

Early detection of earthquake magnitude based on stacked ensemble model

Anushka Joshi ^{*}, Chalavadi Vishnu, C Krishna Mohan

Department of Computer Science and Engineering, IIT Hyderabad, Telangana, India

ARTICLE INFO

Keywords:

Strong motion
Prediction
Magnitude
Machine learning

ABSTRACT

A new machine learning model, named, EEWPEnsembleStack has been developed for predicting the magnitude of the earthquake from a few seconds of recorded ground motion after the arrival of the P phase. The testing and training dataset consists of 2360 and 591 strong-motion records from central Japan recorded by the Kyoshin Network. Eight parameters that are well correlated with the magnitude have been used for training and testing of the model. Feature ablation study using several models shows that a minimum mean absolute error of 0.42 has been obtained for the case when the model has been trained by using all parameters rather than by a single parameter. The model ablation study indicates that among all individually trained single models, the minimum error has been obtained for a Decision Tree regression model. However, the error is minimized when all machine learning models have been together utilized in the EEWPEnsembleStack model for the training purposes. The EEWPEnsembleStack model has been used to predict a 6.3 magnitude earthquake by using its 21 records from various stations that lie within 50 to 150 km epicentral distance. The predicted magnitude from the developed model using weighted magnitude prediction is obtained as 6.4, which is close to the actual magnitude. The comparison of the predicted magnitude of this earthquake from the developed model with that predicted by using popular τ_c and P_d methods clearly indicates the suitability of the developed machine learning model over other conventional models.

1. Introduction

Earthquakes are the worst natural disasters in terms of loss of life and property. Predicting an earthquake magnitude before its major phase strikes any region is a challenging task for reducing damage due to an earthquake. Several earthquake early warning networks (EEWN) have been deployed worldwide, designed to issue warnings to the surrounding region before the arrival of the major destruction phase. Earthquake early warning signals allow people to take effective measures to reduce seismic hazards (Allen and Kanamori, 2003; Odaka et al., 2003; Allen et al., 2009; Behr et al., 2015; Atefi et al., 2017). The earthquake early warning systems (EEWS) are installed and are tested in many seismically active areas, including India (Chamoli et al., 2019), Taiwan (Wu and Teng, 2002; Chen et al., 2016), Japan (Kamigaichi et al., 2009), Turkey (Erdik et al., 2003), Italy (Zollo et al., 2009; Colombelli et al., 2020), California (Allen and Kanamori, 2003; Kuyuk et al., 2014; Kohler et al., 2018), China (Peng et al., 2011; Zhang et al., 2016; Wang et al., 2020), South Korea (Sheen et al., 2017), Mexico (Aranda et al., 1995), Bhucharest (Ionescu et al., 2007). Most of these networks predict the magnitude of the earthquake based on predominant periods or the

amplitudes of the first few seconds of P waves at single or multiple stations (Wu and Zhao, 2006; Wu and Kanamori, 2005; Böse et al., 2009; Brown et al., 2009; Kuyuk and Allen, 2013; Olivieri et al., 2008; Shieh et al., 2008; Wu et al., 2007). The magnitude prediction approaches in most of the recent Earthquake Early Warning (EEW) systems are based on the empirical relationship between P wave onset parameters and the magnitude of the earthquake (Zhu et al., 2021a). The empirical relationships used in most of the EEWS are based on simple linear regression relations (Kanamori, 2005; Wu and Zhao, 2006; Wu et al., 2006, 2007; Zollo et al., 2006; Wang et al., 2009; Lin et al., 2011, Festa et al. 2008).

These methods are based on estimating magnitude using a single parameter which is one of the major disadvantages as a single parameter contains less information than multiple parameters (Zhu et al., 2021b). Moreover, some parameters are affected by signal to noise (SN) ratio and epicentral distance (Carranza et al., 2015). It is seen that the estimation of magnitude by using conventional empirical relations has a serious problem of overestimation for small earthquakes and underestimation for large earthquakes (Zhu et al., 2021b). In this work, the machine learning method for EEW uses the initial phase of data related to the less destructive P phase. The important hidden information from data can be

* Corresponding author.

E-mail address: anushka.j@cs.iitr.ac.in (A. Joshi).

extracted using machine learning tools (Reyes et al., 2013). Machine learning helps to construct models and recognise patterns by a cognitive system using a data set.

Recent studies by Mousavi and Beroza (2020), Perol et al. (2018), Reddy and Nair (2013), Ochoa et al. (2017), Zhu et al. (2021), Zhang et al. (2021) have demonstrated that machine learning and deep learning methods have excellent potential for its use in magnitude estimation in EEW systems. However, machine learning models used for EEWs are primarily based on small to moderate events (Ochoa et al., 2017; Kuang et al., 2021; Zhang et al., 2021). The rapid determination of large earthquake magnitudes ($M_{JMA} \geq 6.5$) using machine learning in EEWs remains elusive (Zhu et al., 2021b). This paper presents an ensemble model consisting of several machine learning models, which is applicable for predicting 3.0 to 7.4 magnitude earthquakes. This paper explores the learning ability of the ensemble model for estimating magnitude based on selected features carefully extracted from the early 3 s of the P phase of the earthquake record. The impact of various features on the magnitude prediction ability of the model has also been studied in detail, together with the relationship of extracted features on the magnitude of the earthquake.

2. Data

The Japanese Islands lie at the junction of four major tectonic plates: the Pacific plate, the Philippine Sea plate, the North American plate and the Eurasian plate (Wei and Seno, 1998), as shown in Fig. 1. Seismic activity in Japan is mainly attributed to the collision and subduction zone between the Eurasian, Philippine Sea, North American, and Pacific plates. In this region, the Pacific and Philippine Sea plates subduct beneath the North American and Eurasian plates. In this region, the Pacific plate subducts beneath the North American plate at a rate of 8 cm/year, while the Philippine Sea Plate subducts beneath the North

American and Eurasian plates beneath Nankai Trough at an approximately 4 cm/year and 4–7 cm/year, respectively (Satake, 2015). These subductions cause strain accumulation on the plate boundaries and generate interplate earthquakes when the accumulated strain is suddenly released (Satake, 2015). The source mechanism of earthquake is mainly related to the type of plate boundaries with which it is associated. The Nankai Trough is among the best-known subduction zones in the world regarding the recurrence of large earthquakes (Satake, 2015). The characteristic features of an earthquake from this subduction zone are ground shaking and aftershock recording in a large area (Satake, 2015). In the present work, the earthquakes originating from the subduction and collision zone of the Philippine Sea Plate and North American Plate with the Eurasian plate have been considered for similarities in their characteristics. Kyoshin Network (K-NET) is a network of strong-motion seismographs installed at approximately 1,000 locations in Japan. The earthquakes recorded by stations of K-NET lying between 30° to 36° N and 128.5° to 138° E have been considered in this work. Fig. 2a shows the location of the epicentres of earthquakes that have occurred in this region from 1996 to 2021.

The location of recording stations that have recorded these earthquakes is shown in Fig. 2b. The time-series recorded by the stations of the K-NET of the National Research Institute for Earth Science and Disaster Prevention (NIED), Japan (Aoi et al., 2011) has been used for extracting features from the early 3 sec of the P phase. The identification of the P phase from the record in this work is based on the short-time-average through long-time-average trigger (STA/LTA) algorithm given by Allen (1978) using a threshold value of 2.0. Most EEWs installed worldwide use three seconds of the P wave records to predict the magnitude (Wu and Zhao, 2006; Wu et al., 2007). In this work, three-second records after the onset of the P phase have been used for training and testing the proposed ensemble model. A total of 2951 earthquake data recorded between January 1996 to August 2021 has

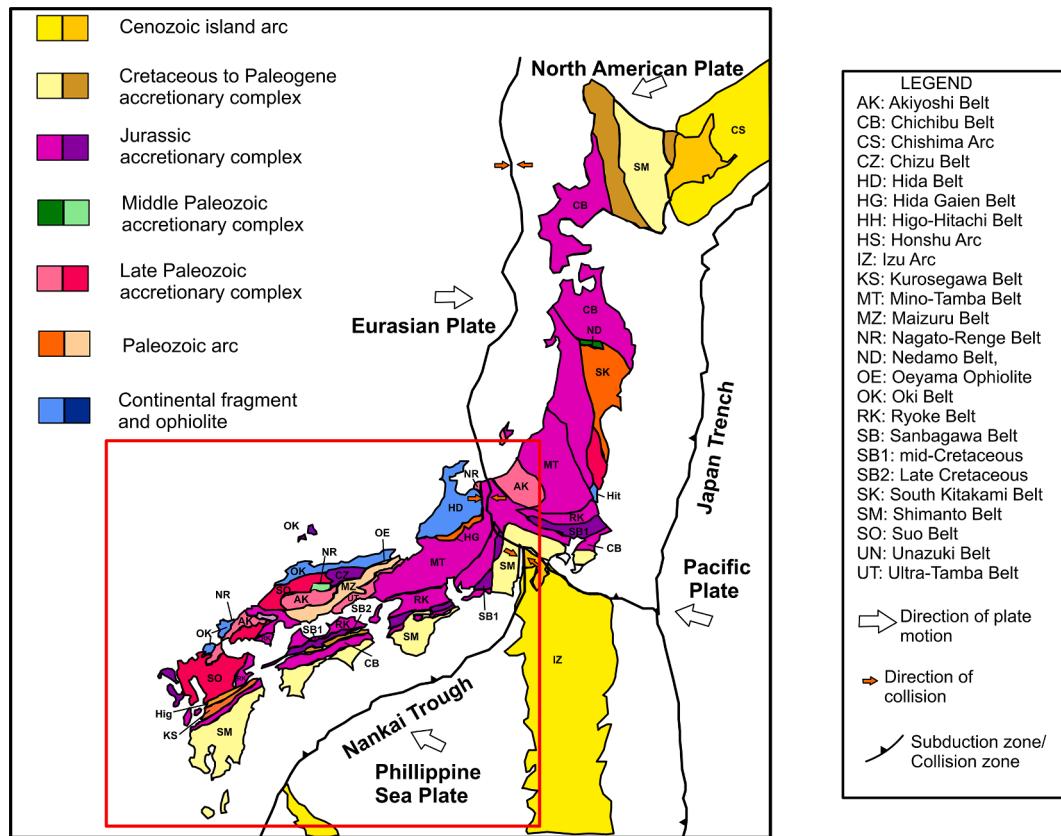


Fig. 1. Geology of Japan's island (modified after Wakita (2013)). The direction of plate motion was taken after Takla et al. (2013). The red rectangular boundary shows the study area used in the present work.

