Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Artificial Intelligence in Seismology-Deep Learning Models for Earthquake Magnitude Prediction

Manka Vasti[1], Dr. Amita Dev[2]

[1]USICT, Guru Gobind Singh Indraprastha University, Delhi, India and GD Goenka University, Gurugram, India, mankavasti@gmail.com
[2]Joint Director, Department of Training and Technical Education, Delhi, India, amita_dev@hotmail.com

Earthquake forecasting has become a continuous evolving discipline for the essential actions to be taken before the natural calamity strikes and raises its consequences on the life on earth, infrastructure and environment. Effective prediction and management of these disasters are critical to minimizing their devastating impact. Artificial Intelligence (AI) has automated several seismological processes due to its powerful algorithms for analysing complex and dynamic datasets, deriving nonlinear relationships in the voluminous datasets and building predictive models using Machine Learning (ML) and Deep Learning (DL) techniques. Deep learning has been applied for problem solving not only in the classification and segmentation problems in seismology but also can be trained for pattern recognition, signal detection, nonlinear modeling, and regression tasks. This research study exercises the power of the deep neural networks and designs the optimized architectures for the real-world Standard Earthquake Dataset (STEAD) waveforms, to predict the magnitude of the earthquake. The three- component (3C), one minute long, 6000 samples per component and with sampling rate of 100 Hz waveforms are used to model the real-world scenario. This study compares comprehensively the performance of various deep neural network model architectures based on their evaluation metrics namely mean absolute error, mean squared error, root mean squared error, mean absolute percentage error, and r-squared score. The results are also visualized in terms of training vs validation losses, prediction vs actual values, residuals values and prediction errors. On comparing different deep learning architectures, it is found that Long Short Term Memory and Bidirectional Long Short Term Memory neural networks outperformed the others in terms of the performance metrics. Long Short Term Memory model edges out the Bidirectional Long Short Term Memory slightly in terms of the simplicity, computational time and efficiency. The other architectures were comparatively less effective in capturing the complex relationships on the given dataset.

Keywords: Deep Learning, Long Short Term Memory, Bidirectional Long Short term Memory, Earthquake Magnitude Prediction, Seismological characterization.

1. Introduction

Seismological prediction has become most complex scientific challenge due to the unpredictable nature of its occurrence. It is very essential to perform seismological characterization so as to safeguard the mankind and the ecosystem. It plays an essential role in mitigating the consequences of the devastating earthquakes and ensuring the safety and well-being of all forms of life. Nowadays, Artificial Intelligence (AI) is transforming every industry sector and operationalizing various domains. AI has become the key solution for complex applications such as pattern recognition, visual perception, signal processing, biological imaging, speech recognition, non-linear modelling, classification, and many more. AI enhanced Seismology emphasizes on developing immaculate machine learning and deep learning model architectures that learn from and make predictions based on data and thus, have shown significant potential to enhance earthquake prediction and modelling. Deep learning is a sub discipline of Artificial Intelligence which works on a very large, complex and non-linear datasets and are based on training neural networks to model high-level abstractions in data. The emerging domain of geotechnical engineering is leveraging artificial intelligence (AI) to identify patterns and make informed decisions.

Historically, the approach to forecasting these events has evolved from reliance on traditional statistical methods to more sophisticated machine learning techniques, which offer the promise of higher accuracy and earlier detection. The increasing application of machine learning (ML) in scientific and technological fields has revolutionized the way complex data are analyzed and interpreted. The size of an earthquake is measured using various attributes such as earthquake magnitude, depth, p-wave and s-wave characteristics, signal-to-noise ratio etc. Earthquake magnitude is most significant amongst all attributes to deduce the nature and size of an earthquake and its related dire consequences. There are multiple scales on which an earthquake's magnitude is measured such as Local Magnitude (M_L), Moment Magnitude (M_W), Surface-Wave Magnitude (M_S), Body-Wave Magnitude (M_b), Duration Magnitude (M_d), Energy Magnitude (M_e) and Coda Magnitude (M_c). The most common scale is Local Magnitude (M_L) developed by Charles F Richter in 1935 and is represented as $M_L=\log_{10}(A)-\log_{10}(A_0)$ [1] measured on a Wood-Anderson instrument, where A is the amplitude of the seismic waves and A_0 is a standard amplitude.

With the evolution in the seismological data collection strategies, billions of bytes of continuous stream of non-linear and highly dynamic data are generated using sophisticated seismometers. As a result, this poses a great challenge for the seismological experts to analyze high voluminous data and derive the indicative attributes. Therefore, there is a critical need for the optimized AI models to process and analyze large scale datasets. This research study performs comprehensive analysis to predict the magnitude of the earthquake by considering the real-world earthquake waveform dataset of Alaska region. The study trains the deep learning model architectures for the real-world scenarios, so as to make it closely aligned with actual conditions and are reflective of real-life challenges and complexities. Various neural network model architectures are trained and optimized using waveforms of Standard Earthquake Dataset (STEAD).

2. Literature Review

In the early era, simple earthquake forecasting methods were adapted in analysing simple observations. Earlier people used to predict earthquakes by observing the atypical behaviour shown by the animals. The period followed when seismograms were predominantly analysed based on the intuition and logical reasoning of seismological experts. Later, there were several methods adopted by experts for earthquake analysis for long and short-term predictions such as mathematical and statistical models to analyze small earthquake data. Nowadays, the enormous collection of data from all over the world and the astonishing evolution of artificial intelligence and computing resources has made it possible to unfold its capability in the field of seismology. AI integrated with traditional seismological approaches, may lead to more accurate, timely, and reliable predictions, ultimately enhancing disaster preparedness and mitigation strategies. The use of artificial intelligence in seismological characterization performs much better and have many advantages over the traditional statistical methods for earthquake detection.

There are several precursors that represent changes within the earth that may lead to the occurrences of earthquakes. The deviations of the physical parameters from their normal value are indicative of some unusual activities in the internal structure of earth. Therefore, the detailed study of the earth is critical in order to lessen the impact of natural disasters such as earthquakes, cyclones, Tsunamis etc. Nowadays, due to the availability of high-resolution data from the global remote sensing devices, it has now become possible to detect the presence of anomalies in real time and to explore their complementary behaviour during the earthquake formative period. Therefore, the early detection of such anomalies may deduce the warning information about the impending natural tremors.

