

Elite GA-based Feature Selection of LSTM for Earthquake Prediction

Zhiwei Ye

Hubei University of Technology

Wuyang Lan

Hubei University of Technology

Wen Zhou (✉ zw_mmwh@hbut.edu.cn)

Hubei University of Technology

Qiyi He

Hubei University of Technology

Liang Hong

Hubei Provincial Geographical National Conditions Monitoring Center

Xinguo Xu

Hubei Provincial Geographical National Conditions Monitoring Center

Yunxuan Gao

Hubei University of Technology

Research Article

Keywords: Earthquake magnitude prediction, Elite genetic algorithm, Long short-term memory, AETA

Posted Date: June 19th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3049982/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Elite GA-based Feature Selection of LSTM for Earthquake Prediction

Zhiwei Ye¹, Wuyang Lan¹, Wen Zhou^{1*}, Qiyi He¹, Liang Hong², Xinguo Xu² and Yunxuan Gao¹

¹School of Computer Science, Hubei University of Technology, No. 28, Nanli Road, Hongshan District, Wuhan, 430068, Hubei, China.

²Hubei Provincial Geographical National Conditions Monitoring Center, No. 50, Zhongnan 1st Road, Wuhan, 430064, Hubei, China.

*Corresponding author(s). E-mail(s): zw_mmwh@hbut.edu.cn;
Contributing authors: hgcsyzw@hbut.edu.cn; lwy@hbut.edu.cn;
20220014@hbut.edu.cn; 147775318@qq.com; 584231247@qq.com;
gyxxxh@gmail.com;

Abstract

Earthquake magnitude prediction is an extremely difficult task that has been studied by various machine learning researchers. However, the redundant features and time series properties hinder the development of prediction models. Elite Genetic Algorithm (EGA) has the advantages in searching optimal feature subsets, meanwhile, Long Short-Term Memory (LSTM) is dedicated to processing time series and complex data. Therefore, we propose an EGA-based feature selection of LSTM model (EGA-LSTM) for time series earthquake prediction. First, the acoustic and electromagnetics data of the AETA system we developed are fused and preprocessed by EGA, aiming to find strong correlation indicators. Second, LSTM is introduced to execute magnitude prediction with the selected features. Specifically, the RMSE of LSTM and the ratio of selected features are chosen as fitness components of EGA. Finally, we test the proposed EGA-LSTM on the AETA data of Sichuan province, including the influence of data in different periods (*timePeriod*) and fitness function weights (ω_a and ω_F) on the prediction results. Linear Regression (LR), Support Vector Regression (SVR), Adaboost, Random Forest (RF), standard GA (SGA), steadyGA,

and three Differential Evolution Algorithms (DEs) are adopted as our baselines. Experimental results demonstrate that all the methods can get the best performance when *timePeriod* = 0 : 00 – 8 : 00, $\omega_a = 1$, and $\omega_F = 0.8$. Moreover, our proposed approach is superior to state-of-the-art approaches on the evaluation indicators MAE, MSE, RMSE, and R_2 . Non-parametric tests reveal that EGA-LSTM is significantly different from others and outperforms the standard LSTM.

Keywords: Earthquake magnitude prediction, Elite genetic algorithm, Long short-term memory, AETA

1 Introduction

Earthquakes are serious, sudden-onset natural disasters. Since 1970, at least 3.3 billion people have died from such natural disasters, and 226 million people are directly affected each year [1]. Earthquakes cause serious human injuries and economic losses due to the destruction of buildings and other rigid structures. Nevertheless, earthquake prediction is essential and necessary to reduce human casualties and economic loss.

Earthquake Moment Magnitude (MW) prediction is one of the important issues in earthquake prediction fields. Many studies have been carried out for finding the relationships in earthquakes and MW prediction. However, the complex and non-linear relationships in earthquakes make MW prediction to be a challenging task. Machine Learning (ML) methods have been adopted for MW prediction due to their high classification accuracy. Adeli et al. [2] calculated the seismic historical indicators, including Gutenberg Richer b-values, time lag, earthquake energy, and mean magnitude based on ML methods to predict MW. Asim et al. [3] investigated the Cyprus earthquake catalog temporally and computed sixty seismic features. Then, these features served as the input instances of Support Vector Machines (SVM) and RF to predict five days-ahead, one week-ahead, ten days-ahead, and fifteen days-ahead with different MW thresholds. Crustal movement is a continuous process makes MW prediction to be a time-series issue. Hence, in [4], Berhich et al. calculated an appropriate feature by historical data to enhance time-series task and implemented LSTM to predict MW with the enhanced features. Cai et al. [5] used three groups of real seismic data: gravity, georesistivity, and water-level dataset to predict MW based on LSTM. They considered the time series of precursors and earthquakes, however, there are irrelevant or even redundant features within the precursors, resulting in low prediction accuracy.

Feature Selection (FS) is a critical issue to improve the classification accuracy of ML methods. FS can be regarded as combinatorial optimization problems and EGA has the advantage in searching optimal feature subset. Kadam et al. [6] proposed EGA for selecting features in arrhythmia classification, and achieved satisfied performance. Thus, we adopt EGA for FS in MW

prediction and design a novel fitness function, which is calculated by the RMSE of LSTM and the ratio of selected features. Finally, considering the time-series effect of electromagnetic and acoustic data from AETA and the abruptness of earthquake, we introduce LSTM to predict magnitude with optimal feature subset selected by EGA.

To verify the performance of the proposed EGA-LSTM, 95 features that are derived from the electromagnetic and acoustic data collected by AETA are adopted as our experimental dataset. After the feature selection process by EGA, the chosen features served as the input data of LSTM model. Besides, five Evolution Algorithms (EAs) and four different ML methods are adopted as our baselines: SGA, steadyGA, and three DEs, LR, SVR, Adaboost, and RF. Five ML evaluation indicators are adopted to assess the performance of those methods. The experimental results demonstrate that our proposed EGA-LSTM is superior to the state-of-the-art methods.

The main contributions of this study are as follows:

- The original dataset is collected by the AETA system developed by our team, which detects electromagnetic and acoustic to predict earthquakes in Sichuan and surroundings.
- To eliminate redundant features from the original 95 features, we propose EGA with a novel fitness function of a specific seismic scene to find the optimal solution.
- Considering earthquake magnitude's time series effect and the abruptness of earthquake, we choose LSTM as prediction model. Other ML methods, such as LR, SVR, Adaboost, and RF are our baselines.
- The statistical measures (MAE, MSE, RMSE, R-square) are used to measure the performance of suggested ML models.

The remainder of the study is arranged as follows. Section 2 depicts the latest and related works. The method will be fully described in section 3. In section 4, the experiments of EGA-LSTM and other EAs and ML methods will be illustrated in detail. The conclusion and future works will be shown in section 5.

2 Related Work

Various approaches have been implemented in earthquake prediction, including feature selection methods and prediction methods.

2.1 Seismic feature selection

Feature selection is a crucial step in building a robust model, especially in high-dimensional data scene [7]. The application of feature selection to earthquake prediction is a complex problem, because different techniques are sensitive to data imbalance and noise. Martínez-Álvarez et al. [8] proposed information gain to evaluate different indicators and removed the low-ranked or null contribution seismic indicators. Roiz-Pagador et al. [9] introduced four different

feature selection methods and nine different EAs in correlation-based feature selection methods.

2.2 Earthquake prediction methods

The main MW prediction methods contain mathematical models, shallow ML techniques, and DL methods.

For mathematical models, the mathematical methods have been implemented to deal with uncertainty problems, such as probability theory and fuzzy sets to predict MW. For instance, Chen et al. [10] focused on the laws of the earthquake time series based on chaos theory and carried on the earthquake forecast simulation through the analysis of real data. The quantitative identification of time-series data result shows that seismic time-series data perform deterministic chaotic characteristics. Cekim et al. [11] captured the dynamics of earthquake occurrence using a novel Singular Spectrum Analysis and Autoregressive Integrated Moving Average method to predict the average and maximum MW in the East Anatolian Fault of Turkey. However, the prediction results proved unsatisfactory due to the strict requirements of the mathematical model on the independence and correlation of the data.

