ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

Earthquake Prediction and Early Warning System Using Hybrid Random Forest and Time-Series Transformer Deep Learning Algorithms

S. Ghanerezgh [1], M. Rabiei [2,*]

[1] M.Sc. Student in Information Technology, Department of Computer Engineering, Faculty of Electrical, Computer and Mechanical Engineering, University of Eyvanekey, Semnan, Iran
[2] Assistant Professor, Department of Computer Engineering, Faculty of Electrical, Computer and Mechanical Engineering, University of Eyvanekey, Semnan, Iran

ARTICLE INFO

Article History:
Received 27 December 2025
Received in revised form 17 January 2026
Accepted 2 March 2026
Available online 9 March 2026

Keywords:
Earthquake prediction, Deep learning, Random Forest Regression, Time-series analysis, Transformer models, Disaster risk management.

ABSTRACT

Earthquakes pose a catastrophic threat to human life and critical infrastructure, necessitating the development of highly accurate early warning systems. The primary objective of this research is to design an earthquake early warning framework capable of predicting both seismic magnitude and occurrence timing by integrating machine learning and deep learning architectures. In this study, a Random Forest Regression model is employed to estimate earthquake magnitudes, while a time-series Transformer algorithm is utilized to forecast occurrence times. The underlying dataset, sourced from seismic networks in Japan, underwent comprehensive preprocessing, normalization, and train-test partitioning. To predict timing, data was converted into a sequential time-series format before feeding it into the Transformer network. The designed Transformer model incorporates multi-head self-attention mechanisms, layer normalization, and aggregation modules, significantly enhancing its temporal forecasting capacity. Operationally, the Random Forest algorithm first identifies seismic events exceeding a magnitude of 5.0, after which the time-series Transformer predicts the exact occurrence window down to the minute. By combining these predictive outputs, the system dynamically generates visual warning alerts based on the forecasted severity and timeline. Empirical evaluations demonstrate strong predictive proficiency. The magnitude estimation model achieved a Mean Squared Error (MSE) of 0.0261, a Mean Absolute Error (MAE) of 0.0883, a Root Mean Squared Error (RMSE) of 0.1615, and a Coefficient of Determination (R^2) of 0.7737. For earthquake timing forecasting, the system yielded an MSE of 0.0007, an MAE of 0.0265, and an RMSE of 0.0264. Implementing this hybrid early warning system offers a vital tool for effective disaster risk management and mitigation strategies in seismically active regions.

[*] Corresponding Author: Mohammad.rabiei@eyc.ac.ir
Assistant Professor, Department of Computer Engineering, Faculty of Electrical, Computer and Mechanical Engineering, University of Eyvanekey, Semnan, Iran

http://dx.doi.org/10.47176/TMI.2026.244946

1. INTRODUCTION

Seismic events represent some of the most catastrophic natural hazards globally, imposing severe threats to human populations and structural infrastructures. Within this context, Earthquake Early Warning (EEW) frameworks play a pivotal role by precisely quantifying seismic magnitudes and generating rapid alert signals. Leveraging deep learning (DL) architectures for real-time detection has emerged as a transformative research paradigm with the capacity to drastically mitigate seismic-induced losses, specifically through the automated estimation of earthquake magnitude, focal depth, and hypocenter localization [1]. Concurrently, recent epochs have witnessed profound breakthroughs in machine learning (ML) and deep learning. As a core subset of artificial intelligence, ML offers robust mathematical tools highly applicable to seismological data processing. DL networks, in particular, demonstrate exceptional proficiency in extracting intertwined spatial and temporal features from complex seismographic waveforms, thereby offering profound insights into seismic source dynamics and accelerating early detection pipelines.

Because tectonic ruptures manifest abruptly, maximizing the operational response window is paramount. Rapid classification algorithms can provide critical lead times ranging from several seconds to minutes prior to the arrival of destructive shear waves, optimizing emergency preparedness and casualty reduction [2]. Driven by this imperative, the core objective of this study is the development of a hybridized EEW framework. The architecture sequentially orchestrates a machine learning-based Random Forest Regression (RFR) model to identify and predict significant seismic events exceeding a magnitude threshold of 5.0, followed by a deep learning-based Time-Series Transformer network to forecast the precise temporal occurrence of the event. The integration of these two distinctive models culminates in a unified, highly responsive emergency alert infrastructure.

2. LITERATURE REVIEW AND TECHNICAL FOUNDATIONS

Within automated alert systems, decoupling the magnitude characteristics from raw waveforms constitutes the primary computational stage [3]. Random Forest Regression, an ensemble machine learning paradigm, excels at mapping non-linear seismic signatures to continuous magnitude outputs [4]. For example, a comparative analysis by Apriani & Wijaya (2021) demonstrated that RFR architectures exhibit superior estimation accuracy when benchmarked against traditional artificial neural networks [5]. To advance wave-packet processing, Münchmeyer et al. introduced the Transformer-based Earthquake Magnitude (TEAM) model, achieving substantial performance gains by handling dynamically varying station densities. Utilizing diverse seismic catalogs from northern Chile, Italy, and Japan, their framework successfully leveraged transfer learning techniques to minimize data variance, validating model efficacy through Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2) [6].

Similarly, deep architectures have been widely adapted for multi-parameter classification. Saad et al. (2021) developed a Convolutional Neural Network (CNN) tailored for Japan's EEW system, extracting high-dimensional feature maps from short-duration waveforms. By processing an 8-second waveform window from three distinct stations, their CNN effectively classified origin time (89.55%), focal depth (92.54%), epicentral location (89.50%), and magnitude (93.67%), surpassing conventional baseline models [7]. In a related development for the Indian subcontinent, Mukherjee et al. (2021) engineered a hybrid CNN-LSTM framework optimized for regional and local seismic catalogs, demonstrating that a specialized CNN achieved 91.10% accuracy, while the recurrent LSTM block achieved 93.99% accuracy, balancing computational throughput with predictive speed [8]. Moving toward vision-based sequence modeling, Saad et al. (2022) deployed a Vision Transformer (ViT) on California seismic datasets, leveraging a 30-second, three-component seismogram to categorize magnitudes in real time while drastically minimizing MAE compared to standard neural topologies [9].

Temporal and spatiotemporal sequence modeling has also seen significant architectural evolution. Zhang & Wang (2023) developed the Earthquake Prediction Transformer (EPT) model using historical catalogs from the China Earthquake Network Center. By implementing a 70:30 train-test partition that boosted structural accuracy by 50%, their multi-directional, multi-head self-attention mechanism effectively resolved long-range temporal dependencies. When benchmarked against Support Vector Machines (SVM), Bidirectional LSTMs (BiLSTM), and Random Forests, the EPT model established superior metrics, yielding a 90% accuracy rate, 86% recall, and a 94% Area

Under the ROC Curve (AUC-ROC) [10]. Addressing specific geographical contexts, Abebe et al. (2023) formulated a multivariate time-series regression task to forecast seismic events (\ge M3.0) in the Horn of Africa across a three-month horizon. Their empirical evaluations using MAE, MSE, and RMSE confirmed that the self-attention Transformer framework significantly outclassed standalone LSTM and BiLSTM alternatives [11].

