[Journal of Soft Computing and Computer Applications](https://jscca.uotechnology.edu.iq/jscca)

Volume 2 | Issue 1 Article 1014

2025

Predicting Earthquake Location Using Convolutional Neural Network-Attention Mechanism Approach

Mohammed A. Jaleel Shaneen
University of Technology – Iraq, Department of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq, cs.22.16@grad.uotechnology.edu.iq

Suhad M. Kadhem
University of Technology – Iraq, Department of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq, suhad.m.kadhem@uotechnology.edu.iq

Recommended Citation
Shaneen, Mohammed A. Jaleel and Kadhem, Suhad M. (2025) "Predicting Earthquake Location Using Convolutional Neural Network-Attention Mechanism Approach," *Journal of Soft Computing and Computer Applications*: Vol. 2: Iss. 1, Article 1014.
DOI: https://doi.org/10.70403/3008-1084.1014

JOURNAL OF SOFT COMPUTING AND COMPUTER APPLICATIONS 2025;2:1014

ORIGINAL STUDY

Predicting Earthquake Location Using Convolutional Neural Network-Attention Mechanism Approach

Mohammed A. Jaleel Shaneen *, Suhad M. Kadhem

University of Technology – Iraq, Department of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq

ABSTRACT

In seismically active areas, earthquake prediction is essential for minimizing potential damages and preserving lives. However, precise forecasts are complicated to achieve because of seismic events' complex and unpredictable nature. The current study presents an advanced prediction approach to address such issues, combining Convolutional Neural Networks (CNNs) and Attention Mechanism (AM). The primary goal is to improve the accuracy of the earthquake predictions and the generalizability across various mainland Chinese regions. AM layer emphasizes significant features for improving the prediction performance, whereas CNNs are utilized to extract spatial features of seismic data. The efficiency and effectiveness of the proposed approach were evaluated by comparing it with several well-known models. Results showed that the proposed approach performed consistently better than others in nine regions, with a reduced Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as well as higher R-squared (R^2) scores, especially in substantial seismic variability regions. Moreover, the proposed approach outperformed the conventional techniques in Region One, achieving an RMSE of 0.020, an MAE of 0.015, and an R^2 value of 0.960. In regions susceptible to seismic events, this all-encompassing approach presents a promising path for earthquake prediction, boosting readiness and risk management methods.

Keywords: Convolution neural networks, Earthquake prediction, Attention mechanism, Feature selection

1. Introduction

Earthquakes are a devastating natural occurrence that can seriously damage human life and infrastructure. These disasters usually occur suddenly and without warning, causing severe suffering to people and massive financial costs to communities [1]. Earthquakes

Received 6 September 2024; revised 24 January 2025; accepted 25 January 2025.
Available online 14 June 2025

* Corresponding author.
E-mail addresses: cs.22.16@grad.uotechnology.edu.iq (M. A. J. Shaneen), suhad.m.kadhem@uotechnology.edu.iq (S. M. Kadhem).

https://doi.org/10.70403/3008-1084.1014

could lead to direct effects. However, they might set off secondary calamities such as landslides, floods, and tsunamis [2–4], and they could also have detrimental impacts on the ecosystems, resulting in issues that include soil liquefaction [5] and surface fault ruptures [6]. Effective methods of prediction [7, 8] and a comprehensive understanding of their direct as well as indirect impacts [9] are vital, taking into consideration the significant damages as well as mortality that are caused by this kind of event.

Accurate earthquake prediction is vital for minimizing its devastating effects, bolstering preparedness, and improving crisis management efforts [10]. Effective forecasting techniques strive to estimate the magnitude and location of upcoming earthquakes within a designated timeframe [11], which can ultimately help save lives and lessen economic damages. Despite numerous proposed methods, accurate earthquake predictions remain challenging due to earthquakes' complex and stochastic nature [12].

The main objective of the current study's proposed approach is to improve the performance and accuracy of earthquake prediction in seismic active regions and generalize the approach to other areas by integrating an attention mechanism with a Convolutional Neural Network (CNN).

The rest of this study is organized as follows: Section 2 reviews the related works and examines earlier studies. In Section 3, the preliminary section presents the important concepts. The suggested approach is explained in Section 4. Section 5 introduces the experimental results, while Section 6 provides a comparative analysis of different approaches. Finally, the conclusion in Section 7 summarizes the study and offers recommendations for future research paths.

2. Related works

Numerous techniques are utilized in earthquake prediction research, and they can be divided into five major groups: mathematical modeling, Shallow Machine Learning (ML), precursor signal analysis, Deep Learning (DL), and hybrid methods. Table 1 provides an overview of such investigations.

2.1. Mathematical modeling

Mathematical modeling methods forecast earthquakes by applying mathematical and statistical methods. Kannan [13] presented a noteworthy research in the category, where he introduced a spatial correlation hypothesis for predicting earthquake locations in important seismic regions. Marisa et al. [14] have estimated Sumatra Island's earthquake using a Poisson Hidden Markov Model (PHMM), which leveraged the statistical approaches for predicting seismic event likelihood. Fadaee & Dehghani [15] employed a bivariate lognormal distribution to create a probabilistic prediction approach and optimize parameters through the use of the maximum likelihood approach. Their model was designed for earthquake prediction in Tehran with magnitudes that range from 6.6 to 6.8 over the next 10 to 15 years. Mathematical modeling methods are frequently dependent upon assumptions like linear and stationarity correlations, which might not always hold for earthquake data, leading to a potential restriction of prediction accuracy [16].

2.2. Precursor signal analysis

Precursor signal analysis focuses on examining some anomalies that could occur before the earthquakes. For example, Uyeda et al. [17] investigated electromagnetic signals as potential short-term earthquake forecast precursors, exploring the potential of the seism-electromagnetic signals for the predictions. Similarly, Li and Parrot [18] presented an analysis of the fluctuations in the ionospheric density, suggesting that the alterations in the ionospheric parameters could play the role of the seismic indicators. In addition, Wikelski et al. [19] studied farm animal behavior as a short-term earthquake prediction means, where they looked into how animal behavior shifts might indicate the upcoming seismic activities. However, the effectiveness of precursor-based methods is usually limited by these precursors' unpredictability and detectability.

2.3. Shallow machine learning methods

Data-driven methods utilizing statistical learning algorithms are called shallow Machine Learning methods. One of the notable research studies in this field was conducted by Asim et al. [20], who utilized the classifiers of ML in addition to historical seismic data for earthquake magnitude predictions. They were successful in earthquakes with 5.5 magnitude or more significant. In order to forecast earthquake magnitude, location, and depth in Indonesia, Murwantra et al. [21] utilized the Naive Bayesian (NB), Support Vector Machine (SVM), and multinomial logistic regression algorithms. They found out that the SVM's performance was superior to the others. Utilizing the historical data for the purpose of improving the accuracy of predictions, Lin [22] presented a probabilistic back-propagation NN model for earthquake predictions in Taiwan. Khalil et al. [23] created a hybrid NN incorporating SVMs with other approaches to predict earthquakes along the Chaman fault in Baluchistan. Even though shallow Machine Learning techniques can spot patterns in data, they frequently encounter some difficulties because the earthquake data is highly complicated, and they might require extensive feature engineering to increase their accuracy.