A list of the various aspects to consider while studying the earth's internal state include basement rock condition, change in soil resistivity, electromagnetic waves emission, electric potential change, variation in the radon gas emission, change in water level in wells and temperature of thermal springs, change in animal behaviour etc. [2].

The internal structure of earth is affected by natural forces such as micro-tremors, microseisms, wave, wind, tide, air pressure, precipitation and a variety of human induced sources such as hydraulic fracturing, oil prospecting explosions, cavern collapsing, mining etc. [3].

The analysis of ten earthquakes in Delhi/NCR of magnitude in the range from 2.5 to 4.5, revealed stress drop of 3-13 MPa using Fourier acceleration spectra [4]. GPS and DEMETER detected ionospheric anomalies and ULF emissions before the December 12, 2009 earthquake, suggesting potential for short-term predictions [5].

Conventional seismic prediction methods rely on data, but lacks accuracy and timeliness, necessitating innovative, precise approaches. Also, it is arduous for a seismological expert to analyse billions of bytes of continuous stream of highly dynamic data using his expertise and logic. The remote sensors above the earth and below the seas generate terabytes of data and it is very challenging to derive the indicative attributes from the continuous stream due to the complex nature of seismic data. Therefore, there is a critical need for pre-trained AI models to process and analyse large-scale seismic data efficiently, identify patterns, and predict potential earthquakes with higher accuracy and timeliness than traditional methods.

The applications of machine learning in seismology primarily focus in producing earthquake early warning systems, ground motion predictions, seismic tomography and illuminating geophysical structure, earthquake geodesy and non-inertial deformation [6]. The conventional approaches may not be adequately sufficient in the scenario where a region is more susceptible to large number of small seismic events [7].

In a deep learning neural network, trained using more than 131,000 mainshock-aftershock pairs to predict aftershocks on a test dataset comprising of more than 30,000 mainshock-aftershock produced results of 0.849 AUC (area under curve) as compared to 0.583 AUC using classic Coulomb failure stress change [8].

A research study presented an early earthquake warning method called PreSeis (Pre-sesimic) based on two-layer feed forward neural network that estimates the earthquake hypocentre location, the moment magnitude, and the expansion of the evolving seismic rupture [9].

A deep Convolutional Neural Network (CNN) trained with 200, 3D synthetic seismic images, adam optimiser with learning rate of 0.0001 and 25 epochs generated an accuracy of 95% and loss converging to 0.01 in the 3D seismic fault segmentation [10].

[11] applied twenty machine learning models such as support vector machine, neural networks, random forest, gradient boosting, logistic regression etc. for seismic facies analyses. The efficiency achieved to predict seismic facies is upto 98.3% with error upto 0.004%.

A research work proposed a framework called PhaseLink to identify for the grid -free phase association [12].To discriminate noise from the earthquake event, a Generative Adversarial Network (GAN) with Random Forest classifier is used to learn the characteristics of the first arrival of p-waves in 3,00,000 waveforms with 99.2% and 98.4% for earthquake p-wave detection and noise detection respectively [13].

In order to handle the challenges of massively generated seismological dataset, a cloud-based analytics system unveils the seismic attributes, selects features, and performs the seismic interpretation process [14]. [15] reviewed some of the finest AI techniques in the application area of pile foundations. ConvNet, a Convolutional Neural Network, is used to detect phases of seismic body wave even when the seismograms are affected by noise [16].

Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), Combine CNN and LSTM, and Attention-based deep learning models were used to predict the number and strength of aftershocks following a major earthquake [17].

A simple convolutional neural network with 67,939 parameters distinguished earthquakes from noise with an accuracy of 96.7%, 95.3%, and 93.2% in training, validation and test dataset. The dataset comprises of three component seismograms of 3,04,878 earthquakes comprising of 6,29,095 earthquake waveforms and 6,15,847 noise waveforms from 1487 broad band receivers [18].

A random forest method for classifying large earthquake occurrences and Long Short-Term Memory (LSTM) provided a rough estimation of earthquake magnitude on large earthquake datasets [19].

A convolutional neural network method which is able to process complex-valued seismic data in the time–frequency domain by using complex convolutional and complex activation functions trained the model and achieved a Mean Absolute Error (MAE) of 4.51 km for epicentral distance, 6.15 km for depth, and 0.26 for magnitude estimation [20].

The study on Turkey's earthquake catalogue dataset from January 2014 to August 2023 leveraged Artificial Neural Network (ANN) for the prediction of earthquake magnitude and risk assessment. The trained ANN model achieved a Root Mean Square Error (RMSE) 0.18 and the value of R-Squared (R2) equal to 0.99 [21].

A novel research study to estimate the earthquakes magnitude based on the LSTM network architecture using two decades of seismic events could perform the short-term earthquake prediction accurately [22].

The relationship between Total Electron Content (TEC) in Ionosphere and earthquake occurrences has been identified using LSTM based prediction models [23]. The study compares the performance metrics of LSTM based predictions with Support Vector Machine, Linear Discriminant Analysis and Random Forest Classifier to evaluate the proposed models for earthquake prediction.

A regressor named MagNet based on convolutional and recurrent neural network models and trained on single station waveforms is capable of predicting the local magnitudes with an average error of nearly zero and 0.2 standard deviation [24].

A temporal-convolutional neural network architecture is trained to learn epicentral distance and P-wave travel time from a single station 1-minute seismograms. The trained neural network achieved the evaluation of epicentral distance and P-wave travel time with mean errors of 0.23 km and 0.03 seconds, and standard deviations of 5.42 km and 0.66 seconds, respectively [25].

An approach comprising of two models, the first one used the LSTM model architecture predicting the year, location, and magnitude of earthquakes and the second model followed decomposition method of dividing the dataset based on magnitude ranges and applying LSTM models to the segmented data. The decomposition model proves to be more efficient, especially in predicting large magnitude earthquakes. The results of the research study of these models are also compared with ANN models and found to result in better performance metrics [26].

