future internet

Article

Earthquake Magnitude Detection Utilizing a Novel Hybrid Earth–Transformer–LSTM Architecture

Amir A. Ghavifekr [1,*], Elman Ghazaei [1], Mohsen Mirzajani [2] and Paolo Visconti [3,*]

[1] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz 51666-16471, Iran; elmangazayi@gmail.com
[2] Department of Civil Engineering, Marand Technical Faculty, University of Tabriz, Tabriz 51666-16471, Iran; m.mirzajani@tabrizu.ac.ir
[3] Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy
* Correspondence: aa.ghavifekr@tabrizu.ac.ir (A.A.G.); paolo.visconti@unisalento.it (P.V.); Tel.: +39-0832-297334 (P.V.)

Abstract

One of the complicated and demanding tasks in seismology is the reliable detection of earthquakes. The key challenge is that the detection models must be applied to a specific region, and models trained on one region may not perform as well in others. The limitations of datasets for most regions of the world pose another task. Comprehensive, high-quality datasets are essential for developing robust earthquake detection algorithms. Despite these challenges, developing effective earthquake detection systems is critically important. This paper proposes a novel deep network, Earth–Transformer–LSTM (ETL), to estimate earthquake magnitude with high precision. The proposed method uses Transformer encoders as its first layer to extract profound features from the dataset. To obtain highly accurate results, the extracted data is used as the input to the Long Short-Term Memory (LSTM) neural network. Additionally, one-dimensional convolution is replaced by Multi-Layer Perceptron (MLP), which performs better in Transformer encoders' feed-forward networks. The Turkey earthquake dataset 2000–2018 was used in this research because significant earthquakes have occurred in this region in recent years. According to the obtained results, the proposed method's Root Mean Squared Error (RMSE) is 0.7, representing a noticeable improvement over advanced conventional models.

Keywords: Earth–Transformer–LSTM (ETL); deep learning; earthquake detection; USGS dataset

Check for updates

Academic Editors: Diego Vergara and Pablo Fernández-Arias

Received: 31 January 2026
Revised: 7 March 2026
Accepted: 9 March 2026
Published: 11 March 2026

1. Introduction

Turkey is one of the most seismically active countries in the world due to its geographical location. Between 2014 and 2024, eleven significant earthquakes occurred in Turkey, most with magnitudes over 6, suggesting that this region could experience similar events in the near future. Such events can affect public safety and the country's economy. Due to the complexity of earthquakes, seismologists are seeking solutions to better understand seismic events as several factors must be considered. Crucial factors in determining the hazard and spatial probability of earthquakes include magnitude, depth, epicenter, seismic gap, earthquake frequency, slope, elevation, and curvature, with the magnitude, depth, and epicenter being the most important. As a result, earthquake detection and investigation of its characteristics have become crucial in modern seismology for purposes such as earthquake early warning systems, understanding the behavior of tectonic plates across the Earth, and monitoring earthquakes on a global scale.

The typical procedure so far has been to monitor earthquakes using seismographs to record them and prepare earthquake catalogs, allowing researchers to gain a deeper understanding of the earthquake process and fault mechanisms. Since these data are precious to the seismological community, the number of seismic stations has increased and device accuracy has also improved. Furthermore, over recent years, several machine learning (ML) and deep learning (DL) algorithms have been developed based on seismic event data and can be used to detect earthquakes. ML and DL are leading technologies suitable for solving problems across various fields of science. Furthermore, these tools are highly effective in formulating theories and implementing different datasets. In seismology, some of these tools have been used to predict earthquakes.

Rojas et al. [1] surveyed recent applications of Artificial Neural Networks (ANNs) for automating seismic data interpretation, particularly for earthquake detection and onset-time estimation. The paper also discusses the challenges faced by human interpreters, especially in detecting small magnitude earthquakes, and how ANNs can address these uncertainties. Mehta et al. [2] developed a normalized ANN to predict earthquakes; using Exploratory Data Analysis (EDA), they removed unnecessary features from the raw data and optimized the model using backpropagation, achieving 94.3% accuracy. Ridzwan and Yusoff [3] reviewed numerous machine learning studies on earthquake prediction and compared models predicting magnitude, trend, and occurrence across regions, identifying top-performing seismic indicators and algorithms. Kavianpour et al. [4] introduced a CNN-BiLSTM-AM approach for earthquake prediction. Convolutional Neural Networks (CNNs) extracted spatial characteristics, Bi-directional Long Short-Term Memory (BiLSTM) captured data's long-term dependencies, and attention mechanism (AM) emphasized highly contributing results. Sadhukhan et al. [5] predicted earthquake magnitudes using Long Short-Term Memory (LSTM), Bi-LSTM, and Transformer models on eight seismic indicators, considering Mean Absolute Error (MAE), Mean Squared Error (MSE), log-cosh loss, and Mean Squared Logarithmic Error (MSLE) metrics.

Doğan and Demir [6] proposed a novel Structural Recurrent Neural Network (SRNN) model to predict earthquakes based on the existence of faults in specific regions, outperforming baseline and the state-of-the-art models. Berhich et al. [7] employed an LSTM network with attention to predict large earthquakes using the Japan earthquake dataset, achieving an MSE of approximately 60%. Saad et al. [8] proposed a Compact Convolutional Transformer (CCT) for earthquake phase arrivals, which outperformed EQTransformer and PhaseNet. Aslam et al. [9] predicted seismic activity in the northern part of Pakistan, using Support Vector Regressor (SVR) and Hybrid Neural Network (HNN) to predict seismic events.

Zhou et al. [10] examined the accuracy of the Haskell-based deterministic dendritic cell algorithm (hDCA) for predicting earthquakes with magnitudes greater than 4.5. They first extracted eight seismic indicators (b-value, magnitude deficit, mean magnitude, and energy release rate) from historical earthquake records, using established geophysical laws such as the Gutenberg–Richter relationship. These indicators were mapped into danger and safe signals that simulate immune responses. A population of artificial dendritic cells then processed these signals, accumulated internal state values (csm and k), and classified each monthly seismic state as normal or anomalous, where an anomaly indicated a high probability of an earthquake with magnitude greater than 4.5 in the following month.

Majhi et al. [11] estimated earthquake magnitudes using a method called Functional Link Artificial Neural Network (FLANN), which combines standard machine learning algorithms with the Moth Flame Optimization (MFO) algorithm. Zhang et al. [12] proposed

https://doi.org/10.3390/fi18030143

a machine learning (ML) algorithm, CNN ONsite Intensity Prediction (CONIP), to forecast seismic intensities and noted that this method could improve the accuracy of the earthquake early warning (EEW) process. Zhu et al. [13] combined RNNs and transfer learning to estimate ground-motion intensity at recording stations in China. The predicted alarms had a 90% success rate. Kudłacik et al. [14] used a Global Navigation Satellite System (GNSS)-based method to detect and characterize mining tremors, agreeing with seismic solutions for earthquake first-epoch detection.

Machine learning and deep learning methods have been applied to enhance discrimination between earthquakes and quarry blasts based on three seismic parameters by M. S. Abdalzaher et al. [15]. They trained and compared six classifiers (XGB, DT, KNN, QD, RF, and LR) on seismic catalog data from Northern California, using multiple evaluation metrics (accuracy; F1-score; Matthews Correlation Coefficient—MCC; Kappa; Area Under the Receiver Operating Characteristic Curve—ROC-AUC). The Extreme Gradient Boosting (XGB) model achieved the best performance. Although the results demonstrate strong classification capability and practical efficiency, the study relies on a limited feature set and a single regional dataset, which may restrict generalizability.

Bilal et al. [16] presented a Batch Normalized Graph Convolutional Neural Network with attention mechanism (BNGCNNATT) to predict earthquake depth and magnitude from seismic data. Saad et al. [17] developed an unsupervised deep learning method for denoising single-channel earthquake data using a UNET-based network with attention mechanisms. The method employs a short-time Fourier transform (STFT) to separate the real and imaginary parts, which are then processed to mute seismic noise while preserving the signals. Using a deep image prior (DIP) approach, the network learns iteratively without labeled data. The algorithm outperforms DeepDenoiser by 1.95 dB in signal-to-noise ratio (SNR) and improves seismic event detection and earthquake location accuracy.

