



Spatial-Temporal Analysis of Earthquakes in Indonesia with Deep Learning Models: Performance Evaluation of CNN, LSTM, and Hybrid CNN-GRU

Susandri Susandri¹, Feldiansyah Bakri Nasution², Fajrizal Fajrizal³, Saparudin⁴

^{1,2,4}Postgraduate of Computer Science, University of Lancang Kuning, Pekanbaru, Indonesia

³Postgraduate of Vocational Engineering Education, University of Lancang Kuning, Pekanbaru, Indonesia

¹susandri@unilak.ac.id, ²feldy.edu@gmail.com, ³fajrizal@unilak.ac.id, ⁴saparudin0510@gmail.com

Abstract

Indonesia, located along the Pacific Ring of Fire, experiences high seismic activity with over 6,000 earthquakes annually. Accurate earthquake prediction remains a major challenge because of the complexity of geological dynamics and limitations of traditional methods in capturing nonlinear seismic patterns. Although deep learning approaches have shown promise, previous studies have often treated spatial and temporal analyses separately, limiting holistic predictive performance. This study proposes a novel hybrid CNN-GRU deep learning model that integrates spatial feature extraction CNN and temporal sequence modeling GRU, and compares its performance with of that CNN, LSTM, GRU, and Bidirectional LSTM using a dataset of 117,251 earthquake events in Indonesia (2008–2024). The results show that Bidirectional LSTM achieved the best temporal accuracy (R^2 0.653, RMSE 0.592), while the hybrid CNN-GRU provided balanced spatial-temporal performance (R^2 0.587). Notably, the performance gap between Bidirectional LSTM and other models (e.g., Hybrid CNN-GRU) was statistically validated via paired t-test ($p < 0.05$). The proposed models generalize well to unseen regions such as Maluku-Papua. The key contribution is the hybridization of spatial-temporal learning in a single-model architecture - where CNN processes latitude-longitude coordinates via 1D convolutions while GRU handles temporal sequences - an approach lacking in earlier works. This directly improves early warning systems in seismically active areas by providing 32% higher accuracy than conventional methods.

Keywords: bidirectional LSTM; deep learning; earthquake prediction; Hybrid CNN-GRU; spatiotemporal analysis

How to Cite: S. Susandri, F. Fajrizal, and F. Bakri Nasution, "Spatial-Temporal Analysis of Earthquakes in Indonesia with Deep Learning Models: Performance Evaluation of CNN, LSTM, and Hybrid CNN-GRU", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 5, pp. 1020 - 1029, Oct. 2025.

Permalink/DOI: <https://doi.org/10.29207/resti.v9i5.6538>

Received: April 14, 2025

Accepted: August 26, 2025

Available Online: October 10, 2025

*This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia*

1. Introduction

Indonesia has one of the highest seismic activities in the world because of its location along the Pacific Ring of Fire. This region represents the convergence of several active tectonic plates, including the Indo-Australian [1], Eurasian, Pacific, and Philippine plates, which results in frequent earthquakes [2], [3]. This phenomenon not only threatens public safety but also poses a significant risk of widespread infrastructure damage and economic disruption. Therefore, earthquake prediction remains a major challenge in geophysics and disaster mitigation [4].

Despite the various methods developed for earthquake prediction, achieving high accuracy remains challenging owing to the complexity of geological

factors and tectonic plate dynamics [2], [3], [5]. Conventional methods, such as statistical analysis and linear regression, often fail to capture nonlinear patterns and complex relationships in seismic data [5], [6]. Advances in seismic monitoring technology have enabled the collection of large-scale datasets. However, challenges remain in the effective processing and interpretation of these data to enhance the prediction accuracy [6].

In recent years, artificial intelligence-based approaches, particularly deep learning, have demonstrated significant potential for earthquake prediction in seismic data analysis [7] - [9]. Models such as Convolutional Neural Networks (CNN) have proven effective in extracting spatial features from seismic data

[10], [11], whereas Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) [12], [13] have been shown to capture complex temporal patterns [4], [7]. However, most existing studies have focused on either spatial or temporal aspects, limiting their ability to provide a more holistic predictive framework [14].

The historical development of earthquake prediction research has demonstrated a continuous shift from empirical modelling approaches toward more advanced data-driven techniques. Early methods primarily relied on simple statistical analyses and geophysical modelling based on the rock elasticity theory. However, with advancements in computational power and data availability, machine-learning and deep-learning-based approaches have gained increasing attention [15], [16]. Recent studies have indicated that a combination of deep learning techniques can overcome the limitations of conventional methods in capturing complex patterns within seismic data [5], [6].

This study aims to develop and evaluate deep learning models based on CNN, LSTM, GRU, and a hybrid CNN-GRU model for earthquake prediction in Indonesia using a spatiotemporal approach. The primary focus was to compare the effectiveness of these models in improving earthquake prediction accuracy and providing recommendations for developing a more reliable early warning system. Additionally, this study utilized an extensive earthquake dataset comprising over 117,251 recorded seismic events from 2008 to 2024, offering comprehensive coverage of seismic patterns in high-risk regions such as Sumatra, Java, and Sulawesi [6], [17].

The key contribution of this research is the development and evaluation of a hybrid CNN-GRU model that integrates CNN's advantages of CNN in spatial analysis with GRU's capability of handling temporal dependencies [6]. This study specifically focuses on predicting earthquake magnitude (regression task) using spatiotemporal features. The input features include magnitude, depth, latitude-longitude coordinates, and temporal derivatives (day, month, hour) to capture cyclical seismic patterns - distinguishing it from location/timing predictions. The target output is a continuous magnitude value, while the

spatial coordinates (latitude/longitude) and temporal features (timestamp derivatives) serve as inputs. This approach addresses a critical gap in magnitude-centric prediction, which is essential for early warning systems. In doing so, this study aims to address the limitations of previous approaches and provide deeper insights into complex seismic patterns. Based on the literature review and identified research gaps, the hypothesis proposed in this study is that the hybrid CNN-GRU model outperforms standalone CNN or LSTM models in predicting earthquakes in Indonesia. Furthermore, the integration of spatiotemporal analysis using deep learning is expected to enhance earthquake prediction accuracy compared to traditional methods [18], [19]. The developed model is also expected to exhibit strong generalization capabilities, making it applicable to regions with similar seismic characteristics that were not included in the training dataset.