A research study pursued the implementation of recurrent neural network named LSTM on the earthquake dataset and found the results to be far better than feed forward neural networks [27].

Deep neural networks used on unlabeled data for performing the clustering of seismic data and optimizing feature representations for specific tasks, achieved high precision supporting earthquake early warning systems and various other applications [28].

A proposed research study introduced a novel prediction method that applies an attention mechanism (AM), convolutional neural network (CNN), and bi-directional long short-term memory (BiLSTM) models. It predicts the number and maximum magnitude of earthquakes in mainland China using the regional earthquake catalog [29].

This study enhances earthquake forecasting by applying deep learning (DL) and machine learning (ML) to lab-simulated earthquakes. It introduces autoregressive models and expands on existing techniques to predict fault zone stress and labquake events, outperforming current methods. These advancements suggest improved accuracy in earthquake prediction, particularly in predicting the time to start and end of failure under various conditions, including pre-seismic creep and aperiodic events, offering promising potential for real-world applications [30].

3. Dataset

The development of advanced techniques has led to collection of large earthquake datasets and more robust models to process and analyse the earthquake signals in real time. For supervised machine learning, labelled dataset is a prerequisite to build a model. There are many labelled datasets available, but the reliability of those datasets is always a matter of concern and thus, imposes a big hurdle in the processing. Therefore, a high-quality, large-scale global data recorded using highly sensitive seismometers are required to study the structural changes within the earth. The dataset used in this research study is a large scale, global dataset named STEAD (Standard Earthquake Dataset) [31] that comprises of 1.2 million samples of waveform data recorded from January 1984 – August 2018, associated with 4.5 lakh earthquakes. Each waveform is one minute long with sampling rate of 100 Hz i.e. 6000 samples / minute. The overall dataset contains 19K hours of earthquake recordings by 2613 seismometers located all over the world. Each earthquake is represented as Numpy arrays containing 6000 samples associated with 60 seconds of ground motion recorded in three components (3C) i.e. north-south(N), vertical(Z) and east-west(E) directions. There are 35 attributes in the metadata associated with each earthquake such as network code, location of the receiver, origin time, epicentral location, depth, magnitude, magnitude type, focal mechanism, arrival times, signal to noise ratio, coda end etc.

To process the above dataset, this research study has used one of the very powerful toolboxes of python programming namely Seisbench [32]. Seisbench is an open-source toolbox that provides support of functions to data analyses using machine learning and deep learning models. The dataset API facilitates the downloading of seismic waveform datasets such as ETHZ, GEOFON, INSTANCE, STEAD, etc. for machine learning and deep learning algorithms. Seisbench datasets consist of two components i.e. the waveform traces and the metadata. Each dataset has two files namely a .csv file for the metadata and a. hdf5 file for the actual waveform traces.

The study creates the most optimised Deep Learning (DL) architecture models using Artificial Neural Network, Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory network and Bi-directional Long Short-Term Memory network on the real-world earthquake waveform dataset derived from Standard Earthquake Dataset (STEAD).

All the neural network architectures are designed to be used for a regression task of predicting the earthquake magnitude. Deep Learning is a multi-layer network model that has the capability to derive features from the raw data thereby enhancing the functionality of the neural network to model complex earthquake data. The deeper the learning, the higher will be the degree of abstraction in the model. The most emphasis in such models are provided to hyperparameters optimization such as number of hidden layers, number of computing units in each layer, activation functions used, optimiser adopted, weight manipulation, no. of epochs, no. of iterations and many more. The training phase, being most critical, requires weight updation in such a way so as to minimize the loss function. Therefore, amongst several optimizers namely Gradient Descent, Stochastic Gradient Descent, Adam stochastic optimization, Adagrad, Adadelta etc., the most suitable to the given application is selected. Deep learning applications generally require multi-processor graphics cards or GPUs for developing the model. It is envisioned to be the future trend to be followed in order to uncover the complex seismic prediction models.

This research study develops deep learning model architectures on Standard Earthquake Dataset (STEAD) to predict the magnitude of earthquakes using three-Component (3C) earthquake waveforms. The dataset comprises of 17,324 waveforms containing 3 Components of 6000 samples each making it a total 31,18,32,000 data points for analyses. The dataset is split into training, validation, and test and the models are trained on 10 epochs, batch size of 64 instances for training, validation and testing. The dataset considered under study is primarily of the Alaska region that is located in the northwest extremity of North America. Alaska is highly seismically active region and had many devastating magnitude earthquakes in the past and also has the potential to generate large magnitude earthquakes in the future. The details of the dataset used for the study are shown in Table 1.

Table 1 : Dataset selection for prediction

Description	Data Samples
STEAD - Original Dataset Traces (earthquake waveforms + noise waveforms)	1.2 million
Earthquake Waveforms traces (3 Component dataset)	10,30,231
Noise Waveforms traces (3 Component dataset)	2,35,426
Total no. of samples per waveforms per component	6,000
Sampling Rate (in Hz)	100
No. of earthquake waveform traces of Alaska (AK) region	141877
No. of earthquake waveform traces of Alaska (AK) region with magnitude greater than 3.0	17324
Total Data points processed for study	17324 x 6000 x 3 = 31,18,32,000
Output	Earthquake Magnitude

4. Evaluation Metrics

In order to build and deploy a generalized model, the evaluation metrics have very high significance in finding the performance of the deep neural networks. The following evaluation metrics are used in this research study to compare the performance of the various deep neural networks used for the earthquake magnitude prediction. N is the total data points, y_j is the actual output and $\hat{y}_j$ is the predicted output:

a) Mean Absolute Error (MAE): Mean Absolute Error is defined as the distance between the Predicted and the Actual Values. MAE measures the average magnitude of errors without considering the direction. The calculation of Mean Absolute Error is shown by equation no. (1).