2.4. Deep learning methods

The capabilities of earthquake predictions have increased significantly due to the latest DL developments. Wang et al. [24] utilized Long Short-Term Memory (LSTM) networks in order to find spatial-temporal correlations between earthquakes in a variety of regions, which represents a noteworthy method example. Through the use of CNNs for learning spatial features, Huang et al. [25] utilized CNNs to analyze image data and predict the magnitudes of significant earthquakes in Taiwan. Bhandarkar et al. [26] employed the LSTM networks to investigate earthquake patterns by identifying the temporal dependencies in the seismic data. The deep Learning Model for Earthquake Prediction (DLEP) was presented by Li et al. [27] and utilizes the CNNs for the process of Feature Extraction (FE) for the integration of the explicit as well as the implicit earthquake features. Through the application of CNNs to the raw waveform data, Jozinovi et al. [28] focused on predicting the intensity of ground shaking throughout earthquakes in Italy. Banna et al. [29] bidirectional LSTM model with attention processes aims to increase prediction accuracy while predicting earthquakes in Bangladesh. Mao et al. [30] examined monthly maximum magnitude predictions in China's North-South Seismic Belt using many DL models to improve accuracy. Quinteros et al. [31] reported preliminary results from a CNN architecture model for utilizing High Rate – Global Navigation Satellite System (HR-GNSS) data to estimate earthquake magnitudes. With the help of the Standard Earthquake Dataset (STEAD), Manka et al. [32] investigated the use of DL models to predict earthquake magnitudes. They found that LSTM networks and Bidirectional LSTM architectures perform better when compared to other models in several metrics of performance.

2.5. Hybrid methods

Combining LSTMs and CNNs has demonstrated the potential in enhancing prediction performance by utilizing both temporal and spatial data. For instance, a Convolutional Neural Network – Bidirectional Long Short-Term Memory (CNN-BiLSTM) model was created by Beroza and Mousavi for estimating earthquake magnitudes using time-frequency features [33]. Nicolis et al. [34] suggested a CNN-LSTM method for the prediction of seismic events by using geographic images in Chile. Related studies along with their approaches and limitations are outlined in Table 1 below.

3. Preliminaries

3.1. Convolutional neural networks

Unlike other methods, DL methods concentrate on extracting high-level aspects [35]. One area of DL is CNNs. Accuracy and great effectiveness are attributes of CNN-based techniques [36]. CNN is distinguished by its excellent scalability and minimal complexity [37]. CNNs have proven helpful in several applications, including time series analysis, image processing, and facial recognition [38]. Pooling, convolutional, and Fully Connected Layers (FCLs) comprise a CNN's architecture. Every convolutional layer consists of a set of learnable filters that automatically extract the local characteristics from the input matrix. Those filters' convolution operations use the concepts of weight sharing and local connectivity, which reduce computational complexity and improve the performance of the model [39]. The pooling layer, which comes after convolutional layers, does downsampling, which lowers the feature map's dimensionality and helps avoid overfitting. The final layers of CNN are usually FCLs, which provide the final output by combining the information that the convolutional layers have extracted [40].

3.2. Attention mechanism

The human visual system inspires an attention mechanism (AM) technique to amplify the importance of essential information [41]. Just as human vision does not process an entire scene simultaneously but focuses on specific areas as needed, AM selects important information while disregarding fewer details. This approach is widely used in the image captioning field [42], machine translation [43], and earthquake prediction field [44].

4. The suggested approach's framework

4.1. Proposed convolutional neural network–attention mechanism

This section describes the architecture and essential elements of the suggested technique of earthquake prediction, which consists of four primary stages: model training, data preparation, evaluation, and testing, as shown in Fig. 1.

A new approach integrating CNNs and AMs methods is proposed to improve prediction accuracy and FE. The FE block, input block, prediction block, and attention block are the four main parts of the suggested approach, as seen in Fig. 2 and Table 2. The FE block uses CNNs for the extraction of the spatial features from input data. The AM then receives such features. The attention module gives the features varying weights at this point, emphasizing the most important elements and assisting the model in producing more precise and accurate predictions. The prediction block, which consists of a FCL and output

Table 1. The related works in earthquake predictions.

Research Category	Ref. no.	Approach/Model/Method	Limitations
Mathematical Method	[13]	Spatial Correlation	It relies on statistical assumptions that may not accurately reflect geological differences in various seismic regions.
	[14]	Poisson Hidden Markov	Assumes earthquakes follow a Poisson distribution, which may not capture the complex dynamics in regions like Sumatra.
	[15]	Probabilistic Model	Based on historical data, which may not account for changing seismic patterns over time, limiting prediction accuracy.
Precursor Method	[17]	Statistical Analysis Method	The unpredictability and occasional absence of precursors can hinder consistent short-term prediction.
	[18]	Statistical Analysis Method	Variability in ionospheric responses to earthquakes makes it challenging to establish reliable correlations.
	[19]	Statistical Analysis Method	Nonseismic factors could influence animal responses, reducing the method's specificity for earthquake prediction.
Shallow ML	[21]	Multinomial Logistic Regression (LR), SVM, NB	High reliance on data quality and extensive feature engineering; struggles with capturing complex seismic interactions.
	[22]	Backpropagation NN	Limited by insufficient data for accurate modeling, overfitting can occur with small datasets.
	[23]	Hybrid Neural Network (HNN)	Model performance heavily depends on the quality and variety of historical data, limiting generalizability.
DL Method	[24]	LSTM	High computational cost and extensive training time; sensitive to temporal data irregularities.
	[25]	CNN	Requires a large dataset of high-quality images, which may not be available in all regions.
	[26]	LSTM	It may overlook spatial aspects, as LSTM focuses primarily on temporal dependencies.
	[27]	CNN	Combining implicit features can introduce noise, reducing model clarity and accuracy.
	[28]	CNN	Limited generalizability due to training on region-specific data (Italy), which may not apply elsewhere.
	[29]	Attention Mechanism with Bi-directional LSTM	High sensitivity to data anomalies can skew attention mechanisms, impacting accuracy.
	[30]	predict maximum magnitude using Multiple DL models	Limited to some regions and may not generalize to other seismic zones.
	[31]	CNN for estimating magnitudes of earthquakes using HR-GNSS	Based on dataset configurations and the quality of data from different regions, performance may vary.
	[32]	predict earthquake magnitudes using LSTM and BiLSTM model	Performance is significantly influenced by dataset quality and may need further optimization for broader use
Hybrid	[33]	CNN and BiLSTM	Requires high computational resources to balance spatial and temporal aspects.
	[34]	CNN and LSTM	Model accuracy varies with geographic diversity, as it may not generalize well across different seismic zones.