Another branch of MW prediction approaches focused on shallow ML techniques. They are driven by data and don't require many parameters. Since there is no definite mathematical relationship between precursor indicators and location and MW, Panakkat et al. [12] introduced three different neural networks: feed-forward Levenberg-Marquardt Back Propagation, RNN, and Radial Basis Function (RBF) neural network to predict MW and location with eight indicators from Gutenberg-Richter power-law. In 2017, Asim et al. [13] proposed different ML approaches including RF and Linear Programming Boost ensemble of decision trees for earthquake prediction in the region of Hindukush. These techniques show significant and encouraging results but they were still not available so far. As Chile is rocked by intraplate, interface, and crustal events, Chanda et al. [14] proposed six different ML methods to predict the total duration and significant duration for the three types of earthquakes. Shah et al. [15] introduced an Improved Artificial Bee Colony algorithm to improve the training process of Multilayer Perceptron. Muhammad et al. [16] studied earthquakes, Radon gas, and the Ionospheric total electron content and used statistical and machine/deep learning methods to find the relationship among the above-mentioned three elements in North Anatolian Fault. Yang et al. [17] proposed an automated regression pipeline approach for high-efficiency earthquake prediction. The core prediction part is training RF with bayesian parameter optimization. Ensemble learning has also been used in the field of earthquake prediction. Asim et al. [18] proposed a novel earthquake predictor system combining seismic indicators with GA and AdaBoost. GA's searching capabilities and boosting of AdaBoost makes the proposed method to be a powerful classifier. In addition, considering the earthquake prediction process is similar to the anomaly detection process of the biological immune system, Zhou et al. [19][20][21][22] introduced dendritic

cells algorithm, artificial macrophage classification optimization method, and artificial antigen-presenting cells approach to earthquake prediction. However, ML methods are unable to learn the complex and nonlinear relationship of earthquake features. Therefore, we need a more powerful technique to solve this problem.

Deep learning methods are experts at solving nonlinear problems because they have many hidden layers and densely connected neurons that preserve complex information. Moustra et al. [23] applied Artificial Neural Network with the history of MW data and seismic electric signals in Greece to predict MW, but the result is not satisfactory. In 2018, Asim et al. [24] computed sixty features by employing seismological concepts, extracted features by Maximum Relevance and Minimum Redundancy (mRMR), and predicted MW with the selected features based on SVR and HNN. In addition, HNN is supported by enhanced Particle Swarm Optimization. Jain et al. [25] proposed a DL method to predict the position and depth of earthquakes. They analyzed the effects of parameters on different datasets. However, these methods didn't catch the time series effects and only utilize DL methods to learn the relationship between features and magnitude. Draz et al. [26] believed that the evolution and appearance of earthquake precursors exhibit complex behavior. Therefore, they used deep machine learning for the detection of ionosphere and atmosphere precursors. This study showcases the importance of machine learning techniques in earthquake detection, which contributes to the understanding of the Lithosphere-Atmosphere-Ionosphere Coupling mechanism. Berhich et al. [27] presented LSTM for earthquake prediction to study the correlations between two divided groups considering their range of MW. To find the specific pattern among magnitude, timing and location. Berhich et al. [28] used LSTM to learn temporal relationships, and the attention mechanism extracts important patterns and information from input features. Kavianpour et al. [29] introduced a novel prediction model based on the attention mechanism, Convolution Neural Network (CNN), and Bidirectional Long Short-Term Memory (BiLSTM) models, to capture the temporal dependencies, which can predict the maximum MW and the number of earthquakes in a specified period in mainland China. Berhich et al. [4] calculated an appropriate feature to enhance time feature and predict MW based on LSTM with this feature. Nevertheless, most of these DL methods didn't consider some precursor features are irrelevant features, even redundant features. These irrelevant features can make DL models difficult to execute learning process, reducing predictive accuracy.

Generally speaking, the MW prediction process contains two steps: finding a valid optimization precursor subset and constructing a predicting model. Before a major earthquake stroke, it is often accompanied by changes in electromagnetic and acoustic signals in AETA. Hence, we have reasons to prove that electromagnetic and acoustic signals are related to MW. As far as authors know, electromagnetic and acoustic signals as precursor features haven't been

researched by LSTM, so we utilize these indicators to predict MW. In addition, considering the existence of redundant features, we propose EGA-LSTM for time series earthquake prediction based on AETA.

3 Methodology

This section describes the proposed model to perform earthquake prediction. EGA-LSTM is presented to predict MW.

3.1 Model overview

A schematic presentation of the proposed model is shown in Fig. 1. First, the dataset from AETA was detected in the Sichuan and surroundingd. This dataset includes two types, electromagnetic and acoustic, with a total of 95 features. AETA has recorded electromagnetic and acoustic signals from January 1, 2017 to December 31, 2022 and detects them every ten minutes. So, there are more than 20000 records. Since MW predicting in a short period makes less sense and human activities affect electromagnetic and acoustic signals, we merge signals from one day into one piece of catalog. The fusing detail will be described in section 3.2. Then, we proposed EGA for feature selection. As Fig. 1 depicts, the selected features are divided into two groups: the training set and the testing set. The time sequence of the train set and test set don't change. In this work, EGA-LSTM has been compared to other classic ML methods, such as LR, SVR, Adaboost, and RF. Finally, different evaluation indicators are applied to evaluate these algorithms.

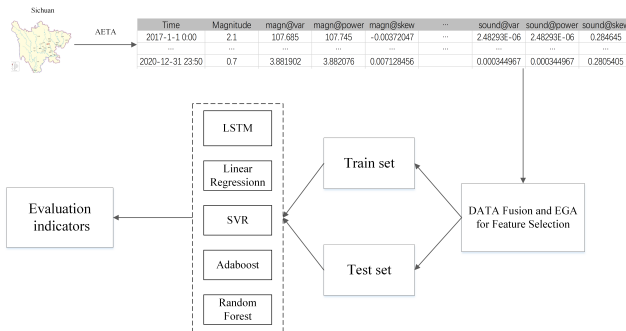


Fig. 1 The program diagram of the proposed earthquake prediction model

3.2 Process of EGA-LSTM

The EGA-LSTM architecture for earthquake magnitude prediction depicts as Fig. 2, meanwhile, the pseudocode of EGA-LSTM is shown in Algorithm 1. Their main idea is as follows. The AETA dataset includes 95 features, 51

features for electromagnetic signals, and 44 features for acoustic signals. However, MW isn't detected in AETA. Hence, the MW data is chosen from China Earthquake Networks Center. As a result, the algorithm merges 95 features and MW according to time. Since the time interval is short and MW prediction in a short period makes less sense. Hence, we select a typical value in every feature to represent a whole day. That means the size of the original one-day data is 144×96 . After fusing phase, the size becomes 1×96 . In addition, electromagnetic and acoustic signals may be affected by human activities, we need to choose a suitable *timePeriod* to present a whole day. Considering the lack of nighttime activities, we choose the data from the *timePeriod* 0:00 to 8:00. Then, the algorithm transforms time series data into supervised learning. In this algorithm, the input variable is electromagnetic and acoustic signals and the output variable is the MW of the following day. The step size is equal to 1. That means the algorithm can predict MW of the next day.