$$\mathrm{MAE} = \frac{1}{N}\sum_{j=1}^{N} |y_j - \hat{y}_j| \tag{1}$$

b) Mean Squared Error (MSE): In Mean Squared Error, the square of the average errors between the predicted and the actual values are considered. The squared term is performed in order to avoid the cancellation due to negative terms. Also, higher magnitude errors when raised to their square, indicates more focus to be put on them. The calculation of Mean Squared Error is shown by equation no. (2).

$$\mathrm{MSE} = \frac{1}{N}\sum_{j=1}^{N} (y_j - \hat{y}_j)^2 \tag{2}$$

c) Root Mean Squared Error (RMSE): It is a metric that is derived from Mean Squared Error by taking the square root of MSE. Root Mean Squared Error is calculated as shown by equation no. (3).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} (y_j - \hat{y}_j)^2} \tag{3}$$

d) Mean Absolute Percentage Error (MAPE): It is defined as the average absolute percentage error between the predicted value and the actual value. MAPE is calculated as per the equation (4).

$$\mathrm{MAPE} = \frac{1}{N}\sum_{j=1}^{N} \frac{|y_j - \hat{y}_j|}{y_j} \text{ X } 100 \tag{4}$$

e) Mean Squared Logarithmic Error (MSLE): It is used to evaluate the model when the data has wide range of values. This metric is particularly useful in the application that may see exponential growth such as population, stock etc.

MSLE is calculated as shown in equation (5):

$$\mathrm{MSLE} = \frac{1}{N}\sum_{j=0}^{N-1}(\log_e(1 + y_j) - \log_e(1 + \hat{y}_j))^2 \qquad (5)$$

f) R-Squared (Coefficient of Determination): It is used to compare the trained model with the baseline model. Also, known as the Goodness of Fit, it is calculated as per the equation no. (6) given below. SSr is the Squared Sum Error of Regression Line. SSm is Squared Sum Error of Mean Line.

$$R^2 = 1 - \frac{SSr}{SSm} \qquad (6)$$

5. Methodology

This section provides the comprehensive details of the methodology adopted for the study. It provides details of the various deep learning model architectures implemented on the earthquake waveform dataset to derive the performance metrics of each of them for effective comparison.

Fig 1. shows the methodology adopted for earthquake magnitude prediction using deep neural networks.

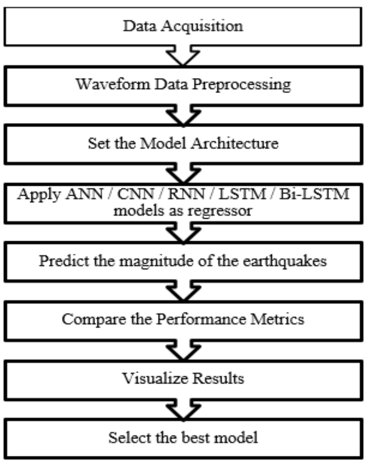

Fig. 1: Methodology proposed for earthquake magnitude prediction using deep neural networks

It includes the following steps:

i. Waveform data Acquisition: The dataset acquired for the study, STanford EArthquake Dataset (STEAD) is the reliable global dataset comprising of over 1 million seismograms associated with 4.5 lakh earthquakes. Each earthquake seismogram is captured across three components i.e. Vertical (Z), North(N) and East(E) also known as 3Cs. The recorded waveform is real world earthquake data consisting of one-minute long, raw waveform with 100Hz sampling rate, resulting in 6000 samples per component per event. Therefore, each earthquake is represented with a total of 18000 samples. The computing resources to process such large dataset is always a big challenge. Therefore, the study is restricted to Alaska earthquake data with earthquake magnitudes > 3.0. This leads to a total of 17,324 earthquakes with 18000 samples per event, makes it total of 31,18,32,000 data samples to process for the study using deep learning models.

ii. Data Pre-Processing: This step includes waveform data normalization, reshaping the data and splitting the dataset into train, test and validation datasets.

- Set the Model Architecture: This requires deciding on the type of input, number of hidden layers, weights initialization, neurons per layers, activation functions, optimizer, loss function, batch size, no. of epochs etc.

- Apply the deep learning models: This research study is based on the implementation of the following are the deep learning models. Each model is trained on the same dataset and is specifically tasked as regressor for the earthquake magnitude prediction. By providing the consistent training, validation and test dataset ensure the selection of the best model architecture capable to capture the underlying patterns in the seismic dataset and provides the most accurate earthquake magnitude predictions

- Artificial Neural Network: A basic feedforward neural network used as a baseline for performance comparison.

- Recurrent Neural Network: This is used for handling sequential data, therefore, RNNs are employed to capture temporal dependencies in the waveform data.

- Convolutional Neural Network: Convolutional Neural Networks are utilized to automatically extract spatial features from the time-frequency representations of the seismic signals.

- Long Short Term Memory: LSTM networks are implemented to model long-term dependencies in the seismic data, which is crucial for capturing patterns in the waveform sequences.

- Bidirectional Long Short Term Memory: The Bidirectional LSTM architecture is tested to capture information from both past and future time steps in the sequence, providing a more comprehensive understanding of the seismic events.

iii. Predict the earthquake magnitude values: After the successful training of the deep learning neural network, the prepared model is used for the prediction on the test dataset.

iv. Compare performance metrics: This step allows for a comprehensive evaluation of each model's effectiveness in the critical regression task by comparing their performance metrics such as training loss, validation loss, test dataset’s Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, Mean Absolute Percentage Error, and R-Squared values.

v. Visualize results: For effective comparison, the results are visualized in terms of training vs validation loss, predicted vs actual values, model’s residual values, model’s error distribution and heatmap to represent the prediction errors.

vi. Select the best Model: By comparing all the performance metrics, the best model architecture for the given dataset is recommended for the most accurate earthquake magnitude predictions.

6. Results

In this section, the performance metrics of various deep learning models are evaluated and compared. The different deep learning models namely ANN, RNN, CNN, LSTM and Bi-LSTM are tasked as regressors to predict earthquake magnitude using the STEAD waveform dataset. Each model was trained, validated and tested on the same pre-processed dataset ensuring same environment across all the models processing. The performance of deep learning models was compared using the key evaluation metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), R-Squared(R^2), training loss and validation loss values. The results were plotted to indicate the strengths and limitations of each model in predicting earthquake magnitudes on the real-world seismic waveform dataset. The results in the study are achieved by training deep learning algorithms on the STEAD dataset using the fourth-generation Tensor Cores, NVIDIA L4 Tensor Core GPU, paired with the CV-CUDA library, Hi-RAM environment of Google Colab Pro.

