



ARTICLE

Error Analysis of Geomagnetic Field Reconstruction Model Using Negative Learning for Seismic Anomaly Detection

Nur Syaiful Afrizal¹, Khairul Adib Yusof^{1,2,*}, Lokman Hakim Muhamad¹,
Nurul Shazana Abdul Hamid^{2,3}, Mardina Abdullah^{2,4}, Mohd Amiruddin Abd Rahman¹,
Syamsiah Mashohor⁵ and Masashi Hayakawa^{6,7}

¹Department of Physics, Faculty of Science, Universiti Putra Malaysia, Seri Kembangan, 43400, Malaysia

²Space Science Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi, 43600, Malaysia

³Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, 43600, Malaysia

⁴Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600, Malaysia

⁵Department of Computer and Communication System Engineering, Faculty of Engineering, Universiti Putra Malaysia, Seri Kembangan, 43400, Malaysia

⁶UEC Alliance Center, Hayakawa Institute of Seismo Electromagnetics Co., Ltd. (Hi-SEM), Chofu, Tokyo, 1820026, Japan

⁷Advanced Wireless & Communications Research Center (AWCC), University of Electro-Communications, Chofu, Tokyo, 182-8585, Japan

*Corresponding Author: Khairul Adib Yusof. Email: adib.yusof@upm.edu.my

Received: 08 April 2025; Accepted: 21 August 2025; Published: 09 December 2025

ABSTRACT: Detecting geomagnetic anomalies preceding earthquakes is a challenging yet promising area of research that has gained increasing attention in recent years. This study introduces a novel reconstruction-based modeling approach enhanced by negative learning, employing a Bidirectional Long Short-Term Memory (BiLSTM) network explicitly trained to accurately reconstruct non-seismic geomagnetic signals while intentionally amplifying reconstruction errors for seismic signals. By penalizing the model for accurately reconstructing seismic anomalies, the negative learning approach effectively magnifies the differences between normal and anomalous data. This strategic differentiation enhances the sensitivity of the BiLSTM network, enabling improved detection of subtle geomagnetic anomalies that may serve as earthquake precursors. Experimental validation clearly demonstrated statistically significant higher reconstruction errors for seismic signals compared to non-seismic signals, confirmed through the Mann-Whitney U test with a p -value of 0.0035 for Root Mean Square Error (RMSE). These results provide compelling evidence of the enhanced anomaly detection capability achieved through negative learning. Unlike traditional classification-based methods, negative learning explicitly encourages sensitivity to subtle precursor signals embedded within complex geomagnetic data, establishing a robust basis for further development of reliable earthquake prediction methods.

KEYWORDS: Error analysis; geomagnetic field; BiLSTM model; negative learning; earthquake precursor

1 Introduction

Research into potential earthquake precursors through the observation of non-seismological (i.e., not related to seismic waves) Earth and space parameters began in late 1980s [1]. These parameters, collectively termed seismo-electromagnetic phenomena, encompass anomalous geoelectric currents, ultra-low frequency (ULF) and radon gas emissions in the lithosphere, over-the-horizon very-high frequency (VHF), and thermal anomalies in the atmosphere, along with very-low-frequency (VLF) transmitter signals



and abnormal changes in total electron content (TEC) in the ionosphere [2,3]. Among various methods, analyzing disturbances in ground-based magnetometer data has gained popularity, thanks to the wealth of publicly available data, such as SuperMAG, employed in this study. In particular, this study focuses on detecting low-frequency fluctuations in the geomagnetic field, typically evolving over multi-hour to multi-day timescales, which may be indicative of lithospheric or ionospheric processes during earthquake preparation [4,5]. Ground-based measurements are deemed more practical for real earthquake prediction compared to satellite measurements due to their ability to trace spatiotemporal variations more precisely [2].

Several mechanisms have been proposed to explain geomagnetic anomalies preceding significant earthquakes, including underground conductivity changes and electrokinetic effects resulting from water-gas migration through the ground medium [6]. These mechanisms are integral to the lithosphere-atmosphere-ionosphere-magnetosphere coupling (LAIMC) theory, which is regarded as the primary framework for understanding various seismo-electromagnetic phenomena [7]. Despite the potential, previous studies have consistently demonstrated that geomagnetic anomalies tend to be relatively weak [8]. To address the challenge of detecting such subtle anomalies, researchers have applied various signal processing techniques and data analysis methods, including polarization ratio [4], diurnal variation range ratio [8], natural time [5] and vertical variance methods [9] analyses. The focus of most methods has been on the vertical component of the geomagnetic field due to its heightened sensitivity to underground structural heterogeneity, particularly during earthquake preparation phases [4,10]. Moreover, the vertical component is less susceptible to disturbances from solar-terrestrial signals, which are more likely to be plane waves affecting only the horizontal component [11]. While these methods have showcased their effectiveness in enhancing earthquake-related geomagnetic anomalies, previous studies have typically been confined to case or regional investigations with limited observation periods [1]. Such limitations may impede progress toward achieving accurate earthquake forecasting, as comprehensive large-scale or global studies are crucial for robust model development.

In the contemporary landscape, artificial intelligence (AI) has emerged as the preferred approach for modeling complex processes, driven by the availability of extensive datasets [12]. Despite this trend, literature on the adoption of AI for non-seismological pre-earthquake anomalies, particularly in the geomagnetic field, remains sparse. A recent prospective study introduced a Convolutional Neural Network (CNN) model trained using data from three magnetometer stations in European tectonic zones, revealing a subpar performance with a maximum accuracy of 66.40% [13]. Other relevant studies have primarily explored seismo-ionospheric anomalous variations induced by earthquakes using TEC time series data. Notably, studies that developed Recurrent Neural Network (RNN)-based models reported higher accuracies of around 78.05% [14] and 91.05% (R^2) [15]. Among the various RNN models, the Long Short-Term Memory (LSTM)-based model stands out as a commonly employed architecture for handling time-series data [16]. The capability of LSTM models in detecting geomagnetic as well as various ionospheric and atmospheric anomalies (e.g., outgoing longwave radiation, air temperature, relative humidity, and vertical TEC) prior to a sizeable earthquake has been demonstrated in the past [17,18]. Bidirectional LSTM (BiLSTM) further improves on regular LSTM by processing input sequences in both forward and backward directions. Equipped with memory cells featuring input, output, and forget gates, a BiLSTM network can selectively retain or release information over extended sequences, proving effective in time series analysis [19].

In earlier attempts at earthquake prediction based on geomagnetic field data, we adopted a direct classification approach, where each event was labelled as seismic or non-seismic depending on its temporal proximity to an earthquake. Our preliminary findings suggested that a simple classification model struggled to detect subtle anomalies thus limiting its performance, probably due to the inherent variability and complexity of the geomagnetic signals. To address these challenges, a reconstruction-based approach was explored as an alternative. It focuses on training models that are able to recognize the pattern of normal

(non-seismic) data and reconstruct them accurately. Through reconstruction error analysis by comparing the actual and reconstructed signals, anomalies potentially associated with seismic activities can be detected. This method avoids reliance on ambiguous class labels and is thus better suited for capturing the weak, nonstationary nature of precursor signals. However, conventional reconstruction-based models typically treat all errors symmetrically, which may limit their sensitivity to subtle seismic deviations. To address this, we employed a negative learning approach, inspired by recent advances in computer vision [20,21], where the model is explicitly penalized for accurately reconstructing known anomalies. While previous work has applied this idea in the image domain for visual anomaly detection, to the best of our knowledge, our study is the first to extend negative learning to geophysical time series data for seismic anomaly detection.

