Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2431~2436
e-ISSN: 2548-6861

Comparison of ARIMA, LSTM and GRU Models for Seismic b-Value Prediction in Southern Sumatera

Rendinis [1]*, Makhsun [2]*, Choirul Basir [3]**
* Master of Informatics Engineering Program, Universitas Pamulang
**Departement of Mathematics, Faculty of Mathematics and Natural Science, Universitas Pamulang
rendinis@bmkg.go.id [1], dosen00345@unpam.ac.id [2], dosen02278@unpam.ac.id [3]

Article Info

Article history:

Received 2026-05-01
Revised 2026-05-09
Accepted 2026-05-25

Keyword:

ARIMA,
B-Value,
Earthquake,
GRU,
LSTM.

ABSTRACT

Indonesia experiences high seismic activity due to its location at the convergence of three major tectonic plates, making continuous monitoring of earthquake potential crucial. A fundamental parameter in seismic hazard analysis is the b-value, which reflects the stress conditions, structural heterogeneity, and magnitude distribution within the Earth's crust. Predicting b-value fluctuations remains challenging due to its highly volatile nature. This study aims to analyze and forecast the seismic b-value in Southern Sumatra by comparing a classical statistical model, ARIMA , with two advanced machine learning architectures, LSTM and GRU. Historical earthquake catalogs from BMKG and NEIC-USGS spanning 1960–2025 were utilized. The data underwent a declustering process using the Reasenberg method to eliminate foreshocks and aftershocks, yielding 15,844 independent events. The monthly b-value was then calculated using Maximum Likelihood Estimation. Furthermore, PSO was applied to tune the hyperparameters of the deep learning models. Evaluation reveals that ARIMA yields the highest predictive accuracy, achieving a Mean Absolute Error (MAE) of 0.11, Root Mean Square Error (RMSE) of 0.16, and Mean Absolute Percentage Error (MAPE) of 13.85%. In contrast, GRU (MAPE 14.30%) and LSTM (MAPE 16.22%) produced smoother predictions but struggled to capture extreme short-term fluctuations. The findings conclude that for volatile and limited-size time series data like regional b-values, the linear approach of ARIMA remains significantly more effective than complex deep learning models.

I. INTRODUCTION

Indonesia is recognized as one of the countries with the highest seismic activity globally due to its strategic position at the confluence of three major tectonic plates: the Eurasian, Indo-Australian, and Pacific plates [1]. These plates are in constant relative motion; for instance, the Indo-Australian plate moves northward, subducting beneath the Eurasian plate, which moves in a southeastward direction. This complex tectonic interaction frequently triggers earthquakes of varying intensities across the archipelago [2]. Southern Sumatra, in particular, exhibits high seismic activity and potential for large-magnitude events due to these active subduction zones and the movement of the Great Sumatran Fault. Given the region's vulnerability and its significant population density, accurate earthquake prediction and analysis have become critical components of disaster preparedness and early warning systems.

In the study of seismicity, one of the most fundamental parameters is the b-value, which is an essential part of the Gutenberg-Richter relationship [3] [4]. The b-value measures the frequency distribution of seismic events relative to their magnitudes and serves as a vital indicator for identifying seismic activity patterns and the probability of large earthquakes in a specific area [5] [6]. Geophysically, the b-value reflects local stress accumulation and the heterogeneity of the Earth's crust. It tends to fluctuate over time based on changes in geological conditions. For example, a low b-value is frequently associated with high tectonic stress and the potential for a major earthquake, whereas a higher b-value

typically indicates a lower stress environment characterized by more frequent small seismic events [7] . Consequently, the accurate monitoring and prediction of b-value fluctuations are indispensable tools for effective disaster risk mitigation planning [8].

However, traditional methods for calculating and predicting the b-value face significant limitations, particularly regarding assumptions of linearity and sensitivity to incomplete data samples. This is where machine learning and deep learning methodologies become highly relevant. Machine learning is capable of identifying complex, non-linear patterns within seismic data that might remain undetected through simple statistical analysis [9] [10]. Furthermore, these techniques allow for the utilization of larger and more diverse datasets, resulting in predictive models that are more adaptive and robust against the ever-changing dynamics of seismic activity [11].

