

# Optimizing Parameters for Earthquake Prediction Using Bi-LSTM and Grey Wolf Optimization on Seismic Data

Guruh Fajar Shidik<sup>1</sup>, Ricardus Anggi Pramunendar<sup>2\*</sup>, Purwanto<sup>3</sup>, Zainal Arifin Hasibuan<sup>4</sup>, Erlin Dolphina<sup>5</sup>, Yupie Kusumawati<sup>6</sup>, Nurul Anisa Sriwinarsih<sup>7</sup>

<sup>1, 2, 3, 4, 5, 6, 7</sup> Center for Intelligent Distributed Surveillance and Security, Universitas Dian Nuswantoro, Semarang, Central Java, 50131, Indonesia

Email: <sup>1</sup>guruh.fajar@research.dinus.ac.id, <sup>2</sup>ricardus.anggi@dsn.dinus.ac.id, <sup>3</sup>purwanto@dsn.dinus.ac.id, <sup>4</sup>zhasibua@dsn.dinus.ac.id, <sup>5</sup>erlindolphina@dsn.dinus.ac.id, <sup>6</sup>yupie.kusumawati@dsn.dinus.ac.id, <sup>7</sup>nurulanisasw@dsn.dinus.ac.id

\*Corresponding Author

**Abstract**—Earthquakes pose a significant threat to societies worldwide, underscoring the urgent need for advanced prediction technologies. This study introduces an optimization technique aimed at reducing the error rate in earthquake prediction by selecting the most suitable parameters for a Bi-LSTM (Bidirectional Long Short-Term Memory) model. Despite Bi-LSTM's promising outcomes, variations in parameters can impact performance, necessitating careful parameter selection. This research employs Grey Wolf Optimization (GWO) to optimize parameters and evaluates its effectiveness against other group optimization approaches to identify the most efficient parameters for earthquake prediction. Additionally, a multiple input multiple output (MIMO) architecture is implemented to enhance prediction accuracy. The evaluation results demonstrate that GWO outperforms other optimization techniques, achieving a reduced loss score of 0.364. The ANOVA method yields a p-value approaching 0, indicating statistical significance. This study contributes to the development of early warning systems for earthquake disasters by emphasizing the importance of parameter optimization in earthquake prediction and showcasing the effectiveness of Bi-LSTM and GWO methodologies.

**Keywords**—Earthquake Prediction; Bi-LSTM; Grey Wolf Optimization; Seismic Data; Parameter Optimization.

## I. INTRODUCTION

Earthquakes, considered perilous natural phenomena [1]–[3], happen abruptly and lead to a significant loss of life, surpassing 50% compared to other natural catastrophes [4]. Between 1998 and 2017, a total of 750,000 lives were claimed worldwide, affecting 125 million people and resulting in injuries, relocation, and property damage [4]. Earthquakes may be categorized into many categories, including tectonic, volcanic, and impact earthquakes [5]. These sorts of earthquakes are most common in areas where tectonic plates come together, particularly the Pacific Plate, which has caused significant disasters like the 1960 Chile Earthquake [1], [4], [6]. Due to increasing magnitudes, earthquakes provide substantial economic and material dangers [7]. This highlights the need for automated early warning systems, particularly in areas with high seismic activity, such as the Pacific Plate [8]. However, the

unexpected nature of the subject makes analysis challenging owing to the many elements at play.

Machine learning (ML) techniques have been extensively used in diverse fields, including hazard anticipation and projection, such as floods, earthquakes, and landslides [9]. The incorporation of ML and deep learning (DL) into seismology has facilitated the creation of applications such as early prediction of casualties, detection of precursor vibrations, and measurement of arrival times [10]–[13], providing an alternative approach to tackle significant obstacles in earthquake engineering [14].

Various ML approaches, such as neural networks (NN) [15], [16], support vector machines (SVM) [17], random forests (RF) [16], [18], and convolutional neural networks (CNN) [19], have been used in the study. Song et al. [17] recommended using SVM for earthquake early warning, while Lin et al. [15] put forth predictions using a two-layer back-propagation neural network (BPNN) for earthquakes in Taiwan. Murwantara et al. [20] performed a comparative examination of ML techniques, such as multinomial logistic regression, naïve Bayes (NB), and SVM, to forecast earthquakes in Indonesia. Although ML algorithms are competent in handling non-stationary earthquake data [16], they have difficulties in accurately recognizing surface-level earthquake features and need complex feature engineering to get excellent prediction results [12], [21].

Deep learning algorithms, developed to tackle optimization difficulties in machine learning algorithms, include intricate structures with strong generalization abilities, significantly improving their learning capacity in comparison to shallow neural networks [22], [23]. Wang et al. [19] proposed the use of CNN to forecast cumulative absolute velocity in earthquakes. In their study, Zhang and Wang et al. [24] suggested the use of multi-modal approaches, which include integrating sequence-to-sequence CNN and Long Short-Term Memory (LSTM) models, to predict earthquakes. Their research shows this approach has great potential for accurately predicting seismic occurrences. Sadhukhan et al. [25] assessed several DL methods, such as LSTM, Bi-LSTM, and Transformer Model, for predicting



earthquake magnitude. The results showed that LSTM performed better than the other models. One study conducted by Berhich et al. [4] used an LSTM network with an attention mechanism to forecast significant earthquakes. This approach led to a notable increase in accuracy. Abebe et al. [26] proposed a DL transformer technique for predicting earthquake magnitude. Their study demonstrated the system's capability compared to LSTM and Bi-LSTM architectures. Yoma et al. [27] introduced a method based on LSTM for accurately determining the location of volcano-seismic events. Their approach outperformed existing methods, attaining more excellent rates of success. Shidik et al. [8] proposed using a blend of activation functions in conjunction with Bi-LSTM to enhance earthquake prediction.

This research aims to identify the most effective technique for analyzing a unique dataset using the selected strategy. The criteria for achieving high performance must be clearly defined. Therefore, this study aims to determine the most practical combination of parameters and develop a novel procedure using an enhanced deep-learning framework to provide predictive seismic variables. Specifically, this study employs the Bi-LSTM (Bidirectional Long Short-Term Memory) architecture within a multiple input multiple output (MIMO) model and applies the Grey Wolf Optimization (GWO) approach for parameter selection to reduce prediction error rates [28], [29]. The GWO approach was chosen for its effectiveness in identifying the optimal combination of parameters, demonstrating its practical utility in diverse scenarios for determining optimum values for complex parameters. This capability makes GWO a compelling choice for this study, as it can facilitate the discovery of optimal parameter configurations for predictive models. Furthermore, GWO offers the advantage of rapid decision-making to achieve optimal outcomes [30]. An expedited optimization process can reduce the time required to find the best solution, thereby enhancing efficiency in meeting objectives. Consequently, the contributions of this research are twofold: first, the integration of GWO into the Bi-LSTM framework to optimize parameters, and second, the development of a novel workflow that significantly reduces the error rate in earthquake predictions.

The following sections delineate the organization of this article. The second sections presents a comprehensive summary of prior research that is relevant to the current topic. The final part outlines the experimental technique and presents our suggested model. The fourth portion gives a comprehensive analysis of the findings and assessment of the conducted experiments, while the fifth section presents a concise overview of the conclusions.

## II. RELATED RESEARCH

Earthquake prediction has been performed in recent years. Numerous research has attempted to use statistical methodologies, including FDL [31], HMM [32], and BLD [33]. Nevertheless, since earthquake data sets exhibit non-stationary properties, this technique fails to provide satisfactory estimates. Murwantara et al. (2020) have suggested alternative research approaches that use machine-learning techniques to address this issue. Their study compares several machine learning methods, including

Multinomial Logistic Regression, Naïve Bayes (NB), and Support Vector Machines (SVM). The research's objective is to calculate the precise coordinates, intensity, and depth of the earthquake by using the time and date of the occurrence, as well as the latitude, longitude, magnitude, and depth of the earthquake's epicenter. The prediction procedure begins with estimating the earthquake's location using latitude and longitude. Subsequently, the prediction outcomes are a reference point for forecasting the earthquake's magnitude and depth. The evaluation findings indicate that the SVM model surpasses other approaches in the overall assessment.