Han et al. [18] developed a DL model for earthquake detection, efficiently detecting small-magnitude events compared to the Short Time Average/Long Time Average (STA/LTA) method. However, phase picking was inconsistent, large earthquakes were indistinguishable, and non-natural earthquakes were undetectable. Abebe et al. [19] used a Transformer model to predict earthquake magnitudes in the Horn of Africa, comparing it with LSTM, BiLSTM, and BiLSTM with attention models. The Transformer outperformed the others, achieving lower error rates. Its ability to capture long-range dependencies and process data in parallel contributed to better accuracy, though challenges remained in predicting both large and small earthquakes.

Laurenti et al. [20] applied the autoregressive (AR) forecasting DL method for labquake prediction. The method has been able to detect the time to the start of failure (TTsF) and the time to the end of failure (TTeF) for labquakes, thereby representing a solution to outperform previous methods of earthquake prediction. Mohammadigheymasi et al. [21] implemented a pair-input DL (PIDL) model and waveform migration location methods for earthquake detection, including the phase and location of earthquakes. Yuan et al. [22] enhanced earthquake detection by combining pretrained DL models, multiple frequency bands, and ensemble estimation. Mousavi et al. [23] presented a global DL model to predict seismic events and phase picking. Their model was capable of predicting and locating earthquakes twice as well as other models, using less than a third of the seismic stations.

This paper proposes a novel deep network, called Earth–Transformer–LSTM (ETL), to detect earthquake magnitude using an earthquake catalog. The term “Earth” in the model’s name is chosen primarily to distinguish our model from others that incorporate both Transformer and LSTM architectures. We aimed to create a unique identifier for our proposed hybrid model, even though it does not specifically include Earth-related features. The ETL incorporates the Transformer encoder and LSTM, which have enhanced the extraction of hidden features of its inputs and representation of sequence data's features. In the Transformer architecture, an LSTM enhances the representation of sequence data features. The main dataset has been split into two subsets: a training set and an evaluation set. The training phase was performed using the Adam optimization algorithm with a learning rate of 1×10^{-4} and a cross-entropy loss function. Compared with previous earthquake detection methods, the ETL introduces a Transformer–LSTM with stronger capabilities, achieving more accurate earthquake magnitude detection than other RNN-based models such as Gated Recurrent Units (GRUs) and CNN-LSTM.

In the rest of the manuscript, Section 2 provides background on data collection and the steps for downloading the datasets. In Section 3, the methodology of the proposed method is explained, including the main concepts of the Transformer encoder and LSTM, which provides a more detailed introduction to the novel ETL approach in Section 4. Results and related discussion are provided in Section 5, along with a comparative analysis for the proposed method. Finally, the main findings of the research work and conclusions are summarized in Section 6.

2. Data Collection Background

Looking at recent earthquakes in Turkey, the country has been annually hit by at least one earthquake of a magnitude greater than 5.5 [24]. These earthquakes usually occur in the North Anatolian Fault zone (NAFZ) and East Anatolian Fault zone (EAFZ) [25,26]. There have been several significant earthquakes in Turkey over the last few decades. The most devastating ones have been those that occurred on 30 October 2020, reported as the "Aegean Sea earthquake (7.0 Mw)"; on 23 November 2022, as the "Gölyaka–Düzce earthquake (6.1 Mw)"; and in February 2023, namely the "Turkey–Syria earthquakes (7.8 Mw, 7.7 Mw, 6.3 Mw)" [27–29].

The Turkey–Syria earthquakes (7.8 Mw, 7.7 Mw, 6.3 Mw) in February 2023 have been described in [27], which focuses on fault interaction and stress transfer along the East Anatolian Fault zone. The focal mechanism inversion reveals a trans-tensional stress regime with left-lateral and normal slip, offering insights into regional tectonic processes and fault interactions. The Aegean Sea earthquake (7.0 Mw), which occurred on 30 October 2020, extensively analyzed in [28], was a typical extensional faulting event characterized by a complex aftershock sequence and high variability in rupture directivity. This event demonstrated strong spatial clustering and temporal decay patterns, which have been used in multiple studies to evaluate predictive models of earthquake catalog behavior. The Gölyaka–Düzce earthquake (6.1 Mw), which occurred on 23 November 2022 and is documented in [29], investigates near-field effects in the 6.1 magnitude earthquake on the North Anatolian Fault zone, focusing on ground-motion intensities and their spatial distribution using the AFAD–Turkish Accelerometric Database. The study challenges previous assumptions by showing that velocity pulses, typically associated with directivity effects, are concentrated in the fault-parallel rather than the fault-normal component, especially for moderate events.

Tectonic plates, their boundaries, and active faults in Turkey are shown in Figure 1. As large cities and dams are located around fault zones, the safety of these cities and infrastructure could be affected by future earthquakes. So, it is crucial to study possible earthquakes and quantify them using advanced numerical techniques.

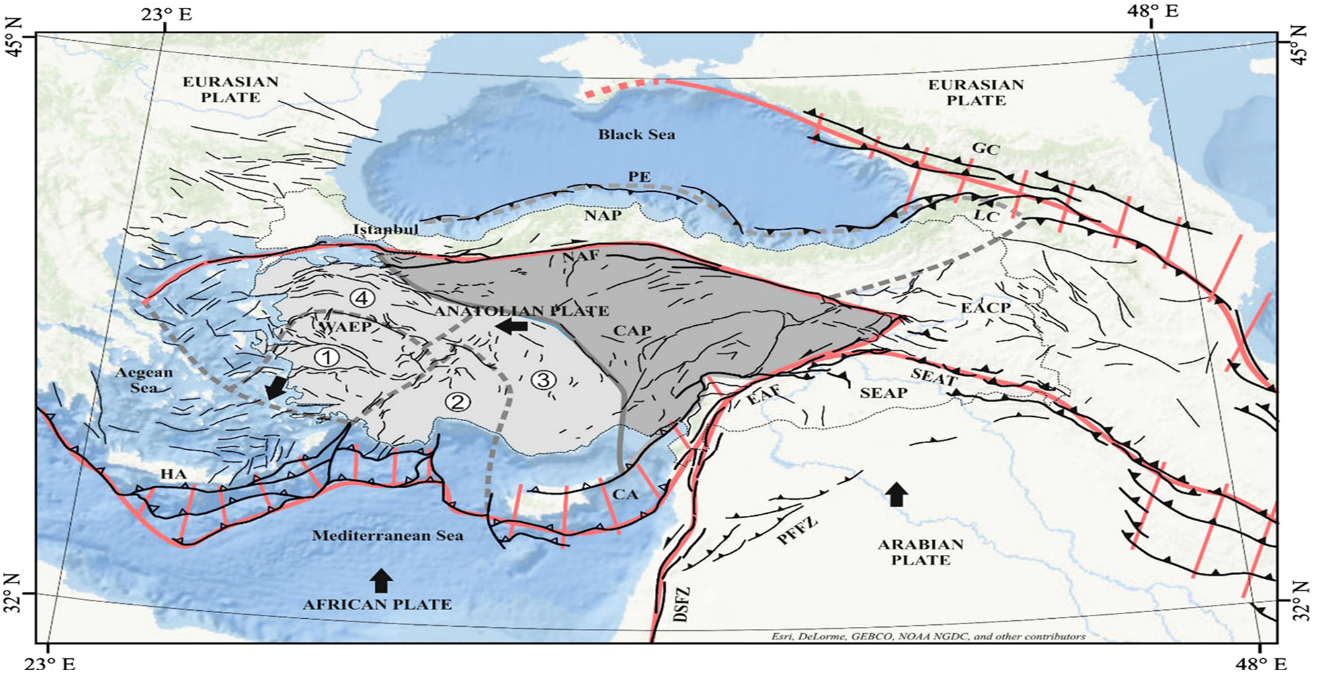

Figure 1. The tectonic plates and active fault map of the Eastern Mediterranean region [30].

Datasets

The data used in this study are sourced from the USGS earthquake catalog [31], covering the period from 1 January 2000 to 31 December 2018, specifically for Turkey. This extensive dataset provides a comprehensive record of seismic events over the specified timeframe. The spatial distribution of these earthquakes is visualized in both 2D and 3D formats [32], as illustrated in Figure 2. These visualizations offer valuable insights into the geographic and depth-related distributions of seismic activity across Turkey, helping to highlight patterns and trends in the dataset. A detailed overview of the dataset, including the number of recorded events, magnitudes, and other relevant attributes, is presented in Table 1, which summarizes the dataset's key characteristics and facilitates a clearer understanding of the data used in this research work.

Figure 3 shows histograms of date, depth, magnitude, and magnitude type. To clarify the depth and magnitude distribution over the considered years, Figure 4 is depicted.

Table 1. Details provided in the Earthquake catalog.

