JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Wardhani, T. P. M., Tahir, Z., Warni, E., Bustamin, A., Imran Oemar, M. A. F., & Kayyum, M. A. (2025). Deep learning approach in seismology: Enhancing earthquake forecasting using K-means clustering and LSTM networks. *Journal of Information and Communication Technology, 24*(1), 29-51. https://doi.org/10.32890/jict 2025.24.1.2

Deep Learning Approach in Seismology: Enhancing Earthquake Forecasting using K-Means Clustering and LSTM Networks

[1]Tyanita Puti Marindah Wardhani, [2]Zulkifli Tahir, [3]Elly Warni, [4]Anugrayani Bustamin, [5]Muhammad Alief Fahdal Imran Oemar & [6]Muhammad Alwi Kayyum
[1,2,3,4,5&6]Department of Informatics, Universitas Hasanuddin, Indonesia

[*1]tyanitaputi@unhas.ac.id
[2]zulkifli@unhas.ac.id
[3]elly@unhas.ac.id
[4]anugrayani@unhas.ac.id
[5]alieffahdal@unhas.ac.id
[6]kayyumma18d@student.unhas.ac.id
*Corresponding author

Received: 28/7/2024 Revised: 14/1/2025 Accepted: 15/1/2025 Published: 31/1/2025

ABSTRACT

Located in the subduction zone of four tectonic plates, the high occurrence of seismic events is a severe threat in Indonesia. Mitigating the adverse effects of such disasters is essential to forecast the likelihood of future earthquakes. Consequently, developing a robust method of forecasting future earthquakes is critical to facilitate prevention and mitigation efforts. A reliable earthquake prediction method is necessary to reduce the after-effects to the greatest extent possible. This study utilises historical seismic and proposes innovative data pre-processing methods using K-means clustering to build a Long Short-Term Memory (LSTM) model for earthquake forecasting to overcome high-disparity locations. Four LSTM layers are embedded with adjusted fine-tuned network hyperparameters to enhance forecasting accuracy. The results attain 0.379816, 0.616292, and 0.414586 for Mean Square Error (MSE), Root MSE, and Mean Absolute Error, respectively, providing significant insights into earthquake prediction. In addition, predicted seismic occurrences are plotted on a map to display their geographic location within the specified research region. This research provides significant value in facilitating the efficient distribution of resources, such as evacuating residents impacted by earthquakes or reinforcing buildings and infrastructure, for emergency responders and policymakers.

Keywords: Deep learning, earthquake forecasting, LSTM, spatial distribution, seismology.

INTRODUCTION

According to the global database of the 21,000 most catastrophic disasters since 1900, 50% of earthquakes that generated the highest number of injuries occurred within the last two decades (Nkurunziza et al., 2022). In that period, earthquakes have contributed to six deadly disasters and 21% of the economic losses. The circum-Pacific seismic belt is the optimal geographical site to research earthquakes and related hazards, including tsunamis, floods, and landslides, as it is the site of more than 80% of the world's most significant earthquakes (Zhang et al., 2018). Numerous incidents of buildings collapsing, damaged infrastructure, and loss of life have been documented as a result of the devastating impact of these earthquakes, which occur frequently (Roque et al., 2024). As a part of the Pacific Rim belt, Indonesia is experiencing a series of catastrophic disaster events that lead to immense destruction, disruption, and loss of life to people's daily lives and livelihood (Dartanto, 2022; Fuady et al., 2021; Haryanto et al., 2020).

Approximately 220,000 Indonesians, primarily in Aceh, claimed their lives in one of the most lethal natural disasters in recent history (Ismail et al., 2018). The 2006 Yogyakarta earthquake, with a magnitude of 6.3, caused the death of more than 5,700 individuals and inflicted damage upon residences, roads, utilities, and other essential infrastructure in densely populated Central Java. As a result, numerous people were left without fundamental requirements and the resources to reconstruct their lives and businesses. The 2009 Padang earthquake, which had a magnitude of 7.6, struck the city of Padang in West Sumatera. It resulted in the death of more than 1,100 individuals and left hundreds requiring medical attention (Putra, 2020). Recently, the 2018 Sulawesi disasters, a catastrophic combination of seismic and hydrological disasters, served as another tragic episode of an extensive chronicle of natural calamities in Indonesia, resulting in a death toll of over 4,300 and extensive destruction to infrastructure and the local economy.

The increased capabilities of machine learning models have generated massive interest in utilising these algorithms for earthquake forecasting. This forecasting is challenging to solve using traditional analytical techniques as complex patterns with extensive seismic data must be detected. Many scholars have proven the performance of various machine learning techniques in predicting earthquake activities (Banna et al., 2021; Bhatia et al., 2023; Chomchit & Champrasert, 2022; Murwantara et al.,2020; Schäfer & Wenzel, 2019). These models improve the efficiency and speed of seismic monitoring and early warning systems by automatically classifying characteristics.

The studies above comprehensively compare machine learning applications across domains, including time series forecasting and sequence prediction. Nevertheless, this literature identifies some significant research gaps linked to the proposed study, as follows:

1. The studies focus on specific datasets or applications, suggesting it is essential to assess the effectiveness of the proposed methodologies across various domains;
2. The interpretability of pre-processing might not be thoroughly addressed, highlighting a need for methods that enhance the model;
3. The impact of diverse data quality and pre-processing methods on the model appears unexplored, suggesting further investigation into data handling practices.

To overcome the technical limitations and gaps of prior studies, this research contribution addresses the following aspects:

1. Utilises K-Means clustering during data pre-processing to slice the dataset into geographic clusters to generate location-specific predictions;
2. Enhance the pre-processing techniques with depth binning and scaling to manage the data disparity from the dataset;
3. Compares the results with certain literature sources;
4. Visualise predicted earthquakes on a map, facilitating aid in promptly identifying emergency responders and policymakers.

This study also employs the Long Short-Term Memory (LSTM) algorithm for the model. This model is a Recurrent Neural Network (RNN) that is designed to analyse temporal patterns and the spread of seismic waves to enhance the method of earthquake forecasting. LSTMs are ideal for earthquake prediction because they can accurately capture the temporal and spatial patterns in seismic data. These models have the ability to acquire knowledge and retain information about the extended connections and complex associations within the data, resulting in earthquake predictions that are more precise and dependable in comparison to the other machine learning techniques (Banna et al., 2021).

RELATED WORKS

A reliable prediction of earthquakes, which has been considered significant for a long time, is a viable challenge. Scholars and intellectuals worldwide have been interested in investigating a roadmap to predict this precarious process. Short-term and medium-term prediction of earthquakes remains elusive; researchers have made significant progress in essential advances to understand the complexities of seismic processes that lead to this natural disaster and to develop probabilistic models to assess seismic risks (He et al., 2022; Kachakhidze et al., 2024; Laurenti et al., 2022; Mukherjee et al., 2022; Murwantara et al., 2020). Further, scientists are utilising massive datasets and advanced computer techniques to enhance the precision and dependability of forecasting models capable of considering the multi-faceted nature of seismicity. One of the key findings of the study performed by Tan et al. (2023) on earthquake prediction is that how it is presented can lead to unexpected changes in communication. Also, the literature on earthquake prediction and communication stresses that a softer approach should be adopted. This respect of dealing with earthquakes accepts failure to predict and the ability to have public expectations impact the disaster planning and mitigation efforts (Li et al., 2023).

Recently, potential machine learning methods have been investigated to enhance earthquake forecasting. Salam et al. (2021) developed a hybrid machine learning method to predict the earthquake magnitude in two weeks. They applied FPA-ELM, which is a combination of the Flower Pollination Algorithm (FPA) with the Extreme Learning Machine (ELM), and FPA-LS-SVM, which is a combination of FPA and the Least Square Support Vector Machine (LS-SVM). This hybrid model resulted in the proposed model outperforming the FPA-ELM model in all criteria for short-term earthquake prediction. Further, Xiong et al. (2021) employed satellite data to investigate seismic data's physical and dynamic changes and create a new machine learning technique called Inverse Boosting Pruning Trees (IBPT). This technique is used to make short-term predictions using satellite data from 1371 earthquakes with a magnitude of six or higher, as these earthquakes significantly affect the environment. The IBPT framework outperformed the selected state-of-the-art methodologies and was the best across all benchmarking datasets.

Furthermore, scholars have employed LSTM to examine the sequential and geographical patterns in intricate data. LSTM models have exhibited remarkable efficacy in collecting patterns and have outperformed conventional machine learning techniques as predictive models for data. LSTM overcomes the vanishing gradient problem, which often occurs in RNNs. An LSTM model has a memory cell that maintains information over a long period. This activity allows the LSTM to aggregate relevant knowledge from previous inputs and retain it effectively (Sherstinsky, 2020). An advantage of LSTM is its capacity to choose, retain, and utilise the most pertinent and consequential information via its memory block. Therefore, LSTM is highly efficient in representing data with long-term dependencies and significant complexity (Ali Khumaidi & Nirmala, 2022).