Negative learning is an emerging approach in deep learning that emphasizes models training on explicitly penalizing undesirable patterns and anomalies. It was inspired by contrastive learning with the objective to minimize intra-class variance (normal and anomalous data) while maximize inter-class separation (normal vs. anomalous data). Negative learning is different compared to traditional anomaly detection method which involves unsupervised or semi-supervised frameworks to model normal and flag deviations. Instead, this method include both normal and anomalous data during training, then, explicitly teaching the model to recognize and penalize anomalous signals. This approach has been applied in time series domain for example in fraud detection [22]. These applications usually involve in training the models to reconstruct normal data accurately while deliberately failing in reconstructing anomalous ones. The major component that enable this is a loss function that penalizes errors asymmetrically based on the classes, forcing the distinction innately. Thus, in the context of rare events such as earthquakes, where precursor are often embedded within complex temporal dependencies, this approach is particularly powerful for identifying subtle variations.

In summary, classification-based approaches offer the advantage of end-to-end detection pipelines and decision boundaries but depend heavily on accurately labeled data, and may generalize poorly due to the temporal uncertainty and weak amplitude of seismic precursors. They often require event-specific tuning and perform inconsistently across geographies. Reconstruction-based models, including the one proposed here, offer an alternative by focusing on the normal signal distribution and identifying deviations without relying on explicit event labels. Our proposed method, enhanced by negative learning, further amplifies class-level separability by penalizing accurate reconstruction of anomalous sequences. However, it does not provide real-time classification or assign probabilistic earthquake risk levels, and is instead best suited for post-hoc anomaly analysis or early-warning signal exploration. This trade-off favors interpretability and generalization but limits deployment as a direct forecasting system.

The work done in the present study is illustrated in Fig. 1, where the steps are labeled from (a) to (f). First, the vertical Z component of the geomagnetic field was acquired from their respective databases in (a), labeled similarly to our preliminary study. Both input and original signal uses the same data, in which it will undergo detrending to isolate short term fluctuations that may indicate anomalies, followed by down-sampling in (b). To introduce perturbation in the input, Gaussian noise were added as shown in (c). Next, training on both seismic and non-seismic dataset to obtain a sequence-to-sequence reconstruction model (d). The model was then tested on unseen data from both seismic and non-seismic datasets that have been segmented to enable temporal analysis as shown in (e). The reconstruction error between the classes was analyzed and compared to assess whether significant differences existed.

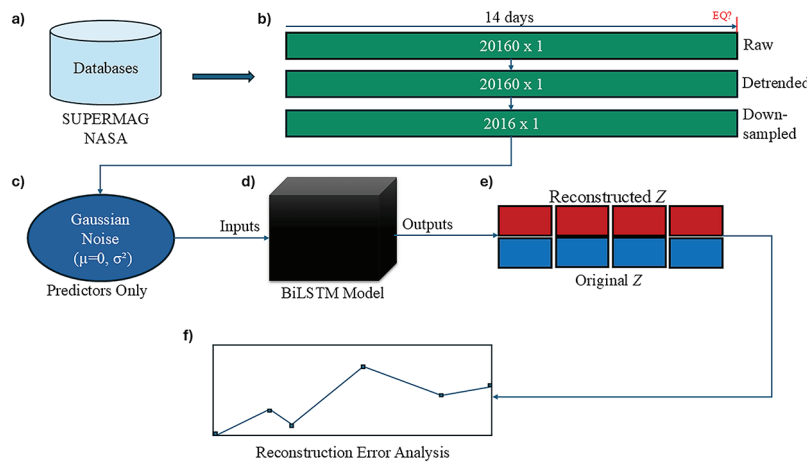


Figure 1: The present study workflow

2 Methodology

2.1 Data Description

Geomagnetic field data with a sampling period of 1 min, covering the period from January 1970 to December 2021, were sourced from the SuperMAG online database (www.supermag.jhuapl.edu, accessed on 12 July 2025). The vast SuperMAG project encompasses 138 magnetometer stations located around the world, allowing for large-scale, worldwide studies to be carried out [23]. The stations are installed and maintained by different organizations; thus, their data might have different configurations. These differences are mitigated by the SuperMAG network through data transformation into a common coordinate system, namely geomagnetic northward N , eastward E , and vertical Z , where the latter was the primary focus of this study. In addition, the data underwent a baseline removal process that includes the daily baseline, annual trend, and residual offset [24]. By removing noise and concentrating on the significant fluctuations in the magnetic field, the reliability and uniformity of the data from different locations were ensured. The study also incorporates earthquake catalog data from the United States Geological Survey (USGS) online database (www.earthquake.usgs.gov, accessed on 12 July 2025). The catalog includes seismic event records providing key information about event locations, depths, magnitudes, and occurrence times.

2.2 Dataset Formation

The datasets used in this study consist of geomagnetic field records combined with earthquake catalog data. We first listed all global earthquake events from 1970 to 2021 with magnitudes of at least $M5.5$ and depths shallower than 50 km. For each earthquake, magnetometer stations located within 200 km of the epicenter were identified. If the stations were operational, geomagnetic field data were extracted for the 14 days leading up to the earthquake. In total, 1478 seismic-class samples were obtained. The 200 km threshold was estimated from the empirical relation $0.025 d \geq M - 4.5$ [3], which implies that $d \leq 200$ km represents the maximum detectable range for the largest recorded earthquake ($M9.5$). The depth criterion ≤ 50 km was chosen because geomagnetic ULF anomalies are most prominent in upper and mid-crustal depths [25,26].

The non-seismic dataset was constructed by selecting periods at the same stations that were not within 200 km of any earthquake. Because geomagnetic behavior is seasonally influenced by solar activity and Earth's axial tilt [27], non-seismic samples were taken from one month or one year before or after the corresponding seismic intervals. Duplicate extractions were avoided to prevent overfitting. This procedure yielded 1478 non-seismic samples. Each seismic sample was paired with a non-seismic sequence from the

same station to reduce temporal bias. For the evaluation set, we ensured station-level balance by including an equal number of seismic and non-seismic samples. A summary of the regional distribution is given in Table 1, while the geographic coverage of magnetometer stations and earthquake epicenters is shown in Fig. 2. For the development set, strict one-to-one matching at the station level was not always possible due to sample limitations. Instead, we maintained global class balance (1024 samples per class for training and 181 per class for validation) while maximizing spatial and temporal coverage. The remaining 273 non-seismic samples, together with their seismic counterparts, were reserved for model evaluation.