Parameter	Value
Start time	01-01-2000 00:00:00
End time	31-12-2018 23:59:59
Min longitude	25.85
Max longitude	45.14
Min latitude	35.67
Max latitude	42.38
Number of raw earthquakes	6679

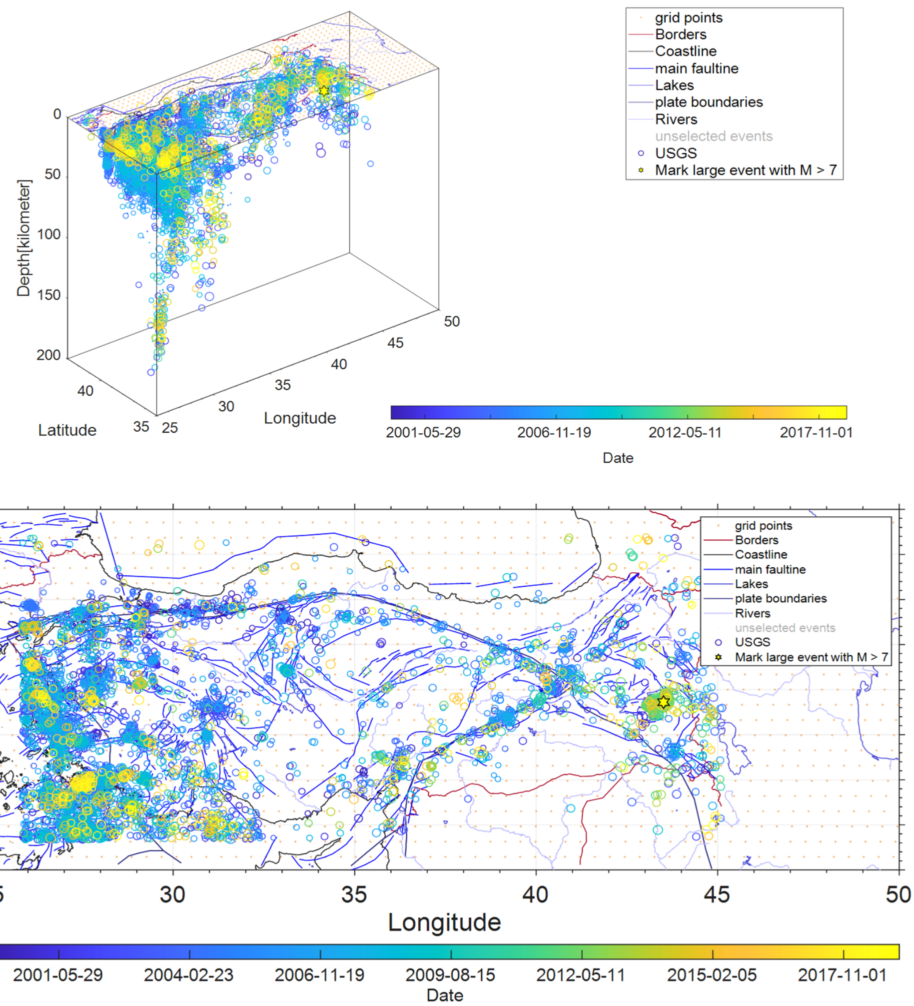

Figure 2. Spatial distribution of earthquakes that occurred in Turkey from 2000 to 2018.

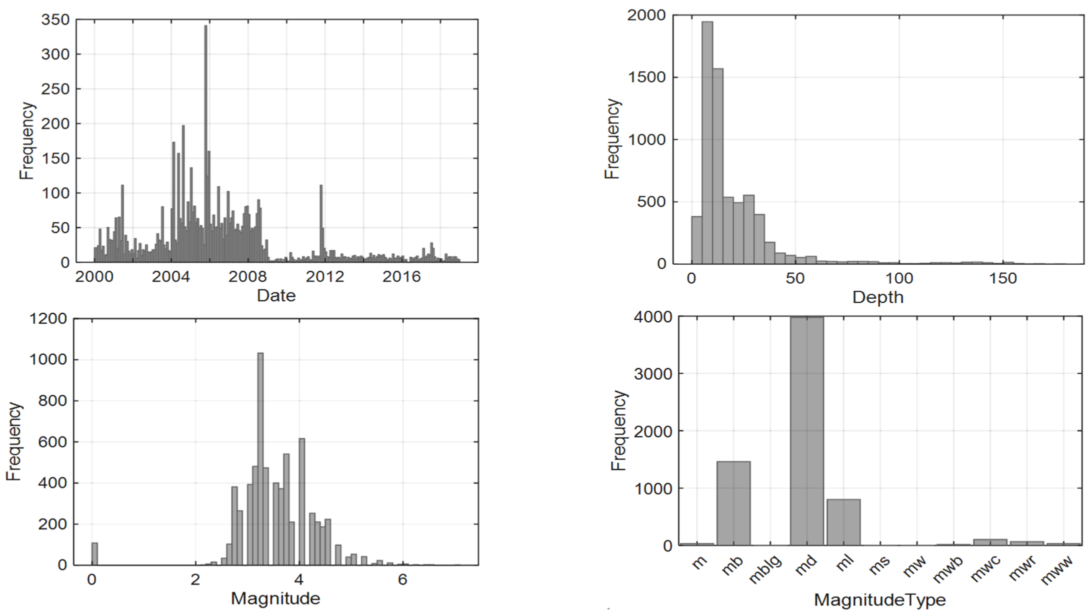

Figure 3. Date, depth, magnitude, and magnitude type histograms for the Turkey dataset.

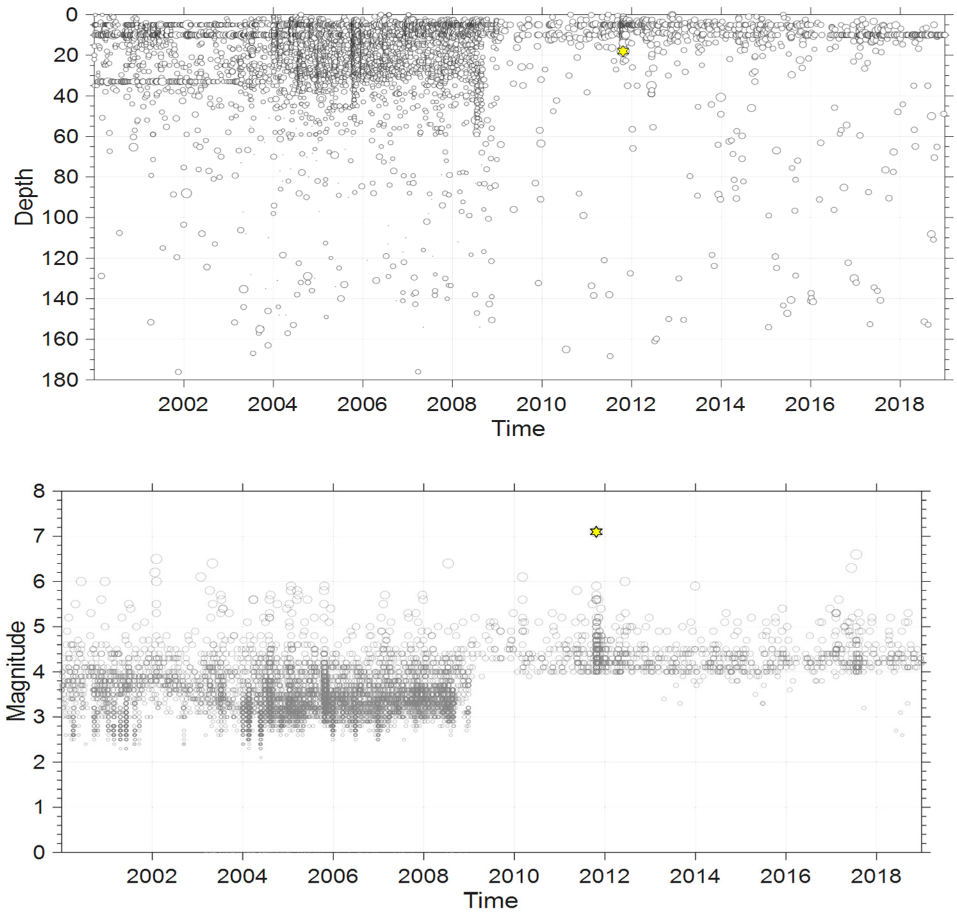

Figure 4. Depth and magnitude distribution between 2000 and 2018 for the Turkey dataset; the star indicates significant events with magnitude greater than 7 (Figures generated by the ZMAP package, https://github.com/zmap/zmap).