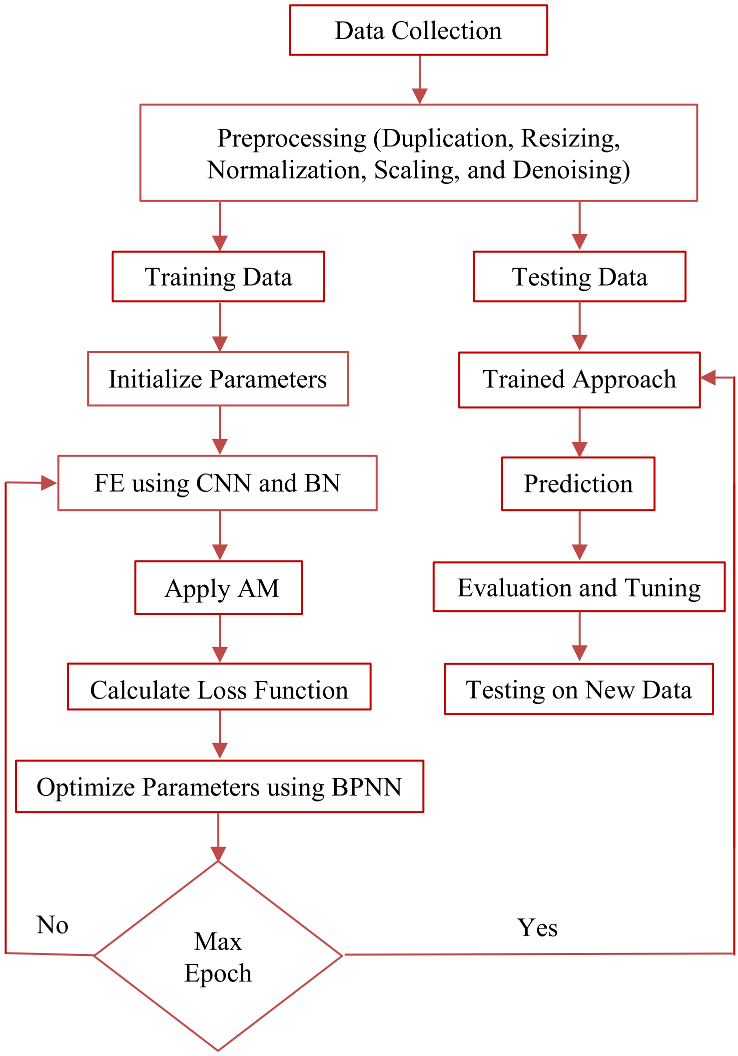

Fig. 1. The flowchart of the suggested approach.

layer, rates the final predictions. Trainable parameters, including loss functions, filter sizes, kernels, and the number of neurons, are present in every block. Reducing prediction errors requires optimizing such parameters. The structural elements of the suggested approach are made to reduce parameters and dimensions of the network. The Rectified Linear Unit (ReLU) activation function has been utilized in the FE block to improve convergence rates by addressing gradient vanishing or explosion problems. Batch Normalization (BN) is a regularization method after every convolutional layer. As seen in Table 2, BN improves the approach's performance by regularizing and lowering internal covariate shifts. Each one of the blocks in the suggested approach is described thoroughly in detail below:

4.1.1. Input block

This block processes earthquake data monthly using duplication, resizing, normalization, scaling, and denoising. Duplication increases dataset size for better model performance; resizing adjusts image dimensions for compatibility; normalization standardizes pixel

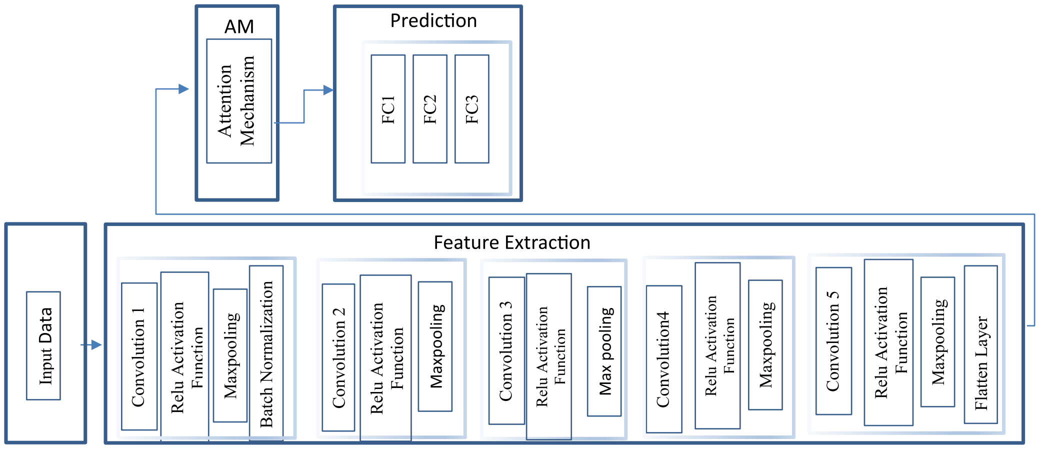

Fig. 2. The architecture of the suggested approach.

values to stabilize training; scaling converts pixel values to a uniform range for efficiency; denoising removes noise to improve image quality and analysis accuracy.

4.1.2. Feature extraction block

Five pooling layers, five convolutional layers, and one flattening layer make up the eleven layers of the 1D CNN used in the block. The first convolutional layer employs a higher filter size for effectively attenuating high-frequency signals than the subsequent layers. Data representation is improved by extracting more complex features that are made possible by integrating numerous pooling and convolutional layers. One of the convolution layers is followed by max-pooling layers to enhance the network's training and generalization capabilities. Through adapting features to the intended distribution, BN normalizes and speeds up training layer by layer. The BN approach is computed as follows: [40]:

$$\mu = \frac{1}{N_{batch}} \sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma^2 = \frac{1}{N_{batch}} \sum_{i=1}^{N} (x_i - \mu)^2 \tag{2}$$

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{3}$$

$$y_i = \gamma \hat{x}_i + \beta \tag{4}$$

In such case, y_i and x_i are the output and input values of i^{th} observation within a mini-batch, whereas Nbatch denotes the mini-batch size. The symbol μ represents the mini-batch sample's mean, its standard deviation by γ, and a small constant called ϵ added for numerical stability—additionally, β functions as a bias parameter and γ as a scaling parameter. The padding type is always used to avoid feature loss throughout convolution operations. The convolution layer's multidimensional output must be transformed into 1D data before being processed and sent to the flattened layer.

Table 2. Suggested approach structures.

Block	Layer	No. of Filters × size × Stride
FE	Convolution one	16 × 5 × 1
	Max-pooling	16/2/1
	Convolution two	32 × 3 × 1
	Max-pooling	32/2/1
	Convolution three	64 × 3 × 1
	Max-pooling	64/2/1
	Convolution four	128 × 3 × 1
	Max-pooling	128/2/1
	Convolution five	256 × 3 × 1
	Max-pooling	256/2/1
Attention	Attention Mechanism	–
Prediction	Fully Connected 1	32
	Fully Connected 2	10
	Fully Connected 3	1

4.1.3. Attention block

At the end of sequence learning phase, an attention layer is incorporated into the attention block. This layer emphasizes the most influential factors affecting the prediction results, enhancing accuracy. An attention mechanism functions by assigning weights to different components and higher weights to more relevant information, optimizing traditional models' performance. The attention function maps the query with a sequence of value-key pairs. The process of calculating attention involves three steps, as depicted in Fig. 3. In the first step, the correlation or similarity between each key and query is computed as follows [42, 43]:

$$s_t = tanh\left(W_h h_t + b_h\right) \tag{5}$$

where s_t represents the attention score value, and b_h, W_h, represent the bias and weight values regarding AM. h_t denotes the input vector. The score value that was obtained from the first stage is normalized in the second phase, and the attention score is converted using the softmax function, as given in the following formula [43]:

$$a_t = \frac{exp\,(s_t)}{\sum_t exp\,(s_t)} \tag{6}$$

The final attention value of the weight coefficient is obtained by weighted summation, as shown below [43]:

$$s = \sum_t a_t h_t \tag{7}$$

The attention mechanism technique is typically utilized after Recurrent Neural Network (RNN) and CNN to focus on an important feature that considerably affects the output variables and increases and enhances the model performance.

