

RESEARCH ARTICLE

Comparative Analysis of LSTM and Bi-LSTM Models for Earthquake Occurrence Prediction in Tokai-Japan Region

Azhari Haris Al Hamdi,^{*} Hapsoro Agung Nugroho, and Benyamin Kusumoputro

Department of Electrical Engineering, Universitas Mercu Buana, Depok, Indonesia

^{*}Corresponding author. Email: azhari.haris@ui.ac.id

Abstract

This study compares the performance of Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) models in predicting earthquake occurrences in the Tokai region, using data from the United States Geological Survey (USGS) dataset. Given the importance of accurate earthquake prediction, particularly in high-risk regions, this research focuses on assessing the effectiveness of each model in identifying occurrence and non-occurrence events. Both models were tuned to optimize sensitivity and specificity through adjustments in sequence length, learning rate, and additional hyperparameters, with results evaluated using metrics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the curve (AUC). Findings reveal that while both models achieved high sensitivity, the LSTM model demonstrated superior specificity and AUC, indicating a more balanced performance in distinguishing between earthquake occurrences and non-occurrences. The results show that LSTM outperforms Bi-LSTM in terms of its classification metrics. LSTM achieved an accuracy of 76%, compared to 55% for Bi-LSTM. For the AUC metric, LSTM scored 66%, while Bi-LSTM scored 67%.

Keywords: LSTM, USGS, Tokai, Earthquake, Deep Learning

1. Introduction

Earthquake prediction is a critical area of research, especially for regions like Tokai, where seismic activity poses a constant threat to communities and infrastructure. Efforts to predict earthquake probabilities in the Tokai region have been informed by advancements in paleo seismology, plate tectonics, tsunami studies, and geodetic

measurements. Rikitake [1] highlights that the mean return period of great earthquakes in the Tokai-Nankai zone is approximately 109 years, with probabilities of a significant earthquake in the Tokai district reaching 35–45% for the decade following 2000, underscoring the urgency of seismic preparedness. The Tokai region, part of the broader Nankai megathrust system, has been a key focus of seismic studies due to its history of recurring great earthquakes, with probabilities of M8-class events within the megathrust estimated to range from 4.3% to 96% over three years [2]. Historical evidence shows that while the Tokai segment did not rupture during the 1944 and 1946 megathrust earthquakes, it has experienced significant seismic activity in earlier events, such as in 1854 and 1707, indicating a history of periodic stress release along the plate boundary [3]. This study chose Tokai Region, Japan because of the availability of high-quality and dense seismic monitoring systems in Japan, including GPS geodetic networks and sensor arrays, allows for precise analysis of seismic activities in the Tokai region. Nishimura [4] detected slow slip events near the Tokai seismic gap using these systems, providing crucial data on precursory phenomena for large earthquakes. In contrast, Indonesia faces challenges in maintaining consistent and comprehensive seismic data coverage.

Accurate prediction of earthquake occurrences can enable timely preparations and potentially save lives by informing preemptive actions. However, predicting earthquakes involves complex challenges due to the multifaceted nature of seismic data, which includes factors like magnitude, depth, frequency, and regional geology. Machine learning models, specifically neural networks, have shown promise in analyzing these complex patterns, enabling data-driven approaches to earthquake prediction.