Time series analysis specifically allows for the identification of seasonal patterns and long-term trends, which are invaluable for forecasting and informed decision-making. It is also effective in detecting anomalies, enabling rapid responses to unexpected geological changes [12] . The Autoregressive Integrated Moving Average (ARIMA) is a classic statistical model widely used for analyzing and projecting time series data. On the other hand, advanced deep learning algorithms, such as Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM), represent types of Recurrent Neural Networks (RNN) specifically designed to model and predict changes in seismic time series [13]. GRU, in particular, can assist in detecting unusual b-value patterns often associated with increased earthquake activity or the early precursors of significant events [6] [14].

Despite these advancements, a research gap exists regarding the comparative application of these models for b-value prediction in Southern Sumatra. There is a need to test whether conventional statistical methods like ARIMA have significant limitations in handling the complexity of seismic time series data compared to deep learning methods. Furthermore, relatively few studies have employed LSTM and GRU algorithms specifically to model b-value fluctuations in the high-activity environment of Southern Sumatra.

To address these issues, this study aims to predict the seismic b-value in Southern Sumatra using ARIMA, LSTM, and GRU methods. By leveraging historical earthquake data from catalogs provided by the National Earthquake Information Center (NEIC-USGS) and the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) from 1960 to 2025, this research seeks to build accurate predictive models. This effort is intended to improve seismic activity monitoring and provide enhanced information for disaster mitigation in the region [15]. The study evaluates the performance of these three distinct approaches using standard metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) to determine the most effective framework for capturing tectonic stress dynamics.

II. METHOD

A. Research Framework

The methodology adopted in this study follows a systematic sequence to ensure robust predictive modeling of the seismic b-value . The complete research design is illustrated in Figure 1.

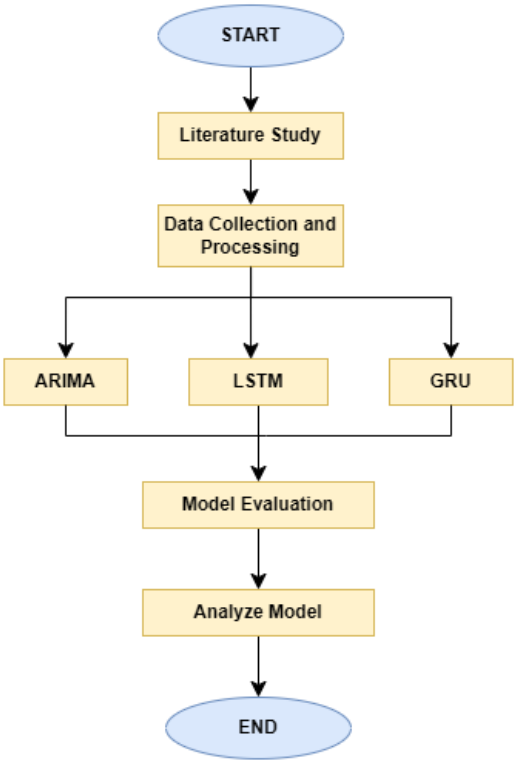

Figure 1.Research Flowchart

As depicted in Figure 1, the research framework consists of several interconnected technical phases. It begins with the acquisition of raw seismic catalogs, followed by a rigorous data preprocessing phase that includes declustering and Magnitude of Completeness (M_c) determination. The core variable, the b-value, is then extracted using Maximum Likelihood Estimation to form a continuous time series. For the forecasting phase, the data is channeled into two parallel pipelines: a classical statistical approach utilizing ARIMA, and a deep learning approach utilizing LSTM and GRU. Finally, the models are evaluated using standard error metrics to determine the most reliable predictive framework.

B. Study Area and Data Acquisition

The study focuses on the Southern Sumatra region, defined by coordinates 1.143° S - 7.144° S and 101.514° E - 106.194°

E. This region is characterized by high tectonic activity due to the subduction of the Indo-Australian plate beneath the Eurasian plate. The dataset comprises 16,167 historical earthquake events recorded between 1960 and 2025, with ≤ 200 km and magnitudes ranging from 1 to 9. Data were retrieved from the catalogs of BMKG and NEIC-USGS.

