A CNN-BiLSTM model with attention mechanism for earthquake prediction

Parisa Kavianpour1 · Mohammadreza Kavianpour2 · Ehsan Jahani1 · Amin Ramezani2

Accepted: 1 May 2023 / Published online: 26 May 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023, corrected publication 2023

Abstract

Earthquakes, as natural phenomena, have consistently caused damage and loss of human life throughout history. Earthquake prediction is an essential aspect of any society’s plans and can increase public preparedness and reduce damage to a great extent. Despite advances in computing systems and deep learning methods, no substantial achievements have been made in earthquake prediction. One of the most important reasons is that the earthquake’s nonlinear and chaotic behavior makes it hard to train the deep learning method. To tackle this drawback, this study tries to take an effective step in improving the performance of prediction results by employing a novel method in earthquake prediction. This method employs a deep learning model based on convolutional neural networks (CNN), bi-directional long short-term memory (BiLSTM), and an attention mechanism, as well as a zero-order hold (ZOH) pre-processing methodology. This study aims to predict the maximum magnitude and number of earthquakes in the next month with the least error. The proposed method was evaluated by an earthquake dataset from nine distinct regions of China. The results reveal that the proposed method outperforms other prediction methods in terms of performance and generalization.

Keywords Earthquake prediction · Convolution neural network · Long short-term memory · Deep learning · Attention mechanism

Mohammadreza Kavianpour
kavianpour@modares.ac.ir

Parisa Kavianpour
p.kavianpour08@umail.umz.ac.ir

Ehsan Jahani
e.jahani@umz.ac.ir

Amin Ramezani
ramezani@modares.ac.ir

1 Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran
2 Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
1 Introduction

Earthquakes are a type of natural catastrophe that can be extremely devastating. They are unexpected events that often occur without warning, imposing human and financial losses on society [1]. In addition, it can cause various other disasters such as flood and avalanche [2], tsunamis [3], and debris collapse [4], and also various consequences for the ecosystem like soil liquefaction [5] and fault rupture [6]. Researchers have been conducting in-depth studies for many years to raise people’s awareness of earthquakes due to the tremendous damage and high mortality [7], as well as the primary and secondary effects of earthquakes [8]. They have proposed various solutions for earthquake prediction [9]. Timely and reasonable prediction is critical to prevent or reduce the destructive effects of this phenomenon, and it aids society in developing more accurate scenarios of the crisis process and taking measures to manage it. In other words, such a prediction improves the level of readiness and speed of action in crisis control in the community. An efficient prediction specifies the magnitude and geographical location of a future earthquake over a period of time [10]. It can save plenty of lives and prevent severe economic losses. Since earthquakes are stochastic and complicated phenomena with a wide variety of parameters that are challenging to examine, various approaches based on the features and parameters affecting the earthquake have been developed. However, a few of these works have achieved accurate predictions [11]. Mathematical modeling, precursor signal analysis, shallow machine learning (ML) techniques, and deep learning (DL) algorithms are some of the most popular methods in earthquake prediction research. Table 1 displays the works related to earthquake prediction in four categories.

The mathematical modeling attempts to formulate earthquake prediction using various statistical and mathematical methods [12]. For instance, Marisa et al. [13] applied a Poisson hidden Markov model (PHMM) to predict the probability of an earthquake on Sumatra island. Dehghani and Fadaee [14] proposed a bivariate lognormal distribution-based probabilistic prediction method. Its parameters are adjusted using the maximum likelihood method. They could predict the probability of an earthquake in Tehran with a magnitude of 6.6–6.8 in the next 10–15 years. Kannan [15] utilizes the spatial connection hypothesis to predict the location of earthquakes in six major seismic regions. Spatial connections theory states that earthquakes within a fault zone are related. Statistical approaches typically need prior assumptions such as stationarity and linear correlation between data, while earthquake data are nonstationarity. Therefore, applying these methods for earthquake prediction frequently does not yield good results [16].

Another branch of research focuses on precursor signals analysis. Radon gas emissions [17], unusual animal behavior [18], electromagnetic signals [19], ionospheric analysis [20], and other anomalous phenomena that may indicate an impending earthquake are examples of precursors. Uyeda et al. [19] provided short-term earthquake predictions based on electromagnetic signal analyses. Li and Parrot [20] claimed that ionospheric density variations could be used as a seismic precursor. In another study, Wikelski et al. [18] examined farm animals’
| Category                        | Authors                        | Year   | Method               | Description                                                                 |
|--------------------------------|-------------------------------|--------|----------------------|-----------------------------------------------------------------------------|
| Mathematical method            | S. Kannan. [15]               | 2014   | Poisson distribution | Predicting the location of the next earthquake according to the theory of spatial connections |
|                                 | Marisa et al. [13]            | 2019   | PHMM                 | Applied a PHMM model to predict the probability of an earthquake in Sumatra island |
|                                 | H. Dehghani and M. J. Fadaee. [14] | 2020   | BLD                  | Predicting the time and magnitude of a future earthquake in Tehran using a model probabilistic |
| Precursor method               | S. Uyeda et al. [19]          | 2009   | statistical analysis | Provide short-term earthquake predictions based on electromagnetic signal analysis |
|                                 | M. Li and M. Parrot. [20]     | 2013   | statistical analysis | Earthquake prediction based on ionospheric density variations               |
|                                 | M. Wikelski et al. [18]       | 2020   | statistical analysis | Investigate the behaviors and activities of farm animals in order to predict short-term earthquakes |
| Shallow machine learning method| I. M. Murwantara et al. [21]  | 2020   | Multinomial LR, SVM and Naive Bayes | Medium- and long-term earthquake prediction in Indonesia using 30 years of historical data |
|                                 | JW. Lin. [23]                 | 2020   | BPNN                 | Probabilistic earthquake prediction in Taiwan using 25 years of historical data |
|                                 | U. Khalil et al. [22]         | 2021   | HNN-SVM              | Earthquake prediction along Chaman fault of Baluchistan based on seismic indicators |
Table 1 (continued)

| Category                      | Authors                      | Year | Method     | Description                                                                 |
|-------------------------------|------------------------------|------|------------|-----------------------------------------------------------------------------|
| Deep learning method          | JP. Huang et al. [27]        | 2018 | CNN        | Major earthquake prediction in Taiwan based on image data                   |
|                               | T. Bhandarkart et al. [30]   | 2019 | LSTM       | Predict future earthquake trends using earthquake historical data           |
|                               | R. Li et al. [29]            | 2020 | CNN        | Earthquake prediction by combining explicit and implicit earthquake features |
|                               | D. Jozinović. [28]           | 2020 | CNN        | Prediction of earthquake ground shaking intensity using raw waveform in Italy |
|                               | Mousavi et al. [34]          | 2020 | CNN-BiLSTM | Earthquake magnitude estimation in Northern California based on earthquake signal |
|                               | Al Banna et al. [32]         | 2021 | BiLSTM-AM  | Earthquake prediction in Bangladesh in the coming month based on eight seismic indicators |
|                               | O. Nicolis et al. [33]       | 2021 | CNN-LSTM   | Earthquake prediction in Chile based on geographic images                   |
|                               | Doğan and Demir. [31]        | 2022 | SRNN       | Earthquake prediction by learning spatial proximity and structural feature based on the fault lines |
behaviors and activities to predict the short-term earthquake. Most of the mentioned techniques rely on particular precursors’ occurrence. Predictions based on precursors rarely produce satisfactory results because they may occur without any further earthquake event and are difficult to detect.