Table 1: Regional distribution of seismic and non-seismic samples in the development and evaluation sets

Set	Region	Seismic	Non-seismic
Development	Australia & New Zealand	53	83
Development	Central Asia	133	231
Development	East Asia Mainland	139	202
Development	Equatorial band	42	35
Development	Other	103	180
Development	Pacific Rim	688	458
Development	Southeast Asia	47	16
	Total	1205	1205
Evaluation	Australia & New Zealand	8	8
Evaluation	Central Asia	17	17
Evaluation	East Asia Mainland	53	53
Evaluation	Equatorial band	2	2
Evaluation	Other	27	27
Evaluation	Pacific rim	156	156
Evaluation	Southeast Asia	10	10
	Total	273	273

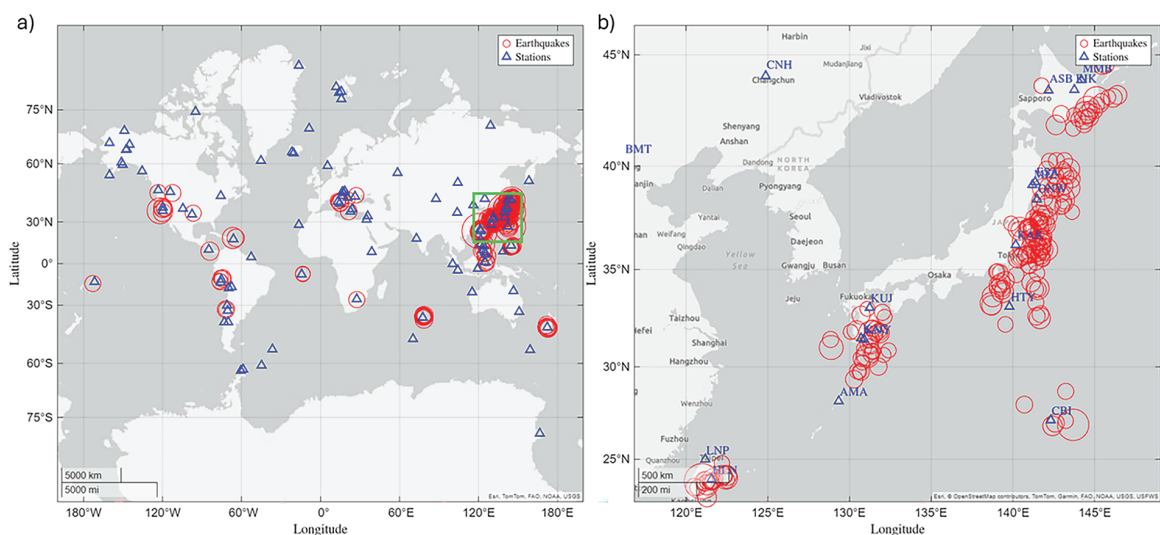


Figure 2: Locations of magnetometer stations (blue triangles) and earthquakes (red circles) included in this study, where (a) shows the global map, while (b) is a zoomed map of the region that is highlighted in green in (a).

It is worth emphasizing that searching the non-seismic data within the same season and balancing the classes for each station were done to ensure that the temporal and spatial characteristics between both classes were as similar as possible. In other words, the spatiotemporal aspect of the data was the control variable, while the presence of earthquakes in the vicinity of the data was the variable being tested in this study. To ensure spatial and temporal representativeness, the dataset was analyzed by geographic region and station coverage. A total of 38 unique magnetometer stations contributed data spanning 7 global regions, including East Asia, Central Asia, South America, the Mediterranean, and Oceania. The most active stations in terms of sample count include KAK (Japan), KNY (Japan), and HTY (Italy), all located in seismically active tectonic regimes. The dataset was constructed by extracting Z-components of the geomagnetic field for a total duration of 14 days. Both the input and the original signals were based on the same data. The Z geomagnetic field were then z-score normalized, where the original signal is subtracted by its mean and then divided by its standard deviation. The normalization was done on the combined development and evaluation datasets, combined across different stations and observation periods to ensure the uniformity of features, i.e., zero mean and unit variance [28].

The Z-component sequence was then detrended to remove slow-varying background trends, such as daily baselines and seasonal fluctuations. This step isolates shorter-term variations relevant to potential geomagnetic anomalies. After detrending, the data was downsampled by a factor of 10, reducing the sequence length from 20,160 (14 days at 1-minute resolution) to 2016. It is worth noting that no ULF components can be extracted from the original signal of 1 min due to the Nyquist frequency being 8.3 mHz. This downsampling improves training efficiency while preserving essential temporal structures, including diurnal and multi-day variations rather than short-term fluctuations [2], and our target signals remain intact within the retained frequency band [4]. The down-sampling reduces the signal length from 20,160 points (1 min $\times$ 24 h $\times$ 14 days) to 2016 points. To prevent the model from trivially learning an identity function, since the input and target are identical, a small amount of zero-mean Gaussian noise ($\sigma^2 = 0.01$) was added to the input during training. This technique introduces controlled variability and promotes generalization by encouraging the model to learn structural features rather than exact replication. The noise level was empirically selected to balance signal preservation with effective regularization, in line with common practices in denoising autoencoders and sequence modeling [22]. To prevent this and to ensure that the model maximizes anomalous errors, an essential aspect for reconstructing the signal accurately, which was negative learning, was implemented at the final stage of the process.

3 Model Development

3.1 Network Architecture

The architecture employed in this study leverages a BiLSTM layer, which is a commonly used layer in an RNN. This network configuration consists of several interconnected layers, structured sequentially for optimal information processing as illustrated in Fig. 3. The first layer is an input sequence layer that accepts a 2016 (signal length) $\times$ 1 (features) array as the input data representation. Subsequently, a BiLSTM layer, comprising 900 hidden units, is utilized to capture temporal dependencies and context within the input data. Following the BiLSTM layer, a fully connected (FC) layer with 600 output nodes is integrated into the architecture to facilitate feature extraction and transformation. To prevent overfitting and enhance generalization, a dropout layer with a dropout rate of 0.2 is introduced. The number of hidden units (in the BiLSTM layer) and number of output nodes (in the first FC layer) were selected through Bayesian optimization. The search explored combinations of hidden units and FC layer widths over the range 100 to 1000 with step size of 100. The total number of 100 trials were conducted which produce the final configuration of 900 hidden units and 600 FC nodes chosen based on the lowest validation loss. Another

FC layer is employed to yield a 2016×1 array as the intermediate output. The last layer is an output sequence layer that reconstructs the signals. The six layers of the network yield a total of around 7.5 million learnable parameters that are capable of capturing temporal information in the input data for pattern recognition tasks.

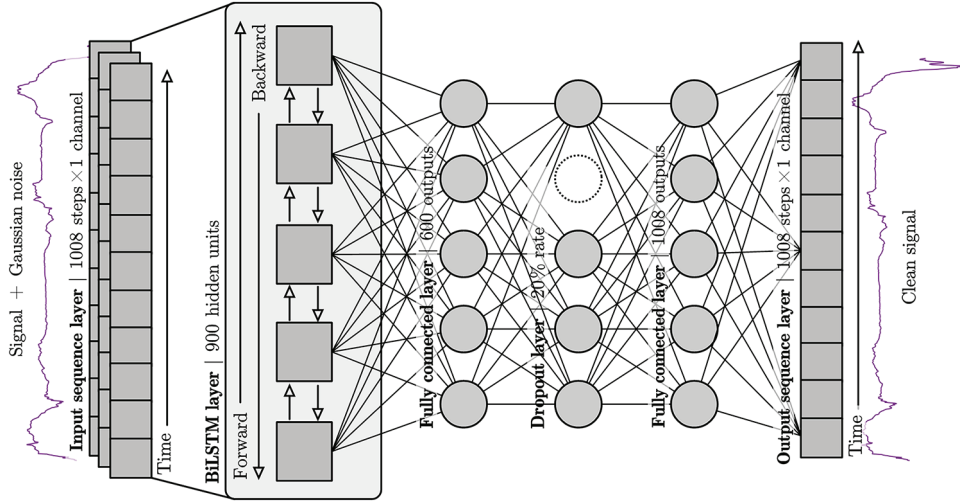


Figure 3: The network architecture consists of the sequence input layer, BiLSTM layer, fully connected layer, dropout layer, fully connected layer, and output sequence layer. Line plots on the left and right represent the dimensions of data on both ends.

