Hybrid CNBLA architecture for accurate earthquake magnitude forecasting

Somia A. Shams, Asmaa Mohamed, Abeer S. Desuky, Gaber A. Elsharawy, Rania Salah El-Sayed
Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt

Article Info

Article history:

Received Oct 28, 2024
Revised Jul 17, 2025
Accepted Sep 14, 2025

Keywords:

Attention mechanism
Bidirectional long short-term memory
Earthquake prediction
Layer normalization
Neural network

ABSTRACT

Earthquake prediction in seismology is challenging due to sudden events and lack of warnings, requiring rapid detection and accurate parameter estimation for real-time applications. This study proposed a novel automatic earthquake detection model to enhance the processing and analysis of seismic data. The hybrid model comprises convolutional layers, normalization techniques, bidirectional long short-term memory (Bi-LSTM) networks, and attention mechanisms, collectively referred to as the hybrid convolutional–normalization–BiLSTM–attention (CNBLA) model. The attention mechanism allows the model to focus on critical segments of seismic sequences, while layer normalization stabilizes training by normalizing activations, thus reducing the effects of input scale variations. This dual approach mitigates the impact of input scale variations and enhances the model's ability to effectively decode complex temporal patterns. The hybrid CNBLA model optimizes the extraction and processing of temporal features from raw waveforms recorded at single stations, thereby improving the accuracy and efficiency of seismic magnitude estimation. The proposed model is evaluated using two datasets: the STEAD and USGS achieving a mean square error (MSE) values 0.054 and 0.0843 and a mean absolute error (MAE) 0.15 and 0.2526 respectively. The hybrid CNBLA model outperforms two baseline models and five state-of-the-art approaches in earthquake magnitude estimation, improving seismic monitoring and early warning systems.

Corresponding Author:

Somia A. Shams
Department of Mathematics, Faculty of Science, Al-Azhar University
Nasr City, Cairo, 11754, Egypt
Email: somia.shams10@azhar.edu.eg

1. INTRODUCTION

Earthquakes occur when the ground moves and shakes, releasing energy stored in rocks. These natural disasters can cause significant damage, financial loss, and injury, and in severe cases, mass casualties [1]. They can cause mass death. Earthquake early warning (EEW) has emerged as an effective technology for mitigating the impact and damage caused by earthquakes. Nine countries or regions currently have operational EEW systems; another thirteen are testing such systems [2]. For as little as a few seconds, EEW systems can alert target locations before the arrival of harmful seismic waves, which is essential for lessening the impact of earthquakes [3]. The regional EEW system provides early warning by determining the properties of an earthquake, including its time, position, and magnitude. Since the earthquake's magnitude plays a crucial role in a warning signal, rapid and precise magnitude estimation is a crucial field of research for regional EEW systems [2].

Magnitude is a logarithmic indicator of earthquake power. The Richter scale, often known as the local magnitude scale, was developed by Richter [4] and has been extensively used in scientific study and for providing quick public earthquake information. People usually feel earthquakes greater than 2.5 but cannot feel earthquakes less than 2.5. Earthquakes that cause significant damage have magnitudes greater than 4.5 [5]. Scientists strive continually to improve methods for predicting earthquake magnitudes to accurately prevent serious consequences. Although it is impossible to prevent earthquakes, timely and accurate warnings can significantly reduce their destructive impact. Various methods exist for predicting earthquake magnitudes, including the use of sensors, devices, magnetic and electrical waves, and seismic indicators derived from historical data. Although no ideal model exists, ongoing trials enhance prediction accuracy [6].

Artificial intelligence, particularly neural networks [7] has shown great promise in solving complex nonlinear problems related to seismic activity. Neural networks, a subset of machine learning, can model intricate patterns and relationships in data, making them well suited for earthquake magnitude prediction [1]. By analyzing vast amounts of seismological data, these models can identify subtle features that traditional methods may overlook. Machine learning and data mining techniques offer robust methods for studying seismic data and indicators, making them effective for handling large datasets [8]. These technologies have revolutionized the field of seismology, providing new insights and improving the accuracy of earthquake predictions. Ensuring the quality and accuracy of the dataset is critical for the performance of predictive models [5].

Deep learning models perform better when interpreting complicated and nonlinear inputs using these layers for dimensionality reduction. Deep learning models like graph neural networks (GNN) [9], multilayer perceptrons [10], long short-term memory (LSTM) [11], Bi-LSTM [12], provide effective approaches for capturing geographical data, including stations and their relationships. Chakraborty *et al.* [13] presented a multitasking deep learning model called the convolutional recurrent model for earthquake identification and magnitude estimation (CREIME). It can perform the following tasks: i) identify the earthquake signal from background seismic noise, ii) calculate the arrival time of the first primary wave (P wave), and iii) estimate the magnitude using the raw three-component waveforms from a single station as the model input. Biases in performance evaluation may arise from variations in preprocessing techniques and input data length when comparing CREIME with other models in the study. This disparity indicates that uniform benchmarks are necessary to guarantee equitable comparisons across various approaches.

Saad *et al.* [14] introduced a model comprising two specialized vision transformer (ViT) networks: one for identifying P-wave arrival times and another for predicting earthquake magnitudes, both engineered to process seismic data fast and reliably. A wider range of performance measurements would be beneficial for the paper, even though the evaluation metrics employed like mean absolute error (MAE) are significant. Mousavi and Beroza [15] used convolutional and recurrent neural networks, namely bidirectional long short-term memory (Bi-LSTM) units, to efficiently predict the correlations between seismic wave amplitudes and magnitudes. The transformer technique was utilized to forecast earthquake magnitudes based on existing data for the Horn of Africa [16]. Several studies have examined the use of various machine learning and deep learning models for earthquake prediction that are summarized in Table 1.

The results of this research have significant consequences for both the scientific community and public safety. The proposed model is built upon six key contributions:

a. Integration of attention mechanism with bidirectional LSTM: This serves as the baseline model for earthquake prediction, leveraging the strengths of both approaches.
b. Enhanced prediction through layer normalization: By replacing the attention mechanism with layer normalization (LN), the study demonstrates the effectiveness of this approach in models that do not utilize attention.
c. Development of the hybrid convolutional–normalization–BiLSTM–attention (CNBLA) model: This model effectively combines the advantages of the attention mechanism and layer normalization, which enhances the stability of the training process.
d. The custom loss function is crafted to allow the model to learn both accurate predictions and the associated uncertainty, specifically addressing aleatoric uncertainty, which refers to the uncertainty inherent in the data itself.
e. Comparative analysis of architectures: The research includes a detailed comparison of various architectures, accompanied by an in-depth discussion of the results obtained.
f. The efficiency of the proposed model is thoroughly evaluated using two different datasets and several performance metrics, including mean square error (MSE), mean absolute error (MAE), standard deviation of mean absolute error (MAE_STD), standard deviation of mean square error (MSE_STD), and mean combination error (MCE).

Improved magnitude estimation can enhance the reliability of EEW systems by providing critical seconds for individuals and infrastructures to take protective actions. Our findings will contribute to ongoing efforts to develop more robust and reliable early earthquake warning systems. It ultimately aiming to reduce the impact of these natural disasters on society.

The structure of the paper is organized as follows: The second section provides the necessary preliminaries, including foundational concepts and definitions pertinent to this study. Section 3 presents the architecture of the proposed Hybrid CNBLA model along with the configuration of the baseline model. Section 4 presents the results and discussion, beginning by describing the dataset used and assessing the model's performance. In Section 5, the work is concluded with a summary of the significant contributions and recommendations for future research directions.