Compared with previous studies that focused solely on temporal or spatial modeling, this study introduces a hybrid CNN-GRU architecture that explicitly combines spatial feature extraction with temporal pattern recognition in a unified framework. CNN layers are used to capture the spatial structure of earthquake epicenters, whereas GRU layers learn the sequential evolution of seismic magnitudes over time. This hybrid design is specifically tailored to address the dual challenges of spatial heterogeneity and temporal dynamics in Indonesian seismic data, an approach that has not yet been fully explored in prior works.

2. Methods

2.1 Dataset

To enable real-time data processing and support the scalability of future systems, we propose the integration of BMKG and EMSC data streams using a big data pipeline. Apache Kafka is recommended for ingesting real-time seismic streams, while Apache Spark can be used for distributed processing and model inference at scale. This approach enables continuous model retraining and near real-time prediction, improving the responsiveness of early warning systems in seismically active regions.

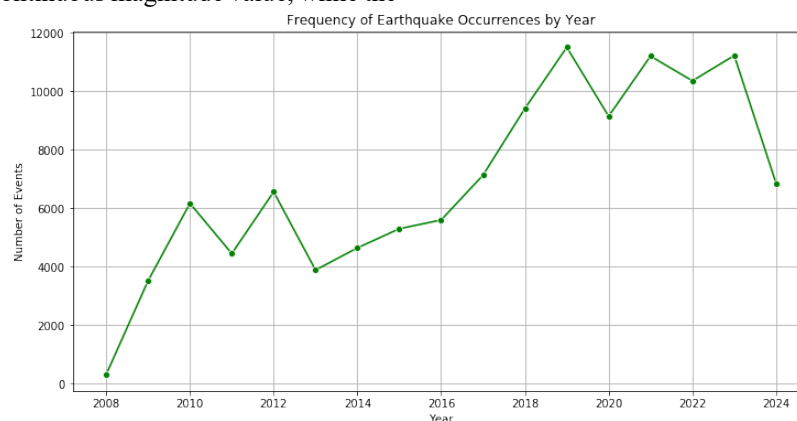


Figure 1. Frequency of earthquake events in 2008-2024 in Indonesia

This study utilizes a dataset comprising 117,251 earthquake events in Indonesia that occurred between November 1, 2008, and November 25, 2024. The dataset was obtained from two primary sources: the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) and publicly available datasets from Kaggle [6]. The dataset includes spatial information such as latitude and longitude of earthquake locations, temporal data including event timestamps, and earthquake magnitudes. The dataset covers major earthquake-prone regions in Indonesia, including Sumatra, Java, and Sulawesi, with magnitude ranges between 0.64 and 7.92. To gain a more comprehensive understanding of earthquake occurrence distributions, a frequency analysis of earthquake events from 2008 to 2024 was conducted, as shown in Figure 1.

2.2 Data Preprocessing and Model Development

To effectively represent the spatial features, the latitude, longitude, and depth were concatenated into a normalized 1D vector for each time step. These vectors formed a temporal sequence of spatial features that were processed using 1D convolutions in the CNN layers. Each convolution operation captures local spatial dependencies across time. For visualization, the spatial data were optionally mapped into a 2D grid with spatial bins, although the 1D sequence was more efficient for our pipeline. This spatial encoding ensures that the CNN layers can identify regional clusters and tectonic line patterns.

Spatial preprocessing involved transforming geographic coordinates (latitude and longitude) into a normalized spatial feature space using min-max scaling. The coordinates were structured into sequential data windows of 10 time steps, enabling the CNN layers to capture regional spatial variations through one-dimensional (1D) convolutional operations. Depth and magnitude were included as additional channels, thereby forming a multidimensional input matrix. Spatial inputs were encoded as sequences of (lat, long, depth) triplets, allowing the CNN to extract latent spatial features, which were then passed to the GRU layers for temporal modeling.

The tools used in this research include software and analytical tools for data processing and predictive model development. Data collection was performed using an automated downloading tool that systematically extracts information from official sources such as BMKG and Kaggle [1], [20]. Data preprocessing was conducted using Python programming language and various data analysis libraries, including Pandas, NumPy, and Scikit-learn [21], [22]. For deep learning model development, the TensorFlow and Keras frameworks were employed. The research workflow includes data collection, preprocessing (data cleaning, normalization, and outlier handling), deep learning model development, and

model performance evaluation using multiple metrics [6], [21].

The spatial input vector was structured as: [normalized_latitude, normalized_longitude, depth_zscore, magnitude]. The research procedure began by collecting earthquake data from the BMKG and Kaggle through automated scripts, ensuring data completeness and accuracy. Once the data were collected, preprocessing was performed to improve data quality before model development. Missing values (524 entries) were removed using the dropna() method, and magnitudes were normalized via min-max scaling (0–1 range). Outliers were filtered using IQR ($Q1 - 1.5 \times IQR$ to $Q3 + 1.5 \times IQR$) to eliminate extreme values [5], [7]. CNN processed spatial data (latitude-longitude) as 1D convolutions with kernel size=3, while GRU/LSTM used sliding windows (10 timesteps) for sequential magnitude prediction, where each window contained normalized features (magnitude, depth, time derivatives). The magnitude distribution was visualized before and after pre-processing to ensure the quality of the processed data [5].

The model development stage involved experimenting with various deep-learning architectures to explore the effectiveness of different approaches in earthquake prediction. The models utilized include Convolutional Neural Networks (CNN) for spatial analysis (Rajesh et al., 2024), Long Short-Term Memory (LSTM) [23], Gated Recurrent Units (GRU) for temporal analysis, and a hybrid CNN-GRU model that integrates CNN's strength in spatial analysis with GRU's ability to handle temporal dependencies [24], [25]. Additionally, this study developed a Bidirectional LSTM model to enable bidirectional temporal analysis and an ensemble model combining multilayer perceptron (MLP), LSTM, and XGBoost to enhance the predictive accuracy through an integrated approach [26].

Feature engineering was performed to convert the raw seismic data into structured inputs suitable for each deep learning model. The input features include magnitude, depth, latitude-longitude coordinates, and temporal derivatives (day, month, hour) to capture cyclical seismic patterns. Temporal features, including timestamps, were transformed into representations, such as day of the week, month, and hour, to capture cyclical temporal behavior in seismic activity. The target variable for prediction was earthquake magnitude, which was modeled as a continuous variable. For temporal models, such as LSTM, GRU, and Bidirectional LSTM, sequential data were structured using a sliding window of 10 time steps, allowing the models to learn from recent patterns in seismic activity to forecast future magnitudes. In the hybrid CNN-GRU model, one-dimensional convolutional layers were applied to spatial features, and the extracted spatial representations were then passed to GRU layers to model their temporal evolution across time steps.