Furthermore, environmental correlations have begun merging with seismic analytics; Sadhukhan et al. (2023) integrated a hybrid Transformer network with recursive mathematical formulations to map global temperature variations alongside calculated seismic variables across catalogs from Indonesia, Japan, and the Indian Himalayas. Their findings indicated that the Transformer variant achieved optimal predictive precision, yielding an MAE of 0.083 and an MSE of 0.015 [12]. Finally, addressing localized spatiotemporal dependencies, Putran (2024) designed a unified CNN-GRU model validated on multi-national datasets from Morocco, Turkey, and Japan. This hybrid configuration allowed for precise early warning windows and superior occurrence-time estimation, noticeably reducing error metrics (MAE, RMSE) and raising the R^2 index relative to traditional standalone algorithms [13]. A comprehensive synthesis of these structural methodologies and their alignment with the current study is contextualized in Table 1.

Table. 1. Related research background

Model accuracy	Evaluation technique	Model	dataset	Research title	Researcher's name - year
93.67%	Accuracy	CNN	Japan	Deep learning approach in earthquake early warning system	Saad (2021)[7]
93.99%	Accuracy	CNN LSTM	india	Earthquake early warning system using deep learning	Mukherjee (2021)[8]
0.42 0.13 0.11	MAE RMSE R^2	TEAM	northern Chile, Japan, and Italy	Predicting the magnitude and location of an earthquake with a transformer	**Münchmeyer** (2021)[6]
0.120	MAE	Transformer (VIT)	china	Detecting earthquake magnitude and location with the VIT model	Saad (2022)[9]
0.083 0.015	MAE MSE	LSTM+BiLSTM Ertransform	Japan and Indonesia, and the Indian Himalayas and the Karakoram Kush	Deep learning model based on climatic and seismic data.	Sadhukhan (2023) [12]
0.271 0.376 0.142	MAE RMSE MSE	BiLSTM+AM+transformer	Horn of Africa	Earthquake magnitude prediction with deep learning.	Abebe (2023)[11]
90%	Accuracy	Transformer(EPT)	China	Earthquake prediction with EPT	Zhang (2023) [10]
98.67%	MAE RMSE R^2	CNN-GRU	Morocco, Japan, and Turkey	Spatio-temporal analysis with a hybrid model CNN-GRU	Puthran (2024) [13]

2.1. Research Motivations and Contributions

Unlike prior literature that predominantly addresses seismic parameter prediction (e.g., location, magnitude, depth, and origin time) in isolation or relies on recursive hybrid algorithms, this study introduces an integrated visual Earthquake Early Warning (EEW) system designed for destructive events (\ge M5.0). The proposed architecture uniquely couples a machine learning-based Random Forest Regression (RFR) model for magnitude estimation with a deep learning-based Time-Series Transformer network for temporal forecasting. The deployment of the time-series Transformer is specifically motivated by its superior capacity to model long-range temporal dependencies, facilitate parallelized computational processing, leverage multi-head self-attention mechanisms, and provide exceptional generalizability and architectural flexibility. To guarantee rigorous error analysis, optimal model convergence, structural change analysis, and reliable predictive outputs, this study evaluates the framework using three standardized metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

3. RESEARCH METHODOLOGY

3.1. Computational Environment and Software Frameworks

The end-to-end implementation, architectural training, and evaluation pipelines of the proposed framework were executed within the Google Colab cloud computing environment. The core predictive algorithms were developed in the Python programming language, leveraging the computational efficiencies of TensorFlow and Keras as the primary deep learning backends.

3.2. Data Acquisition and Dataset Characteristics

The empirical seismic catalog utilized in this research was sourced from the United States Geological Survey (USGS) database, capturing seismic historical records from Japan. The raw dataset comprises 6,268 records compiled in a comma-separated values (CSV) format, with recorded earthquake magnitudes ranging continuously from M4.0 to M7.1. Each distinct row signifies a single, verified seismic event.

Initially, the raw catalog contained 22 descriptive attributes, including: event timestamp, latitude, longitude, focal depth, seismic magnitude, magnitude calculation algorithm, total number of reporting seismological stations, maximum angular gap between monitoring stations, horizontal distance to the epicenter, root mean square (RMS) of the residual travel time (seconds), data contributor identifier, unique event ID, metadata update timestamp, geographic area description, event type, depth error, location error, magnitude error, human review status, reported event location, station count for magnitude estimation, and the specific magnitude reporting grid.

3.3. Data Preprocessing and Feature Selection

To enhance computational efficiency and focus specifically on high-impact seismic events, the dataset was filtered to include only earthquakes with a magnitude greater than or equal to 5.0 (\ge M5.0). Initial data refinement and scrubbing were executed via PowerQuery to eliminate null values, irrelevant inputs, and anomalous data points [14]. Following this initial cleaning phase, data formatting and attribute transformations were finalized using a combination of PowerQuery and the Weka data mining suite.

To mitigate the curse of dimensionality and optimize computational overhead, feature selection was performed within the Weka software environment utilizing the InfoGain AttributeEval methodology [15]. Through this information-gain heuristic, the initial 22 attributes were systematically reduced to 11 core features that exhibit the highest predictive power. The optimized feature space retains the following parameters:

1. Earthquake Timestamp
2. Latitude
3. Longitude
4. Focal Depth
5. Seismic Magnitude
6. Number of Reporting Seismological Stations
7. Maximum Angular Gap Among Monitoring Stations
8. Root Mean Square (RMS) of Residual Travel Time
9. Depth Error
10. Magnitude Error
11. Total Station Count for Magnitude Calculation

3.4. Data Normalization and Partitioning

To stabilize gradient descent during the deep learning phase and alleviate disparate feature scales, the optimized attributes were scaled using the MinMaxScaler function from the Scikit-learn library. This normalization mapping bounds the input feature values within a uniform range of [0, 1], thereby accelerating model convergence.

Following feature scaling, the processed dataset was partitioned into distinct training and testing subsets using an 80:20 ratio, ensuring an independent experimental group for robust validation. The structural partitioning was implemented programmatically via the train_test_split function within the Scikit-learn environment. Finally, the magnitude estimation task was operationalized by executing a Random Forest Regression architecture using the sklearn.ensemble module and the RandomForestRegressor class. The sequential progression of the preprocessing, feature engineering, and modeling phases is systematically cataloged in Table 2.