Table 1. Comparison of existing earthquake forecasting models

Ref.	Techniques	Predicted variables	Range of magnitude	The type of prediction
[16]	Transformer algorithm	Magnitude	Magnitudes >= 3.	Regression
[17]	LSTM, GRU	Magnitude occurrence, location cluster, and time	Magnitude > 5.0	Clustering and regression
[18]	attention-based LSTM	time, magnitude, and location	magnitude > 5	Regression
[19]	Autoregressive integrated moving average (ARIMA) singular spectrum analysis (SSA)	Magnitude	Magnitude > 4	regression
[20]	attention and Bi-LSTM	Earthquake or no earthquake occurrence location occurrence (regression)	Magnitudes between 7 and 7.5	Classification
[21]	GNN with batch normalization and an attention mechanism	depth and magnitude	undefined	regression

2. PRELIMINARIES

2.1. Convolutional neural network

Convolutional neural networks (CNNs) have demonstrated efficacy in various domains, such as image processing, condition monitoring, and time series analysis. A CNN is constructed sequentially, layering three primary components: convolution, pooling, and fully connected (FC) layers [22]. The convolution layers comprise a collection of trainable kernels that are specifically designed to automatically extract local features from the input matrix [23]. These kernels execute convolution operations by utilizing weight sharing and local connection principles, resulting in reduced computational load, decreased model complexity, and improved performance. CNNs have been applied to earthquake prediction by analyzing seismic data, such as waveform signals and spectrograms. CNNs can learn geographical and temporal patterns in seismic data, enabling the detection of earthquake precursors or anomalies in the signals. CNNs, also known as feature learners, can automatically extract relevant features from raw input data [24].

2.2. Long short-term memory

Long short-term memory (LSTM) is a recurrent neural network that retains temporal connections between input items during training. They are widely used to simulate sequential data, such as earthquake signals [25]. LSTM units are effective for magnitude estimation due to their gated mechanism, which includes Tanh and Sigmoid activation functions, making them less sensitive to unnormalized input. The proposed LSTM architecture is illustrated in Figure 1. The LSTM unit comprises a cell: a forget gate, output gate, and input gate. The cell unit is responsible for storing values at each time interval. The gates control information that enters and leaves the rest of the unit. The forget gate (Γ_f) in the memory block structure is managed by a basic one-layer neural network. Equation (1) expresses how this gate operates. Forget gate(Γ_f) determines how much of the section should be retained and what should be discarded. The sigmoid activation function (σ) is used to implement it.

$$\Gamma_f = \sigma(w_f[a^{<t-1>}, x^{<t>}] + b_f \qquad (1)$$

where ($x^{<t>}$) is the current input, ($a^{<t-1>}$) is the previous hidden state, and (w_f and b_f) are the weight matrix and bias vector, which are learned from the input training data. An input gate is a unit in which the previous memory block effect forms new memory. A simple NN with an activation function is called tanh. These operations are calculated by (2), which calculates the candidate cell state, and (3), which calculates how much of the new cell state is retained using the sigmoid function (σ).

$$\tilde{c}^{<t>} = tanh\,(w_c[a^{<t-1>}, x^{<t>}] + b_c) \qquad (2)$$

$$\Gamma_u = \sigma(w_u[a^{<t-1>}, x^{<t>}] + b_u) \tag{3}$$

Output gate (Γ_o): It controls what parts of the cell state will be output, and it is described by (4).

$$\Gamma_o = \sigma(w_o[a^{<t-1>}, x^{<t>}] + b_o) \tag{4}$$

The output gate is the output of the current LSTM block and expressed using (5) and (6).

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \tag{5}$$

$$a^{<t>} = \Gamma_o * tanhc^{<t>} \tag{6}$$

Where ($a^{<t>}$) can be calculated by applying the output gate to the hyperbolic tangent of the cell state.

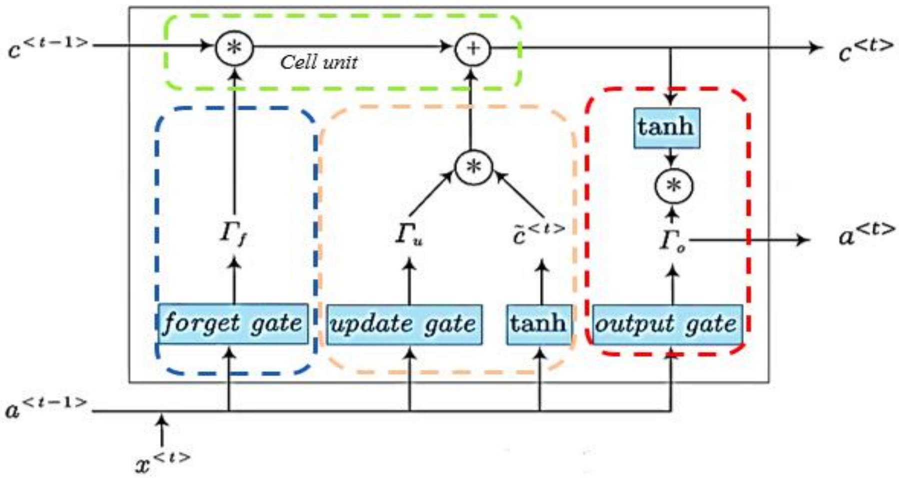

Figure 1. The architecture of LSTM

2.3. Bidirectional long-short term memory

Typically, an individual LSTM only functions in the forward direction of the information value. As a result, there was only one way to deliver the information. Two LSTM layers work together in the Bi-LSTM architecture [12], one layer handling forward information processing, and the second layer handling backward execution, as illustrated in Figure 2. This architecture is superior to single LSTM and RNN algorithms regarding earthquake magnitude prediction due to its ability to use previous and subsequent information.

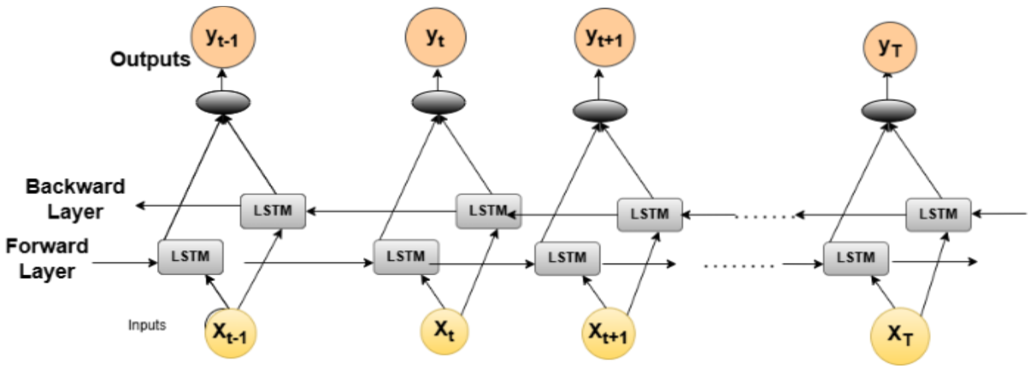

Figure 2. Bidirectional LSTM

2.4. Attention mechanism

Attention mechanism (ATT) plays a crucial role in sequence-to-sequence models, particularly when dealing with long sequences or complex patterns. The attention mechanism allows the model to focus on relevant parts of the input sequence, regardless of its length [26]. It assigns different weights to different time steps to emphasize important information. Note that attention improves model performance by reducing information loss risk. Instead of relying solely on the final hidden state of the LSTM, this model considers all hidden states in each decoding step.

3. RESEARCH METHOD

3.1. Architecture of the proposed hybrid CNBLA model

In the pursuit of advancing the computational processing and analysis of seismic data, this paper introduces a hybrid CNBLA model that integrates the strengths of convolutional layers, Bi-LSTM, and ATT with innovative regularization techniques. The Hybrid CNBLA model amalgamates the key features of the two preliminary models as follows. The first Multi-CNN-Bi-LSTM-ATT model leverages an ATT to selectively emphasize significant segments of input sequences, thereby enhancing the model's ability to handle complex patterns over long durations. In contrast, the second Multi-CNN-LN-Bi-LSTM model incorporates layer normalization (LN) [27] to stabilize the training process and mitigate the impact of input scale variations. By combining these approaches, the proposed hybrid CNBLA model aims to harness the robustness of layer normalization and the precision of attention-based mechanisms to deliver superior performance when analyzing three-channel seismogram data. This integration was designed to optimize the extraction and processing of temporal features, thereby improving the accuracy and efficiency of seismic magnitude estimation. The hybrid CNBLA architecture is shown in Figure 3 provided a comprehensive framework capable of addressing the challenges presented by the input data, leading to a more accurate and reliable prediction model.