In their study, Jung and Oh (2023) implemented a classification model using Deep Neural Network (DNN), RNN, and LSTM to distinguish high and low concentrations of particulate matter in Korea. They also created distinct models that accurately forecasted both low and high levels of particulate matter. The study performance evaluation results of the prediction showed that based on the concentration, the proposed separation prediction model had slightly lower RMSE and Mean Absolute Percentage Error (MAPE) values than the individual neural network models, as indicated by decimal figures. In seismology, Banna et al. (2021) utilised the LSTM model for earthquake occurrence and location. This proposed model used seismic indicators extracted from Bangladesh's earthquake database as input characteristics to forecast earthquakes occurring in the subsequent month. An attention mechanism was included in the LSTM architecture to enhance the accuracy of earthquake prediction in the model, resulting in a 74.67% accuracy rate.

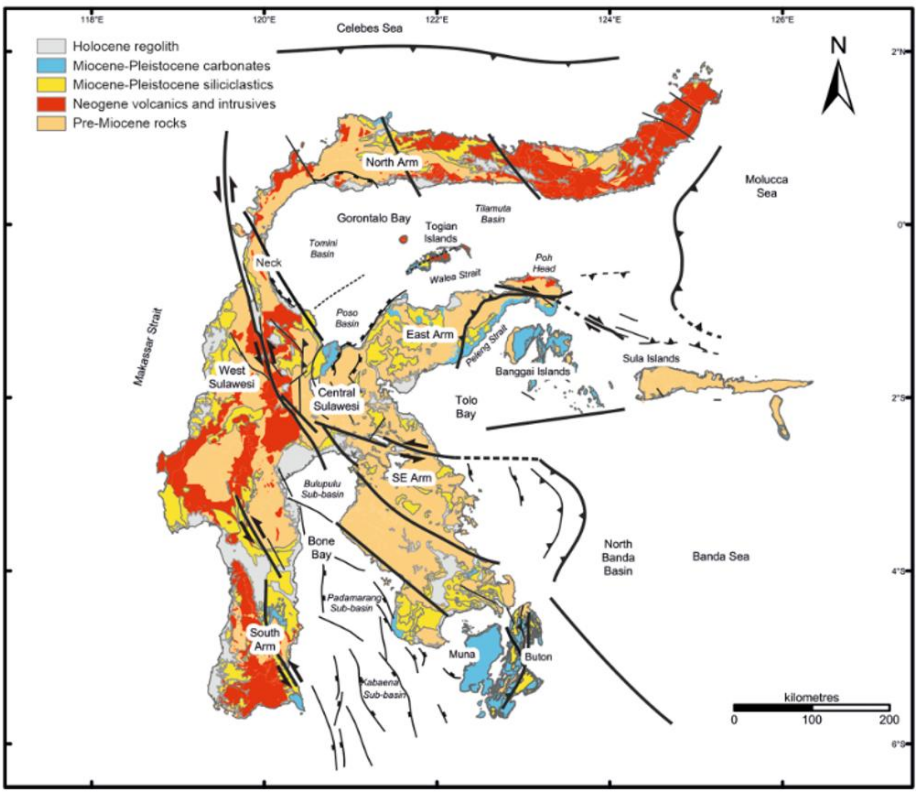
Upon the limitation of the aforementioned related works, this research performed various pre-processing approaches, including depth binning, K-means clustering, and scaling, to mitigate the extensive geolocation disparity combined with the LSTM algorithm for the machine learning model.

THE PROPOSED METHOD

The case study of this research is Sulawesi Island, Indonesia (the latitude is between 6.5^{o} S to 2^{o} N, and the longitude is between 118.5^{o} W to 126^{o} E). Sulawesi is located in the central part of Indonesia and is one of the largest islands in the area. Situated at the junction of three tectonic plates (Pacific Plate, Indo-Australian Plate, and Eurasian Plate), this area experiences more quakes than other regions (Baillie & Decker, 2022). Sulawesi Island also has two active faults, the Palu-Koro and Matano faults. As depicted in Figure 1, the Palu-Koro fault is a very active horizontal fault that is the source of earthquakes stretching from north to south. The Matano fault, located in the southeastern part of the island, is also one of the causes of high earthquake activity.

Figure 1

Active Fault Map of Sulawesi Island (Nugraha et al., 2022)



The extensive geographical location of the case study area results in disparities in parameter values within the geolocation variable. As an initial hypothesis, this disparity will affect the model's accuracy. Furthermore, Indonesia's position among subduction, extension, thrust, and strike-slip fault zones influences the occurrence of earthquakes characterised by significant depth variability and a broad range (Hutchings & Mooney, 2021). This research addresses data pre-processing using several strategies to manage discrepancies in earthquake depth, which will be further discussed.

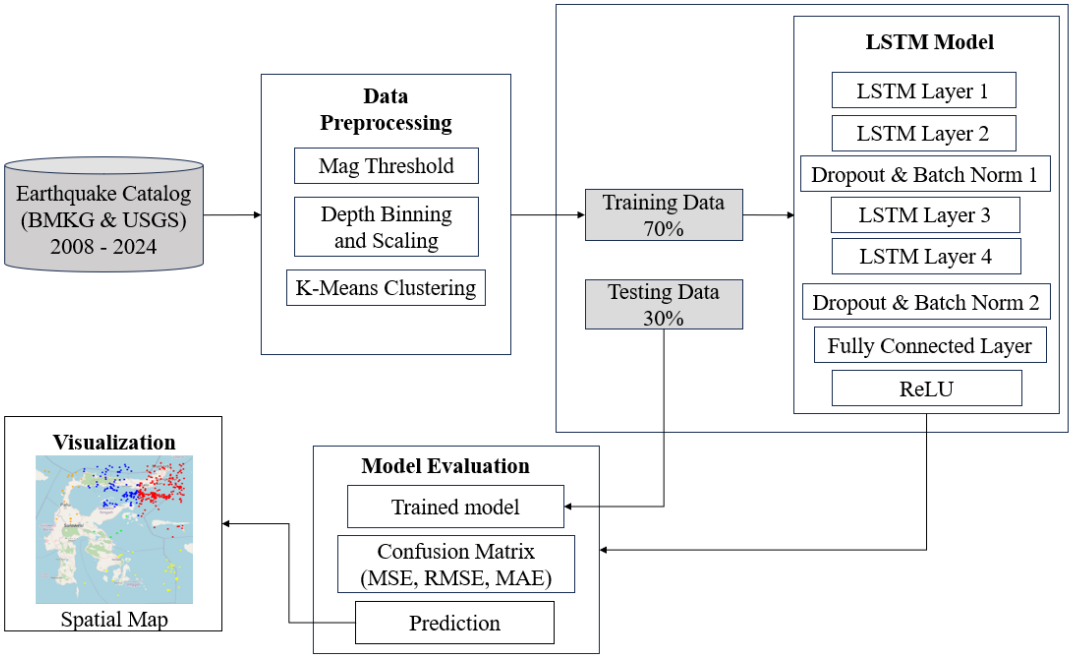
Figure 2 illustrates the proposed method of this research. The earthquake catalogue obtained from BMKG (Indonesia's Meteorology, Climatology, and Geophysical Agency) and USGS (United States Geological Survey) undergoes innovative pre-processing through various techniques to enhance model accuracy, including threshold filtering for magnitude, depth binning, and K-Means Clustering to reduce significant gaps in geolocation and depth features. These pre-processing methods also contribute as a novelty in this study. Subsequently, the dataset is partitioned into training and testing subsets, with the training subset utilised for the LSTM model. The LSTM model is adjusted in terms of layers and hyperparameters for evaluation with the testing data. The evaluation process employs a confusion matrix, which will assess Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE). The final stage of the suggested methodology involves mapping visualisation, which will enhance the precise interpretation of the data.

System Configuration

This research implemented several devices, instruments, and software, including Google Collaboratory and Jupyter Notebook 7.0.3, with the code developed in Python 3.9.1. All research procedures were executed on an Asus ROG Strix G512LI equipped with an Intel(R) Core(TM) i7-10750H CPU operating at 2.60GHz and 16.00 GB of RAM.

Figure 2

Design Method



Data Acquisition

The study exploits seismic data from BMKG and USGS as an Earthquake Catalogue. The dataset contains detailed information about all the seismic activities in Sulawesi, Indonesia, as a research area. This dataset includes the precise date and time of the earthquake (date time), the geographical coordinates of the earthquake (latitude and longitude), the vertical distance of the quake below the Earth's surface (depth), and the intensity of the earthquake (magnitude) spanning 15 years, from 2008 to early 2024. Table 1 displays the detailed samples of each historical earthquake, including its latitude, longitude, depth, and magnitude.

Table 1

Input Dataset

Date Time	Latitude	Longitude	Depth	Magnitude
1/2/2008	0.343	125.794	35.5	4.4
1/3/2008	-5.921	122.662	10	5.4
1/6/2008	-2.163	120.83	35	4.5
1/6/2008	1.184	125.374	35	3.9
1/8/2008	0.521	123.35	268.7	4.6
1/12/2008	-2.255	120.587	25	4.1
1/14/2008	0.189	122.032	214.7	4.3

Data Pre-processing

The initial stage of data pre-processing is data cleaning. This process will clean the data from empty entries, thus ensuring good data quality. This stage includes checking for empty or null data removed from the dataset to ensure data cleanliness.

