial intelligence, other algorithms surpassed its performance. Because most ML algorithms are tailored to a specific dataset or task, combining different ML algorithms and building a hybrid machine learning can improve the outcome by assisting one another in tuning, generalising, or adapting to new tasks which is suitable for predicting the various types of earthquake prediction problem.

Author contributions Siti Harwani Md Yusof provided the idea of writing a review paper, reviewed and edited the draft. Nurafiqah Syahirah Md Ridzwan performed the literature search, data analysis and wrote the initial draft.

Funding This work was supported by the "Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with Project Code: FRGS/1/2020/STG07/USM/03/3."

Data availability Not Applicable.

Code availability Not Applicable.

Declarations

Ethics Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing Interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Al Banna MH, Taher KA, Kaiser MS, Mahmud M, Rahman MS, Hosen ASMS, Cho GH (2020) Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges. IEEE Access 8:192880–192923. https://doi.org/10.1109/ACCESS.2020.3029859

Asencio-Cortés G, Morales-Esteban A, Shang X, Martínez-Álvarez F (2018) Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. Comput Geosci 115(September 2017):198–210. https://doi.org/10.1016/j.cageo.2017.10.011

Asim KM, Martínez-Álvarez F, Basit A, Iqbal T (2017) Earthquake magnitude prediction in Hindukush region using machine learning techniques. Nat Hazards 85(1):471–486. https://doi.org/10.1007/s11069-016-2579-3

Asim KM, Idris A, Iqbal T, Martínez-Álvarez F (2018a) Earthquake prediction model using support vector regressor and hybrid neural networks. PLoS ONE 13(7):1–22. https://doi.org/10.1371/journal.pone.0199004

Asim KM, Idris A, Iqbal T, Martínez-Álvarez F (2018b) Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification. Soil Dyn Earthq Eng 111(February):1–7. https://doi.org/10.1016/j.soildyn.2018.04.020

Asim, Khawaja M, Moustafa, SS, Niaz, IA, Elawadi, EA, Iqbal, T, Martínez-Álvarez, F (2020) Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in Cyprus. Soil Dynamics and Earthquake Engineering, 130(October 2019). https://doi.org/10.1016/j.soildyn.2019.105932

Bao Z, Zhao J, Huang P, Yong S, Wang X (2021) A deep learning-based electromagnetic signal for earthquake magnitude prediction. Sens, 21(13). https://doi.org/10.3390/s21134434

Cao C, Wu X, Yang L, Zhang Q, Wang X, Yuen DA (2021) Long Short-Term Memory Networks for Pattern Recognition of Synthetical Complete Earthquake Catalog. Sustain (Switzerland), 9–14. https://doi.org/10.3390/su13094905

Chanda S, Raghucharan MC, Karthik Reddy KSK, Chaudhari V, Somala SN (2021) Duration prediction of Chilean strong motion data using machine learning. J South Am Earth Sci, 109(October 2020). https://doi.org/10.1016/j.jsames.2021.103253

Cheraghi A, Ghanbari A (2017) Study of risk analysis and earthquake magnitude and timing prediction via tectonic and geotechnical properties of the faults and identifying risky areas in terms of seismicity in larestan city using artificial neural network. QUID: Investigación, Ciencia y Tecnología, No. Extra 1, 2017, Págs. 1137-1142, (1). Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=6158766&info=resumen&idioma=ENG. https://dialnet.unirioja.es/servlet/articulo?codigo=6158766. Accessed 10 Oct 2021

Cicerone RD, Ebel JE, Britton J (2009) A systematic compilation of earthquake precursors. Tectonophysics 476(3):371–396. https://doi.org/10.1016/j.tecto.2009.06.008

Corbi F, Sandri L, Bedford J, Funiciello F, Brizzi S, Rosenau M, Lallemand S (2019) Machine Learning Can Predict the Timing and Size of Analog Earthquakes. Geophys Res Lett 46(3):1303–1311. https://doi.org/10.1029/2018GL081251

Debnath P, Chittora P, Chakrabarti T, Chakrabarti P, Leonowicz Z, Jasinski M, Gono R, Jasińska E (2021) Analysis of earthquake forecasting in India using supervised machine learning classifiers. Sustainability (switzerland) 13(2):1–13. https://doi.org/10.3390/su13020971

Essam, Y, Kumar, P, Ahmed, AN, Murti, MA, El-Shafie, A (2021) Exploring the reliability of different artificial intelligence techniques in predicting earthquake for Malaysia. Soil Dyn Earthquake Eng, 147(May). https://doi.org/10.1016/j.soildyn.2021.106826

Fernández-Gómez, MJ, Asencio-Cortés, G, Troncoso, A, Martínez-álvarez, F (2017) Large earthquake magnitude prediction in Chile with imbalanced classifiers and ensemble learning. Appl Sci (Switzerland), 7(6). https://doi.org/10.3390/app7060625

