Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

New Proposed Model for Predicting Earthquakes Details Using Bi-LSTM

[1]Mohammed A. Jaleel Shaneen, [2]Dr. Suhad Malallah Kadhem

[1]Department of Computer Science, University of Technology, Baghdad, Iraq.

cs.22.16@grad.uotechnology.edu.iq

ORCID https://orcid.org/0009-0008-6946-9444

[2]Department of Computer Science, University of Technology, Baghdad, Iraq.

suhad.m.kadhem@uotechnology.edu.iq

ORCID https://orcid.org/0000-0003-0471-8231

ARTICLE INFO

Received: 22 Dec 2024

Revised: 15 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

An earthquake can be defined as a shaking event that occurs when the tectonic plates of Earth move. Significant harm, including fatalities, structural destruction, and economic effects, may result from such events. The majority of models have only been able to predict certain regions, in spite of multiple attempts to predict such events. For predicting the occurrence and location of earthquakes, this research presents two new models. Bi-directional Long Short-Term Memory (BiLSTM) networks were found to be very appropriate by reviewing the literature because of their efficient memory retention qualities. The best model has been selected by utilizing Keras tuner, which allowed for the selection of different dense layer combinations as well as BiLSTM configurations. The model of choice makes use of seismic markers from earthquake catalog of Bangladesh in order to predict the probability of earthquakes in the future month. An attention process has been incorporated into BiLSTM framework to improve prediction accuracy in the occurrence prediction model, yielding an accuracy rate of 80.1%. Also, an attention mechanism was not included in the location prediction model since it would not improve the performance of BiLSTM architecture and would just add needless complexity. Instead, a regression model has been created by using BiLSTM and dense layers for estimating earthquake epicenter relative to fixed point. Obtaining a root mean square error (RMSE) of 1.1830 as a result.

Keywords: *Earthquake, Attention Mechanism, Occurrence, Location, Bi-LSTM.*

I. Introduction

When tectonic plates move, massive amounts of energy that have been stored inside the Earth are released, causing deadly natural phenomena known as earthquakes. These events generally occur along fault lines where the Earth's crustal rocks slide past one another [1]. The core of the Earth contains molten magma, which produces high temperatures and intense energy. This energy seeks to escape, and fault lines provide the necessary pathways, resulting in seismic activity known as earthquakes.

The consequences of earthquakes can be severe, including massive damage to buildings, loss of lives, and the potential onset of secondary disasters like tsunamis. From 1998 to 2017, earthquakes were responsible for approximately 750,000 deaths and impacted about 125 million people globally. Bangladesh, a densely populated nation in South Asia situated at the convergence of the Indian, Burmese, and Eurasian tectonic plates and traversed by five major fault lines, faces significant risk. A magnitude 7.5 earthquake in Dhaka could lead to an estimated 88,000 fatalities, destroy 72,000 buildings, and cause financial losses around $1.075 billion [2][3]. Developing a precise earthquake prediction system could greatly alleviate these severe impacts.
Due to the impact of deep learning technology across various fields, recent research has increasingly focused on this area [4]. Deep learning has enhanced image analysis capabilities, yielding notable and positive outcomes [5].

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Deep learning techniques are focused in contrast to other approaches, on extracting high-level aspects [6]. Convolutional neural networks (CNNs) and A bidirectional long short-term memory (BiLSTM) are a deep learning branch [7]. BiLSTM is characterized by having a low complexity and high scalability [8]. Deep learning incorporates the feature extraction process directly within the learning model, making it particularly well-suited for handling large datasets [9].

Using an extensive collection of interdisciplinary time series data, this research provides a novel model depending on attention-enhanced LSTM to predict earthquake occurrence as well as location. The attention mechanism, the technique, and the results will all be covered thoroughly in the ensuing sections, with concluding remarks given in the last section.

II. Related Works

Recently, deep learning (DL) methods have significantly advanced various fields, including diagnostics and forecasting. In 2019, notable contributions were made across several domains. Noor et al. [10] (2019) reviewed the application of deep learning for diagnosing neurodegenerative diseases from MRI scans, which improved diagnostic accuracy. Maya and Yu [11] (2019) employed transfer learning techniques for short-term forecasting, demonstrating advancements in predictive modeling. Yahaya et al. [12] (2019) proposed a consensus-based ensemble approach for anomaly detection in daily activities. Additionally, Ye et al. [13] (2019) used attention mechanisms to enhance object transformation in images, and Li et al. [14] (2019) applied attention mechanisms to text and social networks for better user attribute classification. Wang et al. [15] (2019) developed an ensemble model to predict customer churn in search ads, and Orojo et al. [16] (2019) used multi-recurrent networks for predicting crude oil prices. In their evaluation of neural networks (NNs)' efficacy in earthquake prediction, Broccardo and Mignan [17] (2019) noted both their advantages and disadvantages.

In 2020, advancements in earthquake prediction models were prominent. Noor et al. [18] (2020) further explored DL to detect neurological disorders from MRI, focusing on Parkinson's, Alzheimer's, and schizophrenia. Kail et al. [19] (2020) combined recurrent and convolutional neural networks to improve earthquake location prediction. Zhu et al. [20] (2020) developed an ensemble learning system for forecasting wind speeds, showcasing the potential of ensemble techniques. Peng et al. [21] (2020) examined the use of DL in biological data mining (DM), highlighting its versatility. Fabietti et al. [22] (2020) used neural networks (NNs) to detect artifacts in neural recordings, and Al Nahian et al. [23] (2020) introduced an AI-driven emotion-aware fall monitoring system. Lin et al. [24] (2020) developed Backpropagation Neural Network (BPNN) models for predicting earthquake magnitudes, while Cai et al. [25] (2020) created a multi-relation structure RNN for modeling temporal point processes to enhance earthquake prediction accuracy. Berhich et al. [26] (2020) utilized LSTM networks for earthquake forecasting, demonstrating improvements over traditional methods.
Additional improvements in DL applications continued in the year 2021. A CNN-based technique for identifying low-magnitude earthquakes through a multi-level sensor network was presented by Shahen et al. [27] (2021), resulting in improved performance. Mahmud et al. [28] (2021) continued to explore deep learning in biological data mining, while Li and Wu [29] (2021) applied clustering techniques to predict stock market trends. Yahaya et al. [30] (2021) developed new methods for detecting anomalies in daily activities, and Yahaya et al. [31] (2021) introduced an adaptive anomaly detection system for dynamic human activity monitoring. Ren et al. [32] (2021) developed a T-S fuzzy model to enhance system stability in positive hidden Markov systems. Mahmud et al. [33] (2021) applied random forests and LSTM networks to predict tourist arrivals, illustrating how deep learning models can handle complex time-series data.
Recent innovations from 2022 to 2024 have further advanced earthquake prediction. Laurenti et al. [34] (2022) improved lab quake prediction accuracy using DL and autoregressive methods. Wang et al. [35] (2022) introduced EEWNet, a DL model which enhances early warning accuracy and speed for earthquake magnitude prediction. Abri and Artuner [36] (2023) used LSTM-based models to predict earthquakes with high accuracy from ionospheric TEC data. Elbes et al. [37] (2023) combined convolutional and LSTM networks for accurate earthquake prediction using seismic data. Wang et al. [38] (2024) developed a VMD-LSTM model that improves earthquake prediction for time, location, and magnitude by integrating variation mode decomposition with LSTM networks. Briones et al. [39] (2024) demonstrated that incorporating pink noise into LSTM networks provides more accurate earthquake magnitude forecasts. DynaPicker, a DL model for accurate seismic phase selection as

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

well as magnitude estimation via dynamic convolutional neural networks (CNNs), was developed by Li et al. [40] in the year 2024 and used for the aftershocks of Kahramanmaras earthquake.