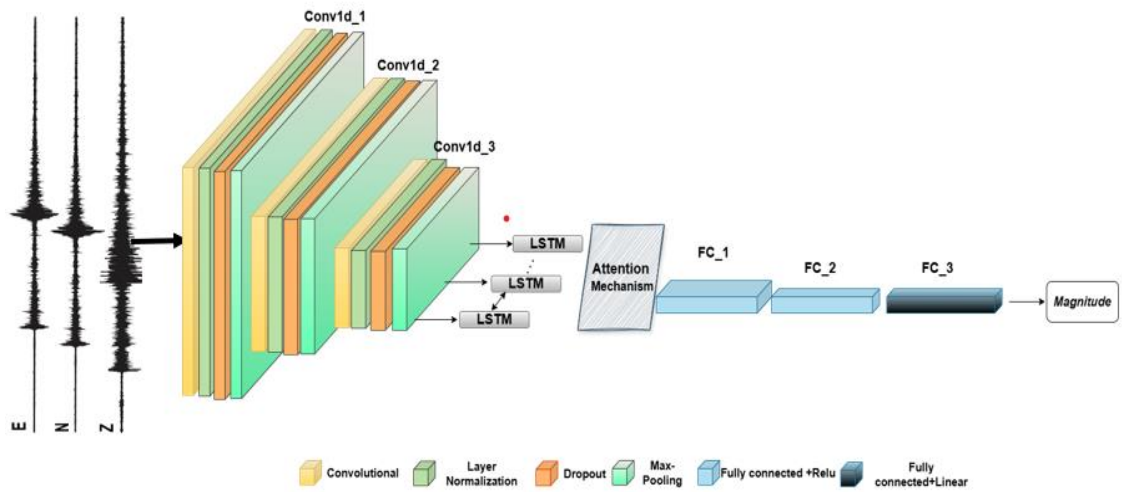

Figure 3. Architecture of hybrid CNBLA model for earthquake prediction

The proposed model used seismograms from three channels, each covering 30 s, to predict earthquake magnitudes. The architecture comprises three convolutional layers with filter counts of 32, 64, and 32, each with a kernel size of 3. Layer normalization was used to improve stability and efficiency during training, and dropout and max pooling were used to decrease spatial dimensions while preserving essential features.

The output of the CNN was used as Input for the Bi-LSTM architecture, which outperformed a single LSTM in predicting earthquake magnitudes. Attention mechanism is added after Bi-LSTM to evaluate all hidden states at each decoding step. The model incorporates two fully connected Dense layers, with the initial layer consisting of 64 units and L2 regularization to minimize overfitting. Dropout regularization is implemented to enhance model generalizability. The final output is created using a fully connected layer with a single neuron, and a linear activation function is used to estimate the output amplitude.

To enable the proposed model to learn both accurate predictions and their inherent aleatoric uncertainty (data noise), we use a custom loss function (7). This loss function achieves this by combining two key components: the weighted squared error and a direct penalty on the predicted uncertainty. This method provided a variant of the common MSE loss but incorporated an additional scaling factor that was dependent on the exponential function of variable s_i. The parameter s_i is extracted from the $y_{hat,i}$ tensor, which represents the model's predictions. The $y_{hat,i}$ tensor is expected to have at least two components along its last axis: the predicted value $y_{hat,i}$ and the secondary parameter s_i. The s_i parameter is used in the custom loss function to model aleatoric uncertainty, which represents the inherent noise or variability in the data. Specifically, s_i influences the weighting of the squared error term and adds a regularization-like term to the loss. The exponential transformation $0.5.e^{s_i}$ adjusts the contribution of the squared error based on the value of s_i.

This loss function is particularly interesting because it allows for dynamic adjustment of the error term based on the value of s_i.

$$L_i = \sum(0.5.e^{s_i}.\left(|y_{true,i} - y_{hat,i}|\right)^2 + 0.5.s_i) \tag{7}$$

Where L_i is the loss of element i in the batch, and s_i is a parameter that modulates the squared error's influence on the total loss. The exponential function ensures that this scaling factor decreases as s_i increases, which can be interpreted as reducing the importance or weight of large errors when s_i is high.

We used an adjustable learning rate technique in our implementation. The proposed model adaptively modifies the learning rate during training, thereby enabling more effective convergence. We calculated the learning rate using (8):

$$lr_t = lr_{t-1} * e^{-\lambda} \tag{8}$$

Where λ is the decay rate, which in this case is set to the square root of 0.1. The value of lambda was determined through extensive experimentation with different configurations. We tested a variety of lambda values to find the optimal setting that balanced convergence speed and stability during training. Our results indicated that a lambda value of 0.1 consistently produced favorable results across multiple training runs, contributing to a more stable learning process. This choice reflects our findings that this particular value allowed the model to converge effectively without oscillations or divergence. The patience parameter was set to 4 epochs, which indicates that the learning rate decreased if no progress was detected during this time frame. The absence of a cooldown period (cooldown=0) permitted the immediate resumption of post-reduction. This adaptive methodology enhances training by achieving a harmonious equilibrium between swift initial learning and meticulous refinement in subsequent phases. This may result in enhanced model performance and expedited convergence. This dual approach leverages the benefits of both normalization and attention mechanisms to ensure robustness and superior performance in sequence processing tasks. All parameters are shown in Tables 2 and 3.

Table 2. Description of proposed model and its parameters

Block	Layers (Name)	Description and parameter values
feature extraction block	conv1d	1D conv layer with 32 filters, kernel size of 3 and L2 regularization = 0.001
		layer_normalization
		Dropout rate of 0.2
		max_pooling1d with a pool size of 4
	conv1d_1	1D conv layer with 64 filters, kernel size of 3 and L2 regularization = 0.001
		layer_normalization_1
		Dropout rate of 0.2
		max_pooling1d with a pool size of 4
	conv1d_2	1D conv layer with 32 filters, kernel size 3, and L2regularization = 0.001
		layer_normalization_2
		Dropout rate of 0.2
		max_pooling1d with a pool size of 4
Sequence learning	Bi-LSTM	The LSTM layer has 100 units, recurrent_dropout=0.2, and dropout regularization rate is 0.2.
Attention block	dense_1	A dense layer with a single output neuron is applied to the input using the tanh activation function.
		Attention
Prediction block	dense_2	Dense layer with 64 units, L2 regularization = 0.001, and the ReLU activation function was applied.
		layer_normalization_3
		Dropout rate of 0.2
	dense_3	Dense layer with 64 units, L2 regularization = 0.001, and ReLU activation function.
		layer_normalization_4
		Dropout rate of 0.2
	dense_4	A dense layer is applied with a linear activation function.

Table 3. Parameter settings for the hybrid CNBLA model

Parameter	Value
No. of training samples	50734
No. of validation samples	7248
No. of testing samples	14495
No. of epochs	200
Batch size	256
L2 regularization strength	0.001
Loss function	The custom loss function as in (7)
optimizer	Adam algorithm
Learning rate	Automatically adjusts the learning rate during training, as shown in (8)
Early stopping	monitor=validation loss, and patience = 5

3.2. Baseline model configuration
3.2.1. Multi CNN-Bi-LSTM-ATT model
The proposed neural network architecture integrates convolutional layers, Bi-LSTM, and an attention mechanism. Following the attention mechanism, the architecture includes three fully-connected dense layers to predict the output magnitude. The attention mechanism helps the model focus on pertinent input, which improves its capacity to decode complex patterns over lengthy periods of time. This selective attention enhances model performance, particularly on sequence-to-sequence tasks.