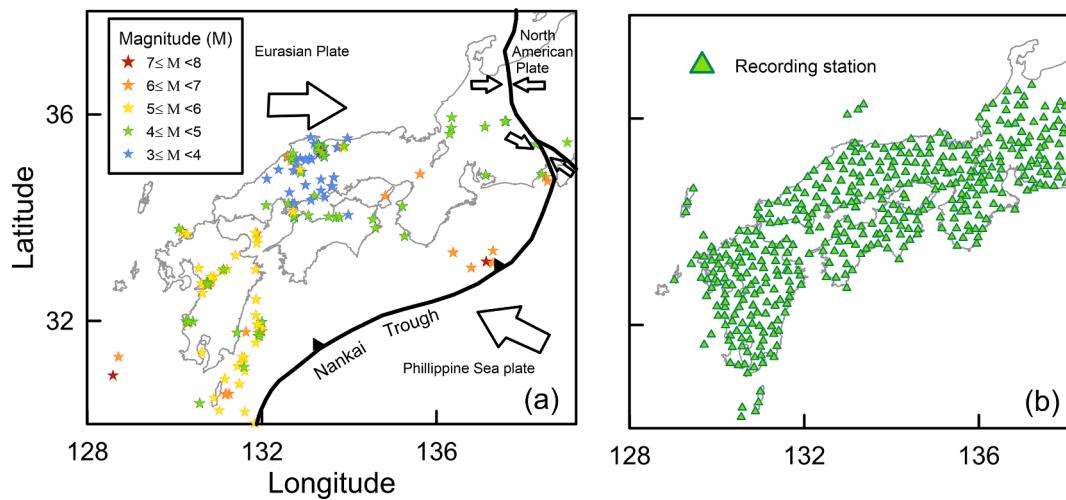


Fig. 2. (a) Location of epicentres and (b) recording stations used in the present work. The stations and epicentres are shown by solid triangles and stars, respectively. Tectonics of the region is taken after [Takla et al. \(2013\)](#).

been used in this work. Fig. 2a and b show earthquakes' locations and the recording stations corresponding to records used in the training and testing data sets. The data recorded within a hypocentral distance of 50 to 150 km has been used to clarify the P phase. The entire dataset consists of records from earthquakes of a magnitude of atleast 3. Eighty percent of the entire data set has been divided into training purposes, while the remaining 20% has been utilised for testing purposes. A total of 2360 earthquake records in the training data set and 591 in the testing data have been randomly selected from the entire data set and are shown in Fig. 3.

3. Methodology

Earthquake magnitude gives an idea about the size of an earthquake and is a prime concern in every EEW system. The earthquake early warning networks installed in different parts of the world predict the magnitude of an earthquake from the early few seconds of the ground motion using features like characteristics period (τ_c) given by Kanamori (2005), predominant period (τ_p) given by Allen and Kanamori (2003), velocity squared integral (IV2) given by Festa et al. (2008) and peak ground displacement (P_d) given by Wu and Kanamori (2005). In the present work, the popular earthquake early warning features like τ_c , τ_p , IV2, and P_d have been used for training and testing of the model. The features are defined as displacement squared integral (ID2) and the ratio of peak ground velocity with peak ground displacements (T_{vd}), have also

been utilised in the present work to train the model. These features are similar to those already defined by Festa et al. (2008). In this work, features extracted from the autocorrelation function (ACF) of displacement records have been utilised. The ACF has been previously used by Sinhal and Khatri (1983) for the discrimination of various sedimentary beds in seismic exploration. In this work, the ratio of the parameter of the ACF function at three lag with zero lag (ACF_1) and the ratio of the area on the positive and negative sides of the abscissa obtained from the displacement function (ACF_2) have been used to train the model. A total of eight features have been selected for the training and testing of the model.

Once the features have been extracted, their dependency on the magnitude of an earthquake can be determined by a correlation study which is an essential task before training any model for predictions. The correlation study is used to project the extent to which features are related to each other. The correlation of various features with the earthquake's magnitude has been determined using the Pearson and Spearman correlation matrix and is shown in Fig. 4a and b. The Spearman correlation can be used as a measure of the monotonic association, while the Pearson correlation is associated with the amount of linear relationship between two parameters (Schober et al., 2018). The value of a correlation coefficient of more than 0.1 to 1.0 amounts to a weak and very strong correlation (Schober et al., 2018). The value of Pearson and Spearman coefficients between different parameters used in this work is shown in Fig. 4a and b, respectively. The Pearson coefficient

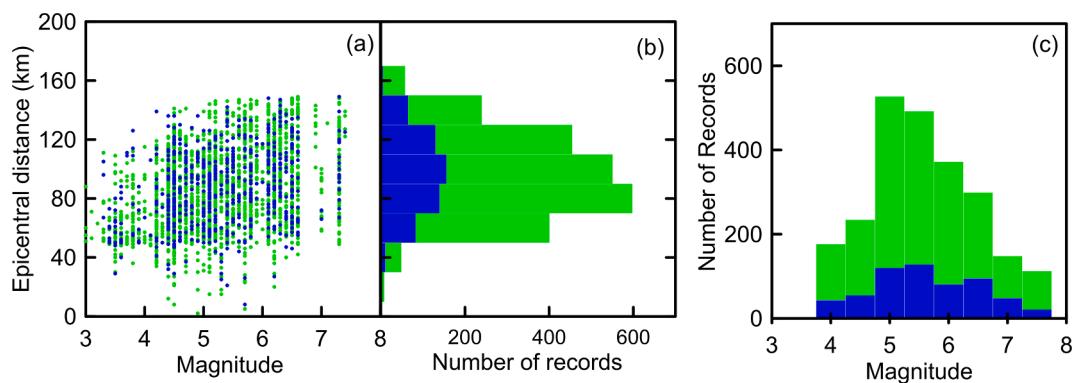


Fig. 3. (a) Distribution magnitude of earthquakes with respect to the epicentral distance of records used in the training and testing data sets, (b) distribution of earthquake records used in the training and testing data sets with respect to the epicentral distance, and (c) distribution of records used in the training and testing data sets with respect to the magnitude of the earthquake. The blue and green colors used in this figure represent values for testing and training data sets, respectively.

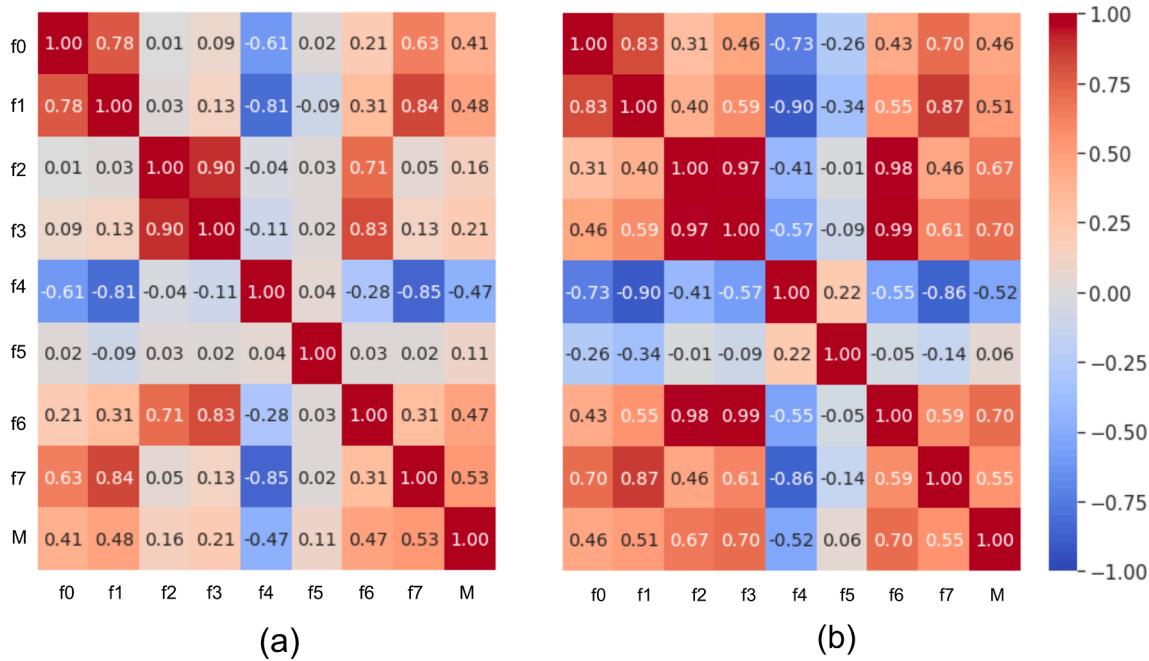


Fig. 4. (a) The Pearson, and (b) the Spearman correlation coefficient matrix heatmap between features, color coding, is from -1 (dark blue) to $+1$ (dark red). The symbol f0, f1, f2, f3, f4, f5, f6, f7, and M used in this figure represent parameters τ_p , τ_c , IV2, ID2, T_{vd} , ACF_2 , P_d , ACF_1 and actual magnitude, respectively.

between different parameters shown in Fig. 4a indicates that various parameters used in this work have weak (0.1 to 0.39) to moderate linear correlation (0.4 to 0.69) with the magnitude of the earthquake. However, the Spearman correlation coefficient between different parameters indicates the presence of a strong to moderate monotonic association with the magnitude of an earthquake. The parameter ACF_2 shows the presence of weak linear correlation with magnitude. This parameter shows a clear presence of monotonic association with other parameters, which have a strong monotonic and linear association with the magnitude of the earthquake. Hence in the present work, all eight parameters have been utilised for the feature ablation study for the training of the machine learning model.

Once the features and their correlation with the magnitude have been established and the model has been finalised, the next task is to reduce the overfitting of the model. In the present work, the K-fold cross-validation has been applied to reduce the overfitting of the model. K-fold cross-validation is based on the holdout concept, in which a single holdout concept is repeated across multiple folds of data. The fraction of data has been used for testing, and the remaining has been used for training the model. The process is repeated using 10 different values of fractions. The following basic steps have been used to obtain a 10-fold cross-validation (Gunes et al., 2017):

1. Divide training data into ten disjoint sets of equal sizes.
2. Prepare the holdout part of the data as the test set. Make ten holdout sets and use them one at a time.
3. Training the model on remaining data that has not been used for the holdout.
4. Test the trained model using a holdout set.
5. Measure the performance of K-fold cross-validation by taking an average of results obtained from ten holdout test sets.