III. ATTENTION MECHANISM

Bahdanau et al. [41] presented the idea of attention to machine translation in the year 2015. This idea was initially created for natural language processing (NLP), but it has now been applied to other machine learning (ML) domains as well [19]. Even while LSTM models are successful, they have trouble focusing across long input sequences and identifying specific segments, which affects their effectiveness. Through enabling the model to focus on the most pertinent portions of input, the attention mechanism helps to overcome such difficulties. Equation (1) illustrates how to determine conditional probability regarding an output event given the input sequence, for example, if $X_1, X_2,...., X_T$ representing input sequence and yi at time i signifies the output sequence.

$$P(y_i \mid y_1, \ldots, y_{i-1}) = g(y_{i-1}, s_i, c_i) \quad (1)$$

Here, s_i represents hidden state that could be computed with the use of equation 2.

$$S_i = f_(y_{i-1}, S_{i-1}, c_i) \quad (2)$$

In this context, the context vector referred by ci, controls how much weight is assigned to each input sequence segment for getting the desired output. The annotations (h_1, h_2, hTx) from which this vector is constructed each provide information regarding the entire sequence, with special attention paid to the region around ith position. Equation (3) provides the value for ci.

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \quad (3)$$

In this context, α_{ij} represents the attention weights applied to each segment of the sequence, which are determined through a softmax operation. Equation (4) provides a mathematical expression for this process.

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \quad (4)$$

The model's alignment score, represented by the term eij, is dependent upon hj and si−1. A feedforward neural network (FFNN), concurrently trained with the entire model, is used to calculate such alignment score. Through such computations, every output is obtained from the input sequence's weighted sum, in which weights are context vector elements. A diagram of the attention mechanism overall is shown in Fig 1. The decoder and encoder are the two primary parts of the architecture. This structure consists of the decoder blocks and encoder blocks (highlighted in red). The decoder Recurrent Neural Network (RNN) receives input from past decoder hidden states and both decoder's input and encoder RNN receive input from preceding hidden states of the encoder. Context vector utilizes hidden states of the encoder in order to produce context for the final output, the decoder input is affected by the preceding context vector's output as well as overall output.

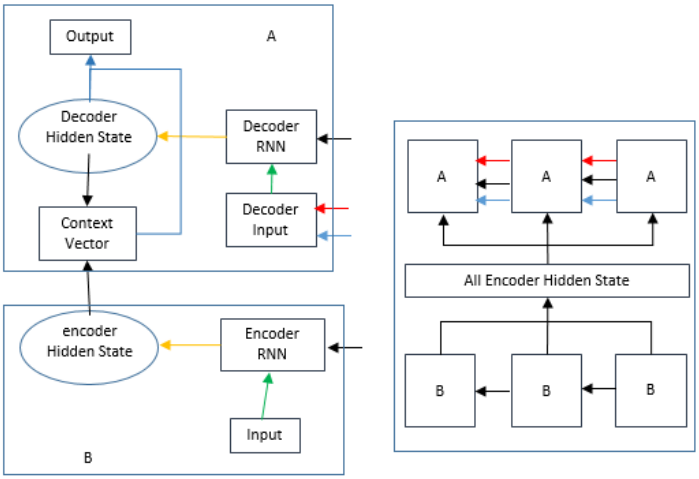

Figure 1. A rchitecture of attention mechanism [16].

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

IV. METHODOLOGY

Data was collected for this study from two different sources and then preprocessed. From the dataset, a total of 8 seismic indicators have been obtained. BiLSTM model with attention has been used to predict the occurrence of earthquakes. Fig 2 shows the analysis pipeline and comprehensive approach. The datasets from USGS and the Bangladesh Meteorological Department that have been used to calculate the seismic indicators had to be cleaned as part of the data processing process. HCTSA library made it easier to calculate multi-domain features, which were then used in conjunction with feature selection algorithms to determine which features were most pertinent. For predicting both the location and occurrence of earthquakes, different prediction models have been created by merging dense layer with BiLSTM models.

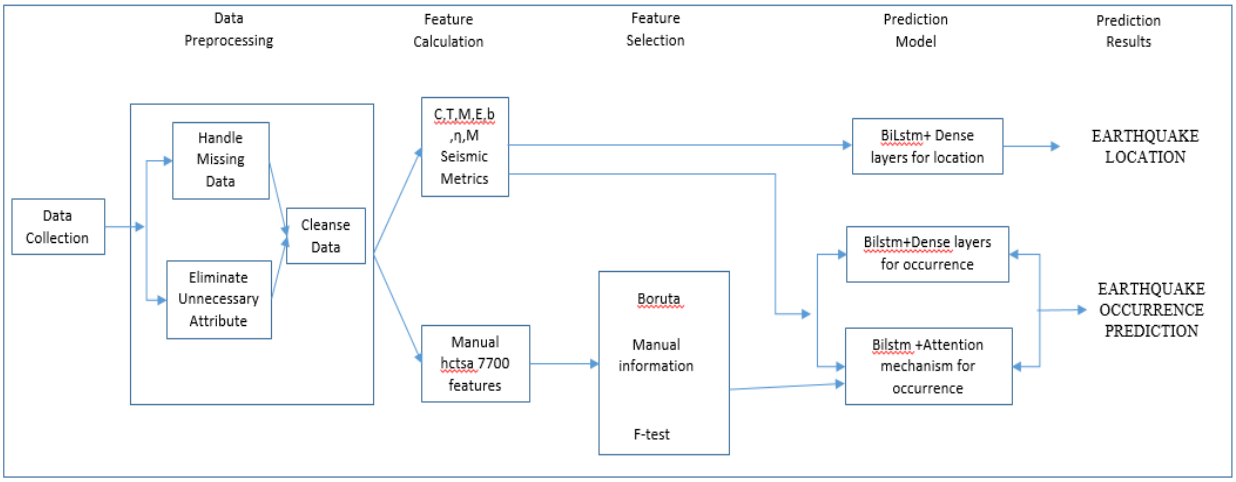

Fig 2. The analysis pipeline and overall methodology.

The occurrence prediction model performed better once the attention mechanism was added. Ultimately, performance criteria have been used for comparing and assessing the models.

A. DATASET COLLECTION

The earthquakes near Bangladesh were the main topic of this case study. The dataset included earthquake catalogs from the U.S Geological Survey (USGS) for the same period of time as well as the Bangladesh Meteorological Department (covering 1950 to 2019) [42]. A total of 6 features are included in the dataset from Bangladesh Meteorological Department: time, date, latitude, longitude, depth, and magnitude. On the other hand, the USGS dataset has 17 more variables, including earthquake ID, depth error, disaster type, and update date. Other attributes were not included, and just the magnitude feature from USGS dataset has been used for consistency's sake. Since many magnitude scales are utilized for measuring earthquakes, the dataset was standardized to a particular scale using the magnitude type parameter, with the Richter scale being the default. The region surrounding Bangladesh, which spans latitudes 18.11° N to 27.11° N and longitudes 87.19° E to 95.36° E, was the subject of data collection. From this location, 1,764 earthquake recordings were taken and utilized for the calculation of features for the prediction analysis.

