

Attention-based Residual Long Short-Term Memory for Earthquake Return Period Prediction in the Sulawesi Region

Muhdad Alfaris Bachmid¹, Daniel Febrian Sengkey^{2*}, Fabian Johan Manoppo³

^{1,2} Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado 95115, North Sulawesi, Indonesia

² Department of Research, Development, and Innovation, Indonesian Artificial Intelligence Society, Jl. Jend. Sudirman Kav 51, Jakarta Selatan 12930, Indonesia

³ Department of Civil Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado 95115, North Sulawesi, Indonesia

**Correspondence: danielsengkey@unsrat.ac.id*

Abstract

Indonesia, particularly the Sulawesi region, experiences significant seismic activity due to its position at the convergence of three major tectonic plates. This study seeks to construct a model for predicting earthquake return periods in the Sulawesi area by employing the Residual Long Short-Term Memory (Residual LSTM) architecture integrated with an attention mechanism. The dataset utilized originates from the United States Geological Survey (USGS), focusing on the Sulawesi Island region within the coordinates of latitude -6.184° to 2.021° and longitude 118.433° to 125.552° , spanning the years 1975 to 2024. The research methodology is structured into three primary phases: (1) data collection and preprocessing, including data cleaning, missing value handling, and normalization, (2) exploratory data analysis to understand seismic data characteristics, and (3) development of the Residual LSTM model with an attention mechanism. The evaluation results show excellent model performance with Train Loss 0.0090, Test Loss 0.0091, Training MAE 0.0698, Testing MAE 0.0717, Training RMSE 0.0947, Testing RMSE 0.0951, and stable Huber Loss of 0.0045 for both training and testing data. The implementation of residual connections successfully addressed the vanishing gradient problem, while the attention mechanism enhanced prediction interpretability. The small discrepancy between the training and testing metrics confirms the model's robust generalization ability, indicating its strong potential for applications in predicting earthquake return periods.

Keywords: *earthquake prediction, LSTM, return period, time series analysis, Sulawesi*

1. Introduction

Lying in one of the most seismically active areas, the intersection of three major tectonic plates: the Eurasian, Indo-Australian, and Pacific, renders Indonesia highly vulnerable to frequent earthquakes with diverse magnitudes and recurrence intervals [1]. Based on the data from the Indonesian Agency for Meteorological, Climatological and Geophysics (Badan Meteorologi, Klimatologi, dan Geofisika [BMKG]), a notable rise in both the frequency and intensity of seismic events across Indonesia has been detected over the past two decades. Updates to earthquake hazard maps further highlight the country's exposure to significant seismic activity, driven by the complex interactions of these tectonic plates [2].

Ongoing seismic activity in highly active regions such as Sulawesi underscores the need for regular updates to seismic parameters to improve the precision of predictions and risk assessments [3]. The significance of earthquake prediction research extends beyond disaster risk reduction to include minimizing economic losses and preserving human lives [4]. There is a pressing need for accurate earthquake prediction models, with one promising strategy involving the application of machine learning and time series analysis to estimate earthquake return periods [5]. The integration of advanced analytical techniques with seismic data provides valuable insights for seismic risk mitigation. Delivering accurate predictive information can enhance disaster management strategies and substantially decrease the risks associated with seismic events [6].

Machine learning has demonstrated its effectiveness across various domains for analyzing and predicting complex, dynamic data. Highlighted the adaptability of machine learning techniques for analyzing large datasets and generating accurate predictions[7]. In the context of earthquake prediction, machine learning offers a powerful means of uncovering latent patterns and improving predictive accuracy[8]. Prior research has demonstrated the success of machine learning techniques in forecasting seismic events, showing that such models can yield predictions closely aligned with actual seismic data[9].

Numerous research efforts have utilized machine learning approaches to address the complexities of earthquake prediction, employing techniques such as Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machines (GBM). Notably, Random Forest has demonstrated considerable effectiveness in improving prediction accuracy within certain seismic zones[10]. Moreover, hybrid architectures such as CNN-BiLSTM models integrated with attention mechanisms have yielded promising outcomes in seismic forecasting applications. Additionally, LightGBM has been implemented to assess seismic stability, leading to enhanced precision in predicting earthquake locations and magnitudes[11].

The earthquake return period is a fundamental concept in seismology, representing the average time interval between major seismic events within a specific geographic area. Previous research has investigated seismic hazards and the recurrence patterns of shallow earthquakes in the Cianjur region, utilizing United States Geological Survey (USGS) earthquake catalog data from a 50-year period (1973–2023) and geographic coordinates between 06°–08°S and 106°–108°E. The analysis indicated return periods ranging from 3.17 to 29.1 years for earthquakes with magnitudes between 5.0 and 6.0, and from 29.1 to 267.38 years for those with magnitudes between 6.0 and 7.0[12].

Recent studies have identified the Long Short-Term Memory (LSTM) algorithm as a prominent approach for stock market forecasting, owing to its robust ability to model and capture temporal dependencies. LSTM has been identified as one of the most effective techniques for achieving high accuracy and low error rates in time-dependent data contexts[13]. Expanding upon these advancements, the current research endeavors to construct a machine learning-based model for predicting earthquake return periods through time series analysis. The model integrates fundamental seismological concepts with data sourced from the United States Geological Survey (USGS) Earthquake Repository, which offers extensive

records on earthquake locations, depths, magnitudes, and long-term seismic trends. This methodological framework aims to address the limitations of prior studies by employing rigorously validated time series analytical techniques.

The core strength of this research lies in the implementation of a Residual Long Short-Term Memory (Residual LSTM) architecture integrated with attention mechanisms, tailored to forecast earthquake return periods within the geologically complex environment of Sulawesi Island. This methodological integration not only enhances predictive precision but also increases the model's contextual relevance for disaster mitigation efforts specific to Indonesia. By generating accurate and reliable return period predictions, this study contributes meaningfully to seismic risk reduction. The proposed model provides a robust tool for guiding resource allocation, supporting timely preventive strategies, and ultimately reducing the consequences of earthquakes, including casualties, infrastructure damage, and economic disruption.