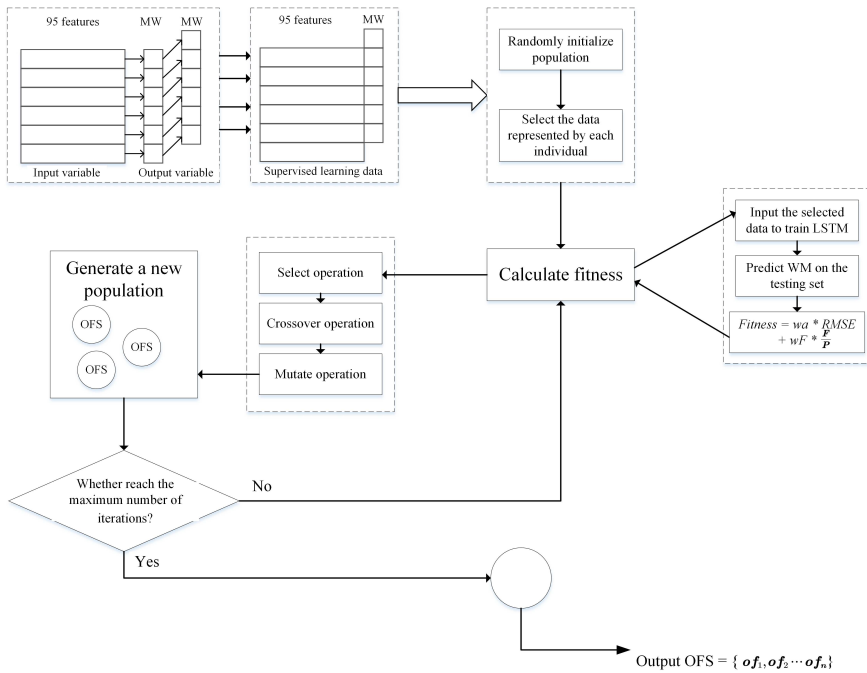


Fig. 2 The EGA-LSTM architecture for earthquake magnitude prediction

After processing the original data, the algorithm executes feature selection. This model adopts EGA to select optimal feature subset, the individual of EGA is applied to binary encoding and the dimension of each individual is 95. If one feature is selected, the corresponding dimension is 1, otherwise 0. With comprehensive consideration of prediction accuracy and time complexity, we proposed a novel fitness function defined as Eq. (1), where ω_a and ω_F is

Algorithm 1 EGA-LSTM

Input: electromagnetic and acoustic signals from AETA**Output:** MW of the next day

- 1: *Maxiter*: The maximum number of iterations
- 2: *Ofs*: The selected optimal feature subset
- 3: Use the max value of the data between 0:00 and 8:00 representing the whole day
- 4: Time series transform into supervised learning sequence, step size equal to 1
- 5: Initialize population and parameter
- 6: **while** $t < \textit{Maxiter}$ **do**
- 7: Calculate fitness for each individual
- 8: **for** each individual **do**
- 9: Retaining the best individuals
- 10: Selected operation
- 11: Crossover operation
- 12: Mutate operation
- 13: **end for**
- 14: Generate new population
- 15: **end while**
- 16: $OFS = \textit{best fitness individual}$
- 17: $MW = LSTM(OFS)$
- 18: **return** MW

the weight factor, F is the number of selected features, and P is the number of all features. The first part of Eq. (1) represents the prediction accuracy and the second part indicates the complexity of the model. In addition, Root Mean Squared Error (RMSE) is an evaluation indicator that is defined as Eq. (2), where \hat{y} is the predicting value and y is the true value. This fitness function enables considerable accuracy and the number of selected features is significantly reduced.

$$\textit{fitness} = \omega_a * RMSE + \omega_F * \left(\frac{F}{P}\right) \quad (1)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}\|_2^2} \quad (2)$$

The algorithm first defines the maximum number of iterations, crossover rate, mutate rate, and the number of individuals. Then the model randomly initializes the population and begins to iterate. The first step of iterating is calculating the fitness of every individual. Hereafter, the algorithm executes the selection operator, crossover operator, and mutation operator to generate a new subpopulation. The loop finishes until the number of iterations reaches the maximum number of iterations. After the iteration phase, the best individual is the selected optimal feature subset. The last step of the algorithm

is predicting MW. At the beginning, the selected optimal feature subsets are normalized by the Min-Max scaler, which is a transform method calculated by Eq. (3). In this model, optimal feature subsets are mapped into the same range which is between 0 and 1, which enables no one feature dominates the others. The next step consists of dividing 70% of the training set and 30% of the testing set. Then LSTM is trained and supported by RMSE for error calculation and evaluation. Finally, after the model is trained, LSTM predicts MW on the testing set, and the error between the real MW and the predicted MW is calculated by different evaluation indicators. They are widely used to evaluate regression models and applied to earthquake prediction. The detailed evaluation indicators will be provided in section 4.2.

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3)$$

3.3 Feature selection of EGA

GA is expert in solving combinatorial optimization problems since its operator was designed based on discrete encoding [30]. Feature selection is a combinatorial optimization problem and non-deterministic polynomial problem. The traditional feature selection method is to find relevant features to the target variables. However, the features among the selected features subset should be weakly correlated. Combining the above-mentioned two conditions, the model can find the relationship between features and target variable. EGA can directly take the optimal gene and directly participate in the next evolution without selection. Thus, we adopt EGA to find the optimal feature subset based on GA's searching ability.

In this method, we set the individual as a 95-dimensional vector and each element is equal to 1 or 0. If the element is equal to 1, it means the corresponding dimension's feature is selected, otherwise without being selected. Initially, we initialize these individuals randomly. Then, calculate the fitness of each individual based on Eq. (1). To get RMSE in Eq. (1), we train a sample LSTM on the training test and predict MW on the testing set. Then calculate RMSE based on Eq. (2). To decrease the time complexity and increase the accuracy, we design a novel fitness function as Eq. (1). The first part represents the prediction accuracy and the second part represents time complexity. After calculating the fitness, the algorithm begins to select individuals. Before every loop, EGA retains the best individuals. The selection part is applied to the roulette wheel to produce two new solutions. The algorithm generates a random probability value and selects the corresponding individual. The next step is the crossover operator. The chosen crossover technique in this model is double-point. There are two crossover points and the chromosomes between two points are swapped only. The mutation operator is inversion mutation randomly generating a gene locus in a low mutate rate and inverting. If the old gene locus value is 1, the new value is 0. After generating a new population, if the number of iterations reaches the maximum, output the

best individual, otherwise recalculate fitness. The best individual gene (output $OFS = \{of_1, of_2, \dots, of_n\}$) is the optimal feature subset.

3.4 The LSTM for MW prediction

LSTM aims to solve the problem of long-term dependence [31]. Hence, we apply LSTM to predict MW with the output OFS of EGA feature selection phase. The diagram of LSTM for MW prediction is shown in Fig. 3. LSTM cell is enhanced by three components called gates, including forget gate, update gate, and output gate and two memory cells: hidden state and internal state.

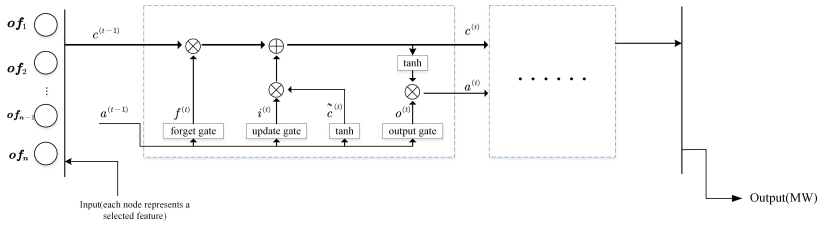


Fig. 3 The LSTM for MW prediction

The selected features are processed by the Min-Max scaler and the squashed data pass the input gate which takes the relevant features from the squashed input data by multiplying them with a sigmoid function. This function maps the relevant features the range 0 to 1. If the value is 0, the network removes the feature, otherwise, the feature pass through the network. The next step is to decide how much memory we need to store in the cell state. Then a tanh layer creates a new vector of new candidate values. The memory gate takes the information stored in the previous state and adds it to the input gate. Since the memory operator is addition instead of multiplication, LSTM avoids the vanishing problem. Moreover, the forget data decides which state of information needs to be memorized or forgotten based sigmoid function. Finally, the output gate decides what the algorithm is going to output based sigmoid function. Then, we put the cell state through tanh layer to push the values to between -1 and 1 and multiply it by the output of the sigmoid gate, so that we only output the parts we are interested in. In the last layer, the output is MW. These operations are described as Eqs. (4-9).