3.2 Negative Learning

In this study, negative learning is applied to detect geomagnetic field anomalies that precedes earthquakes. Model was trained to reconstruct non-seismic (normal) geomagnetic field sequences while being penalized for accurately reconstructing seismic (anomalous) sequences. This is achieved through a composite loss function that comprises three main components.

The first component is weighted reconstruction error that measures the reconstruction error between the original and reconstructed sequences. Errors are weighted by their magnitude in order to prioritize larger deviations that are more likely present in seismic data. This was achieved by scaling the MSE term with a factor proportional to the absolute error as shown in:

$$\text{Weighted MSE} = \frac{1}{T} \sum_{t=1}^T (1 + |\hat{y}_t - y_t|) \cdot (\hat{y}_t - y_t)^2$$

where $\hat{y}_t$ is the constructed value, y_t is the original value. The concept of weighting errors by their magnitude initially stems from focal loss that prioritizes challenging or difficult-to-classify instances in classification modeling. Thus, in anomaly identification, this would ensure that the model focuses on significant deviations rather than smaller noise or fluctuations.

The second aspect is in class-specific penalty that computes reconstruction error separately and combined it with a weighing factor that determine the contribution of seismic anomalous signal. It mirrors contrastive loss that wanted to learn the embeddings or features that separates classes [29]. The equation could be expressed as:

$$\mathcal{L}_{\text{class}} = \text{MSE}_{\text{normal}} + \lambda \cdot \text{MSE}_{\text{anomaly}}$$

where $\lambda = 0.1$ which has been selected based on empirical evaluation during model development, with the goal of maximizing RMSE separation between seismic and non-seismic data. The chosen value yielded the best class separability on the test set while preserving stability in non-seismic reconstructions. Such term is needed due to the indistinguishable nature of anomalous and normal signal when used on a classification model, with an assumption that the model are able to recognize and decides which section contributes as a precursory signal.

Next, the loss function also considered temporal consistency by utilizing temporal smoothness penalty. This concepts has been widely used in sequence modeling to ensure coherence such as in video prediction [30] or CNN outputs stabilization [31]. Physical process such as geomagnetic field evolves smoothly over time which necessitates temporal consistency. Hence, in order to prevent erratic reconstructions due to maximizing error on anomalies, a penalty term is introduced to minimize the difference between two consecutive predictions. This expression is shown as:

$$\mathcal{L}_{\text{temp}} = M \cdot \frac{1}{T-1} \sum_{t=2}^T (\hat{y}_t - \hat{y}_{t-1})^2$$

where the M term controls the strength of the penalty. A coefficient of $M = 0.01$ was selected based on its ability to suppress high-frequency artifacts without oversmoothing genuine variation. Essentially, this term discourages sudden fluctuations in reconstructed signals, following the gradual nature of pre-seismic geomagnetic changes [7]. The total loss function integrates these components such that:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{class}} + \mathcal{L}_{\text{temp}}$$

During model development, we manually tested the contribution of each loss component. The model using only standard MSE (normal learning) produced minimal class-wise RMSE difference. Introducing the class-specific penalty term enhanced anomaly sensitivity by elevating seismic reconstruction error. The temporal smoothness regularizer was found to be essential in mitigating erratic outputs when seismic errors were deliberately amplified. These findings support the conclusion that the full composite loss improves both model discriminability and signal quality.

3.3 Model Training

The BiLSTM network was then trained on the training dataset comprising 1205 non-seismic and seismic samples for updating the model's weights and optimizing its performance. The training hyperparameters were configured to facilitate effective convergence and optimal model performance. An optimal balance between computing efficiency and model update frequency is achieved by selecting a mini-batch size of 128. The initial learning rate is set at 0.0001, and a piecewise learning rate scheduling strategy is adopted, reducing the learning rate by a factor of 0.95 every 30 epochs to fine-tune the model's convergence. Simultaneously, the model's generalization capability was assessed using the validation dataset every 25 epochs. After training, the neural network corresponding to the training iteration with the lowest validation loss was selected. This ensured that the BiLSTM network not only converged effectively but also maintained its ability to generalize to unseen data. The training and validation loss throughout the training process are depicted in Fig. 4, with the lowest loss observed at the 1636th iteration with the values of 5.1626.

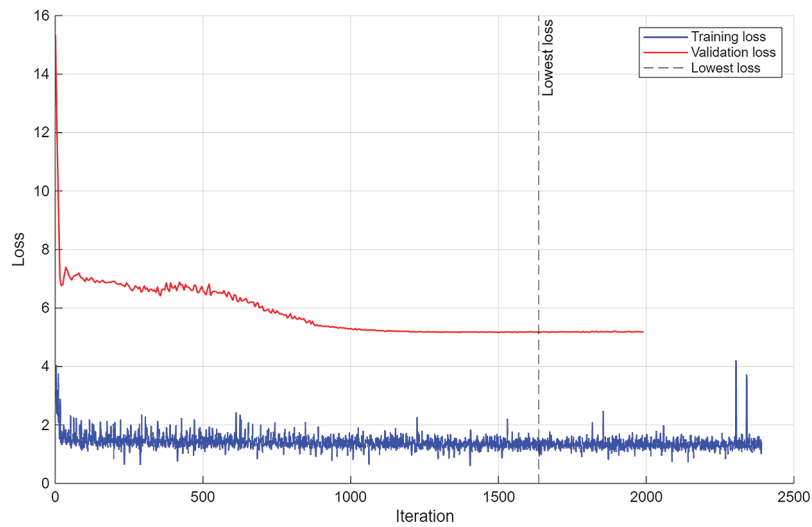


Figure 4: Training and validation loss throughout the training period. The lowest loss value was achieved after 1636 iterations, indicated by the vertical dashed line.

Fig. 5 further illustrates the distinct separation between normal (non-seismic) and anomalous (seismic) MSE values that highlights the effects of negative learning in differentiating the two classes. The MSE for normal samples (represented by the cyan and magenta lines for training and validation, respectively) remains consistently lower compared to anomalous samples (depicted by the green and black dashed lines) throughout the training progress, reflecting the network's ability in identifying anomalous signals. This gap in error values underscores the robustness of the model in distinguishing between normal and anomalous instances, further reinforcing the impact of the training strategy. The vertical dashed line marks the iteration at which the lowest validation loss was achieved, indicating the optimal model state.