4.1.4. Prediction block

This block consists of two FCLs and one FCL (the output layer). The FCLs perform a sequence of nonlinear transformations to the features of the attention layer processes. The final output layer then generates the prediction results.

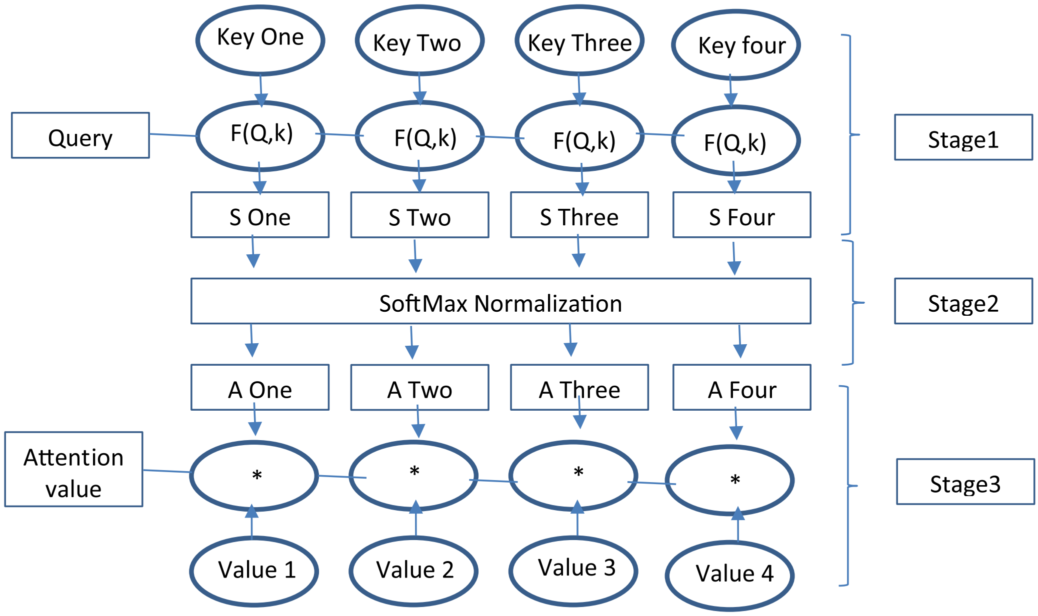

Fig. 3. The steps of determining AM [41].

5. Experimental results

5.1. Division of study area

The study focuses on the Chinese mainland area, which is positioned in the southeastern region of the Eurasian tectonic plate. This region, connected the Indian and Philippine plates and the Mongolia-Siberia subplate, is one of the most seismically active regions in the world. Mainland China has experienced numerous significant earthquakes, including the Xingtai earthquake in 1966, the Daguan earthquake in 1975, the Jilin earthquake in 2002, the Wenchuan earthquake in 2008, the Lushan earthquake in 2013, and the Gorkha earthquake in 2015 with Moment Magnitude (MW) 7.1, 7.2, 8.0, 7.0, and 7.8, respectively. Since 1949, catastrophic earthquakes have occurred in this region, resulting in over 275,000 deaths, which represents 55% of total deaths from different natural disasters in area of mainland China [45]. Given the high frequency of such events, reliable earthquake predictions are crucial to mitigating damage and casualties. The Chinese mainland is split into smaller regions to analyze and predict earthquake locations more accurately. Due to the complex challenge of limited and inadequate data, the study area was segmented into nine smaller regions. These regions are defined by a range of degrees from 75 to 119 and a range of longitudes of degrees from 23 to 45.

5.2. The dataset and data preprocessing

5.2.1. Dataset

A comprehensive earthquake dataset is essential to describe regional seismicity effectively. The United States Geological Survey (USGS) [46], as well as the National Seismological Center (NSC) [47], provided the data used in this investigation as a unified dataset. The dataset used in the presented work includes both temporal and spatial data. A total of 34,326 images that show the geographic distribution regarding earthquakes make up the spatial data, which allows the model to use CNNs to identify spatial patterns.

Table 3. Summarizes the approach dataset.

Aspect	Details
Data Sources	NSC and USGS
Date Range	January 15, 1966 - May 22, 2023
Total Images	34,326
Total Records	11,442
Magnitude Threshold	≥ 3.5
Included Information	Latitude, Longitude, Date/Time, Magnitude, Depth, Station Number
Monthly Data Samples	665

The dataset includes 11,442 records of earthquakes with magnitudes of 3.5 or greater, spanning from January 15, 1966, to May 22, 2023. Each record provides details such as magnitude, longitude range, latitude range, date, depth, time of occurrence, and station number. The present research is primarily focused on spatial data to enhance the model's ability to identify regions that are prone to earthquakes. A summary of key aspects of the data-set is illustrated in Table 3.

5.2.2. Data preprocessing

A number of the preprocessing steps have been undertaken for data preparation for modeling. At first, duplicate records were identified and removed to maintain the accuracy and reliability of the dataset. Data was divided into two subsets: 20% for the testing and 80% for the training model. The resizing, duplication, scaling, normalization, and denoising were employed for the preprocessing, as explained in Section 4.1.1.

5.3. Implementation details

In order to create a well-trained model, it is essential to select the correct hyperparameters. The test-and-error approach is utilized to modify the hyperparameters. The Stochastic Gradient Descent (SGD) technique [48], Root Mean Square Propagation (RMSProp) for DL [49], and Adam stochastic optimization approach [50] were among the methods of optimization that have been compared. Adam was chosen as the optimization algorithm and has been shown to improve the model's accuracy. The Mean Square Error (MSE) metric is used as a back-propagation loss function to update bias and weight values. The learning rate value starts at 0.001 and gradually drops to 0.0001 by the end of the epoch for maintaining a steady learning speed. Concerning the training phase, the model is trained for 150 epochs with a 32-batch size. Each experiment is performed multiple times to reduce random variations. Even though the nine training regions employ the configurations and settings, the training for each region should be divided since geological features vary by region. For each of the nine regions, the prediction results on the test data are acquired independently.

5.4. The evaluation metrics

Three evaluation metrics, i.e., RMSE, MAE, and R-squared (R^2), are used to assess the suggested approach's performance. Below are the formulas for such metrics [51]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - y_i\right)^2} \tag{10}$$

In the equations, n denotes the number of predicted data points; y_i denotes actual values; $\widehat{y}$ represents predicted values and $\bar{y}$ denotes the mean regarding actual values. It is crucial to maximize R^2 while minimizing RMSE and MAE values in order to improve model performance. A complete linear correlation is shown by a value of 1, the R^2 metric's estimate of the strength of linear relation between observed and predicted values. Approach performance is measured by MAE and RMSE, with zero denoting the ideal result and lower numbers indicating higher performance. Evaluation criteria, including precision, accuracy, F1 score, and recall are added along with such metrics for a more thorough examination of the prediction models. The percentage of accurate predictions amongst all of the predictions is referred to as accuracy. The percentage regarding true positive predictions compared to all positive predictions is called the precision. The percentage regarding true positive predictions among all true positive situations is called the recall. Lastly, the F1 score balances recall and precision by representing the harmonic mean of such two criteria. These matrices are computed as follows [51]:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \tag{11}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{12}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{13}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

5.5. Comparison methods

Random Forest (RF), SVM, Decision Trees (DT) [20], Multi-Layer Perceptron (MLP), CNN [25], LSTM [26], and a CNN-Bidirectional LSTM (BiLSTM) combination are among the models commonly used in earthquake prediction that are included in the comparison. A comparison between the new model and existing ones took place using same data. The comparison approaches are divided into distinct classes.