The next research category refers to applying machine learning approaches in predicting earthquakes. These approaches are data-driven, nonparametric, and require fewer a priori assumptions. In their study, Murwantara et al. [21] used three different methods, including multinomial logistic regression (LR), support vector machine (SVM), and Naive Bayes (NB), to predict the magnitude, location, and depth of earthquakes in Indonesia. They confirmed that the SVM algorithm performs better than the other two methods in earthquake prediction. Khalil et al. [22] developed an earthquake prediction approach along the Chaman fault in Baluchistan based on a hybrid neural network (HNN) and SVM approach. Lin. [23] proposed a probability backpropagation neural network (BPNN) for predicting earthquake probabilities in Taiwan. ML approaches have a limited ability to learn nonlinear and complex relationships of earthquake data. In addition, they only can extract shallow features in the dataset and need almost complex feature engineering operations.

DL algorithms have recently made significant progress in solving a wide range of earthquake prediction problems. Because these models include multiple hidden layers and densely connected many neurons, they have a high generalization power, significantly increasing their learning ability compared to shallow networks [24]. Many approaches based on DL have been expanded for earthquake prediction, such as convolutional neural networks (CNN) [25] and long short-term memory (LSTM) networks [26]. Huang et al. [27] suggested the CNN model in their research predicts the magnitude of the major earthquake in Taiwan based on image data. Jozinović et al. [28] offer a CNN-based approach for predicting earthquake ground shaking intensity measurements in Italy. Li et al. [29] introduced a deep learning model for earthquake prediction called DLEP by merging explicit and implicit earthquake characteristics. In DLEP, they employ eight precursory pattern-based indications as the explicit features and utilize a CNN to extract implicit features. Bhandarkar et al. [30] examined the trend of upcoming earthquakes using the LSTM network. Doğan and Demir [31] carried out their studies using a structural recurrent neural network (SRNN) that considers the spatial proximity and structural features. They have focused on two seismic regions in China and Turkey to predict earthquakes. Al Banna et al. [32] have suggested an attention-based bi-directional LSTM architecture for earthquake prediction in Bangladesh in the coming month.

In earthquake prediction, it is critical to collect both the spatial and temporal information of the earthquake data at the same time [31]. Nevertheless, the above studies frequently ignore the spatial or temporal characteristics of earthquake data, which causes some hidden features in the data not to be extracted well. By combining the benefits of CNN and LSTM methods, forecasting performance can be improved. In recent times, several studies for earthquake prediction employing a hybrid CNN and LSTM model have been presented. For example, Nicolis et al. [33] proposed a hybrid CNN-LSTM approach for earthquake prediction in Chile, and they utilized geographic images related to seismic data as input. Their study’s goal has been to predict the location and intensity of
seismic events. Mousavi and Beroza [34] introduced a hybrid method based on the CNN and recurrent neural network (RNN) called CNN-RNN for earthquake magnitude estimation. They examined the time-frequency characteristics of the dominant phases in an earthquake signal from three-component data recorded on individual stations in Northern California and achieved promising results.

The studies mentioned above are concerned with spatio-temporal correlations between earthquake data. Even though spatio-temporal approaches have many advantages over other methods for predicting time series, it is hard to train these models or even figure out the correct pattern of earthquake activity because earthquakes are random and chaotic. This challenge has been neglected in earthquake prediction research. To solve this challenge, in addition to considering the spatio-temporal correlation between earthquake data, this study introduces a new technique called ZOH. The ZOH technique is inspired by the imputation method in missing data and is employed for earthquake data preprocessing to improve the method’s training process. Deep learning methods’ training and prediction results increase with the quality of their input data. Spatio-temporal models do not grasp the significance of features and may omit critical features. The AM can overcome this obstacle by assigning various weights to each feature, enhancing the network’s performance. As a result, this work aims to predict the chaotic behavior of earthquakes by utilizing the spatio-temporal CNN-BiLSTM approach in combination with AM and ZOH techniques to create a method capable of predicting earthquake magnitude for statistically significant periods.

The main contributions of this research can be summed up as follows:

1. A hybrid CNN-BiLSTM-AM method with ZOH technique for earthquake prediction is suggested, which predicts the maximum magnitude and number of earthquakes in one-month period.
2. Because the earthquakes are a complex, nonlinear, and nonstationary time series, zero values in data have a negative impact on model training and make prediction difficult. On the other hand, these values are not part of the model’s behavior. Filling the zero values with the ZOH technique can reduce the effect of earthquake complexity, resulting in better predictions.
3. We have divided China into nine smaller sub-regions to determine the next earthquake’s location range. Each region has a geologically different structure, making it unique in terms of seismic activity. The simulation results demonstrate the proposed method’s efficiency compared to other methods. With the obtained results, it can be seen that our proposed method has achieved good results even in different areas with different data sets.

The rest of this paper is organized as follows. Section 2 is shown the preliminary knowledge. Section 3 is described the proposed method. Experiment and compared method are provided in Sect. 4. Section 5 is presented introduces our case study and results analysis. Section 6 concludes this study.
2 Preliminaries

2.1 Convolutional neural network

CNN has successfully been applied in image processing, condition monitoring, and time series analysis [35]. The CNN architecture is created by stacking three main layers: convolution, pooling, and fully connected (FC). Each convolution layer has a set of learnable kernels whose goal is to extract local characteristics from the input matrix automatically. Kernels perform convolution operations based on two important ideas: weight sharing and local connection, which can aid in reducing computational burden, decreasing model complexity, and enhancing model performance [36]. The pooling layer follows the convolution layer and performs the downsampling operation. One distinguishing characteristic of the pooling layer is that it reduces the feature map’s dimensionality and avoids overfitting. Usually, the FC layers are employed in the last layers of the CNN architecture to learn the nonlinear combination of extracted features by the convolution layer to create the final output [37].