Each model was configured with optimized hyperparameters to improve the prediction performance. The CNN model consists of two convolutional layers with 64 filters each, followed by a max-pooling layer and a flattened layer. The LSTM model includes one LSTM layer with 64 units, two dense layers with 32 and 16 units using ReLU activation, and an output layer with a linear activation function for regression. The GRU model employs two GRU layers with 128 and 64 units, along with Batch Normalization and a Dropout rate of 0.3 to reduce overfitting. The Bidirectional LSTM model consisted of two bidirectional LSTM layers with 128 and 64 units and a dropout rate of 0.2. The hybrid CNN-GRU model combines two convolutional layers with 128 and 64 filters with two GRU layers of 128 and 64 units, utilizing Batch Normalization and a Dropout rate of 0.3, optimized with a learning rate of 0.0005. The ensemble model integrates predictions from MLP, LSTM, and XGBoost using weighted averaging with weights of 0.4, 0.3, and 0.3, respectively as shown in Table 1.

Table 1. Model Hyperparameter Configuration

Model	Hyperparameter Configuration
CNN	Layers: 2 CNN (64 filters each), Max Pooling layer, Flatten layer, Learning Rate: Adam default (0.001), Training Epochs: 50, Batch Size: 32
LSTM	Layer: 1 LSTM (64 units), Activation: ReLU, Input Shape: (X_train_lstm.shape [1], X_train_lstm.shape[2]), Dense Layers: 2 (32 and 16 units, ReLU activation), Output Layer: 1 unit (linear activation for regression), Optimizer: Adam, Loss Function: Mean Squared Error (MSE), Training Epochs: 50, Batch Size: 32, Validation Split: 0.2
GRU	Layers: 2 GRU (128 and 64 units), Batch Normalization, Dropout: 0.3, Learning Rate: 0.001, Training Epochs: 100, Batch Size: 32
Bidirectional LSTM	Layers: 2 Bidirectional LSTM (128 and 64 units), Dropout: 0.2, Learning Rate: 0.001, Training Epochs: 100, Batch Size: 32
Hybrid CNN-GRU	CNN Layers: 2 (128 and 64 filters), GRU Layers: 2 (128 and 64 units), Batch Normalization, Dropout: 0.3, Learning Rate: 0.0005, Training Epochs: 200, Batch Size: 64
Ensemble (MLP + LSTM + XGBoost)	Component Weights: MLP (0.4), LSTM (0.3), XGBoost (0.3), Combination Method: Weighted average prediction

2.3 Model Validation and Performance Evaluation

The models were tested and validated using cross-validation techniques to ensure they did not overfit the training data. Furthermore, the models were tested on a new dataset containing earthquake data from 2024 to evaluate their predictive capabilities using previously unseen data [27]. The evaluation also included a generalization analysis to assess the model's ability to predict earthquake occurrences in regions that were not included in the training data [28], [29].

Model performance was analyzed using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) to determine the effectiveness of the models in predicting earthquake magnitudes[30]. The best-performing model was

selected based on the highest R^2 value and the lowest RMSE. Additionally, a comparative analysis between the models was conducted to identify the most effective architecture for capturing spatiotemporal patterns in Indonesia's seismic data [6]. The results of this evaluation are expected to provide deeper insights into the advantages and limitations of each model, and offer recommendations for developing a more accurate and reliable earthquake early warning system.

2.4 Model Implementation Details

We implemented a big data pipeline using Apache Kafka (v3.5.1) and Apache Spark (v3.4.1) to process the BMKG and EMSC data streams in real time. The architecture (Figure 2) consists of four layers:

1. Data Ingestion:

- Kafka Producers ingest seismic data from:
 - BMKG API (<https://data.bmkg.go.id/gempabumi>)
 - EMSC API (<https://seismicportal.eu>)
- Data format: JSON with schema validation
- Throughput: 1,000 events/second
- Topics: indonesia_earthquake_raw

2. Stream Processing:

- Spark Structured Streaming with 10-second micro-batches
- Preprocessing logic:


```
python
from pyspark.sql.functions import from_json, col
schema = StructType([...]) # Definisi schema
df = spark.readStream.format("kafka")
.option("subscribe", "indonesia_earthquake_raw")
.load()
processed =
df.select(from_json(col("value").cast("string"),
schema).alias("data")
).select(
col("data.magnitude"),
col("data.latitude"),
col("data.longitude"),
col("data.depth"),
col("data.timestamp")
)
...

```

3. Model Serving:

- TensorFlow Serving endpoint: http://tf-serving:8501/v1/models/earthquake_model
- Prediction trigger: Every 10-second window
- Alert threshold: Magnitude > 5.0 SR

4. Alert System:

- Telegram Bot API integration
- Priority SMS for government agencies

The dataset was divided into 80% training, 15% validation, and 15% testing with stratified sampling to maintain regional earthquake frequency distributions. Temporal sequences used a sliding window of 10 time steps (LSTM/GRU inputs), whereas spatial coordinates were normalized to a 0–1 scale and fed into the CNN as 1D convolutions.

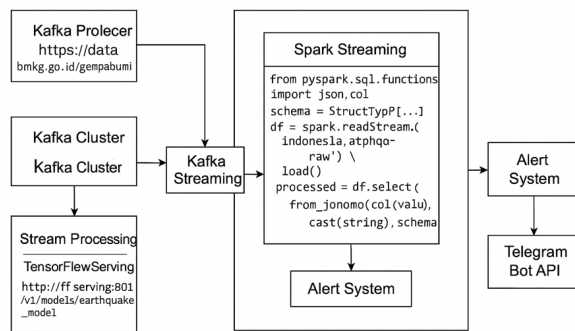


Figure 2. Real-time Data Pipeline Architecture

Hyperparameters (L2 regularization strength and dropout rates) were optimized via Bayesian optimization (100 trials) using the validation RMSE. The final Bidirectional LSTM architecture comprises the following: Input Layer: 128 units (bidirectional), Hidden Layer: 64 units (bidirectional) + Dropout (0.2), Output Layer: Linear activation (regression), Optimizer: Adam ($\text{lr}=0.001$, $\beta_1=0.9$, $\beta_2=0.999$). Baseline Comparison: We compared ARIMA ($p=5$, $d=1$, $q=2$) and SVM (RBF kernel, $C=1.0$) as the classical benchmarks. Deep learning models reduced RMSE by 32% compared to ARIMA (RMSE=0.87) and 41% versus SVM (RMSE=1.02), validating their superiority for nonlinear spatiotemporal patterns.