5.5.1. Shallow machine learning models

RF, SVM, MLP, and DT make up this group. Those methods compare the suggested approach's FE capabilities against traditional ML approaches.

- **SVM**: This algorithm is grounded in statistical learning theory and involves hyperparameters that include the Radial Basis Function (RBF) kernel, a regularization factor (C) of 1, epsilon (ϵ) of 0.01, and gamma (γ) of 0.1.

- **RF**: This ensemble learning method utilizes averaging to improve prediction accuracy and prevent overfitting. In such a study, the maximum depth regarding the trees is configured to 9, and the trees' number in the forest is 100.
- **DT**: This supervised learning method employs a tree structure to create models of regression or classification based upon decision rules derived from the data features. The maximum depth of the tree is configured to 10.
- **MPL:** The input, hidden, and output layers of the MLP are frequently used in earthquake prediction studies. Two hidden layers, which have 15 neurons each, are utilized in the presented work. The model has a sigmoid activation function, 150 epochs, and a learning rate value of 0.01 [52].

5.5.2. Deep learning encompasses techniques

This includes LSTM, CNN, and combining CNN-LSTM. The CNN technique is associated with the proposed approach by the FE component, while the LSTM technique is akin to the sequence learning component of the proposed approach, with BiLSTM being used instead of LSTM. The hybrid CNN-BiLSTM approach integrates both the FE and sequence learning components.

6. Comparative analysis

6.1. Earthquake frequency analysis

Most previous studies have concentrated on predicting earthquake time, location, and size using seismic indicators like depth, magnitude, and geographic location as inputs. To the authors' knowledge, no research has yet been done on earthquake frequency prediction. Earthquake frequency might be a critical element for forecasting seismic activity and providing a more thorough understanding of a region's seismic behavior. To precisely forecast the number of anticipated earthquakes monthly, this case study uses the suggested approach under the same assumptions and conditions. The suggested approach's efficiency, effectiveness, and generalizability are evaluated compared to LSTM, CNN, MLP, RF, SVM, and DT. Table 4 summarizes the findings for predicting the frequency of earthquakes in nine distinct regions. Results have demonstrated that, across all regions, the suggested approach consistently produces superior prediction performance, as evidenced by the greatest R^2 scores and the minimal RMSE and MAE values. This shows that the suggested approach is quite good at predicting the frequency of earthquakes and either beats or is on par with the other models. For example, the suggested approach displays outstanding performance in Region 1, whereas other models show significant prediction error rates, with RMSE of 0.20, R^2 score of 0.960, and MAE of 0.015.

The suggested approach performs better than all comparison approaches in Region 7, which is regarded as one of the most difficult R^2 metrics. In particular, it attains R^2 values that surpass the comparison approaches' maximum and minimum R^2 values, respectively. This outstanding performance demonstrates the efficacy of a suggested approach, especially amid notable seismic abnormalities. The approach's three main components—using CNN to acquire spatial attributes, using AM to highlight the significance of various hidden states, and the concentrated consideration of the most crucial states—are responsible for the improved performance. Such characteristics make the suggested approach more successful than the ML and DL methods. Across all regions, however, the SVM model performs the worst, with the lowest R^2 scores and the greatest MAE and RMSE values. Long-term dependencies in time series data are difficult for SVM to capture, even if it can handle non-linear problems. The mean evaluation indicators for each implemented

Table 4. Evaluation error indexes comparisons for eight approach.

Regions	Metrics	SVM	DT	MLP	RF	CNNs	LSTM	CNN-Bi-LSTM	Proposed approach
Region One	RMSE	0.10	0.09	0.11	0.09	0.10	0.09	0.09	0.020
	MAE	0.06	0.04	0.05	0.04	0.05	0.05	0.04	0.015
	R2	-0.01	0.16	0.05	0.15	0.14	0.17	0.25	0.960
Region Two	RMSE	0.13	0.09	0.15	0.13	0.08	0.07	0.07	0.060
	MAE	0.09	0.11	0.09	0.06	0.05	0.04	0.04	0.029
	R2	0.60	0.66	0.43	0.58	0.83	0.86	0.88	0.920
Region Three	RMSE	0.28	0.26	0.18	0.20	0.20	0.19	0.19	0.065
	MAE	0.15	0.15	0.13	0.11	0.11	0.12	0.12	0.028
	R2	-0.23	-0.03	0.43	0.37	0.41	0.45	0.47	0.940
Region Four	RMSE	0.15	0.15	0.14	0.14	0.13	0.13	0.10	0.090
	MAE	0.09	0.08	0.07	0.08	0.07	0.07	0.07	0.060
	R2	0.01	0.14	0.26	0.21	0.43	0.50	0.61	0.705
Region Five	RMSE	0.20	0.12	0.10	0.11	0.09	0.08	0.09	0.055
	MAE	0.13	0.08	0.07	0.08	0.07	0.07	0.07	0.054
	R2	-0.05	0.58	0.75	0.68	0.70	0.74	0.72	0.910
Region Six	RMSE	0.23	0.20	0.17	0.19	0.20	0.20	0.18	0.085
	MAE	0.14	0.13	0.01	0.13	0.14	0.13	0.13	0.058
	R2	-0.35	0.02	0.21	0.06	0.05	0.10	0.15	0.810
Region Seven	RMSE	0.16	0.16	0.11	0.10	0.09	0.09	0.09	0.089
	MAE	0.10	0.14	0.05	0.04	0.03	0.04	0.04	0.025
	R2	0.13	0.11	0.22	0.37	0.43	0.46	0.48	0.615
Region Eight	RMSE	0.14	0.14	0.13	0.14	0.13	0.13	0.12	0.090
	MAE	0.08	0.07	0.07	0.08	0.06	0.06	0.06	0.055
	R2	0.08	0.18	0.26	0.18	0.39	0.39	0.43	0.675
Region Nine	RMSE	0.12	0.12	0.12	0.12	0.12	0.12	0.10	0.085
	MAE	0.08	0.09	0.10	0.09	0.09	0.09	0.07	0.055
	R2	0.61	0.51	0.48	0.52	0.56	0.57	0.63	0.815

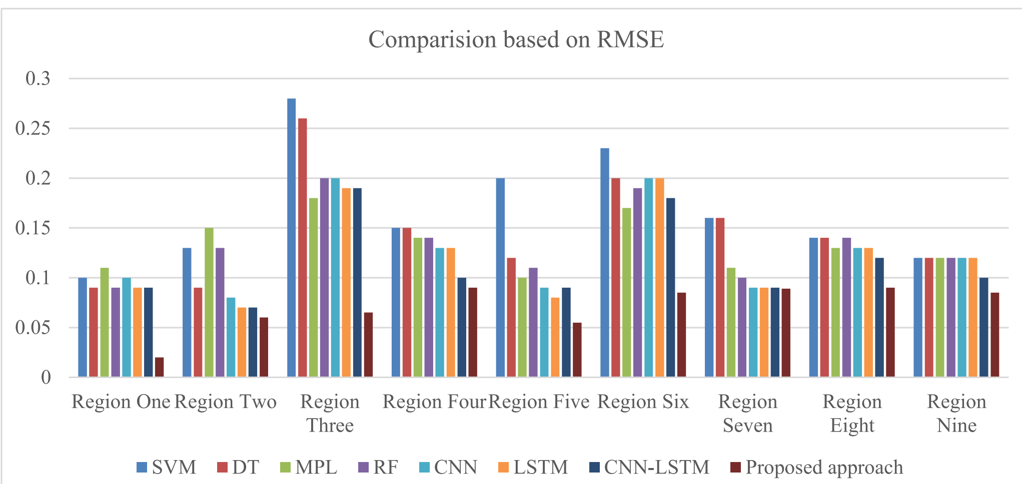

Fig. 4. Comparison of RMSE evaluation metric between the suggested approach and the compared approaches.

approach are compared to further highlight the benefits of the suggested approach. As depicted in Fig. 4, there is a notable disparity in performance among various techniques. DL methods generally exhibit lower average errors and greater predictive potential than shallow ML approaches. This is due to the relative simplicity of shallow learning structures,

Table 5. Evaluation metrics results.