Magnitude Threshold

The next pre-processing phase filters the features used as inputs in the earthquake analysis. In this stage, a filtration process was implemented based on earthquake magnitude. Table 2 explains the classification of the quakes according to their impact (Michigan Tech, 2024). This study employed earthquakes with intensity higher than magnitude 2.5 on the Richter Scale as they potentially cause harm. Whereby only data with earthquake magnitudes equal to or beyond the 2.5 threshold are preserved. Then, the data duplication is performed based on several key features such as datetime, latitude, longitude, depth, and magnitude. Prior to the removal process, duplicate data is counted to give an idea of the level of possible duplication in the dataset.

Table 2

Earthquake Strength Scale in Magnitude

Magnitude	Impact	Estimated Number Per Year
2.5 or less	Typically imperceptible but detectable by seismographs	Millions
2.5 – 5.4	Frequently experienced, although resulting in relatively little harm	500,000
5.5 – 6.0	Slight structural damage to buildings and other structures	350
6.1 – 6.9	Can inflict significant destruction in highly inhabited regions	100
7.0 – 7.9	Significant seismic event with severe harm	10-15
8.0 or higher	Significant seismic event and has the potential to devastate towns close to the epicentre	1 every 1 or 2 years

Depth Binning and Scaling

The next pre-processing stage is the earthquake depth classification process, which divides the depth into several intervals based on certain predetermined boundaries. The earthquake depth grouping boundaries, defined as bin_borders, are based on earthquake types based on their depth (Cui et al., 2022). The data augmentation technique is implemented to resolve the issue of limited data variety resulting from the earthquake classification into only three groups. This study divides each earthquake category into two more specific segments. The first depth category is defined by a value of 30 (represents an earthquake with a depth < 30 km), the second shallow category is denoted by 60 (denotes an earthquake that occurs within a depth of 30 km - 60 km), the initial intermediate category is denoted by 180 (denotes an earthquake that occurs between 61 km -180 km), the second intermediate category defined by 180 (defines an earthquake with a depth ranging from 181 km - 300 km), the next deep category is denoted by 525 (denotes an earthquake that occurs at a depth of 301 km - 525 km), then the depth of the earthquake surpasses this value, it will be classified as a second deep earthquake and is restricted by 1000.

Once the grouping boundaries have been defined, the depth categories are labelled. Following completion of the grouping procedure, these labels are employed to identify the depth categories of earthquakes. The purpose of this clustering is to simplify the analysis and allow the identification of patterns or trends that may exist in the earthquake data. Subsequently, the data is normalised using the Min-Max Scaler technique to standardise the data range to a consistent scale and eliminate outliers, simplifying the model training process. This normalisation results in a range of values from 0 to 1. The equation of the Min-Max Scaler is implemented in Equation 1 as follows:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where,
x_new = Normalisation result
x = Original data value
x_min = Minimum value in the data before normalisation
x_max = Maximum value in the data before normalisation

K-Means Clustering

The K-Means Clustering is applied to divide the case study area into regions, with six clusters selected to manage the data within each cluster. The random_state is set to 42 to guarantee the reproducibility of results, thereby establishing a consistent starting point for the random number generator when the function is re-run. Setting the random_state parameter to 42 ensures that the clustering outcomes are consistently replicated each time the function is executed. The n_init parameter was also set to 10 to obtain the precise results. During the implementation process, clusters are established by randomly selecting centroids, and the clustering procedure is repeated n_init times to achieve the most optimal results.

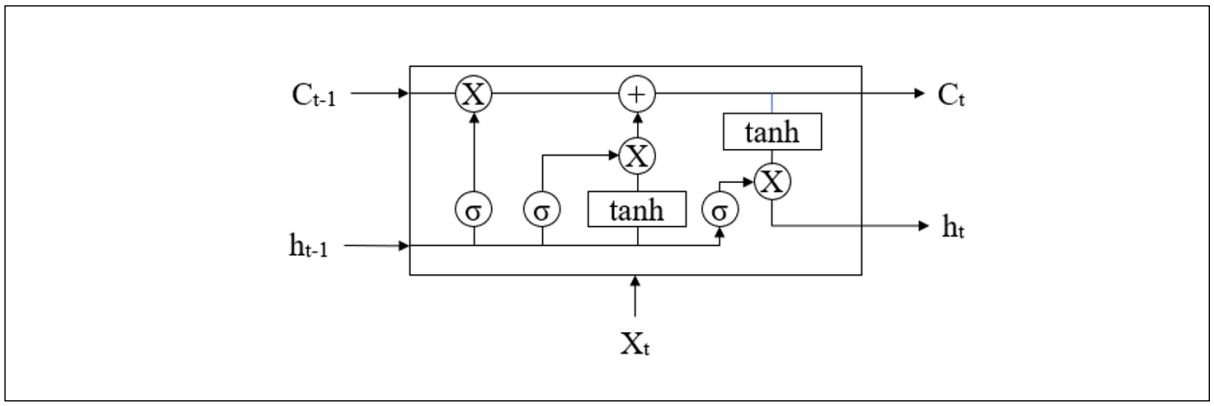
LSTM Model

This research adapts machine learning to the LSTM model. This model is a version of the RNN designed to address the problem of vanishing gradients frequently encountered in RNNs. In LSTM, a memory

cell is added to store information for a more extended period of time. This memory cell allows the LSTM to remember and retain relevant information from previous inputs. One of the advantages of LSTM is its ability to select the most relevant and significant information through its memory block. Thus, LSTM becomes effective in modelling data with long-term dependency and high complexity (Ali Khumaidi & Nirmala, 2022). The LSTM architecture consists of three major layers: the input, output, and hidden layers. The hidden layer is the central part of the LSTM, which has one or more memory cells. Each memory cell controls the information transfer very precisely and may store information for the long term or short periods, also known as forget gate (Wiranda & Sadikin, 2019). Figure 3 portrays the general architecture of LSTM, where C_{t-1} represents the Cell state at the previous time step t-1, h_{t-1} represents the Hidden state at the prior time t-1, X_t is Input data at the current time step (t), σ complies Sigmoid function, tanh equals to Hyperbolic tangent function, X represents Multiplication operation, + represents Addition operation, C_t is Cell state at the current time step (t), and h_t represents the Hidden state at the current time step (t).

Figure 3

LSTM Architecture



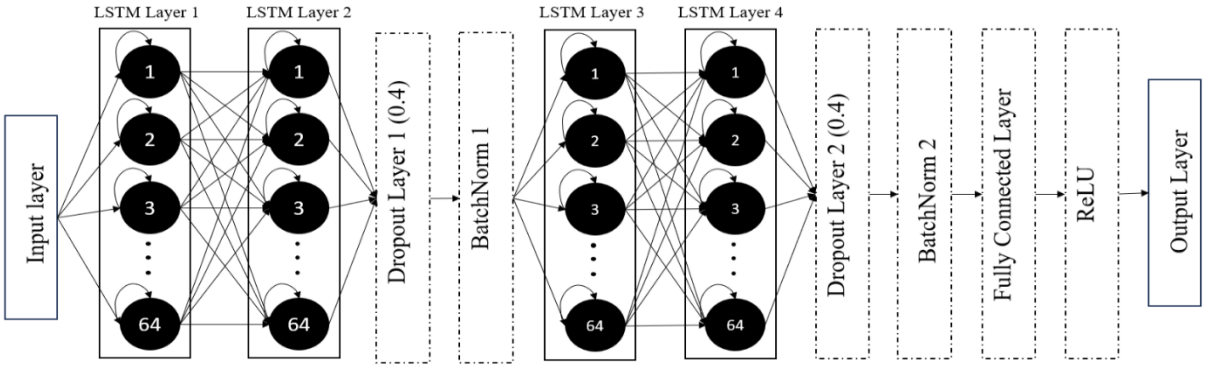
The sigmoid function is crucial in the LSTM method since it efficiently regulates information distribution inside the memory cells. This function transforms values between -1 and 1 into a sequence ranging from 0 to 1. The function also regulates the storage and updating of information in memory cells between consecutive time steps. Additionally, the tanh function is employed in the LSTM design to regulate the values that pass through the network, guaranteeing that they remain within the range of -1 to 1. Thus, the LSTM model has the capability to provide precise predictions, ensure consistency in the transmission of information within the memory cells, and efficiently process and retain information by utilising both of these activation functions.

The forget gate controls the degree to which previous information is preserved or discarded in the memory cell. The purpose of this gate is to enable the LSTM to discard information that is no longer relevant or essential for making more precise predictions. The forget gate (f_t) is a layer that applies the sigmoid function to combine the output from time t-1 with the input at time t and then applies the sigmoid activation function. The output of this gate is consistently constrained between the range of 0 and 1. If the value of ft is 0, the preceding information will be disregarded. However, if the value of f_t is 1, the preceding information will be preserved without alteration.

The model has different layers, and its neurons are arranged in a pattern. Other parameters and hyperparameters of the network need to be fine-tuned. Weights and biases in a neural network are examples of parameters, which are internal variables optimised during the training process. Hyperparameters are predetermined factors that are established before the beginning of training and stay constant throughout the training process. The hyperparameters encompass various factors, such as the lookback value, the number of layers in the model, the number of neurons in each layer (hidden_size), the learning rate (learning_rate), the batch size (batch_size), and the number of training epochs. Figure 4 describes the LSTM Architecture for the Proposed Method of this study.