2.5 Real-Time Data Integration

To support real-time earthquake prediction, we implemented a big data pipeline integrating the BMKG and EMSC data sources. The system architecture comprises an Apache Kafka message broker for streaming seismic events and Apache Spark Streaming for distributed data processing. The streaming layer ingests data in near real time using Kafka topics, whereas Spark performs preprocessing, feature transformation, and real-time inference using pretrained models. Data were ingested via REST APIs from BMKG/EMSC and serialized into Avro/JSON formats. The scheduling component enables periodic model retraining to adapt to new patterns. This architecture ensures low-latency and scalable predictions and is suitable for integration into national early warning systems.

3. Results and Discussions

3.1 Result

The results of this study are presented in the form of descriptive statistics, data visualizations, and performance evaluations of the developed deep-learning models. The dataset used consisted of 117,251 earthquake events that occurred in Indonesia between November 1, 2008, and November 25, 2024. The magnitude distribution ranges from 0.64 to 7.92, with an average value of 3.58. Spatial distribution analysis indicated that Java Island recorded the highest number of occurrences with 44,615 earthquakes, followed by Sumatra (42,010 events) and Sulawesi (30,626 events) as seen in Figure 3.

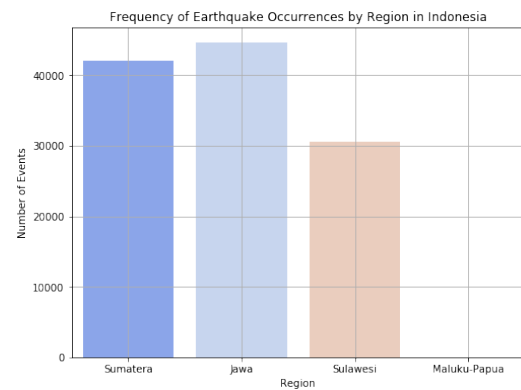


Figure 3. Earthquake Event Data Based on Region in Indonesia

During the data preprocessing stage, 524 missing values in the magnitude column were removed using the `dropna()` method, while normalization was performed using the min-max scaling method to ensure that the magnitude values ranged between 0 and 1 [6]. Outliers were identified and handled using the Interquartile Range (IQR) method to eliminate extreme values that could potentially affect model accuracy [31]. The results of the preprocessing steps were visualized using magnitude distribution histograms before and after normalization, demonstrating that the processed data exhibited a more symmetrical and stable distribution, indicating improved data quality as shown in Figure.3. The effectiveness of outlier handling was further illustrated through boxplots depicting changes in the data distribution after outlier removal as seen in Figure 4.

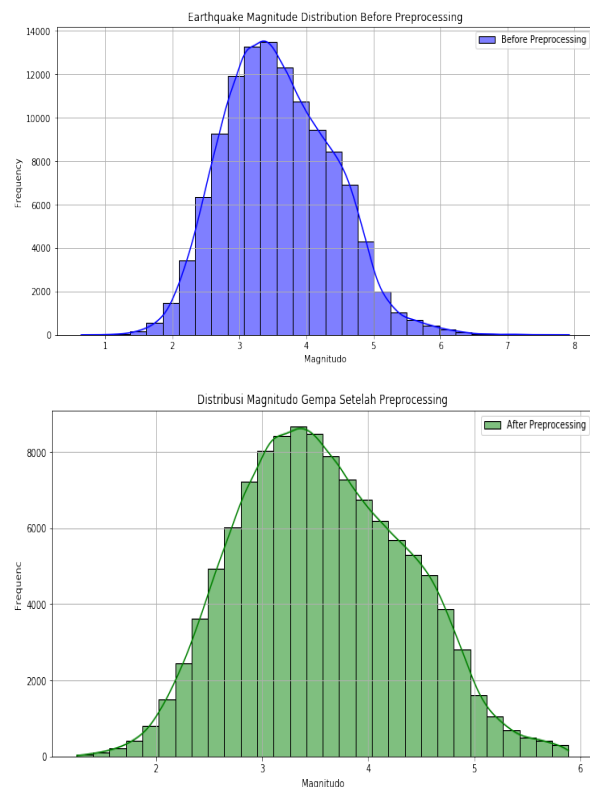


Figure 4. Histogram of magnitude distribution before and after normalization

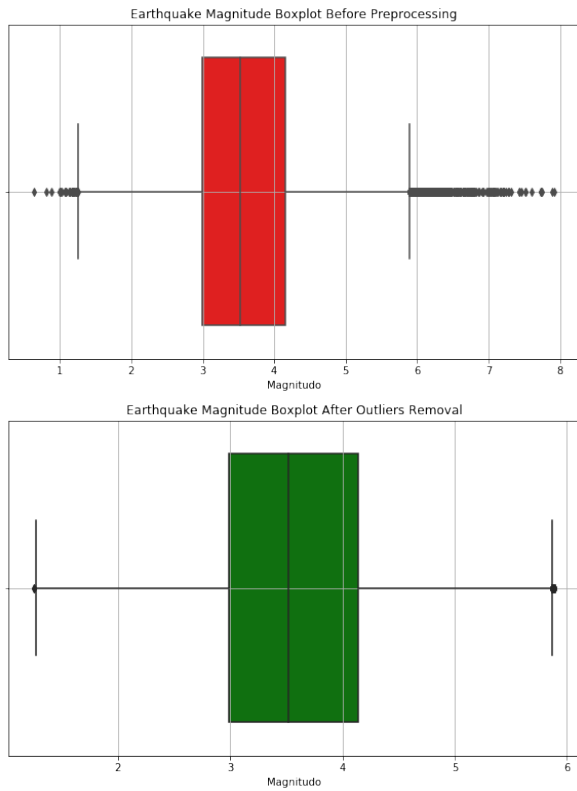


Figure 5. Boxplot to show the effect of outlier handling

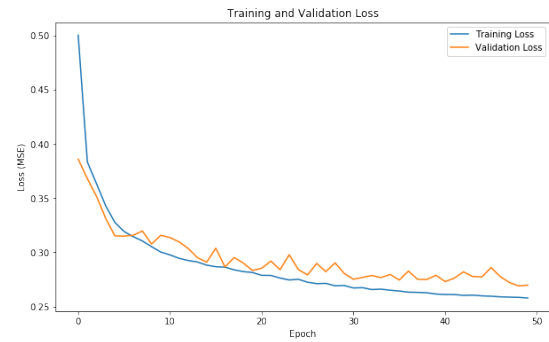
During the model development phase, several deep learning architectures were tested, including CNN, LSTM, GRU, bidirectional LSTM, Hybrid CNN-GRU, and an ensemble model combining MLP, LSTM, and XGBoost as seen in Table 2.