The training and validation loss curves for all five deep learning models are shown in Fig 2 (a) – (e). These loss curves show how well each model learnt from the training dataset and how effectively it generalizes to unseen validation dataset on 10 epochs of training. The LSTM and CNN models have more consistently stable loss curves, while Bi-LSTM stabilizes after a few epochs. ANN and RNN models also indicate smooth decline in loss curves and are effective as baseline models.

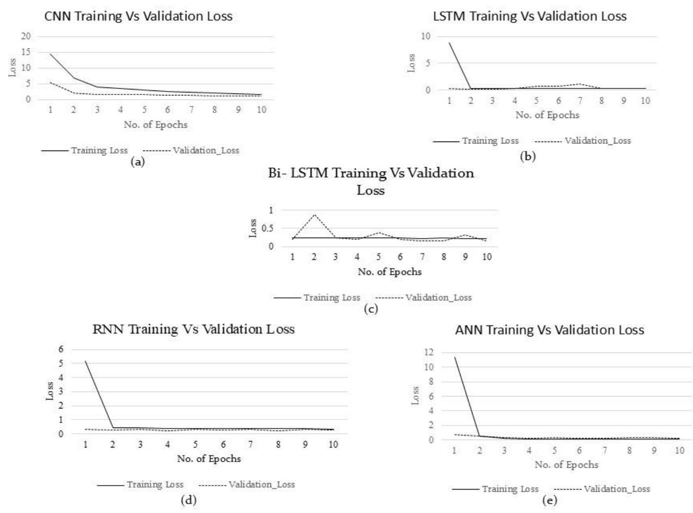

Fig. 2 (a) – (e): Comparison of training vs validation losses over the epochs of training on the earthquake waveform dataset using various deep learning models. (a) Convolutional

Neural Network (b) Long Short Term Memory Neural Network (c) Bidirectional Long Short Term Memory Neural Network (d) Recurrent Neural Network (e) Artificial Neural Network

Fig. 3 (a)-(e), represents the scatter plots between the predicted earthquake magnitude vs the actual earthquake magnitudes for all the five deep learning models. The diagonal line shows the ideal scenario where the predicted values are same as actual values. As seen in Fig. 3, LSTM and Bi-LSTM show better spread around the diagonal, indicating strong predictive accuracy. However, there are some magnitude overestimations by Bi-LSTM and underestimation by LSTM models. LSTM and Bi-LSTM models have the capability to capture temporal dependencies well and therefore, provide more consistent predictions as compared to other models. Though, CNN model also performed well, but it shows some issues with higher magnitude predictions.

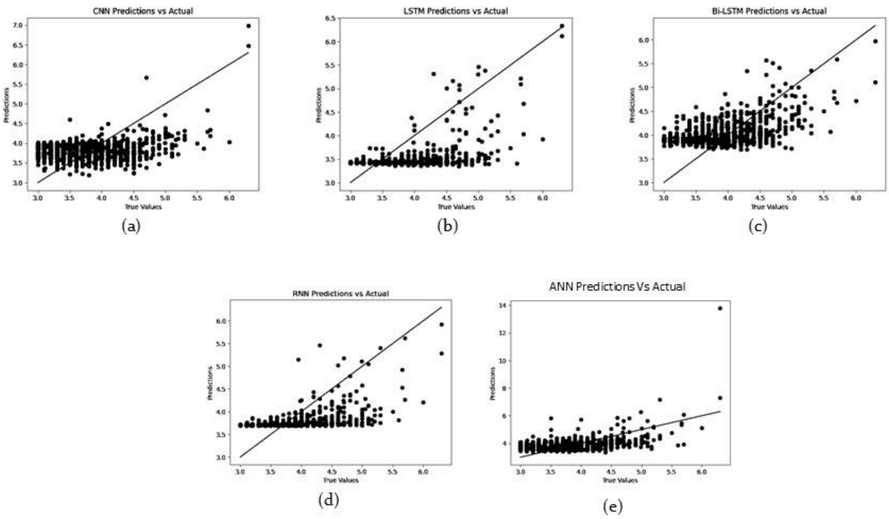

Fig. 3 (a) – (e): A comparison of Predictions Vs Actual values of the various deep learning models.

RNN model has more outliers and ANN indicate many deviations from the actual values and hence, these models are less suitable for accurately predicting earthquake magnitudes.

In Fig. 4 (a) – (e), the plots display the residuals, i.e. the difference between the actual values and the predicted values, versus the predicted values. Residuals help in understanding the accuracy and bias of the model. In the ideal scenario, the residuals are randomly scattered around zero to show that the model's predictions are unbiased and accurate.

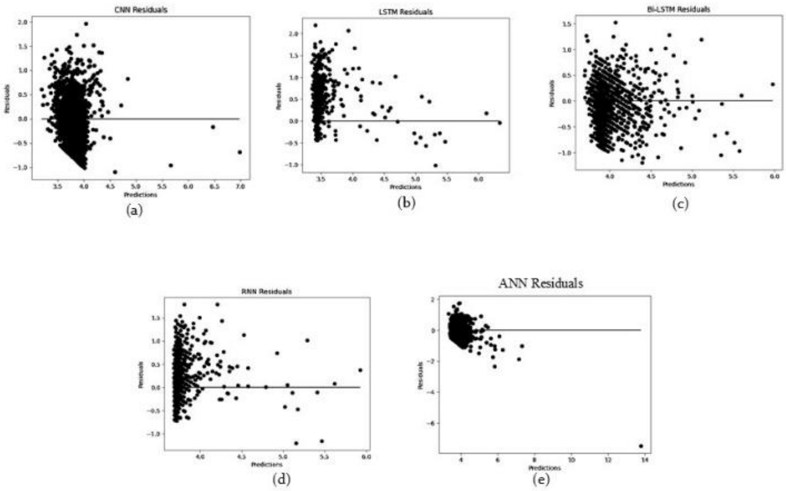

Fig. 4 (a) – (e): A comparison of residuals of the various deep learning models on earthquake waveform dataset

These residual plots indicate that CNN, LSTM and Bi-LSTM models generally provide more reliable and accurate predictions over other models.