C. Data Preprocessing and b-Value Estimation

To ensure the statistical independence of the events, the catalog underwent a declustering process using the Reasenberg algorithm in Zmap V6. Standard interaction parameters were applied, including a maximum spatial interaction distance linked to earthquake magnitude and a temporal window of up to 10 days for aftershock sequence identification. This rigorous process identified and removed 323 dependent events (foreshocks and aftershocks). The dataset was reduced from 16,167 to 15,844 independent events, effectively eliminating approximately 2% of the clustered data that could introduce statistical bias.. The b-value estimation was conducted using a fixed temporal window of 1 month to capture short-term stress fluctuations. Prior to the calculation, the Magnitude of Completeness (M_c) was determined to be 4.0 using the Maximum Curvature method, ensuring that only reliably recorded events were included. The b-value was then calculated using the Maximum Likelihood Estimation (MLE) method based on the Utsu (1967). MLE was chosen over standard linear regression because it provides a statistically robust, unbiased estimation for power-law distributions and assigns appropriate weights to the frequency of events, which is critical for seismic data

$$b = \frac{\log_{10} e}{\overline{M} - M_{min}}$$

Where $\overline{M}$ is the mean` magnitude and M_{min} is the M_c. This process yielded a monthly time series of 646 b-value data points.

D. Time Series Forecasting Models

This study implemented and compared three forecasting architectures:

- ARIMA

To prevent data leakage, a common pitfall in time series forecasting the dataset was strictly split in chronological order without any random shuffling. For the ARIMA model, the first 90% of the historical data was utilized for training, while the remaining 10% was preserved as an out-of-sample test set. For the machine learning models (LSTM and GRU), a sliding window approach with a window size of 6 months was implemented to capture temporal dependencies. The sequential data was chronologically partitioned into 80% for training, 10% for validation (to monitor early stopping and tune hyperparameters), and the final 10% for testing. Stationarity was verified using the Augmented Dickey-Fuller (ADF) test.

- Machine Learning (LSTM and GRU)

Feature engineering incorporated cyclical time features (sin/cos of months), lag features (1 and 2 periods), and rolling means (3 periods) .The dataset utilized a sliding window size of 6 months.

E. Hyperparameter Optimization Using PSO

To maximize performance, Particle Swarm Optimization (PSO) tuned hyperparameters based on minimizing RMSE. The detailed architectural constraints are summarized in Table 1

TABEL I
MACHINE LEARNING ARCHITECTURE AND HYPERPARAMETER SEARCH SPACE

Parameter / Configuration	LSTM Model	GRU Model
PSO Search SPace		
Hidden Layer 1 Units	16 – 64 neurons	32 – 256 neurons
Hidden Layer 2 Units	8 – 32 neurons	16 – 128 neurons
Droput Rate	0.2 – 0.4	0.1 – 0.3
Learning Rate	0.0001 - 0.005	0.0001 – 0.005
Fixed Parameters		
Optimizer	AdamW	Adam
Activation	Tanh	Tanh
Batch Size	32	32
Max Epochs	150	150
Early Stopping Patience	15 epochs (monitor: val_loss)	15 epochs (monitor: val_loss)
Regularization	L2 (0.001) on recurrent layers	Dropout

Through the PSO tuning process, the global optimum hyperparameters were identified. The optimal LSTM configuration consisted of 32 units in the first layer, 16 units in the second layer, a learning rate of 0.001, and a dropout rate of 0.2. The optimal GRU architecture utilized 64 units in the first layer, 32 units in the second layer, a learning rate of 0.001, and a dropout rate of 0.15. For the classical approach, the Grid Search method identified ARIMA (1, 1, 2) as the most parsimonious model based on the lowest Akaike Information Criterion (AIC) score.

The hyperparameter configurations for both LSTM and GRU were determined iteratively through the PSO algorithm to ensure the models reached their global optimum. For the LSTM model, a more constrained search space with AdamW optimization and L2 regularization was implemented to strictly prevent overfitting given the high volatility of the b-value time series. In contrast, the GRU model utilized a broader search space for its hidden units and a standard Adam optimizer, taking advantage of its computationally lighter

architecture. The final training phase for all deep learning models employed an early stopping mechanism to monitor validation loss, ensuring that the training process halted at the most generalizable state before any potential divergence.