2.2 Bidirectional LSTM

When an RNN is utilized to analyze time-series data, some of the neurons’ outputs can be used as input to the neurons again. The RNN structure contains a return loop to use prior information efficiently. However, RNNs have limitations on memory and information storage. As the sequence period increases, the RNN loses the ability of learning information in the past and is lead to gradient vanishing. [38]. The LSTM network is an enhanced version of RNN that solves the problem of gradient disappearance in RNN by introducing the memory cell and gate mechanism. The LSTM structure uses memory cells to remember long-term historical information and regulates this through a gate mechanism. A common LSTM unit has three types of gates: input gate $i_t$, forget gate $f_t$, and output gate $o_t$. These three gates are shown in Fig. 1. In each gate, the state of memory cells is controlled through point-wise multiplication and sigmoid function operations. $x_t$ at the current state and output $h_{t-1}$ from the hidden state of the previous layer are entered all gates as input. The forget gate decides what information should be ignored or kept. Information from the current input $x_t$ and from the previously hidden state $h_{t-1}$ is transmitted by the sigmoid function. Thus, the forget gate’s output value is between zero and one. If the value is near zero, it indicates that the information will be discarded. Otherwise, the closer to one, the more information will be kept. The formula for the forget gate is calculated as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$  \hspace{1cm} (1)

where $\sigma$ is the sigmoid activation function, and $W$ and $b$ denote the weight and bias of any gate unit, respectively. The current input $x_t$ and the previously hidden state $h_{t-1}$ are fed into the sigmoid function. The input gate decides which information is updated by transforming the values from zero to one. Among them, zero denotes unimportant, and one denotes importance. The input gate is formulated as follows:
A single LSTM usually processes information in just one forward direction. In other words, it can only use past information. In contrast, the structure of BiLSTM is such that it has two layers of LSTM, one in the forward direction and the other in the backward direction. Figure 1 depicts the schematic diagram of BiLSTM. The forward LSTM can receive the input sequence’s past data information, whereas the backward LSTM can obtain the input sequence’s future data information. Then the output from both hidden layers is combined. The hidden state $h_t$ of Bi-LSTM at current time $t$ contains both forward $h_f$ and backward $h_b$:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$  \hspace{1cm} (2)

then the current input $x_t$ and hidden state $h_{t-1}$ are fed to the $\tanh$ function. The cell state $\hat{C}_t$ is calculated, and the new value is updated to the cell state.

$$\hat{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$  \hspace{1cm} (3)

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{C}_t$$  \hspace{1cm} (4)

where $\tanh$ is the hyperbolic tangent activation function. $\odot$ is the dot multiplication operation, and $C_t$ is the new memory cell. Finally, the output gate selects the next hidden state. The new memory cell $C_t$ and new hidden state $h_t$ are then passed to the next time step.

$$o_t = \sigma(W_o[h_{t-1}, p_t] + b_o)$$  \hspace{1cm} (5)

$$h_t = o_t \odot \tanh(c_t)$$  \hspace{1cm} (6)

Fig. 1 Schematic diagram of bidirectional LSTM
where $\oplus$ denotes the summation by component, used to sum the forward and backward output components. Bi-LSTM produces better efficiency than LSTM and RNN because it can use both preceding and subsequent information.

### 2.3 Attention mechanism

AM is an appropriate technique for increasing the importance of crucial information inspired by the human visual system [39]. When human vision observes anything in the environment, it does not usually witness a scene from beginning to end but rather focuses on a specific section as needed. Based on this, the AM selectively focuses on some more influential information, dismisses unnecessary information, and boosts desirable information. AM is commonly used in various fields, such as image captioning [40], machine translation [41], and earthquake prediction [42]. The AM technique acts based on weight allocation, determining the most effective information by distributing higher weights. As a result, it has a positive optimization impact on traditional methods. The nature of the attention function can be defined as a mapping from a query to a sequence of key-value pairs. Furthermore, as seen in Fig. 2, the calculation process of AM involves three phases. In the first phase, the similarity or correlation between the query and each key is calculated as follows:

$$ s_t = \text{tanh}(W_h h_t + b_h) $$

where $s_t$ is attention score. $W_h$ and $b_h$ are the weight and bias of AM, respectively. $h_t$ is the input vector. In the second phase, the obtained score in the first stage is normalized, and the softmax function is utilized to convert the attention score as follows:

$$ h_t = h_t \oplus \hat{h}_t $$

\( (7) \)

\( (8) \)

Fig. 2 The step in determining AM
\[
a_t = \frac{\exp(s_t)}{\sum_i \exp(s_i)}
\]  

(9)

Regarding the weight coefficient, the final attention value is obtained by the weighted summation of values as follows:

\[
s = \sum_t a_t h_t
\]  

(10)

The AM technique is usually used after CNN and RNN networks to focus on the features that significantly influence output variables, increasing the method’s performance.

3 The framework of the proposed method

3.1 Proposed method

This section describes the proposed method architecture of earthquake prediction and its main components. To extract features more effectively and enhance prediction performance, we integrate CNN, BiLSTM, and AM into a single framework and suggest a novel CNN-BiLSTM-AM earthquake prediction method. As shown in Fig. 3, the proposed method consists of five basic blocks: input block, feature extraction block, sequence learning block, attention block, and prediction block. In the feature extraction block, CNN is utilized to extract spatial features from the input data, and these retrieved spatial features are sent into the BiLSTM network as input. The BiLSTM is used to learn long-term temporal information in the sequence learning block.
block, and the results are fed as input to the AM layer. In the attention block, AM assigns various weights based on the model’s feature input, highlighting the effect of the more significant component and assisting the model in generating more accurate decisions. Finally, fully connected layers and the output layer are stacked in the prediction block to carry out the final prediction. These parts contain hyperparameters such as kernel size, number of kernels, loss function type, and number of neurons. The proposed method’s prediction error can be reduced by adjusting these parameters properly. The structural details of the proposed method are given in Table 2. In the following, each block of the proposed method is described in detail:

- (1) **Input block** The input data contain the earthquake information per month. We are dealing with time series data containing zero values due to the inconsistent pattern of earthquakes and their nonoccurrence in some months. This factor damages and misleads the model training process, leading to less accurate models and poorer results. The ZOH technique means that a sample value is kept constant in time until a new value is obtained. The last nonzero value from the month prior is used to replace each month’s zero values in this study, which was motivated by this methodology. Therefore, the input data are organized in a way that is appropriate for the deep learning algorithm. This method is a suitable preprocessing technique for earthquake prediction because it helps to improve the training process and prediction performance by reducing the effect of the zero value of the data.

- (2) **Feature extraction block** One-dimensional CNN obtains the feature extraction block with nine layers, including four convolutional layers, four pooling layers, and one Flatten layer. Unlike the subsequent layers, the filter size of

| Block                          | Layer   | Number/ size/stride of kernels or number of neurons |
|-------------------------------|---------|-----------------------------------------------------|
| Feature extraction block      | Conv1   | 16*32*1                                             |
|                               | Max-pooling | 16/2/1                                           |
|                               | Conv2   | 32*3*1                                              |
|                               | Max-pooling | 32/2/1                                           |
|                               | Conv3   | 64*3*1                                              |
|                               | Max-pooling | 64/2/1                                           |
|                               | Conv4   | 128*3*1                                             |
|                               | Max-pooling | 128/2/1                                          |
|                               | Flatten | –                                                   |
| Sequence learning block       | BiLSTM1 | 128                                                 |
|                               | BiLSTM2 | 64                                                  |
| Attention block               | Attention | –                                                   |
| Prediction block              | FC1     | 32                                                   |
|                               | FC2     | 10                                                   |
|                               | FC3     | 1                                                    |
the first convolution layer is selected wide. Compared to small kernels, this structure is superior at damping high-frequency signals [43]. Stacking several convolutions and pooling layers allows higher-level features to be extracted from the input, which helps represent the input data better. The Max-pooling layer is implemented after each convolution layer to reduce the dimensions and parameters within the network. In the feature extraction block, rectified linear unit (ReLU) is used as the activation function to avoid gradient vanishing or explosion problems while enhancing the convergence rate. Following each pooling layer, a batch normalization (BN) algorithm is employed as an effective regularization strategy. In addition to having a regularizing effect, it can reduce the shift of internal covariance, improve the network’s training performance, and increase the generalization capability of the network. BN is a feature normalization method in a layer-by-layer manner that can normalize any intermediate layer of a neural network and reduce sample differences between layers. As a result, this technique aids in the acceleration of the training process [44]. The specific formulas are as follows:

\[
\mu = \frac{1}{N_{\text{batch}}} \sum_{i=1}^{N} x_i
\]  