In recent years, deep learning models such as Long Short-Term Memory (LSTM) [5] and Bidirectional LSTM (Bi-LSTM) [6] have become popular in time-series analysis, including seismic forecasting. Machine learning has been extensively used to analyze complex data, including EEG signals, by employing advanced algorithms for feature extraction and classification. For example, the study demonstrates how various machine learning methods, such as Support Vector Machines and Random Forests, have been successfully applied to classify EEG-based emotional states, highlighting the versatility and adaptability of these methods [7]. In earthquake forecasting, LSTM networks have been designed to improve the accuracy and reliability of earthquake forecasting. These models excel at detecting and analyzing hidden patterns and anomalies within earthquake data, which are essential for making precise predictions in regions prone to seismic activity [8] [9]. In same context, LSTM networks are increasingly utilized in early warning systems to forecast earthquake parameters like magnitude and epicentral distance. By integrating deep learning with traditional machine learning methods, these models enhance the precision of early warnings, helping to mitigate the devastating effects of seismic events [10]. In control systems, Machine learning has been extensively used to address the challenges of modeling complex, nonlinear dynamics, such as those found in quadcopter attitude and altitude control. For example, the application of neural network architectures like Elman Recurrent Neural Networks (ERNN) has demonstrated significant improvements over traditional PID and backpropagation methods, particularly in handling time-varying

dynamics and enhancing control precision [11].

Experimental comparisons have shown that Bi-LSTM models outperform standard LSTM architectures in terms of prediction accuracy, achieving a 37.78% reduction in error rates. However, Bi-LSTM models typically require more data and training time to reach equilibrium due to their bidirectional training approach [12]. These models are particularly suited to handling temporal dependencies in data, which is crucial for recognizing patterns over time. LSTM networks are adept at retaining relevant information over extended sequences, making them highly suitable for sequential data where long-term dependencies play a significant role. Bi-LSTM models, by processing data in both forward and backward directions, are designed to capture a more comprehensive understanding of temporal dependencies, potentially improving predictive accuracy.

Despite the potential of LSTM and Bi-LSTM models, their effectiveness in earthquake prediction can vary based on how well they capture both positive cases (earthquake occurrences) and negative cases (non-occurrences). For practical applications, the balance between sensitivity and specificity is crucial. A model that excessively favors one over the other may either fail to provide early warnings or issue too many false alarms, each with significant real-world implications. Consequently, finding the right balance in model performance metrics is vital for effective earthquake prediction systems, and this study aims to compare LSTM and Bi-LSTM models in this regard.

This paper explores the performance of LSTM and Bi-LSTM models, tuned specifically for earthquake occurrence prediction, using seismic data from the United States Geological Survey (USGS) for the Tokai region. The models are evaluated across multiple metrics, including sensitivity, specificity, predictive values, and area under the curve (AUC). This study seeks to determine which model provides a better balance between accuracy and reliability, thereby offering insights into the applicability of these models in real-world seismic monitoring and prediction efforts.

2. Methodology

2.1 Data Acquisition

The area studied in this paper is in the Tokai region of Japan with period covered is from June 1975 to August 2024. The data used is sourced from USGS [13] with latitude coordinates of 33.943 – 37.475 and longitude coordinates of 135.681 – 144.097. The features considered are time, magnitude, and depth. From the established data limits, the total number of data records is 11336.

2.2 Data Cleansing

Record null data removal is carried out to prevent the inclusion of empty data in the subsequent stages. Data checks are performed to assess consistency over time. At this stage, data is filtered with the condition that the magnitude is greater than 4.4 and the depth is less than 100 km because the threshold magnitude is determined using a magnitude completeness [14], while a depth of 100 km falls into the category of shallow earthquakes that can be felt by humans. All magnitudes within the filtered range will be labeled as 1 (earthquake occurred), while those outside this range will be labeled as 0 (no earthquake occurred).

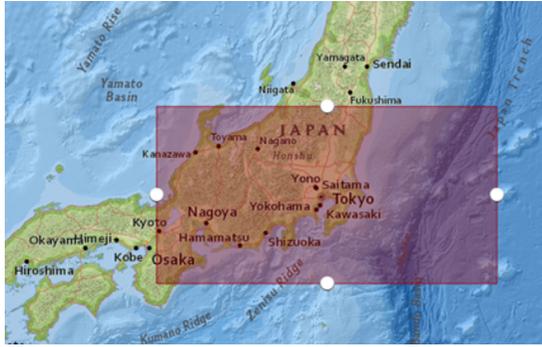


Figure 1. Area Data of Tokai, Japan

2.3 Compute Seismic Parameter

In this study, all earthquake data within the specified filter range (such as based on magnitude and depth) are sampled monthly to calculate seismic parameters. There are 8 parameters necessary calculated to analyze seismic activity [9] [15] i.e. Date Difference, Mean Magnitude, Rate of Square Root of Seismic Energy Released (RSRSER), B-Value, Mean Square Deviation, Magnitude Deficit, Mean Time Between Characteristic or Typical Events (μ value), and Aperiodicity of the mean (c value).

