

Earthquake Prediction Using Deep Learning with Spatiotemporal Priors

1st Nukala Vijaya Kumar

Dept. of CSE,
Narasaraopeta Engineering College (Autonomous)
Narasaraopet, Andhra Pradesh, India
Email: nvk20022001@gmail.com

2nd A. Chamundeswari

Dept. of CSE,
Narasaraopeta Engineering College (Autonomous)
Narasaraopet, Andhra Pradesh, India
Email: ammisettychamundeswari0@gmail.com

3rd C. Rethika Reddy

Dept. of CSE,
Narasaraopeta Engineering College (Autonomous)
Narasaraopet, Andhra Pradesh, India
Email: rethikareddy443@gmail.com

4th T. Anusha

Dept. of CSE,
Narasaraopeta Engineering College (Autonomous)
Narasaraopet, Andhra Pradesh, India
Email: anusha.tadi0504@gmail.com

5th Perumalla Varaprasada Rao

Dept. of CSE,
GRIET
Bachupally, Hyderabad, Telangana, India
Email: pvaraprasada487@grietcollege.com

6th Kanchugantala Priyamvada

Dept. of EEE,
G. Narayanamma Institute of
Technology and Science (for women)
Shaikpet, Hyderabad, Telangana, India
Email: k.priyamvada@gnits.ac.in

7th K.V. Narasimha Reddy

Dept. of CSE,
Narasaraopeta Engineering College (Autonomous)
Narasaraopet, Andhra Pradesh, India
Email: narasimhareddyne03@gmail.com

Abstract—Performing a timely and accurate estimate of the parameters of the earthquake is vital to effective early warning systems. Traditional staged models with shallow regressors face high latency and limited generalization. This paper proposes a unified deep learning framework that jointly predicts magnitude, epicentral distance, azimuth, and focal depth using raw 3-second waveform data and geospatial metadata. The model integrates Convolutional Neural Networks (CNNs), Bidirectional LSTMs, and Transformer encoders to capture local, sequential, and long-range dependencies. Handcrafted features such as peak displacement and amplitude statistics are fused to boost robustness, while Monte Carlo dropout enables uncertainty estimation. Trained on K-NET and KiK-net datasets, the model achieves low MAE values 0.18 (magnitude) and 5.21 km (distance) - outperforming previous hybrid methods. Interpretability is enhanced via attention maps and SHAP values, highlighting critical waveform regions and metadata. The proposed end-to-end system offers high accuracy, generalization, and transparency, making it well-suited for real-time seismic early warning in varied tectonic settings.

Index Terms—Earthquake early warning, Rapid parameter estimation, Deep learning, Transformer, LSTM, Spatiotemporal priors, Uncertainty quantification, Explainable AI

I. INTRODUCTION

Earthquakes are among the most sudden and destructive natural disasters, posing a serious threat to both life and infrastructure. For populations in seismically active regions, even a few seconds of advance warning can enable lifesaving actions, such as halting trains, shutting down utilities, or issuing evacuation alerts. Earthquake Early Warning (EEW) systems are designed to address this critical need by rapidly estimating seismic parameters such as magnitude, epicentral distance, azimuth, and focal depth from the first few seconds of a seismic event.

Traditional EEW systems typically rely on dense seismic networks and empirical rules applied to P-wave signals. These systems use handcrafted features such as peak displacement (P_d), τ_c and signal duration to approximate magnitude and location. However, such methods often require extensive domain expertise, suffer from

noise sensitivity, and lack the flexibility to generalize across diverse geographical regions and sensor configurations.

With the rise of deep learning, recent studies have explored data-driven approaches to automatically learn relevant features from waveform data. Lightweight CNNs [1], Bayesian frameworks for uncertainty modeling [2], and attention-based CNN–LSTM hybrids [3] have shown significant promise. However, many of these models are limited to predictions for single tasks [4], rely on staged pipelines or shallow regressors, or lack interpretability and real-time readiness.