3. Methodology

In this paper, we propose a novel deep network, Earth–Transformer–LSTM (ETL), for estimating the earthquake magnitudes with high precision. Firstly, utilizing Transformer encoders as a first layer could extract many subtle features from sequence data. Secondly, exploiting LSTM could process the extracted data to obtain high accuracy. In other words, in this research we use sequence-ensemble-based models to predict future earthquake magnitudes using our dataset.

3.1. Transformer Encoder

Transformers have two main components: encoder and decoder [33]. These cutting-edge deep learning models have gained significant fame in Natural Language Processing (NLP) tasks such as text classification, summarization, and even one-shot classification. Encoder parts can extract many hidden features from the inputs and transfer them to decoder parts, which can generate the desired outputs. Transformers offer many unique benefits over conventional sequence-based models, such as LSTM and GRU. For example, they can process inputs in parallel, since these models have multi-head attention mechanisms that can assign special attention weights to more important segments of the data.

In more detail, this research focuses on the Transformer encoder because it can extract features from sequence inputs. Transformer encoders have several specialized components that can improve the model's performance; for instance, multi-head attention can adjust attention weights and parameters on the more important sections of the data. Layer normalization also has a positive effect on the model's performance, potentially outweighing the batch normalization's effects. The structure of the Transformer encoder block is shown in Figure 5, while the sequence of steps and related parameters are reported in Table 2.

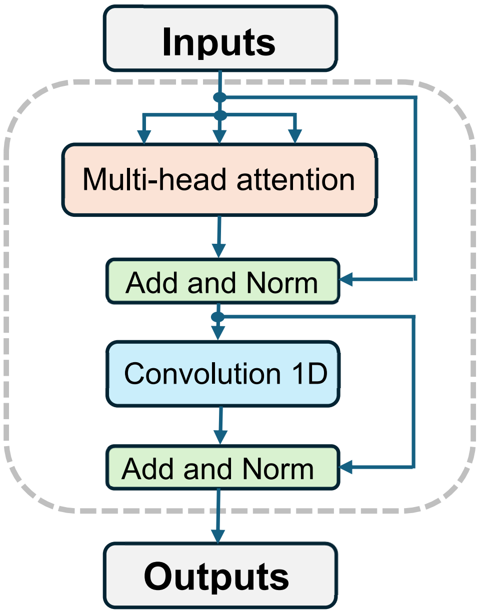

Figure 5. Structure of the Transformer encoder's single block.

Table 2. Sequence of steps and parameters of the proposed Transformer encoder.

Structure of the Transformer Encoder block	
1-	**Inputs:** datapoints (shape = (3,1))
2-	**Outputs:** Encoder output (shape = (100,1))
3-	X = Layer Normalization(inputs)
3-	X = Multi-head attention (X, X, X)
4-	X = Dropout(X)
5-	res = X + inputs
6-	X = Layer Normalization (res)
7-	X = Conv-1D(X)
8-	X = Dropout(X)
9-	X = Conv-1D(X)
10-	Output = x + res
11-	X = Dropout(X)

Multi-head attention is a mechanism used in neural network architectures, particularly in NLP, and is prominently featured in models like Transformer. It enhances the ability of self-attention mechanisms by allowing the model to attend to different parts of the input data simultaneously (Figure 6), effectively enabling it to capture various features of the input's context. The key benefit of multi-head attention is its ability to capture different types of dependencies and relationships within the input data simultaneously. Each attention head can focus on various parts of the input sequence, allowing the model to attend to multiple sides of the context in parallel. This capability can lead to more expressive representations and better performance, particularly in tasks that require understanding long-range dependencies and intricate relationships within the input data. According to Equation (1), the multi-head attention can be computed as follows:

$$MultiHead(Q,K,V) = Concat(head1,\ldots,headh)W^O\left(head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)\right) \tag{1}$$

with $W_i^Q \in \mathbb{R}^{d_{model}\times d_k}$, $W_i^K \in \mathbb{R}^{d_{model}\times d_k}$, $W_i^V \in \mathbb{R}^{d_{model}\times d_v}$ and $W^O \in \mathbb{R}^{hd_v\times d_{model.}}$.

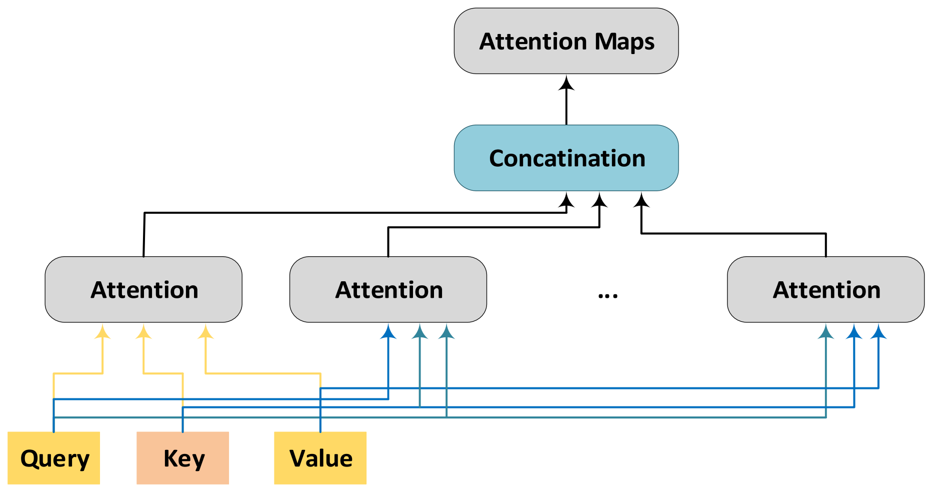

Figure 6. A multi-head attention structure that can compute attention weights in parallel paths.

In this formulation, Q, K, and V denote the query, key, and value matrices, respectively, n is the sequence length, and d_{model} the model embedding dimension. The model employs n attention heads; for each *head* i, W_i^Q, W_i^K and W_i^V are the learnable projection matrices for queries, keys, and values, respectively. In addition, d_k and d_V denote the dimensionality of queries/keys and values per head. The outputs of all heads are concatenated along the feature dimension.

3.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) architecture that is designed to address the vanishing gradient (killing gradient) problem often encountered when training traditional RNNs. LSTM networks are able to learn long-term dependencies in sequential data, making them particularly well-suited for tasks such as speech recognition, language modeling, machine translation, and time-series prediction. LSTM has three main components: input, output, and forget gates. The input gate determines which new information from the current time step should be stored in the memory cell for use by other parts. Moreover, the output gate determines which information from the memory cell should be passed to the next time step's hidden state, which serves as the input to the next cell. Eventually, the forget gate determines which data from the previous time step should be discarded from the memory cell if it is not useful to the model. The LSTM structure is shown in Figure 7.

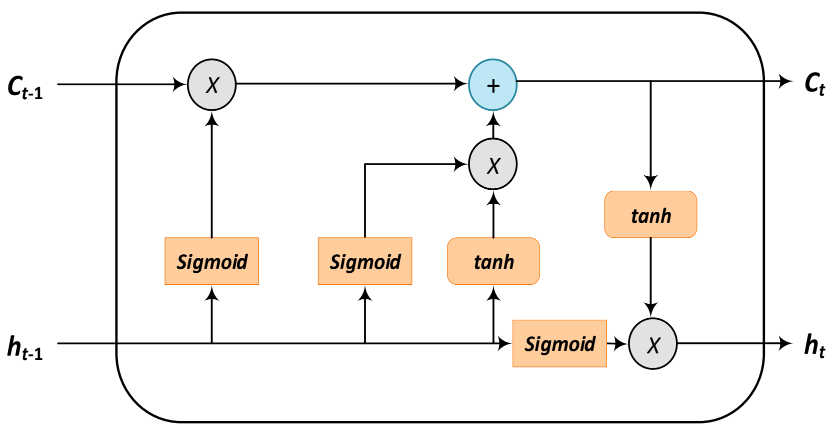

Figure 7. LSTM structure.

4. Earth–Transformer–LSTM (ETL)

In this paper, we propose a novel model combining a Transformer encoder and an LSTM (Figure 8). The primary novelty of our approach lies in the seamless integration of these two distinct architectures. While Transformer encoders have demonstrated remarkable capabilities in extracting contextual features from input data and sequences, they lack the inherent ability to model long-range dependencies effectively. On the other hand, LSTM networks show good performance at capturing sequential patterns and including a steady memory state. This hybrid form aims to utilize the strengths of both models to achieve superior performance. The Transformer encoder structure is equipped with a multi-head attention mechanism, which allows it to extract a wide set of hidden features from the input sequence.