Nevertheless, machine learning models often exhibit limited seismic pattern recognition capabilities and need intricate feature engineering for satisfactory prediction outcomes [21], [34]. Thus, DL approaches address the limited comprehension of shallow features in machine learning models. Abri and Artuner [35] conducted a comparative analysis of LSTM, SVM, and Linear Discriminant Analysis (LDA) for earthquake prediction using ionospheric data. Based on the findings of this study, it is evident that LSTM successfully identified the earthquake that occurred on the previous day with an accuracy rate of 0.82. Still, SVM only attained an accuracy rate of 0.6. The DL technique used in this work exhibits superior and more dependable performance compared to the studied machine learning approaches. In addition, a further study conducted by Bhandarkar et al. [36], directly compared LSTM with Feed Forward Neural Network (FFNN). The results showed that LSTM achieved superior performance than FFNN, with a  $R^2$  score of 59%. Zhang and Wang [24] have suggested an alternative study that employs a fusion of advanced DL methodologies, particularly CNN and LSTM. This study utilizes spatial and temporal data. Both sets of data are transformed into a dataset with four dimensions. The assessment findings indicate that the mean recall and accuracy values are 51.83% and 64.54%, respectively.

Additionally, Kavianpour [21] included attention and zero-order hold (ZOH) processes in their previously presented study model, which initially used two-way-LSTM [21]. This study's objective is to forecast future earthquakes' peak intensity and frequency, using months with minimal margin of error. The findings of the study substantially enhance performance and generality compared to earlier methodologies. LSTM and Bi-LSTM are often used in forecasting, as previously stated. The distinction between the two approaches lies in using Bi-LSTM, which employs two LSTMs to process the input data. The first LSTM operates on the input sequence forward, while the second LSTM processes the input sequence backward [37]. Sadhukhan et al. [25] conducted a study to assess the performance of LSTM and Bi-LSTM architecture. They compared the outcomes of these models with the Transformer Model. This study forecasts the intensity of forthcoming earthquake occurrences by analyzing a dataset of climatic and seismic information gathered from three specific areas: Indonesia, the Himalayan region, and Japan. The study findings indicate that the LSTM architecture performs more than other models, particularly regarding the MAE, MSE, and log-cosh loss metrics.

Other research using LSTM has also been conducted by Liao et al., [38] introduced the attention-based LSTM

(AttLSTM), specifically tailored for predicting bridge behavior during earthquakes. By focusing on relevant information, the AttLSTM significantly enhances prediction accuracy compared to traditional LSTM architecture, as validated with actual bridge data. Berhich et al. [4], which explores an attention-based LSTM network to predict the timing, magnitude, and location of large earthquakes. The results significantly improve prediction accuracy over traditional methods. Research was also conducted by Abebe et al. [26], who proposed a DL transformer algorithm to predict earthquake magnitudes and compared it with several LSTM and Bi-LSTM architectures. These results indicate that the transformer model could be a valuable tool for earthquake prediction. Meanwhile, Yoma et al. [27] proposed an LSTM-based method for localizing volcano-seismic events that is superior to traditional automatic phase picking methods and also shows an 18% higher success rate compared to the CNN approach. Regarding the research conducted by Shidik et al. [8], they proposed a fusion of activation functions for earthquake prediction employing Bi-LSTM, resulting in significantly enhanced performance. The errors were notably reduced compared to those using the original ReLU activation function, with a decrease of up to 4% in MAE and 3% in MSE.

Several optimization techniques have been employed to enhance earthquake prediction performance. For instance, Asim et al. [6] combined seismic indicators with Genetic Programming (GP) and AdaBoost, resulting in a significant improvement in earthquake prediction accuracy. Additionally, Hason et al. [39] utilized Particle Swarm Optimization (PSO) for Peak Ground Acceleration (PGA), considering seismic parameters, yielding low error rates and high correlations. As for other optimization methods to monitor earthquake-affected areas or other events through various other methods [40]–[42].

However, the lack of these related research is use deep learning-based approaches, however most utilize varying parameters, often relying on default configurations. In this study, it is proposed to enhance performance by selecting parameters obtained from previously conducted earthquake prediction methods [8]. This study presents an earthquake prediction process that utilizes Bi-LSTM and optimizes parameter selection using the Gray Wolf Optimization (GWO) approach. The optimal parameters will extract output variables from multiple input variables, including latitude, longitude, magnitude, and depth information [8]. The GWO approach was chosen for parameter selection due to its proven ability to improve the performance of the underlying method [29]. In this research, GWO is used and compared with PSO and weight-optimized PSO optimization methods to identify the most influential parameters in the Bi-LSTM architecture.

### III. PROPOSED RESEARCH

This research establishes an updated workflow for earthquake prediction using deep learning (DL) techniques. The objective is to reduce the error in earthquake forecasting by analyzing data patterns and anomalies. The proposed approach involves optimizing parameters for dense layers, activation functions, and epochs. These specific parameters

are selected based on a comprehensive review of relevant literature and preliminary experiments, which demonstrate their significant impact on minimizing prediction errors. Fig. 1 represents the detailed procedure of the suggested workflow, which includes several phases and offers a clear visual representation to aid in understanding the process.

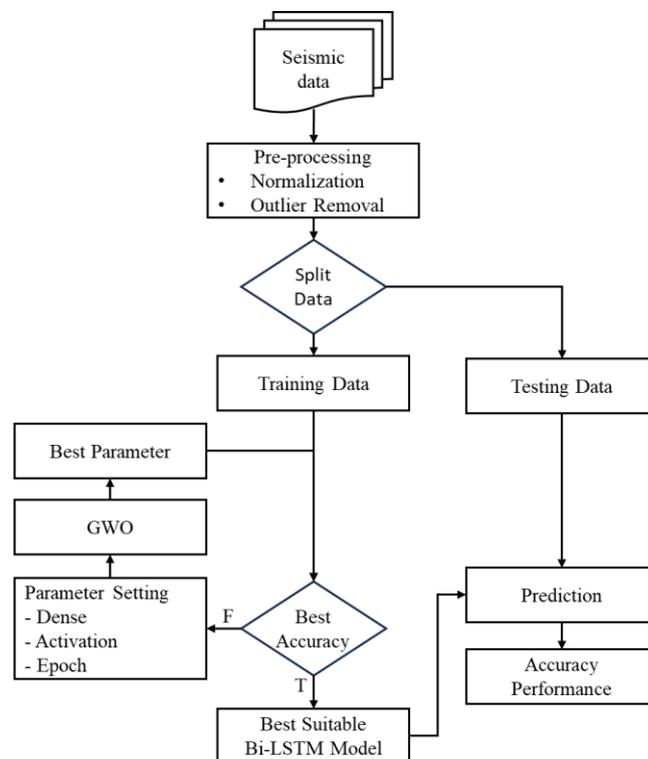


Fig. 1. Illustration of proposed method

The first phase of the proposed process involves gathering seismic data from publicly available sources. Once the dataset has been obtained, the pre-processing step is conducted, which includes normalizing the data and eliminating outliers to ensure data quality and consistency. Subsequently, the second phase begins by partitioning the processed data into a training set and a testing set, with the testing set providing a benchmark for the model used for training. The training method utilizes the Bi-LSTM architecture. The accuracy performance results are calculated to determine the optimal parameters by considering the combination of dense layers, activation functions, and the number of epochs. The best parameter combination is obtained using Grey Wolf Optimization (GWO), which dynamically adjusts these parameters to achieve the best fitness results and reduce forecasting errors. This choice of optimization technique is due to GWO proven effectiveness in handling complex optimization problems. Once the optimal model has been determined through parameter combination, the testing phase is conducted [28], [29]. The testing phase employs benchmarks for evaluating performance, with this research using MAE, MSE, and RMSE as metrics. These metrics were chosen because they are widely recognized and provide comprehensive insights into the model's accuracy and robustness in predicting earthquake occurrences.