F. Evaluation Metrics

Model performance was quantified using the following standard metrics [16]:

1. Root Mean Square Error (RMSE)
   Calculates the square root of the average squared differences, emphasizing larger errors or outliers.

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (1)$$

2. Mean Absolute Error (MAE)
   Measures the average absolute error between actual and predicted values

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \quad (2)$$

3. Mean Absolute Percentage Error (MAPE)
   Measures error in percentage terms to assess relative accuracy.

$$\text{MAPE} = \frac{100\,\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3)$$

III. RESULT AND DISCUSSION

A. Seismic Data Exploration and Regional Seismicity

The initial phase of this research involved a comprehensive spatial analysis of 16,167 seismic events recorded in the Southern Sumatra region from 1960 to 2025. The geographical distribution of these events is visualized in Figure 2 [17].

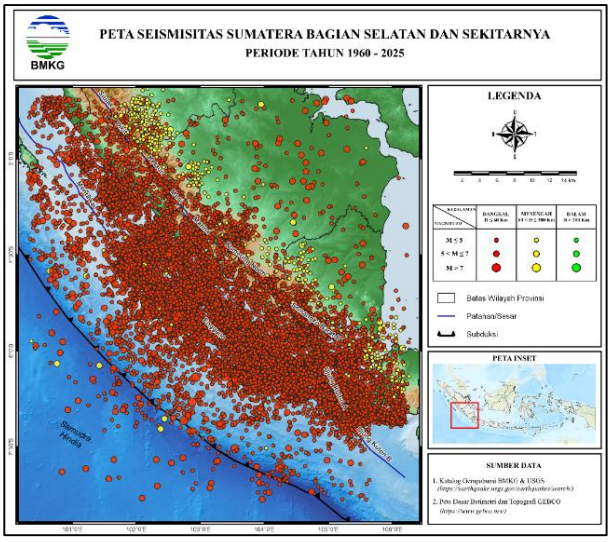

Figure 2. Seismicity Map of Southern Sumatra (1960-2025)

As illustrated in Figure 2, the seismicity is predominantly concentrated along the western coast and southern offshore areas of Sumatra. These locations align closely with the active subduction zone where the Indo-Australian plate subducts beneath the Eurasian plate. The map reveals a high density of shallow earthquakes (red markers, 0–60 km depth), which indicates that tectonic energy release primarily occurs within the brittle upper crust [18] . Furthermore, clusters of activity are observed following major structural lines, including the Great Sumatran Fault and the Mentawai Fault system. To ensure the statistical independence of the events used for b-value calculation, the catalog underwent a declustering process. Using the Reasenberg algorithm, the dataset was reduced to 15,844 independent events, effectively removing approximately 2% of the initial data identified as foreshocks and aftershocks .

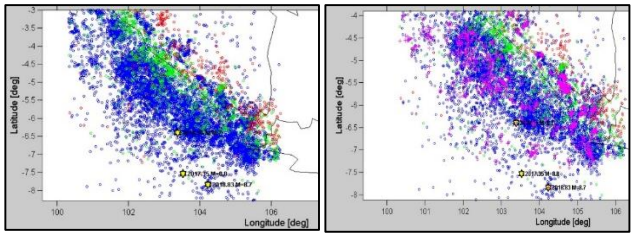

Figure 3. Earthquake Data Comparison: Before (left) and after (right) Declustering

B. Frequency-Magnitude Distrubution and b-value Analysis

The regional b-value estimation was based on M_c of 4.0, determined via the Maximum Curvature method.