Region	Accuracy	Precision	Recall	F1 Score
Region 1	86%	83%	78%	80%
Region 2	83%	81%	76%	78%
Region 3	76%	73%	68%	70%
Region 4	78%	75%	70%	72%
Region 5	84%	82%	76%	78%
Region 6	80%	77%	72%	74%
Region 7	79%	76%	71%	73%
Region 8	78%	75%	70%	72%
Region 9	81%	78%	73%	75%

which often fail to model the complex non-linear relationships inherent in earthquake data. DL approaches, with their multiple hidden layers, are better suited for capturing these complexities and learning features at various levels of abstraction.

Among DL models, LSTM performs better than CNN in time series prediction, as evidenced by its lower RMSE (0.07 vs. 0.08) and MAE (0.04 vs. 0.05) and higher R^2 score (0.86 vs. 0.83). This performance advantage arises because LSTM is more adept at handling long-term dependencies than CNN. The hybrid CNN-BiLSTM model, combining the strengths of both LSTM and CNN, shows the lowest error values and the maximum R^2 score amongst the DL models. However, the proposed approach, for example, in region 2 (RMSE $=$ 0.06, MAE $=$ 0.029, $R^2 = 0.920$), outperforms the CNN-BiLSTM model in predicting frequency. Overall, the proposed approach improves RMSE, MAE, and superior R^2 metrics compared with CNN model, highlighting the effectiveness of the integrated AM techniques. This demonstrates that incorporating AM into the CNN model significantly enhances the prediction accuracy by capturing the complex relationships within the data.

6.2. Earthquake prediction performance

To ensure a thorough performance evaluation of the prediction model, additional metrics are introduced, such as Precision, Accuracy, F1, and Recall for each region, as illustrated in Table 5 and Fig. 5.

Regions with low RMSE and high R^2, such as Regions 1 and 2, achieve the highest accuracy scores (83%–86%), indicating the approach's adequate predictive strength in these areas. In contrast, accuracy in regions 3 and 4 is slightly lower (76%–78%), aligning with higher prediction errors and lower R^2 values, indicating more difficulty in accurate predictions. Precision is notably high (75%–83%) in regions with strong R^2 and low RMSE (e.g., Regions 1, 2, and 5), demonstrating the approach's capability for avoiding false positives in areas with higher prediction quality. Regions with higher errors, like region 3, have somewhat lower precision (73%), reflecting an increase in false positives. Recall values, ranging from 68%–78%, illustrate the approach's success in identifying true positives. Regions with higher R^2 values (e.g., Regions 1 and 2) demonstrate stronger recall, whereas regions facing more significant prediction challenges (e.g., Region 3) exhibit reduced recall. F1 Scores, which indicate the balance between precision and recall, range from 70% to 80%. Regions 1 and 2 achieve higher scores ($\sim$80%), confirming their balanced performance, while Regions 3 and 7 show moderate scores (70%–73%), indicating potential for improvement. These metrics align logically with the error indicators (RMSE, MAE, and R^2) in Table 4. Areas with lower prediction errors (low RMSE and high R^2) correspond to higher values across these metrics, while more difficult-to-predict regions have slightly

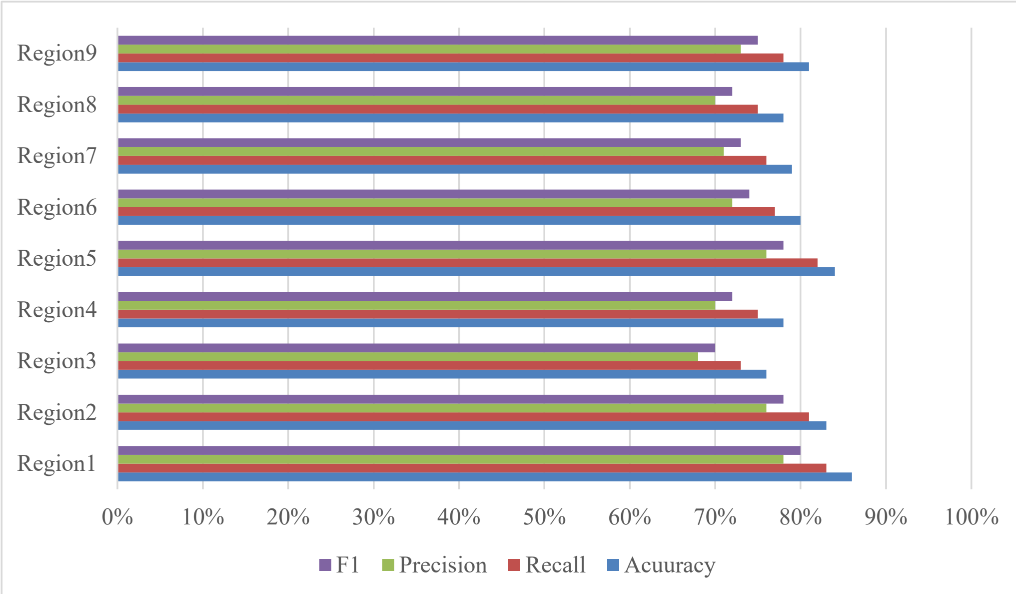

Fig. 5. Evaluation Metrics.

lower scores. Overall, these metrics have provided a realistic assessment of the approach's effectiveness, highlighting strengths and opportunities for enhancing prediction accuracy.

7. Conclusion

The present study introduces an earthquake prediction approach that utilizes a CNN-AM, designed to enhance earthquake location forecasting and magnitude accuracy. Considering the complex and nonlinear nature of the earthquake data, the present study provides a significant advancement in the predictive approach. The CNN-AM approach uses a detailed historical dataset spanning five decades, including monthly earthquake frequency records and the maximum magnitudes. Through the preprocessing of this data for refining input features and using CNNs to extract the spatial characteristics, the approach can effectively identify the relevant patterns. Attention Mechanism integration results in further boosting the performance through the prioritization of the features holding high predictive value, which makes sure that the approach is not only accurate but also generalizes well across a variety of regions. Two case study results demonstrate the CNN-AM approach's superiority over conventional shallow ML and DL models. When predicting earthquake frequency, the suggested approach often beats models like LSTM, CNN, MLP, RF, SVM, and DT. CNN-AM approach demonstrated its efficiency and robustness in earthquake prediction by achieving the MAE and RMSE as well as the greatest R^2 scores. CNN-AM approach has excelled particularly in challenging regions and scenarios that have characterized significant seismic anomalies. Its superior performance in areas with complex seismic patterns reflects its capability for navigating earthquake data intricacies more effectively than the other approach. This can result from the advanced FE capabilities and AM of the approach, collectively enhancing the approach's proficiency in identifying and utilizing the critical patterns within data. Future works could implement LSTM or GRU techniques to sequence data prediction. Employing Transformer Attention could also be utilized to recognize key seismic data elements, as well as integrate various data sources, including seismic and satellite data and others, for advanced model training efficiency.