B. DATA-SET PRE-PROCESSING

Preparing the data is an essential stage in making sure the predictions are accurate. Data have been cross-verified between USGS and Bangladesh Meteorological Department in order to find discrepancies in earthquake catalogs. All magnitudes have been standardized to Richter scale, and missing values were resolved. The time, date, latitude, longitude, magnitude type, magnitude, and depth were among the features that were kept for computation. Both foreshocks and aftershocks were taken out of the dataset. A total of 8 features—also referred to as seismic indicators—were then calculated using the main shocks as a basis.

C. CALCULATION OF SEISMIC FEATURES

Features relevant to earthquake research have been computed in the presented work. A total of eight seismicity indicators value of b (b), MD (i.e., magnitude deficit), MSD (i.e., mean square deviation), MM (i.e., mean

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

magnitude), elapsed days (ED), mean time between characteristic events (MTBCE), coefficient of variation from mean time (CVFMT), and RSRER (i.e., rate of square root of energy released) were identified by Panakkat and Adeli [31] as being utilized for earthquake prediction. Since then, the discipline has largely embraced such indications. As a result, such 8 seismicity indicators have been computed for this study on a monthly basis, taking into account the 50 events that occurred prior to each computation.

The eight seismicity indicators are detailed in the following way:

1) ELAPSED DAYS

This indicator measures the time that had elapsed since the occurrence of the last n earthquakes with magnitudes exceeding a specified threshold. It is calculated using the equation:

$$ED = t_n - t_1 (5)$$

In which the first event occurs at time 1 and the nth event occurs at time tn. n was fixed at 50 in this work. A lower ED value indicates that there have been more large-scale earthquakes recently.

2) MEAN MAGNITUDE (MM)

This represents the average magnitude of n most recent earthquakes measured on the Richter scale. It can be calculated as:

$$MM = \frac{\sum M_i}{n} \quad (6)$$

3) RSRER. Energy (E): The energy released by an earthquake is derived from its magnitude M on the Richter scale, using the formula:

$$\mathrm{E} = 10^{(11.80+1.50\mathrm{M})}\mathrm{ergs} \quad (7)$$

The Value of RSRER could be calculated as follows.

$$\mathrm{RSRER} = \frac{\sum \mathrm{E}^{1/2}}{\mathrm{ED}} \quad (8)$$

4) B-VALUE (b)

b-value: The slope of the log of frequency-magnitude distribution curve is represented by parameter. It comes from the law known as Gutenberg-Richter (G-R)

$$\log_{10} N = a - bM \quad (9)$$

In which N represent the number of events with magnitudes more than or equal to M, and b, a, and C are constants. One method for calculating the value of an is:

$$\mathrm{a} = \frac{\sum (\log_{10} + \mathrm{bMi})}{\mathrm{n}} \quad (10)$$

The i-th magnitude in this case is Mi, and the number of events with magnitudes of Mi or higher is Ni. The equation for calculating b-value as follows.

$$\mathrm{b} = \frac{(\mathrm{n}\sum \mathrm{M_i}\log_{10} \mathrm{N_i}) - \sum \mathrm{M_i}\sum \log_{10} \mathrm{N_i})}{((\sum \mathrm{M_i})^2 - \mathrm{n}\sum \mathrm{M_i^2})} \quad (11)$$

5) MEAN SQUARE DEVIATION (MSD)

Through adding up mMSD values from predicted G-R line, this metric expresses how much of a variation there is from the G-R law. It is stated as

$$\mathrm{MSD} = \frac{\sum (\log_{10} \mathrm{N_i} - (\mathrm{a} - \mathrm{bMM_i})^2}{(\mathrm{n}-1)} \quad (12)$$

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6) MAGNITUDE DEFICIT (MD)

This calculates the discrepancy between the largest magnitude predicted by the G-R equation and the maximum observed magnitude in n events. It is calculable with the help of.

$$MD = M_{\text{max observed}} - M_{\text{max expected}} \quad (13)$$

Mmax expected could be computed as follows.

In which Mmax is the maximum observed magnitude and Mmax is calculated as:

$$\mathrm{M}_{\text{max expected}} = \mathrm{a/b} \ (14)$$

7) MEAN TIME BETWEEN CHARACTERISTIC

EVENTS (MTBCE) Based on elastic rebound hypothesis [43], this feature measures the time interval between significant earthquakes with magnitudes between 7 and 7.5. It is calculated as

$$\mathrm{MTBCE} = \frac{\sum (\mathrm{t_{icharacteristics}})}{\mathrm{n_{characteristics}}} (15)$$

The number of occurrences is ncharacteristics, and time interval between 2 characteristic events is ti characteristics.

8) COEFFICIENT OF VARIATION FROM MEAN TIME (CVFMT)

By comparing time distribution between typical earthquakes to the expected magnitude distribution, this indicator evaluates how variable the distribution is. It is calculated with the help of:

$$\mathrm{MTBCE} = \frac{\sum (\mathrm{t_{icharacteristics}})}{\mathrm{n_{characteristics}}} (16)$$

A total of 495 time-series sequences have been examined for this study, and they were split into testing (30%, or 150 sequences) and training (70%, or 345 sequences) sets. The training procedure was kept apart from the testing data. Furthermore, 7,700 multi-domain features with a sequence length of 50 earthquake magnitudes as time series were produced with the use of HCTSA (i.e., Highly Comparative Time Series Analysis) library [44].

D. SYSTEM CONFIGURATION

Kaggle kernel has been utilized for the experiments in this study, providing 32GB of RAM, an NVIDIA GPU, and a CPU core i9. Python was used to create the earthquake location as well as occurrence prediction models. The libraries Keras, Scikit-learn, Tuner, pandas, NumPy, BorutaPy, and Statsmodels have been utilized for model development, comparisons, and feature calculations.

E. THE MODEL OF EARTHQUAKE OCCURRENCE PREDICTIONS

The architecture regarding the suggested a model for earthquake occurrence prediction is shown in Fig 3. Finding the best LSTM as well as dense layer combinations for the model has been the main goal at first. The best-performing models were identified by experimenting with different configurations using Keras Tuner library. Every configuration was examined over ten trials to guarantee stability, with the goal of the tuning procedure being for maximizing validation accuracy. The best model to predict earthquake magnitudes was chosen after each model underwent 1000 training epochs. An initial LSTM layer of 200 neurons was part of the optimized model architecture, and it was succeeded by two bi-directional LSTM layers with 100 and 50 neurons, respectively. Also, it has two more dense layers with 12 neurons each, a dense layer with 25 neurons, a flatten layer, and a final dense output layer with 2 neurons. Every layer was trainable, and all save the output layer employed the tan h activation function. All LSTM as well as bi-directional layers in such deep model underwent L1 and L2 regularization to mitigate any possible overfitting concerns. The foundation of the earthquake occurrence prediction method was this model. After 10,000 epochs of training with a 0.01 learning rate on the feature set, the model was evaluated on the testing set. Because of its ability to reduce gradient vanishing and explosion issues, the BiLSTM architecture is well-suited for applications in which there is a substantial discrepancy between the state of knowledge today and the past. A total of 3 gates comprise an LSTM cell: forget, input, and output gates, in addition to a cell state. The forget gate eliminates unnecessary data, the output gate chooses the next hidden state information, and the input gate controls the information to be added. The input parameters of the model have been defined in this work by using the 8 seismic features as inputs to LSTM.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Xt = \begin{bmatrix} b\ value \\ MSD \\ MD \\ ED \\ MM \\ RSRER \\ MTBCE \\ MTBCE \\ CVFMT \end{bmatrix} \quad (17)$$

First, a forget gate was used to process the current inputs together with the previous hidden state, ht-1. A forget gate's output is calculated in the following way:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (18)$$