### A. Data Gathering

The earthquake dataset utilized in this investigation was obtained from the Northern California Earthquake Data Center (NCEDC) [43]. This dataset has been used in previous research [8]. The dataset comprises records of seismic events that transpired from January 1, 1800, to January 1, 2008. 18:030 rows and 13 columns in number, the dataset was compiled using a magnitude range of three to ten. The dataset's column comprises the following: date and time, the location of the epicenter, comprehensive details regarding the magnitude and type of magnitude, the number of affected stations, and the distance between the epicenter and the adjacent station.

### B. Data Pre-processing

The pre-processing phase includes the elimination of outliers and normalization. In this investigation, the standard scaler normalization method was selected due to its ability to preserve the distribution consistency of the data. According to prior research, performance enhancement was observed with the implementation of the traditional scaler technique [44]. Equation (1) can be utilized to calculate the standard scaler, which normalizes the  $z_i$  for each  $x_i$  observed from a singular variable with a mean value ( $\mu$ ) and standard deviation ( $\sigma$ ). Following the acquisition of the normalized data, the process of outlier eradication is executed. The utilization of the interquartile range (IQR) is intended to identify outlier data. Outliers are values that fall outside the interval bounded by the 25th to 75th percentiles plus 1.5 times the interquartile range [45]. Additionally, the dataset is partitioned into training, validation, and testing sets with a defined ratio of 70%, 15%, and 15% respectively. This stratified partitioning ensures effective training, validation, and testing of the model.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

### C. Grey Wolf Optimization (GWO)

GWO was employed to identify the most beneficial attributes [28], [29], [46], [47]. The GWO methodology revolves around wolves emulating leadership roles and exhibiting intelligent hunting behaviors observed in nature, such as exploration, encirclement, and ambush tactics. Wolves within the GWO framework are categorized into distinct groups, each serving a specific role. The primary group, the alpha, assumes the highest authority and acts as the principal decision-maker. The beta group is an alpha advisor, while the delta group also contributes to the process. Optimization relies on the collaboration of alpha, beta, and delta wolves [30]. A fourth entity, Omega, is also designated to pursue other canines. In this context, GWO forms a pack during the initial population phase and actively adapts wolf positions to attain the most favorable outcomes. Subsequent measures were implemented to foster the development of new competencies [29], [47]:

- Specify the population's start point (*initpop*) value, establish the maximum iterations in a single process (*maxiter*), and create random integers  $X_i$  and  $Y_i$  to represent the wolf's initial location. The GWO technique prioritizes three fundamental values, alpha, beta, and

delta, linked to the solutions  $X_a$ ,  $X_b$ , and  $X_d$ , respectively. The wolves that are still present, mainly the omega, represent a potential resolution.

- Assign the coefficient vectors as  $\vec{C}$ ,  $\vec{a}$ , and  $\vec{A}$ .
- The precise position of each wolf ( $X$ ) is used as a benchmark for selecting the characteristics within their corresponding range of positions.
- Determine the value of vector  $a$ , which exhibits a linear decline as described by Eq. (2). The *maxiter* is the maximum limit of iterations.

$$\vec{a} = 2 - 1 \times \left( \frac{2}{\text{maxiter}} \right) \quad (2)$$

- Determine the value of  $\vec{A}$  and  $\vec{C}$  by using Eq. (3) and Eq. (4) where  $\vec{r}_1$  and  $\vec{r}_2$  represent random vectors within the range of [0,1].

$$\vec{A} = 2\vec{a} \times \vec{r}_1 - \vec{a} \quad (3)$$

$$\vec{C} = 2\vec{r}_2 \quad (4)$$

- Determine the wolf's motion using Eq. (5) and Eq. (6) then adjust its position correctly.

$$\vec{D} = |\vec{C} \times \vec{X}_p(t) - \vec{X}(t)| \quad (5)$$

$$(\vec{X}t + \vec{X}) = \vec{X}_p(t) - \vec{A} \vec{D} \quad (6)$$

where  $t$  denotes the current iteration,  $\vec{A}$  and  $\vec{C}$  represent coefficient vectors,  $\vec{X}_p$  stands for the prey position vector, and  $\vec{X}$  denotes the wolf position vector. The movement and new positions of the alpha, beta, and delta wolves can be determined using Eq. (7) to (13), given Eq. (5) and (6).

$$\vec{D}_\alpha = (\vec{C}_1 \times \vec{X}_\alpha) - \vec{X} \quad (7)$$

$$\vec{D}_\beta = (\vec{C}_2 \times \vec{X}_\beta) - \vec{X} \quad (8)$$

$$\vec{D}_\delta = (\vec{C}_3 \times \vec{X}_\delta) - \vec{X} \quad (9)$$

$$\vec{X}_1 = \vec{X}_\alpha - (\vec{A}_1 \times \vec{D}_\alpha) \quad (10)$$

$$\vec{X}_2 = \vec{X}_\beta - (\vec{A}_2 \times \vec{D}_\beta) \quad (11)$$

$$\vec{X}_3 = \vec{X}_\delta - (\vec{A}_3 \times \vec{D}_\delta) \quad (12)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (13)$$

- Validate the new solutions result for the coefficient vectors  $\vec{C}$ ,  $\vec{a}$ , and  $\vec{A}$ , then apply penalties if required.
- Determine the most recent fitness values. Due to the latest values exceeding the prior ones, the wolves positions are correspondingly modified.
- Assess the stopping conditions in relation to the maximum iteration value as  $r$ .

### D. Bi-LSTM

The Long Short-Term Memory (LSTM) is a modified version of the Recurrent Neural Network (RNN) [48]. The model exploits the capability of RNN to capture dynamic sequences via network cycles. However, gradients in RNN

frequently vanished and exploded. To address these issues, LSTM was developed with vanishing gradients in particular. The "gated" cell comprises multiple neural layers that include the chain structure of LSTM. Remarkably, an LSTM architecture is composed of three gates such as the forget gate, the input gate, and the output gate.

The forget gate  $f_t$  often utilizes the sigmoid function to decide which data should be eliminated from memory. The decision was made concerning the  $h_{(t-1)}$  and  $x_t$  quantities. The inference value of 0 or 1 is the output of this gate; a value of 0 indicates the removal of the information, and a value of 1 indicates the preservation of all the learned information. Determining the value of  $f_t$  as stated by Siami-Namini et al. [37] shown in Eq. (14).

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{(t-1)} + b_f) \quad (14)$$

The determination of whether the new information will be appended to the LSTM memory is made by the input gate ( $i_t$ ). The two layers that comprise this gate are typically the sigmoid layer and the "tanh" layer. The sigmoid layer is responsible for identifying the values that require updating Eq. (15), whereas the tanh layer represents the candidate values that should be appended to the LSTM memory Eq. (16). The expression representing  $i_t$  is shown below:

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{(t-1)} + b_i) \quad (15)$$

$$\tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{(t-1)} + b_c) \quad (16)$$

$\tilde{C}_t$  designates the vector of new candidate values that will be inserted into LSTM memory, while input gate  $i_t$  signifies the value that requires updating. LSTM memory is updated through the fusion of both layers. The update procedure ( $C_t$ ) consists of the previous information value ( $C_{(t-1)}$ ) multiplied by the forget gate  $f_t$  result in Eq. 2 and the update process ( $C_t$ ) plus the new candidate value ( $i_t \cdot \tilde{C}_t$ ). Equation (17) represents the mathematical consequences of this process.

$$C_t = f_t \cdot C_{(t-1)} + i_t \cdot \tilde{C}_t \quad (17)$$

The output gate ( $o_t$ ) inferred the output contribution of LSTM using a sigmoid layer. In order to produce a value between -1 and 1, the non-linear tanh function is subsequently executed. This result was then multiplied by the result of the sigmoid layer. The equation for this process is as follows by Eq. (18) and Eq. (19).