2. Related Work

Earthquakes occur as a result of stress release within the Earth's crust, producing seismic waves that travel through its internal layers[14]. Indonesia, located at the intersection of four major tectonic plates, is particularly susceptible to frequent seismic activity due to the presence of active fault zones[15]. Accurately estimating earthquake recurrence intervals necessitates the application of advanced time series analysis, which enables the modeling of sequential data for forecasting purposes[16]. The Long Short-Term Memory (LSTM) network, a specialized deep learning architecture tailored for time series analysis, effectively models both short-term and long-term temporal dependencies, rendering it highly appropriate for earthquake prediction tasks.

The integration of Residual Long Short-Term Memory (Residual LSTM) with attention mechanisms further enhances pattern recognition capabilities, particularly in seismically active areas such as Sulawesi. This approach improves the model's ability to retain long-term dependencies and has demonstrated efficacy in various domains. For example, a Bi-LSTM model incorporating attention mechanisms was effectively utilized for sentiment analysis of Indonesia's Digital Identity Application (IKD), achieving an accuracy of 96.06%[17]. These successful implementations underscore the capacity of deep learning models to process complex datasets and extract meaningful patterns across diverse applications.

Return period analysis estimates earthquake

recurrence intervals based on magnitude and frequency. A study on shallow earthquakes in Cianjur, Indonesia, calculated recurrence intervals of 3.17–29.1 years for M5.0–6.0 quakes and 29.1–267.38 years for M6.0–7.0 quakes using USGS data (1973–2023) [12]. Similarly, Efrina et al. analyzed seismotectonic parameters in Nias, revealing critical insights into recurrence patterns.

Machine learning models have demonstrated high accuracy in earthquake prediction. Random Forest has been effective in magnitude forecasting [8], while XGBoost has shown promise in earthquake detection. Somantri et al. [18] optimized Support Vector Machines (SVM) for predicting earthquake magnitudes, achieving strong predictive performance through optimized windowing. Convolutional Neural Networks (CNNs) have also been employed to detect earthquake sources using spatial data. Model validation relies on metrics such as RMSE, MAE, and MAPE to ensure reliability [7].

The Vanishing Gradient Problem (VGP) hampers RNN-based models in capturing long-term dependencies [19]. To address this, Residual LSTM incorporates residual connections to preserve gradient flow, improving accuracy in return period prediction. A parallel can be drawn to Turkish classical music analysis, where complex sequential patterns in makam structures mirror earthquake time-series data challenges [20]. Similarly, Residual LSTM has enhanced transformer polarization current (PDC) forecasting, significantly reducing measurement errors [21]. This motivates its application in earthquake prediction, enabling robust handling of noisy, incomplete data and improving forecasting accuracy for disaster risk mitigation.

3. Methodology

3.1 Research Design

This research centers on forecasting earthquake recurrence intervals through a Residual Long Short-Term Memory model, organized into three principal stages. The first phase involves collecting earthquake data from the USGS earthquake repository, followed by comprehensive preprocessing steps, including data inspection, handling missing values, feature normalization, and feature engineering to create meaningful predictors such as annual average magnitude and depth trends over time.

The second stage entails performing Exploratory Data Analysis (EDA) to extract meaningful insights from the seismic dataset. This encompasses time series analysis for uncovering temporal trends, magnitude distribution

assessment, and geospatial visualization to examine the spatial distribution of earthquake events. Clustering techniques reveal geographical patterns linked to tectonic activity, while variable correlation analysis explores interdependencies among seismic attributes. Additionally, return period estimation is performed using statistical methods to analyze earthquake recurrence intervals.

The third phase concentrates on the development of the Residual LSTM model, incorporating residual connections and an attention mechanism to address the Vanishing Gradient Problem (VGP) in recurrent neural networks. The dataset is divided into training and testing subsets, and the model undergoes thorough evaluation via k-fold cross-validation, utilizing metrics including RMSE, MAE, and Huber Loss. The model's predictions on earthquake recurrence periods are visualized to assess its accuracy and reliability.

This research follows an iterative optimization cycle, where evaluation results refine preprocessing strategies and model parameters, continuously enhancing prediction accuracy. The workflow, depicted in Figure 2, ensures a structured and comprehensive approach to earthquake recurrence prediction, ultimately contributing to improved disaster mitigation strategies.

3.2 Data Description and Collection

The dataset employed in this study was obtained from the USGS Earthquake Catalog, encompassing seismic events in the Sulawesi region spanning the years 1975 to 2024. This dataset provides a comprehensive timeline to analyze long-term seismic trends and recurrence intervals. Key variables include the date and time of occurrence, geographical coordinates (latitude: -6.184° to 2.021°, longitude: 118.433° to 125.552°), and depth (in kilometers), which is crucial for classifying seismic events and assessing their potential impact. Additionally, the dataset records magnitude values (≥ 2.5 on the Richter scale), ensuring the inclusion of significant seismic activities relevant to earthquake prediction models. These parameters collectively support a detailed examination of seismic patterns in the region.

Sulawesi was chosen as the study area because of its distinct tectonic features and elevated seismic activity. The region lies at the convergence of three major tectonic plates: Indo-Australian, Pacific, and Eurasian. The interplay between these tectonic plates governs the activity of the Palu-Koro Fault, characterized by an estimated slip rate of 45 mm per year, thereby

rendering Sulawesi one of Indonesia's most seismically active regions[22].

Additionally, fault movements in Sulawesi occur at a significantly higher rate compared to other regions in Indonesia, making it one of the most seismically active areas in the country. The Palu-Koro Fault, a major left-lateral strike-slip fault in central Sulawesi, exhibits a slip rate of approximately 40–45 mm per year, while the Matano Fault, an extension of the Palu-Koro system, moves at ~20 mm per year [23]. These findings highlight the greater seismic hazard potential in Sulawesi, necessitating advanced predictive models for earthquake return period estimation[24].

The data acquisition from the USGS adheres to a systematic procedure designed to guarantee accuracy and pertinence. Researchers initially established specific search parameters, encompassing geographic boundaries and a time frame. For this study, the parameters targeted Sulawesi, with latitudes spanning from -6.184° to 2.021° , longitudes from 118.433° to 125.552° , and a temporal coverage from 1975 to 2024. The spatial extent of the latitude and longitude ranges,

illustrated in Figure 1, was obtained from the United States Geological Survey (USGS) database.