(11)

\[
\sigma^2 = \frac{1}{N_{\text{batch}}} \sum_{i=1}^{N} (x_i - \mu)^2
\]  

(12)

\[
\tilde{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}
\]  

(13)

\[
y_i = \gamma \tilde{x}_i + \beta
\]  

(14)

where \(N_{\text{batch}}\) defines the size of mini-batch, \(x_i\) and \(y_i\) are the input and output of the \(i\)th observation value in the mini-batch. \(\mu\) denotes the mean value of the mini-batch sample. \(\sigma^2\) is the variance of the mini-batch sample. \(\epsilon\) is a constant close to zero to ensure numerical stability, \(\gamma\) is a scaling parameter, and \(\beta\) is a bias parameter. The padding type is chosen the same to ensure that no features are lost during convolution operations. Since the input data of the BiLSTM network are required to be a one-dimensional array, the multidimensional output data of the convolution layer need to be flattened into one-dimensional data by the Flatten layer.

• **(3) Sequence learning block** The sequence learning block aids in the learning of the temporal patterns of extracted characteristics via the feature extraction block. The sequence learning segment contains two BiLSTM layers and two dropout layers. The dropout technique is used after each BiLSTM layer to prevent the over-fitting problem. Dropout means that instead of training all of the neurons in the network, only a subset of them is chosen randomly and trained.
In short, a certain percentage of neurons in each iteration of the training step take zero output and are inactivated. It drives the network to learn more effective features and boosts the model’s generalization capacity.

- **(4) Attention block** An attention layer is added at the end of the sequence learning block, which assigns different weights to those hidden states that have different contributions to earthquake prediction. So, a high-level vector can be obtained summarizing all the useful earthquake information.

- **(5) Prediction block** The prediction block is made up of two fully connected layers as well as an output layer. The fully connected layer performs a series of nonlinear transformations on the values of the features obtained by the attention block. In the end, the final prediction results are generated.

### 3.2 The overall procedure of the proposed CNN-BiLSTM-AM method

In general, our proposed method in this paper contains three main phases: data preparation, model training, and model evaluation, which is shown in Fig. 4. The first step in the data preparation phase is collecting data from existing sources. Subsequently, the collected data are normalized and divided into two parts of training and testing data. Training data are pre-processed using the ZOH technique. It should be noted that ZOH pre-processing only affects the training process, and no changes have been applied to the test data. The second step is to build and train the proposed method. In this step, first, the parameters of the constructed model are determined with the initial values. The spatial features of the input data are obtained through the feature extraction section and transferred to the layers of the sequence learning. The sequence learning section extracts temporal features. The AM technique highlights the more influential features in the prediction results. The parameters are updated in each epoch utilizing the backpropagation algorithm, and this training process will continue until the maximum epoch is achieved. The third step is model evaluation, in which the trained model is used to predict test data. Finally, the proposed method predicts earthquakes using these three steps.

### 4 Experiment and comparisons

#### 4.1 Division study area

This paper has chosen mainland China as the region of interest, situated southeast of the Eurasian plate. Mainland China is linked to the Siberia-Mongolia sub-plate, the Philippine, and the India plate; it is regarded as one of the most seismically active regions in the world. This country has already experienced many large and destructive earthquakes, such as the 1966 Xingtai earthquake ($M_w 7.4$), the 1974 Daguan earthquake ($M_w 7.1$), the 1976 Tangshan earthquake ($M_w 7.6$), the 2002 Jilin earthquake ($M_w 7.2$), the 2008 Wenchuan earthquake ($M_w 8.0$), and the 2013 Lushan earthquake($M_w$...
Fig. 4 The flow chart of the proposed method
Since 1949, more than 100 catastrophic earthquakes have occurred in mainland China. These earthquakes caused the deaths of more than 270,000 people, accounting for 54% of the overall death toll from natural catastrophes in mainland China [45]. Therefore, reliable and effective predictions in this area can help to reduce the damage and casualties caused by earthquakes. One of the aims of the earthquake prediction problem is to predict and identify regions where major earthquakes occur. The exact location of the epicenter is usually predicted with a relatively high error. On the other hand, due to the significant dimensions of faults, when an earthquake occurs, a relatively large level of a region is affected [46]. Based on studies [26] and [31], in order to analyze and more accurately predict the range of location of the next earthquake, mainland China is divided into nine small areas. The latitude ranges from 23 to 45 degrees, and the longitude ranges from 75 to 119 degrees; the range of latitude and longitude is divided into three equal parts. Figure 5 depicts the nine study areas.

4.2 Dataset and data preprocessing

All earthquake information should be collected in the desired range to characterize regional seismicity’s main features. The database of earthquakes used in this paper has been obtained from the US Geological Survey (USGS) and National Seismological Center (NSC) websites. In order to produce a quality seismic catalog, any duplicate data from the rows are recognized and eliminated. The earthquake catalog consists of 11442 cases with a magnitude greater than or equal to 3.5, from January 15, 1966, to May 22, 2021. Each recorded earthquake contains vital information such as latitude and longitude, time of earthquakes, magnitude and depth of earthquakes, and station number. In addition to the importance of model learning structure, proper input and output adjustment also significantly impacts model learning performance. Therefore, this study intends to investigate its impacts through two case studies with different input and output. In the present study, the input and

Fig. 5 A visual representation of the division of China’s nine study areas
output data for case study 1 are the number of earthquakes per month, and case study 2 is the maximum magnitude of earthquakes per month. Notice that when the one-month interval is considered, there are 665 data samples.

The Min-Max standardization [47] was used to transform the data into a specific range [0, 1]. Furthermore, 80% of the data is utilized for model training, with the remaining 20% used for testing.

4.3 Implementation details

The selection of the model’s hyperparameters is crucial to the well-trained model. Most of the hyperparameters are adjusted through trial and error. For example, several optimization algorithms are compared, including stochastic gradient descent (SGD) [48], RMSprop [49], and Adam [50]. According to the comparison results, Adam can enhance the suggested model’s accuracy and is chosen as an optimization algorithm. Mean square error (MSE) is utilized as a loss function that can be back-propagated to update the weights and biases. The starting learning rate is adjusted to 0.001 and gradually decreases linearly to 0.0001 by the final epoch. It helps to keep a relatively stable pace during the learning process. The epoch number is 150, and the batch size is 32. Each experiment is repeated ten times to limit the influence of random factors on network performance. Since each region has different geological characteristics and activities, separate training is necessary. Therefore, all nine regions use the same configurations and settings to train the network. The prediction results were obtained for each of the nine areas individually on the test dataset.