Ensemble models in machine learning tools consist of different models and are generally used to optimise the prediction error and thus enhance the accuracy of the prediction model, which otherwise is not possible in individual models. It combines the results obtained from various models to generate a better result. The stacking of ensemble models allows the prediction generated by one layer to be used as input

by the other layer. The next layer gives the result based on the predictions generated by the previous layer. In the present work, an ensemble model named “EEWPEnsembleStack” has been designed in Python, which consists of many popular machine learning models. The grid search technique has been applied to determine the best hyperparameters of individual models used in the EEWPEnsembleStack by comparing 10-fold cross-validation results. The optimised set of hyperparameters from grid search has been selected using evaluation metrics such as mean and standard test scores. Various steps that are performed to obtain optimised hyperparameters for the model are shown in Fig. 5.

The EEWPEnsembleStack model incorporates models like AdaBoost, XGBoost, LightGBM regressor, Decision Tree, and Lasso regression, respectively. These models are defined in the following section.

3.1. AdaBoost

AdaBoost is a boosting algorithm introduced by Freund and Schapire (1997). In AdaBoost technique, regression is performed in a training set of x input data containing k features from n samples. The input data during testing is represented by (x, y) , where y is the actual output. The AdaBoost initialises the weights given by a hyperparameter named ‘learning_rate’ in the Python API, which is 0.01. The initial base estimator is selected, which has been used as the default initialisation with default max_depth. The weights are initialised for each data point. In this study, the iterative operation starts with iterators given by the ‘n_estimators’ hyperparameter, which is 100 for AdaBoost 1 and 250 for AdaBoost 2. The other hyperparameters were set as default, given in the Python API. The AdaBoost algorithm uses the linear loss function as a loss estimator for updating weights after each iteration. The final result is predicted by combining the predictions from individual ‘ m ’ models with the help of a weighted majority vote (Hastie et al., 2009) as given below:

$$M(x) = \text{sign} \left(\sum_{i=1}^m \alpha_i M_i(x) \right) \quad (1)$$

where $\alpha_i, i = 1, \dots, m$ are calculated from the Adaboost algorithm given by Hastie et al. (2009), and it also gives weights to the contribution of the respective model $M_i(x), i = 1, \dots, m$. In the AdaBoost algorithm, the

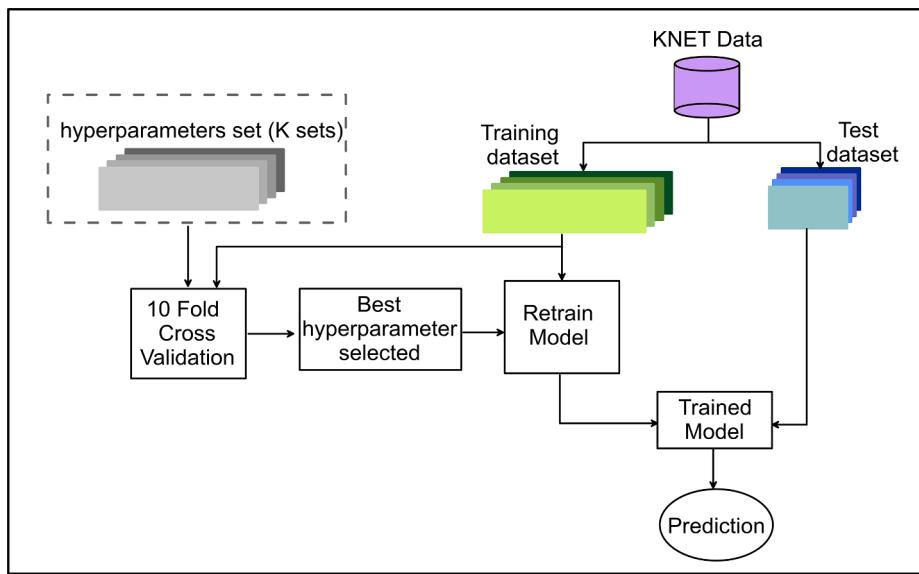


Fig. 5. The flowgraph represents various steps to obtain final optimised hyperparameters for the models using a 10-fold cross-validation step and grid search on hyperparameters.

sampling weight assigned to train the model is iteratively selected using steps defined by [Hastie et al., 2009](#).

3.2. Xgboost

XGBoost is an ensemble machine learning model which consists of ensembles of learning trees developed by [Chen and Guestrin \(2016\)](#). XGBoost is a sparse-aware algorithm that can handle missing and very low-frequency data. The XGBoost is an optimised gradient boosted tree and a regularised model which measures tree complexity while building them. The training loss used in the model is set by a hyperparameter objective function which represents a linear loss. The objective function of the XGBoost model is given below ([Chen et al., 2019](#)):

$$Obj = \sum_{i=1}^N L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

The function L shown in the above equation represents the linear loss function used in the XGBoost model. The complexity of the tree has been calculated by using the following equation ([Chen et al., 2019](#)):

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3)$$

In the above equation, λ represents the L2 normalisation hyperparameter, w_j represents the leaf weight, T represents the number of leaves and γ represents the hyperparameter to be multiplied by T . In the XGBoost model, optimal weight w_j^* are calculated iteratively by using the following minimum objective function:

$$\min Obj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (4)$$

In this equation, G_j and H_j are defined as.

$$G_j = \sum_{i \in I_j} g_i \quad (5)$$

$$H_j = \sum_{i \in I_j} h_i \quad (6)$$

The final value of the hyperparameter selected after grid search for the XGBoost model is given in [Table 1](#). The hyperparameters that are not shown in the [Table 1](#) have been used as default values provided by the

Table 1

Final optimised hyperparameter when all features are passed in the XGboost model.

hyperparameter	Value
max_depth	6
learning_rate	0.2
n_estimators	300
min_child_weight	2
reg_alpha	0.005
reg_lambda	6
min_child_weight	2

python API.

3.3. LightGBM regression

The LightGBM model is among a few machine learning models with fast training speed and less memory usage ([Wang and Wang, 2020](#)). XGBoost uses a level-wise tree growth method, while LightGBM uses a leaf-wise growth method in which the model provides the depth limit. Leaf-wise growth reduces significant errors and gives higher accuracy than the level-wise growth strategy ([Wang and Wang, 2020](#)). The objective function of LightGBM is the same as XGBoost, as shown in eq (2). The objective function used in the LightGBM Regression after applying Taylor series expansion is given by the following equation:

$$Obj_j^{(t)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] \quad (7)$$

The optimised hyperparameters used in LightGBM have been obtained after applying grid search. Two different models, namely, LightGBM 1 and LightGBM 2, have been prepared using different sets of hyperparameters obtained after applying an optimised hyperparameter grid search on the LightGBM model. The value of hyperparameter 'max_depth' in the LightGBM 1 and LightGBM 2 model is obtained as 6 and 3, respectively, after a grid search, while the value of hyperparameter 'Num_boost_round' obtained after a grid search is 300 and 1000 for LightGBM 1 and LightGBM 2 models, respectively. The other hyperparameters are the same as that provided in the default value of the Python API.

3.4. Decision Tree

The decision tree consists of a tree-like model (Bishop, 2006). Decision trees are referred to as classification and regression trees (CART) by Breiman et al. (1984). The root of the decision tree is at the top. The decision tree consists of an internal node, based on the internal node tree is split into edges. The final node is the leaf node from which no further split is made. The cost function of the decision tree used in regression is given below:

$$\text{Cost} = \sum (y_i - \hat{y}_i)^2 \quad (8)$$

where y_i is the actual value and \hat{y}_i is the predicted value. The tree performance is increased by pruning which involves removing branches with low feature importance. The optimised hyperparameter 'max_depth' value is 5 after applying grid search and 10-fold cross-validation step.

3.5. Lasso regression

Least absolute shrinkage and selection operator (Lasso) regression is used for controlling overfitting and feature selection of the model. Linear regression causes the problem of overfitting. Overfitting is the problem that arises when a model shows the perfect fitting of training data, while poor fitting is seen from the same model using the test data. The overfitting in the Lasso regression model is controlled by adding the following regularisation term in the error function (Bishop, 2006):

$$E_D(w) + \alpha E_w(w) \quad (9)$$

In the above equation, α , $E_D(w)$ and $E_w(w)$ represent the regularisation coefficient, data-dependent error and the regularisation term, respectively. The objective function used in the Lasso regression is given as (Tibshirani, 1996):

$$\text{Obj}_{\text{Lasso}} = \sum_{n=1}^N \left[y_n - x_n \cdot \sum_{j=1}^M w_j \right] + \alpha \sum_{j=1}^M |w_j| \quad (10)$$