Table 2. Systematic Data Preprocessing Pipeline

Preprocessing Stage	Operational Procedures and Technical Specifications
Data Filtering	• Isolated seismic records with a magnitude threshold of \ge M5.0 to focus exclusively on high-impact earthquakes.
Data Cleaning	• Utilized **Power Query** to systematically detect and remove duplicate rows, handle missing values (nulls), eliminate invalid/erroneous inputs, and drop structural outliers.
Feature Transformation & Casting	• Standardized attribute data types within Power Query using the following precise schema: - Time \rightarrow **Date/Time** - Latitude, Longitude, Depth, Mag, rms, depthError, magError \rightarrow **Decimal Number** - gap, nst, magNst \rightarrow **Whole Number**
Feature Selection	• Conducted dimensionality reduction via the **Weka** data mining suite. • Implemented the **InfoGain AttributeEval** heuristic combined with a Ranker search method to extract the 11 most predictive core attributes out of the initial 22 variables.
Data Normalization	• Applied the MinMaxScaler function from Python's **Scikit-learn** library. • Mapped all continuous numerical features uniformly into the bounded scalar range of [0, 1] to optimize gradient descent and algorithm convergence.

3.5. Hyperparameter Optimization for Random Forest Regression

To maximize the predictive accuracy of the Random Forest Regression (RFR) architecture, a systematic hyperparameter tuning pipeline was implemented via the GridSearchCV module from the Scikit-learn library. The grid-search mechanism exhaustively evaluates all predefined parameter permutations across the training partition to isolate the optimal structural configuration. Through this cross-validated optimization process, the number of decision trees (n_estimators) was determined to be 200, while the maximum tree depth (max_depth) was configured to None (unlimited), allowing the constituent estimators to expand until all leaves became pure. Following training and optimization, the regression model's predictive efficacy was rigorously quantified on the independent testing partition using four standardized statistical metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

3.6. Sequential Data Transformation and Time-Series Transformer Architecture

Because Transformer networks are inherently non-sequential, the processed seismic dataset had to be restructured into a sequential time-series format prior to modeling the temporal occurrence of earthquakes. This was achieved by defining a fixed sliding window with a specified number of historical time steps, transforming the continuous tabular data into sequential overlapping temporal blocks.

To precisely forecast the exact timestamp of seismic eventuation, a custom 1D Time-Series Transformer architecture was constructed [16]. The structural topology of the proposed deep learning network is sequentially organized as follows:

- **Input Projection Layer:** Encodes the temporal feature slices into a high-dimensional embedding space.
- **Multi-Head Self-Attention Modules:** Captures long-range, non-linear temporal dependencies and correlations across historical seismic patterns.

- **Layer Normalization:** Stabilizes the internal covariate shift and ensures smooth gradient propagation during backpropagation.
- **Global Average Pooling (GlobalAveragePooling1D):** Flattens the temporal dimensions by calculating the mean vector of each feature map, mitigating spatial dimensionality.
- **Regularization (Dropout Layer):** Dynamically penalizes co-adaptation of features to control overfitting.
- **Fully Connected Topologies (Dense & Output Layers):** Maps the pooled representation to a continuous linear scalar denoting the predicted arrival time of the seismic event.

3.7. Theoretical Foundations of Random Forest Regression

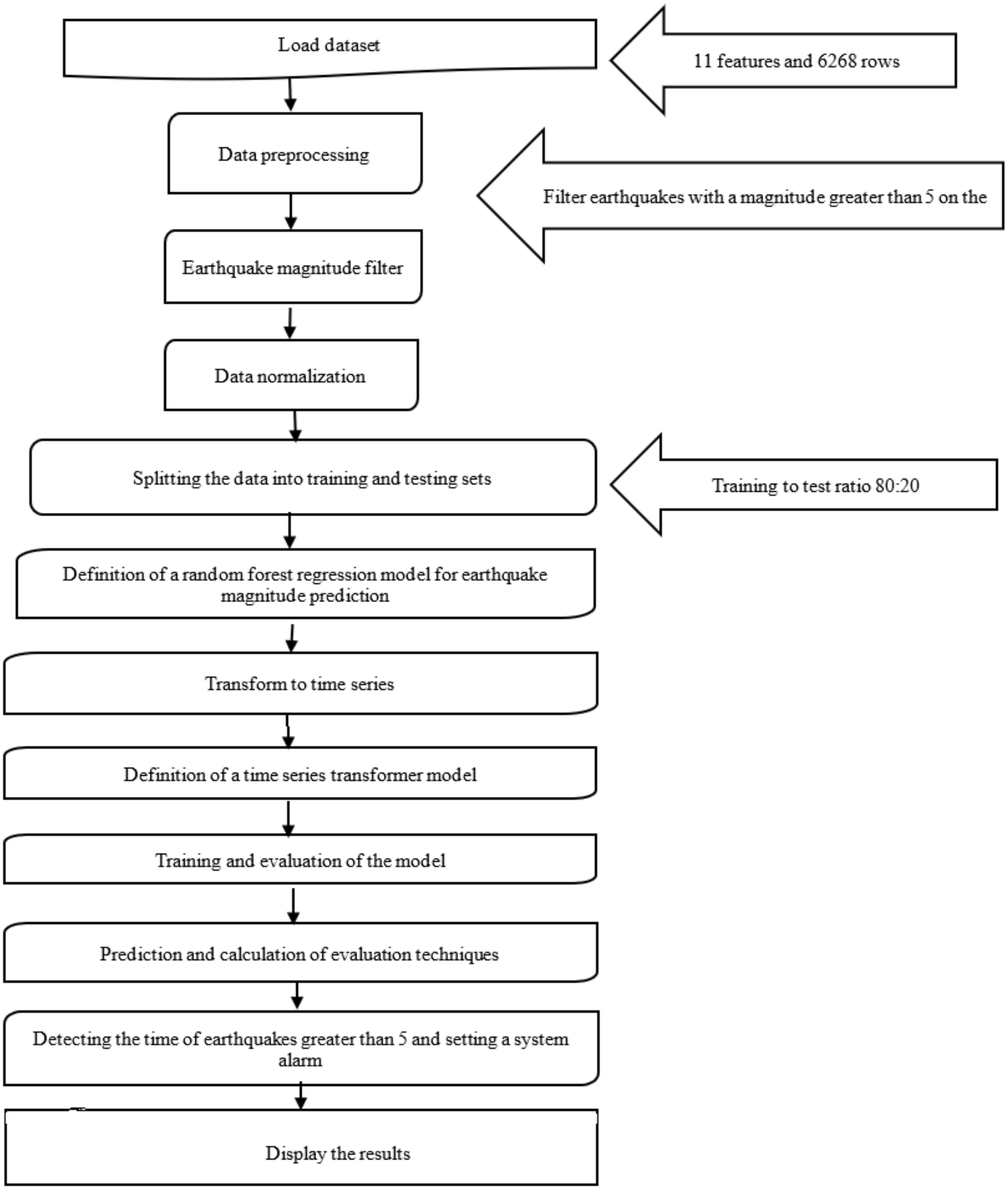

Fig. 1. General state of research methodology

Random Forest Regression operates as an advanced ensemble machine learning paradigm grounded in the principles of bootstrap aggregating (bagging). The algorithm constructs a diverse forest of randomized decision trees during the training phase, combining their individual predictive outputs to yield a robust, continuous final prediction. The primary architectural advantages of this ensemble model include its hierarchical tree structure, inherent resilience against overfitting due to feature sub-sampling, and exceptional performance on noisy, non-linear datasets. Due to these attributes, RFR frameworks have been widely deployed across highly variable prediction tasks, including financial asset forecasting, medical diagnostics, energy demand prediction, and seismological intensity modeling [4]. The holistic, step-by-step procedural workflow of the proposed end-to-end research methodology is visually integrated and conceptualized in Figure 1.