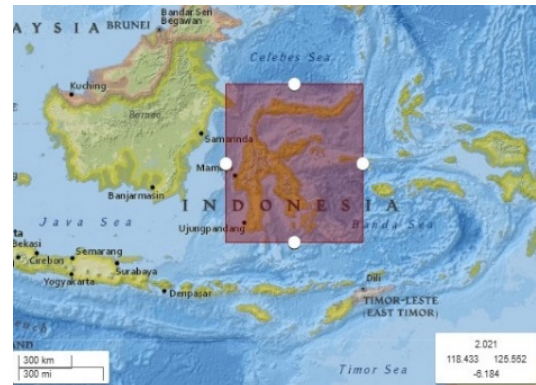


Figure 1. The spatial extent of the study area.

Data were collected through the USGS Earthquake Catalog search tool, which offers detailed information on earthquake events including date, time, latitude, longitude, depth, and magnitude. This tool allowed for the implementation of customized filters aligned with the study's objectives, ensuring the dataset's appropriateness for comprehensive seismic analysis and predictive modeling.

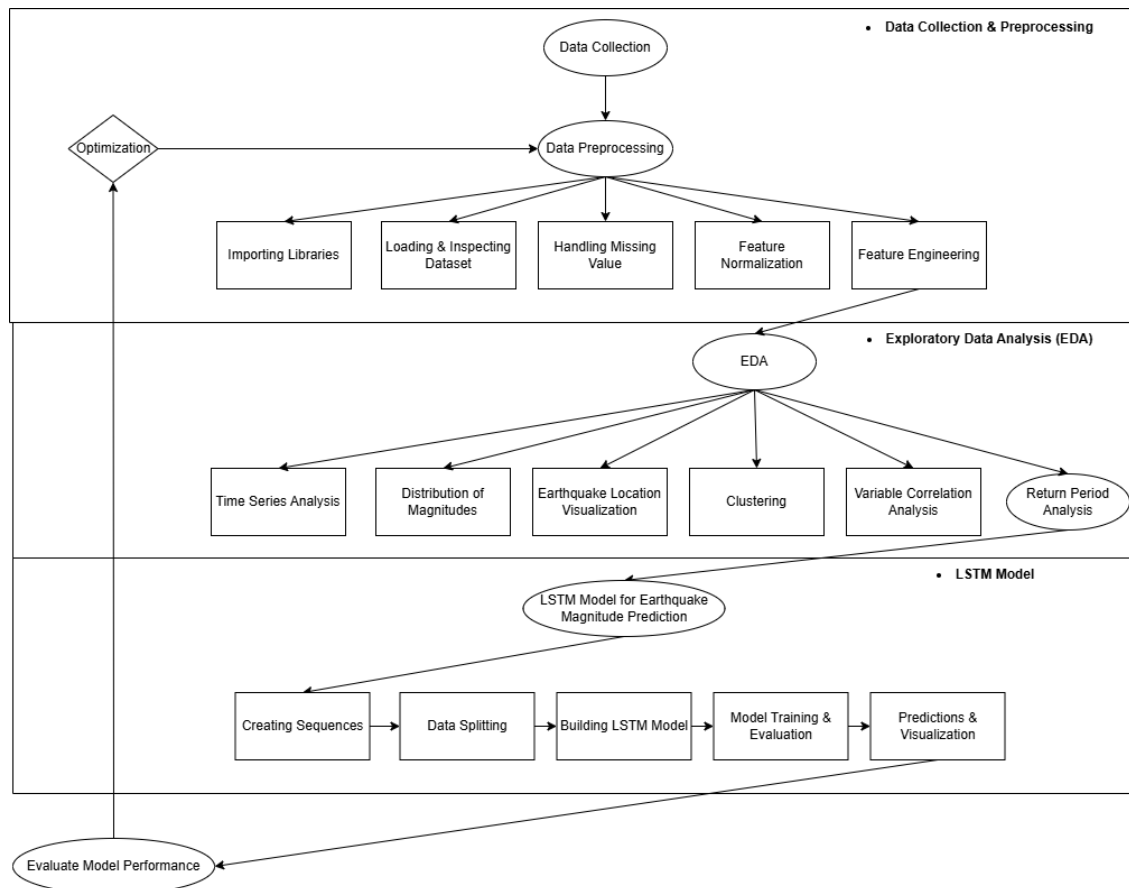


Figure 2. Research workflow.

4. Methodology Implementation

4.1 Data Preprocessing

Data preprocessing is a crucial step in ensuring the integrity and quality of earthquake data before model training. The dataset, obtained from the USGS earthquake repository, consists of 5,702 rows and 21 columns, covering key attributes such as time, latitude, longitude, depth, magnitude (mag), and additional seismic parameters. Initial exploration identified missing values in several columns, necessitating appropriate handling strategies. Missing data was addressed using two approaches, forward and backward fill for sequential continuity, and advanced imputation methods, where numerical attributes were replaced with their median values, and categorical attributes were imputed using mode-based replacement. To enhance computational efficiency, non-essential attributes such as 'id', 'updated', 'place', 'status', and 'color' were removed, reducing dataset dimensionality and retaining only the most relevant predictive features.

To maintain scale consistency, numerical features such as depth, magnitude, and latitude were normalized using the MinMaxScaler, which scales values to a range between 0 and 1. This normalization technique enhanced model convergence and stability, outperforming the StandardScaler. Furthermore, outliers were identified and removed employing the Interquartile Range (IQR) method, whereby data points falling below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ were classified as anomalies. This procedure excluded 150 extreme observations, thereby ensuring that the dataset comprised only statistically significant earthquake events.

4.2 Feature Engineering

To enhance the predictive capability of the model, feature engineering was performed using advanced statistical techniques, particularly Generalized Extreme Value (GEV) distribution analysis, to estimate earthquake recurrence intervals. The first step involved filtering seismic events, where only earthquakes with magnitude ≥ 5.0 were retained, removing lower-magnitude occurrences that do not significantly contribute to recurrence analysis. The filtered data was then fitted to a GEV distribution, estimating three key parameters: shape (ξ), which defines the tail behavior of extreme events, location (μ), representing the central tendency of the magnitude distribution, and scale (σ), describing the dispersion of magnitude values. These parameters

were estimated using the Maximum Likelihood Estimation (MLE) method to ensure optimal distribution fitting.