Figure 4

LSTM Architecture for Proposed Method



This paper's experiments include using four LSTM layers, namely self.lstm1, self.lstm2, self.lstm3, and self.lstm4. In addition, a dropout layer is included in the model, with a dropout rate of 0.4 (nn.Dropout(0.4)). Furthermore, a Batch Normalization layer is inserted between the LSTM 2 and LSTM 3 layers and between the LSTM 4 layer and the nn.linear layer. These additions are intended to mitigate the issue of overfitting. The purpose of this layer is to choose and randomly deactivate certain units throughout their training process. Incorporating the Batch Normalisation (BatchNorm) layer into the model is a crucial step to enhance the efficiency of the training process. BatchNorm is a technique that accelerates the convergence of models, reduces sensitivity to initial parameter values, and mitigates fluctuations in data distribution within each training batch.

In addition, a few other essential factors to consider, such as the lookback value and the number of steps considered in the dataset at the previous time, affect the model's ability to "look" back at the data before making forecasts. Subsequently, the model is constructed using LSTM and dropout layers, wherein the number of layers and the dropout level are determined through parameter modifications. In addition, the model incorporates linear layers and ReLU activation to introduce a nonlinear component.

There is a potential for this to improve the stability and consistency of the model's outcomes. The BatchNorm layer is subsequent by a linear layer named self.fc in the model. The variable "output_size" indicates that the output size for this stratum is 4. This number denotes the aggregate number of input attributes, which includes dimension, depth, breadth, and length. The final layer of the model incorporates non-linearity by employing a rectified linear unit (ReLU) activation function (nn.ReLU()) following the linear layer. The architecture of the LSTM model for earthquake probability prediction comprises all of these components, which collectively constitute the structure.

The constructed LSTM model undergoes two main processes: training and testing. This process is carried out for several epochs using Adam's optimisation algorithm, which is well-known in the deep learning world for its ability to optimise parameters quickly. At the same time, an "early stopping" method is applied to prevent overfitting. The fine-tuned hyperparameters were the number of LSTM layers, lookback, hidden_layer (hidden_size), and epoch. Lookback is a parameter that describes the number of steps considered at the previous time when creating the dataset, whereas, in this study, it is set to 10. With this value, the model can "look" back at 10 previous steps when making predictions. This lookback affects the model's ability to capture long-term patterns in sequential data. Choosing the correct value can considerably impact the model's performance in predicting the likelihood of an earthquake.

The hidden layer is essential for collecting sequential information in continuous data, such as time series. The hidden layer in the model reflects the latent space's dimension. In this case, the hidden_size parameter is set to either 32 or 64, indicating the number of LSTM units the model has in each layer. This size affects the capacity of the model to understand and extract patterns from complex sequence data. In addition, there were improvements in the optimisation algorithm. This study set the learning rate (lr) to 0.001. This setting significantly impacts the model's convergence rate. The model can obtain more precise results and improved convergence with Adam and the appropriate learning rate configuration.

Epochs refer to the number of iterations the model will undergo when processing the entire training dataset. In this case, the number of epochs in this research is measured using various experiments, such as 100 to 5000 epochs. This number of epochs is essential in achieving an adequate convergence rate in the model. By increasing the number of epochs, the model is more likely to comprehend complex patterns in the data. Early halting is another crucial technique. The model consistently compares the current loss with the best_loss. The best_loss is the minimum loss value retained if the current loss value is lower than the prior best_loss. This research also implements early stopping by configuring the hyperparameter for early_stopping_patience to 10 epochs. The training operation will be terminated if there is no improvement in the loss value for 10 consecutive epochs.

The techniques mentioned above are utilised to mitigate overfitting and terminate the training process once the model's performance ceases to improve. This research also employs batch size to determine the number of samples handled simultaneously in one training iteration. The batch size significantly affects both the speed of training and the consumption of computer memory. Thus, this study set the batch size to 32 or 64. The model consistently modifies the parameters based on the training data during the training process to improve its performance. This modification involves systematically processing data in segments, computing error magnitude, and adjusting parameters at each iteration. By iterating through this process over a certain number of epochs, the model may get a deeper understanding of intricate patterns within the data, leading to enhanced accuracy in its predictions. The investigations indicate that the parameter and hyperparameter combinations with the lowest evaluation results are lookback 10, hidden_size 64, epoch 5000, and batch_size 64.

EVALUATION AND RESULTS

The model is evaluated on pre-split test data following the completion of the training process. The model's forecast outputs are analysed and evaluated to ascertain the ability to generalise and perform effectively on unseen data. The model's forecast results on training and test data are assessed using a variety of metrics, comprising RMSE, MAE, and MSE. The MSE determines the degree of agreement between the forecast results and the actual value. RMSE is also used as a standard metric to evaluate prediction models. Technically, RMSE is a square root of MSE. Although it is just a square root operation, it is significant because RMSE corrects for the effect of doubled numbers produced by MSE and gives a more normalised and understandable result. MAE also plays a vital role in seeing the average value of the error between actual and predicted values. MAE is used to measure the degree of prediction inaccuracy in time series analysis (Wang et al., 2021).

The outcome of this assessment will ascertain the model's ability to predict an earthquake's magnitude, depth, latitude, and longitude. According to Davtalab (2023), Table 3 presents the evaluation results for each location, categorised as "very good". Meanwhile, the assessments were carried out by conducting a series of experiments that included changing several parameters, such as the lookback_training (number of time steps), hidden_size (number of units in the LSTM layer), num_epochs (number of training epochs), and batch_size. The layer dropout parameter for this experiment was also set at 0.4 to reduce the possibility of overfitting in the model. The final results presented in the table are the experiment results with the lowest evaluation value of various parameter combinations, namely lookback_training of 10, hidden_size of 64, num_epochs of 5000, and batch_size of 64.

Table 3

Average Evaluation for all Variables per Region

Region	MSE	RMSE	MAE
0	0.860483	0.927622	0.713635
1	0.464615	0.681627	0.430983
2	0.918272	0.958265	0.734752
3	0.511674	0.715314	0.470693
4	0.379816	0.616292	0.414586
5	1.942492	1.393733	0.99777

Figure 5 depicts that throughout the training process, there is a gradual decrease in the loss as the number of iterations (epochs) increases. The loss decreases gradually as the number of iterations (epochs) increases, illustrating the model's ongoing enhancement in its ability to understand the patterns in the training data as time progresses. Each region experiences a different number of epochs because the predefined early-stopping mechanism often stops training. This condition is also influenced by the convergence of the model that has been achieved, although the number of epochs is lower compared to other regions. The comparison graph between actual data and forecast results provides a visual understanding of how much the model can replicate the patterns present in the sequential data. Before displaying, data filtration is performed by only displaying earthquake data with a magnitude equal to or greater than 4.0. Figures 6 (a to f) display a comparison graph of actual data and forecasted data from the training results of each region, with the magnitude of each region starting from Region 0 and moving up to Region 5.

Figure 5

Training Loss for Each Region

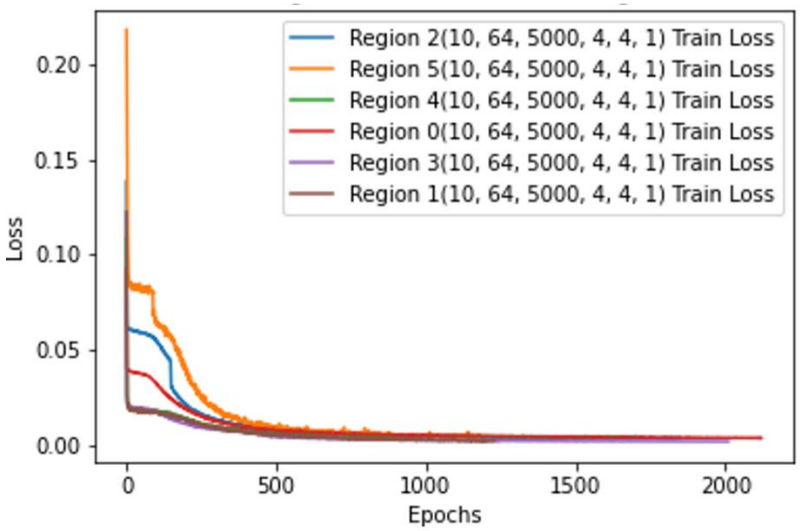
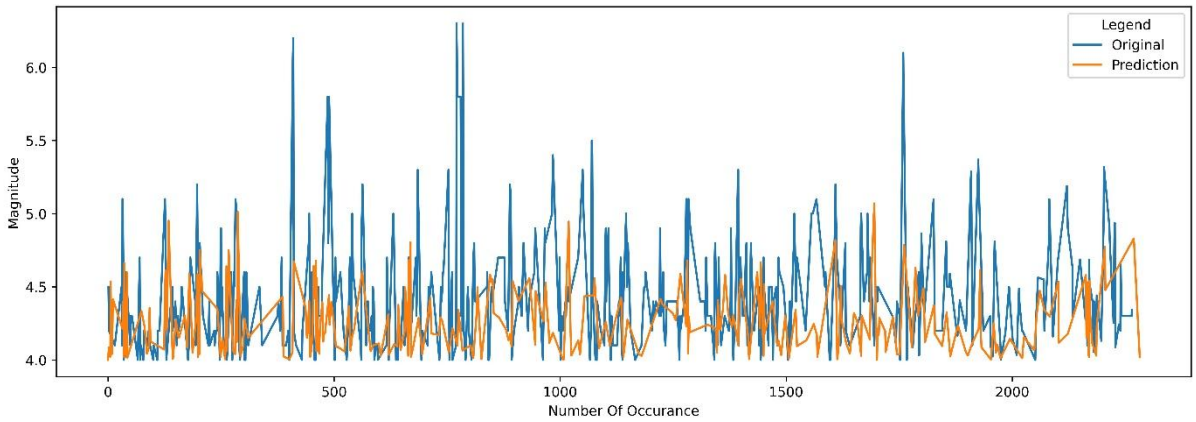
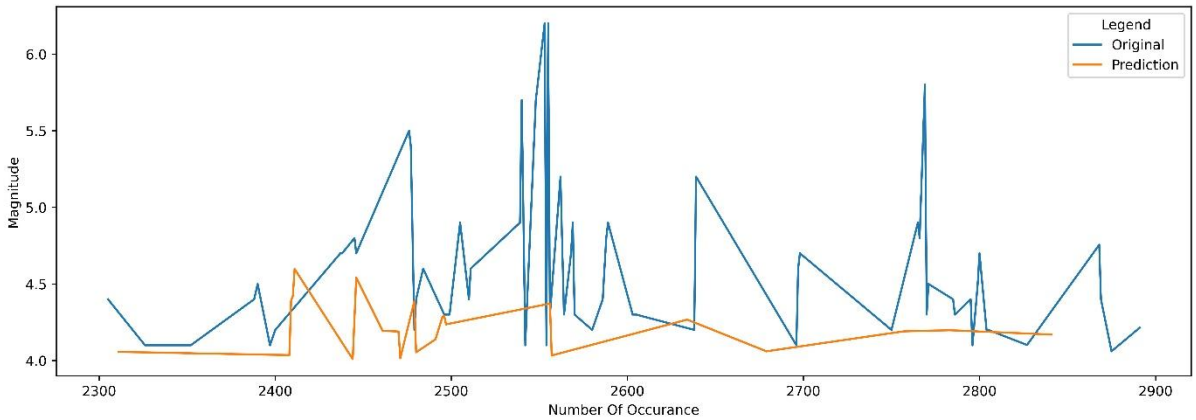