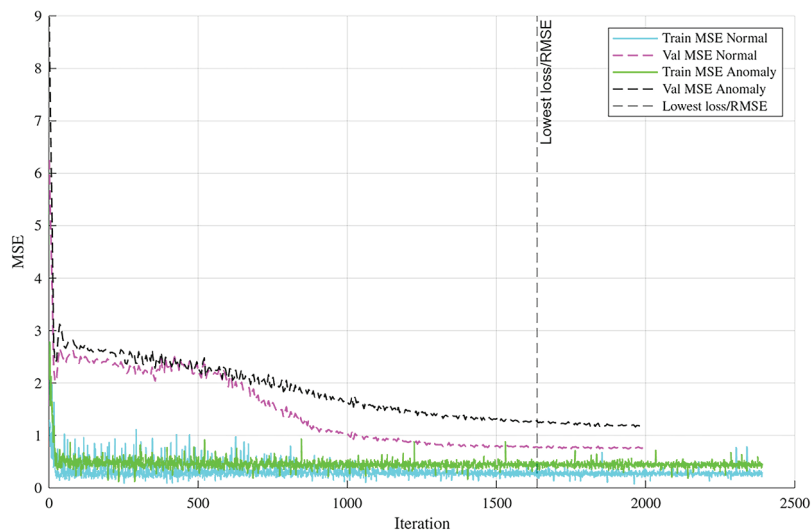


Figure 5: Training and validation loss for normal and anomalous samples. The clear separation highlights the model's ability to distinguish anomalies using negative learning. The vertical dashed line marks the iteration with the lowest validation loss.

4 Model Evaluation and Analysis

The developed model was evaluated on an unseen evaluation dataset comprising seismic and non-seismic classes. The objective was not only to assess the model performance but also to observe significant performance differences between the two classes to test our hypothesis: the model's performance is expected to be worse in the seismic data set since the model has been trained to reconstruct non-seismic better compared to seismic data. If such an observation could be made, we could conclude that the Z geomagnetic field measured in proximity to an earthquake exhibits anomalous and distinguishable characteristics prior to the event when compared with similar data without any nearby earthquake. The input signals were inserted into the model and the output were evaluated against the original by using two metrics to measure the model error. The metrics are RMSE, mean absolute scaled error (MASE), and coefficient of determination (R^2), which are calculated as follows:

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

In the formulae, n is the total number of observations, y_i and $\hat{y}_i$ are the actual and predicted i -th observation, respectively, whereas $\bar{y}$ is the average of actual observational values. RMSE measures the average magnitude of the errors between predicted and actual values whereas MASE evaluates the model accuracy by comparing the mean absolute error to the mean absolute error of a naive (baseline) model. MASE has been formulated as a new method for measuring the accuracy of forecasts by modifying the mean absolute percentage error (MAPE) and symmetric MAPE (SMAPE) methods. It has been found to perform well in measuring the accuracy of forecast data that have intermittent demand or seasonal patterns [32]. For both metrics, their values range from zero to positive infinity, with lower values indicating better accuracy [33]. On the other hand, R^2 quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables, with values between 0 and 1 indicating the percentage of variability in the dependent variable that can be explained by the model [34].

4.1 Case Studies of Model Reconstruction

We first looked at four case studies to exemplify good and poor reconstructions (corresponding to the highest and lowest RMSE, respectively) for seismic and non-seismic classes. Fig. 6 displays the original (blue lines) and the reconstructions (red lines) on the left axis, whereas the MASE of the corresponding time segment (i.e., $\text{MASE}_{\text{segment}}$) is indicated by the grey bars on the right axis. As indicated in the plots, the minimum RMSE for both classes was between 0.2582–0.2609, while the maximum was 0.6260–0.7358, which it was higher for the seismic class in both cases. The reconstructed and original plots are more overlapping for the minimum RMSE (Fig. 6a,b) and vice versa. In Fig. 6c, the signal was noisier than the signal in Fig. 6d, shown by the significantly higher value of RMSE for the former compared to the latter. The observation implies that reconstruction error was not influenced by the signal noise, hence, other factors might be in effect. We then conducted a temporally localized examination by looking at $\text{MASE}_{\text{segment}}$ that reveals irregular variations in error magnitude throughout each timeframe, without apparent temporal coherence or clear anomalies consistently aligned with seismic activity. This irregular pattern of error might result from

the model's negative learning strategy, which inherently amplifies errors during the reconstruction of seismic segments, occasionally leading to notable, localized fluctuations. Nevertheless, the temporal loss strategy appears effective in mitigating abrupt and excessive deviations. The discussed case studies provided insights into the model behavior towards individual data. Meanwhile, analysis of the distribution of error metrics involving all available data revealed the statistical difference between the two classes, which is discussed in the subsequent section.

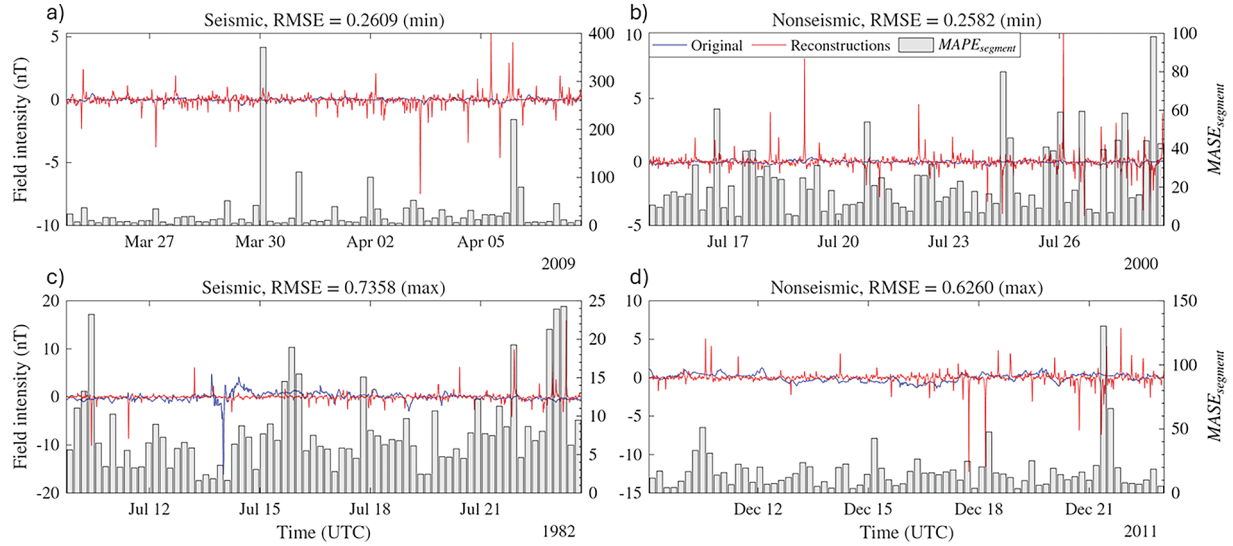


Figure 6: Case studies of earthquakes that record minimum and maximum RMSE for seismic and non-seismic classes.

4.2 Distribution of Error Metrics

In the investigation of error metric distributions, we segmented both the reconstructed and observed signals into intervals of 24-time steps (240 min or 4 h). For each interval, we calculated RMSE, MASE, and R^2 metrics separately and obtained the median value of each metric per sample. Fig. 7 visualizes the resulting histograms of these medians, where seismic data is represented by blue and non-seismic data by red. Examining the RMSE (top panel), a noticeable separation is evident; seismic segments are predominantly shifted toward higher errors, indicating that the model reconstructs seismic segments with less accuracy compared to non-seismic ones. Specifically, the median RMSE value for seismic data (0.3776) is higher than for non-seismic data (0.3579). A Mann-Whitney U test, used to statistically evaluate the difference between these two groups, yields a p -value of 0.003519. This result is significant at the stringent threshold of $p < 0.005$, recently advocated as a more robust standard for statistical significance [35]. Thus, the test strongly supports the notion that the model can effectively differentiate seismic from non-seismic segments based on reconstruction error, potentially enhanced by the model's negative learning characteristic.