4.4 Evaluation metrics

In this study, the three evaluation criteria, including root-mean-square error (RMSE), mean absolute error (MAE), and R-squared ($R^2$), are employed to analyze the performance of the proposed method and compare it fairly with other methods. The equations of RMSE, MAE, and $R^2$ are represented as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$  \hspace{1cm} (15)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$  \hspace{1cm} (16)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$  \hspace{1cm} (17)
In the aforementioned formulas, \( n \) is the number of predicted data; the predictive value is \( \hat{y}_i \); the actual value is \( y_i \), and \( \bar{y} \) is the average of actual value. RMSE, MAE, and \( R^2 \) are the most common evaluation criteria for regression issues used in this study. For the proper performance of the proposed method, RMSE and MAE should be minimized, while \( R^2 \) needs to be maximized. The \( R^2 \) metric indicates the strength of the linear regression between observed and predicted values. When \( R^2 \) equals one, the strongest linear relationship occurs. The RMSE and MAE metrics indicate the performance of the models. Both metrics can have a value ranging from zero to infinite, with zero indicating the highest performance.

4.5 Comparison method

We have used two categories of methods to evaluate the performance of the proposed model: shallow machine learning and deep learning. The methods commonly used in shallow machine learning for earthquake prediction are SVM, multi-layer perceptron (MLP), decision tree (DT), and random forest (RF). In deep learning methods, CNN, LSTM, and CNN-BiLSTM have satisfactory performance at prediction for time series data. Therefore, SVM, MLP, DT, RF, CNN, LSTM, and CNN-BiLSTM are chosen as the comparison methods for our proposed method. All comparison methods utilized a similar dataset and prediction process to the proposed method. As stated, comparison methods are divided into two major categories:

- **Shallow machine learning**: SVM, DT, MLP, and RF are considered in the shallow machine learning category. These methods are proposed to evaluate the feature extraction ability of the proposed method compared to machine learning methods. One of the machine learning approaches capable of classification and regression is the SVM, which is based on statistical learning theory [51]. The main hyperparameters of the SVM model are set as follows: The type of kernel function is chosen radial basis function (RBF) because it has fewer hyperparameters, which decreases the complexity of the model. The regulation factor (C), epsilon (\( \epsilon \)), and gamma (\( \gamma \)) hyperparameters are set to 1, 0.01, and 0.1, respectively. RF is a powerful ensemble learning approach that uses averaging to improve prediction performance while avoiding overfitting. It is extensively used in a diverse variety of regression issues [52]. In this experiment, the maximum depth of the trees and the number of trees in the forest are set to 9 and 100, respectively. A DT algorithm is a supervised learning approach that uses a tree structure to generate regression or classification models. The objective is to build a model that predicts the value of a target variable using simple decision rules inferred from data features. In this experiment, the maximum depth of the tree is chosen to be 10. The MLP approach is also appropriate for regression prediction issues, in which a real-valued quantity is forecasted given a set of inputs and is often used in earthquake prediction studies. The MLP approach is composed of three major layers: one input layer, some hidden layers, and one output layer [53]. In our example, we employ two hidden layers with 15 neurons each. The
learning rate is set at 0.01. The activation function is sigmoid, and the number of epochs is 150.

- **Deep learning**: The deep learning category includes CNN, LSTM, and CNN-BiLSTM methods. The CNN method is related to the feature extraction block of the proposed method. Also, the LSTM method is similar to the proposed method’s sequence learning block, except that LSTM is used instead of BiLSTM. The CNN-BiLSTM approach is built with a combination of feature extraction block and sequence learning block.

In addition, all comparative methods without the ZOH technique are performed to show the effectiveness of the proposed ZOH technique in earthquake prediction.

## 5 Result

### 5.1 Case 1: Number of earthquakes

The majority of research used seismic indicators, magnitude, depth, and geographical location of earthquakes as input and has only been able to predict the earthquake’s time, location, and magnitude. To the best of our knowledge, no predictions concerning the number of earthquakes have been proposed. The number of earthquakes can be an essential factor in predicting an earthquake that can assist in portraying a more accurate picture of a region’s seismicity. The proposed method’s input includes a series of earthquakes that occur each month, calculated based on past earthquakes in each of the nine regions. The output of the proposed method is also the prediction of the number of earthquakes in the following month. This case study is used with the same conditions and assumptions as the proposed method to predict the number of expected earthquakes in a month. The proposed method is compared with CNN-BiLSTM, LSTM, CNN, RF, MLP, DT, and SVM to verify the efficiency, superiority, and generalization ability.

Table 3 shows the results of the proposed method and comparison methods in predicting the number of earthquakes for nine different regions. It is clear from the outcomes that the proposed method consistently achieves the best prediction performance in all regions, with the lowest RMSE and MAE values and the highest $R^2$ score. This result shows that the proposed method appropriately predicts the number of earthquakes and is superior or competitive to other comparison methods. For example, in region 1, the comparison models have substantial prediction errors, but the proposed method has an RMSE of 0.24, an MAE of 0.018, and an $R^2$ of 0.956, which improves prediction outcomes. In one of the most challenging regions in terms of $R^2$ metric, namely region 7, the proposed method’s $R^2$ value is 0.133 and 0.5 greater than the maximum and minimum in comparison methods, respectively. This strong performance demonstrates the effectiveness of the proposed method when there is a significant seismic anomaly. The proposed method’s superior performance is due to three crucial components that can considerably boost prediction performance: 1) adopting the ZOH technique as an appropriate preprocessing, 2) learning the spatial and temporal characteristics using the hybrid CNN-BiLSTM
model, 3) using the AM concept to examine the value of various hidden states, with a greater emphasis on the most significant states. As a result, these components make the proposed method superior to both ML-based and DL-based approaches. On the other hand, the SVM model has the worst prediction performance in all regions, with the highest RMSE and MAE values and the lowest $R^2$ score. The SVM method fails to catch the long-term correlation in the data series, despite its ability to solve nonlinear problems.

Furthermore, we present the mean of the evaluation metrics for each implemented method to highlight our proposed method's overall benefits over other methods for the number of earthquakes in all regions. As shown in Fig. 6, there are evident disparities in the performance of various techniques. Specifically, deep
learning-based methods have a lower average error and more substantial predictive potential than shallow machine learning-based methods. This fact is that the shallow machine learning methods structure is relatively simple, while there is a nonlinear relationship between the earthquake data and the variables related to it. Therefore, deep learning-based approaches are more adaptable in modeling complicated and nonlinear earthquake interactions. They can provide the capability of feature learning at different levels of abstraction by inserting several hidden layers. Additionally, the LSTM can enhance earthquake prediction performance more than the CNN among DL comparison models since the LSTM takes temporal dependence into account. The hybrid CNN-BiLSTM method has the lowest error values and highest $R^2$ score compared with CNN and LSTM single because the CNN-BiLSTM hybrid network addresses the corresponding problem by using the advantages of CNN and BiLSTM. The proposed method performed better than the CNN-BiLSTM method for the number of earthquake predictions. Overall, our proposed method shows improvements of 3.07% and 1.1% beyond the CNN-BiLSTM method on RMSE, MAE, and $R^2$ metrics, respectively. This improvement is due to the effectiveness of the two proposed ZOH and AM schemes. In other words, this demonstrates that introducing AM and ZOH into the original CNN-BiLSTM model can fully explore the complex relationship and enhance the earthquake prediction performance.