Figure 6

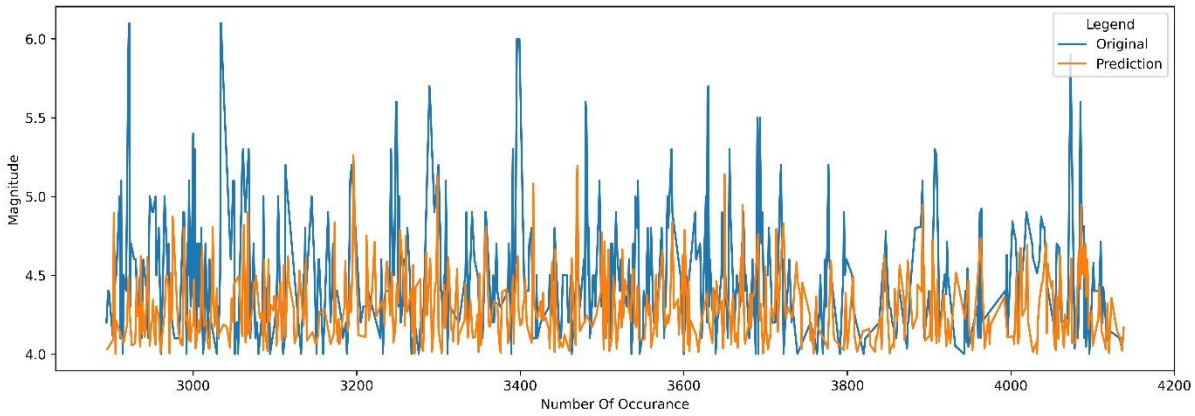
Comparison of Actual and Predicted Earthquake Magnitude Data Based on Regions



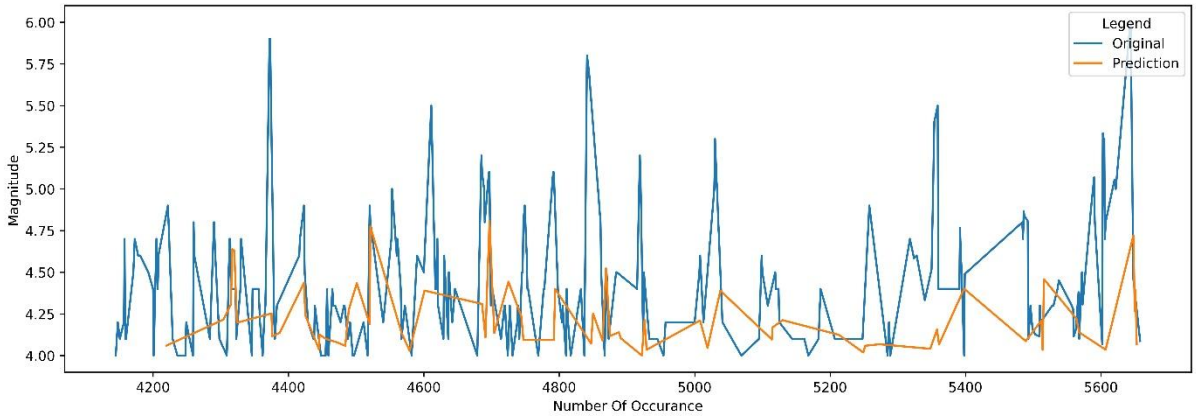
(a) Region 0



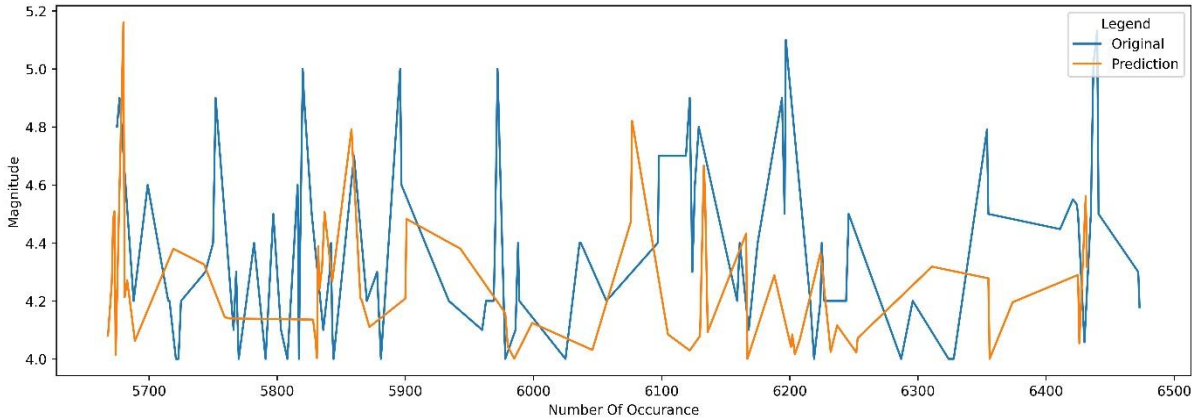
(b) Region 1



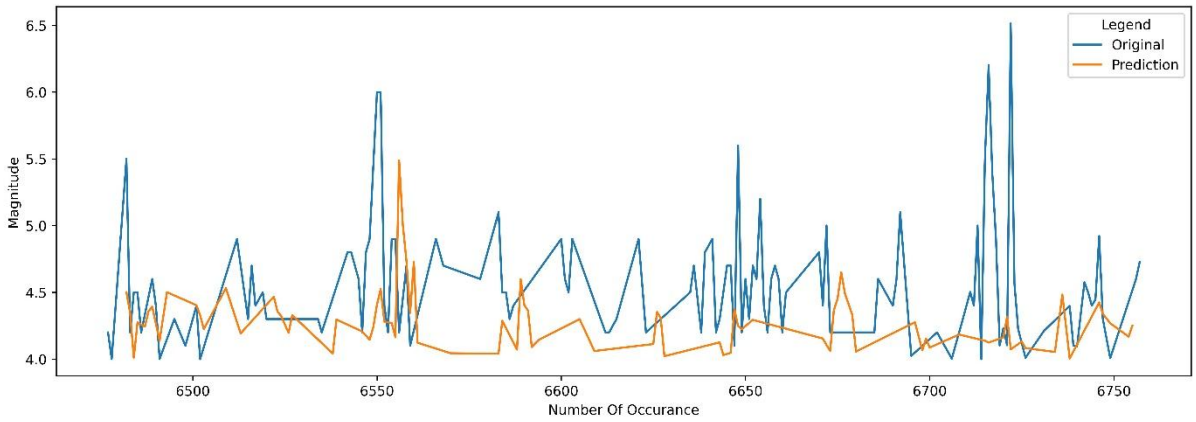
(c) Region 2



(d) Region 3



(e) Region 4



(f) Region 5

From Figure 6, the model can follow the data trend through training. The magnitude parameter of each parameter provides stable evaluation results, where the MSE, RMSE, and MAE obtained range from 0.49 to 0.81. Table 4 shows the detailed evaluation of each variable from all regions. From this table, region 4 has the best metric value. Data analysis was conducted thoroughly, and it was found that the training data of earthquake events in Region 4 tended to be uniform, where most of the earthquakes that occurred were shallow earthquakes with average magnitudes ranging from 3 to 4 on the Richter Scale. This condition allows the model to quickly identify this region and predict future earthquake events more accurately.