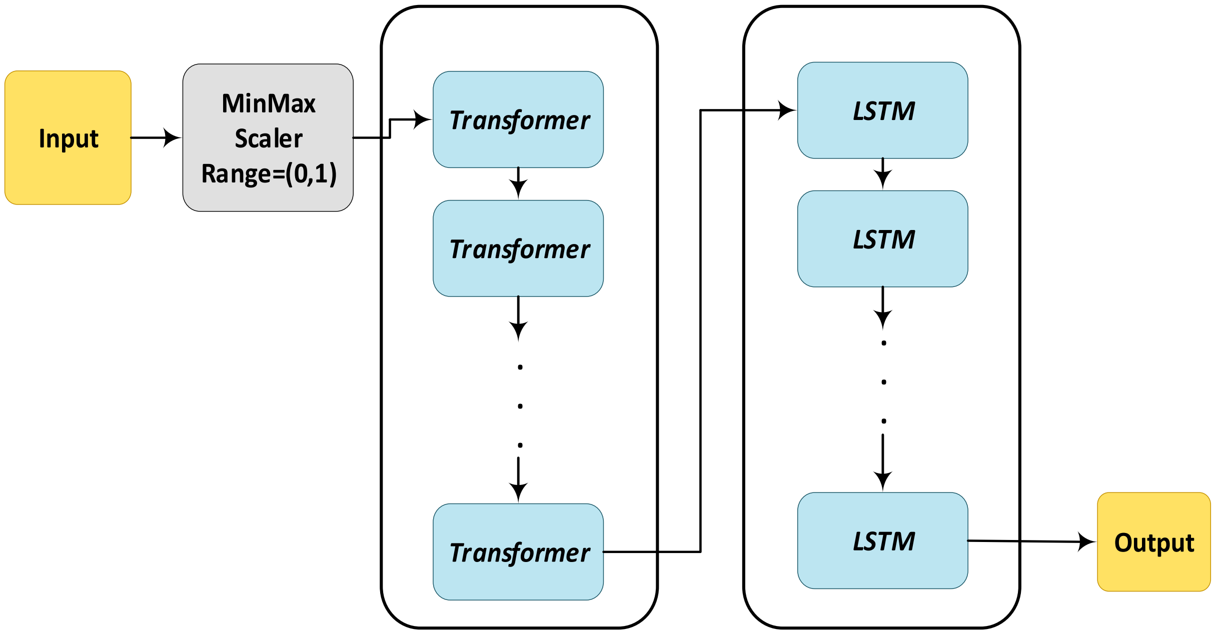

Figure 8. Structure of the developed ETL framework.

This feature is a key merit of the Transformer architecture, as it can cover complex relationships and dependencies within the data that may be difficult for simpler models to learn. By forwarding the Transformer encoder's extracted features to an LSTM layer, our model can further refine and process these representations. The LSTM's ability to maintain internal state and model sequential dependencies can complement the Transformer's strength in modeling global relationships. This combination of techniques enables our model to capture both local and global patterns in the data, potentially improving the detection accuracy. Additionally, we have incorporated one-dimensional convolutional layers into the Transformer encoder's feed-forward networks.

Compared to traditional Multi-Layer Perceptrons (MLPs), the convolutional layers have been shown to exhibit better performance on certain types of sequential data. This architectural choice further enhances the model's capability to extract meaningful features from the input. In the preprocessing stage, we use a MinMax scaler to normalize the raw dataset to a consistent format. This step is crucial for ensuring the model effectively learns from the data, as it helps mitigate the impact of different scales or ranges present in the original features.

Overall, the proposed hybrid model, which we name Earth–Transformer–LSTM (ETL), leverages the strengths of Transformer encoders and LSTMs, along with 1D convolutions and MinMax scaling, to create a powerful and versatile architecture for the task at hand.

https://doi.org/10.3390/fi18030143

The detailed evaluation of the model's performance and its comparative advantages are reported in Table 3.

Table 3. Parameters of the developed ETL model.

Layer	Parameters
Transformer	1816
LSTM_1	4440
LSTM_2	840
Dense	11
Total	7107

Also, we use the MinMax scaler in (2) to scale our dataset to the desired range (0 ÷ 1), which can help prevent overfitting by ensuring that features with different scales do not disproportionately influence the learning process. It also helps stabilize gradient descent, leading to more efficient and effective model training. Overall, using the MinMax scaler enhances the model's performance and robustness.

$$MinMax\ sclaer = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (2)$$

After performing the preprocessing, the dataset is divided into two non-overlapping sets, called the training and test sets, with 80% and 20% of the dataset, respectively. Since the dataset is sequential, the shuffle operation is not applied during splitting, so sequential features are preserved for training the model. This procedure ensures that the sequential features are maintained, which is crucial for effectively training the model. Moreover, as shown in Figure 9, the proposed model takes longitude, latitude, and depth as inputs (3-dimensional vectors) and generates the corresponding estimated magnitude as output (a scalar value).

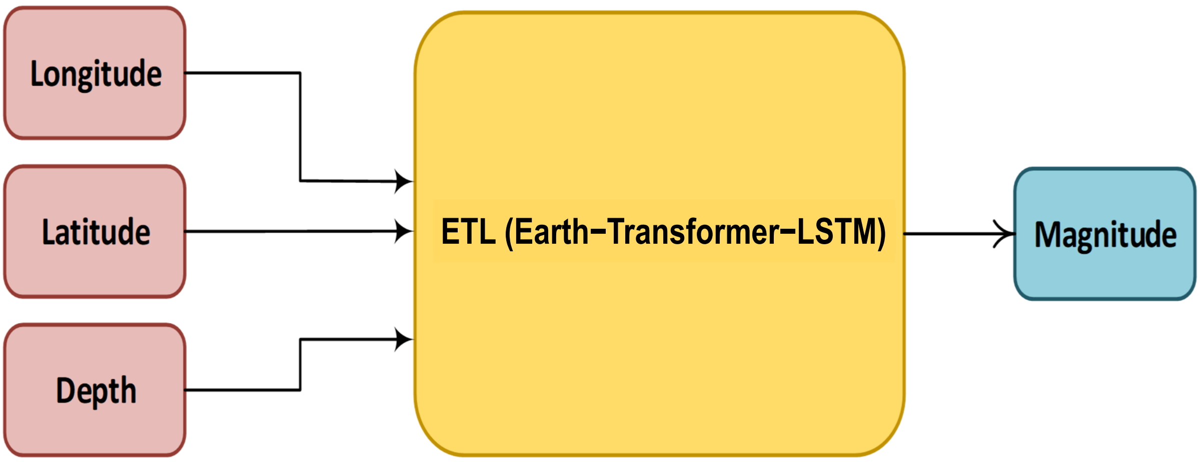

Figure 9. Inputs and outputs of the developed ETL framework.

5. Results

In this paper, Python 3.12 and TensorFlow 2.15.0 are used to implement all models and methods. Moreover, we split the main dataset into two portions—training and testing subsets—with a proportion of 8:2, respectively. Moreover, hyper-parameters can play a crucial role in the models' overall performance, so altering their values can lead to unacceptable results. The hyper-parameters' optimal values for the ETL model are reported in Table 4.

Table 4. ETL hyper-parameters.

	Number of Heads	Number of Transformer Block	LSTM Units	Epochs	Learning Rate	Optimizer
ETL	2	2	10	200	1×10^{-4}	Adam

In this paper, we present a comparison between our proposed hybrid model (ETL model), which combines Transformer encoders and LSTMs, with various benchmark models and other published results. We compute two commonly used performance metrics, Root Mean Squared Error (RMSE) and Mean Squared Error (MSE), to carry out these comparative evaluations. The RMSE metric measures the average magnitude of errors and indicates how well the model's predictions align with the actual target values. It is calculated as the square root of the mean of the squared differences between the predicted and true values. RMSE is a useful metric for evaluating the model's predictive accuracy because it is expressed in the same unit as the target parameter. On the other hand, the average squared difference between the expected and actual values is measured by the MSE metric. Because MSE squares the errors, it gives greater weight to larger deviations, making it more susceptible to outliers or extremely inaccurate predictions. MSE provides insights into the model's ability to minimize overall squared error, an important consideration in many applications. These indices are formulated as follows:

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

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{y} - y_i)^2}{n} \quad (4)$$

We can obtain a more comprehensive understanding of the ETL model's performance by computing both RMSE and MSE. The RMSE provides a sense of the typical magnitude of the prediction errors, while the MSE indicates the model's effectiveness in minimizing the overall squared error.