$$\tilde{c}^{(t)} = \tanh(W_c[a^{(t-1)}, x^{(t)}] + b_c) \quad (4)$$

$$i^{(t)} = \sigma(W_i[a^{(t-1)}, x^{(t)}] + b_i) \quad (5)$$

$$f^{(t)} = \sigma(W_f[a^{(t-1)}, x^{(t)}] + b_f) \quad (6)$$

$$o^{(t)} = \sigma(W_o[a^{(t-1)}, x^{(t)}] + b_o) \quad (7)$$

$$c^{(t)} = i^{(t)} * \tilde{c}^{(t)} + f^{(t)} * c^{(t-1)} \quad (8)$$

$$a^{(t)} = o^{(t)} * \tanh(c^{(t)}) \quad (9)$$

4 Experiment

4.1 Data process

The earthquake information of the regions AETA decting is intercepted from January 1, 2017 to December 31, 2022. AETA is a multi-component seismic monitoring system and it is co-developed by the IMS Laboratory of Peking University and our research group School of Computer Science of Wuhan University for earthquake monitoring and prediction. AETA system mainly collects two categories of data, including electromagnetic and acoustic. It has the advantages of strong system stability, high environmental adaptability, and strong anti-interference ability [32]. To better verify earthquake prediction, the dataset consists four regions: DS1((98°E, 34°N), (101°E, 30°N)), DS2((101°E, 34°N), (104°E, 30°N)), DS3((98°E, 30°N), (101°E, 26°N)), DS4((101°E, 30°N),(104°E, 26°N)). The training data starts on January 1, 2017 and ends on February 1, 2022, and the testing data starts on February 1, 2022 and ends on December 30, 2022. In AETA, electromagnetic and acoustic signals are detected every 10 minutes, so, we selected more than 20000 samples in the four years only at one station. Each catalog contains a list of records: time of earthquake, 51 electromagnetic features, and 44 acoustic features. MW data is maintained on the China Earthquake Network Center (CENC). Their catalog is available over the internet at <http://www.ceic.ac.cn/>. Then, we merge the two tables over the time of earthquake. Table 1 shows the partial earthquake data of DS2 after merging.

Table 1 Partial earthquake data features of DS2 after merging

Day	MW	magn@var	magn@power	magn@skew	sound@var	sound@power	sound@skew
1	0.9	107.685	107.745	-0.003	0.00000248	0.00000248	0.284
2	2.1	22.804	22.809	0.010	0.0000019	0.0000019	0.291
3	0.7	63.482	65.360	-0.047	0.0000019	0.0000019	0.281
...
1078	0.9	1.930	1.930	-0.003	0.000320286	0.000320286	0.305
1079	0.2	1.615	1.615	-0.005	0.000319263	0.000319263	0.303
1080	1	1.608	1.608	-0.004	0.00032357	0.00032357	0.300

After merging the dataset, we fuzze multiple pieces of data from one day into one piece of data. The specific operation consists of two steps: selecting the different period of catalogs as the representative data and selecting the maximum of each feature and MW. In this experiment, we choose the *timePeriods* including: 0:00-8:00, 0:00-12:00, and 0:00-24:00. Then, we transform time-series data into supervised learning data based on MW and the time step is one day. Fig. 4 shows the kurt of sound which is one of the 95 features.

Fig. 5 shows the relationship between partial different features. From Fig. 5(b), we found the electromagnetic absolute mean value and electromagnetic absolute maximum 5% position have a strong correlation. Not only these two features, there are also many characteristics that correlate with each other.

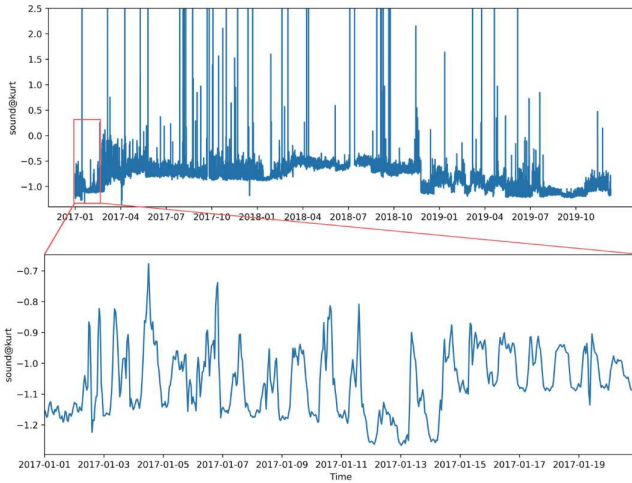


Fig. 4 The wave of the kurt of sound of DS2

Hence, these strong correlation features need to be rejected to decrease time complexity.

4.2 Prediction indicators

To evaluate the performance of our proposed earthquake prediction model, we choose various regression evaluation indicators: mean absolute error (MAE), mean squared error (MSE), RMSE, and R square (R_2). These indicators are calculated as Eqs. (10-12).

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}_i\| \quad (10)$$

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}_i\|_2^2 \quad (11)$$

$$R_2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^n \|y_i - \hat{y}_i\|_2^2}{\sum_{i=0}^n \|y_i - \bar{y}\|_2^2} \quad (12)$$

4.3 Baseline methods

The performance of our proposed EGA-LSTM is compared with other four ML approaches: RF [17], Adaboost [18], LR [33], and SVR [24]. Moreover,

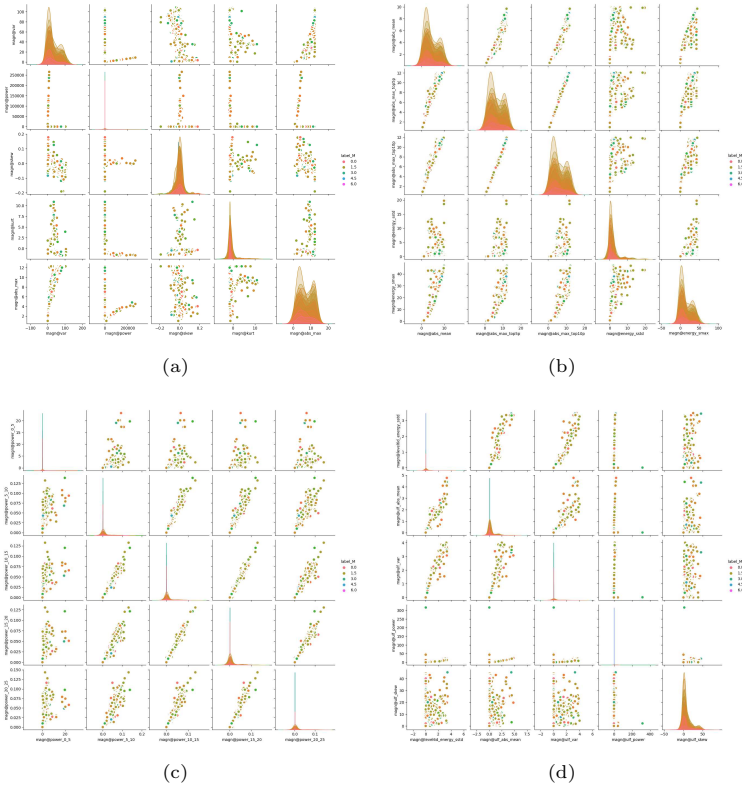


Fig. 5 Analysis of partial features of DS2

we choose two types of GA: SGA, SteadyGA and three different mutation strategies of DEs: operation vector is the best vector and the crossover operation is a random crossover (DE_best_1_b), operation vector is the best vector and the crossover operation is linear order crossover (DE_best_1_L), operation vector is a random vector and the crossover operation is random operation (DE_rand_1_b).