The distribution of prediction errors for all five different models i.e. CNN, LSTM, Bi-LSTM, RNN, and ANN are presented across the subplots in Fig.5 (a)- (e). As can be seen from the Fig.5, LSTM model shows the narrowest spread with a peak near zero. This indicates that LSTM model generalizes well and predicts the most accurate earthquake magnitude values.

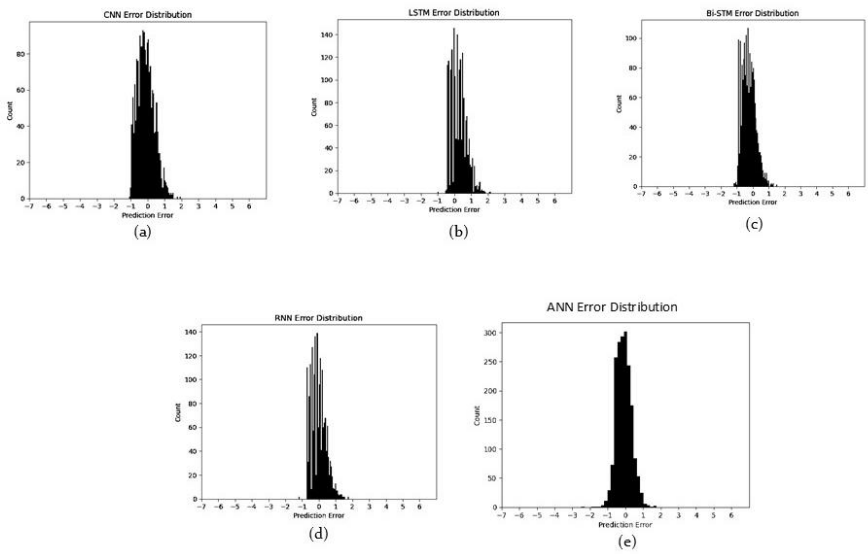

Fig. 5 (a) – (e): A comparison of distribution of error of the various deep learning models

In Fig. 6 (a) – (e), the distribution of prediction errors in heatmaps are shown for CNN, LSTM, Bi-LSTM, RNN, and ANN models is shown. Heatmap shows the error magnitude difference between actual and predicted values for the samples in test dataset. The colour intensity representing the magnitude of the error. Lighter shades indicate larger errors, while darker shades indicate smaller errors.

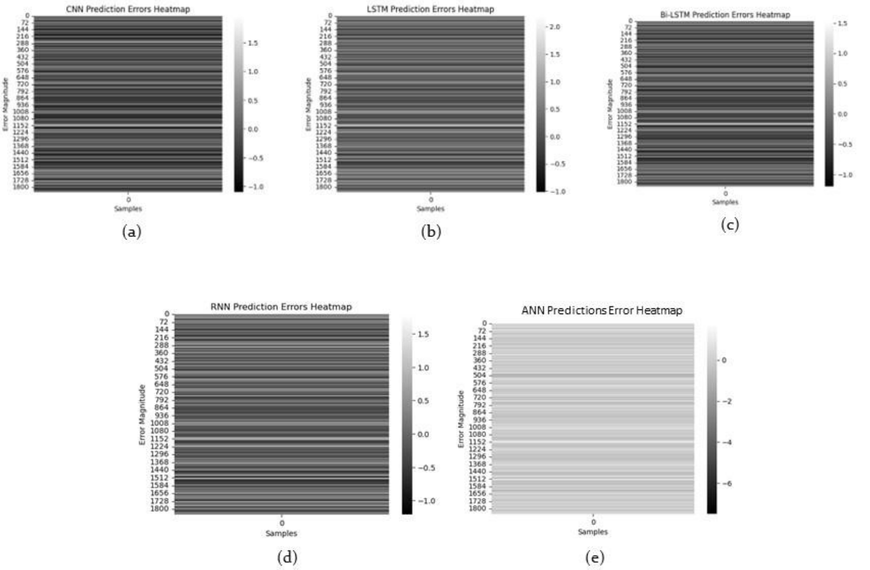

Fig. 6 (a) – (e): Heatmap shows the error magnitude difference between actual and predicted values

The details of the hyperparameters such as the model architecture, optimizer, no. of parameters trained, activation function, loss function etc. are mentioned in Table 2. The research study includes the comparative analysis of the performance metrics of Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (LSTM). Table 2 summarizes the performance of various deep learning models based on the metrics namely training loss, validation loss, mean absolute error, mean squared error, root mean squared error and mean absolute percentage error on the training dataset. Bi-LSTM and LSTM models exhibit the lowest error metrics and validation loss, indicating superior accuracy and generalization.

Table 3 represents the comprehensive comparative analysis of evaluation metrics of various deep learning models implemented on test dataset. Overall, LSTM and Bi-LSTM have better values of performance metrics over other models.