2.3.1 Date Difference (DD)

The difference in dates between the event (with a magnitude equal to or greater than the threshold) and the date of the initial event is defined as

$$t = t_n - t_1 \quad (1)$$

The DD value can serve as a measure of foreshock frequency depending on the threshold value chosen for magnitude [16]. A larger DD value indicates a lower probability of upcoming seismic events, while a smaller DD value suggests a higher probability of foreshock frequency. Date Difference can serve as a measure of foreshock frequency depending on the threshold value chosen for magnitude. A larger DD value indicates a lower probability of upcoming seismic events, while a smaller T value suggests a higher probability of foreshock frequency.

2.3.2 Mean Magnitude

The average of earthquake magnitudes on the Richter scale for the last n events can be calculated using the formula:

$$\bar{M} = \frac{\sum M_i}{n} \quad (2)$$

This average, along with the DD value (which measures the frequency of foreshocks), becomes an important parameter for predicting the occurrence of a major earthquake in certain regions. According to the accelerated release hypothesis and its

modifications, the energy released from a fractured fault increases exponentially as the time until the earthquake gets closer. This means that the observed magnitudes of foreshocks typically increase sharply just before a major earthquake occurs. This phenomenon reflects an acceleration pattern in energy release from the fault line, which can serve as an early warning for an impending major earthquake.

2.3.3 Rate of Square Root of Seismic Energy Released

The rate of square root of seismic energy released over a period t ($dE^{\frac{1}{2}}$) is calculated as:

$$dE^{\frac{1}{2}} = \frac{dE^{\frac{1}{2}}}{t} \quad (3)$$

where $E^{\frac{1}{2}}$ represents the square root of seismic energy E , which is derived from the corresponding Richter magnitude using the empirical relationship:

$$E = 10^{11.8+1.5M} \text{ ergs} \quad (4)$$

2.3.4 B-Value

The b-value represents the slope coefficient of the log frequency of earthquakes and is derived from the Gutenberg-Richter (G-R) inverse power law.

$$\log_{10}N = a \pm bM \quad (5)$$

The values of a and b are obtained from plotting the magnitude class values against the frequency of each class. The resulting slope becomes the value of b , and the intersection point becomes the value of a . In general, the a -value and b -value can be formulated as follows:

$$a = \frac{\sum \log_{10}N_i + bM_i}{n} \quad (6)$$

$$b = \frac{(n\sum M_i \log_{10}N_i) - (\sum M_i \sum \log_{10}N_i)}{(\sum M_i)^2 - n\sum M_i^2} \quad (7)$$

2.3.5 Mean Square Deviation

The mean square deviation (η value) measures how closely observed earthquake data follow the Gutenberg-Richter inverse power law, which describes the relationship between earthquake magnitude and frequency.

$$\eta = \frac{\sum (\log_{10}N_i - (a - bM_i))^2}{n - 1} \quad (8)$$

2.3.6 Magnitude Deficit

This is the residual of the maximum magnitude observed in n events and the largest magnitude based on G-R law. This can be represented as the following equation

$$MD = M_{\text{maximumobserved}} - M_{\text{maximumexpected}} \quad (9)$$

With as

$$M_{\text{maximumobserved}} = \frac{a}{b} \quad (10)$$

2.3.7 Mean Time Between Characteristic or Typical Events (μ value)

The μ value, or mean time between characteristic or typical events, represents the average interval between significant seismic events among the last n occurrences. This hypothesis suggests that, after a fault releases stress in a major earthquake, stress gradually builds up again until another large release occurs. Studies by [17] in the Parkfield area found relatively constant intervals between these large, recurring earthquakes, which are referred to as “characteristic events” due to their similar magnitudes and periodic occurrence. For instance, earthquakes with magnitudes in a specific range (such as 7 to 7.5) might be grouped as a characteristic magnitude.