3.2.2. Multi CNN-LN-Bi-LSTM model
This model modifies the Multi-CNN-Bi-LSTM-ATT model by substituting the attention mechanism with layer normalization (LN) to demonstrate its effectiveness in a model without attention. LN is applied after each learnable layer, such as CNN and fully connected layers. LN facilitates training by normalizing the activations within each layer, thereby stabilizing the learning process and making it less sensitive to the scale of the input features; however, it does not support capturing long-term dependencies or improving the emphasis on critical inputs. Its primary job in this situation is to improve stability and training effectiveness; however, it does not provide any additional input weighting capabilities, as attention mechanisms do.

4. RESULTS AND DISCUSSION
4.1. Datasets description
4.1.1. Stanford earthquake dataset (STEAD) datasets
The STEAD dataset [5] is a large-scale, global dataset containing two classes of waveforms: Seismic noise and local earthquake waveforms, which are recorded at local distances (within 350 km of earthquakes). STEAD includes approximately 1.2 million waveforms recorded by seismometers located worldwide and resampled at 100 Hz, with a duration of 60 s (6000 features). The local earthquake category comprises approximately 1,050,000 three-component seismograms linked to 450,000 earthquakes that occurred between January 1984 and August 2018. The seismic noise class comprises approximately 100,000 waveforms recorded in the United States and Europe since 2000. We require seismic waveforms from continuous time series stored in the archives of the earthquake data management center (IRIS DMC), which is a collaboration of many research organizations. There are three categories of access states: manual selections, which human analysts choose; automatic selections, which are determined by automatic algorithms; and automatic pickers, which are selected using an AI-based model. The STEAD dataset comprises separate arrays with three waveforms

Representing three-component seismograms. Each waveform has 6000 characteristics. To prepare for training, the proposed model does not use all the data from the STEAD dataset. It carefully selects a smaller portion based on specific rules. These rules are designed to ensure data quality and relevance to earthquakes. One example of these rules is selecting only data inputs labeled "trace_category" as "earthquake_local". The columns *p_travel_seconds*, *source_distance_km*, *source_magnitude*, and *source_depth_km* are not empty. We used approximately 300,000 earthquake waveforms recorded at less than 1-degree epicentral distances. The entire waveform (from 1 second before P to end of S coda) was equal to or less than 30 s. The magnitude distribution of the occurrences is shown in Figure 4. All waveforms were band-pass filtered between 1.0 to 40.0 Hz with a signal-to-noise ratio greater than 20 db. To test the impact of factors such as magnitude type, side effects, regional effects, and site-dependent learning, we divided the data into smaller subsets ranging from 60K to 140K. We used 70% of each subset for training and 10% and 20% for validation and testing regressivity.

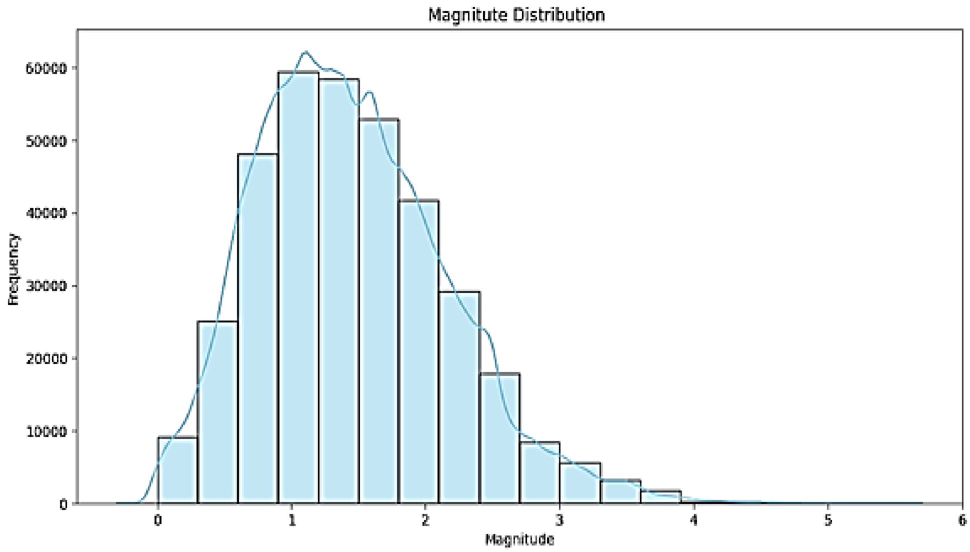

Figure 4. Magnitude distribution

4.1.2. USGS dataset

The United States Geological Survey (USGS) provides real-world earthquake data on historical earthquakes. The dataset used for this research from *https://www.kaggle.com/datasets/warcoder/earthquake-dataset* which contains earthquake data collected from the USGS website by Kaggle. This dataset includes a record of title, magnitude, date, time, intensity, maximum estimated instrumental intensity, tsunami, The total number of seismic stations, The largest azimuthal gap between azimuthally adjacent stations, depth, latitude, longitude and country of every earthquake.

4.2. Performance evaluation

We evaluated the performance of our models using the mean square error (MSE) [28], MAE, mean absolute error standard deviation (MAE_STD), mean square error standard deviation (MSE_STD), and mean combination error (MCE). Let n be the number of samples, and the values y_1, y_2, y_3, …, y_n be samples ones that were observed in the dataset, and let the values $\hat{y}_1$, $\hat{y}_2$, $\hat{y}_3$, …, $\hat{y}$ the ones that were predicted by the model. MSE, MAE_STD, E_STD, and (MCE) can be calculated using (9)-(13).

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

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{10}$$

$$MAE\_STD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i| - MAE)^2} \tag{11}$$

$$MSE\_STD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left(y_i - \hat{y}_i\right)^2 - MSE\right)^2} \tag{12}$$

$$MCE = \propto * MAE + (1 - \propto) * \sqrt{MSE} \tag{13}$$

where: α is a weighting factor between 0 and 1

4.3. Experimental results and discussion

All tests were performed using Google's Kaggle, the Keras framework, which can operate on top of TensorFlow, and the Python programing language. All experiments were conducted on an NVIDIA Tesla K80 graphical processing unit (GPU) with 32 GB RAM. Furthermore, the Windows 10 operating system was used.

This section presents three adapted prediction models. The first model is the Multi CNN-Bi-LSTM-ATT, which incorporates an attention mechanism. The second model is the Multi-CNN-LN-Bi-LSTM, which utilizes layer normalization. Finally, the hybrid CNBLA model integrates the strengths of the two previous approaches. Several experiments were carried out to explore the hyperparameter values of the proposed model. The parameter values that achieved the highest performance in this study are summarized in Table 3.

4.3.1. Multi CNN-Bi-LSTM-ATT model for earthquake prediction

In this section, we assess the performance of the Multi CNN-Bi-LSTM-ATT model for earthquake prediction. Figure 5 offers a compelling visual overview of the model's effectiveness, demonstrating its superior accuracy, consistency, and efficiency in predicting earthquake magnitudes. The figure features a line graph that depicts the training and validation loss of the proposed model across several epochs. The x-axis represents the number of epochs, while the y-axis displays the corresponding loss values. A lower loss value indicates improved model performance. The blue line illustrates the training loss, which begins at a high value in the early epochs and decreases rapidly. As training progresses, the loss continues to decline, albeit at a slower rate.

It was found that the model learned effectively from the training set after approximately 80 epochs. In the testing phase, Multi CNN-Bi-LSTM-ATT achieved MAE=0.19, MAS_STD=0.21, MSE=0.083, MSE_STD=0.22, and MCE=0.03. Figure 6 illustrates the relationship between the measured and predicted magnitudes. The horizontal axis represents the measured magnitude documented in the earthquake event catalog measurements, while the vertical axis represents the predicted magnitude produced by the model. The computed linear regression line is displayed by the diagonal line at the center of the plot. Thus, we can deduce that the CNN-Bi-LSTM-ATT model demonstrates a satisfactory fit for the data because the data point is close to the theoretical regression line.