Conversely, for the MASE metric (middle panel), the median values for seismic (7.2950) and non-seismic (7.3214) data are closely aligned, suggesting a negligible distinction. The Mann-Whitney U test further confirms this lack of significant difference (p -value = 0.611960). Similarly, the R^2 metric (bottom panel) also shows minimal separation between seismic (median = 0.0839) and non-seismic (median = 0.0860) segments, resulting in another statistically non-significant difference (p -value = 0.301603). Therefore, while RMSE distinctly captures differences between seismic and non-seismic signals, neither MASE nor R^2 provides substantial discriminatory power in this analysis. We subsequently examined the temporal evolution of

these metrics, aiming to determine whether increases in reconstruction error could precede seismic events, potentially offering predictive insights.

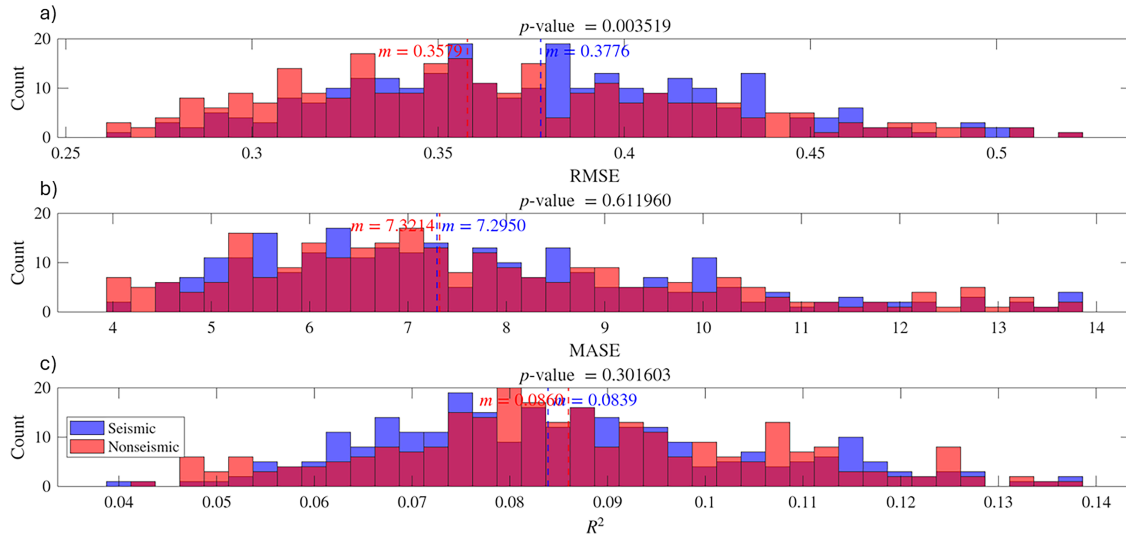


Figure 7: Histograms of median values of (a) RMSE, (b) MASE and (c) R^2 . The blue and red histograms correspond to seismic and non-seismic classes, respectively, where the median of medians, m , and the p -value for the Mann-Whitney U test are included for each metric.

4.3 Superposed Epoch Analysis

The temporal analysis was conducted based on the superposed epoch analysis (SEA), a statistical technique designed to investigate recurring trends by aligning events of interest to a common reference point. By averaging or superposing these aligned data points, underlying patterns and collective behaviors of the events can potentially be revealed [36]. This approach was implemented by computing the summation of the metric values across segments that coincide temporally relative to each earthquake event, as defined by:

$$s(t) = \sum_{i=1}^n y_i(t - t_r) \quad (4)$$

where $y_i(t - t_r)$ represents the shifted time series. The resulting summed values for each metric (RMSE, MASE and R^2) are shown in Fig. 8, with seismic and non-seismic datasets depicted by blue and red lines, respectively.

For RMSE, both classes exhibit irregular fluctuations over time without a consistent or monotonic trend leading up to the earthquake. While there is a general shift toward higher reconstruction error in seismic data, this shift is not localized to a specific temporal window relative to the event onset. This finding suggests that, although our model effectively distinguishes seismic from non-seismic sequences in a statistical sense, it does not identify a temporally fixed precursor signature. Likewise, analysis of MASE and metrics also demonstrates inconsistent temporal patterns, without identifiable anomalies or temporal alignment specifically preceding earthquakes. To systematically quantify the temporal differences between seismic and non-seismic classes, the normalized difference (ND) was calculated using the following equation:

$$ND = \frac{\epsilon_{i,\text{seismic}} - \epsilon_{i,\text{nonseismic}}}{\epsilon_{\max} - \epsilon_{\min}} \times 100 \quad (5)$$

where $\epsilon_{i,\text{seismic}}$ and $\epsilon_{i,\text{nonseismic}}$ are the error metrics at the i -th time segment for seismic and non-seismic datasets, respectively. The ND, represented by grey bars in Fig. 8, further confirms this irregularity, exhibiting no consistent temporal peaks or anomalies indicative of seismic precursors.

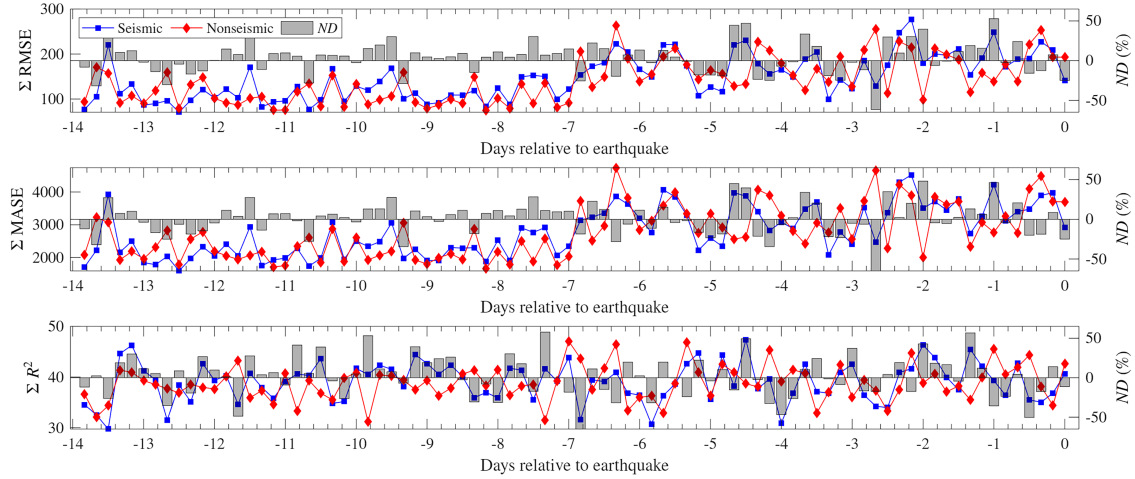


Figure 8: Superposed values (left axis) and normalized difference (right axis) of (a) RMSE, (b) MASE, and (c) R^2 . The blue and red marked lines correspond to seismic and non-seismic classes, whereas the grey bars represent ND.

These results reinforce the notion that geomagnetic anomalies, if associated with seismic activity, do not manifest at uniform time intervals before earthquakes. Rather, the detected differences in reconstruction error reflect broader class-level distinctions between anomalous and normal geomagnetic behavior. This outcome is consistent with previous findings reporting variable and case-dependent lead times, ranging from a few days [5,37] to several weeks or even months [8].