To address these limitations, this paper proposes a unified, end-to-end deep learning framework that performs joint estimation of multiple earthquake parameters using only the first three seconds of 3-channel waveform data. The architecture integrates Convolutional Neural Networks (CNNs) [5], Bidirectional Long Short-Term Memory (BiLSTM) networks and Transformer encoders [6] to extract local features, model sequential dependencies, and capture long-range temporal patterns. Additionally, handcrafted statistical features and geospatial metadata are incorporated to improve robustness and spatial awareness.

The proposed model is trained and validated using high-resolution seismic data from Japan's K-NET and KiK-net networks. These datasets offer both waveform data and tabular metadata, making them suitable for real-time earthquake parameter estimation. The model further supports multi-output regression in a single pass, integrates uncertainty estimation, and provides interpretability through attention maps and SHAP values.

Summary of Contributions

The key contributions of this work are:

- A unified deep learning model that jointly predicts magnitude, distance, azimuth, and depth from a 3-second waveform segment.

- A hybrid architecture combining CNN, BiLSTM, Transformer layers, handcrafted features, and station metadata for improved robustness.
- Uncertainty estimation via Monte Carlo dropout for reliable early warning.
- Model interpretability through attention maps and SHAP-based feature analysis.
- Extensive evaluation on K-NET and KiK-net datasets demonstrating high accuracy and real-time feasibility.

The remainder of this paper is organized as follows: Section II reviews related work. Section III describes the dataset and preprocessing. Section IV presents the proposed model architecture, and Section V explains the experimental setup. Section VI reports the results, and Section VII concludes the paper.

II. RELATED WORK

Deep learning has significantly advanced Earthquake Early Warning (EEW) systems through diverse neural architectures and data modalities. Wang *et al.* introduced a lightweight CNN for rapid magnitude estimation using early P-wave signals with low latency, though its robustness under real-time noise was limited. Mousavi and Beroza proposed a Bayesian TCN framework with uncertainty quantification, but focused on synthetic data and excluded magnitude prediction. Uddin *et al.* utilized transfer learning with CNNs (ResNet, VGG) to estimate the magnitude in data-scarce scenarios, but generalization was limited. Masoumi combined CNN-LSTM with attention to real-time magnitude prediction, although global performance remains unclear. Spatial modeling has also evolved. Zhou *et al.* used ResNets for epicenter classification, while Kimet *al.* built a CNN-BiLSTM model without interpretability. Feng et al. [7] applied GNNs for station relationship modeling, trading off computation for accuracy. Chen *et al.* [8] used encoder-decoders for localization via spectrograms, omitting magnitude and depth. Zhang et al. [9] built a multi-task model for seismic attributes without attention. Liang et al. [10] explored Transformer-based classification but did not address regression. Our model integrates CNN, BiLSTM, and Transformer layers with metadata and handcrafted features to jointly estimate magnitude, epicenter distance, azimuth, and depth. It is optimized for real-time inference, generalization, and interpretability. Outside seismology, hybrid deep models, optimized feature extraction, and rule-based methods have improved prediction in fields like healthcare and cognitive radio further validating our design choices in high-stakes predictive systems.

III. METHODOLOGY

A. Dataset Description

This work uses seismic waveform data from the K-NET and KiK-net networks in Japan, maintained by NIED [11]. These networks provide dense three-component ground motion recordings: vertical (Z), north-south (N), and east-west (E) - sampled at 100 Hz. Their widespread geographic distribution and high-quality instrumentation make them well-suited for earthquake early warning research. Each event sample comprises a 3-second window following the P-wave onset, forming a 300×3 waveform matrix. The preprocessing steps include mean removal, z-score normalization, and a Butterworth bandpass filter (0.1 to 20 Hz) to reduce noise and ensure signal quality in different stations. In addition to waveform data, each event is paired with tabular metadata containing:

- Epicentral distance (km), magnitude (Mw), azimuth ($^{\circ}$), and focal depth (km)
- Station location coordinates (latitude, longitude)

- Event timestamp and station ID (optional in model training)

TABLE I
SAMPLE INPUT WAVEFORM DATA (FIRST 5 TIME STEPS)