Using the estimated GEV parameters, the earthquake return period was determined through the GEV quantile function, specifically focusing on the 99th percentile, which represents extreme seismic events expected to occur once every 100 years. To enhance temporal resolution and provide a more detailed analysis of recurrence intervals, the return period was further converted into months and years using the transformation $\text{Return Period (months)} = \text{Return Period (years)} \times 12$. This conversion allows for finer-scale assessments of earthquake recurrence trends, improving the model's ability to capture variations over different timeframes and enhancing its applicability for long-term seismic risk analysis.

This transformation allowed a more granular analysis of seismic patterns. The newly engineered features, including `return_period (years)`, `return_period_months`, and `significant_magnitude` (a binary indicator where 1 represents significant earthquakes), were incorporated into the dataset to enhance model interpretability. These additional features played a critical role in improving model accuracy by enabling a more detailed representation of earthquake recurrence patterns, ultimately contributing to more reliable earthquake forecasting.

4.3 Return Period Analysis

The calculation of the earthquake return period is conducted using the quantile function of the Generalized Extreme Value (GEV) distribution, utilizing previously estimated distribution parameters. The return period represents the expected recurrence interval of extreme earthquake events exceeding a predefined magnitude threshold. In this study, the 99th percentile quantile function is employed, corresponding to earthquakes projected to occur once every 100 years. By leveraging this statistical approach, the model provides a probabilistic estimate of recurrence intervals for significant seismic events, offering a robust foundation for long-term seismic hazard assessment.

To enhance temporal granularity, the return period, initially expressed in years, is further converted into months and years. This transformation improves the model's ability to capture short-term variations in earthquake recurrence, ensuring more detailed insights into seismic trends. The conversion is performed using the formula $\text{Return Period (months)} = \text{Return Period (years)} \times 12$, enabling flexible temporal

analysis for disaster preparedness and risk mitigation efforts.

Following the return period estimation and conversion, newly derived features are incorporated into the dataset to improve predictive accuracy. These include `return_period` (the estimated recurrence interval in years), `return_period_months` (the same interval expressed in months), and `significant_magnitude` (a binary indicator, where 1 represents high-magnitude earthquakes exceeding the threshold and 0 denotes less significant tremors). The addition of these features allows the model to focus on high-impact seismic events, reinforcing its capability to analyze spatiotemporal earthquake patterns and enhancing the reliability of long-term earthquake recurrence predictions.

4.4 LSTM Model for Earthquake Magnitude Prediction

4.4.1 Attention Layer Creating Sequence

The `AttentionLayer` initializes its weight and bias parameters, where weights are sampled from a random normal distribution and biases are initialized to zero. Within the call method, attention scores are calculated by applying the tanh activation function to the matrix product of the input and the trained weights, followed by the addition of the bias term. Subsequently, the attention weights are obtained using the softmax function, ensuring that they sum to one for each input sequence.

Utilizing these attention weights, a context vector is generated through the weighted sum of the input, enabling the model to emphasize the most relevant portions of the data when predicting earthquake magnitudes. This attention mechanism not only improves the model's interpretability but also enhances its ability to capture complex patterns present in seismic data.

For preparing seismic data to train the LSTM model, the `create_sequences` function is applied to

generate input sequences from the dataset. This function accepts the dataset and the desired sequence length as inputs, producing two arrays: one containing the input sequences and another with the corresponding target outputs.

The sequence generation process involves determining the total number of sequences based on the dataset size and the specified sequence length. Empty arrays are initialized to store both sequences and targets. During iteration over the dataset, each sequence is extracted as a contiguous segment of the defined length, while the target corresponds to the immediately subsequent data point following the sequence.

The output of this function is a three-dimensional array shaped as `(number_of_sequences, sequence_length, number_of_features)`, which conforms to the input requirements of the LSTM model. This structure allows the LSTM to process the data sequentially and effectively learn temporal dependencies within the seismic dataset.

4.4.2 Model Architecture

The Residual Long Short-Term Memory (LSTM) model enhanced with an attention mechanism, as depicted in Figure 3, is designed to predict the earthquake return period by capturing temporal patterns in seismic data. The architecture begins with an input layer (`seq_length, n_features`) (`seq_length, n_features`), followed by Layer Normalization to stabilize training. The first LSTM layer (64 units, ReLU, `return_sequences=True`) processes sequential data, followed by a dropout layer (0.2) to prevent overfitting. A dense layer (`n_features n_features`) is added before a residual connection, which helps maintain gradient flow and improve learning efficiency.

A custom attention layer refines feature importance, followed by a second LSTM layer (64 units, ReLU, `return_sequences=True`). Another dropout layer (0.2) and a Global Average Pooling layer reduce dimensionality while preserving key

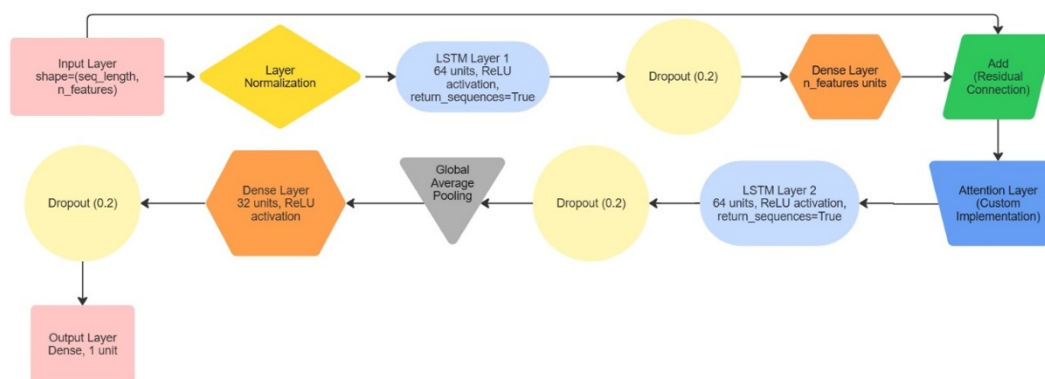


Figure 3. Model Architecture

Table 1. Model performance analysis.