$$\mu = \frac{\sum (t_i \text{ characteristics})}{n_{\text{characteristics}}} \quad (11)$$

2.3.8 Aperiodicity of the Mean (c -value)

The coefficient of variation of the mean time between characteristic events (μ), also called the aperiodicity of the mean (c -value), measures how closely the timing of earthquakes in a seismic region follows a regular, characteristic pattern. It is calculated as the standard deviation of the observed times divided by μ . A high c -value indicates a large variation from the average time between characteristic events, while a low c -value suggests that these events occur at more regular intervals.

$$c = \frac{\text{standarddeviationobsvertime}}{\mu} \quad (12)$$

2.4 Data Preprocessing

In the selected dataset, not all months have earthquake event data. At this stage, data with Null values were removed. From the calculated seismic parameters, the dataset initially contained 938 records monthly [9]. After removing the null values, the dataset without null value had 582 records monthly with sample in Table 1. Once the data was cleaned of null values, all features were normalized using Min-Max scaling to a range of -1 to 1 with Equation 13, as the next step involves modeling with the tanh activation function. Before proceeding to the training phase, the data was split into 80% for training and 20% for validation and testing [9]. The splitting process was conducted without using random methods to preserve the time series sequence.

Splitting process was performed once without cross-validation due to the small dataset size.

$$\hat{x} = 2 \frac{x - x_{min}}{x_{max} - x_{min}} + 1 \tag{13}$$

Table 1. Sample data after preprocessing

No	timestamp	DD	MM	RSRER	b-value	MSD	MD	μ	c	earthquake
1	1948-08-31 00:00:00+00:00	10	6.263	1.26E+10	0.771	-0.145	-0.089	5.000	2.000	1
2	1953-11-30 00:00:00+00:00	3	6.441	3.05E+11	0.586	-0.398	0.457	0.333	0.471	1
3	1953-12-31 00:00:00+00:00	15	6.130	4.27E+09	1.010	-0.082	-0.058	1.000	14.000	1
.
.
.
579	2024-05-31 00:00:00+00:00	7	4.567	9.47E+08	0.977	-0.574	-0.377	3.000	0.000	0
580	2024-06-30 00:00:00+00:00	27	4.670	1.25E+09	0.699	-0.231	-0.140	2.667	3.018	1
581	2024-07-31 00:00:00+00:00	20	4.627	1.41E+09	0.783	-0.269	-0.763	1.600	2.154	1
582	2024-08-31 00:00:00+00:00	13	4.667	6.33E+08	0.560	0.045	-0.274	6.500	0.500	1

2.5 Modeling LSTM and Bi-LSTM

After data splitting, the cleaned data proceeded to the modeling stage. A sequence length of 50 was chosen for this phase [9]. The modeling was implemented using the TensorFlow Keras framework, with multiple rounds of trial and error to identify the optimal number of layers, loss function, and optimizer. In the initial trials, 100 epochs were run to assess the metrics used. The metrics applied at this stage included True Positive (TP), True Negative (TN), False Negative (FN), False Positive (FP), accuracy, and AUC. All layers used the tanh activation function, except for the final activation layer, which used the sigmoid function, as the output represents the probability of an earthquake occurring.