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{(t-1)} + b_o) \quad (18)$$

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

$H_t$  is the inference result of the non-linear tanh function, with a range of -1 to 1, while  $o_t$  represents the value of the output gate.

A singular LSTM typically only operates in the forward direction of information value. Consequently, the information was transmitted in a unidirectional manner. Concurrently, the Bidirectional LSTM architecture comprises two LSTM layers, with one layer responsible for forward processing of information and the other for backward execution. Because preceding and succeeding information

can be utilized, this architecture is more efficient than singular LSTM and RNN [49].

The multi-input, multi-output principle is integrated into this study to generate multiple prediction outcomes within a single learning cycle. This principle represents an alteration to the machine learning architecture that enables the generation of various outputs from prediction models [50]. The present study introduced the novel workflow Bi-LSTM, incorporating a multi-input, multi-output principle into its output architectures.

The Bi-LSTM architecture with a modification-based multi-input multi-output principle is illustrated in Fig. 2. Four nodes in the input layers ( $x_i$ ) represent each variable utilized in this investigation. The input block sends every possible combination of the input variable into the sequence learning block, consisting of Bi-LSTM layers featuring Tanh Activation. Furthermore, each of the four nodes in the output block represents a predicted variable  $y_i$ .

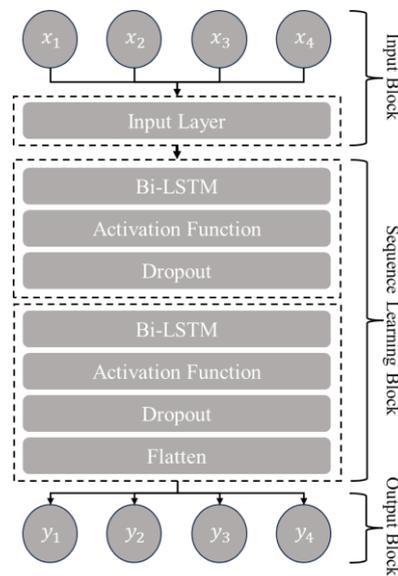


Fig. 2. Architecture of the Bi-LSTM method

### E. Activation Function

Activation functions are mathematical functions that determine the output values of each node or neuron within the layers of an artificial neural network. These functions play a central role in controlling the flow of information and gradients within the network. In this study, three activation functions are provided, namely ReLU (Rectified Linear Activation) [14], [51], Tanh (Hyperbolic Tangent Activation) [14], [52], Sigmoid (Logistic Activation) [7], [14] whose equations are shown sequentially in (20) – (22).

$$f(x) = \max(0, x) \quad (20)$$

$$f(x) = \tanh(x) \quad (21)$$

$$f(x) = 1/(1 + (1 \times e)^{-x}) \quad (22)$$

### F. Evaluation

The earthquake-predicting model will be evaluated using two loss functions, namely mean square error (MSE) and mean absolute error (MAE). Both evaluation parameters indicate the performance of the earthquake forecasting

model, explicitly describing the amount of inaccuracy in the model. A model's prediction performance improves as the error value of both loss functions decreases.

The MSE, as defined by Eq. (23), calculates the average of the squared differences between the actual value ( $y$ ) and the projected value ( $\hat{y}$ ) for all samples in a dataset with a total of  $n$  samples [25]. The RMSE, as defined by Eq. (24), is the square root of the MSE. It offers a more intuitive understanding of the error since it is expressed in units that align with the target variable [53]. Meanwhile, the MAE is calculated using Eq. (25). It measures the average absolute difference between the predicted values ( $\hat{y}$ ) and the actual values ( $y$ ), describing the mistake without taking into account its direction [37]. A decrease in the value of these three measures corresponds to an reduce the error rate of the model's earthquake predictions.

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2 \quad (23)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y - \hat{y})^2} \quad (24)$$

$$MAE = \frac{1}{n} \sum |y - \hat{y}| \quad (25)$$

### G. Experiment Design

In the experimental design, the Bi-LSTM architecture undergoes optimization using three different approaches like Gray Wolf Optimization (GWO) [29], [54], Particle Swarm Optimization (PSO) [55], [56], and PSO with adaptive inertia weighting improvement (AIWPSO) [57]. The Bi-LSTM architecture is employed to predict earthquakes based on the variables Latitude (degrees), Longitude (degrees), Depth (km), and Magnitude [8], [58].

A comparison is conducted among these methodologies to determine the most effective approach for optimizing the Bi-LSTM architecture. The experiments entail a series of trials wherein specific parameters are adjusted to identify the optimal configuration for the Bi-LSTM architecture. The optimization method carefully selects crucial parameters. The parameters for the Bi-LSTM architecture consist of the following. Dense ranges from 10 to 100, increasing by 10 at each step. Epochs are arranged in multiples of 5, starting from 5 to 50. Finally, there are 3 types of activation functions used, namely Tanh, Sigmoid, and Relu. The Bi-LSTM architecture is trained using the Adam optimizer, where the learning rate, momentum, and epoch are its parameters [59], [60]. The parameters used in this experiment are default, with a learning rate of 0.0001 and momentum of 0.9, while the best epoch is the sought-after parameter value. Beside, each optimization algorithm is tested with varying parameters, including population size and the number of iterations in the optimization method. In addition, sensitivity analysis on key parameters such as the number of epochs, size of dense layers, and activation functions to evaluate the robustness of the model.

The results obtained from these approaches are meticulously analyzed to determine the most effective way to optimize the Bi-LSTM architecture for the given dataset. To

attain optimal results, a comparative analysis of optimization methods is conducted, with training and validation accuracies meticulously recorded for thorough analysis. The experiments were performed on hardware that met specific requirements, including an Intel i7-8700 processor with 32GB RAM and a 3060TI 12GB GPU operating on the Windows 11 platform.

## IV. EXPERIMENT AND RESULTS

This section presents and analyzes the results of various optimization methods, including GWO, AIWPSO, and PSO, using tables to visualize optimal parameter configurations and performance outcomes. Evaluation metrics include loss, MAE, MSE, and RMSE. Detailed interpretation of these results follows in the discussion section to provide insights into their significance and implications for earthquake prediction accuracy.

### A. GWO

The results from the GWO experiments are shown in Table I and Table II. GWO successfully identifies optimal parameters for the Bi-LSTM architecture, achieving the lowest loss value at configuration ID 2 with parameters: 20 dense layers, 10 epochs, and tanh activation. This configuration resulted in a loss value of 0.364. MAE, MSE, and RMSE results varied with different configurations, with the best MAE observed at ID 9 and the best MSE and RMSE at ID 10. The ANOVA analysis (Table II) reveals significant differences between groups of evaluation metrics (p-value = 2.46E-05) and between types of evaluation metrics (p-value = 4.59E-43), indicating that parameter changes significantly impact evaluation results.