Metrics	Training Data	Test Data
Train Loss	0.0090	0.0091
Train MAE	0.0698	0.0717
Train RMSE	0.0947	0.0951
Huber Loss	0.0045	0.0045

Table 2. Architecture and characteristics of the LSTM Residual model and the LSTM model.

Characteristics	Model 1 (Residual LSTM + Attention)	Model 2 (LSTM)
Main Architecture	LSTM with residual connection	2 layers (64 units per layer)
LSTM Layers	2 layers (64 units per layer)	2 layers (64 units per layer)
Dropout Rate	0.2 in each layer	0.2 in each layer
Layer Normalization	Yes	Yes
Attention Mechanism	Yes	Yes
Dense Layers	32 units with ReLU	32 units with ReLU
Output Layer	Dense(1)	Dense(1)

Table 3. Performance metrics of the LSTM Residual model and the LSTM model.

Metrics	Model 1 (Residual LSTM + Attention)	Model 2 (LSTM)
Train RMSE	0.0947	0.0951
Test RMSE	0.0951	0.0952
Train MAE	0.0698	0.0707
Test MAE	0.0717	0.0722
Train Huber Loss	0.0045	0.0045
Test Huber Los	0.0045	0.0045

Table 4. Comparison of earthquake return period prediction models and seismic data processing frameworks.

Method	Approach	Primary Objective	Input Data	Output
Residual LSTM + Attention Mechanism	Deep Learning (RNN-based)	Predicting earthquake return periods based on historical seismic data	Time-series seismic data	Estimated time of the next earthquake event
PhaseNet	Deep Learning (CNN-based)	Automatic detection and picking of seismic phase arrival times	Raw seismic waveforms	Arrival times of P and S waves
SeisBench	Machine Learning Framework	Evaluation and development of AI models for seismic analysis	Seismic datasets and earthquake catalogs	Varies depending on the selected model

information. A dense layer (32 units, ReLU) further extracts relevant features before a final dropout layer (0.2). The model concludes with an output dense layer (1 unit) to estimate the earthquake return period. The integration of residual connections and attention mechanisms in Figure 3 enhances feature extraction and improves predictive accuracy.

4.4.3 Building the LSTM Model

The proposed Residual Long Short-Term Memory (LSTM) model integrated with an attention mechanism is designed to effectively capture temporal dependencies while addressing the vanishing gradient issue commonly encountered in recurrent neural networks (RNNs). The architecture comprises two stacked LSTM layers, each consisting of 64 units activated by ReLU functions to introduce non-linearity and

enhance learning capabilities. The first LSTM layer is configured with the parameter `return_sequences=True`, ensuring that the sequential output is passed on to the subsequent layer. To mitigate overfitting, a dropout rate of 20% is applied immediately following the first LSTM layer.

To improve gradient propagation and facilitate the learning of long-term dependencies, a residual connection is incorporated between the input layer and the output of the first LSTM layer. Since residual connections require matching dimensions, a Dense layer is employed to adjust the output dimensions of the dropout layer to align with the input dimensions. This residual mapping allows the model to retain essential information from the original input sequence while enabling deeper feature extraction, ultimately improving training stability and model convergence.

Subsequent to the residual connection, the

output is directed through an Attention Layer that calculates attention scores by performing matrix multiplication between the input sequence and the learned attention weights. These scores are then normalized using the softmax activation function, ensuring that the attention weights sum to one over the sequence length. This attention mechanism allows the model to emphasize crucial time steps, thereby enhancing predictive accuracy by focusing on important temporal features within earthquake recurrence intervals.

The second LSTM layer, identical to the first with 64 units and a 20% dropout rate, processes the refined feature representation obtained from the attention layer. After passing through the LSTM layers, Global Average Pooling (GlobalAveragePooling1D) is applied, reducing the sequence dimension to a single feature vector that encapsulates the extracted temporal patterns.

The model then incorporates fully connected Dense layers to generate the final output. The first Dense layer (32 units, ReLU activation) refines feature representations, followed by an additional dropout layer to mitigate overfitting. The final output layer is a single-unit Dense layer, aligning with the regression task of predicting earthquake return periods. The model is compiled using the Adam optimizer with a learning rate of 0.001, and Mean Squared Error (MSE) is selected as the loss function, ensuring penalization of large errors in prediction.

To mitigate overfitting and improve generalization, EarlyStopping is employed to monitor the validation loss (val_loss), terminating training if no improvement occurs within a specified number of epochs. This approach ensures optimal model training by avoiding excessive iterations that may lead to performance deterioration.

4.5 Reproducible Research

The data and code utilized in this study are publicly available on GitHub at <https://github.com/muhdadbachmid/EQ-ReturnPeriod-AttentionRLSTM>.

5. Model Training and Evaluation

The Residual LSTM with Attention Mechanism was trained using the Adam optimizer, selected for its efficiency in handling large-scale datasets and its adaptive learning rate properties, which facilitate stable convergence. The learning rate was set to 0.001, ensuring a balance between convergence speed and stability. The model's loss function was Mean Squared Error (MSE), chosen

for its effectiveness in minimizing prediction deviations by penalizing larger errors more significantly. To prevent overfitting, the EarlyStopping callback was implemented, monitoring the validation loss (val_loss) and terminating training if no improvement was observed after 15 consecutive epochs. The restore_best_weights=True parameter was applied to ensure that the model retained the optimal weight configuration corresponding to the lowest validation loss, thereby maintaining predictive accuracy. Additionally, min_delta=1e-4 was defined as the minimum required improvement in validation loss to prevent premature stopping while avoiding excessive training.

The training phase was conducted over 150 epochs with a batch size of 32, balancing computational efficiency and gradient stability. To evaluate model generalization, 20% of the training dataset was set aside for validation, enabling real-time monitoring of predictive performance. Upon completion of training, model evaluation was conducted using both the training and test datasets, where the generated predictions were compared against actual values. The assessment framework incorporated Train Loss, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Huber Loss as the primary evaluation metrics. The performance results, summarized in Table 1, demonstrated the model's ability to minimize errors across both datasets. The small discrepancy between training MAE (0.0698) and test MAE (0.0717) indicated that overfitting was effectively mitigated, confirming strong generalization capabilities. Furthermore, the Huber Loss maintained a consistent value of 0.0045 across both training and testing datasets, confirming the model's robustness in managing outliers, an essential aspect given the presence of extreme values in seismic data.