Long Short-Term Memory (LSTM) architecture addresses the limitations of traditional recurrent neural networks (RNNs) in learning long-term dependencies due to vanishing or exploding gradients. This is achieved by incorporating memory cells equipped with self-recurrent connections and multiplicative input, forget, and output gates. These gates control the flow of information, enabling the network to selectively retain or discard information over extended sequences. By using a truncated gradient method, LSTM ensures stable error propagation through its Constant Error Carousel (CEC), allowing efficient learning over time lags exceeding 1,000 steps. Experiments demonstrate the effectiveness of LSTM in solving complex temporal tasks that are challenging for conventional RNN algorithms. The LSTM cell is equipped with three gates: the input gate (i_t), forget gate (f_t), and output gate (o_t), which control the flow of information. The three gates correlation aligns with Equations 14–19.

$$f_t = \sigma \left(W_f \cdot [h_{t-1}, x_t] + b_f \right) \quad (14)$$

$$i_t = \sigma \left(W_i \cdot [h_{t-1}, x_t] + b_i \right) \quad (15)$$

$$\tilde{c}_t = \tanh \left(\sigma \left(W_f \cdot [h_{t-1}, x_t] + b_f \right) \right) \quad (16)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (17)$$

$$o_t = \sigma \left(W_o \cdot [h_{t-1}, x_t] + b_o \right) \quad (18)$$

$$h_t = o_t \odot \tanh(c_t) \quad (19)$$

h_t is the output list that carries the parameters produced by the LSTM to the neural network. LSTM and Bi-LSTM models have the same configuration of cells and gates; however, Bi-LSTM includes LSTM configurations in the hidden state for both forward and backward directions. Bi-LSTM (Bidirectional LSTM) networks extend LSTM by processing input sequences in both forward and backward directions. This bidirectionality captures both past and future dependencies, offering a more comprehensive understanding of temporal patterns. In the study, Bi-LSTM significantly improved forecasting accuracy throughout the entire month, as it utilized information from both past and future states simultaneously [18].

This study references from [9] and [15] to simplify the model and use it as a benchmark due to the small dataset employed. In the study by Dhurandhar et al. [19], it is stated that simple models can outperform complex models when dealing with small datasets. Additionally, from Wang et al. [20] highlights that small batch sizes yield better performance compared to larger batch sizes. Figure 2 and 3 shows the final model structure for this study. The LSTM model included 2 layers, while the Bi-LSTM model had a single layer to observe the output sequence consistency (100) for both LSTM and Bi-LSTM. At the final stage, training and validation were conducted over 1000 epochs. Dropout applied in each LSTM layer and first dense layer after flatten layer as regularization techniques. Loss function using mean squared error [21]. The prediction results indicate the probability of an earthquake occurring with a threshold of a magnitude greater than 4.4 and a depth less than 100 km. Table 2 shows the Applied hyperparameter employed in model.

This research utilized a Mac Studio M2 Max to run Jupyter Notebook for predicting earthquake occurrence probabilities as local system configuration. The Mac Studio is equipped with 32 GB of RAM, a 24-core GPU, and a 12-core CPU with a unified memory architecture. Supporting libraries used include pandas, NumPy, scikit-learn and matplotlib.