Acknowledgment

The researchers express their sincere thanks and appreciation to all those who contributed to completing this research.

Conflict of interest

The authors declare no conflicts of interest.

Authors contribution

Mohammed A. Jaleel Shaneen: study conception, design, data collection, analysis, and interpretation of results

Suhad M. Kadhem: Supervision of the research, Draft manuscript preparation, reviewed the results and approved the final version of this manuscript.

Data Availability

The data was collected from the USGS available at https://www.usgs.gov/ and from NSC at: https://seismo.gov.in/data-portal.

References

1. A. S. N. Alarifi, N. S. N. Alarifi, and S. Al-Humidan, "Earthquakes magnitude prediction using artificial neural network in northern red sea area," *Journal of King Saud University - Science*, vol. 24, no. 4, pp. 301–313, Oct. 2012, doi: 10.1016/j.jksus.2011.05.002.
2. P. Cui *et al.*, "The Wenchuan earthquake (May 12, 2008), Sichuan Province, China, and resulting geohazards," *Natural Hazards*, vol. 56, no. 1, pp. 19–36, April 2009, doi: 10.1007/s11069-009-9392-1.
3. N. Jain, D. Virmani, and A. Abraham, "Proficient 3-class classification model for confident overlap value based fuzzified aquatic information extracted tsunami prediction," *Intelligent Decision Technologies*, vol. 13, no. 3, pp. 295–303, Sep. 2019, doi: 10.3233/IDT-180003.
4. P. Cui, L.-Z. Xiang, and Q. Zou, "Risk assessment of highways affected by debris flows in Wenchuan earthquake area," *Journal of Mountain Science*, vol. 10, no. 2, pp. 173–189, April 2013, doi: 10.1007/s11629-013-2575-y.
5. R. Verdugo and J. González, "Liquefaction-induced ground damages during the 2010 Chile earthquake," *Soil Dynamics and Earthquake Engineering*, vol. 79, pp. 280–295, Dec. 2015, doi: https://doi.org/10.1016/j.soildyn.2015.04.016.
6. J. D. Bray, "Developing mitigation measures for the hazards associated with earthquake surface fault rupture," Presented at Seismic Fault-Induced Failures—Possible Remedies for Damage to Urban Facilities Workshop, Japan, 2001.
7. R. Bilham, "The seismic future of cities," *Bulletin of Earthquake Engineering*, vol. 7, no. 4, pp. 839–887, Sep. 2009, doi: 10.1007/s10518-009-9147-0.
8. N. N. Ambraseys and C. P. Melville, *A History of Persian Earthquakes*. Cambridge, UK: Cambridge University Press, 2005.
9. J. Jia, "Earthquake damages," in *Modern Earthquake Engineering*. Heidelberg, German: Springer, Oct. 2016, pp. 413–431, Springer, Oct. 2016, doi: 10.1007/978-3-642-31854-2_13.
10. G. A. Sobolev, "Methodology, results, and problems of forecasting earthquakes," *Herald of the Russian Academy of Sciences*, vol. 85, no. 2, pp. 107–111, May 2015, doi: 10.1134/S1019331615020069.

11. R. Kail, E. Burnaev, and A. Zaytsev, "Recurrent convolutional neural networks help to predict location of earthquakes," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, Sep. 2021, Art. no. 8019005, doi: 10.1109/LGRS.2021.3107998.
12. G. Otari and R. V. Kulkarni, "A review of application of data mining in earthquake prediction," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 2, pp. 3570–3574, 2012.
13. S. Kannan, "Innovative mathematical model for earthquake prediction," *Engineering Failure Analysis*, vol. 41, pp. 890–895, June 2014, doi: 10.1016/j.engfailanal.2013.10.016.
14. Marisa, U. A. Sembiring and H. Margaretha, "Earthquake probability prediction in Sumatra Island using Poisson hidden Markov model (HMM)," in *Proc. AIP Conf*. vol. 2192, Art. no. 090006, Dec. 2019, doi: 10.1063/1.5139176.
15. H. Dehghani and M. J. Fadaee, "Probabilistic prediction of earthquake by bivariate distribution," *Asian Journal of Civil Engineering*, vol. 21, no. 6, pp. 977–983, May 2020, doi: 10.1007/s42107-020-00254-y.
16. K. Aggarwal, S. Mukhopadhyay, and A. K. Tangirala, "Statistical characterization and time-series modeling of seismic noise," 2020, *arXiv:*2009.01549.
17. S. Uyeda, T. Nagao, and M. Kamogawa, "Short-term earthquake prediction: Current status of seismo-electromagnetics," *Tectonophysics*, vol. 470, no. 3-4, pp. 205–213, May 2009, doi: 10.1016/j.tecto.2008.07.019.
18. M. Li and M. Parrot, "Statistical analysis of an ionospheric parameter as a base for earthquake prediction," *Journal of Geophysical Research: Space Physics*, vol. 118, no. 6, pp. 3731–3739, May 2013, doi: 10.1002/jgra.50313.
19. M. Wikelski *et al.*, "Potential short-term earthquake forecasting by farm animal monitoring," *Ethology*, vol. 126, no. 9, pp. 931–941, July 2020, doi: 10.1111/eth.13078.
20. K. M. Asim, F. Martínez-Álvarez, A. Basit, and T. Iqbal, "Earthquake magnitude prediction in Hindukush region using machine learning techniques," *Natural Hazards*, vol. 85, no. 1, 471–486, Sep. 2017, doi: 10.1007/s11069-016-2579-3.
21. I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data," Telkomnika Telecommunication, Computing, Electronics and Control, vol. 18, no. 3, pp. 1331–1342, June 2020, doi: 10.12928/TELKOMNIKA.v18i3.14756.
22. J.-W. Lin, "Researching significant earthquakes in Taiwan using two back-propagation neural network models," *Natural Hazards*, vol. 103, no. 3, pp. 3563–3590, July 2020, doi: 10.1007/s11069-020-04144-z.
23. U. Khalil *et al.*, "Integrated support vector regressor and hybrid neural network techniques for earthquake prediction along Chaman fault, Baluchistan," *Arabian Journal of Geosciences*, vol. 14, no. 21, Oct. 2021, Art. no. 2192, doi: 10.1007/s12517-021-08564-4.
24. Q. Wang, Y. Guo, L. Yu, and P. Li, "Earthquake prediction based on spatio-temporal data mining: An LSTM network approach," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, no. 1, pp. 148–158, April 2017, doi: 10.1109/TETC.2017.2699169.
25. J. P. Huang, X. A. Wang, Y. Zhao, C. Xin, and H. Xiang, "Large earthquake magnitude prediction in Taiwan based on deep learning neural network," *Neural Network World*, vol. 28, no. 2, pp. 149–160, 2018, doi: 10.14311/NNW.2018.28.009.
26. T. Bhandarkar, V. K., N. Satish, S. Sridhar, R. Sivakumar, and S. Ghosh, "Earthquake trend prediction using long short-term memory RNN," International Journal of Electrical and Computer Engineering, vol. 9, no. 2, pp. 1304–1312, April 2019, doi: 10.11591/ijece.v9i2.pp1304-1312.
27. R. Li, X. Lu, S. Li, H. Yang, J. Qiu, and L. Zhang, "DLEP: A deep learning model for earthquake prediction," in *Proc. IEEE 2020 Int. Joint Conf. on Neural Networks (IJCNN)*, Glasgow, UK, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207621.
28. D. Jozinovic, A. Lomax, I. Štajduhar, and A. Michelini, "Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network," *Geophysical Journal International*, vol. 222, no. 2, pp. 1379–1389, May 2020, doi: 10.1093/gji/ggaa233.
29. H. Banna, T. Ghosh, K. A. Taher, M. S. Kaiser, M. Mahmud, "An earthquake prediction system for Bangladesh using deep long short-term memory architecture," in *Proc. Int. Conf. on Machine Learning, Internet of Things and Big Data (ICMIB2020)*, in Intelligent Systems, Lecture Notes in Networks and Systems, S. K. Udgata, S. Sethi, and S. N. Srirama, Eds. vol. 185, pp. 465–476, doi: 10.1007/978-981-33-6081-5_41.
30. N. Mao, K. Sun, and J. Zhang, "Monthly maximum magnitude prediction in the north–south seismic belt of China based on deep learning," *Applied Sciences*, vol. 14, no. 19, Oct. 2024, Art. no. 9001, doi: 10.3390/app14199001.
31. C. Quinteros-Cartaya, J. Köhler, W. Li, J. Faber and N. Srivastava, "Exploring a CNN model for earthquake magnitude estimation using HR-GNSS data," *Journal of South American Earth Sciences*, vol. 136, April 2024, Art. no. 104815, doi: 10.1016/j.jsames.2024.104815.