Table 4

Evaluation Metrics for Each Variable in Each Region

Region	Variable	MSE	RMSE	MAE
0	Magnitude	0.591329	0.768979	0.611358
1		0.552555	0.74334	0.561848
2		0.526931	0.7259	0.581129
3		0.501177	0.707939	0.564438
4		0.497223	0.705141	0.571101
5		0.814602	0.902553	0.759535
0	Depth	1.68718	1.298915	1.023679
1		0.098098	0.313205	0.067415
2		1.798241	1.340985	1.05986
3		0.329706	0.5742	0.175797
4		0.05686	0.238454	0.02225
5		3.25162	1.803225	1.121396
0	Latitude	0.525441	0.724873	0.565164
1		0.913579	0.955813	0.691434
2		0.677822	0.8233	0.621143
3		0.974679	0.987258	0.769179
4		0.261013	0.510894	0.378209
5		1.033347	1.016537	0.817086
0	Longitude	0.637982	0.798738	0.65434
1		0.294228	0.542428	0.403234
2		0.670093	0.818592	0.676876
3		0.241132	0.491052	0.373358
4		0.704167	0.839147	0.686783
5		2.670401	1.634136	1.293063

Comparing results with earlier studies in earthquake prediction is crucial due to the absence of a standardised dataset and the variability in performance metrics employed. Table 5 presents the results of prior research employing different machine learning methods compared to this study. Data analysis indicates that the LSTM model, enhanced by innovative pre-processing methods, outperformed other experimental scenarios and existing literature methodologies. The implementation method enables us to align each intended output with suitable hyperparameters based on its characteristics and complexity.

Table 5

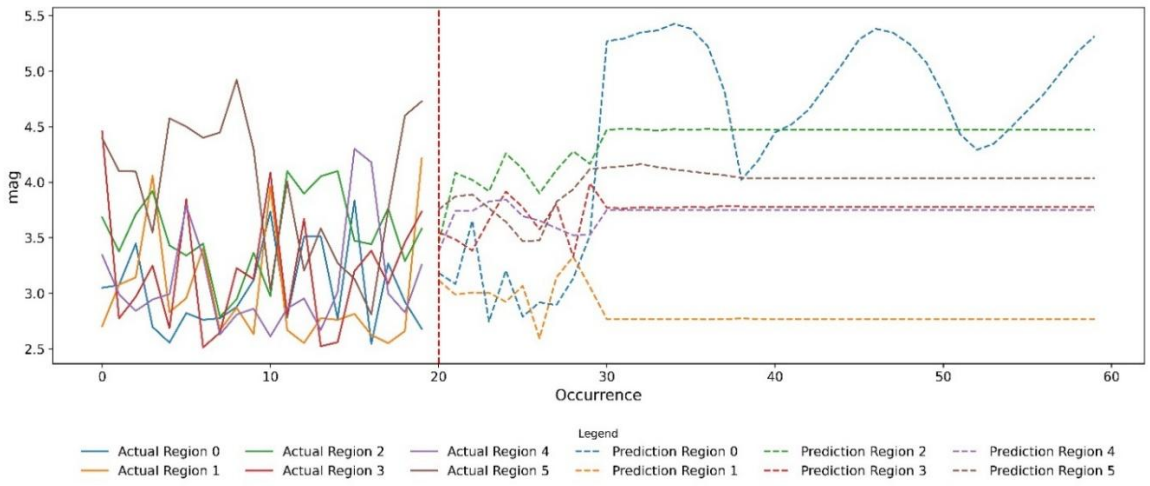
Performance Results from Literature Compared to the Proposed Study

	Model	MSE	RMSE	MAE
Murwantara et al. (2020)	Multinomial Logistic Regression	0.604094	0.777235	0.61487
	SVM	0.564013	0.751008	0.598473
	Na¨ıve Bayes	0.851585	0.922814	0.716253
Banna et al. (2021)	Attention-based Bi-directional Long-Short Term Memory	1.557900	1.248180	Undefined
Bhatia et al. (2023)	Adaptive Neuro-Fuzzy Inference System (ANFIS)	Undefined	0.571000	0.763000
Proposed Method	K-Means Clustering and 4 layers LSTM	0.379816	0.616292	0.414586

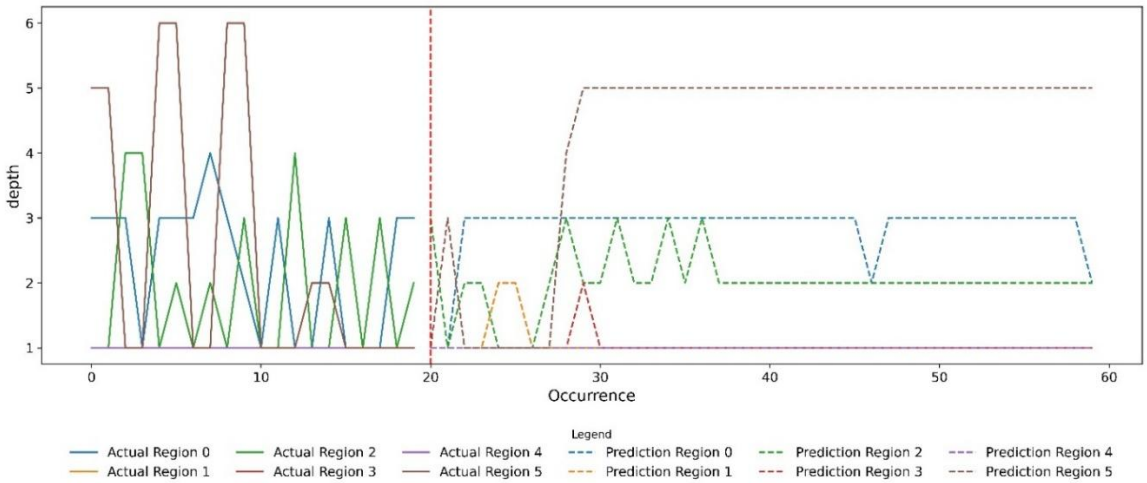
Furthermore, the results of plotting earthquakes on actual data and model training outcomes are also available to enhance comprehension of the accuracy with which the forecast model can reproduce earthquake patterns. The earthquake markers on the map accurately depict the precise geographical location of each seismic event. This visualisation is a concrete example of the system's ability to forecast the location of earthquakes in a variety of geographical regions. The subsequent results illustrate the representation of earthquake coordinates by utilising data obtained from model training. The experiment involves predicting 240 forthcoming earthquake occurrences. The graphs below serve as a visualisation of the tabular data that was previously provided. Figure 7 and Figure 8 illustrate the progression of specific values over a period of time. The vertical axis of the graph reflects a specific value of magnitude, depth, latitude, and longitude, respectively, while the horizontal axis shows the quake occurrences. The graph encompasses the whole data period indicated in the blind prediction, from start to finish.

Figure 7

The Magnitude and Depth Forecast Results Graph of 240 Predicted Earthquake Occurrences for All Regions