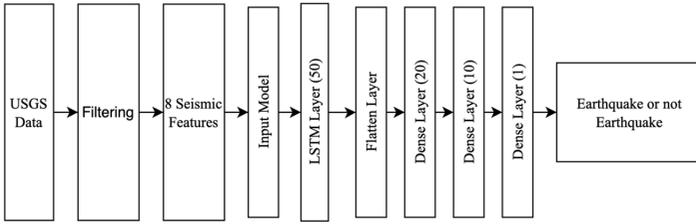


Figure 2. Modeling LSTM

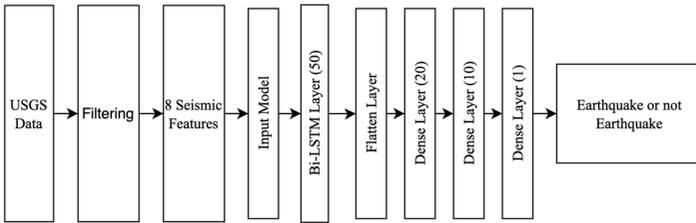


Figure 3. Modeling Bi-LSTM

3. Result Analysis

The metrics used to evaluate the performance of the model include True Negative (TN), False Positive (FP), False Negative (FN), True Positive (TP), Sensitivity (Sn), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Area Under the Curve (AUC), and Accuracy. The metric results for both models are presented in the Table 3. Using a threshold of magnitude more than 4.4 and depth less than 100 km, the LSTM model demonstrated superior sensitivity (0.92) compared to Bi-LSTM (0.6), correctly identifying 46 true positives (TP) with only 4 false negatives (FN) shows in Figure 4 (a). However, its specificity (0.2941) was relatively low, with 12 false positives (FP) and only 5 true negatives (TN). This indicates that while LSTM effectively detects earthquake occurrences, it struggles to differentiate non-occurrence cases.

Table 2. Applied Hyperparameter

Hyperparameter	Name (value)
Dropout LSTM	LSTM/Bi-LSTM Layer Dropout (20%)
Dropout Dense	Dense Layer Dropout (30%)
Learning Rate	Adam (1e-3)
Loss function	Mean Squared Error
Activation Function	tanh and sigmoid
Epoch	1000
Batch size	64

Conversely, the Bi-LSTM model exhibited greater specificity (0.6471), correctly classifying 11 true negatives and reducing the false positives to 6. However, its sensitivity dropped to 0.52, with only 26 true positives and 24 false negatives shows in Figure 4 (b). This trade-off highlights that Bi-LSTM is better at filtering out non-occurrences but less effective at capturing earthquake events, which could limit its applicability in scenarios requiring high sensitivity, such as early warning systems.

In terms of predictive values, both models demonstrated strengths in different areas. The LSTM model achieved a Positive Predictive Value (PPV) of 0.7931, indicating high reliability in its positive predictions, while its Negative Predictive Value (NPV) was moderate at 0.5556. The Bi-LSTM model showed a slightly higher PPV of 0.8125, suggesting its predictions for occurrences were slightly more accurate, but its NPV was considerably lower at 0.3143. This contrast underlines the differences in how each model prioritizes sensitivity versus specificity.

Table 3. LSTM and Bi-LSTM classification metric result

Model	TN	FP	FN	TP	Sn	Sp	PPV	NPV	AUC	Accuracy
LSTM	5	12	4	46	0.92	0.2941	0.7931	0.5556	0.6641	0.76
Bi-LSTM	11	6	24	26	0.52	0.6471	0.8125	0.3143	0.6694	0.55

The Area Under the Curve (AUC) metric further reflects the overall discriminative ability of the models. The LSTM achieved an AUC of 0.6641, slightly outperforming Bi-LSTM, which had an AUC of 0.6694. These AUC values indicate moderate performance for both models, with LSTM leaning towards better sensitivity and Bi-LSTM achieving a stronger balance with higher specificity. While LSTM may be more suitable for applications focused on capturing occurrences, Bi-LSTM's higher specificity makes it preferable in scenarios where reducing false positives is crucial. Future work could focus on optimizing both architectures to improve their trade-offs, such as incorporating attention mechanisms or refining hyperparameters for better balance in sensitivity and specificity.