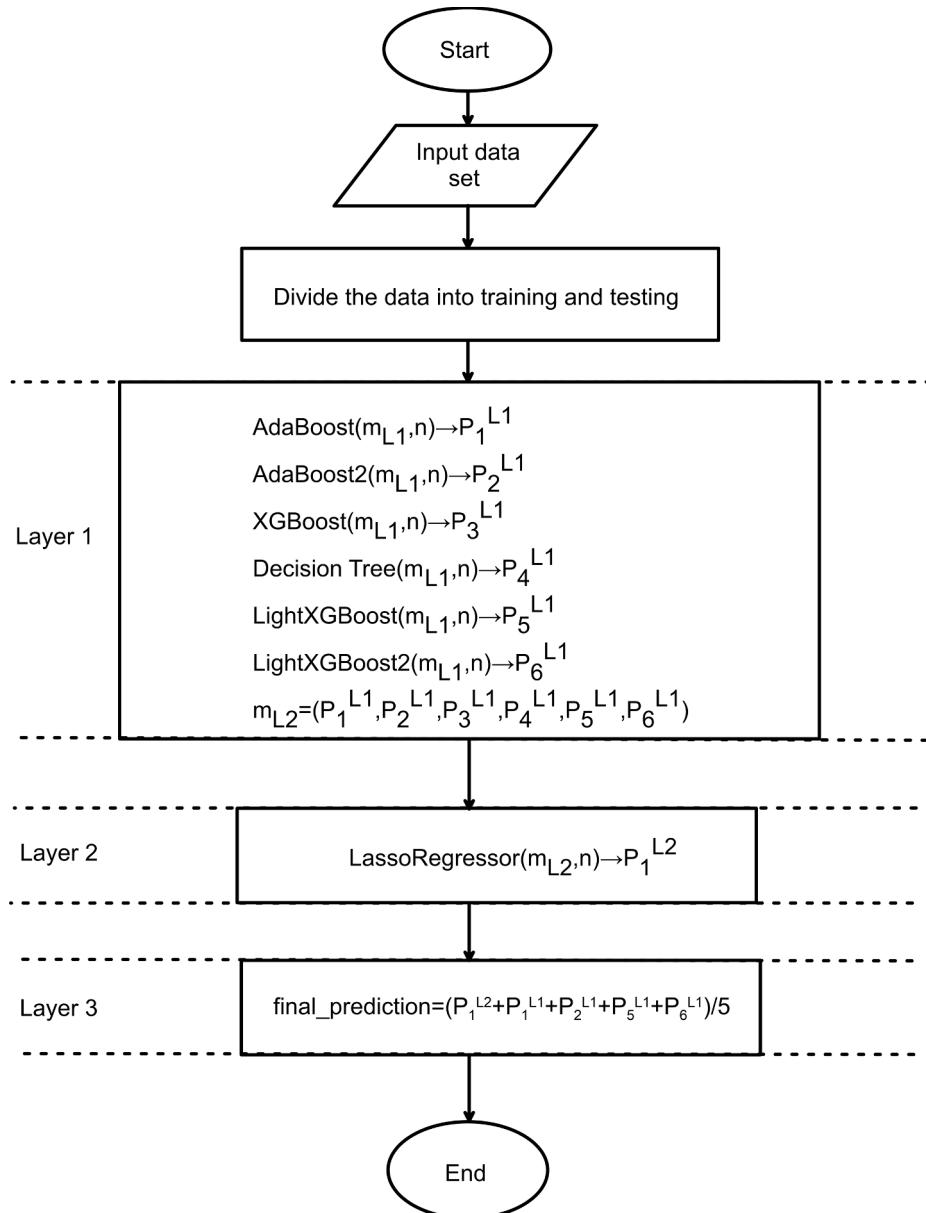


Fig. 6. The flowchart represents the steps to obtain the output using the EEWEnsembleStack model. Layers 1, 2, and 3 represent different layers of the EEWEnsembleStack model.

In this equation, M and N represent the total number of features and data, and the parameter α represents the penalty term. The Lasso regressor uses the max-iter, which controls the number of iterations used in the gradient descent. The final values of α and 'max_iter' obtained after optimised hyperparameter grid search and K fold cross-validation search in the present work are 0.1 and 5, respectively.

3.6. EEWPEnsembleStack

In the EEWPEnsembleStack model, six prediction results were generated by AdaBoost 1, AdaBoost 2, LightGBM 1, LightGBM 2, Decision Tree, and XGBoost from the first layer are passed to the second layer. In the second layer, the Lasso regression model uses the predictions from six models of the first layer as an input. It produces output and passes it to the third layer along with predictions from the first layer. The first layer predictions of LightGBM 1, LightGBM 2, AdaBoost 1, and AdaBoost 2 are passed to the third layer since the predicted results from these models were comparatively better than other models of the first layer. The three-layer algorithm predicts results with optimised error and better prediction than the individual models. The flowchart of the EEWPEnsembleStack model is given in Fig. 6. The architecture of this model has been visualised in Fig. 7.

4. Validation of the model

The EEWPEnsembleStack model has been trained with the randomly selected '2360' records of different earthquakes, and the trained model has been used for the prediction of magnitude using different features extracted from '591' records in the test data set. The accuracy of the prediction has been evaluated using R square (R^2) and means absolute error (MAE) defined by the following formula:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (11)$$

$$MAE = \frac{\sum|y_i - \hat{y}_i|}{n} \quad (12)$$

In equation (11), SS_{res} and SS_{tot} represent the sum of square residuals and average total, respectively. In equation (12), y_i and \hat{y}_i represent predicted and actual magnitude, respectively, and n represents the total number of records in the test data set. The value of R^2 lies between 0 and

1. The value close to 1 indicates that the prediction is equal to the actual magnitude, while the negative value of R^2 indicates worse predictions. Eight features have been extracted from 3 s of the P phase in the record, and the trained EEWPEnsembleStack model has been used to predict the magnitude. In this work, the model ablation study has been conducted to study the contribution of various individual models used in the EEWPEnsembleStack model. Table 2 shows the ablation study result after removing various models from different layers. The performance has been analysed using evaluation metrics defined by the R^2 and MAE score. The result of the model ablation study in Fig. 8a shows that the LightGBM 1, 2 and Lasso regression of Layer 2 models have the highest effect on the EEWPEnsembleStack model. Table 2 also indicates that even if one model from any layer of EEWPEnsembleStack is removed, then it will cause a negative impact on R^2 or MAE score. The study shows that the EEWPEnsembleStack model using all eight models gives the best R^2 score of value 0.63 and a minimum MAE score of value 0.419, and thus, it is suitable for a single station waveform's feature study. It is seen from Fig. 8 a and b that the features like P_d , ID2, IV2 give minimum MAE score and maximum R^2 scores.

The ablation study of input features has been performed to compare the evaluation metrics obtained from the EEWPEnsembleStack model

Table 2
Model ablation study.

Models used in EEWPEnsembleStack	R2 score	MAE score
[AdaBoost 1, AdaBoost 2, XGBoost, Decision Tree, LightGBM 1, Lasso Regression, and Layer 3 Algo]	0.63	0.422
[AdaBoost 1, AdaBoost 2, XGBoost, Decision Tree, LightGBM 2, Lasso Regression, and Layer 3 Algo]	0.63	0.423
[AdaBoost 1, AdaBoost 2, XGBoost, Decision Tree, Lasso Regression, and Layer 3 Algo]	0.62	0.431
[XGBoost, Decision Tree, LightGBM 1, LightGBM 2, Lasso Regression, and Layer 3 Algo]	0.62	0.418
[AdaBoost 1, AdaBoost 2, XGBoost, Decision Tree, LightGBM 1, LightGBM 2, and Layer 3 Algo]	0.62	0.428
[AdaBoost 1, AdaBoost 2, XGBoost, Decision Tree, and Layer 3 Algo]	0.58	0.465
[XGBoost, Decision Tree, LightGBM 1, LightGBM 2, and Layer 3 Algo]	0.6	0.436
[AdaBoost 1, AdaBoost 2, XGBoost, Decision Tree, LightGBM 1, LightGBM 2, Lasso Regression, and Layer 3 Algo]	0.63	0.419

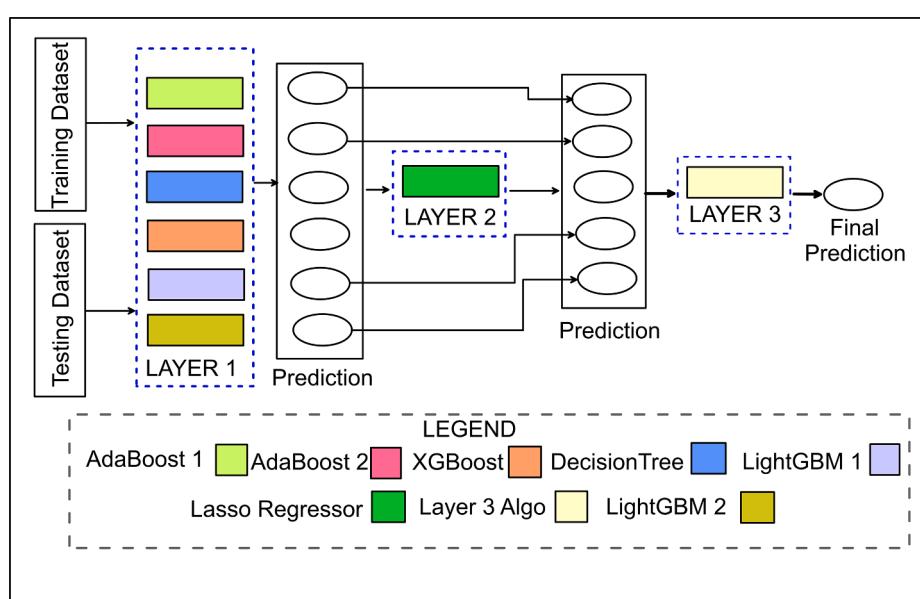


Fig. 7. The EEWPEnsembleStack model architecture, where the circle represents the prediction by the previous model attached to the circle, the rectangle represents the model, and the model colour is used to identify the name of the model, which is given in the legend.