4. STRUCTURAL MECHANICS AND COMPUTATIONAL MODELING OF THE TIME-SERIES TRANSFORMER

4.1. Architectural Advantages Over Recurrent Neural Networks

The Transformer model fundamentally operates on an encoder-decoder architecture driven by non-sequential self-attention mechanisms. The encoder segment maps the input sequence into a continuous high-dimensional vector representation. At the computational core of this topology lies the multi-head attention mechanism, which facilitates the simultaneous and parallelized calculation of attention weights across multiple temporal slices.

When applied to sequential forecasting, time-series Transformers offer profound advantages over classical recurrent network topologies (e.g., RNNs, LSTMs, and GRUs). In recurrent architectures, backpropagation through long sequences frequently causes the error gradients to diminish or grow exponentially, culminating in the well-documented vanishing or exploding gradient problems. Conversely, the self-attention mechanism within Transformer architectures establishes direct mathematical pathways across disparate temporal states. This structure allows gradients to flow directly through the layers without cumulative attenuation, thereby effectively mitigating gradient instability and allowing the model to capture complex long-range dependencies and highly non-linear patterns within seismic sequences. Consequently, historical benchmarks demonstrate that time-series Transformers yield significantly higher predictive accuracy compared to recurrent alternatives [9]. Generally, the functional scope of time-series Transformers spans four primary domains: sequential forecasting, anomaly detection, categorical classification, and discrete event prediction [17].

4.2. Model Operationalization and Training Parameters

The proposed deep learning framework was rigorously trained on the optimized historical training partition. The learning sequence was parameterized by defining a structured batch size and a fixed number of training epochs to ensure adequate loss convergence while preventing overfitting. The network's optimization pipeline was governed by the Adam optimizer due to its adaptive learning rate properties, coupled with a Mean Squared Error (MSE) loss function to mathematically penalize substantial outliers during the temporal regression phase.

Following the training phase, the predictive error and generalizability of the network were quantified using three standardized validation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Once validated, the continuous magnitude estimations derived from the Random Forest Regression model were integrated with the temporal forecasts outputted by the Time-Series Transformer. This unified computational output was then mapped to a rule-based logic structure to dynamically trigger multi-tiered visual alert classifications (alarms) based on the estimated earthquake severity and localized arrival window. Finally, the end-to-end framework was integrated into a real-time monitoring dashboard designed to seamlessly ingest, process, and update incoming seismological stream data. The precise graphical layer sequence and operational topology of the utilized Time-Series Transformer network are visually represented in Figure 2.

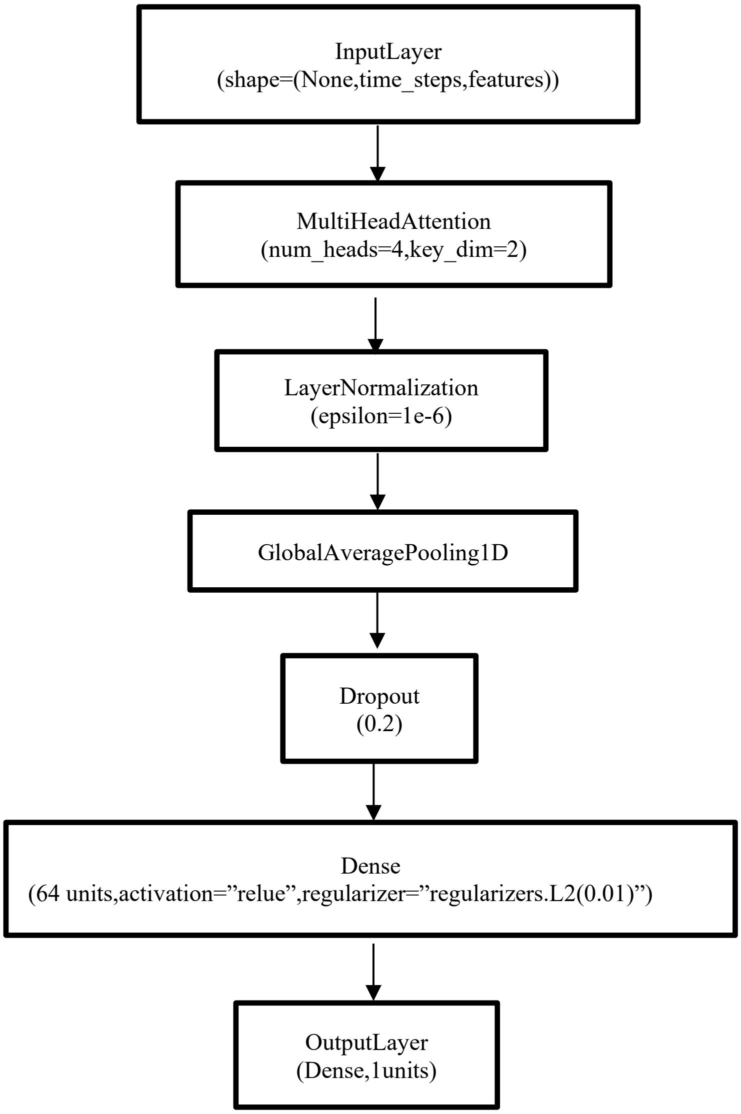

Fig. 2. Structure of the time series transformer model

5. RESULTS AND DISCUSSION

The primary objective of this investigation was to architect a robust, multi-tiered Earthquake Early Warning (EEW) infrastructure by concurrently forecasting seismic temporal occurrence and magnitude. To realize this, the proposed framework achieved high-precision outcomes by shifting from conventional empirical approaches to advanced learning paradigms—specifically utilizing an ensemble Random Forest Regression (RFR) model for continuous magnitude estimation and a deep Time-Series Transformer network for temporal sequencing. By integrating rigorous data normalization techniques with strategic train-test partitioning, the developed models demonstrated exceptional predictive synergy and structural stability.

The empirical findings from this study underscore the transformative role that machine learning and deep learning architectures can play in real-time seismological disaster mitigation. The RFR model accurately quantified seismic magnitudes, serving as a critical threshold-trigger for the downstream visual alarm system. To maximize the operational efficacy of the framework, hyperparameter tuning was conducted to locate the optimal global minimum of the loss surface [18]. Specifically, a comprehensive grid-search approach via GridSearchCV was executed to tune the core hyperparameter bounds of the ensemble estimator. The exact optimized hyperparameter topology utilized to finalize the RFR magnitude prediction model is systematically cataloged in Table 3.