Florido E, Asencio-Cortés G, Aznarte JL, Rubio-Escudero C, Martínez-Álvarez F (2018) A novel tree-based algorithm to discover seismic patterns in earthquake catalogs. Comput Geosci 115(March):96–104. https://doi.org/10.1016/j.cageo.2018.03.005

Hajikhodaverdikhan P, Nazari M, Mohsenizadeh M, Shamshirband S, Chau K (2018) Earthquake prediction with meteorological data by particle filter-based support vector regression. Engineering

Applications of Computational Fluid Mechanics 2060:679–688. https://doi.org/10.1080/19942060.2018.1512010
Huang JP, Wang XA, Zhao Y, Xin C, Xiang H (2018) Large earthquake magnitude prediction in Taiwan based on deep learning neural network. Neural Netw World 28(2):149–160. https://doi.org/10.14311/NNW.2018.28.009
Karimzadeh S, Matsuoka M, Kuang J, Ge L (2019) Spatial Prediction of Aftershocks Triggered by a Major Earthquake : A Binary Machine Learning Perspective. International Journal of Geo-Information Article 8(10):462
Khosravikia F, Clayton P (2021) Machine learning in ground motion prediction. Comput Geosci, 148(June 2020). https://doi.org/10.1016/j.cageo.2021.104700
Kubo H, Kunugi T, Suzuki W, Suzuki S, Aoi S (2020) Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. Sci Rep 10(1):1–12. https://doi.org/10.1038/s41598-020-68630-x
Lin J, Chao C, Chiou J (2018) Determining Neuronal Number in Each Hidden Layer Using Earthquake Catalogues as Training Data in Training an Embedded Back Propagation Neural Network for Predicting Earthquake Magnitude. IEEE Access 6:52582–52597. https://doi.org/10.1109/ACCESS.2018.2870189
Majhi SK, Hossain S, Padhi T (2020) MFOFLANN : moth flame optimized functional link artificial neural network for prediction of earthquake magnitude. Evolving Syst, 1997. https://doi.org/10.1007/s12530-019-09293-6
Murwantara IM, Yugopuspito P, Hermawan R (2020) Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data. Telkomnika (Telecommun Comput Electron Control) 18(3):1331–1342. https://doi.org/10.12928/TELKOMNIKA.v18i3.14756
Rafiei MH, Adeli H (2017) NEEWS: A novel earthquake early warning model using neural dynamic classification and neural dynamic optimization. Soil Dyn Earthq Eng 100(February):417–427. https://doi.org/10.1016/j.soildyn.2017.05.013
Rahmat B, Joelianto E, Afiadi F, Fandenza ADL, Kurniawan RA, Puspaningrum EY, Nugroho B, Kartika DSY (2020) Comparison of B-Value Predictions as Earthquake Precursors using Extreme Learning Machine and Deep Learning. Int Indonesia J 12(1):47–52. Retrieved from https://www.researchgate.net/publication/349517987_Comparison_of_B-Value_Predictions_as_Earthquake_Precursors_using_Extreme_Learning_Machine_and_Deep_Learning. Accessed 10 Oct 2021
Rouet-Leduc B, Hulbert C, Lubbers N, Barros K, Humphreys CJ, Johnson PA (2017) Machine Learning Predicts Laboratory Earthquakes. Geophys Res Lett 44:9276–9282
Salam MA, Ibrahim L, Abdelminaam DS (2021) Earthquake Prediction using Hybrid Machine Learning Techniques. Int J Adv Comput Sci Appl 12(5):654–665. https://doi.org/10.14569/IJACSA.2021.0120578
Shodiq MN, Kusuma DH, Rifqi MG (2018) Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia. INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION 2:37–43
Vardaan K, Bhandarkar T, Satish N, Sridhar S, Sivakumar R, Ghosh S (2019) Earthquake trend prediction using long short-term memory RNN. Int J Electric Comput Eng 9(2):1304–1312. https://doi.org/10.11591/ijece.v9i2.pp1304-1312
Wang Y, Wang Z, Cao Z, Lan J (2017) Deep learning for magnitude prediction in earthquake early warning. IEEE Trans Emerg Top Comput 8(1):148–158. https://doi.org/10.1109/TETC.2017.2699169
Xiong P, Tong L, Zhang K, Shen X, Battiston R, Ouzounov D, Iuppa R, Crookes D, Long C, Zhou H (2021) Towards advancing the earthquake forecasting by machine learning of satellite data. Sci Total Environ, 771. https://doi.org/10.1016/j.scitotenv.2021.145256
Yousefzadeh M, Hosseini SA, Farnaghi M (2021) Spatiotemporally explicit earthquake prediction using deep neural network. Soil Dyn Earthquake Eng, 144(February). https://doi.org/10.1016/j.soildyn.2021.106663

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.