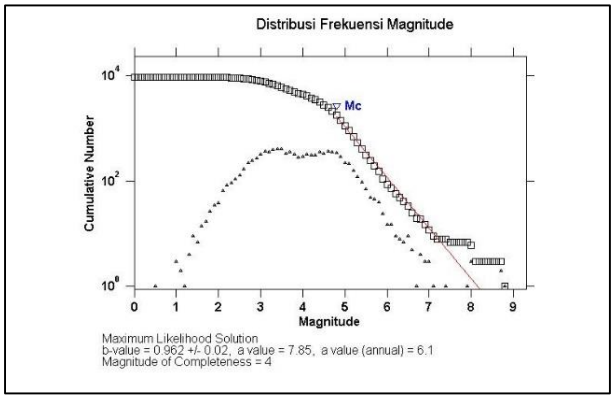

Figure 4. Frequency-Magnitude Distribution (FMD) for Southern Sumatra

The calculated regional b-value of 0.962 indicates a highly active tectonic environment . the b-value serves as a fundamental physical parameter reflecting the stress state of the crust, where values near 1.0 represent a balanced state of energy release [7] . Furthermore, the temporal fluctuations of the b-value, visualized in Figure 6, provide insights into stress accumulation cycles.

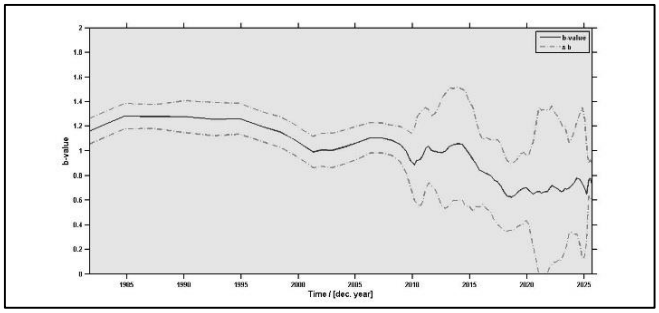

Figure 5. Temporal Fluctuations of b-value over 65 years

As shown in Figure 5, significant drops in the b-value occurred prior to major earthquakes in 2017 and 2018 . This phenomenon aligns with the observations of Lacidogna et al. (2023) and Ito & Kaneko (2023), who noted that a temporal decrease in the b-value often acts as a precursor to large-magnitude seismic events due to increasing tectonic stress.

C. Model Perfomance Evaluation and Comparative Analysis

The predictive efficacy of ARIMA, GRU, and LSTM models was quantified using standard error metrics. The comparative results are summarized in Table II.

TABEL II
MLP MODEL PERFORMANCE BASED ON STATION LOCATION

Model	MAE	RMSE	MAPE(%)
ARIMA (1,1,2)	0.11	0.16	13.85%
GRU	0.12	0.18	14.30%
LSTM	0.14	0.19	16.22%

Based on table II show The results demonstrate that the ARIMA model (1, 1, 2) achieved the highest accuracy. Based on the criteria proposed by Rachim et al. (2020), a MAPE below 15% as achieved by ARIMA (13.85%) and GRU (14.30%) indicates a good level of predictive accuracy for time series data [19].

D. Discussion

The finding that a classical statistical approach like ARIMA outperforms deep learning architectures is a significant observation in this study.

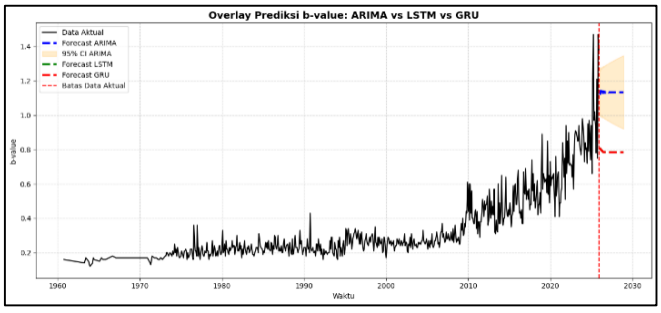

Figure 6. Overlay of Actual and Predicted b-values

Although ARIMA (MAPE 13.85%) statistically outperformed GRU (MAPE 14.30%), the margin of difference is relatively small. A critical analysis of the overlay visualization (Figure 6) reveals the distinct behavioral differences between the models. The deep learning models, particularly GRU, excel at capturing the long-term non-linear macro-trend of the b-value. However, they suffer from an 'over-smoothing' effect; optimizing for the global Mean Squared Error forces the network to treat extreme, short-term b-value spikes as noise. Consequently, GRU struggles to reach the peak values during sudden stress releases.