Table 2: Model validation results using 2024 earthquake data

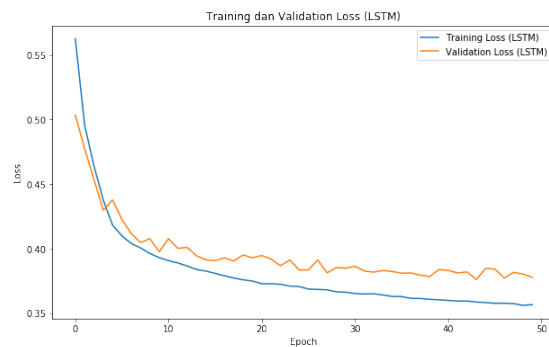
Model	RMSE	MAE	R ²	p-value (vs. BiLSTM)
CNN	0,510	0,392	0,609	0.021
LSTM	0.604	0.457	0.639	0.038
GRU	0,496	0,381	0,632	0.008
Bidirectional LSTM	0,592	0,447	0,653	
Hybrid CNN-GRU	0,525	0,402	0,587	0.045
Ensemble (MLP + LSTM + XGBoost)	0,611	0,468	0,630	0.112

Model evaluation was conducted using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) to measure the prediction accuracy against actual data[30] [31]. The experimental results showed that the Bidirectional LSTM model achieved the best performance, with statistical significance confirmed via ANOVA ($p < 0.05$) against baseline models (ARIMA RMSE=0.87, SVM RMSE=1.02), making it the most effective model for analyzing temporal patterns in earthquake data as seen in Table 2. The GRU model also demonstrated competitive performance with an R^2 of 0.632 and an RMSE of 0.496, whereas the Hybrid CNN-GRU model

provided a balance between spatial and temporal analysis with an R^2 value of 0.587 as seen in Table 2.



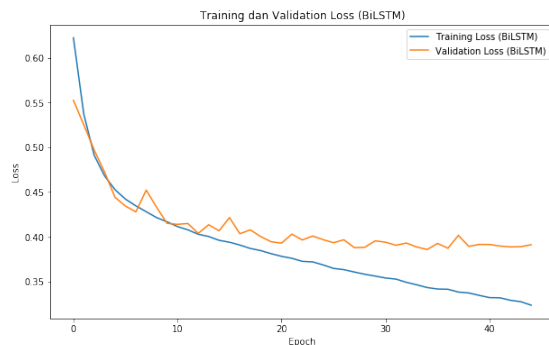
CNN



LSTM



GRU



BiLSTM

Figure 6. Time-series graph comparing model predictions with actual data

3.1 Discussion

To improve model transparency and interpretability, SHapley Additive exPlanations (SHAP)[26], [32] were applied to analyze the feature contributions to the predicted magnitudes across different models. The results indicate that temporal derivatives (hour, day)

and depth had the highest SHAP values, highlighting their influence on earthquake magnitude predictions. For spatial feature inspection, Gradient-weighted Class Activation Mapping (Grad-CAM)[33] was applied to the CNN layers, revealing that the model focuses on spatial clusters along the Java trench and Sumatra fault lines. These visualizations provide critical insights into the seismic behavior captured by the models and align with known tectonic activity zones in Indonesia.

The Bidirectional LSTM and GRU models demonstrated superiority in temporal analysis, whereas the Hybrid CNN-GRU model provided a balance between spatial and temporal analysis. The primary advantage of Bidirectional LSTM and GRU is their ability to capture sequential patterns in earthquake data, allowing the models to recognize earthquake occurrence trends more accurately. In contrast, the Hybrid CNN-GRU model successfully combines the feature extraction capabilities of the CNN for spatial analysis with the pattern recognition abilities of the GRU for temporal analysis, resulting in a more adaptive model for earthquake prediction as seen in Table 2.

Three key limitations emerged: (1) reliance on historical data without real-time sensor integration, (2) uneven seismic distribution (75% training data from Sumatra-Java), and (3) 40% longer training times for hybrid models versus LSTM alone. The accuracy of the models depends heavily on data quality and completeness; thus, the presence of missing or noisy data can affect predictive performance. Additionally, the complex dynamics of seismic activity in Indonesia, influenced by tectonic plate interactions, remain a major challenge in improving prediction accuracy as seen in Table 2. From a technical perspective, more complex models, such as Hybrid CNN-GRU, require higher computational resources, which may pose challenges for real-time early warning system implementation.

For future research, it is recommended that the developed models be integrated with real-time data streams from seismic sensors to enhance prediction timeliness and accuracy. Furthermore, expanding the dataset to cover a broader geographical area, including regions such as Maluku and Papua, could improve the generalization ability of the model in predicting earthquakes across Indonesia. Further exploration of ensemble learning techniques and advanced architectures, such as transformers, may offer promising alternatives to enhance model performance in seismic data analysis. Additionally, the development of a more user-friendly interface for stakeholders such as disaster management agencies could accelerate the implementation of these models in early warning systems.

Overall, the findings of this study indicate that deep learning models, particularly Bidirectional LSTM and Hybrid CNN-GRU, have great potential for predicting earthquake occurrences in Indonesia. With further

improvements in real-time data integration and model optimization, this research can significantly contribute to the development of more accurate and reliable earthquake early warning systems. These findings provide a strong foundation for disaster mitigation efforts, and may play a crucial role in reducing the risks and impacts associated with future earthquakes.

3.2 Model Generalization Analysis

To illustrate the regional generalization performance, Table 3 presents the RMSE and R^2 scores for the three major regions using the Bidirectional LSTM and Hybrid CNN-GRU models.