As a comparison for these results, several studies with similar classification systems have been conducted by previous papers. From Vardaan et al. [22], using 2 LSTM layers with 40 neurons and the Adagrad optimizer achieved an accuracy of 58.67%. In Aslam et al. [23], a proposed ANN model with 2 connected layers, 50 neurons, a sigmoid activation function at the output, and RMSprop as the optimizer achieved an accuracy of 61.34%. Number of LSTM layers used matches this study, but the difference lies in the use of dense layers after flattening. Wang et al. [24] employed 1 LSTM layer and 2 dense layers with 256 and 64 neurons, respectively, achieving an accuracy of 54.67%. These results are still far from the accuracy achieved by both the LSTM and Bi-LSTM models in this study. Banna et al. [9], the model used 1 LSTM layer, 2 Bi-LSTM layers, 1 multiplicative attention layer, and 3 dense layers, resulting in an accuracy of 74% and an AUC of 71%. These results are close to the performance of the LSTM model in this paper. This study used the same preprocessing techniques [9] [15] and dataset sources from USGS [13] but applied them to a different location.

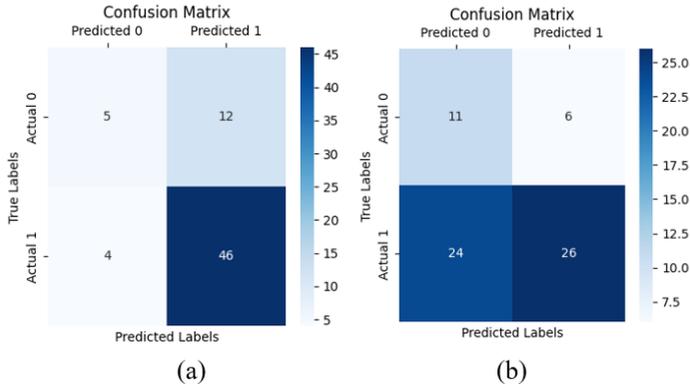


Figure 4. Confusion Metric result from model using (a) LSTM (b) Bi-LSTM

4. Conclusion

This study explored the performance of LSTM and Bi-LSTM models in predicting earthquake occurrences, focusing on their strengths and trade-offs. The results demonstrate that a simple model architecture produces better performance compared to a more complex architecture. However, this strength comes at the cost of a higher false positive rate, which may reduce reliability in distinguishing non-occurrence events. In contrast, Bi-LSTM demonstrated better balance by reducing false positives, but it struggled to capture all occurrences, potentially limiting its effectiveness in scenarios requiring high sensitivity.

In future works, we will use ensemble and attention models. This has the potential to be applied in real-world cases, but further tuning is required. The results highlighted that LSTM excels in detecting occurrences, making it highly suitable for applications where identifying potential events is critical, such as early warning systems. These findings emphasize the complementary nature of LSTM and Bi-LSTM models, with each offering distinct advantages depending on the application. LSTM’s capability to prioritize event detection makes it ideal for risk mitigation systems, while Bi-LSTM’s balanced performance is better suited for applications requiring precision and fewer false alarms. Future work should aim to refine these models through advanced techniques such as attention mechanisms and hybrid architectures, improving their ability to handle the complexities of earthquake prediction while maintaining robust and reliable performance across different use cases.

References

- [1] T. Rikitake. “Probability of a great earthquake to recur in the Tokai district, Japan: reevaluation based on newly-developed paleoseismology, plate tectonics, tsunami study, micro-seismicity and geodetic measurements”. In: *Earth, Planets and Space* (2014), pp. 147–157.
- [2] Y. Fukushima, T. Nishikawa, and Y. Kano. “High probability of successive occurrence of Nankai megathrust earthquakes”. In: *Scientific Reports* 13.1 (2023), p. 63.
- [3] S. Aoi *et al.* “Stress transfer in the Tokai subduction zone from the 2009 Suruga Bay earthquake in Japan”. In: *Nature Geoscience* 3 (2010), pp. 496–500.