5 Conclusion and Recommendation

This research successfully developed and validated a reconstruction-based anomaly detection approach augmented by negative learning, significantly improving the identification of geomagnetic anomalies preceding seismic events. By deliberately maximizing reconstruction errors for seismic signals, the BiLSTM model demonstrated a statistically significant ability to differentiate between seismic and non-seismic classes, as evidenced by the Mann-Whitney U test results (p -value = 0.0035 for RMSE). The effectiveness of this strategy underscores the advantage of explicitly penalizing correct reconstructions of seismic anomalies, thereby enhancing model sensitivity to subtle earthquake precursors often overlooked in conventional methods.

Given the promising outcomes achieved in this study, future research efforts should further enhance model performance through targeted improvements. Expanding the datasets to include a more comprehensive range of seismic events across diverse geographic and tectonic contexts could strengthen generalizability and further validate the robustness of the negative learning approach. Additionally, optimizing the negative learning framework—such as fine-tuning the composite loss function and adjusting hyperparameters—may further heighten sensitivity to subtle anomalies and reduce false positives. Future work could also extend this research by benchmarking the proposed method against baseline models such as CNN-LSTM or Transformer-based architectures, to evaluate whether negative learning within a reconstruction framework offers a feasible and interpretable alternative for detecting earthquake-related anomalies. Integrating complementary geophysical indicators, including ionospheric total electron content, radon gas emission, or thermal infrared anomalies, may provide additional contextual information and enhance overall anomaly detection

capability. Finally, the adoption of more expressive architectures, such as hybrid or attention-based networks, could improve the model's capacity to capture complex temporal dependencies, potentially advancing the accuracy and robustness of future earthquake forecasting systems.

Acknowledgement: This work was supported by the Ministry of Higher Education through Universiti Putra Malaysia (UPM). For the ground magnetometer data we gratefully acknowledge: INTERMAGNET, Alan Thomson; CARISMA, PI Ian Mann; CANMOS, Geomagnetism Unit of the Geological Survey of Canada; The S-RAMP Database, PI K. Yumoto and Dr. K. Shiokawa; The SPIDR database; AARI, PI Oleg Troshichev; The MACCS program, PI M. Engebretson; GIMA; MEASURE, UCLA IGPP and Florida Institute of Technology; SAMBA, PI Eftyhia Zesta; 210 Chain, PI K. Yumoto; SAMNET, PI Farideh Honary; IMAGE, PI Liisa Juusola; Finnish Meteorological Institute, PI Liisa Juusola; Sodankylä Geophysical Observatory, PI Tero Raita; UiT the Arctic University of Norway, Tromsø Geophysical Observatory, PI Magnar G. Johnsen; GFZ German Research Centre For Geosciences, PI Jürgen Matzka; Institute of Geophysics, Polish Academy of Sciences, PI Anne Neska and Jan Reda; Polar Geophysical Institute, PI Alexander Yahnin and Yaroslav Sakharov; Geological Survey of Sweden, PI Gerhard Schwarz; Swedish Institute of Space Physics, PI Masatoshi Yamauchi; AUTUMN, PI Martin Connors; DTU Space, Thom Edwards and PI Anna Willer; South Pole and McMurdo Magnetometer, PI's Louis J. Lanzarotti and Alan T. Weatherwax; ICESTAR; RAPIDMAG; British Antarctic Survey; McMac, PI Dr. Peter Chi; BGS, PI Dr. Susan Macmillan; Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN); MFGI, PI B. Heilig; Institute of Geophysics, Polish Academy of Sciences, PI Anne Neska and Jan Reda; University of L'Aquila, PI M. Vellante; BCMT, V. Lesur and A. Chambodut; Data obtained in cooperation with Geoscience Australia, PI Andrew Lewis; AALPIP, co-PIs Bob Clauer and Michael Hartinger; MagStar, PI Jennifer Gannon; SuperMAG, PI Jesper W. Gjerloev; Data obtained in cooperation with the Australian Bureau of Meteorology, PI Richard Marshall.

Funding Statement: This work was funded by the Ministry of Higher Education through Universiti Putra Malaysia (UPM) under Grant FRGS/1/2023/STG07/UPM/02/4.

Author Contributions: Nur Syaiful Afrizal: Conceptualization, Investigation, Methodology, Software, Visualization, Writing—original draft. Khairul Adib Yusof: Funding acquisition, Investigation, Methodology, Software, Visualization, Writing—review & editing. Lokman Hakim Muhamad: Investigation, Visualization, Writing—review & editing. Nurul Shazana Abdul Hamid, Mardina Abdullah and Syamsiah Mashohor: Conceptualization, Methodology, Validation, and Writing—review & editing. Mohd Amiruddin Abd Rahman and Masashi Hayakawa: Funding acquisition, Project administration, Validation and Writing—review & editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The original datasets used in this study are publicly available from open databases: 1-minute geomagnetic field data from the SuperMAG global magnetometer network (<https://supermag.jhuapl.edu/mag>, accessed on 20 July 2025), global geomagnetic indices from the NASA OMNIWeb Service (https://omniweb.gsfc.nasa.gov/form/omni_min.html, accessed on 20 July 2025), and earthquake catalogue from the United States Geological Survey database (<https://earthquake.usgs.gov/earthquakes/search>, accessed on 20 July 2025). The data processed and analysed in this study are available from the corresponding author on reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Chen H, Han P, Hattori K. Recent advances and challenges in the seismo-electromagnetic study: a brief review. *Remote Sens.* 2022;14(22):5893. doi:10.3390/rs14225893.
2. Hayakawa M, Schekotov A, Izutsu J, Nickolaenko AP, Hobara Y. Seismogenic ULF/ELF wave phenomena: recent advances and future perspectives. *Open J Earthq Res.* 2023;12(3):45–113. doi:10.4236/ojer.2023.123003.