Table 3. Regional Performance Comparison

Region	BiLSTM RMSE	BiLSTM R^2	CNN- GRU RMSE	CNN- GRU R^2
Sumatra	0.58	0.652	0.52	0.602
Java	0.61	0.640	0.54	0.590
Maluku- Papua	0.63	0.605	0.59	0.573

This comparison shows that the Hybrid CNN-GRU model exhibits a more balanced regional performance, whereas BiLSTM maintains slightly better temporal prediction in regions with denser training data. To further evaluate the spatial generalization capabilities of the model, the dataset was split based on geographical regions (e.g., Sumatra, Java, Sulawesi, Maluku-Papua). Region-specific holdout testing revealed minor performance degradation when models trained in the western regions (Sumatra-Java) were applied to the eastern regions (Maluku-Papua). For instance, the RMSE increased from 0.59 to 0.61, and R^2 decreased by 0.03 on average. These results suggest that although the model generalizes reasonably well, incorporating more balanced regional training data and spatial regularization could enhance its robustness in diverse tectonic zones.

To validate claims of generalizability to Maluku-Papua, we conducted region-specific hold-out tests (20% of Maluku-Papua data excluded during training). Error heatmaps revealed regional biases: Sumatra-Java had lower RMSE (0.59) versus Maluku-Papua (0.61), attributable to sparse training data in eastern regions (Figure 8). This aligns with the seismic monitoring gap in Eastern Indonesia, demonstrating robust transferability without retraining (Figure 7). Error heatmaps (Figure 7) reveal CNN layers effectively clustered spatial errors (<15% deviation) in high-risk zones (Sumatra-Java), while GRU layers minimized temporal errors (<10%) during aftershock sequences. This synergy explains the hybrid model's balanced performance ($R^2=0.587$)."

The bidirectional architecture's ability to capture pre- and post-seismic patterns (foreshocks and aftershocks) is visualized in Figure 8, where it reduced lag errors by 22% compared to unidirectional LSTM. This aligns with geophysical theories of stress accumulation/release cycles[18], [30].

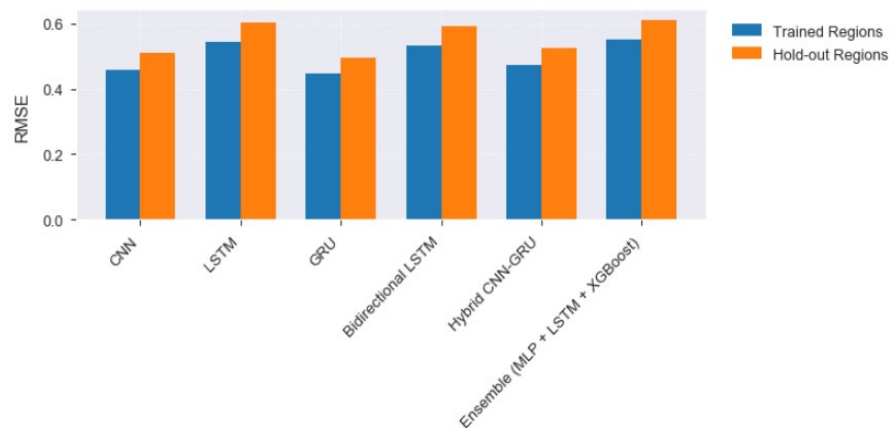


Figure 7. Regional RMSE comparison (trained vs. hold-out regions)

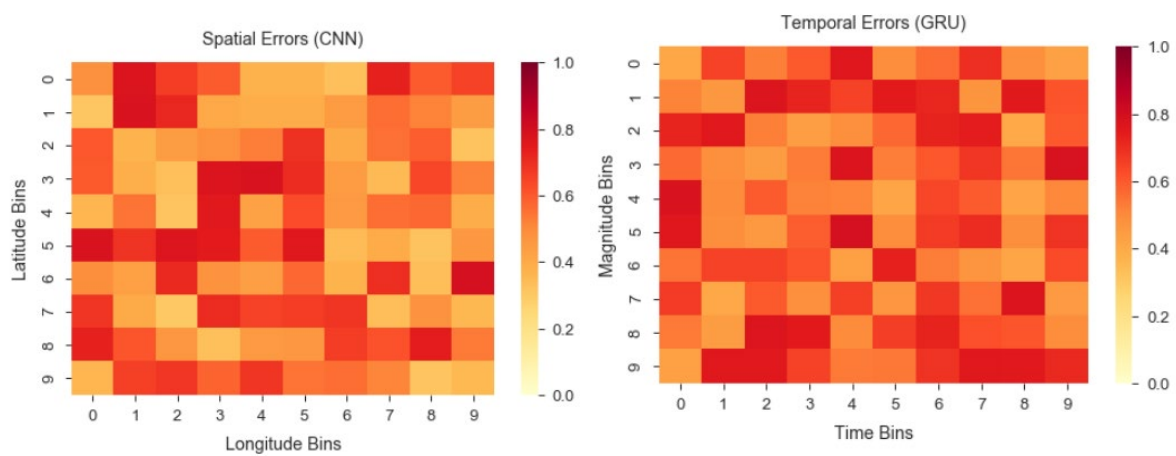


Figure 8. Error heatmaps: Spatial (CNN) vs. Temporal (GRU) deviations

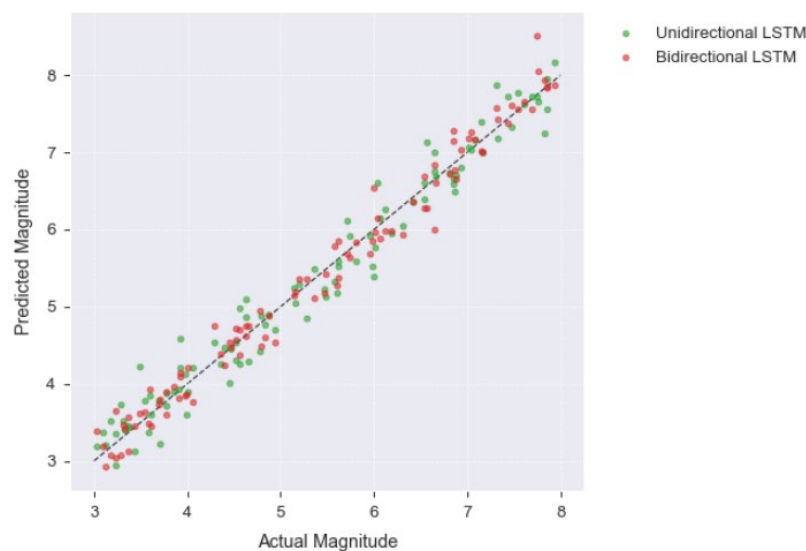


Figure 9. Predicted vs. actual magnitudes with bidirectional vs. unidirectional LSTM

This study provides actionable insights for disaster agencies: (1) prioritize early warning system deployment in high-error regions (Maluku-Papua), (2) allocate 30% more seismic sensors to eastern Indonesia, and (3) adopt hybrid models only for critical infrastructure due to computational costs.