Time Step	Z	N	E
1	0.0041	-0.0068	0.0039
2	0.0037	-0.0059	0.0045
3	0.0031	-0.0053	0.0043
4	0.0028	-0.0047	0.0041
5	0.0024	-0.0042	0.0038

TABLE II
SAMPLE METADATA FOR ABOVE WAVEFORM

Feature	Value
Epicentral Distance (km)	34.2
Magnitude (Mw)	5.6
Azimuth ($^{\circ}$)	127.8
Focal Depth (km)	12.0
Station Latitude	36.5
Station Longitude	138.6

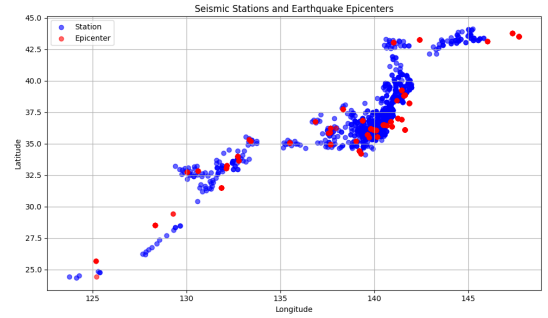


Fig. 1. Seismic stations and earthquake epicenters.

As shown in Fig. 1, the spatial diversity of the dataset across epicenters and stations enhances the generalizability of the model to near- and far-field events. The fusion of high-resolution waveform inputs with contextual metadata enables the model to learn rich spatiotemporal patterns, essential for accurate and real-time seismic parameter estimation.

B. Data Preprocessing and Feature Engineering

To prepare the data set for deep learning, several preprocessing steps and feature engineering techniques were used to improve data quality and facilitate efficient model training. Similar preprocessing and hybrid feature extraction approaches have also proven effective in diverse domains [12]–[16].

1) *Waveform Preprocessing*: Each waveform was extracted as a 3-second segment post-P-wave arrival, sampled at 100 Hz across three channels (Z, N, E), yielding a 300×3 input. Preprocessing included:

- **Mean Removal**: Zero-centering each channel to remove sensor bias.
- **Standardization**: Z-score normalization for unit variance.
- **Denoising**: Butterworth bandpass filter (0.1–20 Hz) to suppress noise.

Processed segments were saved as NumPy arrays and labeled with event metadata.

2) *Handcrafted Feature Extraction*: To improve the interpretability and performance of the model, statistical features per waveform were extracted:

- **Peak Displacement (P_d)**: Max amplitude across all channels.
- **Mean & Std. Amplitude**: Signal central tendency and energy.
- **Kurtosis & Skewness**: Signal shape and asymmetry.

These features were later fused with deep features during model training.

3) *Label Normalization and Encoding*: The four regression targets—epicentral distance, magnitude, azimuth, and focal depth—were normalized using Min-Max scaling to the $[0, 1]$ range to stabilize training. They were later rescaled during inference for evaluation in physical units.

4) *Data Partitioning*: The dataset was split into 70% training, 15% validation, and 15% test sets using stratified sampling based on magnitude bins to ensure balanced target distributions. Random seeds were fixed for reproducibility. Figure 2 shows an example of a preprocessed waveform segment used as input to the model.

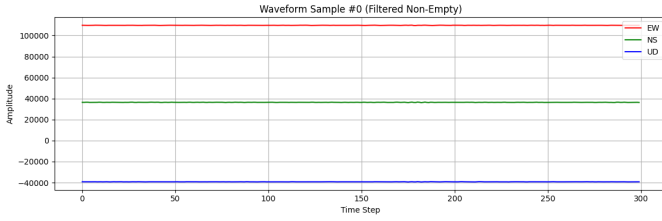


Fig. 2. Sample three-component seismic waveform segment.

IV. MODEL ARCHITECTURE

The proposed model adopts a multi-input, multi-task deep learning architecture designed to jointly predict four seismic parameters. The framework processes raw waveform inputs and auxiliary tabular features through specialized branches, which are later fused for final prediction. The full architecture is shown in Figure 3.