- [4] T. Nishimura. "Slow Slip Events in the Kanto and Tokai Regions of Central Japan Detected Using Global Navigation Satellite System Data During 1994–2020". In: *Geochemistry, Geophysics, Geosystems* 22.2 (2021).
- [5] S. Hochreiter and J. Schmidhuber. "Long Short-term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.
- [6] M. Schuster and K. Paliwal. "Bidirectional recurrent neural networks". In: *IEEE Transactions on Signal Processing* 45.11 (1997), pp. 2673–2681.
- [7] T. D. Kusumaningrum, A. Faqih, and B. Kusumoputro. "Emotion Recognition Based on DEAP Database using EEG Time-Frequency Features and Machine Learning Methods". In: *Journal of Physics: Conference Series*. Vol. 1501. 1. 2020.
- [8] M. A. Rusho et al. "ADVANCED EARTHQUAKE PREDICTION: UNIFYING NETWORKS, ALGORITHMS, AND ATTENTION-DRIVEN LSTM MODELLING". In: *International Journal of GEOMATE* (2024), pp. 135–142.
- [9] M. H. A. Banna et al. "Attention-Based Bi-Directional Long-Short Term Memory Network for Earthquake Prediction". In: *IEEE Access* 9 (2021), pp. 56589–56603.
- [10] A. Joshi, P. Singh, and B. Raman. "A Deep Attention Model for Onsite Estimation of Earthquake Epicenter Distance and Magnitude". In: *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024).
- [11] B. Kamanditya and B. Kusumoputro. "Elman Recurrent Neural Networks Based Direct Inverse Control for Quadrotor Attitude and Altitude Control". In: *2020 International Conference on Intelligent Engineering and Management (ICIEM)*. London, UK, 2020.
- [12] S. Siami-Namini, N. Tavakoli, and A. S. Namin. "The Performance of LSTM and BiLSTM in Forecasting Time Series". In: *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA, 2019.
- [13] USGS. *Search Earthquake Catalog*. <https://earthquake.usgs.gov/earthquakes/search/>. Accessed: 17-08-2024.
- [14] S. Lapins et al. "A Little Data Goes a Long Way: Automating Seismic Phase Arrival Picking at Nabro Volcano With Transfer Learning". In: *JGR Solid Earth* 126.7 (2021).
- [15] A. Panakkat and H. Adeli. "Neural network models for earthquake magnitude prediction using multiple seismicity indicator". In: *International Journal of Neural Systems* 17.01 (2007), pp. 13–33.
- [16] A. Mignan and M. Broccardo. "Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on Their Limitations". In: *Seismological Research Letters* 91.4 (2020), pp. 2330–2342.
- [17] Y. Y. Kagan and D. D. Jackson. "Long-term probabilistic forecasting of earthquakes". In: *Journal of Geophysical Research: Solid Earth* 99.B7 (1994), pp. 13685–13700.
- [18] D.-M. Petroşanu and A. Pirjan. "Electricity Consumption Forecasting Based on a Bidirectional Long-Short-Term Memory Artificial Neural Network". In: *Sustainability* 13.1 (2021), p. 104.
- [19] A. Dhurandhar, K. Shanmugam, and R. Luss. *Leveraging Simple Model Predictions for Enhancing its Performance*. 2019.
- [20] T. Wang et al. "Large Batch Optimization for Object Detection: Training COCO in 12 minutes". In: *Computer Vision–ECCV 2020: 16th European Conference*. Glasgow, UK, 2020.
- [21] H. A. Nugroho et al. "Ensemble Deep Learning NARX for Estimating Time Series of Earthquake Occurrence". In: *2023 3rd International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*. Singapore, 2023.
- [22] K. Vardaan et al. "Earthquake trend prediction using long short-term memory RNN". In: *International Journal of Electrical and Computer Engineering* 9.2 (2019), pp. 1304–1312.
- [23] B. Aslam et al. "Seismic investigation of the northern part of Pakistan using the statistical and neural network algorithms". In: *Environmental Earth Sciences* 80.59 (2021).
- [24] Q. Wang et al. "Earthquake Prediction Based on Spatio-Temporal Data Mining: An LSTM Network Approach". In: *IEEE Transactions on Emerging Topics in Computing* 8.1 (2017), pp. 148–158.