Table 2 : Comparative Analysis of Performance Metrics of Various Deep Learning Algorithms for Earthquake Magnitude Prediction on Training dataset

Name of the Neural Network	Model Architecture	Optimiser	Param #	Training Loss	MAE	MSE	RMSE	MAPE	Validation Loss
Convolutional Neural Network (CNN)	Conv1D(32 – 64 – 128 – 64 – 32 – 1) (Relu, Linear, Batch Normalization)	Adam(learning Rate = 0.001), Loss = “Mean Squared Error”	2,51,15,393	1.6482	0.6163	0.7808	0.8833	16.1306	1.04
Long ShortTerm Memory (LSTM)	LSTM (64 – 1) (Relu, Linear, Batch Normalization)	Adam(learning Rate = 0.001), Loss = “Mean Squared Error”	17,729	0.2790	0.4130	0.2714	0.5208	11.1092	0.2579
Bidirectional Long ShortTerm Memory (LSTM)	Bi-LSTM (64 – 1 (Relu, Linear, Batch Normalization)	Adam(learning Rate = 0.001), Loss = “Mean Squared Error”	35,457	0.2318	0.3786	0.2233	0.4725	10.2138	0.1628
Recurrent Neural Network(RNN)	LSTM (64 – 1 (Relu, Linear, Batch Normalization)	Adam(learning Rate = 0.001), Loss = “Mean Squared Error”	4,673	0.3463	0.4587	0.3417	0.5846	12.2990	0.2663
Artificial Neural Network	256 – 128 – 64 – 32 – 1 (Relu, Linear, Batch Normalization)	Adam(learning Rate = 0.001), Loss = “Mean Squared Error”	11,71,905	0.1272	0.4878	0.1272	0.3566	7.7800	0.2766

Table 3 : Comparative Analysis of Performance Metrics of Various Deep Learning Algorithms for Earthquake Magnitude Prediction on Test dataset

Name of the Neural Network	MAE	MSE	RMSE	MAPE	R2
Convolutional Neural Network (CNN)	0.419	0.261	0.511	0.116	0.932
Long Short Term Memory (LSTM)	0.404	0.2	0.44	0.1027	0.892
Bidirectional Long Short Term Memory (LSTM)	0.412	0.2	0.44	0.1178	0.893
Recurrent Neural Network (RNN)	0.363	0.21	0.45	0.0987	0.819
Artificial Neural Network (ANN)	0.35	0.217	0.466	0.0958	0.887

7. Conclusion

Earthquake characterization and prediction is very crucial for saving humanity, infrastructure and the ecosystem. In the growing need for accurate and timely seismological characterization, this research study leverages AI techniques to transition the processes from the obsoleted traditional statistical methods to more advanced AI methodologies applied on the massive datasets. The highly dynamic, non linear and complex datasets generated from the modern seismological practices necessitates a drift from the traditional paradigms to more advanced seismological practices and thereby indicating the significance of integrating AI into seismological practices. As the volume and complexity of seismological data continue to grow, traditional methods become more insufficient for timely and accurate earthquake prediction. AI-enhanced seismology not only improves the ability to predict seismic events but also plays a vital role in mitigating their consequences.

This research study has delved into the implementation of optimized neural network architectures using STEAD for analysing complex, nonlinear datasets to predict seismic events effectively. The research study emphasizes on designing, training and optimizing neural network models to predict earthquakes magnitude. Using a comprehensive comparative analysis of different model architectures namely Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, Bidirectional LSTM (Bi-LSTM) networks, Recurrent Neural Networks (RNN), and Artificial Neural Networks (ANN), the study evaluates their performance basis various performance key indicators such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results revealed that the LSTM and Bi-LSTM models outperformed other architectures, demonstrating superior accuracy and generalization in predicting earthquake magnitudes. The LSTM model is also slightly better over the Bi-LSTM model in terms of simplicity, computational efficiency, and overall performance, making it the most effective architecture for this specific task. The CNN model also performed well, especially in explaining the variance in the data, as indicated by its high R^2 value. However, it exhibited slightly higher error metrics compared to the LSTM models. The RNN model, while effective, showed more variability in its predictions, making it less reliable for capturing the complex relationships inherent in seismic data. The ANN model, although demonstrating the lowest MAE and MAPE, had a slightly lower R^2 value, and therefore, it may not capture the full complexity of the data as effectively as the LSTM-based models.

In conclusion, the findings of this research underscore the potential of AI-driven models to advance the field of seismology. The optimized deep learning architectures explored in this study offer significant improvements in earthquake prediction, providing a more reliable and efficient means of analysing complex seismic data. As seismological practices continue to evolve, the integration of AI will be indispensable in addressing the challenges posed by the dynamic and nonlinear nature of seismic phenomena, ultimately contributing to a safer and more resilient world.

To further illustrate the comparison in various models, the plots of training vs validation loss, predicted vs actual values, model’s residual values, model’s error distribution and heatmap to represent the prediction errors were shown. The plots indicate that out of all the models, LSTM and Bi-LSTM models performed the best as compared to other models on the given dataset.

Acknowledgement

Statements and Declarations

Competing Interests:

Both the authors have no competing financial interests for this paper.

Conflict of Interest:

The authors declare that they have no conflict of interest.

Contribution by the Authors:

Dr. Amita Dev (PhD Supervisor): Conceptualization, Supervision, Reviewer of the Paper

Manka Vasti (PhD Candidate): Investigation, Methodology, Analysis, Results, Writing and Editing Paper.

Ethics Statement

No ethical approval was required as this study involved publicly available, non-sensitive earthquake data. All methods comply with relevant data protection and usage guidelines. No human subjects were involved.

References

1. Green, D. N., Luckett, R., Baptie, B., & Bowers, D. (2020). A UK local seismic magnitude scale, MLP, using P-wave amplitudes. Geophysical Journal International, 223(3), 2054-2065. https://doi.org/10.1093/gji/ggaa438
2. Alarifi, A.S.N., Alarifi, N.S.N. & Al-Humidan, Saad. (2012). Earthquakes magnitude predication using artificial neural network in northern Red Sea area. Journal of King Saud University - Science. 24. 301–313. 10.1016/j.jksus.2011.05.002.
3. Kitagawa, G., Takanami, T., Matusumoto, N.(2001). Signal Extraction Problems in Seismology. International Statistical Review, 69,1,129-152.
4. Bansal, B.K., Singh, S.K., Suresh G., Mittal H.A.(2021). Source and Ground Motion Study of Earthquakes in and Near Delhi (the National Capital Region), India. Nat Hazards (Dordr). 2021 Nov 30:1-21. doi: 10.1007/s11069-021-05121-w. Epub ahead of print. PMID: 34866792; PMCID: PMC8631568.
5. Karia, S., Sarkar, S., Pathak, K., Sharma, A.K., Ranganath, H., Gwal, A.K. (2013). Analysis of space- and ground-based parameters prior to an earthquake on 12 December 2009, International Journal of Remote Sensing, 34:21,7779-7795.
6. Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J., Gerstoft, P. (2018). Machine Learning in Seismology: Turning Data into Insights. Seismological Research Letters 2018, 90 (1): 3–14. doi: https://doi.org/10.1785/0220180259
7. Jiao, P., Alavi, A.H., (2020). Artificial intelligence in seismology: Advent, performance and future trends. Geoscience Frontiers,11, 739-744. https://doi.org/10.1016/j.gsf.2019.10.004.
8. DeVries, P.M. R., Viégas, F., Wattenberg, M., Meade, B. J.(2018). Deep learning of aftershock patterns following large Earthquakes, Nature, Springer, 560, 632-639.
9. Böse, M., Wenzel, F., Erdik, M.(2008). PreSEIS: A Neural Network-Based Approach to