3. Hayakawa M. Earthquake prediction with radio techniques. 1st ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015. doi:10.1002/9781118770368.
4. Yusof KA, Abdullah M, Hamid NSA, Ahadi S, Yoshikawa A. Correlations between earthquake properties and characteristics of possible ULF geomagnetic precursor over multiple earthquakes. *Universe*. 2021;7(1):20. doi:10.3390/universe7010020.
5. Potirakis SM, Schekotov A, Contoyiannis Y, Balasis G, Koulouras GE, Melis NS, et al. On possible electromagnetic precursors to a significant earthquake ($M_w = 6.3$) occurred in Lesvos (Greece) on 12 June 2017. *Entropy*. 2019;21(3):241. doi:10.3390/e21030241.
6. Pulinets S, Ouzounov D, Karelin A, Davidenko D. Lithosphere-atmosphere-ionosphere-magnetosphere coupling—a concept for pre-earthquake signals generation. In: Ouzounov D, Pulinets S, Hattori K, Taylor P, editors. *Geophysical monograph series*. 1st ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2018. p. 77–98. doi:10.1002/9781119156949.ch6.
7. Hayakawa M, Izutsu J, Schekotov A, Yang S-S, Solovieva M, Budilova E. Lithosphere-atmosphere-ionosphere coupling effects based on multiparameter precursor observations for February–March 2021 earthquakes ($M \sim 7$) offshore Tohoku, Japan. *Geosciences*. 2021;11(11):481. doi:10.3390/geosciences11110481.
8. Yusof KA, Abdullah M, Hamid NSA, Ahadi S, Ghamry E. Statistical global investigation of pre-earthquake anomalous geomagnetic diurnal variation using superposed epoch analysis. *IEEE Trans Geosci Remote Sens*. 2022;60:1–13. doi:10.1109/TGRS.2021.3093555.
9. Abraham A, Asha VA, Ligi C, Renuka G, Blessy V. Isolation of earthquake precursors from interplanetary magnetic field-geomagnetic field coupling. *Indian J Phys*. 2020;94(8):1147–51. doi:10.1007/s12648-019-01557-w.
10. Feng L, Qu R, Ji Y, Zhu W, Zhu Y, Feng Z, et al. Multistationary geomagnetic vertical intensity polarization anomalies for predicting $M \geq 6$ earthquakes in Qinghai, China. *Appl Sci*. 2022;12(17):8888. doi:10.3390/app12178888.
11. Hirano T, Hattori K. ULF geomagnetic changes possibly associated with the 2008 Iwate-Miyagi Nairiku earthquake. *J Asian Earth Sci*. 2011;41(4–5):442–9. doi:10.1016/j.jseaes.2010.04.038.
12. Sarker IH. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Comput Sci*. 2022;3(2):158. doi:10.1007/s42979-022-01043-x.
13. Petrescu L, Moldovan IA. Prospective neural network model for seismic precursory signal detection in geomagnetic field records. *Mach Learn Knowl Extr*. 2022;4(4):912–23. doi:10.3390/make4040046.
14. Tsai TC, Jhuang HK, Ho YY, Lee LC, Su WC, Hung SL, et al. Deep learning of detecting ionospheric precursors associated with $M \geq 6.0$ earthquakes in Taiwan. *Earth Space Sci*. 2022;9(9):e2022EA002289. doi:10.1029/2022EA002289.
15. Xiong P, Long C, Zhou H, Zhang X, Shen X. GNSS TEC-based earthquake ionospheric perturbation detection using a novel deep learning framework. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2022;15:4248–63. doi:10.1109/JSTARS.2022.3175961.
16. Siami-Namini S, Tavakoli N, Namin AS. The performance of LSTM and BiLSTM in forecasting time series. In: 2019 IEEE International Conference on Big Data (Big Data); 2019 Dec 9–12; Los Angeles, CA, USA. Piscataway, NJ, USA: IEEE; 2019. p. 3285–92. doi:10.1109/BigData47090.2019.9005997.
17. Draz MU, Shah M, Jamjareegulgarn P, Shahzad R, Hasan AM, Ghamry NA. Deep machine learning based possible atmospheric and ionospheric precursors of the 2021 $M_w 7.1$ Japan earthquake. *Remote Sens*. 2023;15(7):1904. doi:10.3390/rs15071904.
18. Akhoondzadeh M, De Santis A, Marchetti D, Wang T. Developing a deep learning-based detector of magnetic, Ne, Te and TEC anomalies from Swarm satellites: the case of $M_w 7.1$ 2021 Japan earthquake. *Remote Sens*. 2022;14(7):1582. doi:10.3390/rs14071582.
19. Khan M, Wang H, Riaz A, Elfatyany A, Karim S. Bidirectional LSTM-RNN-based hybrid deep learning frameworks for univariate time series classification. *J Supercomput*. 2021;77(7):7021–45. doi:10.1007/s11227-020-03560-z.
20. Munawar A, Vinayavekhin P, De Magistris G. Limiting the reconstruction capability of generative neural network using negative learning. In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). Piscataway, NJ, USA: IEEE; 2017. p. 1–6. doi:10.1109/MLSP.2017.8168155.

21. Lee J-H, Astrid M, Zaheer MZ, Lee S-I. Deep visual anomaly detection with negative learning. arXiv:2105.11058. 2021.
22. Pang G, Shen C, Cao L, Van den Hengel A. Deep learning for anomaly detection: a review. *ACM Comput Surv.* 2022;54(2):1–38. doi:10.1145/3439950.
23. Orr L, Chapman SC, Gjerloev JW, Guo W. Network community structure of substorms using SuperMAG magnetometers. *Nat Commun.* 2021;12(1):1842. doi:10.1038/s41467-021-22112-4.
24. Gjerloev JW. The SuperMAG data processing technique. *J Geophys Res Space Phys.* 2012;117(A9):JA017683. doi:10.1029/2012JA017683.
25. Rawat G, Chauhan V, Dhamodharan S. Fractal dimension variability in ULF magnetic field with reference to local earthquakes at MPGO, Ghuttu. *Geomatics Nat Hazards Risk.* 2016;7(6):1937–47. doi:10.1080/19475705.2015.1137242.
26. Toker M, Şahin Ş. Upper- to mid-crustal seismic attenuation structure above the mantle wedge in East Anatolia, Turkey: imaging crustal scale segmentation and differentiation. *Phys Earth Planet Inter.* 2022;329–330:106908. doi:10.1016/j.pepi.2022.106908.
27. Marques De Souza Franco A, Hajra R, Echer E, Bolzan MJA. Seasonal features of geomagnetic activity: a study on the solar activity dependence. *Ann Geophys.* 2021;39(5):929–43. doi:10.5194/angeo-39-929-2021.
28. Gaur D, Folz J, Dengel A. Training deep neural networks without batch normalization. arXiv:2008.07970. 2020. Available from: <https://arxiv.org/abs/2008.07970>.
29. Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2006 Jun 17–22; New York, NY, USA. Piscataway, NJ, USA: IEEE; 2006. Vol. 2. p. 1735–42. doi:10.1109/CVPR.2006.100.
30. Lotter W, Kreiman G, Cox D. Deep predictive coding networks for video prediction and unsupervised learning. arXiv:1605.08104. 2017. doi:10.48550/arXiv.1605.08104.
31. Eilertsen G, Mantiuk RK, Unger J. Single-frame regularization for temporally stable CNNs. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. Piscataway, NJ, USA: IEEE; 2019. p. 11168–77. doi:10.1109/CVPR.2019.01143.
32. Mathai AV, Agarwal A, Angampalli V, Narayanan S, Dhakshayani E. Development of new methods for measuring forecast error. *Int J Logist Syst Manage.* 2016;24(2):213. doi:10.1504/IJLSM.2016.076472.
33. Hodson TO. Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geosci Model Development.* 2022 Mar;15(14):5481–7. doi:10.5194/gmd-2022-64.
34. Onyutha C. From R-squared to coefficient of model accuracy for assessing “goodness-of-fits”. *Geosci Model Dev Discuss.* 2020 May:1–25. doi:10.5194/gmd-2020-51.
35. Benjamin DJ, Berger JO, Johannesson M, Nosek BJ, Wagenmakers E-J, Berk R, et al. Redefine statistical significance. *Nat Hum Behav.* 2017;2(1):6–10. doi:10.1038/s41562-017-0189-z.
36. Manu V, Balan N, Zhang Q-H, Xing Z-Y. Double superposed epoch analysis of geomagnetic storms and corresponding solar wind and IMF in solar cycles 23 and 24. *Space Weather.* 2023;21(3):e2022SW003314. doi:10.1029/2022SW003314.
37. Stănică DA. Anomalous geomagnetic signal emphasized before the Mw8.2 Coastal Alaska earthquake, occurred on 29 July 2021. *Entropy.* 2022;24(2):274. doi:10.3390/e24020274.