Table 3. Optimized Hyperparameter Configurations for the RFR Model

Hyperparameter	Optimized Operational Value	Functional Specification within the Model
n_estimators	200	Total number of randomized decision trees generated in the ensemble forest.
max_depth	None	Unlimited tree depth, allowing maximum leaf node purity during partitioning.
Loss Function	Mean Squared Error (MSE)	Mathematical criterion used to measure and penalize node splitting errors.

To validate the generalizability of the trained ensemble model and guarantee that the predictive outputs were not an artifact of overfitting, the framework's performance was rigorously evaluated using k-fold cross-validation alongside independent test set verification. The k-fold cross-validation technique is a standard statistical paradigm used to accurately estimate model prediction error, effectively minimizing evaluation variance and preventing structural overfitting.

In this research, since the grid-search optimization pipeline incorporated an internal cross-validation factor of CV = 3, the subsequent validation phase strictly designated k = 3 to maintain computational consistency across the comparative experiments. The complete performance evaluation metrics of the RFR configuration, alongside its cross-validated performance indicators for earthquake magnitude estimation, are systematically detailed in Table 4.

Table 4. Comprehensive Performance Metrics and k-fold Cross-Validation Results for the RFR Model

Model Evaluation Framework	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Coefficient of Determination (R2)
Random Forest Regression (Test Set)	0.0261	0.0883	0.1615	0.7737
k-fold Cross-Validation (k=3)	0.0240	0.0891	0.1538	0.7514

5.1. Comparative Analysis of Magnitude Filtering and Noise Reduction

To comprehensively validate the operational parameters of the predictive framework, this study investigated the mathematical implications of applying a magnitude threshold filter (\ge M5.0) against modeling the raw, unfiltered seismic catalog.

When executing the model without magnitude constraints, the primary theoretical advantage is the preservation of a larger sample size, which expands the data volume available for architectural training. However, low-magnitude seismic records ($M < 5.0$) inherently introduce high levels of high-frequency stochastic noise and descriptive anomalies into the dataset. This variance severe severely distorts the regression surface, ultimately degrading the network's predictive capabilities and compromising its capacity to reliably generalize to catastrophic, higher-magnitude events.

Conversely, enforcing a rigid structural filter restricted to events exceeding magnitude 5.0 dramatically isolates critical seismic features. This filtering approach acts as an effective noise reduction mechanism, filtering out low-amplitude perturbations that do not contribute to destructive earthquake characteristics. Consequently, the convergence accuracy of the framework increases significantly, while structural prediction error decreases.

Empirical distributions comparing the actual and predicted magnitudes under the filtered schema demonstrate excellent alignment and strong model efficacy. This validation proves that threshold filtering is essential for stabilizing high-magnitude early warnings. For comparative reference, the empirical distributions of the actual versus predicted magnitudes utilizing the unfiltered dataset are visualized and analyzed in Figures 3 and 4.

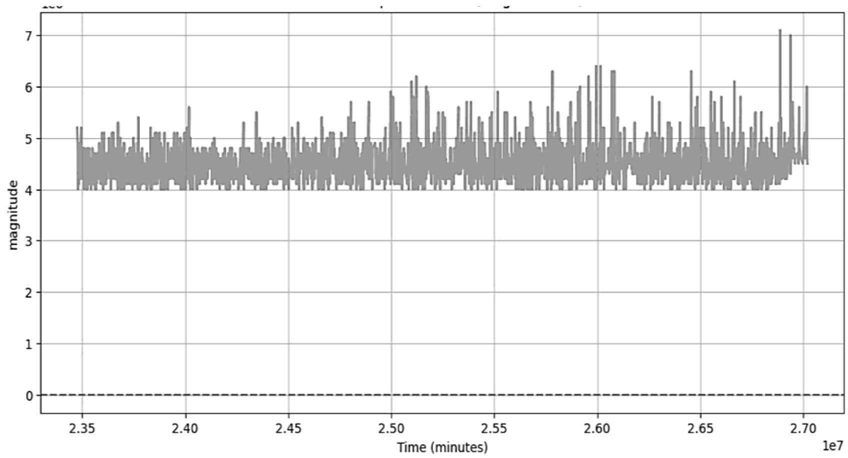

Fig. 3. Magnitude chart of real data without filtering

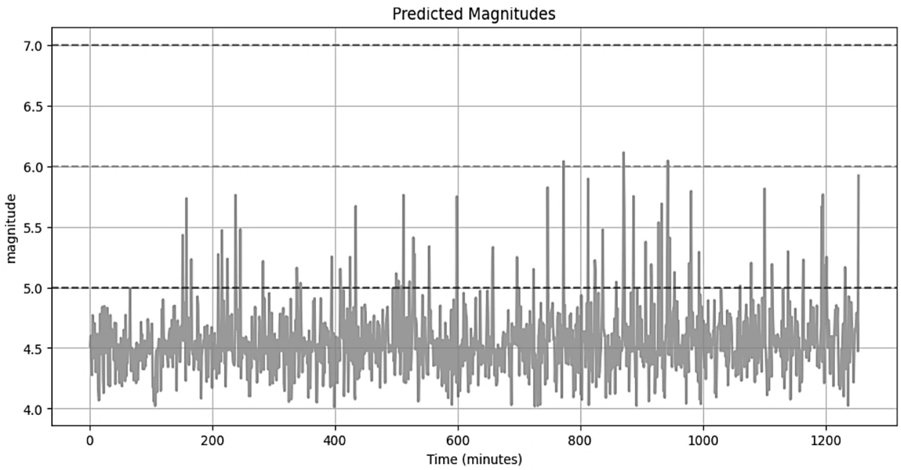

Fig. 4. Magnitude chart of predicted data without filtering

The Random Forest Regression model continuously predicts magnitudes greater than 5, as illustrated in the diagram. For better understanding, the warning system uses a color scheme for magnitudes 5, 6, and 7 Figure 5 shows the actual magnitude diagram, while Figure 6 shows the predicted magnitude diagram.