The benchmark models have been selected carefully to ensure a fair and meaningful comparison. These models represent the state of the art in the field and cover a range of architectural designs, including both traditional and modern techniques. By benchmarking our ETL model against these established methodologies, we can assess the degree of improvement or novelty our hybrid architecture brings to the study. Firstly, we compare our model results with those of the Stacked Normalized Recurrent Neural Network (SN-RNN) [33], which combines multiple RNN-based models to improve overall performance. The SN-RNN has three main components—RNN, GRU, and LSTM. In detail, data is processed inside this RNN stack, and each section has unique benefits. For example, the GRU (middle part) could prevent the model from overfitting, representing a kind of ensembling model. Moreover, the SN-RNN also uses layer normalization, which offers many unique advantages, such as preventing overfitting compared to batch normalization [34]. However, as shown in Table 5, our model performs better than SN-RNN.

Table 5. ETL and SN-RNN results.

	RMSE	MSE
SN-RNN	3.16	9.985
ETL	0.70	0.49

Secondly, within a unified experimental framework, we compare the results of the developed ETL model with other RNN-based models, including LSTM, Gated Recurrent Units (GRUs), CNN-LSTM, and Bi-directional LSTM (BiLSTM). This operating mode ensures consistency in data preprocessing, model selection, optimization strategy, and statistical evaluation algorithms across all models presented in this study. GRU is an improved version of LSTM that helps protect the model from overfitting and other training-related issues. CNN-LSTM can be used for many tasks, such as video classification, and it can extract hidden features from the input via its convolutional layers; these features are then passed to LSTM layers for final processing. BiLSTM can perform conventional LSTM in two directions (first-to-last and last-to-first), and it concatenates the results from both directions. As shown in Table 6, our proposed ETL model (a hybrid Transformer encoder and LSTM) outperforms the other benchmark models.

Table 6. ETL and conventional RNN-based models comparison results.

	RMSE	MSE
LSTM + MinMax scaler	0.95	0.91
GRU + MinMax scaler	0.73	0.54
CNN-LSTM + MinMax scaler	0.87	0.76
BiLSTM + MinMax scaler	0.75	0.56
ETL (Our model)	0.70	0.49

According to Figure 10, our model also has a swift convergence, which can reach its optimal loss (RMSE) at lower epochs (approximately 10 epochs out of 200). In Figure 11, an example of the detected earthquake magnitudes for 10 samples is illustrated.

To further demonstrate the advantages of the proposed ETL architecture, we have analyzed two complementary performance metrics: the relative improvement percentage and training convergence behavior. As reported in Table 6, ETL achieves an RMSE of 0.70 compared to a 0.95 value for LSTM, 0.73 achieved by GRU, 0.87 by CNN-LSTM, and 0.75 by BiLSTM. The obtained result for the ETL model, proposed in this research work, corresponds to relative RMSE reductions of:

- 26.3% improvement over LSTM;
- 4.1% improvement over GRU;
- 19.5% improvement over CNN-LSTM;
- 6.7% improvement over BiLSTM.

Similarly, compared to the SN-RNN model (RMSE value equal to 3.16), the ETL framework reduces prediction error by approximately 77.8%, indicating a substantial gain in predictive accuracy. In addition to lower error values, Figure 10 shows that the ETL model converges to a stable state within approximately 10 epochs, even though it was trained for 200 epochs. This rapid stabilization suggests improved optimization dynamics and learning efficiency, likely due to the complementary integration of Transformer-based global attention and LSTM-based sequential memory. These results provide additional evidence that ETL's superiority is not limited to marginal numerical improvements in RMSE but also reflects enhanced learning stability and stronger representational capability compared to standalone RNN-based architectures.

The principal novelty of the developed ETL framework lies in integrating a Transformer encoder with an LSTM-based model to simultaneously learn long-range spatial correlations among earthquake catalog events via self-attention and temporal evolution patterns via gated recurrent memory. This strategy distinguishes ETL from previous models that rely solely on recurrent learning or convolutional pattern extraction. In terms of robustness, the ETL demonstrated:

https://doi.org/10.3390/fi18030143

➢ Significantly lower RMSE compared to standalone recurrent models under both standard random splitting and temporal hold-out testing;
➢ Stable predictive performance when evaluated on moderate-to-strong magnitude subsets (Mw $\geq$ 5.0), illustrating better generalization compared to benchmark models whose error grew substantially with the magnitude.

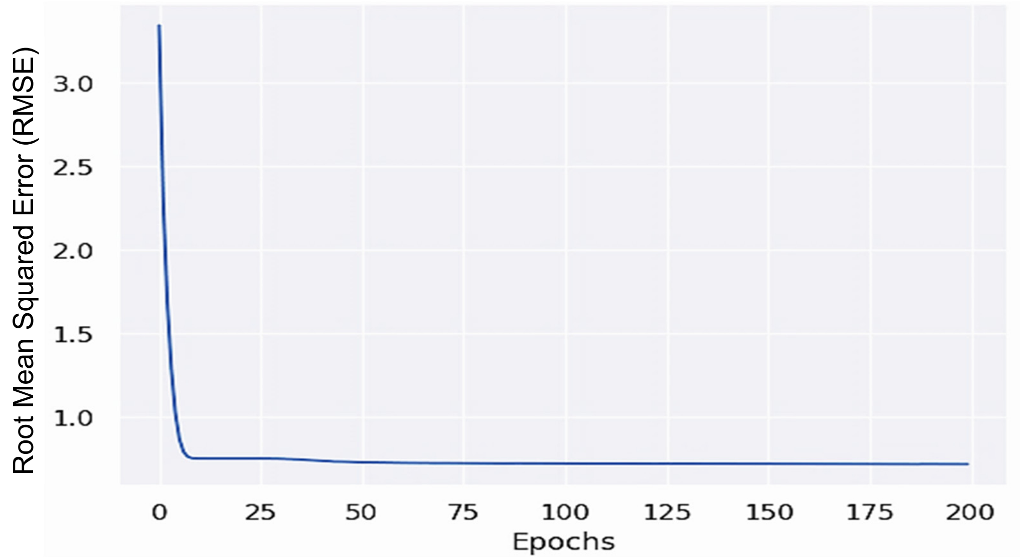

Figure 10. ETL training loss (RMSE).

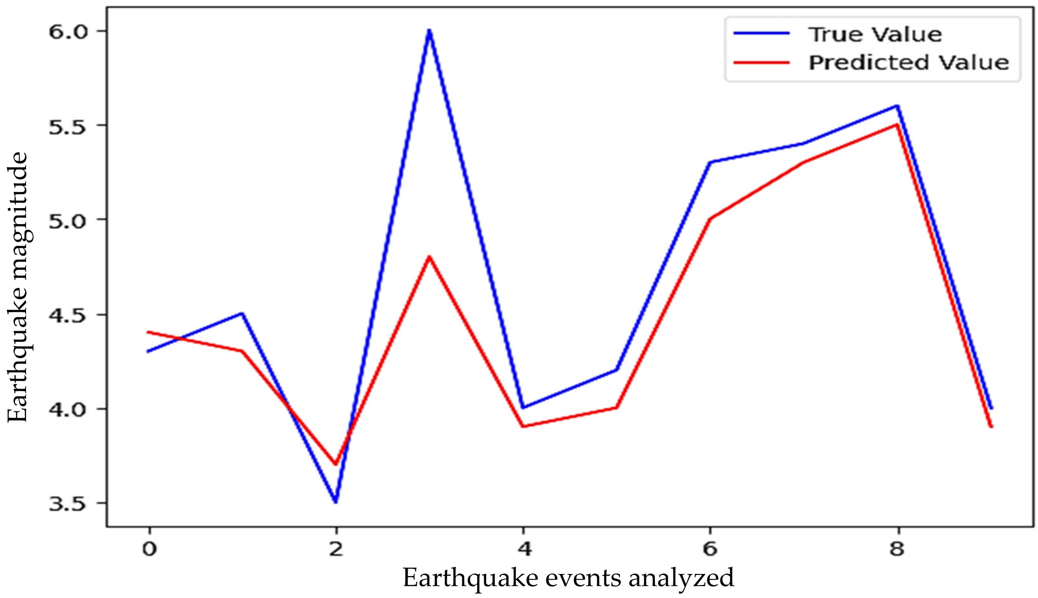

Figure 11. Examples of predicted and actual magnitude values for occurred earthquakes.