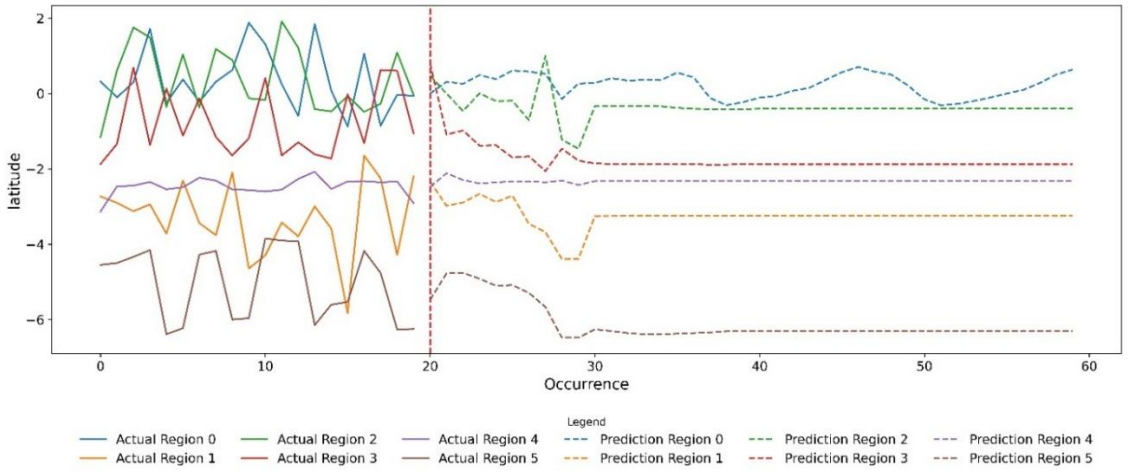
(a) The Magnitude



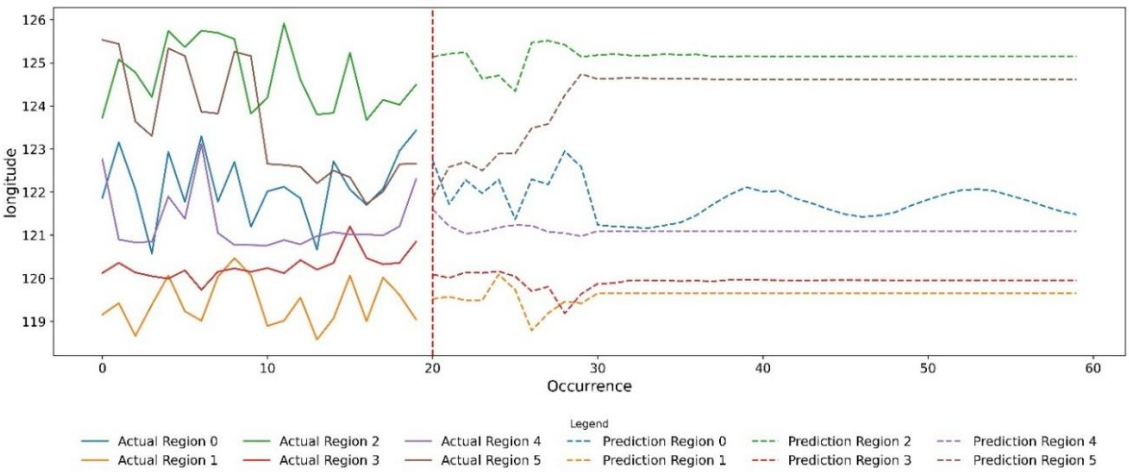
(b) The Depth

Figure 8

The Latitude and Longitude Forecast Results Graph of 240 Predicted Earthquake Occurrences for All Regions



(a) The Latitude

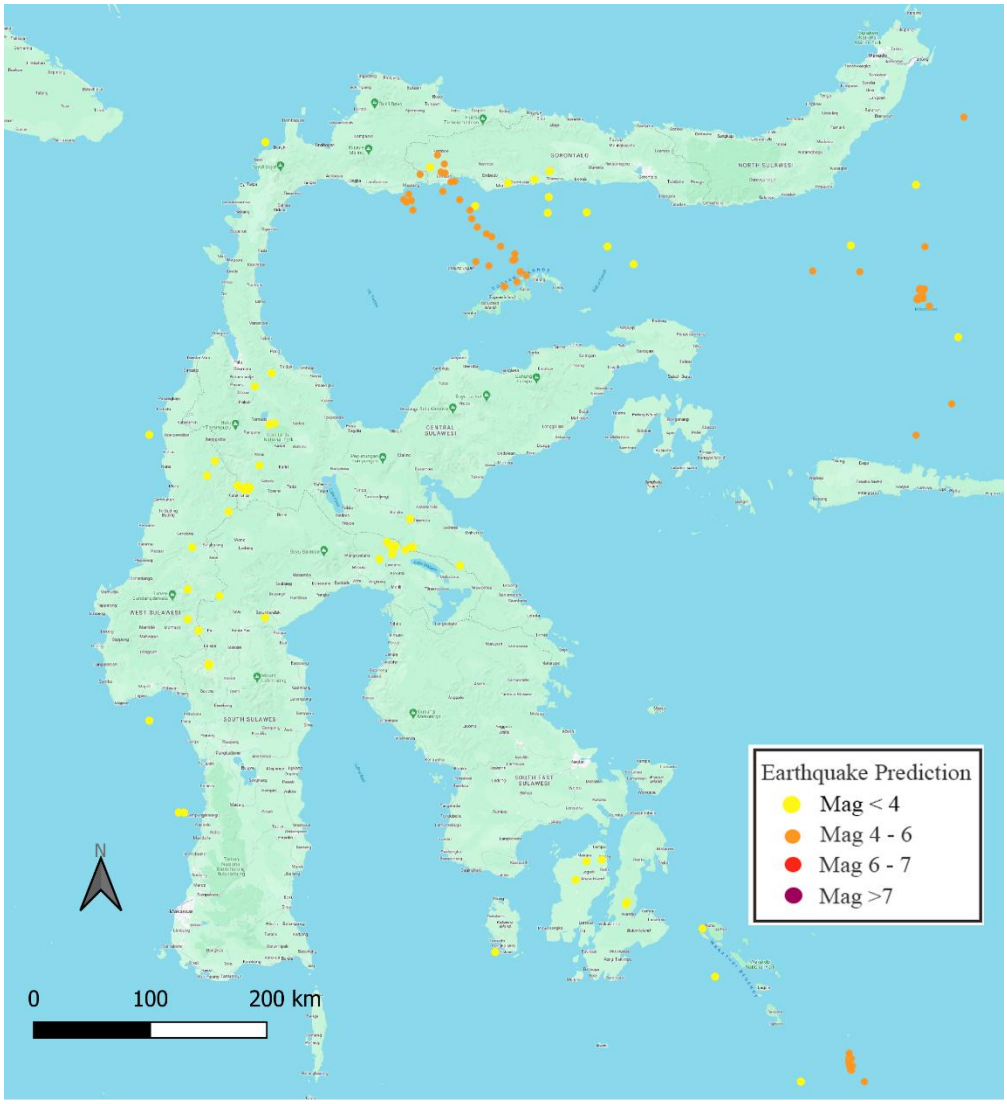


(b) The Longitude

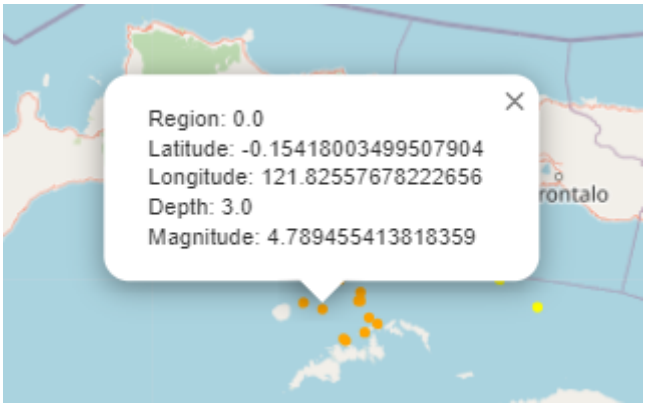
These graphs display the trends or patterns from the fluctuations in each predicted variable throughout time. The provided graphs represent the latitude, longitude, depth, and magnitude data obtained from the forecast. The displayed data includes the most recent 240 data points incorporating the raw data that enable the identification of the pattern in its emergence. The patterns produced by both information sets can be analysed by comparing the actual data with the predictions provided without any prior information. This data enables us to determine if the blind prediction outcomes exhibit a similar pattern to the actual data or whether a substantial disparity exists between them. Hence, the precision of the model in forecasting the probability of earthquakes can be assessed by utilising the existing historical data. Alongside the graph, Figure 9 also displays the outcomes of the earthquake data plotted on a map. In the prior earthquake point charting, the colours and sizes of the points have the same significance. They reflect the magnitude level and depth of the earthquake.

Figure 9

The outcomes of the Earthquake Data Plotted on a Map



(a) Visualisation of seismic event locations based on blind prediction outcomes



(b) Information attached to each occurrence

The earthquake prediction model, utilising LSTM, K-Means clustering, and a four-layer LSTM architecture, demonstrates promising performance metrics. These results present satisfactory accuracy in forecasting earthquake occurrences, particularly considering the intricacies of seismic data. The MSE values indicate that the model proficiently reduces the squared discrepancies between anticipated and actual values, which is essential for accurately capturing the intricacies of earthquake data. The RMSE values corroborate this result as a definitive measure of average prediction error. A reduced RMSE signifies that the model's predictions align closely with actual data, as seen in Figures 6 to 8. These discoveries are essential in earthquake applications since accurate forecasts aid in reducing catastrophe risks. The significance of MAE lies in its provision of a precise statistic for understanding the average error magnitude. This finding indicates that, on average, the model's predictions are adequate for early warning systems in earthquake forecasting, where even little discrepancies can significantly impact decision-making processes.

Furthermore, combining K-Means clustering with depth binning in the pre-processing phase signifies a noteworthy advancement in handling seismic datasets. The significance of K-Means clustering resides in its ability to deepen the model's understanding of specific patterns identified by the regions, which is essential considering Indonesia's varied geological landscape. Clustering decreases noise and enhances incoming data's clarity, enhancing predictive capabilities. Depth binning proves to be beneficial for handling the diverse depths associated with earthquakes. Utilising depth bins allows the model to more accurately comprehend the correlation between seismicity and depth, thereby avoiding the oversimplifications that would result from uniformly addressing all depth ranges. Stratified sampling enhances the model's understanding of complex patterns, augmenting its predictive abilities.

CONCLUSION

Earthquake datasets are intricate and intricate collections of information. They necessitate suitable and robust algorithms for their training. This article utilises hyperparametric optimisation in the proposed system for earthquake forecasting. An LSTM model has been developed to predict the probability of earthquakes in Sulawesi Island and its neighbouring areas. The LSTM architecture has many layers, including LSTM layers with diverse LSTM units, a dropout layer, Batch Normalization (BatchNorm), and a linear layer with ReLU activation function. In addition, modifications have been made to the parameters and hyperparameters of the model, including the number of LSTM layers, lookback, hidden_size, number of epochs, and optimisation method, to improve its performance.