5.1.1 Analysis of the effect of ZOH and AM in the process of training and testing

Figures 7 and 8 illustrate the comparison of the proposed method and CNN-BiLSTM in predicting the number of earthquakes in Region 5 during the methods’ training and testing phase. The prediction results are mainly dependent on pattern sequences in the time series. Nevertheless, identifying the appropriate pattern in time series
sequences is challenging because of earthquakes’ complex and nonlinear characteristics. As can be seen in Fig. 8, the combination of ZOH and AM techniques has been able to have a positive impact on the training and testing process, and it helps the method learn a complex seismic pattern successfully. Therefore, when the method is well-trained, its efficiency and performance at the output increase, making more accurate predictions. As shown in Fig. 7, the proposed method is better fitting than CNN-BiLSTM and, in most cases, closer to the observed values.

5.1.2 Analysis of the effect of ZOH and AM on predicting the number of earthquakes based on $R^2$

An ablation study has been considered to evaluate the performance of ZOH and AM techniques in terms of $R^2$ metric, which is illustrated in Fig. 9. From the unevenness of the radar diagram, it can be seen the effect of adding each of the ZOH and AM techniques in improving the results of the CNN-BiLSTM method.

5.1.3 Analysis of the effect of choosing the preprocessing method

The proposed approach has been compared with the CCN-BiLSTM+AM+Max, CNN-BiLSTM+AM+Mean, and CNN-BiLSTM+AM methods to assess the effectiveness of the ZOH preprocessing method provided in this research for predicting the number of earthquakes. The CCN-BiLSTM+AM+Max approach is similar to the proposed method, except that instead of ZOH, zero values are replaced with the highest value of the training data. Similar to the proposed method, the CCN-BiLSTM+AM+Mean method substitutes the mean value of the training data for zero values in place of ZOH. In the suggested method, we do not apply ZOH preprocessing in the CNN-BiLSTM+AM method either. Figure 10 depicts the simulation results according to the $R^2$ criteria. As can be observed, the proposed method’s $R^2$ value is much higher than that of other pre-processing techniques like maximum and mean. Additionally, the results of the comparative approaches are not robust across regions, in contrast to those of the proposed method. The CNN-BiLSTM+AM+Mean approach performs better in some regions than other comparable methods, while CNN-BiLSTM+AM performs better in some regions than

![Fig. 7 The comparison of proposed method and CNN-BiLSTM method for prediction of earthquake’s number in region 5](image-url)
A CNN-BiLSTM model with attention mechanism for earthquake…

comparative methods. One of these methodologies’ shortcomings is their inability to adapt to changing locations and shifting data distribution. In contrast, the proposed method performs well across all areas.

5.1.4 Analysis of the effect of region division

It has always been crucial for deep learning approaches to be generalizable. The appropriate performance for a portion of the given dataset is one of the generalizability criteria for the proposed method. The study region in Figure 5 is divided into 36 equal portions to examine the impact of different area divisions on the effectiveness of the proposed method. A region that is a subset of region 4 with latitude
ranges from 89.66 to 96.99 and longitude ranges from 23 to 26.67 has been chosen to more accurately assess the performance of the proposed approach on the sub-dataset. The performance of the proposed method and comparison methods for predicting the number of monthly earthquakes in this region is shown in Fig. 11. As can be seen, even though our dataset has changed in general, which has resulted in changing the probability distribution of data in the training and testing phase, the proposed method has performed better than other methods. The better performance of the proposed method shows the high generalizability of this method, even for smaller areas.

5.2 Case 2: Maximum magnitude of earthquake

A major earthquake with a magnitude greater than five can cause significant financial and human losses and cause great concern to society. The existence of these destructive earthquakes in the history of mainland China highlights the significance of the maximum earthquake magnitude prediction in this earthquake-prone area. Also, a timely prediction of an earthquake’s maximum magnitude can be an effective step in giving early earthquake warnings and reducing damage caused by its occurrence. Because major earthquakes occur in the region over a period of months, a prediction time span of one month is appropriate in this study [54]. However, there are situations where more than one major earthquake strikes the region in a month, but the algorithm only counts one. The overall percentage of months with major earthquakes is about 35%, with the other months recording mainly
low-level earthquakes. In case study 2, we evaluate the performance of our proposed method for earthquake’s maximum magnitude prediction. The input data contain the sequence data of the maximum magnitude of earthquakes every month for a span of time, and the output data include the maximum magnitude of earthquakes predicted for the next month.

Table 4 shows the simulation results of the proposed method and comparison models for predicting the maximum earthquake magnitude in nine areas. The results demonstrate that the proposed method consistently performs the best prediction in all areas, with the least RMSE, MAE values, and highest $R^2$ value. This superiority is due to the three advantages of the proposed method: The spatial features and temporal correlations extracted by utilizing CNN-BiLSTM are considered to reflect the hidden features of the earthquake, and the AM can highlight features that are important for earthquake prediction. The ZOH technique can help to reduce the effect of data’s zero value on the network training process and reduce prediction error. For example, the proposed method enhances results for regions 1 and 2 compared to other methods’ performance. The $R^2$ value of proposed methods in these regions is 0.69 and 0.545 greater than the best $R^2$ value of the two categories, machine learning-based and deep learning-based methods. Moreover, the proposed method has shown the best prediction performance in one of the most challenging regions, such as region 9 with $R^2$ of 0.801, while the worst prediction performance in this region is related to the DT model with $R^2 = 0.111$. The proposed method has improved the $R^2$ value by 0.69 and 0.545 higher than the best $R^2$ value of the two categories, machine learning-based and deep learning-based. This high performance indicates the proposed method’s efficiency when there is a substantial distribution difference between train and test data. The difference in the probability and statistical distribution of the data is related to variances in seismic behavior in various regions caused by changes in the arrangement of their tectonic plates.

To show the results more clearly in Table 4, we calculate the values of the evaluation indices of all methods as averages and draw them in Fig. 12. Overall, all the deep learning models outperform the traditional models, and the proposed method has the best performance. Take RMSE and MAE as an example. SVR, DT, MLP, RF, CNN, LSTM, CNN-BiLSTM obtain (0.326, 0.23), (0.272, 0.209), (0.309, 0.192), (0.264, 0.212), (0.229, 0.176), (0.217, 0.166), and (0.191, 0.149) at RMSE and MAE, respectively, while proposed method reduces RMSE and MAE to (0.074, 0.076), respectively. For $R^2$, the proposed method also achieves a higher value (0.906) than the maximum values of the other comparable models. These minimum values of RMSE and MAE, and maximum value of $R^2$ confirm the effectiveness of the ZOH and AM-based prediction method. While the proposed method has more parameters than comparison methods, the computational cost is justified by the proposed method’s superior performance over other methods.