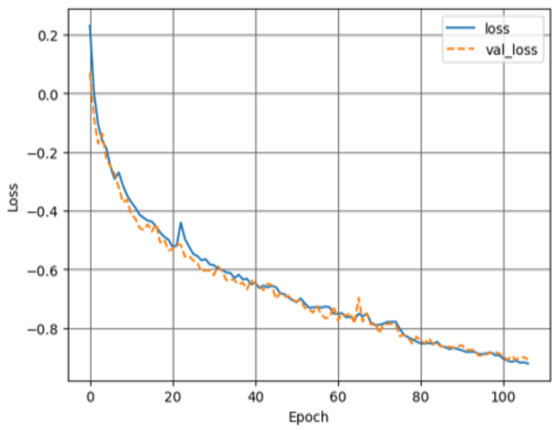

Figure 5. Training of the Multi-CNN-Bi-LSTM-ATT model and validation loss over a no. of epochs

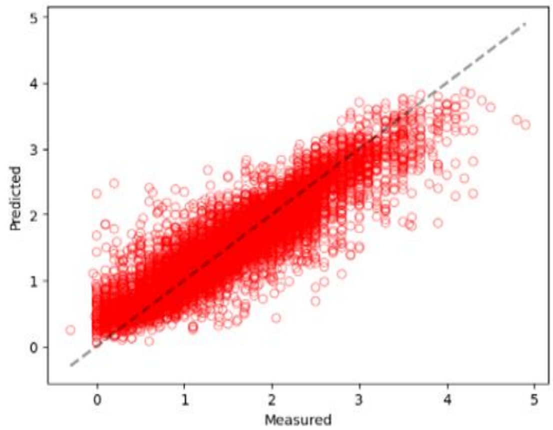

Figure 6. Measured and predicted magnitudes of the Multi CNN-Bi-LSTM-ATT model

4.3.2. Multi-CNN-LN-Bi-LSTM model for enhanced earthquake prediction

The next step investigates the effect of incorporating a normalization layer to enhance the performance of earthquake prediction. Figure 7 demonstrates that the model has learned effectively from the training set. Around epoch 35, the training loss plateaued, while the validation loss continued to fluctuate. This behavior indicates that the model is successfully learning from the data during the training process. The Multi-CNN-LN-Bi-LSTM model achieved in the testing phase, MAE=0.16, MAS_STD=0.18, MSE=0.058, MSE_STD=0.17, and Mean combination error=0.017. Figure 8 illustrates the relationship between the measured and predicted magnitudes. The computed linear regression line is displayed by the diagonal line at the center of the plot. Thus, we can deduce that Multi- CNN-LN-Bi-LSTM model demonstrates a satisfactory fit for the data because the data point is close to the theoretical regression line.

4.3.3. Hybrid CNBLA model for improving earthquake prediction

A novel Hybrid CNBLA model is proposed, which combines the advantageous features of convolutional layers, Bi-LSTM, and ATT with creative regularization techniques. The combination of LN and ATT reduces the influence of input scale fluctuations, allowing the model to train successfully while not negatively influencing the data. This combination improves the model's robustness, allowing it to better tackle the problems posed by seismic data, resulting in higher generalization to previously unexplored data, as shown in the following result. The inclusion of both strategies enables more efficient learning, where LN supports faster convergence and AT ensures that the model learns from the most relevant input, resulting in faster weight modifications during training. Our hybrid CNBLA model aims to leverage the resilience of layer normalization and the accuracy of attention-based processes to achieve enhanced performance when analyzing three-channel seismogram data. To enhance the precision and effectiveness of seismic magnitude estimation, this integration was developed to optimize the extraction and processing of temporal information. The implementation of the hybrid CNBLA architecture depicted in Figure 9 offers a comprehensive framework that effectively tackles the issues posed by the input data, resulting in a prediction model that is both more precise and reliable. Figure 10 Illustrates the relationship between the measured and predicted magnitudes. The computed linear regression line is displayed by the diagonal line at the center of the plot. Thus, we can deduce that the Hybrid CNBLA model exhibits a satisfactory fit for the data because the data point is close to the theoretical regression line. In the testing phase, the hybrid proposed model outperformed the others, with the lowest MAE=0.15, MAS_STD=0.17, MSE=0.054, MSE_STD=0.15, and MCE=0.015.

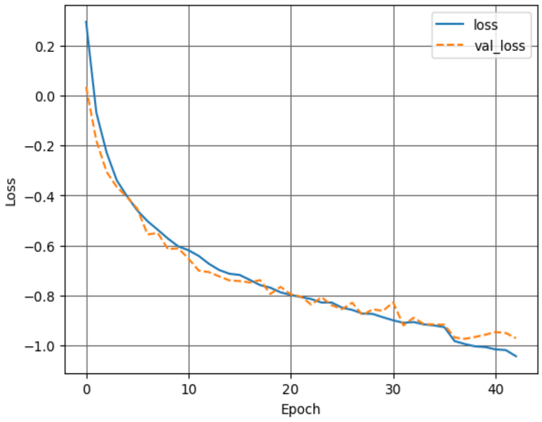

Figure 7. Multi-CNN-LN-Bi-LSTM 's training and validation loss over a no. of epochs

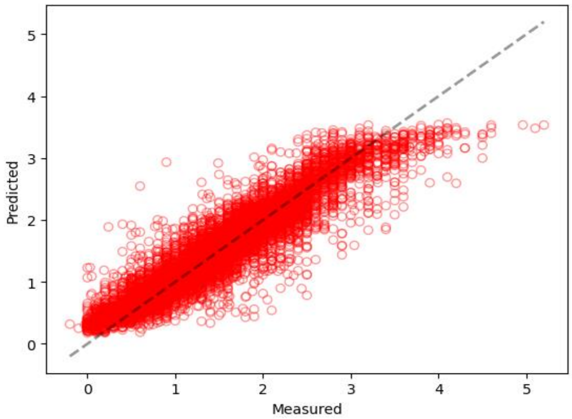

Figure 8. Measured and predicted magnitude estimation by Multi-CNN-LN-Bi-LSTM

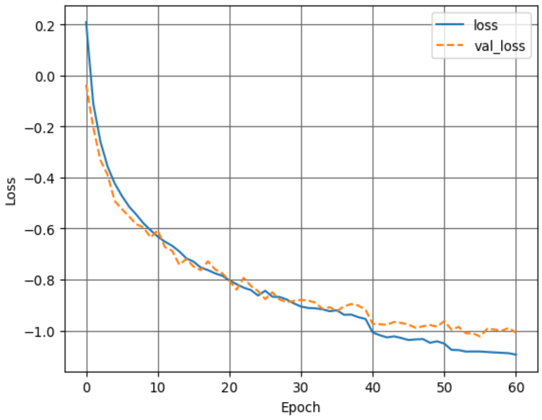

Figure 9. Hybrid CNBLA training and validation loss over a no. of epochs

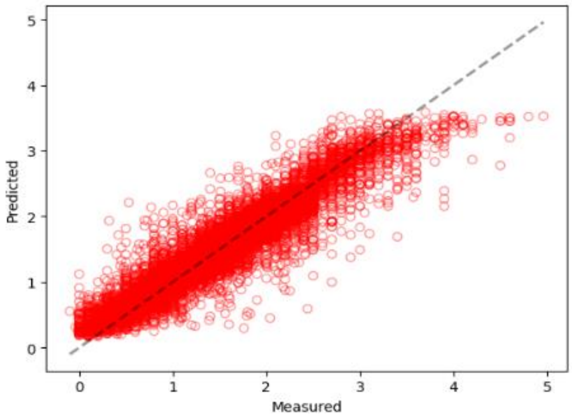

Figure 10. Measured and predicted magnitude estimation of Hybrid CNBLA

4.4. Comparative study and discussion

Table 4 presented a comparative analysis of various model architectures on the STEAD dataset during testing phase. The comparison of the proposed model and the baseline models using the performance metrics are summarized in Table 4.