TABLE I. THE BEST PARAMETER OF GWO

Configur- ation ID	Number of GWO Evaluations	Number of Dense Layers	Number of Bi-LSTM Epoch	Activation Function
1	1	40	25	tanh
2	2	20	10	tanh
3	3	40	20	tanh
4	4	10	35	sigmoid
5	5	30	30	tanh
6	6	20	25	tanh
7	7	60	20	tanh
8	8	20	15	relu
9	9	50	35	tanh
10	10	40	30	tanh

TABLE II. THE BEST VALUE OF GWO EVALUATION

Configuration ID	loss	mae	mse
1	0.365	0.708	0.888
2	0.364	0.712	0.889
3	0.365	0.712	0.892
4	0.375	0.733	0.900
5	0.374	0.735	0.887
6	0.368	0.710	0.899
7	0.378	0.738	0.910
8	0.387	0.749	0.922
9	0.366	0.702	0.907
10	0.365	0.713	0.881

### B. AIWPSO

Table III and Table IV display the results from AIWPSO experiments, showing that AIWPSO also identifies optimal

parameters for the Bi-LSTM architecture. The best configuration is at configuration ID 4, with 10 dense layers, 25 epochs, and sigmoid activation, achieving a loss value of 0.377. Unlike GWO, the best results for all four evaluation metrics were consistent at this configuration. ANOVA analysis (Table IV) shows significant differences in response to parameter changes ( $p$ -value =  $1.41E-09$ ) and between types of evaluation metrics ( $p$ -value =  $1.91E-32$ ), confirming the impact of parameter changes on evaluation results.

TABLE III. THE BEST PARAMETER OF AIWPSO

Configuration ID	Number of AIWPSO Evaluations	Number of Dense Layers	Number of Bi-LSTM Epoch	Activation Function
1	1	10	15	sigmoid
2	2	10	35	relu
3	3	80	15	relu
4	4	10	25	sigmoid
5	5	30	15	relu
6	6	50	15	sigmoid
7	7	50	30	tanh
8	8	70	40	tanh
9	9	30	15	relu
10	10	50	5	tanh

TABLE IV. THE BEST VALUE OF AIWPSO EVALUATION

Configuration ID	loss	mae	mse
1	0.409	0.773	0.985
2	0.395	0.757	0.958
3	0.470	0.851	1.143
4	0.377	0.722	0.921
5	0.382	0.733	0.925
6	0.396	0.755	0.953
7	0.394	0.757	0.936
8	0.392	0.754	0.943
9	0.405	0.774	0.958
10	0.423	0.799	1.015

### C. PSO

Table V and Table VI present the PSO experiment results. The optimal configuration was identified at configuration ID 2, with 20 dense layers, 15 epochs, and tanh activation, resulting in a loss value of 0.360. The best loss and MAE values were found at ID 9. ANOVA analysis (Table VI) indicates significant differences between optimization methods ( $p$ -value =  $3.55E-10$ ) and between types of evaluation metrics ( $p$ -value =  $6.50E-35$ ), suggesting the significant influence of parameter changes on evaluation metrics.

TABLE V. THE BEST PARAMETER OF PSO

Configuration ID	Number of PSO Evaluations	Number of Dense Layers	Number of Bi-LSTM Epoch	Activation Function
1	1	40	15	sigmoid
2	2	20	15	tanh
3	3	10	20	relu
4	4	30	10	relu
5	5	40	30	tanh
6	6	50	30	tanh
7	7	40	15	relu
8	8	10	35	tanh
9	9	20	15	tanh
10	10	40	20	tanh

TABLE VI. THE BEST VALUE OF PSO EVALUATION

Configuration ID	loss	mae	mse
1	0.393	0.753	0.953
2	0.360	0.701	0.890
3	0.395	0.747	0.964
4	0.415	0.784	0.990
5	0.372	0.728	0.894
6	0.392	0.745	0.948
7	0.437	0.815	1.047
8	0.368	0.726	0.890
9	0.369	0.727	0.883
10	0.402	0.769	0.945

### D. Comparison of GWO, PSO, and AIWPSO based on Performance Matrix

Comparison results of evaluation matrices using loss in Fig. 3 show that GWO often provides the best loss value results. Meanwhile, the results of ANOVA analysis between GWO, PSO, and AIWPSO based on loss performance metrics show slight variations in the average values between the configurations of each algorithm. Anova's results revealed significant differences between the optimization algorithms ( $p$ -value= $8.97E-03$ ). This shows that the choice of optimization algorithm significantly influences the performance metrics. However, no significant differences were observed between each parameter configuration ( $p$ -value= $0.540$ ), indicating that variations in each configuration of each method did not significantly influence the results.

Fig. 4 compares the MAE results among the three optimization methods: GWO, AIWPSO, and PSO, revealing that GWO achieved the best average MAE value of 0.721. ANOVA analysis indicates significant differences between the optimization algorithms ( $p$ -value= $0.0114$ ), emphasizing the considerable impact of the choice of optimization algorithm on MAE results. However, there is no significant difference observed between the parameter configurations of each optimization algorithm ( $p$ -value= $0.617$ ), suggesting that variations in these configurations do not significantly influence MAE outcomes. Similarly, Fig. 5 compares the MSE results of GWO, AIWPSO, and PSO, with GWO demonstrating the best performance with an average MSE of 0.898. While ANOVA analysis shows significant differences between optimization algorithms ( $p$ -value= $0.0097$ ), there is no significant difference between parameter configurations ( $p$ -value= $0.45$ ), indicating that fine-tuning specific parameters within each method may not lead to substantial improvements in performance. Overall, these ANOVA results underscore the importance of selecting the right optimization approach, with GWO emerging as the most suitable option among the three methods considered in this study for achieving optimal performance outcomes.

Variations in the number of dense layers, epochs, and types of activation functions significantly influence evaluation metrics such as loss, MAE, and MSE in the Bi-LSTM model. Adjusting the number of dense layers can impact the trade-off between bias and variance, thus affecting existing error levels. Moreover, selecting an appropriate number of epochs enables the model to converge on patterns present in the data without overfitting or underfitting, as reflected in these evaluation metrics. The choice of activation function type is also critical as it can influence the

convergence speed during training and the model's ability to capture non-linear relationships in the data, thereby affecting resulting error levels. Understanding how these parameters influence evaluation metrics allows researchers to make more informed decisions when fine-tuning the Bi-LSTM model to reduce prediction errors.

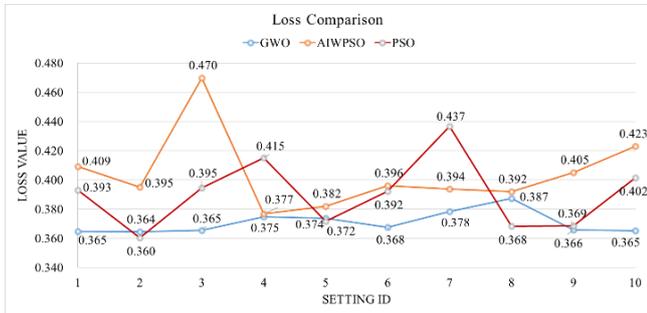


Fig. 3. Loss Comparison

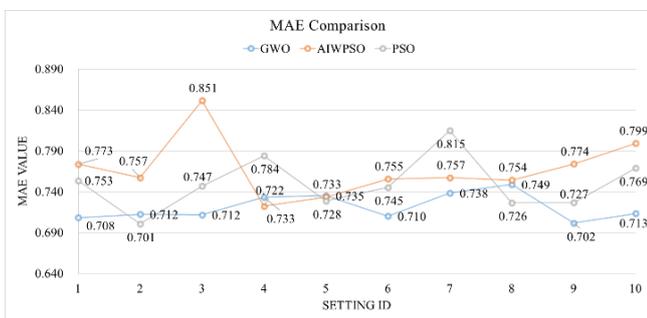


Fig. 4. MAE Comparison

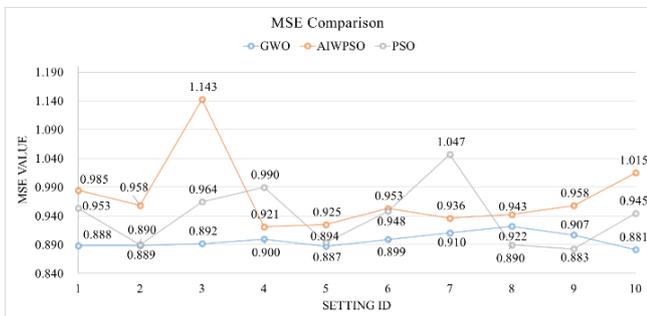


Fig. 5. MSE Comparison

**E. Comparison of Predictions and Targets**

Fig. 6 to Fig. 9 illustrate the comparison between predictions and target data for latitude, longitude, depth, and magnitude. GWO successfully achieved the best parameter configurations, resulting in predictions that closely matched the actual data, demonstrating the model's predictive accuracy. The best performance loss value obtained was 0.364. These results indicate that the predictions still fall within the data range, affirming the model's ability to generate accurate predictions based on the data.