5.2.1 Analysis of the effect of ZOH and AM in the process of training and testing

Figures 13 and 14 demonstrate the comparison of the proposed method with CNN-BiLSTM in predicting the maximum magnitude of earthquakes in region 1 during the methods’ training and testing phase. As can be seen, the high performance of the
The proposed method in the training process and the effective learning of the complex earthquake pattern have caused the proposed method to perform the earthquake prediction process with less error. As a result, when compared to the CNN-BiLSTM method, our proposed method can correctly detect the trend of earthquake time series, especially all peaks. Because the peaks actually represent the maximum magnitude of the monthly earthquakes, which is the purpose of our issue.
5.2.2 Analysis of the effect of ZOH and AM on predicting the maximum magnitude of earthquakes

In order to investigate the effect of ZOH and AM techniques and thus gain a better understanding of their performance, a radar diagram is drawn in Fig. 15. By comparing the results in Fig. 15, we realize that while AM can enhance the model’s performance, it is not as efficient as the ZOH technique.
5.2.3 Analysis of the effect of choosing the preprocessing method

The proposed method is compared with three other approaches similar to Subsect. 5.1.3 in case study 1 to assess the performance of the ZOH preprocessing method for predicting the maximum magnitude of earthquakes. This comparison takes its cue from the imputation method in missing data problems, which substitutes the average or maximum value instead of data with a zero value. Figure 16 depicts the comparison’s outcomes based on the $R^2$ criterion. As can be observed, the suggested method has outperformed other methods over all regions. These findings highlight the significance of using the appropriate preprocessing technique. In some cases, CNN-BiLSTM+AM outperformed approaches that use maximum or average. In light of these findings, it could be concluded that not all preprocessing techniques yield effective outcomes.
5.2.4 Analysis of the effect of region division

The studied area is considered similar to Subsect. 5.1.4 of case study 1 to evaluate the performance of the proposed method for predicting the maximum magnitude of earthquakes in smaller areas. The results of this experiment are given in Fig. 17. Despite the shrinking of the areas, which makes earthquake prediction difficult for any method, the proposed method has performed better than others. One of the challenges of small regions is the lack of sufficient high-magnitude earthquake data, which harms the performance of deep learning and machine learning methods. Despite this problem, the proposed method has been able to perform robustly even for smaller areas, which shows its high generalizability.
Fig. 15  Radar diagram of ablation study in terms of $R^2$ metric for maximum magnitude prediction

Fig. 16  Comparison between different preprocessing techniques for predicting the maximum magnitude of earthquakes based on $R^2$ criteria
6 Conclusion

Due to the nonlinearity and complexity of earthquake data, this paper presents a novel CNN-BiLSTM-AM method and an efficient general framework for earthquake prediction in terms of number and maximum magnitude. The number and maximum magnitude of earthquakes that occurred each month over the past 50 years are considered the model’s input features, making the model extract efficient information from the historical data. As a new data processing technique, ZOH is presented to train the network better and reduce the prediction difficulty. After data preprocessing, CNN is used to extract spatial characteristics and then fed them into BiLSTM to model temporal characteristics corresponding to the input time series data. AM is designed for BiLSTM to focus on the characteristics that significantly contribute to the prediction results to improve the model’s accuracy. Finally, the output of the AM is sent to the fully connected layers to obtain the final result. The complex spatio-temporal correlations among the input data are identified by effectively training the proposed model. The simulation results in two case studies reveal that the proposed method has the best performance compared to other shallow machine learning and deep learning approaches.

Author Contributions PK helped in writing—original draft preparation, conceptualization, methodology, software, visualization. MK contributed to supervision, methodology, software, visualization, writing—review and editing. EJ and AR were involved in writing—review and editing, validation, investigation.

Funding The authors did not receive support from any organization for the submitted work.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.
Declarations

Conflict of interest  Authors declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethical approval  The authors confirm that they approve and accept all the ethical rules of the journal. Also, the authors confirm that this article has not been published anywhere before.