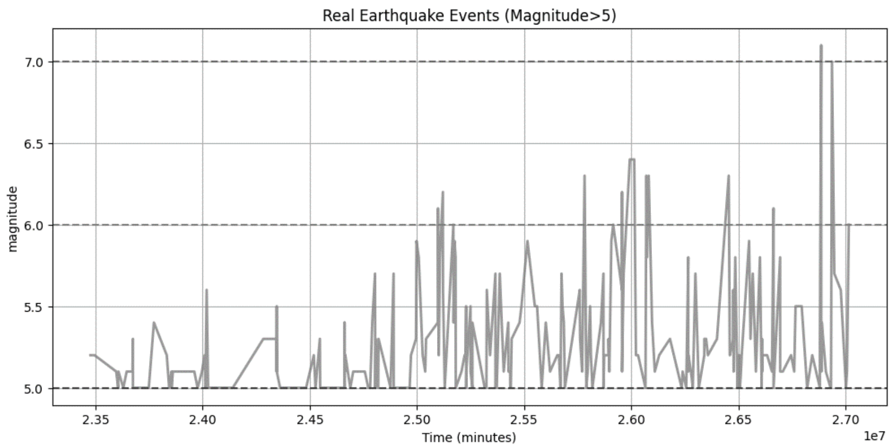

Fig. 5. Magnitude chart of real data with filtering

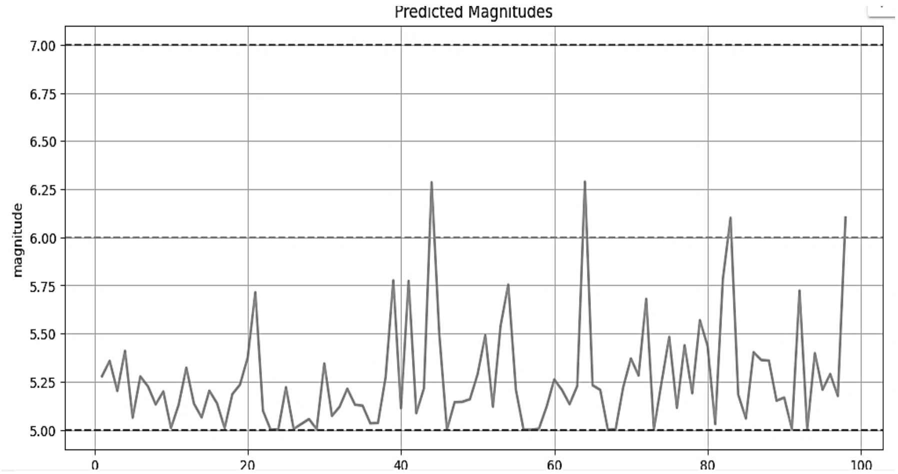

Fig. 6. Magnitude chart of predicted data with filtering

5.2. Hyperparameter Configuration and Optimization of the Time-Series Transformer

For the temporal forecasting pipeline, the window length of the Time-Series Transformer model was systematically configured to 7 time steps. This sequential depth enables the self-attention mechanism to effectively extract temporal dependencies from historical seismic intervals to predict future eventuation timestamps.

To govern the internal dimensionality and enforce structural generalization, a Dropout layer with a regularization rate of 0.2 was integrated immediately following the GlobalAveragePooling1D module, dynamically mitigating the risk of co-adaptation and overfitting during the gradient update phases. Within these dense processing blocks, the Rectified Linear Unit (ReLU) activation function was deployed to ensure robust non-linear mapping while fundamentally preventing the vanishing gradient problem. Furthermore, to penalize excessive weight magnitudes and maintain structural equilibrium across the network topologies, an L_2 weight regularization penalty (ridge regression constraint) was imposed with a factor of 0.01. The stochastic training sequence was operationalized across 100 training epochs, utilizing an optimal mini-batch size of 32 samples per iteration to balance computational throughput with stable loss convergence (epochs=100, batch_size=32). The complete architectural and optimization hyperparameter schema utilized for the proposed Time-Series Transformer network is systematically compiled in Table 5.

Table 5. Complete Structural and Optimization Hyperparameters for the Time-Series Transformer Model

Hyperparameter Component	Configured Operational Value	Functional Role and Optimization Objective
time_steps	7	Sets the sequential window length for analyzing historical seismic patterns.
num_heads	8	Specifies the number of parallel attention heads to capture diverse temporal features.
key_dim	4	Dictates the dimensionality of the query and key vectors within each attention head.
dropout	0.2	Regularization fraction injected to penalize overfitting and boost generalization.
regularizers.l2	0.01	Weight decay penalty applied to constrain structural variance and maintain stability.
Optimizer	Adam	Adaptive stochastic gradient descent algorithm utilized for efficient backpropagation.
Loss Function	Mean Squared Error (MSE)	Mathematical criterion used to compute and minimize temporal regression errors.
Training Epochs	100	Total number of full computational iterations over the training partition.
Mini-Batch Size	32	Number of training instances processed concurrently per gradient update.

5.3. Predictive Performance Evaluation of the Temporal Network

To rigorously verify the capacity of the Time-Series Transformer in forecasting the precise timestamp of seismic eventuation, the trained model was subjected to comprehensive statistical error analysis. The framework's predictive accuracy was validated across multiple operational dimensions using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Furthermore, the convergence properties of the training cycle were evaluated by monitoring the loss trajectory across the training, validation, and independent testing partitions. The comprehensive empirical performance indicators yielded by the deep learning framework are systematically cataloged in Table 6.

Table 6. Multi-Partition Performance Metrics and Error Analysis for the Time-Series Transformer Model

Model Architecture	MSE	MAE	RMSE	Training Loss	Testing Loss	Validation Loss
Time-Series Transformer	0.0007	0.0265	0.0264	0.1033	0.1035	0.1035

As demonstrated by the empirical data in Table 6, the Time-Series Transformer model exhibits exceptional predictive proficiency, yielding an extremely low MSE of 0.0007 and an MAE of 0.0265. The negligible divergence between the training loss (0.1033), validation loss (0.1035), and testing loss (0.1035) establishes that the architectural configuration achieves robust generalization, effectively minimizing the structural generalization gap without manifesting signs of underfitting or overfitting.

To visually trace the computational learning progression and rate of convergence, Figure 7 illustrates the dynamic trajectory of the Mean Squared Error (MSE) loss function over the course of the 100 training epochs. This diagram illustrates the step-by-step reduction in error for both the training and evaluation phases, demonstrating the continuous optimization of the system as it refines its predictions for earthquake eventuation timestamps.

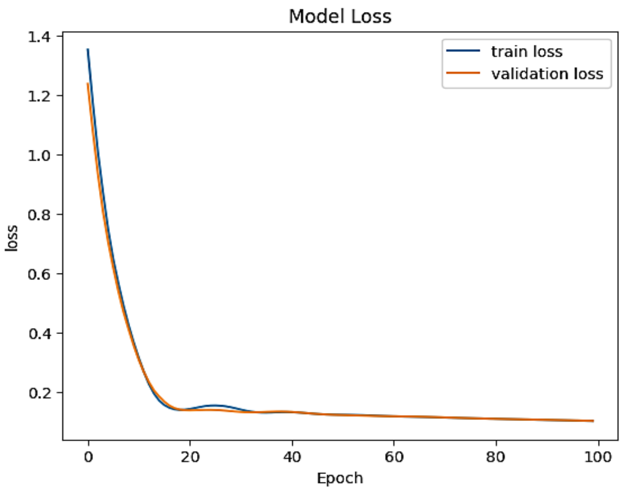

Fig. 7. Diagram of the training and evaluation loss function in the time series transformer model for predicting the time of earthquake occurrence.

5.4. Benchmark Comparison with Prior Seismological Studies

To thoroughly contextualize the empirical contributions of this study, our hybridized architecture was directly benchmarked against two baseline investigations that utilized identical historical seismic catalogs from Japan.