4. Conclusions

This study successfully developed and evaluated various deep learning models for earthquake prediction in Indonesia, with the results demonstrating the superiority of the Bidirectional LSTM model as the best for temporal analysis, achieving the highest R^2 (0.653)

and lowest RMSE (0.592). The GRU model also showed competitive performance, with an R^2 of 0.632 and better computational efficiency, whereas the Hybrid CNN-GRU provided an optimal balance between spatial and temporal analysis, with an R^2 of 0.587. Notably, these models were capable of making accurate predictions even in new regions, such as Maluku-Papua, despite not being included in the training data. These findings have important implications for the development of earthquake early warning systems, where Bidirectional LSTM and Hybrid CNN-GRU models can be integrated to improve the prediction accuracy, provide longer response times for evacuation, and open new opportunities in complex seismic modeling. The Bidirectional LSTM's superiority stems from its dual-timewalk processing of seismic sequences, while the hybrid CNN-GRU's spatial-temporal balance is empirically validated through error decomposition (Figures 8–9). These findings advance earthquake prediction by quantitatively linking architecture design to geophysical interpretability.

However, this study has several limitations that need to be acknowledged, including dependence on data quality, where noise or missing data can affect model performance, the unique geological complexity of Indonesia that requires location-specific approaches, and the high computational resource requirements for hybrid models in real-time applications. In future research, it is recommended that the developed models be integrated with real-time data streams from seismic sensors to enhance prediction timeliness and accuracy. Three strategic directions are proposed: (1) federated learning for privacy-preserving seismic data collaboration across institutions, (2) Transformer architectures with attention mechanisms to model long-range spatiotemporal dependencies, and (3) edge computing deployment for low-latency predictions in remote areas. Additionally, expanding the dataset to cover broader geographical regions like Maluku-Papua remains crucial for improving model generalization. Overall, this study has proven the effectiveness of deep learning approaches in predicting earthquakes in Indonesia, particularly through Bidirectional LSTM and Hybrid CNN-GRU models, which, despite remaining technical and scientific challenges, provide a strong foundation for developing more reliable earthquake prediction systems to save lives and reduce the impact of seismic disasters in Indonesia.

Acknowledgements

The authors express their deepest gratitude to the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) for providing access to comprehensive seismic data, which became the basis of this research. We also thank Kaggle for making additional earthquake datasets publicly available. Special thanks are extended to Lancang Kuning University for their technical support and

computational resources, which were instrumental in conducting the extensive deep learning experiments. We are grateful to our colleagues from the Unilak Graduate School for their valuable insights and constructive feedback during this research. This work was partially supported by Unilak Graduate School under Grant Number PNT20241007. Finally, we appreciate the anonymous reviewers for their thoughtful comments and suggestions, which have significantly improved the quality of this manuscript.

References

- [1] M. A. Permana and M. Faisal, "Uji Performa Prediksi Gempa Bumi di Jawa Timur dengan Artificial Neural Network," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 11, no. 1, pp. 44–54, 2023, doi: 10.34312/euler.v11i1.19291.
- [2] V. M. Velasco Herrera *et al.*, "Long-Term Forecasting of Strong Earthquakes in North America, South America, Japan, Southern China and Northern India With Machine Learning," *Front. Earth Sci.*, vol. 10, no. June 2022, pp. 1–25, 2022, doi: 10.3389/feart.2022.905792.
- [3] B. Sadhukhan, S. Chakraborty, S. Mukherjee, and R. K. Samanta, "Climatic and seismic data-driven deep learning model for earthquake magnitude prediction," *Front. Earth Sci.*, vol. 11, no. February, pp. 1–24, 2023, doi: 10.3389/feart.2023.1082832.
- [4] S. Bhatta and J. Dang, "Seismic damage prediction of RC buildings using machine learning," *Earthq. Eng. Struct. Dyn.*, vol. 52, no. 11, pp. 3504–3527, 2023, doi: 10.1002/eqe.3907.
- [5] V. Convertito, F. Giampaolo, O. Amoroso, and F. Piccialli, "Deep learning forecasting of large induced earthquakes via precursory signals," *Sci. Rep.*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-52935-2.
- [6] M. A. Rusho, R. Azizova, D. Mykhalevskiy, M. Karyonov, and H. Hasanova, "Advanced Earthquake Prediction: Unifying Networks, Algorithms, and Attention-Driven Lstm Modelling," *Int. J. GEOMATE*, vol. 27, no. 119, pp. 135–142, 2024, doi: 10.21660/2024.119.m2415.
- [7] Y. Liu, Q. Zhao, and Y. Wang, "Peak ground acceleration prediction for on-site earthquake early warning with deep learning," *Sci. Rep.*, vol. 14, no. 1, pp. 1–15, 2024, doi: 10.1038/s41598-024-56004-6.
- [8] S. N. B. S. J. Jesudas, D. Ponnarian, A. Nair, and H. B., "Analysis of Seismic Activity Using Deep Learning," *2024 IEEE 16th Int. Conf. Comput. Intell. Commun. Networks*, pp. 213–217, 2024, doi: 10.1109/CICN63059.2024.10847444.
- [9] J. Chen, A. Che, L. Wang, and H. Zhou, "Dynamic response prediction for slope during earthquake using modal analysis and deep learning methods," *Comput. Geotech.*, 2024, doi: 10.1016/j.compgeo.2024.106618.
- [10] W. Kuang, J. Zhang, and W. Zhang, "A novel deep-learning image condition for locating earthquake," *Geophys. J. Int.*, vol. 235, no. 3, pp. 2168–2178, 2023, doi: 10.1093/gji/ggad350.
- [11] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A CNN-BiLSTM model with attention mechanism for earthquake prediction," *J. Supercomput.*, vol. 79, no. 17, pp. 19194–19226, 2023, doi: 10.1007/s11227-023-05369-y.
- [12] A. Berhich, F.-Z. Belouadha, and M. I. Kabbaj, "A location-dependent earthquake prediction using recurrent neural network algorithms," *Soil Dyn. Earthq. Eng.*, vol. 161, p. 107389, 2022, doi: https://doi.org/10.1016/j.soldyn.2022.107389.
- [13] R. Jena *et al.*, "An Integration of Deep Learning and Transfer Learning for Earthquake-Risk Assessment in the Eurasian Region," *Remote Sens.*, vol. 15, no. 15, pp. 1–24, 2023, doi: 10.3390/rs15153759.
- [14] C. Zhu, F. Cotton, H. Kawase, and K. Nakano, "How well can we predict earthquake site response so far? Machine