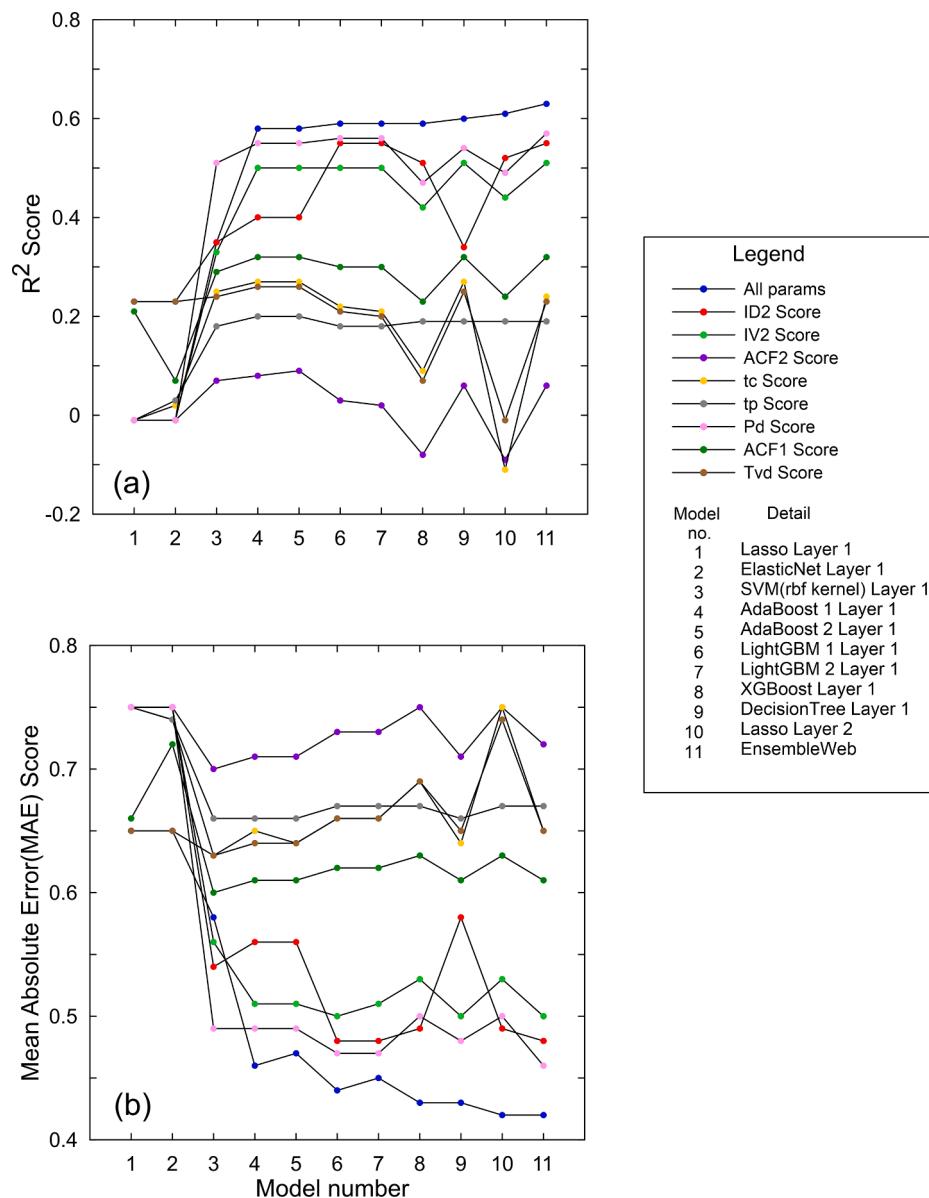


Fig. 8. The comparison of various features and models based on (a) R^2 and (b) MAE score.

trained using different features. The feature ablation study has been done using the EEWPEnsembleStack model trained using different feature sets. Table 3 investigates the effect of various sets of features on the EEWPEnsembleStack model. Table 3 shows that Set 5, which consists of all features, gives the highest R^2 score and lowest MAE score of value 0.63 and 0.419, respectively, among all the sets considered in the training of models. The result shows that the accuracy in prediction increases with an increase in the number of input features used for training the model.

The Shapley (SHAP) values in the LightGBM 1, LightGBM 2, and

XGBoost models give information on the relevance of the feature used for the prediction of magnitude. The SHAP values of various features have been computed using the method given by Lundberg and Lee (2017) for the trained LightGBM 1, LightGBM 2, and XGBoost models and are shown in Fig. 9. It is seen from Fig. 9a, b and c that ID2 and ACF₂ features in the LightGBM 1, ID2 and ACF₂ features in the LightGBM 2 and ID2 and P_d features in the XGBoost are important estimators for magnitude study.

5. Results

The EEWPEnsembleStack model developed in this work consists of seven machine learning models and one final prediction algorithm in the third layer. The input to the EEWPEnsembleStack model consists of eight different extracted features from the early 3 s of the P phase of strong motion records. The EEWPEnsembleStack model has been trained for each of eight different features computed from the early three seconds after onset of the P phase in the record. The predicted magnitudes obtained from the trained model using '591' records in the test data set have been compared with the actual magnitude in Fig. 10.

Table 3
Ablation study result of input parameters.

Set Number	Input parameters	R^2 score	MAE score
1	[IV2, ID2]	0.56	0.470
2	[ID2, P _d]	0.57	0.470
3	[ID2, IV2, P _d]	0.59	0.460
4	[ID2, IV2, P _d , ACF ₁]	0.60	0.440
5	[ID2, IV2, P _d , ACF ₁ , ACF ₂ , T _{vd} , τ _p , and τ _c]	0.63	0.419

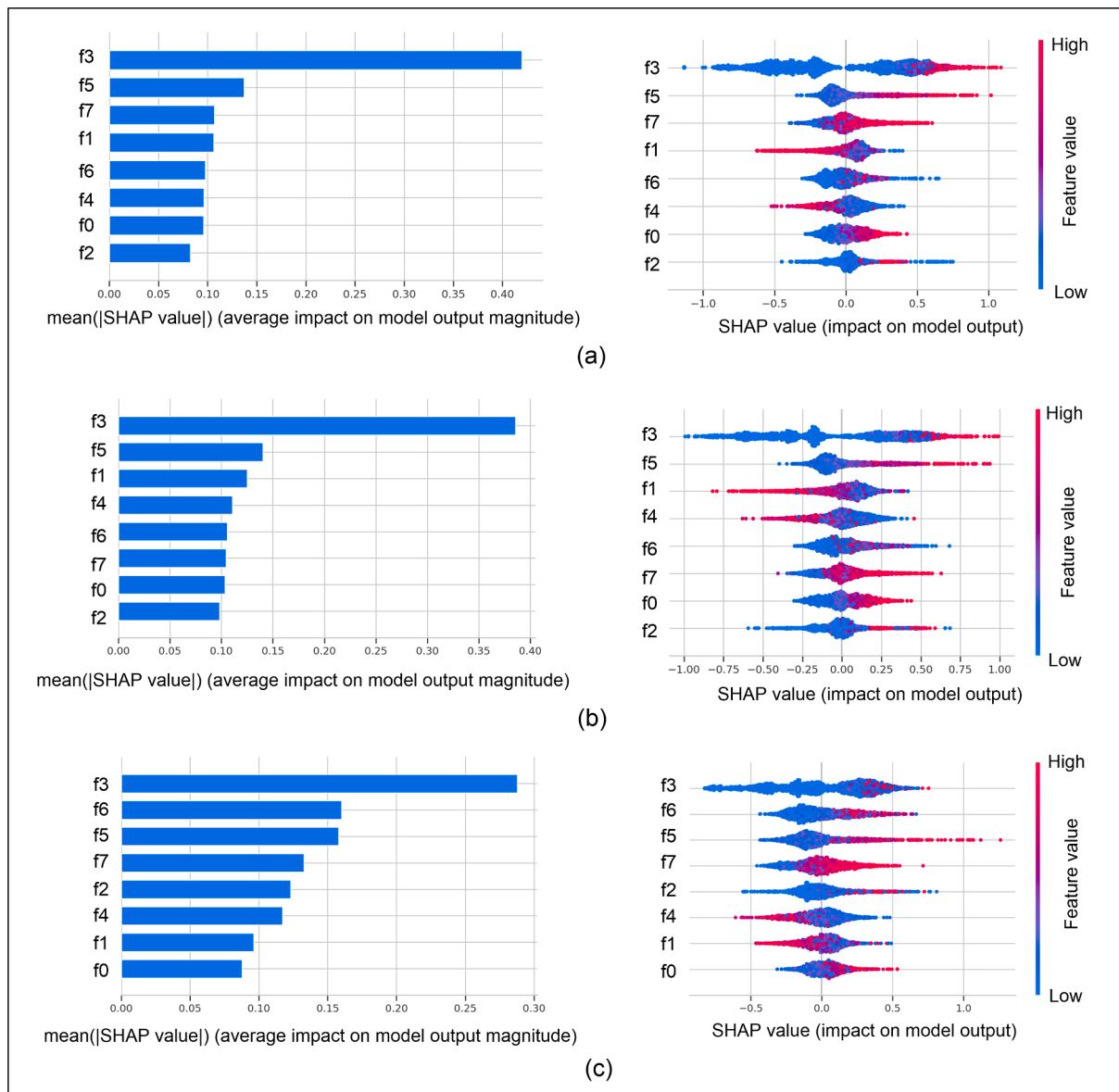


Fig. 9. The result obtained using Shapley values applied to training data with (a) LightGBM 1 model, (b) LightGBM 2 model, and (c) XGBoost model. The parameter f_i represents the name of the feature defined already in Fig. 3.

It is seen that the predicted magnitude is close to the actual magnitude when the EEWPEnsembleStack model is trained with each of the eight features simultaneously. Fig. 10a, b, c, d, e, f, g, h, and i show that the features P_d , $ID2$, and $IV2$ act as promising features when used independently for the training of the EEWPEnsembleStack model. The importance of a single feature in the EEWPEnsembleStack model for magnitude prediction on the basis of evaluation metrics results which are arranged in decreasing order, is P_d , $ID2$, $IV2$, ACF_1 , τ_c , T_{vd} , τ_p , and ACF_2 , respectively. Comparison of prediction error with magnitude and hypocentral distances for all trained ensemble models based on P_d , $ID2$, $IV2$, P_d regression relation and all features are shown in Fig. 11. The predictions obtained from these trained models have been compared with that obtained from the following regression relation of P_d based on the Japanese data given by Jin et al. (2013):

$$M = 0.91 \log_{10}(P_d) + 0.48 \log_{10}(hy) + 5.65 \pm 0.56 \quad (13)$$

In the above expression, hy is the hypocentral distance in km. The following formula for the computation of mean (μ) and standard deviation (σ) of the prediction error has been used in the present work:

$$\mu = \frac{\sum(M_{pred} - M_{actual})}{n} \quad (14)$$

$$\sigma = \sqrt{\frac{\sum(|M_{pred} - M_{actual}|^2)}{n}} \quad (15)$$

It is seen from Fig. 11 that the prediction error obtained from EEWPEnsembleStack models trained by using P_d , $ID2$ and $IV2$ features is -0.02 ± 0.60 , 0.02 ± 0.62 and -0.02 ± 0.63 , respectively. This is among the lowest prediction errors obtained in models trained using single features. The average prediction error obtained from the EEWPEnsembleStack model based on training by using all features is -0.01 ± 0.5 , which is the minimum among all models trained independently with a single feature. Further, the prediction has also been made using the same test data using the regression relation given by Jin et al. (2013), which is higher among all trained models. This is mainly due to the linear regression relation between P_d and magnitude used in the relation by Jin et al. (2013). The value of the Spearman and Pearson correlation coefficient of P_d parameter with magnitude is 0.47 and 0.7, respectively.