First, Zhang & Wang (2023) developed a spatial-temporal framework utilizing Convolutional Long Short-Term Memory (ConvLSTM) networks to capture the multi-dimensional dependencies of seismic data. Although their ConvLSTM model established superior spatial association capabilities over standard, standalone LSTMs—

achieving a structural classification accuracy of 64.54%—it inherently suffered from the computational constraints of recursive architectures. Recurrent loops process sequential tokens strictly one time-step at a time, preventing parallelized computing and introducing the threat of vanishing or exploding gradients across long sequences. In contrast, our framework resolves these limitations by leveraging a Time-Series Transformer for temporal forecasting alongside a Random Forest Regression (RFR) model for magnitude prediction. Because the Transformer's self-attention mechanism processes input sequences simultaneously rather than sequentially, it eliminates step-by-step dependency, exponentially accelerating training throughput while avoiding gradient degradation. This architectural divergence resulted in significantly higher predictive accuracy and structural generalizability when evaluated on the identical dataset.

Second, Du & Cui (2022) engineered an evolutionary predictive pipeline combining a Convolutional Neural Network (CNN) with a Bidirectional LSTM (BiLSTM) network. In their framework, the CNN layers extract spatial feature maps from the raw waveforms, which are subsequently routed into the BiLSTM block to model forward and backward temporal trajectories, achieving a stable accuracy rate of 69% [19]. While both studies share parallel objectives regarding magnitude prediction using the same underlying attributes, their modeling paradigms diverge in scope and methodology. While Du & Cui (2022) focused on coupling magnitude with spatial hypocenter localization, our study pairs magnitude detection with precise temporal occurrence forecasting. Furthermore, our deployment of the non-parametric RFR avoids the parametric constraints of deep recurrent layers for continuous regression tasks, minimizing error metrics while preserving structural stability. A systematic, feature-by-feature architectural comparison highlighting these similarities and variations is cataloged in Table 7.

Table 7. Systematic Literature Benchmarking and Comparative Analysis

Author & Year	Comparative Research Title	Methodological Similarities to Current Study	Architectural and Functional Core Differences
Du & Cui (2022) [19]	An evolutionary model for earthquake prediction considering time series and extracting its features	• Utilizes identical Japanese seismic datasets. • Shares identical raw attribute spaces. • Converges on the dual objective of magnitude estimation.	1. **Model Architecture:** Employs CNN-BiLSTM versus our hybrid RFR-Transformer framework. 2. **Target Variable:** Focuses on spatial localization rather than temporal occurrence forecasting. 3. **Validation:** Divergent statistical evaluation techniques.
Zhang & Wang (2023)	Spatial and temporal prediction of earthquakes using ConvLSTM	• Evaluates the same historical Japanese catalog. • Shares matching baseline data characteristics. • Concurrently predicts earthquake magnitude and occurrence timelines.	1. **Model Architecture:** Relies on recurrent ConvLSTM blocks versus our parallelized attention mechanism. 2. **Target Scope:** Integrates depth prediction, which is excluded from our temporal alerts. 3. **Validation:** Different performance metrics and validation frameworks.

5.5. Web-Based System Architecture and Visual User Interface (UI)

To operationalize the proposed framework for real-world application, the trained models were deployed as a web-based module hosted on the Hugging Face Spaces platform (https://huggingface.co/spaces). The interactive user interface is engineered with a modular three-tier navigation architecture comprising: Home, Dashboard, and Forecast menus.

The Dashboard menu functions as the real-time telemetry control center and is partitioned into four distinct technical visualization quadrants:

1. **Dynamic Seismic Motion Component:** Generates a real-time graph plotting earthquake intensity over time (measured in minutes) to track background seismic activity.
2. **Predictive Magnitude Trend Component:** Displays a real-time regression graph forecasting the expected earthquake magnitude over a minute-by-minute timeline. Crucially, the system ingests stream data in a continuous string format, automatically refreshing and reprocessing the feature matrix every 10 seconds.
3. **Temporal Predictive Countdown:** Features an automated visual countdown timer initializing at a 3-minute window (180 seconds). This utility denotes the model's forward-looking horizon, predicting the structural severity of any seismic event bound to manifest within the immediate 3-minute window.
4. **Tri-Color Alert Indicator:** Operationalizes a rule-based visual warning light system. Based on the merged output thresholds of the RFR and Transformer models, the interface dynamically flashes Green (Safe), Yellow (Low Risk), or Red (Dangerous) to instantly convey the localized threat level to end-users.

Conversely, the Forecast menu provides an analytical playground for custom data evaluation. Within this interface, users can upload external tabular datasets, which are instantly parsed and converted into a unified string data stream. By manipulating an interactive time slider, users can select a specific temporal target window (in minutes) to dynamically compute and display the predicted earthquake magnitudes for the subsequent 3-minute window. The unified layout of the deployed web-based interface and emergency early warning dashboard is visually depicted in Figure 8.