Six EAs adopt binary encoding and set the dimension of individuals to 95 corresponding features. 1 means the corresponding dimension's feature is selected, otherwise without being selected. The number of individuals is set to 10 and the maximum of iteration is set to 20. The threshold of stagnation is set to 0.000001 and the maximum evolutionary stagnation counter is set to 10. In GAs, the crossover rate and mutation rate are set to 0.7 and 0.01. In DEs, the scaling factor and crossbreeding rate are both set to 0.5.

The RF is combined with 100 decision trees. Other parameters are set as follows: max_depth = 10, min_samples_split=2, min_samples_leaf=1. The Adaboost's base learner is decision trees, the number of boosting is 100, and the learning rate is set to 1 and the boosting algorithm is based on the probability of prediction error.

The LR's loss function adopts the least square method. In SVR, the kernel function adopts the radial basis function, the kernel factor is set to the reciprocal of the product of the number of features and the variance of the eigenvector and penalty parameter is set to 1.

The LSTM in our experiment has 100 hidden layers, one output layer, and the activation functions. The learning rate is set to 0.00045, the batch-size is set to 32 and the number of training epochs are set to 120 epochs. The initial weights are random values and the bias is set to 0.

4.4 Result and analysis

This section will demonstrate the result in two subsections. In section 4.4.1, the fitness and RMSE in different periods and EAs are demonstrated. In section 4.4.2, EGA-LSTM is compared with other nine methods based on four metrics. In section 4.4.3, two non-parametric tests were used to prove that EGA-LSTM is different other nine algorithms and EGA-LSTM is superior to LSTM in different datasets.

4.4.1 Different periods and EAs

To find the suitable parameters, we conducted a multi-group ablation experiment with DS2. Tables 2, 3, and 4 refer to the result for time period (*timePeriods*) of 0:00-8:00, 0:00-12:00, and 0:00-24:00, respectively. Fitness is calculated by Eq. (1). Fig. 6 shows the process of decreasing in EGA searching.

Table 2 demonstrates the result for the *timePeriod* of 0:00-8:00, it is worth noting that EGA performs well than other EAs in most groups. Especially, in the group of $\omega_a = 1$, $\omega_F = 0.8$, RMSE is the lowest of all EAs which is a desirable property and in predicting MW. The difference in indicators of F , $RMSE$, and $Fitness$ is markable, producing a better difference greater than 0.03 with the second Fitness (DE_rand_1_b) of RMSE and greater than 0.08 of Fitness. Meanwhile, the number of selected features is only 38 which is lower than 95 (without selecting). Table 3-4 describe *timePeriods* of 0:00-8:00 and 0:00-24:00, respectively, we found RMSE stays between 0.103 and 0.110 and F decreases in different degrees.

By combining the three tables, we take a comprehensive view of F and RMSE. Along with lower the number of selected features, RMSE still performs well. Thus, we conclude that during the period of 0:00-8:00, EGA performs most stable and suitable. In Fig. 5(b), we found that there is a strong correlation between the electromagnetic absolute mean value and the electromagnetic absolute maximum 5% position. To increase predicting accuracy and decrease time complexity, one of these two features needs to be removed. And the best group's result shows the target reached.

Fig. 7 shows the trend of RMSE and Fitness scores as ω_F increased. With ω_F increasing, the penalties of F . Therefore, the number of selected features decreased and the trend of Fitness increased. However, the change of RMSE is not obvious. It always fluctuates in a low range.

Table 2 Fitness and RMSE in different EAs, parameters, and *timePeriod* (0:00-8:00) in DS2

type of EA	$\omega_a = 1, \omega_F = 0$			$\omega_a = 1, \omega_F = 0.2$			$\omega_a = 1, \omega_F = 0.5$			$\omega_a = 1, \omega_F = 0.8$		
	F	RMSE	Fitness	F	RMSE	Fitness	F	RMSE	Fitness	F	RMSE	Fitness
EGA	55	0.103	0.103	39	0.104	0.184	39	0.104	0.299	38	0.101	0.417
SGA	49	0.108	0.108	38	0.108	0.190	37	0.105	0.310	37	0.104	0.424
steadyGA	56	0.104	0.104	39	0.103	0.185	38	0.107	0.307	38	0.106	0.426
DE_best_1.b	72	0.106	0.106	40	0.105	0.189	33	0.109	0.282	44	0.106	0.476
DE_best_1.L	67	0.110	0.110	49	0.104	0.207	38	0.105	0.305	39	0.103	0.431794
DE_rand_1.b	63	0.110	0.110	45	0.103	0.198	38	0.104	0.304	38	0.105	0.425

Table 3 Fitness and RMSE in different EAs, parameters, and *timePeriod* (0:00-12:00) in DS2

type of EA	$\omega_a = 1, \omega_F = 0$			$\omega_a = 1, \omega_F = 0.2$			$\omega_a = 1, \omega_F = 0.5$			$\omega_a = 1, \omega_F = 0.8$		
	F	RMSE	Fitness	F	RMSE	Fitness	F	RMSE	Fitness	F	RMSE	Fitness
EGA	49	0.106	0.106	59	0.107	0.252	39	0.106	0.311	38	0.108	0.722
SGA	43	0.108	0.108	71	0.107	0.256	39	0.107	0.312	39	0.107	0.697
steadyGA	49	0.107	0.107	49	0.111	0.215	43	0.105	0.331	43	0.105	0.636
DE_best_1.b	56	0.107	0.107	76	0.104	0.264	43	0.106	0.333	43	0.107	0.722
DE_best_1.L	48	0.109	0.109	54	0.106	0.220	43	0.106	0.332	42	0.105	0.619
DE_rand_1.b	57	0.110	0.110	55	0.108	0.2244	39	0.105	0.310	39	0.106	0.737

Table 4 Fitness and RMSE in different EAs, parameters, and *timePeriod* (0:00-24:00) in DS2

type of EA	$\omega_a = 1, \omega_F = 0$			$\omega_a = 1, \omega_F = 0.2$			$\omega_a = 1, \omega_F = 0.5$			$\omega_a = 1, \omega_F = 0.8$		
	F	RMSE	Fitness	F	RMSE	Fitness	F	RMSE	Fitness	F	RMSE	Fitness
EGA	50	0.105	0.105	36	0.104	0.184	37	0.107	0.318	38	0.106	0.409
SGA	63	0.105	0.105	38	0.106	0.181	40	0.104	0.299	36	0.108	0.428
steadyGA	58	0.108	0.108	44	0.112	0.204	42	0.106	0.327	43	0.107	0.469
DE_best_1.b	55	0.103	0.103	39	0.105	0.187	38	0.108	0.308	35	0.103	0.397
DE_best_1.L	56	0.106	0.106	43	0.104	0.194	31	0.107	0.270	39	0.104	0.432
DE_rand_1.b	57	0.105	0.105	45	0.109	0.203	43	0.103	0.329	40	0.106	0.443

4.4.2 Comparisons among ML Methods

To verify the performance of EGA-LSTM, we compare it with some state-of-the-art ML methods: RF, Adaboost, LR, SVR, and LSTM.

Table 5 demonstrates the effectiveness between EGA-LSTM and other nine approaches at *timeperiod* of 0:00-8:00 at the region of DS1. It can be found that EGA-LSTM obtains better performance on all indicators. Meanwhile, all the approaches perform badly with the indicator R_2 . Totally speaking, EGA-LSTM outperforms other methods.

Referring to Table 6, we can find that for the MAE, MSE, RMSE and R_2 , EGA-LSTM performs best among the four predictors. For RMSE, it generates a difference greater than 0.011 units with the second best predictor (EGA-SVR). Therefore, EGA-LSTM is an appropriate method to predict MW. The predictor with the worst result is EGA-LR, which obtains much worse performance on the four indicators than other four approaches. That means the time sequence between seismic and precursory features cannot be ignored. It is worth noting that EGA-LSTM's performance for DS2 is the best among the four regions.