The Residual LSTM with Attention Mechanism demonstrated strong capability in capturing complex temporal dependencies within earthquake recurrence data. The low RMSE values (0.0947 for training and 0.0951 for testing) indicated that prediction errors remained controlled, supporting the effectiveness of the model architecture. The integration of attention mechanisms enabled the model to selectively focus on significant temporal patterns, thereby enhancing interpretability and optimizing predictive accuracy. Furthermore, the inclusion of Huber Loss contributed to greater stability by reducing the influence of anomalous seismic events, ensuring that the model maintained predictive consistency despite fluctuations in data distributions.

The findings indicate that the Residual LSTM with Attention Mechanism is capable of producing

highly accurate and generalizable predictions for earthquake recurrence intervals. The model effectively minimizes errors while maintaining resilience against extreme values, demonstrating its applicability in forecasting seismic events with a high degree of reliability. These results suggest that deep learning approaches can provide robust, data-driven insights into earthquake recurrence trends, contributing to enhanced decision-making in earthquake risk assessment and mitigation strategies.

The proposed model, which integrates time-series feature extraction, LSTM-based deep learning, and attention-based optimization, is designed to be adaptable beyond Sulawesi and can be applied to other earthquake-prone regions such as Sumatra, Java, Bali, and Papua. These regions, like Sulawesi, are situated along the Ring of Fire, where active subduction zones and transform faults contribute to frequent and high-magnitude seismic events. However, to ensure the model's robustness and accuracy across different regions, several adaptations are necessary.

First, regional data adaptation is required, as each region exhibits distinct seismic characteristics, including variations in fault movement rates, stress accumulation patterns, and recurrence intervals. Incorporating region-specific historical earthquake data can enhance the model's predictive performance. Second, hyperparameter fine-tuning may be necessary to adjust LSTM architecture, learning rates, and attention weights, ensuring that the model effectively captures the unique seismic behavior of different regions. Finally, validation using diverse datasets is essential to assess the model's generalization capability. These adaptations will enable the model to be scalable and applicable to various seismically active regions, supporting more accurate earthquake forecasting and enhanced disaster risk mitigation strategies.

6. Model Comparison of Earthquake Prediction Models

6.1 Comparison with Standard LSTM Models

In this study, the model architecture design is a crucial step in generating accurate and reliable predictions, particularly in analyzing earthquake data with complex spatiotemporal patterns. Table 2 presents a comparative overview of two implemented model architectures, Model 1 (Residual LSTM + Attention) and Model 2 (LSTM). This comparison aims to identify the contribution of each architectural element in enhancing model performance, including their ability to capture long-term data patterns, mitigate

the vanishing gradient problem, and produce more robust predictions.

Model 1 (Residual LSTM + Attention) offers significant advantages over Model 2 (LSTM) due to the incorporation of residual connections and an attention mechanism. This combination not only improves prediction accuracy but also enhances the model's ability to capture both temporal and spatial dependencies in complex earthquake data. In contrast, Model 2, which adopts a standard LSTM architecture, is simpler but less effective in leveraging intricate data patterns.

Table 3 presents the performance evaluation results of Model 1 (Residual LSTM + Attention) and Model 2 (LSTM) based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Huber Loss for both training and testing datasets. The results indicate that Model 1 outperforms Model 2 in terms of RMSE and MAE. For the training data, Model 1 achieves an RMSE of 0.0947, which is slightly lower than Model 2's RMSE of 0.0951. A similar trend is observed in the testing data, where Model 1 records an RMSE of 0.0951, compared to 0.0952 for Model 2.

In terms of MAE, Model 1 also demonstrates lower absolute errors, with 0.0698 on the training data and 0.0717 on the test data, whereas Model 2 reports MAE values of 0.0707 and 0.0722 for training and test data, respectively. However, for Huber Loss, both models produce identical results, 0.0045 for both training and test data, indicating similar capabilities in handling extreme errors or outliers.

Overall, these findings suggest that Model 1, with residual connections and an attention mechanism, provides more accurate predictions than Model 2, albeit with marginal differences. This highlights the importance of adopting a more complex architecture, such as attention mechanisms, to enhance the model's ability to capture intricate spatiotemporal patterns in earthquake data. As reflected in Table 3, the superior performance of Model 1 demonstrates that this approach is more effective for earthquake prediction applications.

6.2 Comparison of Earthquake Return Period Prediction Models and Seismic Data Processing Frameworks

The comparison of earthquake return period prediction models and seismic data processing frameworks, as detailed in Table 4, highlights the distinct objectives and methodologies of each approach. The Residual LSTM with Attention Mechanism employs a Recurrent Neural Network (RNN) architecture, specifically Long Short-Term Memory (LSTM) networks, enhanced with

residual connections and an attention mechanism to predict earthquake recurrence periods. By analyzing time-series seismic data, this model estimates the timing of future seismic events. The integration of residual connections mitigates the vanishing gradient problem, while the attention mechanism improves model interpretability by emphasizing significant features, leading to enhanced prediction accuracy.

In contrast, PhaseNet is a Convolutional Neural Network (CNN)-based model designed for the automatic detection and picking of seismic phase arrival times, specifically P and S waves. This model processes raw seismic waveforms to accurately determine arrival times, which are crucial for earthquake location and characterization. PhaseNet demonstrates high accuracy and real-time processing capabilities, making it valuable for rapid seismic event analysis [25]. Meanwhile, SeisBench serves as a comprehensive machine learning framework that facilitates the evaluation and development of AI models for seismic analysis. It provides access to various seismic datasets and earthquake catalogs, supporting tasks such as event detection and phase picking. The framework's versatility allows researchers to implement and benchmark different models, contributing to advancements in seismological studies [26].

7. Conclusion

The Residual LSTM with an integrated attention mechanism exhibited remarkable performance in forecasting earthquake recurrence intervals in the Sulawesi region. This is supported by consistently strong evaluation metrics: a Train Loss of 0.0090 and Test Loss of 0.0091, Training MAE of 0.0698 and Testing MAE of 0.0717, Training RMSE of 0.0947 and Testing RMSE of 0.0951, and a stable Huber Loss of 0.0045 for both training and testing datasets. These results highlight the model's accuracy, generalization ability, and resilience in handling the variability and extremities characteristic of seismic data.