In which bf stands for the forget gate's bias and Wf stands for the weights connected to it. LSTM cell's memory, or cell state, is updated under the direction of input gate. To decide how to change the cell state based on current input and hidden state, this gate employs the sigmoid as well as tanh functions. These functions' outputs are computed as follows:

The sigmoid function output:

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (19)$$

The tanh function output:

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \quad (20)$$

In this procedure, the weights of input gate as well as cell state are represented by Wi and WC, respectively, whereas their biases are indicated by bi and bC. The output from the input gate is multiplied point-wise by the output from forget gate for updating the state of the cell. The update could be written as follows if the prior cell state was Ct and the current cell state is Ct:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (21)$$

The next equations are used by the output gate for determining the subsequent hidden state:

$$o_t = \sigma(W_0 \times [h_{t-1}, x_t] + b_o) \quad (22)$$

The presented work defines the output gate's sigmoid output (ot), weights (Wo), hidden state (ht), and bias (bo). The attention layer subsequently processed the hidden state. Luong attention, sometimes referred to as multiplicative attention, was the attention mechanism used since it was more computationally efficient than additive attention. With an attention width set to the preceding 20 inputs, this attention layer has been positioned before the flatten layer. To this layer, L2 and L1 regularization were also applied. A 10,000 period training of the model produced notable performance gains. The model was then tested by benchmarking it against current earthquake prediction research, and it proved to be quite successful. To improve model performance and convergence, a total of 7,700 features were computed and normalized in order to evaluate the influence of multi-domain time-series features. ANOVA F-test [46], Boruta [47], and mutual information [45] were among the feature selection strategies used. The Boruta algorithm focused on picking the most significant qualities, it only highlighted two key features. In contrast, mutual information and ANOVA F-test found the top 20 traits. The suggested attention-LSTM architecture was after that combined with the chosen features from mutual information and ANOVA F-testing for predicting the occurrence of earthquake.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

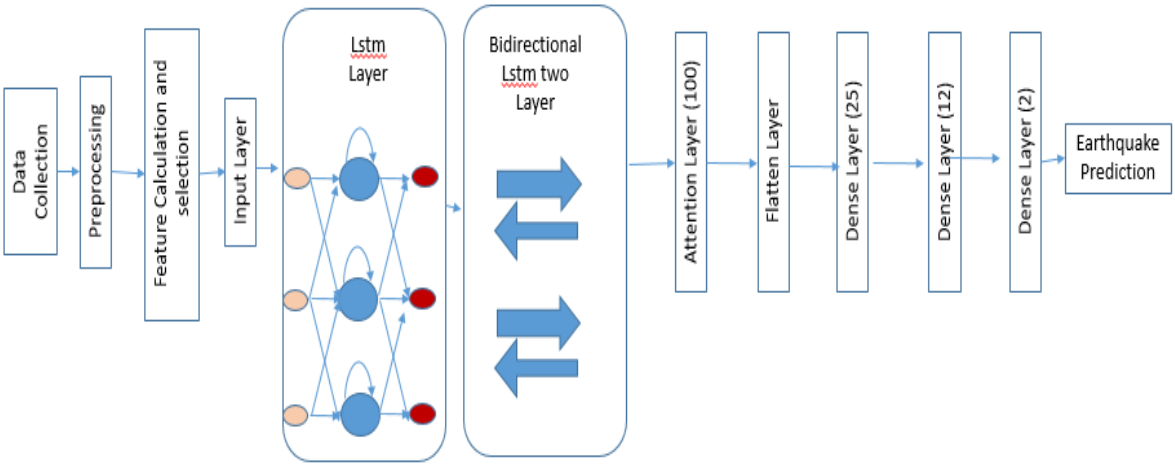

Figure3. Earthquake occurrence prediction model suggested.

F. EARTHQUAKE LOCATION PREDICTION MODEL

A different model, as shown in Fig 4, was used to predict earthquake location. LSTM layer, bidirectional layers, and dense layers are used to create this model. This model does not make use of the attention mechanism. This model calculates distance between Dhaka city and the earthquake epicenter, instead of predicting exact longitude and latitude. This method works well since the effects of an earthquake cover a wide area, therefore identifying the affected area is adequate. Equation (24) is the mathematical expression of Campbell's equation [48], which serves as the foundation for calculating the distance between two geographical sites.

$$D = 68.9722 \cdot \Gamma \quad (23)$$

where,

$\Gamma=[\cos^{-1}\{(\sin a)(\sin b)+(\cos b)(\cos a)(\cos P)\}]$ (24)

with a & b represent latitude of 1st & 2nd point,

With the use of Keras Tuner, the best configuration for the location prediction model was determined to consist of 2 bidirectional LSTM layers with 100 and 50 neurons each, after which there was an LSTM layer with 200 neurons. In addition, the model has 2 dense layers with 12 and 25 neurons each, as well as a flatten layer. The final output layer was a single neuron with no activation function, a regression model. The optimization objective for the tuner was to minimize validation loss. Every layer made use of the tanh activation function, with the exception of the output layer. A total of 10000 epochs were used to train this model. The RMSE and MSE were used to assess performance.

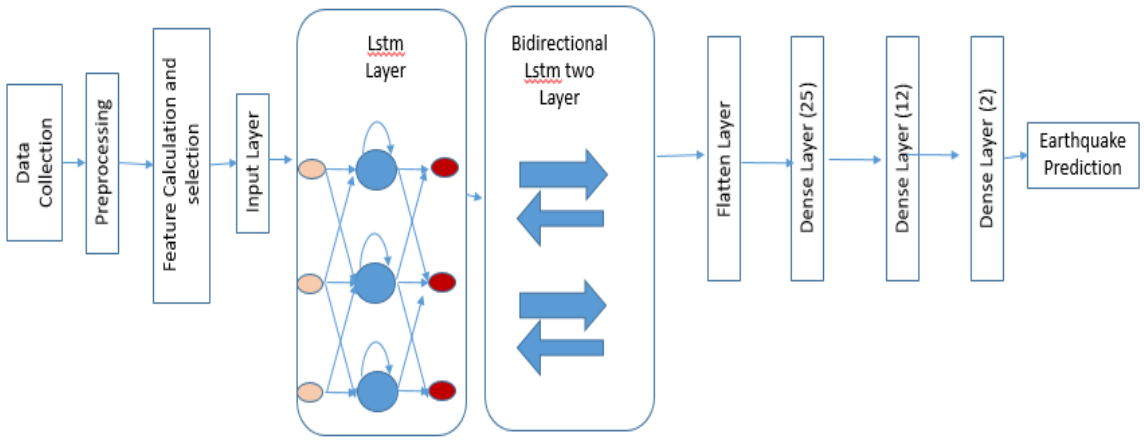

Figure4. The suggested model for location prediction.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

V. ANALYSIS OF RESULTS

A. EARTHQUAKE OCCURRENCE PREDICTION RESULTS

1. **Performance of BiLSTM Model:**

Training and Testing: The BiLSTM model has been trained for 10,000 epochs and evaluated on unseen data. Figure 5 displays confusion matrix, while Figure 6 shows the ROC curve. After evaluating various learning rates, a rate of 0.01 was identified as the most effective after attention mechanism, as outlined in Table 1. The model achieved before attention mechanism approximately 75% accuracy, with a sensitivity (Sn) of 0.8912, indicating strong performance on positive samples. However, it exhibited a lower specificity (Sp) of 0.6054, suggesting a higher rate of false alarms. Area under the ROC curve (AUC) has been 0.75.