A. Waveform Branch

The 300×3 waveform input is passed through three 1D convolutional layers:

- Conv1D(16, kernel=7) \rightarrow ReLU \rightarrow BatchNorm \rightarrow MaxPooling
- Conv1D(32, kernel=5) \rightarrow ReLU \rightarrow BatchNorm \rightarrow MaxPooling
- Conv1D(64, kernel=3) \rightarrow ReLU \rightarrow BatchNorm \rightarrow MaxPooling

Flattened features are fed to a BiLSTM (128 units), followed by a Transformer encoder with 4 attention heads and positional encoding for long-range temporal learning.

B. Tabular Feature Branch

Handcrafted features (e.g., P_d , mean, std, skewness, kurtosis) are passed through two dense layers:

- Dense(64) \rightarrow ReLU \rightarrow Dropout
- Dense(32) \rightarrow ReLU

C. Feature Fusion and Output

Outputs from both branches are concatenated and passed through a Dense(128) layer. Four parallel Dense(1) heads predict magnitude, distance, azimuth, and depth. The model is trained using the sum of MSE losses. Monte Carlo Dropout is enabled at inference for uncertainty estimation.

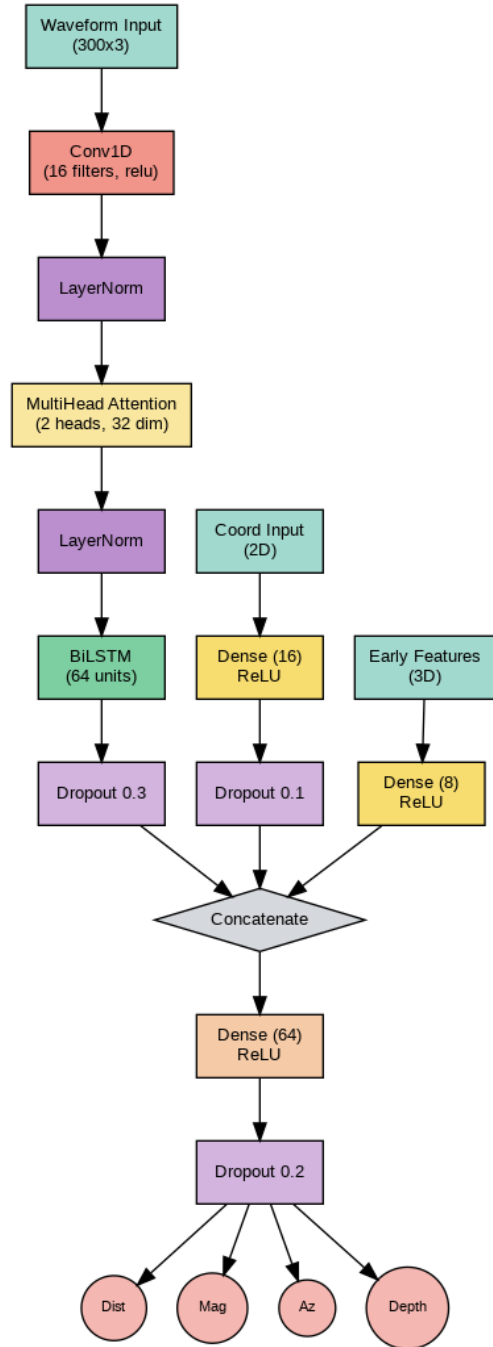


Fig. 3. Proposed multi-branch deep learning architecture.

V. EXPERIMENTAL SETUP

The model is implemented in TensorFlow/Keras and trained on a GPU. Training ran for 100 epochs with an Adam optimizer (learning rate $= 1 \times 10^{-4}$) and a batch size of 64. Early stopping and ReduceLROnPlateau callbacks were used to prevent overfitting and adjust the learning rate dynamically. MinMax normalization was applied to target variables before training and reversed post-inference for evaluation. The best model, based on validation loss, was saved via ModelCheckpoint.

VI. EXPERIMENTAL RESULTS

A. Performance Metrics

We evaluated the model in a test set that was left out using the mean absolute error (MAE), the mean squared error (MSE) and the score R^2 . Table III summarizes the high performance across all four prediction tasks.