Table 4. Comparative results of different models for proposed model on STAND dataset

Dataset	Model Architecture	MAE	MAE_STD	MSE	MSE_STD	MCE
STEAD	LSTM [16]	0.2234	0.2304	0.103	0.2599	1.0
	Bi-LSTM [16]	0.222	0.2226	0.099	0.231	1.0
	Bi-LSTM-attention [16]	0.224	0.227	0.102	0.243	1.0
	Transformer [16]	0.20879	0.2088	0.0872	0.2095	1.0
	MagNet [15]	0.201	0.22	0.089	0.228	0.02
	Multi-CNN-Bi-LSTM-ATT (baseline_model)	0.19	0.21	0.083	0.226	0.03
	Multi-CNN-LN-Bi-LSTM (baseline_model)	0.16	0.18	0.058	0.17	0.017
	Hybrid CNBLA (Proposed Model)	0.15	0.17	0.054	0.15	0.015

The model of LSTM establishes the baseline with an MAE of 0.2234 and an MSE of 0.103. This model is followed by a bidirectional variant, Bi-LSTM, which achieves a slightly improved MAE of 0.222 and an MSE of 0.099. These results suggest a marginal benefit in capturing temporal dependencies with bidirectional processing, but the improvements are limited. Incorporating an attention mechanism into the Bi-LSTM model results in an MAE of 0.224 and MSE of 0.102, which are comparable to the standard Bi-LSTM. This indicates that the simple addition of attention alone does not substantially enhance the model's performance. The attention mechanism in this context may not effectively capture the long-term dependencies within the seismic sequences. Transformer architecture [16] significantly improves, achieving an MAE of 0.20879 and an MSE of 0.0872. The transformer's self-attention mechanism seems to better understand the relationships in the seismic data, leading to better performance compared to recurrent-based architectures. The model also demonstrates greater stability with lower standard deviations (MAE_STD: 0.2088, MSE_STD: 0.2095), highlighting the transformer's consistent predictive ability. The author in [15] designed a regressor (MagNet) combining convolutional and recurrent neural networks which increase architectural complexity to efficiently predict the correlations between seismic wave amplitudes and magnitudes. The model achieves an MAE of 0.20 and an MSE of 0.089 when using this architecture for the STEAD dataset.

The multi-CNN-Bi-LSTM-ATT architecture further improves the results, obtaining an MAE of 0.19 and an MSE of 0.083. This model has convolutional layers for extracting features, Bi-LSTM for recognizing temporal patterns, and an attention mechanism to pay attention to important inputs. The combination of these components enhances the model's ability to extract meaningful features from seismic data, as reflected in the reduced errors. The standard deviations for this architecture are 0.21 for MAE and 0.226 for MSE, indicating reasonable stability. The Multi-CNN-LN-Bi-LSTM model, which replaces the attention mechanism with layer normalization (LN), demonstrates a notable leap in performance. This model achieves an MAE of 0.16 and an MSE of 0.058, demonstrating the impact of LN in stabilizing the training process and improving both accuracy and consistency. The standard deviations of MAE and MSE further decrease to 0.18 and 0.17, respectively. The MCE for this model is also significantly lower at 0.017, underscoring the importance of layer normalization in this architecture.

The hybrid CNBLA proposed model, which integrates convolutional layers, Bi-LSTM, attention, and layer normalization, achieves the best overall performance. This model records an MAE of 0.15 and an MSE of 0.054, with the lowest standard deviations (MAE_STD: 0.17, MSE_STD: 0.15) and a remarkably low MCE of 0.015. When you combine layer normalization and attention mechanisms, they make training go faster and predictions more accurate. This lets the model work well with seismic data.

A clear trend emerges when examining the MCE across models. The basic architectures, including LSTM, Bi-LSTM, and transformer, exhibit relatively high MCEs around 1.0. However, the integration of more advanced components—such as attention mechanisms, convolutional layers, and layer normalization—leads to dramatic improvements. The Multi-CNN-Bi-LSTM-ATT model reduces the MCE to 0.03, while the Multi-CNN-LN-Bi-LSTM achieves 0.017. The hybrid CNBLA model further reduces the MCE to 0.015, demonstrating the effectiveness of hybrid architectures in capturing complex seismic patterns. Figure 11 illustrates the efficiency of the proposed hybrid model, which exhibits the lowest deviation metrics (MAE, MAE_STD, MSE, and MSE_STD) when evaluated with an unknown test dataset, in comparison to other state-of-the-art models.

Comparing the suggested model with two baseline models and five state-of-the-art techniques shows how effective it is at estimating earthquake magnitude. For routine seismic monitoring and early warning systems, our hybrid CNBLA model performs better, with MSE and MAE values of 0.054 and 0.15, respectively. According to this comparison investigation, better seismic data analysis performance is correlated with more complex architecture. When moving from simple LSTM models to more complex hybrid architectures, MAE, MSE, and MCE consistently improve; each change improves accuracy and stability. The hybrid CNBLA proposed model was also tested using USGS dataset. Table 5 presents a comparative study between the hybrid CNBLA proposed model and the other adopted ones using this dataset.

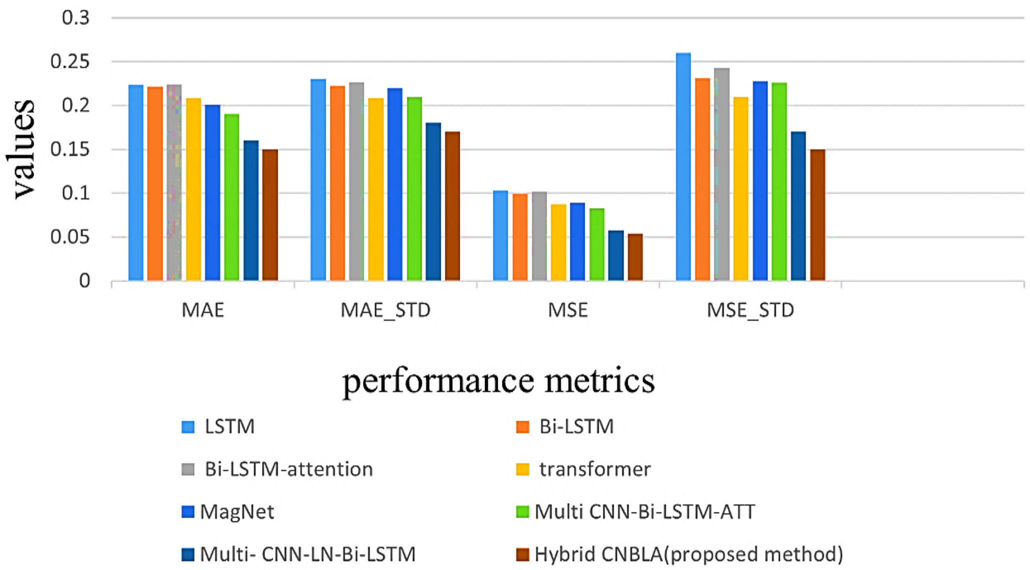

Figure 11. Comparison of proposed model and state-of-the-art on STEAD dataset

Table 5. Comparative results of different models for proposed model on USGS dataset

Dataset	Model architecture	MAE	MAE_STD	MSE	MSE_STD
USGS	Bi-LSTM [16]	0.277642	0.1728135	0.10694	0.1158977
	Bi-LSTM-attention [16]	0.2670741	0.1660221	0.0988919	0.1027874
	Transformer [16]	0.2582806	0.145861	0.082423	0.07721
	MagNet [15]	0.2626354	0.1502673	0.0915576	0.0811363
	Multi-CNN-Bi-LSTM-ATT (baseline_model)	0.2589762	0.1487923	0.0892078	0.0777827
	Multi-CNN-LN-Bi-LSTM (baseline_model)	0.2581094	0.1416009	0.0866713	0.077774
	Hybrid CNBLA (Proposed Model)	**0.2526728**	**0.143346**	**0.0843916**	**0.0759424**