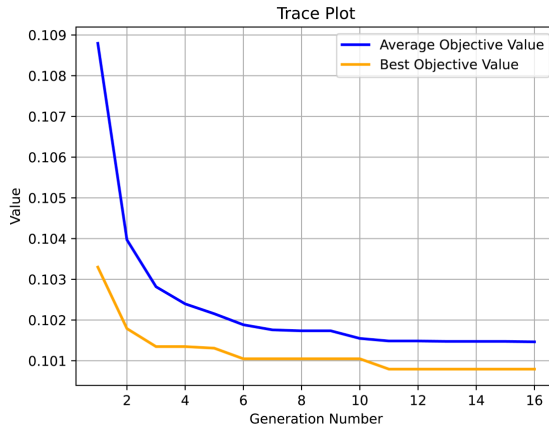


Fig. 6 The process of decreasing in EGA searching

With the reference to DS3(see Table 7), it is worth noting that LSTM without EGA perform best than EGA-LSTM. However EGA-LSTM's results still perform superior than others for the rest indicators. Intuitively, the difference between EGA-LSTM and other algorithms is not obvious, especially EGA-SVR. More statistical comparison details can be seen in section 4.5.

Particular results for DS4 are shown in Table 8. EGA-LSTM perform better than other nine methods for four metrics.

From a joint analysis of the four tables, a conclusion can be easily concluded. For the four indicators, EGA-LSTM is the best one of them. Therefore, it can be concluded that EGA-LSTM is the most precise and stable algorithms for all datasets.

Fig. 8 shows the fitting curves of EGA-LSTM. Since we choose the suitable parameter and loss function, this model is not overfitting and underfitting. The real MW and predicted MW on different methods are shown on Fig. 9, and the result demonstrates that all methods perform well in the range of a low MW. However, the high MW can rarely be predicted accurately.

Table 5 Comparison of effectiveness between EGA-LSTM and other approaches in DS1

	MAE	MSE	RMSE	R_2
RF	0.146422	0.033577	0.183239	-0.057718
Adaboost	0.149681	0.033896	0.184108	-0.067774
LR	0.148834	0.035783	0.189163	-0.127214
SVR	0.144821	0.034977	0.187022	-0.101843
LSTM	0.1469	0.03385	0.183984	-0.06633
EGA-RF	0.147592	0.034421	0.18553	-0.084325
EGA-Adaboost	0.146576	0.033112	0.181968	-0.043088
EGA-LR	0.206224	0.06769	0.260173	-1.132344
EGA-SVR	0.14293	0.032739	0.180938	-0.031317
EGA-LSTM	0.140621	0.031833	0.178418	-0.002788

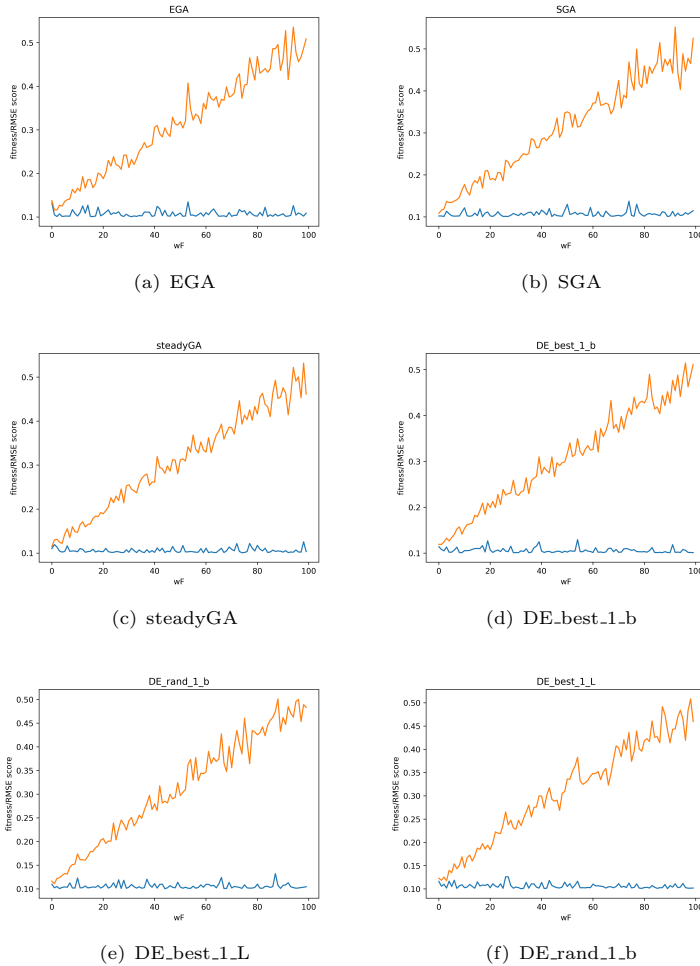


Fig. 7 The trend of RMSE and Fitness score as ω_F

4.5 Non-parametric tests

Since some of the algorithms used were deterministic, we adopt hypothesis tests. To prove the difference of multiple algorithms, we test the performance of multiple algorithms on multiple datasets. Earthquake MW datasets can be used Kruskal-Wallis H test which is a kind of non-parametric test because they did not assume a normal distribution. Therefore, Kruskal-Wallis H test was used to determine whether EGA-LSTM's prediction accuracy significantly different from other algorithms on datasets DS1, DS2, DS3, and DS4. We set the confidence level α 0.05. Finally, we designed that the distribution function of RMSE for group i follows the form $F_i(RMSE) = F(RMSE - \mu_i)$, where

Table 6 Comparison of effectiveness between EGA-LSTM and other approaches in DS2

	MAE	MSE	RMSE	R_2
RF	0.090146	0.012795	0.113116	-0.241414
Adaboost	0.085846	0.011548	0.107461	-0.120391
LR	0.085367	0.012644	0.112446	-0.226753
SVR	0.079376	0.010625	0.103077	-0.030835
LSTM	0.080103	0.010495	0.102447	-0.018286
EGA-RF	0.080753	0.010902	0.104411	-0.0577
EGA-Adaboost	0.089556	0.012042	0.109736	-0.168331
EGA-LR	0.145741	0.029862	0.172806	-1.897247
EGA-SVR	0.078192	0.010457	0.102259	-0.014552
EGA-LSTM	0.077376	0.010226	0.101125	0.007829

Table 7 Comparison of effectiveness between EGA-LSTM and other approaches in DS3

	MAE	MSE	RMSE	R_2
RF	0.13752	0.026332	0.162271	-0.17191
Adaboost	0.138929	0.02688	0.163952	-0.196316
LR	0.152271	0.037694	0.194151	-0.677603
SVR	0.128642	0.023442	0.153108	-0.0433
LSTM	0.121392	0.023618	0.153683	-0.051143
EGA-RF	0.139164	0.027206	0.164944	-0.210831
EGA-Adaboost	0.142054	0.028007	0.167354	-0.246474
EGA-LR	0.160076	0.038701	0.196726	-0.722408
EGA-SVR	0.121748	0.022916	0.15138	-0.019885
EGA-LSTM	0.122361	0.022642	0.150472	-0.007679

μ_i represents the RMSE of a specific algorithm in our experiments. We set the null hypothesis(H_0) and alternative hypothesis(H_1) as Eqs. (13-14):

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_{10} \quad (13)$$

$$H_1 : \text{not all } \mu_i \text{ are equal} \quad (14)$$

The p-value and statistical quantities of Kruskal-Wallis H test from the experimental results was 0.0005 and 29.30. The p-value is less than 0.05 which indicating that we can reject the H_0 and receive H_1 (that indicates EGA-LSTM is different from other algorithms).