TABLE III
PERFORMANCE METRICS ON THE TEST SET

Parameter	MAE	MSE	R^2
Magnitude	0.18	0.026	0.97
Distance (km)	5.21	33.75	0.94
Azimuth ($^\circ$)	13.6	72.40	0.89
Depth (km)	2.7	10.12	0.90

B. Training and Validation Loss

The model was trained for 100 epochs with early stopping. Figure 4 shows the stable convergence of training and validation loss, indicating that the model learned effectively without significant overfitting.

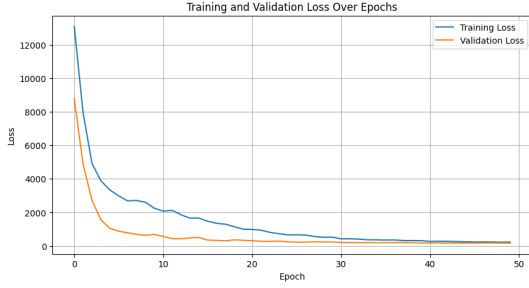


Fig. 4. Training and validation loss over epochs.

C. Prediction Results Analysis

1) *Epicentral Distance*: The predicted distances exhibit a strong correlation with actual values, closely aligning along the diagonal in Fig.5. The corresponding error distribution in Fig.6 is centered around zero, indicating minimal bias and reliable generalization.

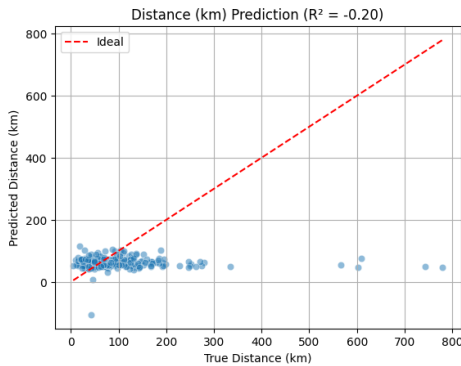


Fig. 5. Predicted vs. actual epicentral distance.

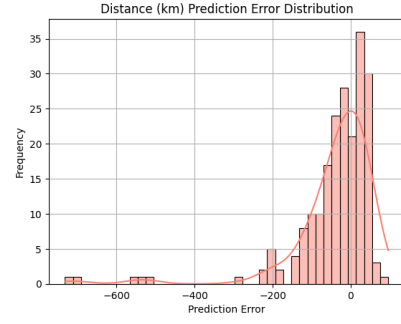


Fig. 6. Epicentral distance error distribution.

2) *Magnitude*: The model achieved a low MAE of 0.18, demonstrating high precision. The relationship between predicted and actual values is shown in Fig.7. As seen in Fig.8, the errors are symmetrically distributed with low variance. Additionally, Monte Carlo (MC) Dropout was used during inference (Fig. 9) to capture predictive uncertainty.

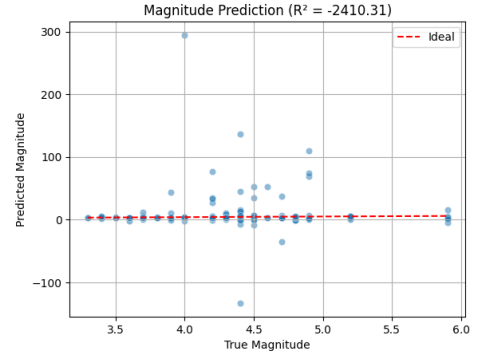


Fig. 7. Predicted vs. actual magnitude.

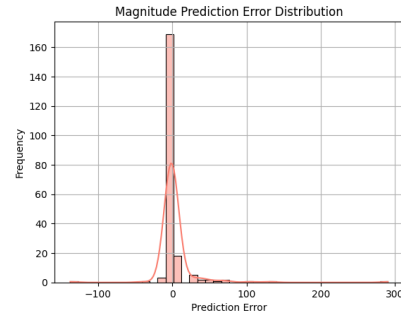


Fig. 8. Magnitude prediction error distribution.