The model's strength lies in its residual connections, which effectively address the vanishing gradient problem by allowing gradients to flow more smoothly through the network layers. This facilitates efficient learning of long-term dependencies, which are essential in modeling seismic recurrence intervals. Furthermore, the integration of an attention mechanism empowers the model to dynamically focus on the most relevant portions of the input sequence. By assigning adaptive weights to different time steps, the attention layer not only enhances predictive performance but also increases model

interpretability, enabling a clearer understanding of which temporal patterns most influence the earthquake recurrence predictions.

The model's stability is further evidenced by the negligible difference between training and testing metrics, indicating robust generalization performance. Moreover, its consistent results across multiple evaluation metrics and resilience to outliers, as demonstrated by the low Huber Loss, highlight the dependability of this method.

This study contributes a significant methodological advancement by proposing a novel framework for forecasting earthquake recurrence. The successful integration of Residual LSTM with an attention mechanism validates the effectiveness of deep learning in analyzing seismic data, thereby facilitating wider applications in earthquake prediction research.

8. Future Works

This study acknowledges several limitations that warrant consideration for future model enhancement. A primary limitation lies in the exclusion of comprehensive geological information, such as rock formations and fault line characteristics, which are critical to accurately modeling earthquake recurrence. Furthermore, the omission of environmental variables, including surface alterations and anthropogenic influences, may hinder the model's ability to generalize effectively. Incorporating these elements in future work is expected to strengthen the model's robustness and predictive precision.

Based on the research findings and performance evaluations, several important recommendations are proposed to improve the earthquake return period prediction model. Firstly, integrating the Residual LSTM model with other machine learning approaches, such as Resilient Propagation (RPROP), is essential for enhancing predictive accuracy and speeding up convergence. RPROP has proven effective in seismic activity prediction, and its combination with Residual LSTM can result in a more adaptive and robust model. Additionally, the application of ensemble learning techniques, like XGBoost, can further boost predictive performance by capturing complex, non-linear relationships within the data. The integration of Residual LSTM for modeling temporal dependencies and XGBoost for recognizing non-linear patterns is expected to strengthen the overall predictive capability of the system.

Developing a multi-task learning architecture that simultaneously predicts both earthquake recurrence periods and magnitudes is strongly advised. This can be realized by using a shared

feature extraction network, coupled with distinct output layers that are individually optimized for each task. By learning from both tasks simultaneously, the model can achieve more comprehensive seismic predictions while improving overall accuracy. Moreover, incorporating multidimensional data sources such as real-time tectonic plate movements, surface deformations, and meteorological information is vital for improving the robustness of predictions. These diverse data streams offer a comprehensive perspective on seismic events and their causative factors. Consequently, it is imperative to develop a high-performance data integration pipeline that can efficiently process and manage these heterogeneous sources, while implementing rigorous quality control measures to maintain prediction reliability.

The establishment of a real-time modular processing system, incorporating an interactive dashboard and early warning functionalities, is essential for effective practical deployment. Such a system should be capable of handling multiple data streams, delivering insightful visualizations, and generating automated reports to convey predictions clearly to relevant stakeholders. An interactive dashboard would enable users to monitor seismic risk levels in real time, evaluate potential threats proactively, and support data-driven decision-making for disaster preparedness. Furthermore, integrating a continuous performance monitoring framework with real-time evaluation metrics and anomaly detection would be critical for sustaining long-term model reliability. This system should also support automatic model updates based on newly acquired seismic data, thereby preserving predictive accuracy and adaptability over time.

To ensure the sustainability and adaptability of the system, the implementation of a knowledge management framework is recommended. This should include version-controlled model repositories, automated documentation, and training materials for end users. By maintaining comprehensive documentation and facilitating knowledge transfer, stakeholders and system operators can effectively manage, update, and optimize the predictive model in the future.

In conclusion, the gradual implementation of these recommendations, emphasizing model integration, incorporation of multidimensional data sources, and the establishment of continuous monitoring mechanisms, is anticipated to markedly enhance the accuracy and robustness of earthquake prediction models. By integrating real-time data sources, refining model architectures, and ensuring consistent evaluation and adaptation, the predictive capabilities of the model can be enhanced, leading to better-informed disaster risk management

strategies. These improvements are particularly crucial for Indonesia, a seismically active region, where robust predictive models can contribute to effective disaster preparedness and risk mitigation strategies. Ultimately, continuous advancements in these models will play a vital role in reducing earthquake-related risks and minimizing the potential impact on communities and infrastructure.

9. References

- [1] M. V. Brilliantina, H. Pratiwi, and Y. Susanti, "Analisis Seismisitas pada Data Gempa Bumi di Provinsi Maluku Utara," *Prosiding Seminar Pendidikan Matematika dan Matematika*, vol. 4, May 2021, doi: 10.21831/pspmm.v4i2.174.
- [2] M. Ridwan, A. Soehaimi, S. R. Sinung Baskoro, Y. Sopian, R. Setianegara, and A. Cita, "Pengembangan Peta Bahaya Gempabumi di Batuan Dasar untuk Daerah Cilacap dan Sekitarnya," *Jurnal Geologi dan Sumberdaya Mineral*, vol. 24, no. 1, pp. 31–38, Jan. 2023, doi: 10.33332/jgsm.geologi.v24i1.704.
- [3] J. R. Talumepa, F. J. Manoppo, and L. D. K. Manaroinson, "Respon Spektra pada Jembatan Ir. Soekarno Manado," *Jurnal Sipil Statik*, vol. 7, no. 7, 2019.
- [4] N. Nursyabani, R. E. Putera, and K. Kusdarini, "Mitigasi Bencana Dalam Peningkatan Kewaspadaan Terhadap Ancaman Gempa Bumi Di Universitas Andalas," *Jurnal Ilmu Administrasi Negara ASIAN (Asosiasi Ilmuwan Administrasi Negara)*, vol. 8, no. 2, pp. 81–90, Sep. 2020, doi: 10.47828/jianaasian.v8i2.12.
- [5] M. Nurtas, Z. Zhantaev, and A. Altaibek, "Earthquake time-series forecast in Kazakhstan territory: Forecasting accuracy with SARIMAX," *Procedia Comput Sci*, vol. 231, pp. 353–358, 2024, doi: 10.1016/j.procs.2023.12.216.
- [6] J. D. Dorathi Jayaseeli, D. Malathi, M. Sisodia, and C. Ruvinga, "Time series analysis for predicting environmental earthquakes using ARIMA model," *Journal of Green Engineering*, vol. 10, no. 1, pp. 76–90, 2020.
- [7] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan machine learning dalam berbagai bidang," *Jurnal Khatulistiwa Informatika*, vol. 5, no. 1, p. 490845, 2020.
- [8] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS: Indonesia*