Earthquake Early Warning for Finite Faults. Bulletin of the Seismological Society of America. 98 (1): 366–382. doi: https://doi.org/10.1785/0120070002.

10. Wu, X., Liang, L., Shi, Y., Fomel, S. (2019). FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics, 84(3), 35-45.
11. Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., (2018). Seismic facies analysis using machine learning. Geophysics, 83(5), 83-95.
12. Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. Journal of Geophysical Research: Solid Earth, 124, 856–869. https://doi.org/10.1029/2018JB016674
13. Li, Z., Meier, M., Hauksson, E., & Zhan, Z. & Andrews, J. (2018). Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning. Geophysical Research Letters. 45. 10.1029/2018GL077870.
14. Huang, L., Dong, X., & Clee, T., (2017). A scalable deep learning platform for identifying geologic features from seismic attributes. The Leading Edge. 36. 249-256. 10.1190/tle36030249.1.
15. Shahin, M.A. (2016). State-of-the-art review of some artificial intelligence applications in pile foundations, Geoscience Frontiers, Volume 7, Issue 1, 2016, Pages 33-44, ISSN 1674-9871, https://doi.org/10.1016/j.gsf.2014.10.002.
16. Ross, Z. E., Meier, M. and Hauksson, E. (2018). P-Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning. Journal of Geophysical Research: Solid Earth, 123, 5120–5129. https://doi.org/10.1029/2017JB015251.
17. Witze, A. (2023). AI predicts how many earthquake aftershocks will strike - and their strength. Nature. https://doi.org/10.1038/d41586-023-02934-6
18. Magrini, F., Jozinovi, D., Cammarano, F., Michelini, A., Boschi, L. (2020). Local earthquakes detection: A benchmark dataset of 3-component seismograms built on a global scale. Artificial Intelligence in Geosciences.1, 1-10.
19. Wang, X., Zhong, Z., Yao, Y., Li, Z., Zhou, S., Jiang, C., & Jia, K. (2023). Small Earthquakes Can Help Predict Large Earthquakes: A Machine Learning Perspective. Applied Sciences, 13(6424). https://doi.org/10.3390/app13116424
20. Ristea, N.-C., & Radoi, A. (2022). Complex neural networks for estimating epicentral distance, depth, and magnitude of seismic waves. IEEE Geoscience and Remote Sensing Letters, 19, 1-5. https://doi.org/10.1109/LGRS.2021.3059422
21. Biswas, S., Kumar, D., Nas, M., Softa, M., Akgün, E., & Bera, U. (2023). Evaluation of artificial neural networks in predicting earthquake magnitudes and assessing risks in Türkiye. In Proceedings of DEU International Symposium Series on Graduate Researches-2023 (pp. 251-259). Geo Science & Technology, İzmir, Türkiye, December 7-8, 2023.
22. González, J., Yu, W., & Telesca, L. (2019). Earthquake magnitude prediction using recurrent neural networks. Proceedings, 24(22). https://doi.org/10.3390/IECG2019-06213
23. Abri, R., & Artuner, H. (2022). LSTM-based deep learning methods for prediction of earthquakes using ionospheric data. Gazi University Journal of Science, 35(4), 1417-1431. https://doi.org/10.35378/gujs.950387
24. Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. Geophysical Research Letters, 47(1), e2019GL085976. https://doi.org/10.1029/2019GL085976
25. Mousavi, S. M., & Beroza, G. C. (2020). Bayesian-deep-learning estimation of earthquake location from single-station observations. IEEE Transactions on Geoscience and Remote Sensing, 58(11), 8211–8224. https://doi.org/10.1109/TGRS.2020.298877
26. Berhich, A., Belouadha, F.-Z., & Kabbaj, M. I. (2020). LSTM-based models for earthquake prediction. In Proceedings of the 2020 International Conference on Advanced Communication

Technologies and Networking (pp. 1-7). https://doi.org/10.1145/3386723.3387865
27. Bellamkonda, S., Settipalli, L., Vedantham, R., & Vemula, M. K. (2021). An enhanced earthquake prediction model using long short-term memory. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(14), 2397–2403. https://doi.org/10.17762/turcomat.v12i14.10665
28. Mousavi, S. M., Zhu, W., Ellsworth, W., & Beroza, G. (2019). Unsupervised clustering of seismic signals using deep convolutional autoencoders. IEEE Geoscience and Remote Sensing Letters, 16(11), 1693-1697. https://doi.org/10.1109/LGRS.2019.2909218
29. Kavianpour, P., Kavianpour, M., Jahani, E., & Ramezani, A. (2023). A CNN-BiLSTM model with attention mechanism for earthquake prediction. The Journal of Supercomputing, 79, 1-33. https://doi.org/10.1007/s11227-023-05369-y
30. 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, 598, 117825. https://doi.org/10.1016/j.epsl.2022.117825
31. Mousavi, S. M., Sheng, Y., Zhu, W., Beroza, G. C. (2019). STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI, in IEEE Access, vol. 7, DOI: 10.1109/ACCESS.2019.2947848
32. HelMUG (Helicoidal Model User Group). (2021). Seisbench: A benchmarking toolbox for seismic waveform analysis (Version 1.0) [Software]. GitHub.https://github.com/seisbench/seisbench https://doi.org/10.1007/978-981-15-0372-6_1