32. M. Vasti and A. Dev, "Artificial intelligence in seismology-deep learning models for earthquake magnitude prediction," *Nanotechnology Perceptions,* vol. 20, no. S9, pp. 899–917, 2024.
33. S. M. Mousavi and G. C. Beroza, "A machine-learning approach for earthquake magnitude estimation," *Geophysical Research Letters*, vol. 47, no. 1, Dec. 2019, Art. no. e2019GL085976, doi: 10.1029/2019GL085976.
34. O. Nicolis, F. Plaza, and R. Salas, "Prediction of intensity and location of seismic events using deep learning," *Spatial Statistics*, vol. 42, April 2021, Art. no. 100442, doi: 10.1016/j.spasta.2020.100442.
35. C. Quinteros-Cartaya, J. Köhler, W. Li, J. Faber, and N. Srivastava, "Exploring a CNN model for earthquake magnitude estimation using HR-GNSS data," *Journal of South American Earth Sciences*, vol. 136, April 2024, Art. no. 104815, doi: 10.1016/j.jsames.2024.104815.
36. M. Elbes, S. AlZu'bi, and T. Kanan, "Deep learning-based earthquake prediction technique using seismic data," in *Proc. 2023 IEEE Int. Conf. on Multimedia Computing, Networking and Applications (MCNA)*, Valencia, Spain, pp. 103–108, doi: 10.1109/MCNA59361.2023.10185869.
37. N. Van Quan, H.-J. Yang, K. Kim, and A.-R. Oh, "Real-time earthquake detection using convolutional neural network and social data," in *Proc. 2017 IEEE 3rd Int. Conf. on Multimedia Big Data (BigMM)*, Laguna Hills, CA, USA, pp. 154–157, doi: 10.1109/BigMM.2017.58.
38. Y. Li, Y. He, and M. Zhang, "Prediction of Chinese energy structure based on convolutional neural network-long short-term memory (CNN-LSTM)," *Energy Science & Engineering*, vol. 8, no. 8, pp. 2680–2689, April 2020, doi: 10.1002/ese3.698.
39. W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, no. 1, Nov. 2020, Art. no. 6622927, doi: 10.1155/2020/6622927.
40. R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 2, pp. 415–433, Feb. 2020, doi: 10.1007/s00477-020-01776-2.
41. K. Miao, Q. Hua, and H. Shi, "Short-term load forecasting based on CNN-BiLSTM with Bayesian optimization and attention mechanism," in *Proc. Int. Conf. on Parallel and Distributed Computing, Applications and Technologies*, in Lecture Notes in Computer Science (LNTCS), Y. Zhang, Y. Xu, and H. Tian, Eds. vol. 12606, pp. 116–128, doi: 10.1007/978-3-030-69244-5_10.
42. Y. Chu, X. Yue, L. Yu, M. Sergei, and Z. Wang, "Automatic image captioning based on ResNet50 and LSTM with soft attention," *Wireless Communications and Mobile Computing*, vol. 2020, no. 1, Oct. 2022, Art. no. 8909458, doi: 10.1155/2020/8909458.
43. B. Zhang, D. Xiong, J. Xie, and J. Su, "Neural machine translation with GRU-gated attention model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4688–4698, Jan. 2020, doi: 10.1109/TNNLS.2019.2957276.
44. H. Al Banna *et al.*, "Attention-based bidirectional long-short term memory network for earthquake prediction," *IEEE Access*, vol. 9, pp. 56589-56603, April 2021, doi: 10.1109/ACCESS.2021.3071400.
45. X. Xu *et al.*, "Seismo-geological signatures for identifying M 7.0 earthquake risk areas and their preliminary application in mainland China," *Seismology and Geology*, vol. 39, no. 2, pp. 219–275, April 2017, doi: 10.3969/j.issn.0253-4967.2017.02.001.
46. United States Geological Survey, Jan. 1966 – May 22, 2023, "USGS Earthquake Data," United States Geological Survey. [Online]. Available: https://www.usgs.gov.
47. National Seismological Center, Jan. 1966 – May 22, 2023, "NSC Earthquake Data Portal," National Seismological Center. [Online]. Available: https://seismo.gov.in/data-portal.
48. P. G. A. Njock, S.-L. Shen, A. Zhou, and H.-M. Lyu, "Evaluation of soil liquefaction using AI technology incorporating a coupled ENN/T-SNE model," *Soil Dynamics and Earthquake Engineering*, vol. 130, Mar. 2020, Art. no. 105988, doi: 10.1016/j.soildyn.2019.105988.
49. D. Xu, S. Zhang, H. Zhang, and D. P. Mandic, "Convergence of the RMSProp deep learning method with penalty for nonconvex optimization," *Neural Networks*, vol. 139, pp. 17–23, July 2021, doi: 10.1016/j.neunet.2021.02.011.
50. Y. Ding, J. Chen, and J. Shen, "Conditional generative adversarial network model for simulating intensity measures of aftershocks," *Soil Dynamics and Earthquake Engineering*, vol. 139, Dec. 2020, Art. no. 106281, doi: 10.1016/j.soildyn.2020.106281.
51. G. James, D. Witten, T. Hastie, and R. Tibshirani, *Introduction to Statistical Learning: With Applications in R*. 2nd Edition, New York, NY, USA, Springer, 2013, doi: 10.1007/978-1-0716-1418-1.
52. J. Mahmoudi, M. A. Arjomand, M. Rezaei, and M. H. Mohammadi, "Predicting the earthquake magnitude using the multi-layer perceptron neural network with two hidden layers," *Civil Engineering Journal*, vol. 2, no. 1, pp. 1–12, Jan. 2016, doi: 10.28991/cej-2016-00000008.