- Journal Information System*, vol. 6, no. 2, pp. 83–89, 2023.
- [9] A. Fauzan and D. Ahmad, “Analisis Hasil Prediksi Magnitudo Gempa di Wilayah Kota Padang Menggunakan Teknik Random Forest,” *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 4, no. 3, pp. 1569–1576, Dec. 2023, doi: 10.46306/lb.v4i3.450.
- [10] Kholiq Budiman and Yahya Nur Ifriza, “Analysis of earthquake forecasting using random forest,” *Journal of Soft Computing Exploration*, vol. 2, no. 2, Sep. 2021, doi: 10.52465/josce.v2i2.51.
- [11] C. Wang, C. Li, S. Yong, X. Wang, and C. Yang, “Time Series and Non-Time Series Models of Earthquake Prediction Based on AETA Data: 16-Week Real Case Study,” *Applied Sciences*, vol. 12, no. 17, p. 8536, Aug. 2022, doi: 10.3390/app12178536.
- [12] S. Sutrisno, A. Tjahjono, and F. R. Putra, “Analysis of return period and seismic risk of Shallow Earthquake occurrence in Cianjur and surrounding areas,” *Journal of Natural Sciences and Mathematics Research*, vol. 9, no. 2, pp. 109–116, Dec. 2023, doi: 10.21580/jnsmr.2023.9.2.18098.
- [13] R. B. Wiranata and A. Djunaidy, “The Stock Exchange Prediction using Machine Learning Techniques: A Comprehensive and Systematic Literature Review,” *Jurnal Ilmu Komputer dan Informasi*, vol. 14, no. 2, pp. 91–112, Jul. 2021, doi: 10.21609/jiki.v14i2.935.
- [14] H. H. Abbas, Nurbaeti, and Andi Asrina, “Mitigasi Bencana Gempa Bumi dengan Metode Learning by Doing,” *Window of Health: Jurnal Kesehatan*, pp. 475–485, Jan. 2022, doi: 10.33096/woh.vi.139.
- [15] S. Al Faridzi *et al.*, “Pengolahan Data: Pemahaman Gempa Bumi di Indonesia Melalui Pendekatan Data Mining,” *Jurnal Pengabdian Kolaborasi dan Inovasi IPTEKS*, vol. 2, no. 1, pp. 262–270, Feb. 2024, doi: 10.59407/jpki2.v2i1.506.
- [16] E. Nur Cahyo and E. Susanti*, “Analisis Time Series Untuk Deep Learning Dan Prediksi Data Spasial Seismik: Studi Literatur,” *Jurnal Teknologi*, vol. 15, no. 2, pp. 124–136, Jan. 2023, doi: 10.34151/jurtek.v15i2.3581.
- [17] R. Onsu, D. Sengkey, and F. Kambey, “Implementasi Bi-LSTM dengan Ekstraksi Fitur Word2Vec untuk Pengembangan Analisis Sentimen Aplikasi Identitas Kependudukan Digital,” *Jurnal Teknologi Terpadu*, vol. 10, pp. 46–55, Jul. 2024, doi: 10.54914/jtt.v10i1.1225.
- [18] O. Somantri, S. Purwaningrum, and R. Riyanto, “Model Support Vektor Machine (Svm) Berdasarkan Parameter Windows Untuk Prediksi Kekuatan Gempa Bumi,” *JTT (Jurnal Teknologi Terapan)*, vol. 8, no. 1, pp. 17–24, 2022.
- [19] A. Rehmer and A. Kroll, “On the vanishing and exploding gradient problem in Gated Recurrent Units,” *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1243–1248, 2020, doi: 10.1016/j.ifacol.2020.12.1342.
- [20] F. K. Mirza, A. F. Gürsoy, T. Baykaş, M. Hekimoğlu, and Ö. Pekcan, “Residual LSTM neural network for time dependent consecutive pitch string recognition from spectrograms: a study on Turkish classical music makams,” *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 41243–41271, Oct. 2023, doi: 10.1007/s11042-023-17105-y.
- [21] A. Vatsa, A. S. Hati, P. Kumar, M. Margala, and P. Chakrabarti, “Residual LSTM-based short duration forecasting of polarization current for effective assessment of transformers insulation,” *Scientific Reports*, vol. 14, no. 1, p. 1369, Jan. 2024, doi: 10.1038/s41598-023-50641-z.
- [22] Y. Serhalawan and P.-F. Chen, “Seismotectonics of Sulawesi, Indonesia,” Jan. 24, 2023, doi: 10.22541/essoar.167458063.37633079/v1.
- [23] A. Khairi, M. Awaluddin, and B. Sudarsono, “Analisis deformasi seismik sesar matano menggunakan GNSS dan interferometrik SAR,” *Jurnal Geodesi Undip*, vol. 9, no. 2, pp. 32–42, 2020.
- [24] “Paleoseismic studies reveal the high seismic hazard potential of the Matano fault, Sulawesi, Indonesia,” *Past Global Changes Magazine*, vol. 32, no. 1, pp. 14–15, May 2024, doi: 10.22498/pages.32.1.14.
- [25] W. Zhu and G. C. Beroza, “PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method,” *Geophys J Int*, Oct. 2018, doi: 10.1093/gji/ggy423.
- [26] J. Woollam *et al.*, “SeisBench—A Toolbox for Machine Learning in Seismology,” *Seismological Research Letters*, vol. 93, no. 3, pp. 1695–1709, May 2022, doi: 10.1785/0220210324.