**F. Comparison Result with Other Researcher Based on Performance Metric**

In comparison with previous researchers, this study shows a significant improvement in performance (Fig. 10). Sadhukhan et al. [25] used the Bi-LSTM Model and obtained MAE (Mean Absolute Error) of 1.499. Meanwhile, Shidik et al. [8]. also utilizing the Bi-LSTM Model, achieved slightly

better results with MAE of 1.441. In this study, the proposed method demonstrates superior performance, with MAE values of 0.702. This indicates that the proposed approach yields more accurate predictions in the same context. Such improvement suggests that the proposed method has better capabilities in modeling and predicting complex data. The significant difference in the outcomes could indeed be attributed to the utilization of varying parameters, thus potentially enabling the proposed approach to outperform.

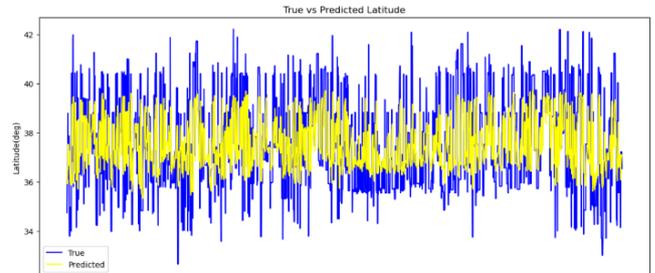


Fig. 6. Latitude Predicted

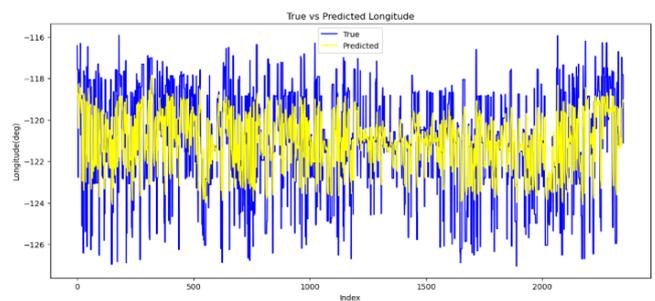


Fig. 7. Longitude Predicted

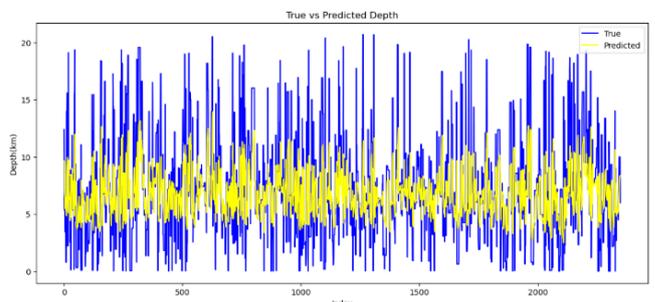


Fig. 8. Depth Predicted

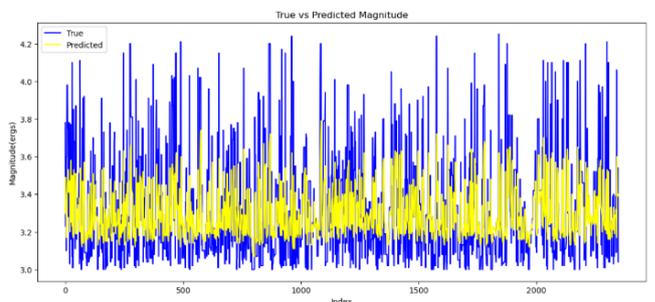


Fig. 9. Magnitude Predicted

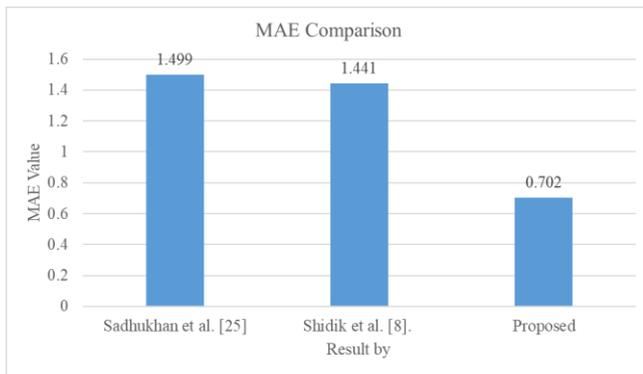


Fig. 10. Research Comparison

### G. Discussion

The comparative analysis reveals that the Grey Wolf Optimization (GWO) algorithm consistently achieves superior results across all evaluated parameters, including loss, Mean Absolute Error (MAE), and Mean Squared Error (MSE). ANOVA analysis indicates substantial disparities among the optimization techniques, demonstrating that GWO generally outperforms Particle Swarm Optimization (PSO) and Adaptive Inertia Weight PSO (AIWPSO) in all assessment measures. The choice of optimization procedure significantly impacts the model's performance, as evidenced by the low p-values observed in the ANOVA study. However, there are no noteworthy disparities in parameter choices for each algorithm, suggesting that changes in algorithm parameters do not substantially affect the comparative results between optimization approaches. Nonetheless, the assessment findings may be influenced by conducting tests with various evaluation criteria and parameter configurations.

In this experiment, GWO outperforms PSO and AIWPSO due to its higher convergence rate, more efficient trade-off between exploration and exploitation, smaller population size, and greater flexibility to environmental changes. This is evident from the fluctuations in parameters, as GWO aims to achieve an equilibrium between exploration by altering parameter values and exploitation by assessing potential solutions. The application of GWO to obtain the best parameters in the Bi-LSTM method was successful, yielding results that closely align with actual data. It can be concluded that using a Dense parameter of 20, an epoch of 10, and tanh activation provides accurate prediction results from the Bi-LSTM method based on Latitude, Longitude, Depth, and Magnitude.

### V. CONCLUSION

The Grey Wolf Optimization (GWO) algorithm effectively identified optimal parameters for the Bi-LSTM architecture in seismic data prediction by utilizing datasets from the Northern California Earthquake Data Center (NCEDC). GWO consistently outperformed AIWPSO and PSO across various evaluation metrics, including loss, Mean Absolute Error (MAE), and Mean Squared Error (MSE). This superiority is attributed to GWO's efficient balance between exploration and exploitation, smaller population size, and adaptability to environmental changes. Consequently, GWO enabled the Bi-LSTM method to achieve more accurate

predictions based on Latitude, Longitude, Depth, and Magnitude. This research advances the understanding of parameter optimization in deep learning frameworks for seismic prediction and demonstrates the practical utility of GWO in achieving accurate predictive modeling outcomes. However, the study is limited by the specific dataset and parameter configurations used, suggesting that future research should explore different datasets and more extensive parameter tuning to validate and extend these findings. Future work could also investigate the integration of GWO with other advanced optimization techniques and apply it to different predictive modeling domains to enhance performance. Overall, this work contributes to the field by highlighting the significance of parameter optimization in predictive modeling and demonstrating the effectiveness of GWO in achieving accurate predictions.

### ACKNOWLEDGMENT

This research received support and funding from Universitas Dian Nuswantoro. We extend our gratitude to our colleagues at the Research Center for Intelligent Distributed Surveillance and Security, particularly those specializing in Artificial Intelligence Studies in Nature Conservation and Natural Disasters, for their valuable contributions.