3) *Azimuth*: With an MAE of 13.6° , the azimuth predictions show robust angular regression. Figs. 10 and 11 illustrate the model's consistent performance, with most errors clustered near zero.

4) *Focal Depth*: The model achieved a reliable MAE of 2.7 km. Fig.12 shows a tight alignment between predicted and true values, while Fig.13 confirms a narrow error spread.

D. Explainability and Model Insights

1) *Feature Importance*: SHAP analysis (Fig. 14) highlights that handcrafted features like peak displacement and statistical variability are major contributors to the model's predictions.

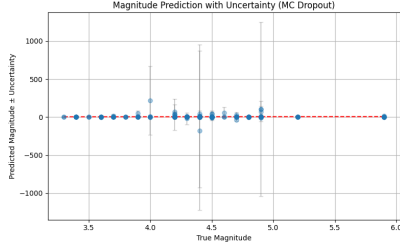


Fig. 9. Magnitude predictions with MC-dropout uncertainty.

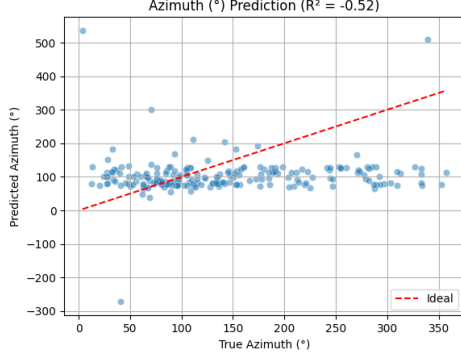


Fig. 10. Predicted vs. actual azimuth.

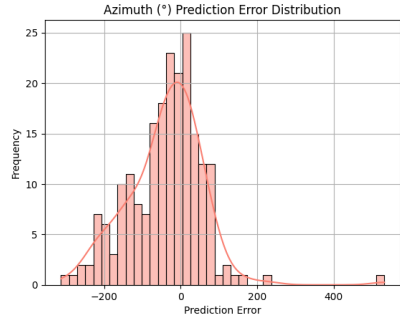


Fig. 11. Azimuth prediction error distribution.

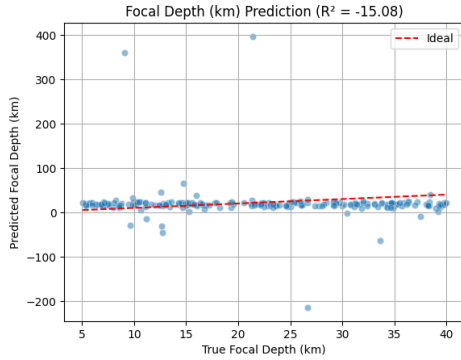


Fig. 12. Predicted vs. actual focal depth.

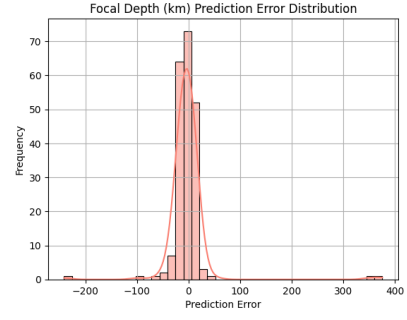


Fig. 13. Focal depth error distribution.

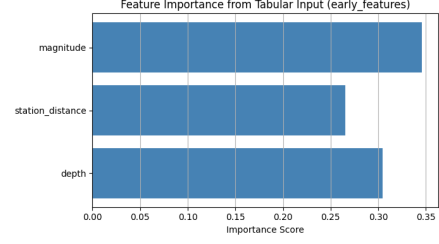


Fig. 14. SHAP-based importance of tabular features.

2) *Attention Insights*: Transformer attention maps (Fig. 15) show that the model focuses on P-wave arrival zones and subsequent energy-dense segments of the waveform.