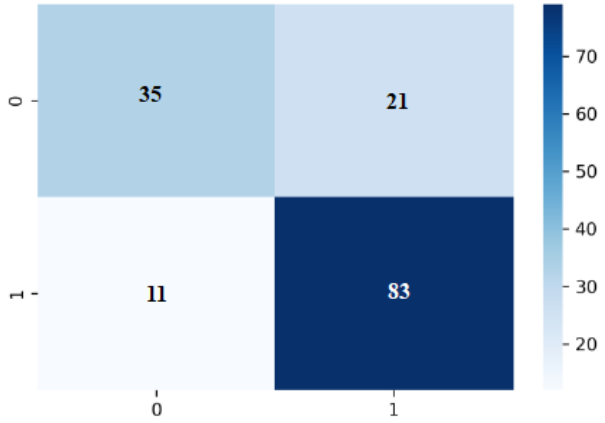

Figure5. Confusion matrix of BiLSTM. accurately classified 83 out of 94 earthquake events, with 35 instances of false alarms in earthquake detection.

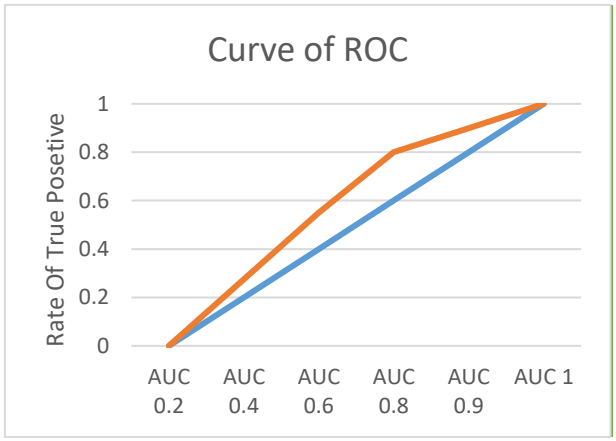

Figure 6. The suggested model's ROC curve, which had provided a 0.75 AUC.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

TABLE1. Change in accuracy depending on the learning rate choice.

Learning Rate value	Accuracy value
0.10	0.6867
0.050	0.7067
0.030	0.7434
0.010	0.80.1
0.001	0.6434

Attention Mechanism: Integrating an attention mechanism significantly enhanced the model's performance, as illustrated in Figure 7. This addition enabled the model to accurately predict 120 out of 150 events, reducing the number of false positives to 21, compared to 35 with the standard BiLSTM model. AUC improved to 0.75%, and accuracy 0.81% The attention-based model required 22.368 seconds to train for 100 epochs, which is only 2 minutes and 24 seconds longer than the training time for the BiLSTM model without attention. This additional training time is considered acceptable given the model's monthly prediction cycle**.** Table 2. s h o w Performance of attention-based LSTM model.

TABLE2. A ttention-based LSTM Model Performance.

The Metrics	Value
True Negative	35
False Positive	21
False Negative	9
True Positive	83
Sensitive (Sn)	0.8912
Specific (Sp)	0.6054
P0	0.7890
P1	0.7800
AUC	0.75

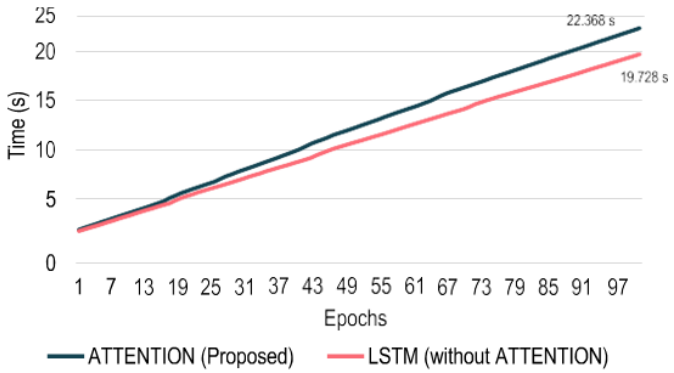

Figure 7. A ttention mechanism impacts in training time cases after the addition with the BiLSTM.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2.Feature Selection Techniques:
Mutual Information: Out of 7,700 features, the mutual information method selected the best 20 features, yielding a 70.67% accuracy model. However, this model demonstrated a tendency to predict earthquakes excessively, classifying 148 out of 150 samples as earthquake events.

Boruta: The Boruta method selected only 2 features from the 7,700, producing a model with 72% accuracy but a high false alarm rate. The model exhibited a sensitivity of 0.9500 and a specificity of 0.2300.

F-test: When using the top 20 features selected by the F-test, the model achieved 70.67% accuracy with a high sensitivity of 0.9500. However, it showed very low specificity of 0.2300 and a negative predictive value of 0.6000, indicating suboptimal performance for non-earthquake events.

3. Comparison with Other Models:

Narayanakumar's LM Model: This model, comprising three layers with 12 neurons in each hidden layer, attained an accuracy of 61.87% (see Figure 9(a)). It exhibited a tendency to overlook earthquakes, incorrectly classifying 37 earthquake events as non-earthquakes.
Bhandarkar's Model: With the use of a dropout layer and two 40-neuron LSTM layers, this model's accuracy was 58.67% (see Fig 9(b)). It was typified by a high false alarm rate.

Aslam's ANN Model: This model, which included two fully connected layers with sigmoid activation, reached an accuracy of 61.34% but frequently predicted all events as earthquakes (see Figure 9(c)).

Wang's Model: Featuring a single LSTM layer followed by two dense layers, this model achieved a 54.67% accuracy (illustrated in Figure 9(d)).

4. Performance Comparison:
Proposed Model vs. Existing Models: The suggested model outperformed all other models in terms of accuracy as well as average parameters such as the positive predictive value, specificity, sensitivity, and negative predictive value, as shown in Fig 8. In particular, it showed a 15% improvement in the Unweighted Average Recall (UAR) over other models and 18.23% higher accuracy than Narayanakumar's LM model.

Multi-Domain Feature Models: The suggested attention-based model's performance is contrasted with that of initial LSTM models and models that make use of multi-domain features in Fig 10. While the suggested model obtained 80.1% accuracy, attention-based model with Boruta feature selection only managed to reach 72%. The suggested model's UAR was 0.75, which is 7% higher than the basic LSTM model.

Machine Learning (ML) Classifiers: When compared to ML-based earthquake prediction models (see Figure 11), the suggested model demonstrated superior performance. Among ML classifiers, the Random Forest (RF) algorithm performed the best, yet the proposed model was 16% more accurate and 12% better in UAR compared to the second-best ML model. The Logistic Regression (LR) classifier showed the lowest performance.