### REFERENCES

- [1] O. Nicolis, F. Plaza, and R. Salas, "Prediction of intensity and location of seismic events using deep learning," *Spatial Statistics*, vol. 42, 2021, doi: 10.1016/j.spasta.2020.100442.
- [2] L. B. Elvas, B. M. Mataloto, A. L. Martins, and J. C. Ferreira, "Disaster management in smart cities," *Smart Cities*, vol. 4, no. 2, pp. 819–839, 2021, doi: 10.3390/smartcities4020042.
- [3] D. N. Ford and C. M. Wolf, "Smart Cities with Digital Twin Systems for Disaster Management," *Journal of Management in Engineering*, vol. 36, no. 4, pp. 1–10, 2020, doi: 10.1061/(asce)me.1943-5479.0000779.
- [4] A. Berhich, F. Z. Belouadha, and M. I. Kabbaj, "An attention-based LSTM network for large earthquake prediction," *Soil Dynamics and Earthquake Engineering*, vol. 165, 2022, p. 107663, 2023, doi: 10.1016/j.soildyn.2022.107663.
- [5] S. M. Mousavi and G. C. Beroza, "Machine Learning in Earthquake Seismology," *Annual Review of Earth and Planetary Sciences*, vol. 51, pp. 105–129, 2023, doi: 10.1146/annurev-earth-071822-100323.
- [6] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification," *Soil Dynamics and Earthquake Engineering*, vol. 111, pp. 1–7, 2018, doi: 10.1016/j.soildyn.2018.04.020.
- [7] M. Moustra, M. Avraamides, and C. Christodoulou, "Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15032–15039, 2011, doi: 10.1016/j.eswa.2011.05.043.
- [8] G. F. Shidik *et al.*, "LUTanh Activation Function to Optimize Bi-LSTM in Earthquake Forecasting," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 1, pp. 572–583, 2024, doi: 10.22266/ijies2024.0229.48.
- [9] B. Dey, P. Dikshit, S. Sehgal, V. Trehan, and V. Kumar Sehgal, "Intelligent solutions for earthquake data analysis and prediction for future smart cities," *Computers and Industrial Engineering*, vol. 170, 2022, doi: 10.1016/j.cie.2022.108368.
- [10] Z. Qadir, F. Ullah, H. S. Munawar, and F. Al-Turjman, "Addressing disasters in smart cities through UAVs path planning and 5G communications: A systematic review," *Computer Communications*, vol. 168, pp. 114–135, 2021, doi: 10.1016/j.comcom.2021.01.003.
- [11] K. Sharma, D. Anand, M. Sabharwal, P. K. Tiwari, O. Cheikhrouhou, and T. Frikha, "A Disaster Management Framework Using Internet of

- Things-Based Interconnected Devices,” *Mathematical Problems in Engineering*, vol. 2021, 2021, doi: 10.1155/2021/9916440.
- [12] G. C. Beroza, M. Segou, and S. Mostafa Mousavi, “Machine learning and earthquake forecasting—next steps,” *Nature Communications*, vol. 12, no. 1, pp. 10–12, 2021, doi: 10.1038/s41467-021-24952-6.
- [13] V. Chernykh, A. Stepnov, and B. O. Lukyanova, “Data preprocessing for machine learning in seismology,” *CEUR Workshop Proceedings*, vol. 2930, pp. 119–123, 2021.
- [14] C. Ning, Y. Xie, and L. Sun, “LSTM, WaveNet, and 2D CNN for nonlinear time history prediction of seismic responses,” *Engineering Structures*, vol. 286, 2023, doi: 10.1016/j.engstruct.2023.116083.
- [15] J. W. Lin, “Researching significant earthquakes in Taiwan using two back-propagation neural network models,” *Natural Hazards*, vol. 103, no. 3, pp. 3563–3590, 2020, doi: 10.1007/s11069-020-04144-z.
- [16] Y. Essam, P. Kumar, A. N. Ahmed, M. A. Murti, and A. El-Shafie, “Exploring the reliability of different artificial intelligence techniques in predicting earthquake for Malaysia,” *Soil Dynamics and Earthquake Engineering*, vol. 147, 2021, doi: 10.1016/j.soildyn.2021.106826.
- [17] J. Song, J. Zhu, Y. Wang, and S. Li, “On-site alert-level earthquake early warning using machine-learning-based prediction equations,” *Geophysical Journal International*, vol. 231, no. 2, pp. 786–800, Jul. 2022, doi: 10.1093/gji/ggac220.
- [18] L. Izquierdo-Horna, J. Zevallos, and Y. Yopez, “An integrated approach to seismic risk assessment using random forest and hierarchical analysis: Pisco, Peru,” *Heliyon*, vol. 8, no. 10, p. e10926, Oct. 2022, doi: 10.1016/j.heliyon.2022.e10926.
- [19] Y. Wang, Q. Zhao, K. Qian, Z. Wang, Z. Cao, and J. Wang, “Cumulative absolute velocity prediction for earthquake early warning with deep learning,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 39, no. 11, pp. 1724–1740, 2023, doi: 10.1111/micc.13065.
- [20] I. M. Murwantara, P. Yugopuspito, and R. Hermawan, “Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1331–1342, 2020, doi: 10.12928/TELKOMNIKA.v18i3.14756.
- [21] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, “Earthquake Magnitude Prediction using Spatio-temporal Features Learning Based on Hybrid CNN- BiLSTM Model,” in *2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1–6, Dec. 2021, doi: 10.1109/ICSPIS54653.2021.9729358.
- [22] G. Gürsoy, A. Varol, and A. Nasab, “Importance of Machine Learning and Deep Learning Algorithms in Earthquake Prediction: A Review,” in *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 1–6, May 2023, doi: 10.1109/ISDFS58141.2023.10131766.
- [23] P. Kavianpour, M. Kavianpour, and A. Ramezani, “Deep Multi-scale Dilated Convolution Neural Network with Attention Mechanism: A Novel Method for Earthquake Magnitude Classification,” in *2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1–6, Dec. 2022, doi: 10.1109/ICSPIS56952.2022.10043978.
- [24] Z. Zhang and Y. Wang, “A Spatiotemporal Model for Global Earthquake Prediction Based on Convolutional LSTM,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023, doi: 10.1109/TGRS.2023.3302316.
- [25] B. Sadhukhan, S. Chakraborty, S. Mukherjee, and R. K. Samanta, “Climatic and seismic data-driven deep learning model for earthquake magnitude prediction,” *Frontiers in Earth Science*, vol. 11, pp. 1–24, Feb. 2023, doi: 10.3389/feart.2023.1082832.
- [26] E. Abebe, H. Kebede, M. Kevin, and Z. Demissie, “Earthquakes magnitude prediction using deep learning for the Horn of Africa,” *Soil Dynamics and Earthquake Engineering*, vol. 170, p. 107913, 2023, doi: 10.1016/j.soildyn.2023.107913.
- [27] N. B. Yoma *et al.*, “End-to-end LSTM based estimation of volcano event epicenter localization,” *Journal of Volcanology and Geothermal Research*, vol. 429, 2022, doi: 10.1016/j.jvolgeores.2022.107615.
- [28] R. A. Pramonendar *et al.*, “Integrating Grey Wolf Optimizer for Feature Selection in Birdsong Classification Using K-Nearest Neighbours Algorithm,” *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 6, pp. 695–705, Dec. 2023, doi: 10.22266/ijies2023.1231.58.
- [29] P. I. Santosa and R. A. Pramonendar, “A Robust Feature Construction for Fish Classification Using Grey Wolf Optimizer,” *Cybernetics and Information Technologies*, vol. 22, no. 4, pp. 152–166, Nov. 2022, doi: 10.2478/cait-2022-0045.
- [30] A. Alzaqebah, I. Aljarah, O. Al-Kadi, and R. Damaševičius, “A Modified Grey Wolf Optimization Algorithm for an Intrusion Detection System,” *Mathematics*, vol. 10, no. 6, pp. 1–16, 2022, doi: 10.3390/math10060999.
- [31] A. C. Boucouvalas, M. Gkasios, N. T. Tselikas, and G. Drakatos, “Modified-Fibonacci-Dual-Lucas method for earthquake prediction,” in *Third international conference on remote sensing and geoinformation of the environment (RSCy2015)*, vol. 9535, pp. 400–410, 2015.
- [32] M. Marisa, U. A. Sembiring, and H. Margaretha, “Earthquake probability prediction in sumatra island using Poisson hidden Markov model (HMM),” in *AIP Conference Proceedings*, vol. 2192, no. 1, 2019.
- [33] H. Dehghani and M. J. Fadaee, “Probabilistic prediction of earthquake by bivariate distribution,” *Asian Journal of Civil Engineering*, vol. 21, no. 6, pp. 977–983, 2020, doi: 10.1007/s42107-020-00254-y.
- [34] A. Kundu, S. Ghosh, and S. Chakraborty, “A long short-term memory based deep learning algorithm for seismic response uncertainty quantification,” *Probabilistic Engineering Mechanics*, vol. 67, p. 103189, 2022, doi: 10.1016/j.probenmech.2021.103189.
- [35] R. Abri and H. Artuner, “LSTM-Based Deep Learning Methods for Prediction of Earthquakes Using Ionospheric Data,” *Gazi University Journal of Science*, vol. 35, no. 4, pp. 1417–1431, 2022, doi: 10.35378/gujs.950387.
- [36] T. Bhandarkar, V. K. N. Satish, S. Sridhar, R. Sivakumar, and S. Ghosh, “Earthquake trend prediction using long short-term memory RNN,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 2, p. 1304, 2019, doi: 10.11591/ijece.v9i2.pp1304-1312.
- [37] S. Siami-Namini, N. Tavakoli, and A. S. Namin, “The Performance of LSTM and BiLSTM in Forecasting Time Series,” in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 3285–3292, 2019, doi: 10.1109/BigData47090.2019.9005997.
- [38] Y. Liao, R. Lin, R. Zhang, and G. Wu, “Attention-based LSTM (AttLSTM) neural network for Seismic Response Modeling of Bridges,” *Computers and Structures*, vol. 275, 2023, doi: 10.1016/j.compstruc.2022.106915.
- [39] M. M. Hason, A. N. Hanoon, and A. A. Abdulhameed, “Particle swarm optimization technique-based prediction of peak ground acceleration of Iraq’s tectonic regions,” *Journal of King Saud University - Engineering Sciences*, vol. 35, no. 7, pp. 463–473, 2023, doi: 10.1016/j.jksues.2021.06.004.
- [40] P. K. E S, V. N. Thatha, G. Mamidiseti, S. V. Mantena, P. Chintamaneni, and R. Vatambeti, “Hybrid deep learning model with enhanced sunflower optimization for flood and earthquake detection,” *Heliyon*, vol. 9, no. 10, p. e21172, 2023, doi: 10.1016/j.heliyon.2023.e21172.
- [41] A. M. Chung Baek, E. Park, M. Seong, J. Koo, I. D. Jung, and N. Kim, “Multi-objective robust parameter optimization using the extended and weighted k-means (EWK-means) clustering in laser powder bed fusion (LPBF),” *Expert Systems with Applications*, vol. 236, 2024, doi: 10.1016/j.eswa.2023.121349.
- [42] Y. Xu, H. Cao, J. Shi, S. Pei, and K. She, “A comprehensive multi-parameter optimization method of squeeze film damper-rotor system using hunter-prey optimization algorithm,” *Tribology International*, vol. 194, p. 109538, 2024, doi: 10.1016/j.triboint.2024.109538.
- [43] NCEDC, “Northern California Earthquake Data Center,” *UC Berkeley Seismol. Lab.*, 2014.
- [44] P. Ferreira, D. C. Le, and N. Zincir-Heywood, “Exploring Feature Normalization and Temporal Information for Machine Learning Based Insider Threat Detection,” *2019 15th International Conference on Network and Service Management (CNSM)*, pp. 1–7, 2019, doi: 10.23919/CNSM46954.2019.9012708.
- [45] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, “An outliers detection and elimination framework in classification task of data