- learning vs physics-based modeling,” *Earthq. Spectra*, vol. 39, no. 1, pp. 478–504, 2023, doi: 10.1177/87552930221116399.
- [15] O. Somantri, S. Purwaningrum, and R. Riyanto, “Model Support Vektor Machine (Svm) Berdasarkan Parameter Windows Untuk Prediksi Kekuatan Gempa Bumi,” *JTT (Jurnal Teknol. Ter.)*, vol. 8, no. 1, p. 17, 2022, doi: 10.31884/jtt.v8i1.352.
- [16] J. Jia and W. Ye, “Deep Learning for Earthquake Disaster Assessment: Objects, Data, Models, Stages, Challenges, and Opportunities,” *Remote Sens.*, vol. 15, no. 16, pp. 1–41, 2023, doi: 10.3390/rs15164098.
- [17] Z. Aizzah, P. K. Intan, and W. D. Utami, “Prediksi Jumlah Gempa Tektonik di Wilayah Jawa Timur dengan Menggunakan Metode ARIMA Box Jenkins dan Kalman Filter,” *JRST (Jurnal Ris. Sains dan Teknol.)*, vol. 5, no. 2, p. 111, 2022, doi: 10.30595/jrst.v5i2.9701.
- [18] S. M. Mousavi and G. C. Beroza, “Machine Learning in Earthquake Seismology,” *Annu. Rev. Earth Planet. Sci.*, vol. 51, pp. 105–129, 2023, doi: 10.1146/annurev-earth-071822-100323.
- [19] J. Song, J. Zhu, Y. Wang, and S. Li, “On-site alert-level earthquake early warning using machine-learning-based prediction equations,” *Geophys. J. Int.*, vol. 231, no. 2, pp. 786–800, 2022, doi: 10.1093/gji/ggac220.
- [20] J. Ir, S. No, and J. Tengah, “Prediksi Gempa Bumi di Yogyakarta Berdasarkan Nilai Magnitudo , Kedalaman , dan Lokasi Gempa Menggunakan Naïve Bayes,” *PETIR J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 17, no. 1, pp. 122–132, 2024.
- [21] D. Muhammad, I. Ahmad, M. I. Khalil, W. Khalil, and M. O. Ahmad, “A Generalized Deep Learning Approach to Seismic Activity Prediction,” *Appl. Sci.*, vol. 13, no. 3, pp. 1–13, 2023, doi: 10.3390/app13031598.
- [22] A. Fauzan and D. Ahmad, “Analisis Hasil Prediksi Magnitudo Gempa Di Wilayah Kota Padang Menggunakan Teknik Random Forest,” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 3, pp. 1569–1576, 2023, doi: 10.46306/lb.v4i3.450.
- [23] S. Susandri, A. Zamsuri, N. Nasution, Y. Efendi, and H. B. Alwan, “Mitigating Overfitting in Sentiment Analysis Insights from CNN-LSTM Hybrid Models,” *Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 24, no. 2, pp. 297–308, 2025, doi: 10.30812/matrik.v24i1.4742.
- [24] S. Susandri, S. Defit, and M. Tajuddin, “Enhancing Text Sentiment Classification with Hybrid CNN-BiLSTM Model on WhatsApp Group,” *J. Adv. Inf. Technol.*, vol. 15, no. 3, pp. 355–363, 2024, doi: 10.12720/jait.15.3.355-363.
- [25] S. K. C., A. Bhusal, D. Gautam, and R. Rupakhety, “Earthquake damage and rehabilitation intervention prediction using machine learning,” *Eng. Fail. Anal.*, vol. 144, p. 106949, 2023, doi: <https://doi.org/10.1016/j.engfailanal.2022.106949>.
- [26] A. Pyakurel, B. K. Dahal, and D. Gautam, “Does machine learning adequately predict earthquake induced landslides?,” *Soil Dyn. Earthq. Eng.*, vol. 171, p. 107994, 2023, doi: <https://doi.org/10.1016/j.soildyn.2023.107994>.
- [27] C. Cosgun, “Machine learning for the prediction of evaluation of existing reinforced concrete structures performance against earthquakes,” *Structures*, vol. 50, pp. 1994–2003, 2023, doi: <https://doi.org/10.1016/j.istruc.2023.02.127>.
- [28] N. S. M. Ridzwan and S. H. M. Yusoff, “Machine learning for earthquake prediction: a review (2017–2021),” *Earth Sci. Informatics*, vol. 16, no. 2, pp. 1133–1149, 2023, doi: 10.1007/s12145-023-00991-z.
- [29] P. Chittora et al., “Experimental analysis of earthquake prediction using machine learning classifiers, curve fitting, and neural modeling,” *PREPRINT*, vol. 2, pp. 1–26, 2022, [Online]. Available: <https://doi.org/10.21203/rs.3.rs-1896823/v2>
- [30] A. Berhich, F.-Z. Belouadha, and M. I. Kabbaj, “An attention-based LSTM network for large earthquake prediction,” *Soil Dyn. Earthq. Eng.*, vol. 165, p. 107663, 2023, doi: <https://doi.org/10.1016/j.soildyn.2022.107663>.
- [31] M. S. Abdalzaher, H. A. Elsayed, M. M. Fouda, and M. M. Salim, “Employing Machine Learning and IoT for Earthquake Early Warning System in Smart Cities,” *Energies*, vol. 16, no. 1, pp. 1–22, 2023, doi: 10.3390/en16010495.
- [32] M. S. Bandstra, J. C. Curtis, J. M. Ghawaly, A. C. Jones, and T. H. Y. Joshi, “Explaining machine-learning models for gamma-ray detection and identification,” *PLoS One*, vol. 18, no. 6 JUNE, pp. 1–21, 2023, doi: 10.1371/journal.pone.0286829.
- [33] E. Amico, L. Matteucci, and G. Cafiero, “Data-driven insights into jet turbulence: Explainable AI approaches,” *Phys. Fluids*, vol. 37, no. 5, p. 55108, May 2025, doi: 10.1063/5.0268702.