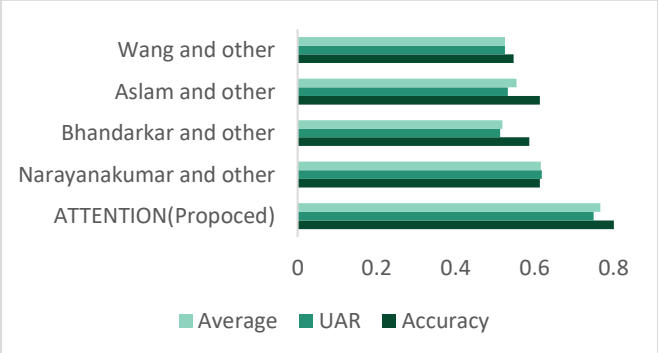

Figure8. Comparing the suggested model's performance to those of models Aslam *etal*., Wang *etal*., Narayanakumar *etal*., and Bhandarkar *etal*. When it came to earthquake occurrence prediction, the suggested architecture outperformed all of the previously described models according to results of UAR and accuracy.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

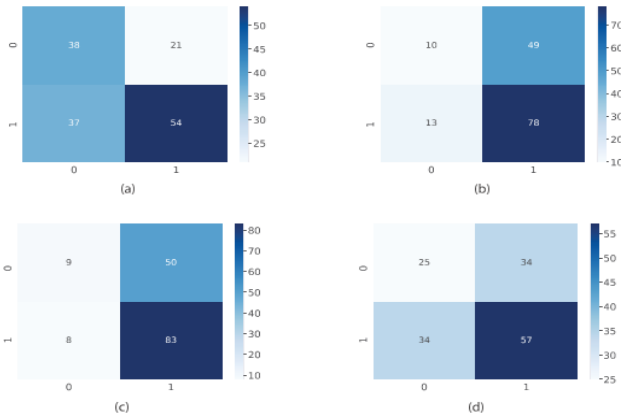

Figure (9) (a) Narayanakumar's model confusion matrix: identified Correctly 54 from 91 earthquake events. (b) Bhandarkar et al 's LSTM confusion matrix: Demonstrates higher sensitivity but a high rate of false alarms. (c) Confusion matrix of Aslam et al.'s proposed architecture: Predicted 133 out of 150 events as earthquakes, with many false alarms. (d) Confusion matrix of Wang et al.'s proposed model: Correctly identified 57/91 earthquake events.

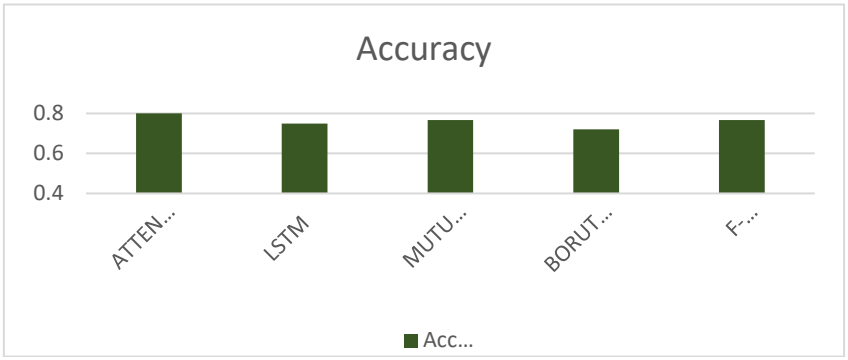

Figure10. Comparison of the suggested model of earthquake predictions with multi-domain feature models and LSTM: The suggested model, using eight indicators, outperformed other models in accuracy, sensitivity, and specificity for predicting earthquake occurrences

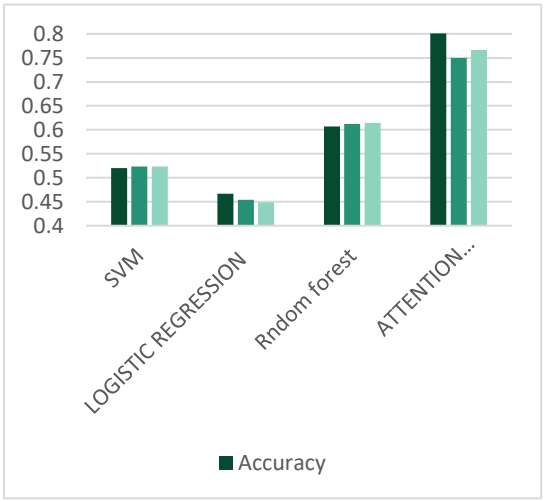

Figure11. Comparison of the suggested models to Machine Learning-based models: While the Random Forest (RF) classifier performed best among the ML models, it did not surpass the proposed model in predicting earthquake occurrences.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

B. LOCATION PREDICTION RESULTS

The predicting of the distance between Bangladesh's capital city Dhaka and the epicenter of the largest earthquake each month is the main goal. Since an earthquake usually affects several hundred kilometers, the distance from Dhaka is determined using the latitude and longitude of such earthquakes. It is considered practical to calculate distances from Dhaka, a significant urban hub with vital infrastructure. The location prediction model's performance was assessed by calculating MSE and RMSE. In contrast to the model of occurrence prediction, the location prediction model does not include an attention mechanism since doing so would have only increased complexity and not improved the efficiency of the BiLSTM architecture. As a result, the attention layer was left out. A different set of 150 data was used to test the regression model after it had been trained for 10,000 epochs. The comparison of the actual and predicted earthquake locations is illustrated in Figure 12, where the predicted locations are represented by an orange line and the actual locations by a blue line. When earthquakes happen close to Dhaka, the model shows good forecasts that closely match the expected values. Yet, the model's projections are less accurate for earthquakes that occur further out from the city center. For earthquake location prediction, this model's MSE of 1.400 and RMSE of 1.1830 are excellent.

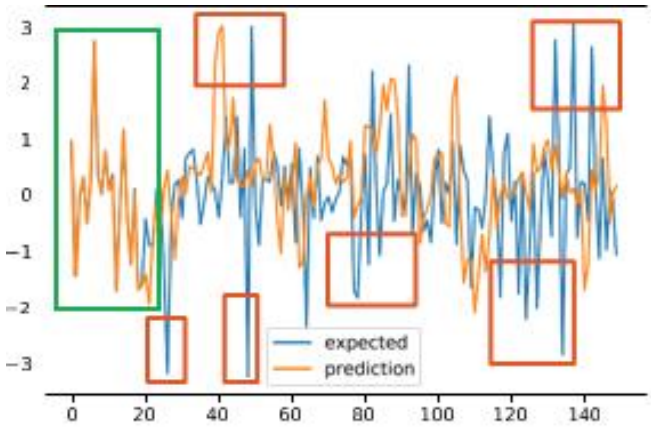

Figure12. Analyses of Location Prediction Results: The predicted locations closely matched the expected locations.

In the areas indicated by the green box, the predicted distances align closely with the expected values, demonstrating high accuracy in these instances. Conversely, the red-boxed regions show peaks in the distance predictions. Those peaks are uncommon and challenging to predict, therefore the model usually treats them as outliers. However, a few of such outlier events are successfully predicted by the suggested model.

It is noteworthy that the model's performance decreases for earthquakes located at greater distances from the city center. This reduced accuracy is anticipated, as the influence of earthquake energy diminishes with increasing distance. Despite this limitation, the proposed model remains effective for predicting earthquakes near the city, where its performance is dependable and valuable for location prediction.

VI. CONCLUSION

The Earth provides essential resources for life, but natural disasters, like earthquakes, could have devastating effects on human civilizations. Throughout history, many empires have been obliterated by these events, which not only destroy infrastructure but also lead to significant loss of life. The region studied in this research has a history of major earthquakes and is expected to experience similar events in the future. One of the major challenges with earthquakes is the absence of reliable precursors, making effective prediction models crucial.

This work employed two models to predict both earthquake occurrences and locations, analyzing historical earthquake data from Bangladesh as a time series. Based on literature, BiLSTM emerged as a powerful algorithm for time-series analysis. The first model integrated an attention mechanism into the BiLSTM framework, achieving an impressive accuracy of 80.1% in predicting earthquake occurrences with eight seismic indicators. Comparisons with various machine learning algorithms showed that the proposed model significantly outperformed its counterparts.

The second model effectively predicted earthquake locations with a RMSE of 1.400, using BiLSTM combined with dense layers but without an attention mechanism. The primary aim of the presented work was developing a robust earthquake prediction system and determine the most effective feature set. Although the proposed models demonstrated strong performance for the work area, there is potential for further accuracy improvements. Future research aimed at refining these models could significantly advance the field of earthquake prediction.