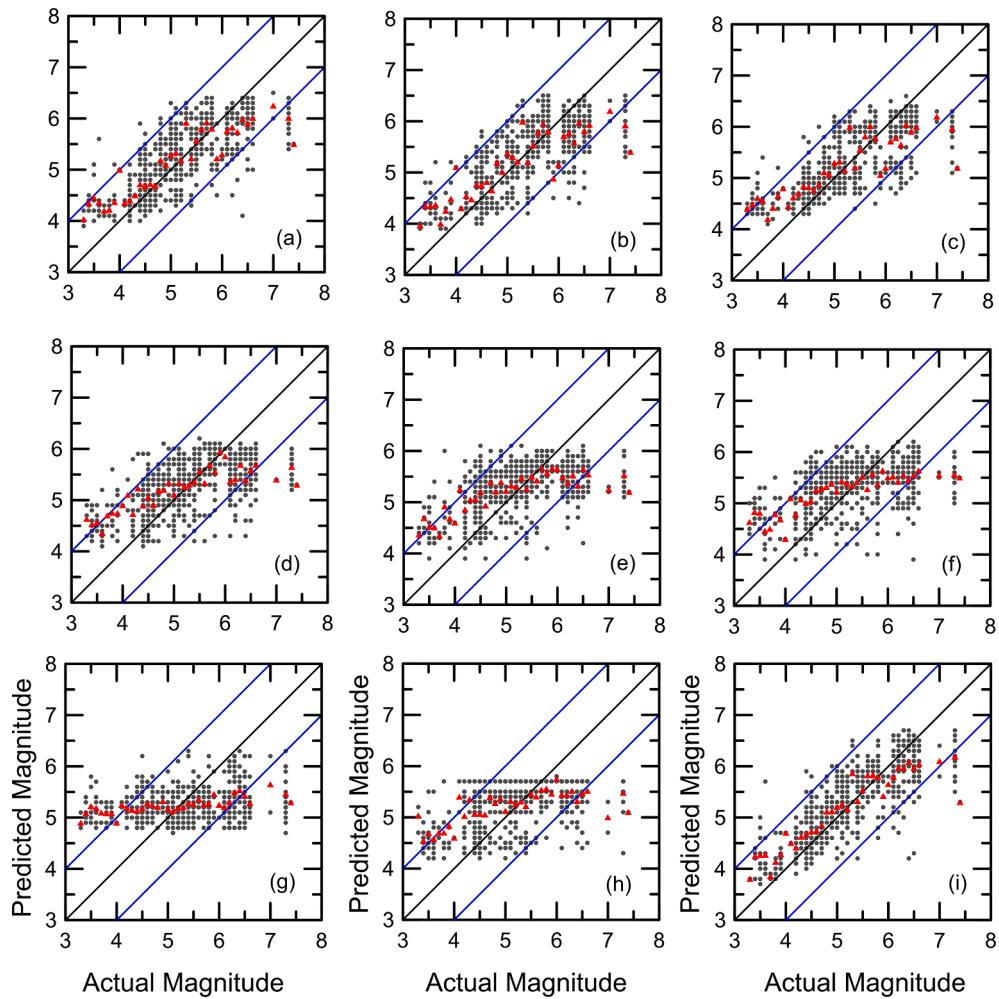


Fig. 10. Comparison of actual and predicted magnitude for test data set using EEWPEnsembleStack model based on (a) P_d , (b) IV2, (c) ID2, (d) ACF₁, (e) τ_c , (f) T_{vd} , (g) ACF₂, (h) τ_p , and (i) all features, respectively.

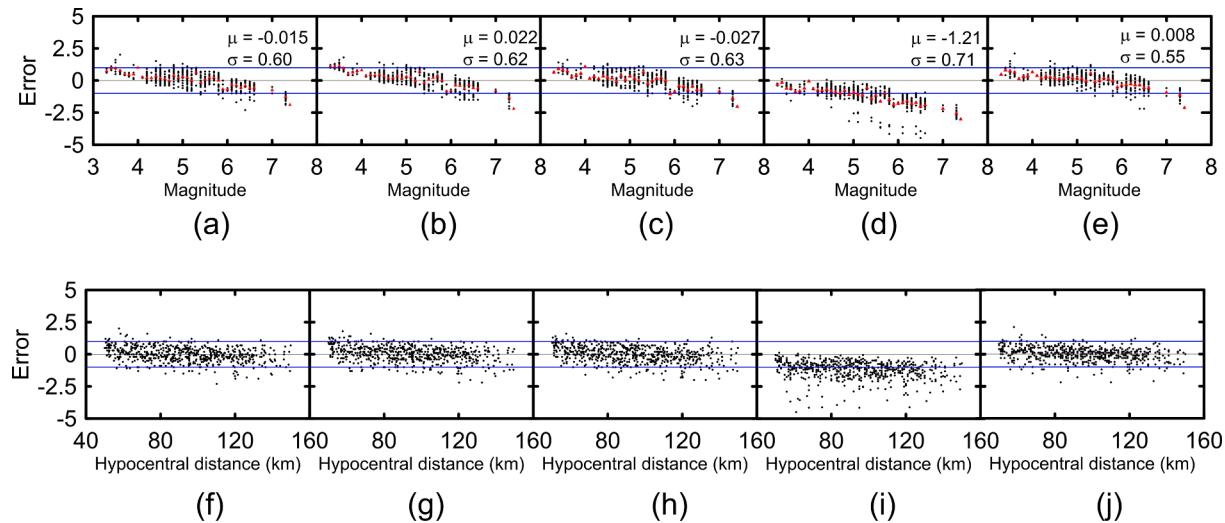


Fig. 11. The comparison of (a) Magnitude plot of prediction error with respect to magnitude obtained in the EEWPEnsembleStack model trained by using (a) P_d , (b) ID2, (c) IV2, (d) regression relation based on P_d given by Jin et al. (2013), and (e) all eight features. The plot of prediction error with respect to magnitude obtained in the EEWPEnsembleStack model trained by using (f) P_d , (g) ID2, (h) IV2, (i) regression relation of P_d given by Jin et al. (2013), and (j) all eight features.

This clearly indicates that P_d has moderate linear correlation and strong monotonic correlation with magnitude, which is the main source of high error obtained while using simple linear regression relation for calculation of magnitude.

The comparison of prediction error with actual magnitude for different models in Fig. 11a, b, c, and e show that the prediction error is steadily increasing with an increase in magnitude in all cases. The comparison of prediction error with respect to the actual magnitude presented in Fig. 11a, b, c and e show that the value of saturation magnitude has been effectively increased in the case of the EEWPEnsembleStack model based on all features. Fig. 11d, and i show that a high amount of discreteness is exhibited in the prediction error obtained by the conventional P_d methods based on the regression relation, and the average prediction error increases with increasing magnitude and hypocentral distance. The average prediction error using the EEWPEnsembleStack model shown in Fig. 11f, g, h, and j, remain nearly equal to zero for all hypocentral distances, which indicates that the errors related to the estimation of magnitude remain unaffected by choice of the hypocentral distance in the EEWPEnsembleStack model. It is seen clearly that though the predictions by all trained EEWPEnsembleStack models are far better than the regression relation, a minimum error has been obtained when trained EEWPEnsembleStack model based on all features have been used.

6. Discussion

The results obtained by the EEWPEnsembleStack model confirm that the predictions are close to actual values when the trained model based on all features has been used. For training and testing purposes, the data set consists of records from stations and earthquakes randomly selected from the entire data set. The primary purpose of the EEW systems is to predict the magnitude of the earthquake based on the early P phase of the records. This requires needs of testing developed algorithm for EEW based on the dataset of the same earthquake. In order to test a real scenario, an earthquake of magnitude 6.3 recorded at 21 stations within an epicentral distance of 50 to 150 km has been used for testing the trained EEWPEnsembleStack model using all eight features. The location of nearfield stations and the epicentre of this earthquake have been shown in Fig. 12a. This earthquake has not been included in the training data set, and hence the possibility of biases in prediction can be ruled out. The early three sec of strong motion records shown in Fig. 12b have been used for the prediction of the magnitude of this earthquake. The magnitude of this earthquake has also been computed from τ_c and P_d features using the conventional approach of regression relation. The following regression relation is given by Jin et al. (2013), based on τ_c has been used to calculate magnitude:

$$M = 2.94 \log_{10}(\tau_c) + 5.3 \pm 0.46 \quad (16)$$

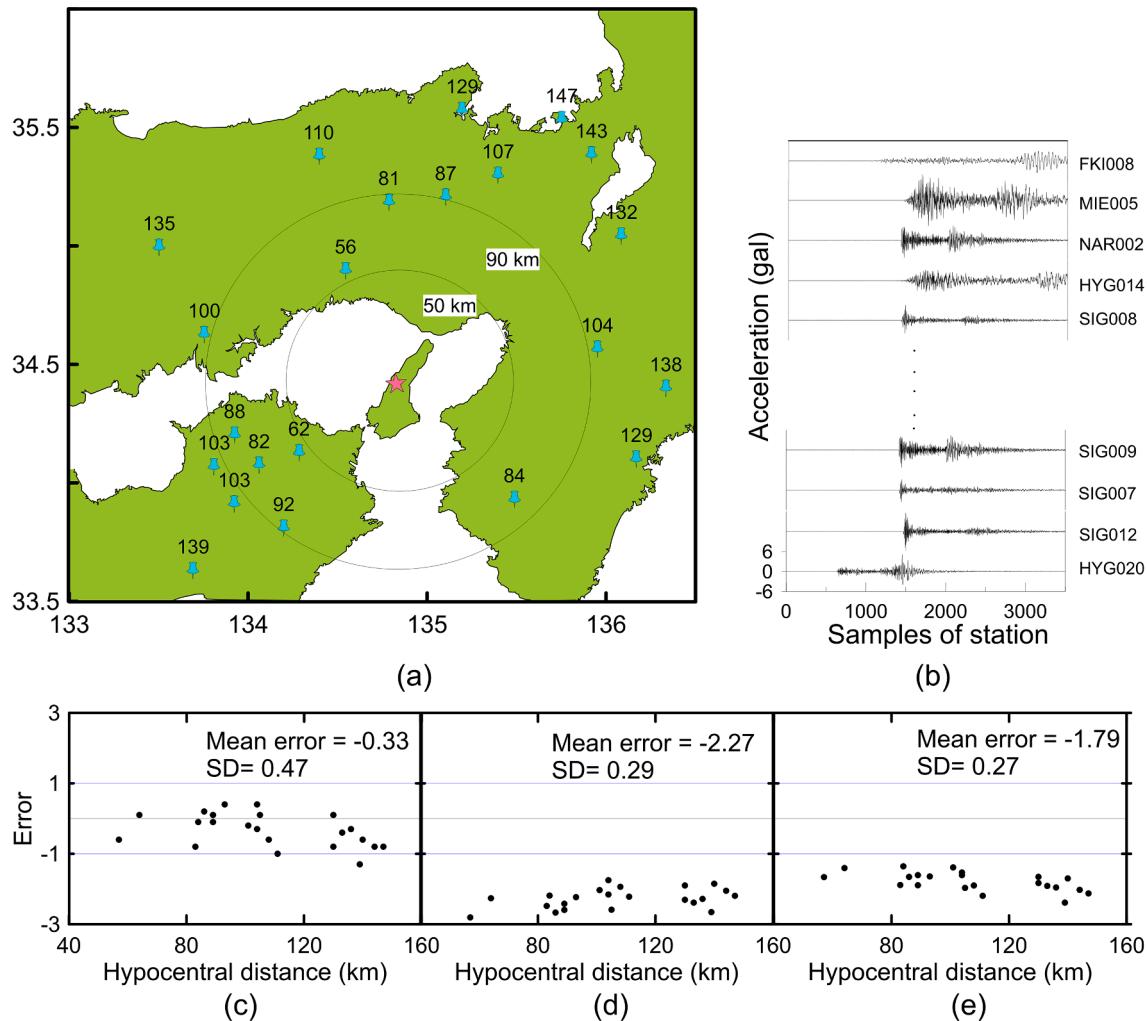


Fig. 12. (a) The location of the station that has recorded an earthquake of magnitude 6.3 shown by the solid star where the green colour represents the land region and white colour represents the ocean region, (b) strong motion records of this earthquake at some of the stations, distribution of predicted error with respect to hypocentral distance using (c) EEWPEnsembleStack prediction, (d) traditional τ_c method and (e) traditional P_d method.