Conversely, ARIMA utilizes its differencing component (d=1) and linear moving average to react more aggressively to recent localized shifts. While ARIMA is not inherently superior in modeling complex temporal dynamics, its parsimonious nature prevents the over-smoothing of high-frequency fluctuations, making it marginally more accurate and computationally efficient for this specific, limited-sample geophysical dataset.

To address the significance of the performance differences, it is crucial to highlight the narrow margin in MAPE between ARIMA (13.85%) and GRU (14.30%). Given the relatively limited sample size of the dataset (646 monthly points), the 0.45% difference indicates that while ARIMA yields the lowest absolute error metrics, the performance of the GRU model is highly competitive and practically comparable. Therefore, the superiority of ARIMA observed in this study represents a marginal advantage specific to the high short-term volatility and limited temporal scale of the current regional b-value data, rather than an absolute algorithmic dominance [20]. It is highly plausible that in scenarios involving significantly larger seismic datasets, the capacity of GRU to map complex non-linear dependencies would outpace the linear constraints of ARIMA [11].

IV. CONCLUSION

This study successfully developed and evaluated time series predictive models for the seismic b-value in the Southern Sumatra region using three distinct methodologies: ARIMA, LSTM, and GRU. The analysis of the historical seismic catalog (1960–2025) reveals that the b-value in this region is highly dynamic and volatile, reflecting complex tectonic stress changes over time.

The performance evaluation using MAE, RMSE, and MAPE metrics identifies the classical statistical model, ARIMA (1, 1, 2), as the most effective architecture for this specific dataset. ARIMA achieved the lowest error rates (MAPE 13.85%), providing a more stable and accurate representation of b-value fluctuations compared to deep learning models. While LSTM and GRU demonstrated the ability to follow general macro-trends, they were less responsive to the extreme, non-linear spikes characteristic of regional b-value data. This research concludes that for seismological datasets with high volatility and relatively

limited temporal samples, parsimonious linear models like ARIMA remain superior to more complex deep learning architectures. These findings provide a methodological foundation for enhancing earthquake monitoring and disaster risk mitigation strategies in Southern Sumatra.

REFERENCES

[1] B. Rahmat, F. Afiadi, and E. Joelianto, "Earthquake Prediction System using Neuro-Fuzzy and Extreme Learning Machine," in *Proceedings of the International Conference on Science and Technology (ICST 2018)*, Paris, France: Atlantis Press, 2018. doi: 10.2991/icst-18.2018.95.

[2] Y. Wang, S. Li, and J. Song, "Exploring Magnitude Estimation for Earthquake Early Warning Using Available P-Wave Time Windows Based on Chinese Strong-Motion Records," *Pure Appl. Geophys.*, vol. 179, no. 11, pp. 4037–4052, Nov. 2022, doi: 10.1007/s00024-022-03062-4.

[3] R. Ito and Y. Kaneko, "Physical Mechanism for a Temporal Decrease of the Gutenberg-Richter *b* -Value Prior to a Large Earthquake," *J. Geophys. Res. Solid Earth*, vol. 128, no. 12, Dec. 2023, doi: 10.1029/2023JB027413.

[4] G. Lacidogna, O. Borla, and V. De Marchi, "Statistical Seismic Analysis by b-Value and Occurrence Time of the Latest Earthquakes in Italy," *Remote Sens. (Basel).*, vol. 15, no. 21, p. 5236, Nov. 2023, doi: 10.3390/rs15215236.

[5] M. A. Bilal, Y. Ji, Y. Wang, M. P. Akhter, and M. Yaqub, "An Early Warning System for Earthquake Prediction from Seismic Data Using Batch Normalized Graph Convolutional Neural Network with Attention Mechanism (BNGCNNATT)," *Sensors*, vol. 22, no. 17, p. 6482, Aug. 2022, doi: 10.3390/s22176482.

[6] B. Rahmat *et al.*, "Comparison of B-Value Predictions as Earthquake Precursors using Extreme Learning Machine and Deep Learning," *Internetworking Indonesia Journal*, vol. 12, no. 1, pp. 47–52, 2020.