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, had reviewed the results and approved the final version of this manuscript

REFERENCES

[1] N. Absar, S. N. Shoma, and A. A. Chowdhury, ‘‘Estimating the occurrence probability of earthquake in Bangladesh,’’ Int. J. Sci. Eng. Res, vol. 8, no. 2, pp. 1–8, 2017.
[2] A. A. Zaman, S. Sifty, N. J. Rakhine, A. Abdul, R. Amin, M. Khalid,
M. H. Tanvir, K. Hasan, and S. Barua, ‘‘Earthquake risks in Bangladesh and evaluation of awareness among the university students,’’ J. Earth Sci. Climatic Change, vol. 9, no. 7, p. 482, 2018, doi: 10.4172/2157- 7617.1000482.
[3] M. Rahman, S. Paul, and K. Biswas, ‘‘Earthquake and Dhaka city— An approach to manage the impact,’’ J. Sci. Found., vol. 9, nos. 1–2, pp. 65–75, Apr. 2013, doi: 10.3329/jsf.v9i1-2.14649.
[4] W. J. Hadi, S. M. Kadhem, and A. R. Abbas, ‘‘A survey of deepfakes in terms of deep learning and multimedia forensics,’’ Int. J. Electra. Compute. Eng., vol. 12, no. 4, pp. 4408–4414, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4408-4414.
[5] S. T. Ahmed and S. M. Kadhem, ‘‘Using machine learning via deep learning algorithms to diagnose the lung disease based on chest imaging: A survey,’’ Int. J. Interactive Mobile Technol., vol. 15, no. 16, pp. 95–112, Aug. 2021, doi:10.3991/ijim.v15i16.24191.
[6] Ali A. Hussan, Shaimaa H. Shaker, Akbas Ezaldeen Ali (2024). Strangeness detection from crowded video scenes by hand-crafted and deep learning features. *Journal of Soft Computing and Computer Applications*, 1(1), Article 1005. Available at: https://jscca.uotechnology.edu.iq/jscca/vol1/iss1/6
[7] Shahad Fadhil Abbas, Shaimaa Hameed Shaker, Firas. A. Abdullatif (2024). Face mask detection based on deep learning: A review. *Journal of Soft Computing and Computer Applications*, 1(1), Article 1006. Available at: https://jscca.uotechnology.edu.iq/jscca/vol1/iss1/7'
[8] Hussein K. Abdul Atheem , Israa T. Ali, Faiz A. Al Alawy (2024). A comprehensive analysis of deep learning and swarm intelligence techniques to enhance vehicular ad-hoc network performance. Journal of Soft Computing and Computer Applications, 1(1), Article 1004. Available at: https://jscca.uotechnology.edu.iq/jscca/vol1/iss1/5
[9] S. T. Ahmed and S. M. Kadhem, ‘‘Early Alzheimer’s disease detection using different techniques based on microarray data: A review,’’ Int. J. Online Biomed. Eng., vol. 18, no. 4, pp. 106–126, Mar. 2022, doi: 10.3991/ijoe.v18i04.27133.
[10] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, M. Mahmud, and S. Al Mamun, ‘‘Detecting neurodegenerative disease from MRI: A brief review on a deep learning perspective,’’ in Brain Informatics (Lecture Notes in Computer Science), P. Liang, V. Goel, and C. Shan, Eds. Cham, Switzerland: Springer, 2019, pp. 115–125, doi: 10.1007/978-3-030-37078-7_12.
[11] M. Maya and W. Yu, ‘‘Short-term prediction of the earthquake through neural networks and meta-learning,’’ in Proc. 16th Int. Conf. Electr. Eng., Comput. Sci. Automat. Control (CCE), Sep. 2019, pp. 1–6, doi: 10.1109/ICEEE.2019.8884562.
[12] S. W. Yahaya, A. Lotfi, and M. Mahmud, ‘‘A consensus novelty detec- tion ensemble approach for anomaly detection in activities of daily living,’’ Appl. Soft Comput., vol. 83, Oct. 2019, Art. no. 105613, doi: 10.1016/j.asoc.2019.105613.

[13] Z. Ye, F. Lyu, L. Li, Y. Sun, Q. Fu, and F. Hu, ‘‘Unsupervised object transfiguration with attention,’’ Cognit. Comput., vol. 11, no. 6, pp. 869–878, Dec. 2019, doi: 10.1007/s12559-019-09633-3.
[14] Y. Li, L. Yang, B. Xu, J. Wang, and H. Lin, ‘‘Improving user attribute classification with text and social network attention,’’ Cognit. Comput., vol. 11, no. 4, pp. 459–468, Aug. 2019, doi: 10.1007/s12559-019-9624-y.
[15] Q.-F. Wang, M. Xu, and A. Hussain, ‘‘Large-scale ensemble model for customer churn prediction in search ads,’’ Cognit. Comput., vol. 11, no. 2, pp. 262–270, Apr. 2019, doi: 10.1007/s12559-018-9608-3.
[16] O. Orojo, J. Tepper, T. M. McGinnity, and M. Mahmud, ‘‘A multi- recurrent network for crude oil price prediction,’’ in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Dec. 2019, pp. 2953–2958, doi: 10.1109/SSCI44817.2019.9002841.
[17] A. Mignan and M. Broccardo, ‘‘Neural network applications in earthquake prediction (1994–2019): Meta-analytic insight on their limitations,’’ 2019, arXiv:1910.01178. [Online]. Available: http://arxiv.org/abs/1910.01178, doi: 10.1785/0220200021.
[18] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, and M. Mahmud, ‘‘Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer’s disease, Parkinson’s disease and schizophrenia,’’ Brain Informat., vol. 7, no. 1, pp. 1–21, Dec. 2020, doi: 10.1186/s40708-020-00112-2.
[19] R. Kail, A. Zaytsev, and E. Burnaev, ‘‘Recurrent convolutional neural net- works help to predict location of earthquakes,’’ 2020, arXiv:2004.09140. [Online]. Available: http://arxiv.org/abs/2004.09140
[20] L. Zhu, C. Lian, Z. Zeng, and Y. Su, ‘‘A broad learning system with ensemble and classification methods for multi-step-ahead wind speed prediction,’’ Cognit. Comput., vol. 12, no. 3, pp. 654–666, May 2020, doi: 10.1007/s12559-019-09698-0.
[21] L. Peng, L. Wang, X.-Y. Ai, and Y.-R. Zeng, ‘‘Forecasting tourist arrivals via random forest and long short-term memory,’’ Cognit. Comput., vol. 13, pp. 1–14, Jun. 2020, doi: 10.1007/s12559-020-09747-z.
[22] M. Fabietti, M. Mahmud, A. Lotfi, A. Averna, D. Guggenmo, R. Nudo, and
M. Chiappalone, ‘‘Neural network-based artifact detection in local field potentials recorded from chronically implanted neural probes,’’ in Proc. IJCNN, 2020, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207320.
[23] M. J. Al Nahian, T. Ghosh, M. N. Uddin, M. M. Islam, M. Mahmud, and M. S. Kaiser, ‘‘Towards artificial intelligence driven emotion aware fall monitoring framework suitable for elderly people with neurological disorder,’’ in Proc. Int. Conf. Brain Informat. Cham, Switzerland: Springer, 2020, pp. 275–286, doi: 10.1007/978-3-030-59277-6_25.
[24] J.-W. Lin, ‘‘Researching significant earthquakes in Taiwan using two back- propagation neural network models,’’ Natural Hazards, vol. 103, no. 3, pp. 3563–3590, 2020, doi: 10.1007/s11069-020-04144-z.
[25] H. Cai, T. T. Nguyen, Y. Li, V. W. Zheng, B. Chen, G. Cong, and X. Li, ‘‘Modeling marked temporal point process using multi-relation structure RNN,’’ Cognit. Comput., vol. 12, no. 3, pp. 499–512, May 2020, doi: 10.1007/s12559-019-09690-8.
[26] A. Berhich, F.-Z. Belouadha, and M. I. Kabbaj, ‘‘LSTM-based models for earthquake prediction,’’ in Proc. 3rd Int. Conf. Netw., Inf. Syst. Secur., 2020, pp. 1–7, doi: 10.1145/3386723.3387865.
[27] Shaheen, A., bin Waheed, U., Fehler, M., Sokol, L., & Hanafy, S. (2021). Groningen Net: Deep Learning for Low-Magnitude Earthquake Detection on a Multi-Level Sensor Network. *Sensors, 21(23), 8080*. https://doi.org/10.3390/s21238080.
[28] M. Mahmud, M. S. Kaiser, T. M. McGinnity, and A. Hussain, ‘‘Deep learning in mining biological data,’’ Cognit. Comput., vol. 13, no. 1, pp. 1–33, Jan. 2021, doi: 10.1007/s12559-020-09773-x.
[29] X. Li and P. Wu, ‘‘Stock price prediction incorporating market style cluster- ing,’’ Cognit. Comput., vol. 13, pp. 1–18, Jan. 2021, doi: 10.1007/s12559- 021-09820-1.
[30] S. W. Yahaya, A. Lotfi, and M. Mahmud, ‘‘Detecting anomaly and its sources in activities of daily living,’’ Social Netw. Comput. Sci., vol. 2, no. 1, pp. 1–18, Feb. 2021.
[31] S. W. Yahaya, A. Lotfi, and M. Mahmud, ‘‘Towards a data-driven adaptive anomaly detection system for human activity,’’ Pattern Recognit. Lett., vol. 145, pp. 1–8, May 2021, doi: 10.1016/j.patrec.2021.02.006.
[32] C. Ren, S. He, X. Luan, F. Liu, and H. R. Karimi, ‘‘Finite-time L2-gain asynchronous control for continuous-time positive hidden Markov jump systems via T–S fuzzy model approach,’’ IEEE Trans. Cybern., vol. 51, no. 1, pp. 77–87, Jan. 2021, doi: 10.1109/TCYB.2020.2996743.2921238.
[33] M. Mahmud, M. S. Kaiser, T. M. McGinnity, and A. Hussain, ‘‘Deep learning in mining biological data,’’Cognit. Comput., vol. 13, no. 1, 2021, doi: 10.1007/s12559-020-09773-x.
[34] Laurenti, L., Tinti, E., Galasso, F., Franco, L., & Marone, C. (2022). Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth and Planetary Science Letters*, 570, 117825. https://doi.org/10.1016/j.epsl.2022.117825