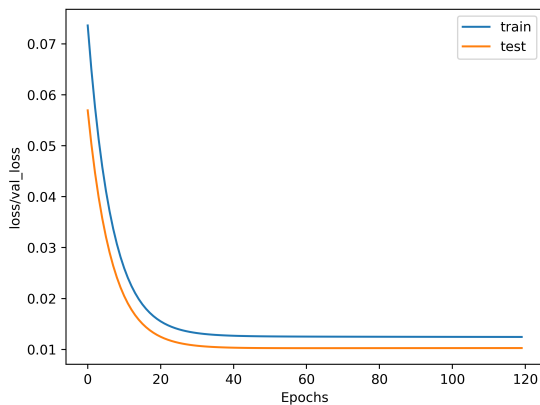
In order to verify the effectiveness of EGA-LSTM versus LSTM, another non-parametric statistical test, the Wilcoxon signed-rank test, was conducted for four different regions between the proposed EGA-LSTM and LSTM. The hypotheses are list as Eqs. (15-16):

$$H_0 : \mu_{LSTM} < \mu_{EGA-LSTM} \quad (15)$$

$$H_1 : \mu_{LSTM} \geq \mu_{EGA-LSTM} \quad (16)$$

Table 8 Comparison of effectiveness between EGA-LSTM and other approaches in DS4

	MAE	MSE	RMSE	R_2
RF	0.109322	0.020374	0.142739	-0.184501
Adaboost	0.107933	0.01903	0.137949	-0.106342
LR	0.107009	0.020991	0.144883	-0.220349
SVR	0.103363	0.018632	0.136498	-0.083187
LSTM	0.100836	0.017347	0.131709	-0.00851
EGA-RF	0.106098	0.019892	0.141041	-0.156479
EGA-Adaboost	0.105577	0.01833	0.13539	-0.065664
EGA-LR	0.11569	0.023812	0.154312	-0.384367
EGA-SVR	0.100364	0.017603	0.132678	-0.023402
EGA-LSTM	0.099404	0.017231	0.131267	-0.001748

**Fig. 8** Plot of epoch and loss of the LSTM model

The μ_{LSTM} and $\mu_{EGA-LSTM}$ represent the RMSE of the two models. H_0 means LSTM perform better than EGA-LSTM and H_1 the opposite. The description and quartile statistics are presented in Table 9 for the two methods. Sub-Table 9(a) shows the basic test statistics, such as maximum, minimum, standard deviation, and other values for the pairwise comparison of EGA-LSTM and LSTM based on each evaluation indicators separately. It can be observed from Table 11 for 20 independent and replicate experiments, that RMSE of EGA-LSTM is better than the opposed to its counterpart LSTM.

In sub-Table 9(b) of Wilcoxon signed-rank, positive ranks, negative ranks are presented for each region of measures corresponding to EGA-LSTM and LSTM. The Wilcoxon signed-rank shows that the results from different regions are different. Checking the Wilcoxon signed-rank's boundary table, the critical value in case of one-tailed test with confidence level of 0.01 is 43, and in case of the confidence level of 0.05 is 60. In DS1, the test statistic is 28, therefore we reject the null hypothesis with 99% certainty that EGA-LSTM is superior to

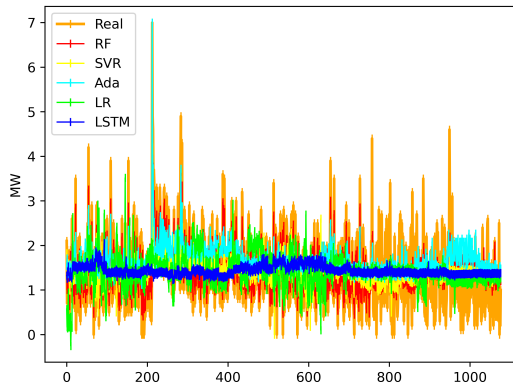


Fig. 9 Comparison between real and predicted MW

LSTM. DS4 is the same as DS1. The test statistic of DS2 is 58 with is greater than 43, we can not reject the null hypothesis with the confidence level of 0.01. However, 58 is less than 60 which is the critical value with confidence level of 0.05. DS3 is the same az DS2. In summary, they all indicate strong evidence for the null hypothesis, that EGA-LSTM is superior than the standard LSTM.

Table 9 Results of Wilcoxon signed-rank test in terms of statistical measures

	N	Mean	Std.	Min	Max	25th	Percentiles	
							50th	75th
a) Descriptive Statistics								
1) DS1								
LSTM-RMSE	20	0.181	4.81E-6	0.178	0.186	0.179	0.181	0.182
EGA-LSTM-RMSE	20	0.179	1.40E-6	0.177	0.182	0.178	0.179	0.180
2) DS2								
LSTM-RMSE	20	0.105	1.50E-5	0.100	0.114	0.102	0.103	0.108
EGA-LSTM-RMSE	20	0.104	2.30E-5	0.100	0.11	0.101	0.102	0.107
3) DS3								
LSTM-RMSE	20	0.157	6.17E-6	0.153	0.162	0.155	0.156	0.158
EGA-LSTM-RMSE	20	0.155	2.04E-6	0.152	0.157	0.154	0.155	0.156
4) DS4								
LSTM-RMSE	20	0.132	0.10E-6	0.130	0.134	0.131	0.132	0.132
EGA-LSTM-RMSE	20	0.131	0	0.130	0.133	0.131	0.131	0.131
b) Wilcoxon Signed-Ranks Test								
1) DS1								
RMSE: LSTM - EGA-LSTM		Positive Ranks	15 ^a	12.13	182	a. RMSE: LSTM > EGA-LSTM		
		Negative Ranks	5 ^b	5.60	28	b. RMSE: LSTM < EGA-LSTM		
		Ties	0 ^c			c. RMSE: LSTM = EGA-LSTM		
		Total	20					
2) DS2								
RMSE: LSTM - EGA-LSTM		Positive Ranks	13 ^a	10.62	138			
		Negative Ranks	7 ^b	10.29	72			
		Ties	0 ^c					
		Total	20					
3) DS3								
RMSE: LSTM - EGA-LSTM		Positive Ranks	13 ^a	12.85	167			
		Negative Ranks	7 ^b	6.14	43			
		Ties	0 ^c					
		Total	20					
4) DS4								
RMSE: LSTM - EGA-LSTM		Positive Ranks	16 ^a	11.50	184			
		Negative Ranks	4 ^b	6.50	26			
		Ties	0 ^c					
		Total	20					
RMSE: LSTM - EGA-LSTM	DS1	DS2			DS3		DS4	
c) Testing								
Test statistic	28		58		47		26	

5 Conclusion and Future Works

Aiming at solving the problem that the redundant features and time series properties hinder the development of earthquake magnitude prediction models, we propose an EGA-LSTM for time series earthquake prediction. First, the acoustic and electromagnetics data of AETA system we developed are fused and preprocessed by EGA to find the strong correlation indicators. Second, since the EGA has the advantages in searching optimal feature subset, we adopt it to selected features. Then, LSTM is implemented to execute magnitude prediction with the selected features, to process time series and complex data. Specifically, we chose RMSE of LSTM and the ratio of selected features as the fitness components of EGA. Finally, we test the proposed EGA-LSTM on the AETA data of Sichuan and Yunnan province. Experimental results demonstrate that all the methods can get the best performance when $timePeriod = 0 : 00 - 8 : 00$, $\omega_a = 1$, and $\omega_F = 0.8$. Moreover, our proposed EGA-LSTM obtains satisfying performance than state-of-the-art approaches on the evaluation indicators MAE, MSE, RMSE, and R_2 .

However, due to the data of medium and large earthquakes belonging to small samples, in order to be able to predict medium and large earthquakes, the model proposed in this study needs to be improved for the data in AETA. Our future work will focus on how to process the time series of small samples to obtain an effective and usable magnitude prediction model, which is suitable for medium and large earthquakes. Apart from that, the theoretical analysis

and more complicated earthquake prediction scenarios of our method can also be part of future work.