[7] C. H. Scholz, "On the stress dependence of the earthquake b-value," *Geophys. Res. Lett.*, vol. 42, no. 5, pp. 1399–1402, Mar. 2015, doi: 10.1002/2014GL062863.

[8] K. Lee, J. Oh, H. Lee, and K. You, "Earthquake Magnitude Estimation Using a Total Noise Enhanced Optimization Model," *Sensors*, vol. 19, no. 6, p. 1454, Mar. 2019, doi: 10.3390/s19061454.

[9] K. Arvanitakis, I. Karydis, K. L. Kermanidis, and M. Avlonitis, "A machine learning approach for asperities' location identification," *Evolving Systems*, vol. 10, no. 1, pp. 41–50, Mar. 2019, doi: 10.1007/s12530-017-9204-x.

[10] J. B. Rundle, A. Donnellan, G. Fox, and J. P. Crutchfield, "Nowcasting Earthquakes by Visualizing the Earthquake Cycle with Machine Learning: A Comparison of Two Methods," *Surv. Geophys.*, vol. 43, no. 2, pp. 483–501, Apr. 2022, doi: 10.1007/s10712-021-09655-3.

[11] S. M. Mousavi and G. C. Beroza, "A Machine-Learning Approach for Earthquake Magnitude Estimation," *Geophys. Res. Lett.*, vol. 47, no. 1, Jan. 2020, doi: 10.1029/2019GL085976.

[12] B. H. Mustawinar, K. Kardiana, Y. Yuliani, and D. R. Arifanti, "Peramalan Cadangan Devisa Menggunakan Metode Double Smoothing Exponential dan Metode Fuzzy Time Series," *Venn: Journal of Sustainable Innovation on Education, Mathematics and Natural Sciences*, vol. 3, no. 3, pp. 105–114, Aug. 2024, doi: 10.53696/venn.v3i3.79.

[13] A. P. Meriani and A. Rahmatulloh, "Perbandingan Gated Recurrent Unit (GRU) Dan Algoritma Long Short Term Memory (LSTM) Linear Refression Dalam Prediksi Harga Emas Menggunakan Model Time Series," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan. 2024, doi: 10.23960/jitet.v12i1.3808.

[14] R. E. Abercrombie, "Resolution and uncertainties in estimates of earthquake stress drop and energy release," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2196, p. 20200131, May 2021, doi: 10.1098/rsta.2020.0131.

[15] M. S. Abdalzaher, M. Krichen, and F. Falcone, "Emerging technologies and supporting tools for earthquake disaster management: A perspective, challenges, and future directions," *Progress in Disaster Science*, vol. 23, p. 100347, Oct. 2024, doi: 10.1016/j.pdisas.2024.100347.

[16] C. JIANG, L. FANG, L. FAN, and B. LI, "Comparison of the earthquake detection effects of phaseNet and EQTransformer considering the Yangbi and Maduo Earthquakes," *Earthquake Science*, vol. 34, no. 0, pp. 20210038–20210038, 2021, doi: 10.29382/Q20210038.

[17] Y. Zhong and Y. J. Tan, "Deep-Learning-Based Phase Picking for Volcano-Tectonic and Long-Period Earthquakes," *Geophys. Res. Lett.*, vol. 51, no. 12, Jun. 2024, doi: 10.1029/2024GL108438.

[18] G. Cremen and C. Galasso, "Earthquake early warning: Recent advances and perspectives," *Earth. Sci. Rev.*, vol. 205, p. 103184, Jun. 2020, doi: 10.1016/j.earscirev.2020.103184.

[19] F. Rachim, T. Tarno, and S. Sugito, "Perbandingan Fuzzy Time Series Dengan Metode Chen Dan Metode S. R. Singh (Studi Kasus : Nilai Impor di Jawa Tengah Periode Januari 2014 – Desember 2019)," *Jurnal Gaussian*, vol. 9, no. 3, pp. 306–315, Aug. 2020, doi: 10.14710/j.gauss.v9i3.28912.

[20] I. W. McBrearty and G. C. Beroza, "Earthquake Phase Association with Graph Neural Networks," *Bulletin of the Seismological Society of America*, vol. 113, no. 2, pp. 524–547, Apr. 2023, doi: 10.1785/0120220182.