The optimisation above implies a significant result of the model. According to the assessment findings, the model has significant results in forecasting the probability of an earthquake. The average evaluation in the MSE is 0.379816, the RMSE is 0.616292, and the MAE is 0.414586. This model outperformed other experimental scenarios and existing literature methodologies. Modifying the hyperparameters of the LSTM model and leveraging specific data pre-processing procedures can enhance the model's accuracy in forecasting earthquake events. However, LSTM has limitations in event prediction, particularly with rare event forecasting due to class imbalance. The model's lack of interpretability may impede comprehension of seismic occurrences. Moreover, the model depends on the accessibility and quality of historical data, which can differ markedly between locations.

In order to overcome the limitations above, future research must conduct further investigations or include alternative algorithms for comparison with the current LSTM model. It is necessary to conduct a comparative analysis of several algorithms to evaluate the performance of the LSTM model and aid in choosing the most effective model for earthquake forecasting. Additional pre-processing methods may concentrate on refining the number of clusters within K-Means, investigating various bin sizes for depth, and delving into alternative deep learning architectures to augment prediction accuracy. Further, extending the duration of data collection is expected to reveal long-term patterns and trends that may need to be more evident in shorter timeframes.

ACKNOWLEDGMENT

This work was supported by the Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Hasanuddin through a research grant (No: 00310/UN4.22/PT.01.03/2024).

REFERENCES

Ali Khumaidi, & Nirmala, I. A. (2022). *Algoritma long short term memory dengan hyperparameter tuning: Prediksi penjualan produk*.

Baillie, P., & Decker, J. (2022). Enigmatic Sulawesi: The tectonic collage. *Berita Sedimentologi*, *48*(1), 1–30. https://doi.org/10.51835/bsed.2022.48.1.388

Banna, M. H. Al, Ghosh, T., Nahian, M. J. Al, Taher, K. A., Kaiser, M. S., Mahmud, M., Hossain, M. S., & Andersson, K. (2021). Attention-based bi-directional long-short term memory network for earthquake prediction. *IEEE Access*, *9*, 56589–56603. https://doi.org/10.1109/ACCESS.2021.3071400

Bhatia, M., Ahanger, T. A., & Manocha, A. (2023). Artificial intelligence based real-time earthquake prediction. *Engineering Applications of Artificial Intelligence*, *120*(September 2022), 105856. https://doi.org/10.1016/j.engappai.2023.105856

Chomchit, P., & Champrasert, P. (2022). *Strong-motion earthquake prediction model using convolutional extreme learning machine*. 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 756–759. https://doi.org/10.1109/ITC-CSCC55581.2022.9894973

Cui, X., Li, Z., & Hu, Y. (2022). Similarity of shallow and deep earthquakes in seismic moment release. *Nature Geoscience*, 0–32.

Dartanto, T. (2022). Natural disasters, mitigation and household welfare in Indonesia: Evidence from a large-scale longitudinal survey. *Cogent Economics and Finance*, *10*(1). https://doi.org/10.1080/23322039.2022.2037250

Davtalab, R. (2023). *What's the acceptable value of Root Mean Square Error (RMSE, Sum of Squares due to Error (SSE) and adjusted R-square*. https://www.researchgate.net/post/Whats-the-acceptable-value-of-Root-Mean-Square-Error-RMSE-Sum-of-Squares-due-to-error-SSE-and-Adjusted-R-square/63cf02d06e7af9e6ee0db6d4/citation/download

Fuady, M., Munadi, R., & Fuady, M. A. K. (2021). Disaster mitigation in Indonesia: between plans and reality. *IOP Conference Series: Materials Science and Engineering*, *1087*(1), 012011. https://doi.org/10.1088/1757-899x/1087/1/012011

Haryanto, B., Lestari, F., & Nurlambang, T. (2020). *Extreme events, disasters, and health impacts in Indonesia bt - extreme weather events and human health: International case studies* (R. Akhtar (ed.); pp. 227–245). Springer International Publishing. https://doi.org/10.1007/978-3-030-23773-8_16

He, M., Ren, S., & Tao, Z. (2022). Cross-fault Newton force measurement for earthquake prediction. *Rock Mechanics Bulletin, 1*(1), 100006. https://doi.org/10.1016/j.rockmb.2022.100006

Hutchings, Sean, & Mooney, Walter. (2021). The Seismicity of Indonesia and Tectonic Implications. *Geochemistry, Geophysics, Geosystems. 22.* https://doi.org/10.1029/2021GC009812

Ismail, N., Okazaki, K., Ochiai, C., & Fernandez, G. (2018). Livelihood strategies after the 2004 Indian Ocean tsunami in Banda Aceh, Indonesia. *Procedia Engineering*, *212*, 551–558. https://doi.org/10.1016/j.proeng.2018.01.071

Jung, Y. Jin, & Oh, C. Heon. (2023). Concentration separation prediction model to enhance prediction accuracy of particulate matter. *Journal of Information and Communication Technology, 22*(1), 77–96. https://doi.org/10.32890/jict2023.22.1.4

Kachakhidze, M., Kachakhidze-murphy, N., Ramishvili, G., & Khvitia, B. (2024). Earthquakes three-stage early warning and short-term prediction. *Prevention and Treatment of Natural Disasters, 3*(1), 84–94.

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

Li, X., Wei, Y., Ming, Z., Cong, H., Zheng, X., & Chang, Q. (2023). Reliability prediction and evaluation of communication base stations in earthquake prone areas. *Scientific Reports, 13*(1), 1–9. https://doi.org/10.1038/s41598-023-35841-x

Michigan Tech. (2024). *Earthquake magnitude scale*. https://www.mtu.edu/geo/community/seismology/learn/earthquake-measure/magnitude/

Mukherjee, S., Gupta, P., Sagar, P., Varshney, N., & Chhetri, M. (2022). A novel Ensemble Earthquake Prediction Method (EEPM) by combining parameters and precursors. *Journal of Sensors*, 2022. https://doi.org/10.1155/2022/5321530

Murwantara, I. M., Yugopuspito, P., & Hermawan, R. (2020). Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data. *Telkomnika (Telecommunication Computing Electronics and Control), 18*(3), 1331–1342. https://doi.org/10.12928/TELKOMNIKA.v18i3.14756

Nkurunziza, J. M. V., Udahemuka, J. C., Umutesi, F., & Dusenge, J. B. (2022). Earthquake early warning system: A Solution for life rescue in health facilities and risks mitigation for the population of the Virunga Region. *Global Clinical Engineering Journal*, *5*(2), 9–28. https://doi.org/10.31354/globalce.v5i2.143

Nugraha, A. M. S., Hall, R., & BouDagher-Fadel, M. (2022). The Celebes Molasse: A revised Neogene stratigraphy for Sulawesi, Indonesia. *Journal of Asian Earth Sciences*, *228*, 105140. https://doi.org/10.1016/j.jseaes.2022.105140

Putra, R. R. (2020). Damage investigation and re-analysis of damaged building affected by the ground motion of the 2009 Padang earthquake. *International Journal of GEOMATE, 18*(66), 163–170.

https://doi.org/10.21660/2020.66.Icee2nd

Roque, P. J. C., Violanda, R. R., Bernido, C. C., & Soria, J. L. A. (2024). Earthquake occurrences in the Pacific Ring of Fire exhibit a collective stochastic memory for magnitudes, depths, and relative distances of events. *Physica A: Statistical Mechanics and Its Applications, 637,* 1–24. https://doi.org/10.1016/j.physa.2024.129569

Salam, M. A., Ibrahim, L., & Abdelminaam, D. S. (2021). Earthquake prediction using hybrid machine learning techniques. *International Journal of Advanced Computer Science and Applications, 12*(5), 654–665. https://doi.org/10.14569/IJACSA.2021.0120578

Schäfer, A. M., & Wenzel, F. (2019). Global megathrust earthquake hazard—maximum magnitude assessment using multi-variate machine learning. *Frontiers in Earth Science, 7*(June), 1–19. https://doi.org/10.3389/feart.2019.00136

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404,* 132306. https://doi.org/10.1016/j.physd.2019.132306

Tan, M. L., Vinnell, L. J., Valentin, A. P. M., Prasanna, R., & Becker, J. S. (2023). The public's perception of an earthquake early warning system: A study on factors influencing continuance intention. *International Journal of Disaster Risk Reduction, 97*(June), 104032. https://doi.org/10.1016/j.ijdrr.2023.104032

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. *Nature Communications, 12*(1), 1–11. https://doi.org/10.1038/s41467-021-27553-5

Wiranda, L., & Sadikin, M. (2019). Penerapan Long Short Term Memory pada data time series untuk memprediksi penjualan produk PT. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika, 8*(3), 184–196.

Xiong, P., Tong, L., Zhang, K., Shen, X., Battiston, R., Ouzounov, D., Iuppa, R., Crookes, D., Long, C., & Zhou, H. (2021). Towards advancing the earthquake forecasting by machine learning of satellite data. *Science of the Total Environment, 771,* 145256. https://doi.org/10.1016/j.scitotenv.2021.145256

Zhang, Y. X., Goebel, T., Peng, Z., Williams, C., Yoder, M., & Rundle, J. (2018). Earthquakes and Multi-hazards around the Pacific Rim, Vol. 1: Introduction. *Pure and Applied Geophysics*, *174*(6), 2195–2198. https://doi.org/10.1007/s00024-017-1580-4