References

1. Alarifi AS, Alarifi NS, Al-Humidan S (2012) Earthquakes magnitude prediction using artificial neural network in northern red sea area. J King Saud Univ -Sci 24(4):301–313
2. Cui P, Chen X-Q, Zhu Y-Y, Su F-H, Wei F-Q, Han Y-S, Liu H-J, Zhuang J-Q (2011) The Wenchuan earthquake (May 12, 2008). Sichuan province, China, and resulting geohazards. Nat Hazards 56(1):19–36
3. Jain N, Virmani D, Abraham A (2019) Proficient 3-class classification model for confident overlap value based fuzzified aquatic information extracted tsunami prediction. Intell Dec Technol 13(3):295–303
4. Cui P, Xiang L-Z, Zou Q (2013) Risk assessment of highways affected by debris flows in Wenchuan earthquake area. J Mt Sci 10(2):173–189
5. Verdugo R, González J (2015) Liquefaction-induced ground damages during the 2010 Chile earthquake. Soil Dyn Earthq Eng 79:280–295
6. Bray JD, et al (2001) Developing mitigation measures for the hazards associated with earthquake surface fault rupture. In: Workshop on seismic fault-induced failures-possible remedies for damage to urban facilities. University of Tokyo Press, pp 55–79
7. Bilham R (2009) The seismic future of cities. Bull Earthq Eng 7(4):839–887
8. Jia J (2017) Earthquake damages. Modern earthquake engineering. Springer, Berlin, pp 413–431
9. Yousefzadeh M, Hosseini SA, Farnaghi M (2021) Spatiotemporally explicit earthquake prediction using deep neural network. Soil Dyn Earthq Eng 144:106663
10. Kail R, Burnaev E, Zaytsev A (2021) Recurrent convolutional neural networks help to predict location of earthquakes. IEEE Geosci Remote Sens Lett 19:1–5
11. Kavianpour P, Kavianpour M, Jahani E, Ramezani A (2021) Earthquake magnitude prediction using spatiotemporal features learning based on hybrid CNN- BiLSTM model In: 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS), pp 1–6
12. Boucouvalas A, Gkasios M, Tselikas N, Drakatos G (2015) Modified-fibonacci-dual-lucas method for earthquake prediction. In: Third International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2015), vol 9535, p 95351A, International Society for Optics and Photonics
13. Marisa, Sembiring UA, Margaretha H (2019) Earthquake probability prediction in sumatra island using Poisson hidden Markov model (HMM). In: AIP Conference Proceedings, vol 2192, p 090006, AIP Publishing LLC
14. Dehghani H, Fadaee MJ (2020) Probabilistic prediction of earthquake by bivariate distribution. Asian J Civ Eng 21:977–983
15. Kannan S (2014) Innovative mathematical model for earthquake prediction. Eng Fail Anal 41:89–95
16. K. Aggarwal, S. Mukhopadhyay, and A. K. Tangirala, “Statistical characterization and time-series modeling of seismic noise,” arXiv preprint arXiv:2009.01549, 2020
17. Woith H (2015) Radon earthquake precursor: a short review. Eur Phys J Spec Top 224(4):611–627
18. Wikelski M, Mueller U, Scocco P, Catorci A, Desinov LV, Belyaev MY, Keim D, Pohlsmeier W, Fechteler G, Martin Mai P (2020) Potential short-term earthquake forecasting by farm animal monitoring. Ethology 126(9):931–941
19. Uyeda S, Nagao T, Kamogawa M (2009) Short-term earthquake prediction: current status of seismo-electromagnetics. Tectonophysics 470(3–4):205–213
20. Li M, Parrot M (2013) Statistical analysis of an ionospheric parameter as a base for earthquake prediction. J Geophys Res Space Phys 118(6):3731–3739
21. Murwantara IM, Yugopuspito P, Hermawan R (2020) Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data. Telkomnika 18(3):1331–1342
22. Khalil U, Aslam B, Kazmi ZA, Maqsoom A, Qureshi MI, Azam S, Nawaz A (2021) Integrated support vector regressor and hybrid neural network techniques for earthquake prediction along Chaman fault, Baluchistan. Arab J Geosci 14(21):1–15
23. Lin J-W (2020) Researching significant earthquakes in Taiwan using two back-propagation neural network models. Nat Hazards 103(3):3563–3590
24. Kavianpour P, Kavianpour M, Ramezani A (2022) Deep multi-scale dilated convolution neural network with attention mechanism: a novel method for earthquake magnitude classification. In 2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp 1–6
25. Wang Y, Li X, Wang Z, Liu J (2022) Deep learning for magnitude prediction in earthquake early warning. Gondwana Res. https://doi.org/10.1016/j.gr.2022.06.009
26. Wang Q, Guo Y, Yu L, Li P (2017) Earthquake prediction based on spatio-temporal data mining: an LSTM network approach. IEEE Trans Emerg Top Comput 8(1):148–158
27. Huang J, Wang X, Zhao Y, Xin C, Xiang H (2018) Large earthquake magnitude prediction in Taiwan based on deep learning neural network. Neural Netw World 28(2):149–160
28. Jozinović D, Lomax A, Štajduhar I, Michелиni A (2020) Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. Geophys J Int 222(2):1379–1389
29. Li R, Lu X, Li S, Yang H, Qiu J, Zhang L (2020) Dlep: a deep learning model for earthquake prediction. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp 1–8, IEEE
30. Bhandarkar T, Satish N, Sridhar S, Sivakumar R, Ghosh S et al (2019) Earthquake trend prediction using long short-term memory RNN. Int J Electr Comput Eng 9(2):10304–1312
31. Doğan A, Demir E (2022) Structural recurrent neural network models for earthquake prediction. Neural Comput Appl 34:1–14
32. Al Banna H, Ghosh T, Taher KA, Kaiser MS, Mahmud M et al (2021) An earthquake prediction system for Bangladesh using deep long short-term memory architecture. Intelligent systems. Springer, Singapore, pp 465–476
33. Nicolis O, Plaza F, Salas R (2021) Prediction of intensity and location of seismic events using deep learning. Sapat Stat 42:100442
34. Mousavi SM, Beroza GC (2020) A machine-learning approach for earthquake magnitude estimation. Geophys Res Lett 47(1):20190e2019GL085976
35. Ghorvei M, Kavianpour M, Beheshiti MT, Ramezani A (2023) Spatial graph convolutional neural network via structured subdomain adaptation and domain adversarial learning for bearing fault diagnosis. Neurocomputing 517:44–61
36. Lu W, Li J, Li Y, Sun A, Wang J (2020) A CNN-LSTM-based model to forecast stock prices. Complexity 2020:1–10
37. Ghorvei M, Kavianpour M, Beheshiti MT, Ramezani A (2021) An unsupervised bearing fault diagnosis based on deep subdomain adaptation under noise and variable load condition. Meas Sci Technol 33(2):025901
38. Miao K, Hua Q, Shi H (2020) Short-term load forecasting based on CNN-BILSTM with Bayesian optimization and attention mechanism. In: International Conference on Parallel and Distributed Computing: Applications and Technologies, pp 116–128, Springer
39. Ghorvei M, Kavianpour M, Beheshiti MT, Ramezani A (2022) Synthetic to real framework based on convolutional multi-head attention and hybrid domain alignment. In: 2022 8th International Conference on Control, Instrumentation and Automation (ICCIA), pp 1–6
40. Chu Y, Yue X, Yu L, Sergei M, Wang Z (2020) Automatic image captioning based on resnet50 and LSTM with soft attention. Wiral Commum Mobile Comput 2020:1–7
41. Zhang B, Xiong D, Xie J, Su J (2020) Neural machine translation with GRU-gated attention model. IEEE Trans Neural Netw Learn Syst 31(11):4688–4698
42. Al Banna MH, Ghosh T, Al Nahian MJ, Taher KA, Kaiser MS, Mahmud M, Hossain MS, Andersson K (2021) Attention-based bi-directional long-short term memory network for earthquake prediction. IEEE Access 9:56589–56603
43. Zhang W, Li C, Peng G, Chen Y, Zhang Z (2018) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mech Syst Signal Process 100:439–453
44. Kavianpour M, Ghorvei M, Kavianpour P, Ramezani A, Beheshti MT (2022) An intelligent gearbox fault diagnosis under different operating conditions using adversarial domain adaptation. In: 2022 8th International Conference on Control, Instrumentation and Automation (ICCIA), pp 1–6, IEEE

45. Xi-wei X, Xi-yan W, Gui-hua Y, Xi-bin T, Kang L (2017) Seismo-geological signatures for identifying m≥ 7.0 earthquake risk areas and their preliminary application in mainland China. Seismol Egol 39(2):219

46. Al Banna MH, Taher KA, Kaiser MS, Mahmud M, Rahman MS, Hosen AS, Cho GH (2020) Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges. IEEE Access 8:192880–192923

47. Kavianpour M, Ramezani A, Beheshti MT (2022) A class alignment method based on graph convolution neural network for bearing fault diagnosis in presence of missing data and changing working conditions. Measurement 199:111536

48. Zhang J (2022) Exploration on coal mining-induced rockburst prediction using internet of things and deep neural network. J Supercomput 78:1–21

49. Zhu D, Zheng C, Wang W, Xie H, Cheng G, Wang FL, Wei M (2022) Rapid fault extraction from seismic images via deep learning. J Electron Imaging 31(5):051423

50. Kavianpour M, Ghorvei M, Ramezani A, Beheshti MT (2021) Intelligent fault diagnosis of rolling bearing based on deep transfer learning using time-frequency representation. In: 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS), pp 1–6, IEEE

51. van den Akker D (2022) Earthquake prediction: a MLP and SVM comparison

52. Berhich A, Belouadha F-Z, Kabbaj MI (2022) A location-dependent earthquake prediction using recurrent neural network algorithms. Soil Dyn Earthq Eng 161:107389

53. Malkoçoğlu ABV, Orman Z, Samli R (2022) Earthquake risk prediction with artificial intelligence methods. In: Analyzing Multidisciplinary Uses and Impact of Innovative Technologies, pp 156–171, IGI Global

54. Asim K, Martínez-Álvarez F, Basit A, Iqbal T (2017) Earthquake magnitude prediction in Hindukush region using machine learning techniques. Nat Hazards 85(1):471–486

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.