6. Discussion

There are many complexities involved in earthquakes: the magnitude, depth, location, and other seismological features which generally depend on the region's tectonic characteristics and geology. Because of this, earthquake detection systems are crucial to seismology, and learning models should be trained using powerful algorithms. For this reason, the study should focus on a specific region, and a dense seismic network is needed to increase the model accuracy. One option for overcoming these difficulties is to introduce new methods that perform better on the given seismic datasets.

This research article proposes a novel deep network, Earth–Transformer–LSTM (ETL), to estimate earthquake magnitude with high precision. The first layer of the model includes

Transformer encoders to extract subtle features from dataset; LSTM and one-dimensional convolution have been used as the other layers. The model has been trained on an earthquake dataset downloaded from the USGS for Turkey's earthquakes. This country is an active seismic region and has experienced very strong earthquakes in recent years.

The existing literature on earthquake magnitude estimation and seismic parameter prediction has explored various ML and DL techniques. Traditional statistical models and SVM-based approaches focus on feature-engineered catalog attributes but struggle to generalize nonlinear spatio-temporal relationships. RNN-based models, particularly LSTM and GRU networks, have demonstrated some success in capturing temporal dependencies; specifically, these architectures are effective at learning sequential patterns but can face limitations when modeling long-range spatial dependencies among seismic event locations. Hybrid models that combine CNNs with recurrent units (e.g., CNN-LSTM) have been proposed to integrate feature extraction with temporal learning; however, CNN layers primarily capture local patterns and require large, labeled waveform datasets for optimal performance.

Unlike prior studies, the proposed ETL model integrates self-attention mechanisms from the Transformer architecture with sequential gating from LSTM units. This architectural combination allows the ETL framework to jointly model spatial and temporal structures more effectively than either component alone. In contrast, standalone LSTM/GRU models lack explicit mechanisms to capture spatial interactions beyond the sequential axis, whereas CNN-based hybrids do not integrate learned long-range dependencies with the same level of flexibility as Transformers. This design introduces two key advantages:

- ✓ Global Spatial Attention: The Transformer encoder enables direct modeling of relationships among events through self-attention, allowing the model to learn how each event's location influences magnitude prediction without relying only on local sequential context.
- ✓ Temporal Memory Integration: The LSTM units preserve long-term temporal dependencies, which are crucial in earthquake catalogs where event occurrence patterns may reflect clustering, aftershock sequences, and periodicity.

The novelty of this research lies in combining two distinct sequence-based models for earthquake magnitude detection, which we have named ETL (Earth–Transformer–LSTM). The ETL has two components: the Transformer encoder and the LSTM. The Transformer encoder can extract many hidden features from its inputs by utilizing an attention mechanism (multi-head attention) and convolutional layers, which are then used in the feed-forward section. Moreover, the encoded features are transferred to the next RNN-based LSTM model, which can process these encrypted features across multiple sequences, thereby improving the conventional Transformer encoder results and making the ETL suitable for predicting future earthquake magnitudes. Thus, our model obtains RMSE and MSE metrics of 0.7 and 0.49, respectively. Moreover, we split the main dataset into two subsets: training and evaluation. For the training phase, we used the Adam optimizer (learning rate = 1×10^{-4}) and the cross-entropy loss function. Finally, our model's earthquake detection results are more accurate than those of other RNN-based models, such as GRU and CNN-LSTM.

A detailed comparative analysis has been conducted against several representative benchmark models, including standalone Recurrent Neural Networks (LSTM and GRU), hybrid CNN-LSTM architectures, and the recently proposed SN-RNN model [33]. While RNN-based approaches such as LSTM and GRU are effective at modeling sequential dependencies, their performance is limited by their ability to extract complex global relationships from high-dimensional seismic data. Similarly, CNN-LSTM models rely heavily on convolutional feature extraction, which may not fully capture long-range dependencies inherent in seismic time-series signals. The SN-RNN model improves upon traditional RNNs by incorporating noise-robust mechanisms [33]; however, its reported RMSE of 3.16 remains significantly higher than that achieved by the proposed ETL framework.

This result demonstrates that the ETL algorithm can estimate earthquake magnitude much more precisely than other methods that have used similar datasets. Currently, the model does not incorporate explicit information about the earthquake's location or time in its process. Indeed, longitude and latitude are partitions of our inputs, and the proposed model processes the data sequentially (in time). The primary goal of the model is to estimate earthquake magnitude based on available seismic data features. However, we acknowledge that incorporating location and time information could enhance the model's predictive capabilities and provide a more comprehensive analysis. We plan to explore the inclusion of spatial and temporal features in future iterations of the model to address these aspects. As a potential research topic, it may be helpful to explore using these time-series-based algorithms to adapt and improve earthquake detection across different regions and datasets.

Despite the promising performance of the proposed Earth–Transformer–LSTM (ETL) model, several limitations should be acknowledged. Firstly, the model has been trained and evaluated using a region-specific dataset (Turkey, 2000–2018), which may limit its generalizability to other seismic regions with different tectonic characteristics and seismic patterns. Secondly, the input features are restricted to longitude, latitude, and depth, without incorporating additional geophysical, temporal, or waveform-based parameters that may further enhance predictive capability. Thirdly, the dataset size (6679 events) is relatively small for deep learning applications, and the absence of cross-regional validation or external testing may affect the model's robustness assessment. Furthermore, although the hybrid Transformer-LSTM architecture improves accuracy, it increases architectural complexity and computational cost compared to simpler RNN-based models. Finally, the study focuses solely on magnitude estimation and does not address related tasks, such as real-time detection, phase picking, or integration with early warning systems.

The dataset's time window (2000–2018) was selected to ensure catalog completeness, magnitude homogeneity, and data stability, as more recent major seismic sequences are subject to ongoing revisions and reprocessing (including magnitude recalibration, relocation updates, and aftershock reclassification). Since, therefore, earthquake catalogs are frequently updated after large events, restricting the analysis to a consolidated, quality-controlled period enhances the methodological consistency and reproducibility of the prediction results from ML and DL models.

Another reason to limit the dataset to the end of 2018 is to facilitate a direct comparison with similar studies already published and related to the same period. This timeframe allows us to benchmark our method against existing research that utilizes datasets from that era. In this context, the primary objective of this study is to introduce and evaluate a novel hybrid deep learning architecture under controlled and stable catalog conditions, rather than to provide a real-time operational forecasting system. Establishing architectural robustness on a consolidated dataset is a necessary first step before extending the framework to continuously evolving catalogs. Obviously, future work will extend the ETL framework to incorporate post-2018 seismic sequences to further assess temporal generalization under evolving tectonic conditions.

7. Conclusions

This paper presents an innovative deep learning framework, which we named Earth–Transformer–LSTM (ETL), for accurate earthquake magnitude detection in seismically active areas. The suggested method effectively captured complex temporal dependencies in seismic data by using Transformer encoders for deep feature extraction, LSTM networks for sequential modeling, and an MLP-based feed-forward architecture instead of traditional convolutional layers. Experiments conducted on the Turkey earthquake dataset demonstrated that the ETL model enhances the prediction accuracy, yielding RMSE and MSE values of 0.7 and 0.49, respectively. The model was trained using the Adam optimizer on separated training and evaluation datasets and outperforms existing RNN-based approaches such as GRU and CNN-LSTM.

Future studies can extend this research in several important directions. Firstly, incorporating additional spatial–temporal features (e.g., event time, inter-event intervals, fault proximity, tectonic indicators) may improve the model's predictive performance and physical interpretability. Secondly, evaluating the ETL framework on multi-regional or global datasets would provide insight into its generalization capability across different seismic environments. Transfer learning or domain adaptation techniques could also be explored to adapt the model to regions with limited seismic data. Future work may also investigate lightweight or optimized versions of the architecture to reduce computational complexity for real-time earthquake early warning (EEW) applications. Finally, uncertainty quantification methods and explainable AI (XAI) techniques could be incorporated to improve model transparency and reliability for practical seismological deployment.