[35] Wang, Y., Li, X., Wang, Z., & Liu, J. (2022). Deep learning for magnitude prediction in earthquake early warning.
https://doi.org/10.1016/j.gr.2022.06.009.
[36] Abri, R., & Artuner, H. (2023). LSTM-Based Deep Learning Methods for Prediction of Earthquakes Using Ionospheric Data. *Hacettepe University, Department of Computer Engineering, Ankara, Turkey*.
[37] Elbes, M., AlZu'bi, S., & Kanan, T. (2023). Deep Learning-Based Earthquake Prediction Technique Using Seismic Data. *2023 International Conference on Multimedia Computing, Networking and Applications (MCNA)*. DOI: 10.1109/MCNA59361.2023.10185869.
[38] Wang, Q., Zhang, Y., Zhang, J., & He, X. (2024). On the use of VMD-LSTM neural network for approximate earthquake prediction. *DOI: 10.1007/s11069-024-06724-9*.
[39] Briones Zúñiga, J. L., & Arancel Llerena, F. F. (2024). Deep learning algorithms to forecast the magnitude of earthquakes in Peru. *DOI:* https://doi.org/10.18687/LACCEI2023.1.1.168.
[40] Li, W., Chakraborty, M., Köhler, J., Quinteros-Cartaya, C., Rümpker, G., & Srivastava, N. (2024). Earthquake monitoring using deep learning with a case study of the Kahramanmaras Turkey earthquake aftershock sequence. *Solid Earth, 15*. https://doi.org/10.5194/se-15-197-2024.
[41] D. Bahdanau, K. Cho, and Y. Bengio, ‘‘Neural machine translation by jointly learning to align and translate,’’ 2014, arXiv:1409.0473. [Online].
Available: http://arxiv.org/abs/1409.0473
[42] (2020). Search Earthquake Catalog. Accessed: May 11, 2020. [Online].Available: https://earthquake.usgs.gov/earthquakes/search/
[43] H. F. Reid, ‘‘The elastic-rebound theory of earthquakes,’’ Bull. Dept. Geol., Univ. California Publication, Washington, DC, USA, Tech. Rep., 1911, pp. 413–444, vol. 6, no. 19.
[44] B. D. Fulcher and N. S. Jones, ‘‘Hctsa: A computational framework for automated time-series phenotyping using massive feature extraction,’’ Cell Syst., vol. 5, 2017, doi: 10.1016/j.cels.2017.10.001.
[45] T. Huijskens. (2017). Mutual Information-4 Feature Selection.
Accessed: Oct. 12, 2020. [Online]. Available: https://thuijskens.github.dio/2017/10/07/feature-selection/
[46] S. K. Gajawada. (2019). Anova for Feature Selection in Machine Learning.
Accessed: Oct. 12, 2020. [Online]. Available: https://towardsdatascience. com/anova-for-feature-selection-in-machine-learning-d9305e228476
[47] D. Dutta. (2016). How to Perform Feature Selection (i.e. Pick Important
Variables) Using Boruta Package in R? Accessed: Oct. 12, 2020. Available: https://www.analyticsvidhya.com/blog/2016/03/select- important-variables-boruta-package.
[48] J. Campbell, Map Use and Analysis. Dubuque, IA, USA: Wm C Brown
Dubuque, 1993.

Authors
1. Mohammed A. Jaleel Shaneen

Department of Computer Science, University of Technology, Baghdad, Iraq.
cs.22.16@grad.uotechnology.edu.iq
ORCID https://orcid.org/0009-0008-6946-9444
Born in 1993, holds a Master's degree in Artificial Intelligence in 2019 and is now a PhD student in Artificial Intelligence in the Department of Computer Science at the University of Technology in Baghdad, Iraq.

Journal of Information Systems Engineering and Management
2025, 10(54s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2. Dr. Suhad Malallah Kadhem[2]

Department of Computer Science, University of Technology, Baghdad, Iraq.
suhad.m.kadhem@uotechnology.edu.iq
ORCID https://orcid.org/0000-0003-0471-8231

Suhad Malallah Kadhem earned his Ph.D. in Computer Science from the Department of Computer Science at the Technology University. Suhad earned his bachelor's and master's degree in Computer Science from the University of Technology (UOT), Baghdad, Iraq in 1997. Suhad is a faculty member in the Computer Science Department at UOT since 1997,where she became a Head of artificial intelligent branch at UOT in 2003. Her research interested focus on artificial intelligence, natural language processing (especially Arabic language processing) and computer security (especially steganography).