In Table 5, the hybrid CNBLA model outperforms both the basic Bi-LSTM and its attention-enhanced version in terms of accuracy and consistency. While adding attention to Bi-LSTM reduces MAE by 3.7% and MSE by 7.5%, hybrid CNBLA further improves these metrics—achieving an additional 5.4% reduction in MAE and 14.7% in MSE. It also shows the lowest MSE standard deviation, indicating more stable performance. When comparing our hybrid CNBLA with the transformer and MagNet, two distinct tendencies become evident. The transformer, although proficient in collecting long-range dependencies through self-attention, produces a MAE of 0.2583 and a mean squared error (MSE) of 0.0824. In contrast, MagNet's convolution-driven design attains a somewhat elevated MAE of 0.2626 and an MSE of 0.0916. Our hybrid CNBLA demonstrates superior performance, achieving an MAE of 0.2527 and an MSE of 0.0844. It enhances the transformer by almost 2.2 percent in MAE and exceeds MagNet by over 3.8 percent in MAE and 7.9 percent in MSE. This suggests that although the transformer is proficient in modeling global temporal structures and MagNet in extracting localized waveform features, neither can independently capture the multi-scale patterns seen in seismic data as successfully as our integrated approach.

Secondly, in terms of stability throughout numerous iterations, the Transformer's MSE standard deviation of 0.0772 and MagNet's 0.0811 both surpass the Hybrid CNBLA's 0.0759. This indicates that the incorporation of convolutional feature extractors (as utilized in MagNet) and recurrent-attention algorithms (as employed in the Transformer) inside a unified, normalized framework results in reduced average errors and enhanced reliability in performance. This increased robustness is essential for earthquake forecasting: early-warning systems require accurate magnitude estimations and consistent model behavior under different initial conditions. The Hybrid CNBLA leverages the complementary strengths of CNNs, Bi-LSTMs, layer normalization, and attention, surpassing each particular baseline and providing a balanced solution unattainable by either pure Transformer or pure convolutional architectures alone. Figure 12 shows a comparative performance matrix of seven deep-learning architectures—Bi-LSTM, Bi-LSTM+Attention, Transformer, MagNet, Multi-CNN-Bi-LSTM+ATT, Multi-CNN-LN-Bi-LSTM, and the proposed Hybrid CNBLA—on the USGS earthquake dataset (2001–2023). Across all models, we report MAE, standard deviation of MAE (MAE_STD), mean squared error (MSE), and standard deviation of MSE (MSE_STD).

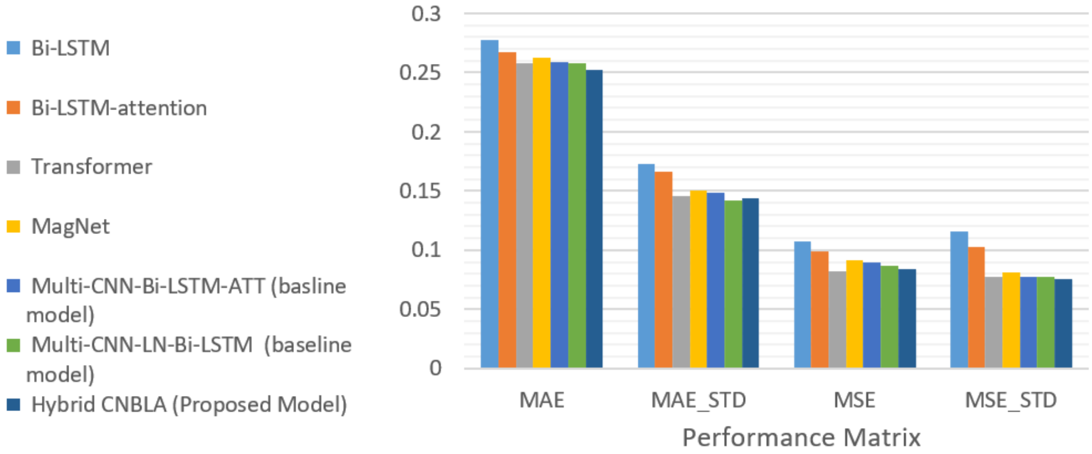

Figure 12. Comparison of proposed model and state-of-the-art on USGS dataset

The results clearly indicate that recurrent models enhanced with attention or convolutional preprocessing outperform the plain Bi-LSTM, but the most significant gains arise from the integrated Hybrid CNBLA design. Attention alone yields modest improvements by allowing the network to focus on salient temporal features, cutting MAE and MSE by approximately 3.7% and 7.5%, respectively. Convolutional layers in the multi-CNN hybrids contribute to local feature extraction, further reducing error and variability.

The Transformer's strong performance underscores the value of self-attention for capturing global dependencies, achieving the lowest variability in MAE. However, the Hybrid CNBLA combines the best of both worlds: convolutional preprocessing filters out noise and emphasizes local seismic patterns, bidirectional LSTMs model sequence context, layer normalization stabilizes training, and attention highlights critical time steps. This synergy results in the lowest average errors and the tightest error distributions (MAE_STD and MSE_STD), demonstrating both high accuracy and robustness.

4.5. Architectural considerations and scalability

Our proposed hybrid model, integrating (CNN, BiLSTM, Attention) stands out for its superior performance in intricate earthquake magnitude prediction. This advanced design is critical for discerning the subtle, long-range connections within extensive seismic data, a capability vital for precise earthquake precursor detection. While this powerful architecture introduces a computational footprint, notably a dominant $O(L^2)$ complexity from the attention mechanism for long sequences, this is a necessary trade-off for achieving unparalleled accuracy in such a critical application. Despite these inherent complexities, our model is engineered for substantial scalability through its robust implementation within the TensorFlow Keras framework. We employ a multifaceted strategy:

a. Implemented distributed training using *tf.distribute*. Strategy for enhancing parallel processing on GPUs/TPUs.
b. Enhanced data management utilizing the *tf.data* API for efficient batch processing and streaming, alleviating memory limitations.
c. Enhancement of deployment using quantization and pruning to reduce model size and improve inference speed for real-time applications.

This comprehensive method confirms that the model's intentional complexity is justified by performance criteria and that its design is inherently suitable for scaling, ensuring practical viability and extensibility for future large-scale seismological applications.

5. CONCLUSION

This paper presents a new automatic earthquake detection model that uses a hybrid neural network architecture to enhance the processing of seismic data. The proposed model efficiently improves the extraction and analysis of temporal characteristics from raw seismic waveforms by using convolutional layers, Bi-LSTM, and attention processes. The proposed model demonstrates superior performance compared to the other models in magnitude estimation when evaluated against two datasets: the STEAD and USGS datasets. It is evaluated using two datasets, resulting in MSE of 0.054 and 0.0843, and MAE values of 0.15 and 0.2526, respectively. The findings emphasize the capacity of sophisticated neural network structures to enhance the precision and effectiveness of earthquake magnitude forecasts. This strategy not only improves the regular monitoring of seismic activity and boosts earthquake's dependability early warning systems, providing crucial seconds for taking precautionary measures. Subsequent efforts will focus on enlarging the dataset and investigating supplementary improvements to neural networks to enhance earthquake prediction capabilities to a greater extent.

REFERENCES

[1] 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, pp. 471–486, Jan. 2017, doi: 10.1007/s11069-016-2579-3.

[2] G. Cremen and C. Galasso, "Earthquake early warning: Recent advances and perspectives," *Earth-Science Reviews*, vol. 205, Jun. 2020, doi: 10.1016/j.earscirev.2020.103184.