The average predicted magnitude had been obtained from different records by using the EEWPEnsembleStack model, τ_c , and P_d regression relation given in equation (13) and (16), respectively, is 6.0 ± 0.5 , 4.0 ± 0.3 , and 4.5 ± 0.3 , respectively. The prediction error obtained from the EEWPEnsembleStack model is -0.33 , which is the least as compared to the τ_c and P_d methods as shown in Fig. 12c, d, and e. The comparison of results obtained through the developed model in this work with that obtained by using linear regression relation based on P_d , and τ_c parameters indicate that mean error is high when linear regression relations have been used. The objective of EEW systems is to compute reliable magnitude from records closer to the epicentre of the earthquake so that early warning can be provided to other stations. The prediction of the magnitude of this earthquake based on records within different epicentral distances has been shown in Table 4. It has been seen that even by using only two records that lie within a distance of 50 to 80 km, the predicted magnitude from EEWPEnsembleStack is obtained as 6.1 ± 0.5 , which is fairly close to the actual magnitude, thereby establishing the efficacy of the EEWPEnsembleStack model of predicting the magnitude of an earthquake based on single or multiple records.

The recordings from the same earthquake have been used for obtaining weighted magnitude. In this attempt, the weighted magnitude has been calculated by using all accelerograms of this earthquake recorded after 3, 6, 9, 12, and 15 secs from the origin time. Weights are iteratively selected to minimise the difference between the actual and weighted magnitude of multiple stations. The following equation has been used for the calculation of weighted magnitude:

$$M_{wn} = \frac{\left[\sum_{i=1}^n w_i \sum_{j=1}^{p_i} M_{ij} \right]}{n} \quad (17)$$

Where, w_i denotes the value of weight given to the i^{th} group and p_i is the number of stations in i^{th} groups. M_{wn} is the weighted earthquake magnitude obtained from n^{th} group. The weights are given by w_i , $i \in \{0, n\}$, where n is the number of groups. Weights are iteratively selected to minimise the difference between actual and weighted magnitude obtained from multiple stations. The final value of iteratively selected weights is denoted by w_1, w_2, w_3, w_4 and w_5 in this work is 1.04, 1.05, 1.06, 1.07 and 1.08, respectively. Fig. 13 shows the plot of average and weighted magnitude obtained for the 6.3 magnitude earthquake at different times after the arrival time of the first signal in the network recording a 6.3 magnitude earthquake. The dashed red line indicates the 6.3 magnitude earthquake. The blue line with a solid blue circle denotes the weighted magnitude while the green line with a solid green circle denotes the average magnitude in this figure.

7. Conclusions

This paper describes the ensemble model for the early prediction of earthquake magnitude based on features extracted from the 3 s of the P phase of the record. The ensemble model developed in this paper utilises popular machine learning models like AdaBoost, XGBoost, Decision-Tree, LightGBM and Lasso Regression. The model ablation study indicates that the minimum error has been obtained for a Decision Tree regression model among all individually trained single models. However, the minimum error has been obtained when all machine learning models have been utilised for training purposes. The feature and model ablation study shows that the least error has been obtained for cases when the ensemble model utilising all machine learning models has been trained by eight extracted features rather than an individual model trained by a single feature. The absolute mean prediction error of 0.42 has been obtained by using 591 records of 137 earthquakes in the developed ensemble model. Twenty-two strong motion records from a test earthquake of magnitude 6.3 (M_{JMA}) have been used for testing the developed model. It has been seen that the model has effectively predicted the magnitude of this earthquake as 6.0 ± 0.5 . The magnitude predicted by using popular τ_c , and P_d methods for this earthquake have

Table 4

The availability of records within different epicentral distances and its dependency on predicted magnitude using EEWPEnsembleStack mode for magnitude 6.3 earthquakes.

Range of epicentral distance (km)	Total records available	Predicted magnitude from EEWPEnsembleStack model
50 to 80 km	2	6.1 ± 0.5
50 to 90 km	7	6.1 ± 0.4
50 to 100 km	9	6.2 ± 0.4
50 to 110 km	14	6.1 ± 0.4
50 to 150 km	22	6.0 ± 0.5

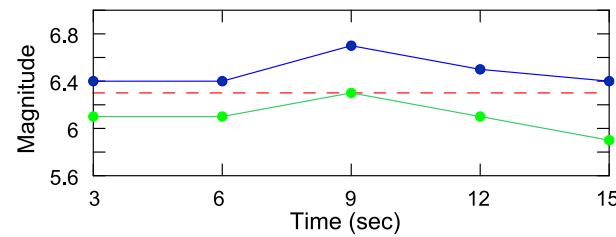


Fig. 13. The weighted and average magnitude obtained at different times after the arrival time of the first signal in the network recording a 6.3 magnitude earthquake. The dashed red line indicates the 6.3 magnitude earthquake. The blue line with a solid blue circle denotes the weighted magnitude while the green line with a solid green circle denotes the average magnitude in this figure.

been obtained as 4.0 ± 0.3 and 4.5 ± 0.3 , respectively, which clearly establishes the efficacy of the developed model for EEW systems. The study indicates that the weighted result of multiple magnitudes for the earthquake of magnitude 6.3 is obtained as 6.4.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Authors thankfully acknowledge National Research Institute for Earth Science and Disaster Resilience (NIED) for providing earthquake data. Authors are thankful to anonymous reviewers, Dr. Nuretdin Kaymakci (Associate editor) and Dr. Ibrahim Uysal (Co Editor in Chief) for their valuable suggestions for improvement of this manuscript.

Availability of data and material

The data used in this research work is taken from National Research Institute for Earth Science and Disaster Resilience (NIED).

Code availability (software application or custom code): Code used in this paper has been developed in Python.

Authors Contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Anushka Joshi and Vishnu Chalavadi. The first draft of the manuscript was written by Anushka Joshi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics Approvals

The work presented in this paper does not involve research using

humans and/or animals.