- mining,” *Decision Analytics Journal*, vol. 6, p. 100164, 2023, doi: 10.1016/j.dajour.2023.100164.
- [46] S. Gupta and K. Deep, “A novel Random Walk Grey Wolf Optimizer,” *Swarm and Evolutionary Computation*, vol. 44, pp. 101–112, 2019, doi: 10.1016/j.swevo.2018.01.001.
- [47] S. M. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey Wolf Optimizer,” *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.
- [48] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [49] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, “A CNN-BiLSTM model with attention mechanism for earthquake prediction,” *Journal of Supercomputing*, vol. 79, no. 17, pp. 19194–19226, 2023, doi: 10.1007/s11227-023-05369-y.
- [50] M. Zandie, H. K. Ng, S. Gan, M. F. Muhamad Said, and X. Cheng, “Multi-input multi-output machine learning predictive model for engine performance and stability, emissions, combustion and ignition characteristics of diesel-biodiesel-gasoline blends,” *Energy*, vol. 262, p. 125425, 2023, doi: 10.1016/j.energy.2022.125425.
- [51] R. M. González, M. A. B. González, A. M. Cruz, A. R. González, and A. L. Pérez, “Classification of land use and vegetation with convolutional neural networks,” *Revista Mexicana de Ciencias Forestales*, vol. 13, no. 74, pp. 97–119, 2022, doi: 10.29298/rmcf.v13i74.1269.
- [52] B. Xia, F. Kong, J. Zhou, X. Wu, and Q. Xie, “Land Resource Use Classification Using Deep Learning in Ecological Remote Sensing Images,” *Computational Intelligence and Neuroscience*, vol. 2022, 2022, doi: 10.1155/2022/7179477.
- [53] R. Rakholia, Q. Le, B. Quoc Ho, K. Vu, and R. Simon Carbajo, “Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam,” *Environment International*, vol. 173, p. 107848, 2023, doi: 10.1016/j.envint.2023.107848.
- [54] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. D. S. Coelho, “Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization,” *Expert Systems with Applications*, vol. 47, pp. 106–119, 2016, doi: 10.1016/j.eswa.2015.10.039.
- [55] S. W. Lin, K. C. Ying, S. C. Chen, and Z. J. Lee, “Particle swarm optimization for parameter determination and feature selection of support vector machines,” *Expert Systems with Applications*, vol. 35, no. 4, pp. 1817–1824, 2008, doi: 10.1016/j.eswa.2007.08.088.
- [56] D. N. Tuyen *et al.*, “A novel approach combining particle swarm optimization and deep learning for flash flood detection from satellite images,” *Mathematics*, vol. 9, no. 22, 2021, doi: 10.3390/math9222846.
- [57] A. Nickabadi, M. M. Ebadzadeh, and R. Safabakhsh, “A novel particle swarm optimization algorithm with adaptive inertia weight,” *Applied Soft Computing Journal*, vol. 11, no. 4, pp. 3658–3670, 2011, doi: 10.1016/j.asoc.2011.01.037.
- [58] G. Asencio-Cortés, F. Martínez-Álvarez, A. Morales-Esteban, and J. Reyes, “A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction,” *Knowledge-Based Systems*, vol. 101, pp. 15–30, 2016, doi: 10.1016/j.knosys.2016.02.014.
- [59] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.
- [60] S. R. Young, D. C. Rose, T. P. Karnowski, S. H. Lim, and R. M. Patton, “Optimizing deep learning hyper-parameters through an evolutionary algorithm,” in *Proceedings of the workshop on machine learning in high-performance computing environments*, pp. 1–5, 2015.