[3] J. A. Strauss and R. M. Allen, "Benefits and costs of earthquake early warning," *Seismological Research Letters*, vol. 87, no. 3, pp. 765–772, May 2016, doi: 10.1785/0220150149.

[4] C. F. Richter, "An instrumental earthquake magnitude scale," *Bulletin of the Seismological Society of America*, vol. 25, no. 1, pp. 1–32, 1935, doi: 10.1785/bssa0250010001.

[5] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "Stanford earthquake dataset (STEAD): A global data set of seismic signals for AI," *IEEE Access*, vol. 7, pp. 179464–179476, 2019, doi: 10.1109/ACCESS.2019.2947848.

[6] T. L. Chin, C. Y. Huang, S. H. Shen, Y. C. Tsai, Y. H. Hu, and Y. M. Wu, "Learn to detect: Improving the accuracy of earthquake detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8867–8878, Nov. 2019, doi: 10.1109/TGRS.2019.2923453.

[7] K. Gurney, *An introduction to neural networks*. CRC Press, 2014.

[8] R. Mallouhy, C. A. Jaoude, C. Guyeux, and A. Makhoul, "Major earthquake event prediction using various machine learning algorithms," in *6th International Conference on Information and Communication Technologies for Disaster Management, ICT-DM 2019*, Dec. 2019, doi: 10.1109/ICT-DM47966.2019.9032983.

[9] D. Bacciu, F. Errica, A. Micheli, and M. Podda, "A gentle introduction to deep learning for graphs," *Neural Networks*, vol. 129, pp. 203–221, Sep. 2020, doi: 10.1016/j.neunet.2020.06.006.

[10] M. P. A. van den Ende and J. P. Ampuero, "Automated seismic source characterization using deep graph neural networks," *Geophysical Research Letters*, vol. 47, no. 17, Sep. 2020, doi: 10.1029/2020GL088690.

[11] J. Schmidhuber, "Tutorial on LSTM recurrent nets," IDSIA, 2003.

[12] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A CNN-BiLSTM model with attention mechanism for earthquake prediction," *Journal of Supercomputing*, vol. 79, no. 17. pp. 19194–19226, 2023, doi: 10.1007/s11227-023-05369-y.

[13] M. Chakraborty *et al.*, "CREIME—A convolutional recurrent model for earthquake identification and magnitude estimation," *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 7, Jul. 2022, doi: 10.1029/2022JB024595.

[14] O. M. Saad, Y. Chen, A. Savvaidis, S. Fomel, and Y. Chen, "Real-time earthquake detection and magnitude estimation using vision transformer," *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 5, May 2022, doi: 10.1029/2021JB023657.

[15] S. M. Mousavi and G. C. Beroza, "A machine-learning approach for earthquake magnitude estimation," *Geophysical Research Letters*, vol. 47, no. 1, Jan. 2020, doi: 10.1029/2019GL085976.

[16] E. Abebe, H. Kebede, M. Kevin, and Z. Demissie, "Earthquakes magnitude prediction using deep learning for the Horn of Africa," *Soil Dynamics and Earthquake Engineering*, vol. 170, p. 107913, Jul. 2023, doi: 10.1016/j.soildyn.2023.107913.

[17] A. Berhich, F. Z. Belouadha, and M. I. Kabbaj, "A location-dependent earthquake prediction using recurrent neural network algorithms," *Soil Dynamics and Earthquake Engineering*, vol. 161, Oct. 2022, doi: 10.1016/j.soildyn.2022.107389.

[18] A. Berhich, F. Z. Belouadha, and M. I. Kabbaj, "An attention-based LSTM network for large earthquake prediction," *Soil Dynamics and Earthquake Engineering*, vol. 165, p. 107663, Feb. 2023, doi: 10.1016/j.soildyn.2022.107663.

[19] M. Merdasse, M. Hamdache, J. A. Peláez, J. Henares, and T. Medkour, "Earthquake magnitude and frequency forecasting in northeastern Algeria using time series analysis," *Applied Sciences (Switzerland)*, vol. 13, no. 3, Feb. 2023, doi: 10.3390/app13031566.

[20] M. H. Al Banna *et al.*, "Attention-based bi-directional long-short term memory network for earthquake prediction," *IEEE Access*, vol. 9, pp. 56589–56603, 2021, doi: 10.1109/ACCESS.2021.3071400.

[21] M. A. Bilal, Y. Ji, Y. Wang, M. P. Akhter, and M. Yaqub, "An early warning system for earthquake prediction from seismic data using batch normalized graph convolutional neural network with attention mechanism (BNGCNNATT)," *Sensors*, vol. 22, no. 17, Sep. 2022, doi: 10.3390/s22176482.

[22] O. M. El Zein and N. E. Ghannam, "An advanced approach for accurate pneumonia detection using combined deep convolutional neural networks," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 3094–3105, Jun. 2024, doi: 10.11591/ijece.v14i3.pp3094-3105.

[23] S. A. El-aal, R. S. El-Sayed, A. A. Alsulaiman, and M. A. Razek, "Using deep learning on retinal images to classify the severity of diabetic retinopathy," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7. pp. 346–355, 2024, doi: 10.14569/IJACSA.2024.0150734.

[24] A. W. Salehi *et al.*, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability (Switzerland)*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.

[25] S. M. Mousavi, W. Zhu, Y. Sheng, and G. C. Beroza, "CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection," *Scientific Reports*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-45748-1.

[26] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[27] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv:1607.06450*, 2016.

[28] S. A. El-aal, A. Deifalla, N. I. Ghali, and A. A. Naim, "A deep learning model for the punching shear strength of prestressed concrete slabs," *International Journal of Machine Learning and Cybernetics*, 2025, doi: 10.1007/s13042-025-02687-w.

BIOGRAPHIES OF AUTHORS

Somia A. Shams is an assistant lecturer in computer science at the Faculty of Science, Al-Azhar University, Cairo, Egypt. She received her B.Sc. and M.Sc. degrees from the Department of Mathematics, Faculty of Science, Al-Azhar University, in 2011 and 2022, respectively. She can be contacted at email: somia.shams10@azhar.edu.eg.

Asmaa Mohamed received the B.Sc. degree in science, in 2011, the M.Sc. degree in computer science, in 2017, and the Ph.D. degree in computer science, in 2023. She is currently a lecturer in computer science at the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. She has published several research papers in the field of AI, machine learning, internet of things, and data mining. She can be contacted at email: asmaamohamed89@azhar.edu.eg.

Abeer S. Desuky received a B.Sc. degree in science in 2003, and M.Sc. and Ph.D. degrees in computer science in 2008 and 2012, respectively. She is currently a professor of computer science at the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. She has published several research papers in the fields of artificial intelligence, machine learning, meta-heuristic optimization, and data mining and analysis. She is also a supervisor for master's and Ph.D. theses. She is a reviewer for many Scopus-indexed journals, such as IEEE Access and PeerJ. She can be contacted at email: abeerdesuky@azhar.edu.eg.

Gaber A Elsharawy received B.Sc. and Ph.D. degrees in computer science in 1986 and 1993, respectively. He became an associate professor in 2004 and a professor in 2009. He is currently a professor of computer science at the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. He has published several research papers in the fields of artificial intelligence, machine learning, meta-heuristic optimization, and data mining and analysis. He is also a supervisor for master's and Ph.D. theses. He is a reviewer for many Scopus-indexed journals. He can be contacted at email: Gaber.Ahmed@buc.edu.eg.

Rania Salah El-Sayed became an associate professor of computer science in 2021. She earned her Ph.D. in pattern recognition and her M.Sc. in network security from Al-Azhar University in 2013 and 2009, respectively. She obtained her B.Sc. in mathematics and computer science from Al-Azhar University in 2004. In 2012, she achieved CCNP Security certification from Cisco. Her research interests include pattern recognition, machine learning, and network security. She can be contacted at email: rania5salah@azhar.edu.eg.