Author Contributions: Conceptualization, A.A.G., E.G., M.M., and P.V.; methodology, A.A.G. and P.V.; validation, A.A.G. and P.V.; formal analysis, E.G. and M.M.; investigation, A.A.G., E.G., and P.V.; software, E.G. and M.M.; resources, M.M.; data curation, M.M.; writing—original draft preparation, A.A.G., E.G., and M.M.; writing—review and editing, A.A.G., E.G., M.M., and P.V.; supervision, A.A.G. and P.V.; funding acquisition, A.A.G. and P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: This research is supported by the research grant of the University of Tabriz [Number 1469-78].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rojas, O.; Otero, B.; Alvarado, L.; Mus, S.; Tous, R. Artificial neural networks as emerging tools for earthquake detection. *Comput. Y Sist.* **2019**, *23*, 335–350. [CrossRef]
2. Mehta, D.; Das, P.P.; Ghosh, S.; Mishra, S.; Alkhayyat, A.; Sharma, V. A Normalized ANN Model for Earthquake Estimation. In Proceedings of the 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 4–6 May 2023; pp. 151–155. [CrossRef]
3. Ridzwan, N.S.M.; Yusoff, S.H.M. Machine learning for earthquake prediction: A review (2017–2021). *Earth Sci. Inform.* **2023**, *16*, 1133–1149. [CrossRef]
4. Kavianpour, P.; Kavianpour, M.; Jahani, E.; Ramezani, A. A CNN-BiLSTM model with attention mechanism for earthquake prediction. *J. Supercomput.* **2023**, *79*, 19194–19226. [CrossRef]
5. Sadhukhan, B.; Chakraborty, S.; Mukherjee, S. Predicting the magnitude of an impending earthquake using deep learning techniques. *Earth Sci. Inform.* **2023**, *16*, 803–823. [CrossRef]
6. Doğan, A.; Demir, E. Structural recurrent neural network models for earthquake prediction. *Neural Comput. Appl.* **2022**, *34*, 11049–11062. [CrossRef]
7. Berhich, A.; Belouadha, F.-Z.; Kabbaj, M.I. An attention-based LSTM network for large earthquake prediction. *Soil Dyn. Earthq. Eng.* **2023**, *165*, 107663. [CrossRef]

8. Saad, O.M.; Chen, Y.; Siervo, D.; Zhang, F.; Savvaidis, A.; Huang, G.-C.D.; Igonin, N.; Fomel, S.; Chen, Y. EQCCT: A production-ready earthquake detection and phase picking method using the compact convolutional transformer. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4507015. [CrossRef]
9. Aslam, B.; Zafar, A.; Khalil, U.; Azam, U. Seismic activity prediction of the northern part of Pakistan from novel machine learning technique. *J. Seismol.* **2021**, *25*, 639–652. [CrossRef]
10. Zhou, W.; Dong, H.; Liang, Y. The deterministic dendritic cell algorithm with Haskell in earthquake magnitude prediction. *Earth Sci. Inform.* **2020**, *13*, 447–457. [CrossRef]
11. Majhi, S.K.; Hossain, S.S.; Padhi, T. MFOFLANN: Moth flame optimized functional link artificial neural network for prediction of earthquake magnitude. *Evol. Syst.* **2020**, *11*, 45–63. [CrossRef]
12. Zhang, H.; Melgar, D.; Sahakian, V.; Searcy, J.; Lin, J.-T. Learning source, path and site effects: CNN-based on-site intensity prediction for earthquake early warning. *Geophys. J. Int.* **2022**, *231*, 2186–2204. [CrossRef]
13. Zhu, J.; Li, S.; Wei, Y.; Song, J. On-site instrumental seismic intensity prediction for China via recurrent neural network and transfer learning. *J. Asian Earth Sci.* **2023**, *248*, 105610. [CrossRef]
14. Kudłacik, I.; Kapłon, J.; Kazmierski, K.; Fortunato, M.; Crespi, M. First feasibility demonstration of GNSS-seismology for anthropogenic earthquakes detection. *Sci. Rep.* **2023**, *13*, 20905. [CrossRef] [PubMed]
15. Abdalzaher, M.S.; Krichen, M.; Fouda, M.M. Enhancing earthquakes and quarry blasts discrimination using machine learning based on three seismic parameters. *Ain Shams Eng. J.* **2024**, *15*, 102925. [CrossRef]
16. Bilal, M.A.; Ji, Y.; Wang, Y.; Akhter, M.P.; Yaqub, M. An early warning system for earthquake prediction from seismic data using Batch Normalized Graph Convolutional Neural Network with Attention Mechanism (BNGCNNATT). *Sensors* **2022**, *22*, 6482. [CrossRef]
17. Saad, O.M.; Chen, Y.; Savvaidis, A.; Chen, W.; Zhang, F.; Chen, Y. Unsupervised deep learning for single-channel earthquake data denoising and its applications in event detection and fully automatic location. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5921310. [CrossRef]
18. Han, J.; Kim, S.; Sheen, D.H.; Lee, D.; Lee, S.J.; Yoo, S.H.; Park, D. Seismic event and phase detection using deep learning for the 2016 Gyeongju earthquake sequence. *Geosci. J.* **2023**, *27*, 285–295. [CrossRef]
19. Abebe, E.; Kebede, H.; Kevin, M.; Demissie, Z. Earthquakes magnitude prediction using deep learning for the Horn of Africa. *Soil Dyn. Earthq. Eng.* **2023**, *170*, 107913. [CrossRef]
20. Laurenti, L.; Tinti, E.; Galasso, F.; Franco, L.; Marone, C. Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth Planet. Sci. Lett.* **2022**, *598*, 117825. [CrossRef]
21. Mohammadigheymasi, H.; Shi, P.; Tavakolizadeh, N.; Xiao, Z.; Mousavi, S.M.; Matias, L.; Fernandes, R. IPIML: A deep-scan earthquake detection and location workflow Integrating Pair-Input deep learning model and Migration Location method. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5914109. [CrossRef]
22. Yuan, C.; Ni, Y.; Lin, Y.; Denolle, M. Better Together: Ensemble Learning for Earthquake Detection and Phase Picking. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5920217. [CrossRef]
23. Mousavi, S.M.; Ellsworth, W.L.; Zhu, W.; Chuang, L.Y.; Beroza, G.C. Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* **2020**, *11*, 3952. [CrossRef] [PubMed]
24. Kartal, R.F.; Kadirioğlu, F.T.; Kılıç, T. An improved earthquake and moment tensor catalogues (M $\geq$ 4.0) for Anatolian region and its near vicinity (2013–2019). *Arab. J. Geosci.* **2022**, *15*, 1632. [CrossRef]
25. Güvercin, S.E.; Karabulut, H.; Konca, A.Ö.; Doğan, U.; Ergintav, S. Active seismotectonics of the East Anatolian Fault. *Geophys. J. Int.* **2022**, *230*, 50–69. [CrossRef]
26. Bohnhoff, M.; Martínez-Garzón, P.; Bulut, F.; Stierle, E.; Ben-Zion, Y. Maximum earthquake magnitudes along different sections of the North Anatolian fault zone. *Tectonophysics* **2016**, *674*, 147–165. [CrossRef]
27. Över, S.; Demirci, A.; Özden, S. Tectonic implications of the February 2023 Earthquakes (Mw7.7, 7.6 and 6.3) in south-eastern Türkiye. *Tectonophysics* **2023**, *866*, 230058. [CrossRef]
28. Ganas, A.; Elias, P.; Briole, P.; Valkaniotis, S.; Escartin, J.; Tsironi, V.; Karasante, I.; Kosma, C. Co-seismic and post-seismic deformation, field observations and fault model of the 30 October 2020 Mw = 7.0 Samos earthquake, Aegean Sea. *Acta Geophys.* **2021**, *69*, 999–1024. [CrossRef]
29. Türker, E.; Yen, M.; Pilz, M.; Cotton, F. Significance of Pulse-Like Ground Motions and Directivity Effects in Moderate Earthquakes: The Example of the Mw 6.1 Gölyaka–Düzce Earthquake on 23 November 2022. *Bull. Seismol. Soc. Am.* **2024**, *114*, 955–964. [CrossRef]
30. Emre, Ö.; Duman, T.Y.; Özalp, S.; Şaroğlu, F.; Olgun, Ş.; Elmacı, H.; Çan, T. Active fault database of Turkey. *Bull. Earthq. Eng.* **2018**, *16*, 3229–3275. [CrossRef]
31. Available online: https://earthquake.usgs.gov/ (accessed on 5 May 2025).
32. Wiemer, S. A software package to analyze seismicity: ZMAP. *Seismol. Res. Lett.* **2001**, *72*, 373–382. [CrossRef]

33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. Advances in Neural Information Processing Systems, Proceeding of the Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2017; Volume 30, pp. 5998–6008.
34. Bilal, M.A.; Wang, Y.; Ji, Y.; Pervez Akhter, M.; Liu, H. Earthquake detection using stacked normalized recurrent neural network (SNRNN). *Appl. Sci.* **2023**, *13*, 8121. [CrossRef]