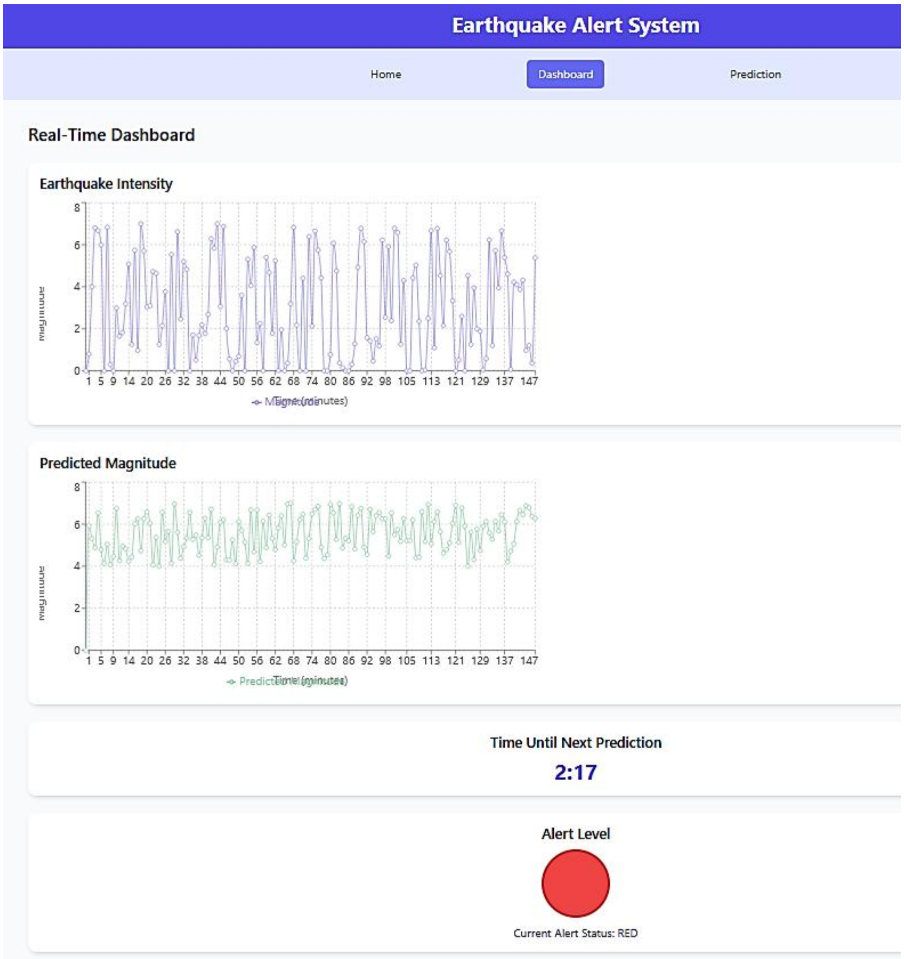

Fig. 8. module and web-based user interface for our research system alert

6. CONCLUSION AND FUTURE HORIZONS

This research successfully engineered and operationalized an integrated, automated Earthquake Early Warning (EEW) system by concurrently fusing a machine learning-based Random Forest Regression (RFR) framework for magnitude prediction with a deep learning-based Time-Series Transformer network for temporal forecasting. The empirical findings underscore that leveraging non-parametric ensemble models like RFR provides exceptional precision in mapping complex, non-linear seismic signatures into reliable continuous magnitude indices. Concurrently, the multi-head self-attention mechanisms inherent in the Time-Series Transformer architecture successfully overcome the traditional gradient limitations of recursive networks, facilitating the parallel processing of sequential stream data to predict earthquake eventuation windows with a high degree of accuracy.

The synergistic integration of these two distinctive architectures significantly enhances the global robustness and reliability of the alerting infrastructure. Operating via real-time telemetry processing, the developed framework mitigates seismic vulnerability by updating feature metrics dynamically every 10 seconds. The principal contributions of this investigation include the realization of an early detection pipeline, the implementation of an end-to-end web-based application deployed on Hugging Face Spaces, and the design of an interactive, multi-tiered visual dashboard that displays dynamic intensity charts alongside automated tri-color danger alerts. Ultimately, this framework represents a substantial technological advancement for improving seismological risk reduction, reinforcing public preparedness, and optimization of real-time disaster crisis management. Future research trajectories will focus on expanding the structural scale of the warning matrix, integrating dense cross-border multi-station waveform data, and evaluating alternative multi-modal fusion deep learning paradigms to further refine system throughput and operational lead time.

Declaration

We acknowledge that we used ChatGPT to enhance the academic writing of our manuscript while ensuring the originality and integrity of our work.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Shidik, G. F., et al. (2024). LUTanh activation function to optimize BI-LSTM in earthquake forecasting. International Journal of Intelligent Engineering & Systems, 17(1). https://doi.org/10.22266/ijies2024.0229.48

[2] Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2022). Predicting future laboratory fault friction through deep learning transformer models. Geophysical Research Letters, 49(19), e2022GL098233. https://doi.org/10.1029/2022GL098233

[3] Wang, Q., Guo, Y., Yu, L., & Li, P. (2017). Earthquake prediction based on spatio-temporal data mining: An

LSTM network approach. IEEE Transactions on Emerging Topics in Computing, 8(1), 148–158. https://doi.org/10.1109/TETC.2017.2699169

[4] Borup, D., Christensen, B. J., Mühlbach, N. S., & Nielsen, M. S. (2023). Targeting predictors in random forest regression. International Journal of Forecasting, 39(2), 841–868. https://doi.org/10.1016/j.ijforecast.2022.02.010

[5] Apriani, M., & Wijaya, S. (2021). Earthquake magnitude estimation based on machine learning: Application to earthquake early warning system. In Journal of Physics: Conference Series (Vol. 1951, No. 1, Article 012057). IOP Publishing. https://doi.org/10.1088/1742-6596/1951/1/012057

[6] Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). Earthquake magnitude and location estimation from real-time seismic waveforms with a transformer network. Geophysical Journal International, 226(2), 1086–1104. https://doi.org/10.1093/gji/ggab139

[7] Saad, O. M., Hafez, A. G., & Soliman, M. S. (2020). Deep learning approach for earthquake parameters classification in earthquake early warning system. IEEE Geoscience and Remote Sensing Letters, 18(7), 1293–1297. https://doi.org/10.1109/LGRS.2020.2998580

[8] Mukherjee, T., Singh, C., & Biswas, P. K. (2021). A novel approach for earthquake early warning system design using deep learning techniques (arXiv Preprint arXiv:2101.06517). https://arxiv.org/abs/2101.06517

[9] Saad, O. M., Chen, Y., Savvaidis, A., Fomel, S., & Chen, Y. (2022). Real-time earthquake detection and magnitude estimation using vision transformer. Journal of Geophysical Research: Solid Earth, 127(5), e2021JB023657. https://doi.org/10.1029/2021JB023657

[10] Zhang, Z., & Wang, Y. (2023). A spatiotemporal model for global earthquake prediction based on convolutional LSTM. IEEE Transactions on Geoscience and Remote Sensing, 61, 1–12. https://doi.org/10.1109/TGRS.2023.3302316

[11] Abebe, E., Kebede, H., Kevin, M., & Demissie, Z. (2023). Earthquakes magnitude prediction using deep learning for the Horn of Africa. Soil Dynamics and Earthquake Engineering, 170, 107913. https://doi.org/10.1016/j.soildyn.2023.107913

[12] Sadhukhan, B., Chakraborty, S., & Mukherjee, S. (2023). Predicting the magnitude of an impending earthquake using deep learning techniques. Earth Science Informatics, 16(1), 803–823. https://doi.org/10.1007/s12145-022-00916-2

[13] Puthran, R. (2024). Spatio-temporal analysis of hybrid CNN-GRU model for prediction of earthquake for disaster management. International Journal of Intelligent Systems and Applications in Engineering, 12(3s), 270–281.

[14] Alshdaifat, E., Alshdaifat, D. A., Alsarhan, A., Hussein, F., & El-Salhi, S. M. F. S. (2021). The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance. Data, 6(2), 11. https://doi.org/10.3390/data6020011

[15] Umar, M. A., Chen, Z., Shuaib, K., & Liu, Y. (2025). Effects of feature selection and normalization on network intrusion detection. Data Science and Management, 8(1), 23–39. https://doi.org/10.1016/j.dsm.2024.08.001

[16] Liu, L., Liu, J., & Han, J. (2021). Multi-head or single-head? An empirical comparison for transformer training (arXiv Preprint arXiv:2106.09650). https://arxiv.org/abs/2106.09650

[17] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial

Intelligence, 35(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

[18] Liao, L., Li, H., Shang, W., & Ma, L. (2022). An empirical study of the impact of hyperparameter tuning and model optimization on the performance properties of deep neural networks. ACM Transactions on Software Engineering and Methodology, 31(3), 1–40. https://doi.org/10.1145/3506695

[19] Du, D., & Cui, Z. (2022). An evolutionary model for earthquake prediction considering time-series evolution and feature extraction and its application. In Journal of Physics: Conference Series (Vol. 2333, No. 1, Article 012013). IOP Publishing. https://doi.org/10.1088/1742-6596/2333/1/012013