References

- Allen, R.M., Kanamori, H., 2003. The potential for earthquake early warning in southern California. *Sci.* 300, 786–789.
- Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* 68, 1521–1532. <https://doi.org/10.1007/BF02247958>.
- Allen, T.I., Marano, K.D., Earle, P.S., Wald, D.J., 2009. PAGER-CAT: a composite earthquake catalog for calibrating global fatality models. *Seismol. Res. Lett.* 80, 57–62.
- S, Aoi, T, Kunugi, H, Nakamura, H, Fujiwara, 2011. Deployment of new strong motion seismographs of K-NET and KiK-net. In: Akkar et al (Eds.) Earthquake data in engineering seismology, geotechnical, geological, and earthquake engineering 14. Springer, Berlin, pp. 167–186. doi:[10.1007/978-94-007-0152-6_12](https://doi.org/10.1007/978-94-007-0152-6_12).
- Aranda, J.M.E., Jimenez, A., Ibarrola, G., Alcantar, F., Aguilar, A., Inostroza, M., Maldonado, S., 1995. Mexico City seismic alert system. *Seismol. Res. Lett.* 66, 42–53.
- Atefi, S., Heidari, R., Mirzaei, N., Siahkoohi, H.R., 2017. Rapid Estimation of Earthquake Magnitude by a New Wavelet-Based Proxy. *Seismol. Res. Lett.* 88, 1527–1533.
- Behr, Y., Clinton, J., Kästli, P., Cauzzi, C., Racine, R., Meier, M.A., 2015. Anatomy of an earthquake early warning (EEW) alert: Predicting time delays for an end-to-end EEW system. *Seismol. Res. Lett.* 86, 830–840.
- Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.
- Böse, M., Hauksson, E., Solanki, K., Kanamori, H., Wu, Y.M., Heaton, T.H., 2009. A new trigger criterion for improved real-time performance of onsite earthquake early warning in southern California. *Bull. Seismol. Soc. Am.* 99, 897–905.
- Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman and Hall.
- Brown, H., Allen, M., Richard, M., Grasso, V.F., 2009. Testing ElarmS in Japan. *Seismol. Res. Lett.* 80, 727–739.
- Carranza, M., Buñorn, E., Zollo, A., 2015. Testing the earthquake early-warning parameter correlations in the southern Iberian Peninsula. *Pure Appl. Geophys.* 172, 2435–2448.
- Chamoli, B.P., Kumar, A., Chen, D.Y., Gairola, A., Jakka, R.S., Pandey, B., Kumar, P., Rathore, G., 2019. A prototype earthquake early warning system for northern India. *J. Earthq. Eng.* <https://doi.org/10.1080/13632469.2019.1625828>.
- Chen, S., Jiang, C., Zhuang, J., 2016. Statistical evaluation of efficiency and possibility of earthquake predictions with gravity field variations and its analytic signal in western China. *Pure. Appl. Geophys.* 173, 305–319. <https://doi.org/10.1007/s00024-015-1114-x>.
- Chen, T., Guestrin, C., 2016. XGBoost: a Scalable Tree Boosting System. *KDD '16*. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
- Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C., Liu, R., 2019. XGBoost-Based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access.* 7, 13149–13158. <https://doi.org/10.1109/ACCESS.2019.2893448>.
- Colombelli, S., Carotenuto, F., Elia, L., Zollo, A., 2020. Design and implementation of a mobile device app for network-based earthquake early warning systems (EEWSs): application to the PRESTO EEWS in southern Italy. *Nat. Hazards Earth Syst. Sci.* 20, 921–931.
- Erdik, M., Fahjan, Y., Ozel, O., Alcik, H., Mert, A., Gul, M., 2003. Istanbul earthquake rapid response and the early warning system. *Bull. Earthq. Eng.* 1, 157–163.
- Festa, G., Zollo, A., Lancieri, M., 2008. Earthquake magnitude estimation from early radiated energy. *Geophys. Res. Lett.* 35, L22307. <https://doi.org/10.1029/2008gl035576>.
- Freund, Y., Schapire, R.E., 1997. 1997. A Decision-Theoretic Generalisation of On-Line Learning and an Application to Boosting. *COLT*.
- Gunes, F., Wolfinger, R., Tan, P., 2017. Stacked Ensemble Models for Improved Prediction Accuracy. *SAS-2017*.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning. Springer-Verlag.
- Ionescu, C., Böse, M., Wenzel, F., Marmureanu, A., Grigore, A., Marmureanu, G., 2007. An early warning system for deep Vrancea (Romania) earthquakes. In: Gasparini, P., Manfredi, G., Zschau, J. (Eds.), Earthquake Early Warning Systems. Springer, Berlin and Heidelberg, pp. 343–349.
- Jin, X., Zhang, H., Li, J., Wei, Y., Ma, Q., 2013. Earthquake magnitude estimation using the τ_c and P_d method for earthquake early warning systems. *Earthq. Sci.* 26, 23–31.
- Kamigaitchi, O., Saito, M., Doi, K., Matsumori, T., Tsukada, S., Takeda, K., Shimoyama, T., Nakamura, K., Kiyomoto, M., Watanabe, Y., 2009. Earthquake early warning in Japan: warning the general public and future prospects. *Seismol. Res. Lett.* 80, 717–726. <https://doi.org/10.1785/gssrl.80.5.717>.
- Kanamori, H., 2005. Real-time seismology and earthquake damage mitigation. *Annu. Rev. Earth Planet. Sci.* 33, 195–214.
- Kohler, M.D., Cochran, E.S., Given, D., Guiwits, S., Neuhauser, D., Henson, I., Hartog, R., Bodin, P., Kress, V., Thompson, S., 2018. Earthquake early warning ShakeAlert system: west coast wide production prototype. *Seismol. Res. Lett.* 89, 99–107.
- Kuang, W., Yuan, C., Zhang, J., 2021. Network-based earthquake magnitude determination via deep learning. *Seismol. Res. Lett.* 92, 2245–2254. <https://doi.org/10.1785/0220200317>.
- Kuyuk, H., Allen, R., 2013. A global approach to provide magnitude estimates for earthquake early warning alerts. *Geophys. Res. Lett.* 40, 6329–6333.
- Kuyuk, H.S., Allen, R.M., Brown, H., Hellweg, M., Henson, I., Neuhauser, D., 2014. Designing a network-based earthquake early warning algorithm for California: Elarms-2. *Bull. Seismol. Soc. Am.* 104, 162–173.
- Lin, T.L., Wu, Y.M., Chen, D.Y., 2011. Magnitude estimation using initial P-wave amplitude and its spatial distribution in earthquake early warning in Taiwan. *Geophys. Res. Lett.* 38 <https://doi.org/10.1029/2011GL047461>.
- Lundberg, S.M., Lee, S., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777.
- Mousavi, S.M., Beroza, G.C., 2020. A machine-learning approach for earthquake magnitude estimation. *Geophys. Res. Lett.* 47 <https://doi.org/10.1029/2019gl085976>.
- Ochoa, L.H., Niño, L.F., Vargas, C.A., 2017. Fast magnitude determination using a single seismological station record implementing machine learning techniques. *Geod. Geodyn.* 9, 34–41.
- Odaka, T., Ashiya, K., Tsukada, S., Sato, S., Ohtake, K., Nozaka, D., 2003. A new method of quickly estimating epicentral distance and magnitude from a single seismic record. *Bull. Seismol. Soc. Am.* 93, 526–532.
- Olivieri, M., Allen, R.M., Wurman, G., 2008. The potential for earthquake early warning in Italy using Elarms. *Bull. Seismol. Soc. Am.* 98, 495–503.
- Peng, H., Wu, Z., Wu, Y.M., Yu, S., Zhang, D., Huang, W., 2011. Developing a prototype earthquake early warning system in the Beijing capital region. *Seismol. Res. Lett.* 82, 394–403.
- Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. *Sci. Adv.* 4 <https://doi.org/10.1126/sciadv.1700578>.
- Reddy, R., Nair, R.R., 2013. The efficacy of support vector machines (SVM) in robust determination of earthquake early warning magnitudes in central Japan. *Jour. Earth Syst. Sci.* 122, 1423–1434.
- Reyes, J., Morales-Esteban, A., Martínez-Álvarez, F., 2013. Neural networks to predict earthquakes in Chile. *Appl. Soft. Comput.* 13, 1314–1328.
- Satake, K., 2015. Geological and historical evidence of irregular recurrent earthquakes in Japan. *Phil. Trans. R. Soc. A* 373, 20140375. <https://doi.org/10.1098/rsta.2014.0375>.
- Schober, P., Boer, C., Schwarte, L., 2018. Correlation Coefficients: appropriate Use and Interpretation. *Anesthesia and Analgesia.* 126, 1763–1768.
- Sheen, D.H., Park, J.H., Chi, H.C., Hwang, E.H., Lim, I.S., Seong, Y.J., Pak, J., 2017. The first stage of an earthquake early warning system in South Korea. *Seismol. Res. Lett.* 88, 1491–1498.
- Shieh, J., Wu, Y., Allen, R.M., 2008. A comparison of τ_c and P_{max} for magnitude estimation in earthquake early warning. *Geophys. Res. Lett.* 35 <https://doi.org/10.1029/2008GL035611>.
- Sinhal, A., Khattri, K.N., 1983. Application of seismic reflection data to discriminate subsurface lithostratigraphy. *Geophysics.* 48, 498–1513.
- Takla, E.M., Yumoto, K., Okano, S., Uozumi, T., Abe, S., 2013. The signature of the 2011 Tohoku mega earthquake on the geomagnetic field measurements in Japan. *NRIAG. J. Astronomy and Geophys.* 2, 185–195.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B (Methodol.).* 58, 267–288.
- Wakita, K., 2013. Geology and tectonics of Japanese islands: a review - the key to understanding the geology of Asia. *J. Asian Earth Sci.* 72, 75–87.
- Wang, Y., Li, S., Song, J., 2020. Threshold-based evolutionary magnitude estimation for an earthquake early warning system in the Sichuan-Yunnan region. *China. Sci. Rep.* 10, 21055. <https://doi.org/10.1038/s41598-020-78046-2>.
- Wang, W., Ni, S., Chen, Y., Kanamori, H., 2009. Magnitude estimation for early warning applications using the initial part of P waves: a case study on the 2008 Wenchuan sequence. *Geophys. Res. Lett.* 36 <https://doi.org/10.1029/2009GL038678>.
- Wang, Y., Wang, T., 2020. Application of improved LightGBM model in blood glucose prediction. *Appl. Sci.* 10, 3227. <https://doi.org/10.3390/app10093227>.
- Wei, D., Seno, T., 1998. Determination of the Amurian Plate Motion. In: Flower, M., Series, GeoDynamics (Eds.), Mantle Dynamics and Plate Interactions in East Asia. AGU.
- Wu, Y.M., Kanamori, H., 2005. Experiment on an onsite early warning method for Taiwan early warning system. *Bull. Seismol. Soc. Am.* 95, 347–353.
- Wu, Y.M., Kanamori, H., Richard, M.A., Hauksson, E., 2007. Determination of earthquake early warning parameters, τ_c and P_d , for southern California. *Geophys. J. Int.* 170, 711–717.
- Wu, Y.M., Teng, T.L., 2002. A virtual sub-network approach to earthquake early warning. *Bull. Seismol. Soc. Am.* 92, 2008–2018.
- Wu, Y.M., Yen, H.Y., Zhao, L., Huang, B.S., Liang, W.T., 2006. Magnitude determination using initial P waves: a single-station approach. *Geophys. Res. Lett.* 33 <https://doi.org/10.1029/2005GL025395>.
- Wu, Y., Zhao, L., 2006. Magnitude estimation using the first three second P-wave amplitude in earthquake early warning. *Geophys. Res. Lett.* 331 <https://doi.org/10.1029/2006GL026871>.
- Zhang, H., Jin, X., Wei, Y., Li, J., Kang, L., Wang, S., Huang, L., Yu, P., 2016. An earthquake early warning system in Fujian, China. *Bull. Seismol. Soc. Am.* 106, 755–765.
- X, Zhang, M., Zhang, & X, Tian, (2021). Real-time earthquake early warning with deep learning: Application to the 2016 M 6.0 Central Apennines, Italy Earthquake. *Geophys. Res. Lett.*, 48, 2020GL089394. doi: [10.1029/2020GL089394](https://doi.org/10.1029/2020GL089394).
- Zhu, J., Li, S., Song, J., Wang, Y., 2021a. Magnitude estimation for earthquake early warning using a deep convolutional neural network. *Front. Earth Sci.* 9 <https://doi.org/10.3389/feart.2021.653226>.

- Zhu, J., Li, S., Song, J., 2021b. Magnitude estimation for earthquake early warning with multiple parameter inputs and a support vector machine. *Seismol. Res. Lett.* 93, 126–136. <https://doi.org/10.1785/0220210144>.
- Zollo, A., Iannaccone, G., Lancieri, M., Cantore, L., Convertito, V., Emolo, A., Festa, G., Gallović, F., Vassallo, M., Martino, C., 2009. Earthquake early warning system in southern Italy: methodologies and performance evaluation. *Geophys. Res. Lett.* 36 <https://doi.org/10.1029/2008GL036689>.
- Zollo, A., Lancieri, M., Nielsen, S., 2006. Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records. *Geophys. Res. Lett.* 33 <https://doi.org/10.1029/2006GL027795>.