6 Ethical Approval

Not applicable.

7 Competing interests

Not applicable.

8 Authors' contributions

Zhiwei Ye and Wuyang Lan and Wen Zhou: Wrote the main manuscript text. Qiyi He: Supervision, Writing- Reviewing and Editing, Funding acquisition. Liang Hong and Xinguo Yu and Yunxuan Gao: Reviewing. All authors reviewed the manuscript.

9 Availability of data and materials

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

10 Funding

The authors want to thank NSFC- <http://www.nsf.gov.cn/> for the support through Grants Number 61877045 and 62202147, and Fundamental Research Project of Shenzhen Science and Technology Program for the support through Grants Number JCYJ20160428153956266, and Research project of the Natural Resources Department of Hubei Province for the support through Grants Number ZRZY2023KJ13.

References

- [1] Bank, W., Nations, U.: Natural Hazards, Unnatural Disasters: the Economics of Effective Prevention. The World Bank, ??? (2010)
- [2] Adeli, H., Panakkat, A.: A probabilistic neural network for earthquake magnitude prediction. *Neural networks* **22**(7), 1018–1024 (2009)
- [3] Asim, K.M., Moustafa, S.S., Niaz, I.A., Elawadi, E.A., Iqbal, T., Martínez-Álvarez, F.: Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in cyprus. *Soil Dynamics and Earthquake Engineering* **130**, 105932 (2020)

- [4] Berhich, A., Belouadha, F.-Z., Kabbaj, M.I.: Lstm-based earthquake prediction: enhanced time feature and data representation. *International Journal of High Performance Systems Architecture* **10**(1), 1–11 (2021)
- [5] Cai, Y., Shyu, M.-L., Tu, Y.-X., Teng, Y.-T., Hu, X.-X.: Anomaly detection of earthquake precursor data using long short-term memory networks. *Applied Geophysics* **16**, 257–266 (2019)
- [6] Kadam, V.J., Yadav, S.S., Jadhav, S.M.: Soft-margin svm incorporating feature selection using improved elitist ga for arrhythmia classification. In: *Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) Held in Vellore, India, December 6-8, 2018, Volume 2*, pp. 965–976 (2020). Springer
- [7] John, G.H., Kohavi, R., Pfleger, K.: Irrelevant features and the subset selection problem. In: *Machine Learning Proceedings 1994*, pp. 121–129. Elsevier, ??? (1994)
- [8] Martínez-Álvarez, F., Reyes, J., Morales-Esteban, A., Rubio-Escudero, C.: Determining the best set of seismicity indicators to predict earthquakes. two case studies: Chile and the iberian peninsula. *Knowledge-Based Systems* **50**, 198–210 (2013)
- [9] Roiz-Pagador, J., Chacon-Maldonado, A., Ruiz, R., Asencio-Cortes, G.: Earthquake prediction in california using feature selection techniques. In: *International Workshop on Soft Computing Models in Industrial and Environmental Applications*, pp. 728–738 (2021). Springer
- [10] Chen, Y., Zhang, J., He, J.: Research on application of earthquake prediction based on chaos theory. In: *2010 International Conference on Intelligent Computing and Integrated Systems*, pp. 753–756 (2010). IEEE
- [11] Cekim, H.O., Tekin, S., Özel, G.: Prediction of the earthquake magnitude by time series methods along the east anatolian fault, turkey. *Earth Science Informatics* **14**(3), 1339–1348 (2021)
- [12] Panakkat, A., Adeli, H.: Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *International journal of neural systems* **17**(01), 13–33 (2007)
- [13] Asim, K., Martínez-Álvarez, F., Basit, A., Iqbal, T.: Earthquake magnitude prediction in hindukush region using machine learning techniques. *Natural Hazards* **85**(1), 471–486 (2017)
- [14] Chanda, S., Raghucharan, M., Reddy, K.K., Chaudhari, V., Somala, S.N.: Duration prediction of chilean strong motion data using machine learning.

Journal of South American Earth Sciences **109**, 103253 (2021)

- [15] Shah, H., Ghazali, R.: Prediction of earthquake magnitude by an improved abc-mlp. In: 2011 Developments in E-systems Engineering, pp. 312–317 (2011). IEEE
- [16] Muhammad, A., Kùlahcı, F., Birel, S.: Investigating radon and TEC anomalies relative to earthquakes via AI models. *J. Atmos. Sol. Terr. Phys.* **245**(106037), 106037 (2023)
- [17] Yang, F., Kefalas, M., Koch, M., Kononova, A.V., Qiao, Y., Bäck, T.: Auto-rep: An automated regression pipeline approach for high-efficiency earthquake prediction using lanl data. In: 2022 14th International Conference on Computer and Automation Engineering (ICCAE), pp. 127–134 (2022). IEEE
- [18] Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F.: Seismic indicators based earthquake predictor system using genetic programming and adaboost classification. *Soil Dynamics and Earthquake Engineering* **111**, 1–7 (2018)
- [19] Zhou, W., Liang, Y., Dong, H., Tan, C., Xiao, Z., Liu, W.: A numerical differentiation based dendritic cell model. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1092–1098 (2017). IEEE
- [20] Zhou, W., Dong, H., Liang, Y.: The deterministic dendritic cell algorithm with haskell in earthquake magnitude prediction. *Earth Science Informatics* **13**(2), 447–457 (2020)
- [21] Zhou, W., Zhang, K., Ming, Z., Chen, J., Liang, Y.: Immune optimization inspired artificial natural killer cell earthquake prediction method. *The Journal of Supercomputing* **2022**, 1–23 (2022)
- [22] Zhou, W., Liang, Y., Wang, X., Ming, Z., Xiao, Z., Fan, X.: Introducing macrophages to artificial immune systems for earthquake prediction. *Applied Soft Computing* **122**, 108822 (2022)
- [23] Moustra, M., Avraamides, M., Christodoulou, C.: Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. *Expert systems with applications* **38**(12), 15032–15039 (2011)
- [24] Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F.: Earthquake prediction model using support vector regressor and hybrid neural networks. *PLoS one* **13**(7), 0199004 (2018)

- [25] Jain, R., Nayyar, A., Arora, S., Gupta, A.: A comprehensive analysis and prediction of earthquake magnitude based on position and depth parameters using machine and deep learning models. *Multimedia Tools and Applications* **80**(18), 28419–28438 (2021)
- [26] Draz, M.U., Shah, M., Jamjareegulgarn, P., Shahzad, R., Hasan, A.M., Ghamry, N.A.: Deep machine learning based possible atmospheric and ionospheric precursors of the 2021 mw 7.1 japan earthquake. *Remote Sensing* **15**(7) (2023). <https://doi.org/10.3390/rs15071904>
- [27] Berhich, A., Belouadha, F.-Z., Kabbaj, M.I.: Lstm-based models for earthquake prediction. In: *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, pp. 1–7 (2020)
- [28] Berhich, A., Belouadha, F.-Z., Kabbaj, M.I.: An attention-based LSTM network for large earthquake prediction. *Soil Dyn. Earthq. Eng.* **165**(107663), 107663 (2023)
- [29] Kavianpour, P., Kavianpour, M., Jahani, E., Ramezani, A.: A cnn-bilstm model with attention mechanism for earthquake prediction. *arXiv preprint arXiv:2112.13444* (2021)
- [30] Jh, H.: *Adaptation in natural and artificial systems*. Ann Arbor (1975)
- [31] Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**(8), 1735–1780 (1997)
- [32] Wanga, J., Yong, S., *et al.*: An aeta electromagnetic disturbance anomaly extraction method based on sample entropy. In: *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, pp. 2265–2269 (2021). IEEE
- [33] Adeli, H., Panakkat, A.: A probabilistic neural network for earthquake magnitude prediction. *Neural Networks* **22**(7), 1018–1024 (2009)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [EQTransformer-master.zip](#)
- [areafeature.rar](#)