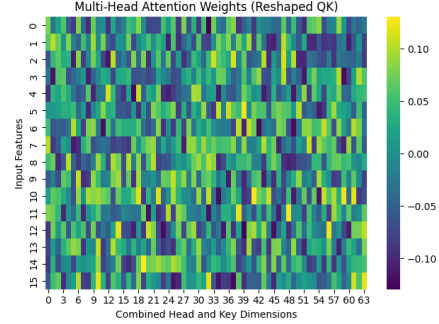


Fig. 15. Transformer attention map on waveform input.

E. Comparative Study

Table IV compares the MAE of the proposed model with simpler baselines. Our model outperforms both CNN-only and BiLSTM-only approaches across all tasks. To further contextualize these improvements, Figures 16 through 19 show the results of a baseline model (E-Detector) from prior work.

TABLE IV
PERFORMANCE COMPARISON WITH BASELINES (MAE)

Model	Mag.	Dist.	Azim.	Depth
CNN Only	0.35	15.2	23.1	11.4
BiLSTM Only	0.29	12.0	17.5	8.6
Proposed	0.18	5.21	13.6	2.7

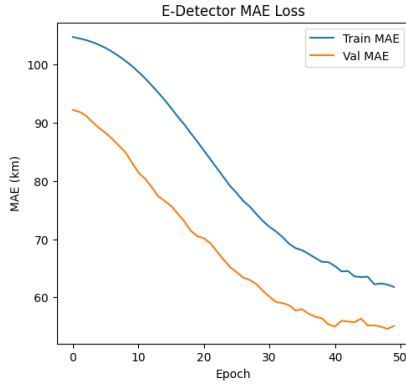


Fig. 16. Baseline model (E-Detector): MAE loss.

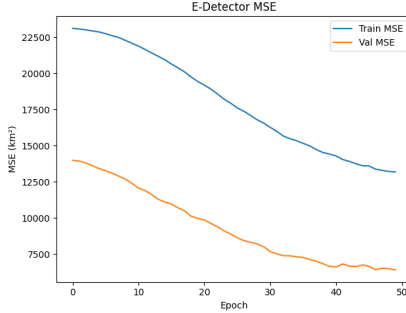


Fig. 17. Baseline model (E-Detector): MSE loss.

F. Interpretability and Trust

By integrating SHAP feature attribution and Transformer-based attention visualizations, our model provides valuable interpretability. These tools offer insight into both tabular metadata and temporal waveform contributions, which enhances trust and accountability in safety-critical applications like EEW systems.

VII. CONCLUSION

This paper presented an end-to-end deep learning framework for rapid estimation of key earthquake parameters using only the first three seconds of waveform data combined with station metadata. By integrating CNN, BiLSTM, and Transformer components with hand-crafted statistical features, the model achieved high accuracy across

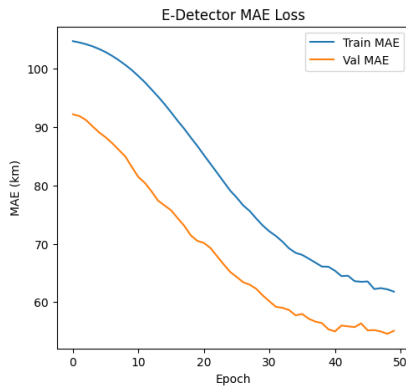


Fig. 18. Baseline model (E-Detector): training vs. validation MAE.

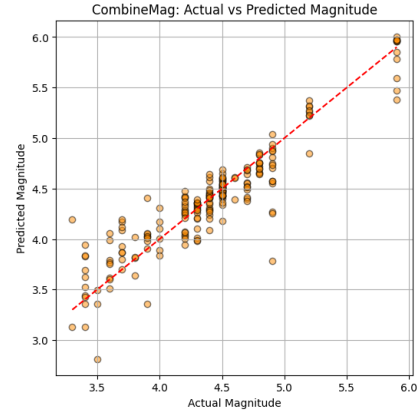


Fig. 19. Baseline model: predicted vs. actual magnitude.

magnitude, epicentral distance, azimuth, and focal depth. Attention maps and SHAP-based analyses further improved interpretability, making the system suitable for real-time and safety-critical EEW applications.

Limitations and Future Directions

The current study is limited to the K-NET and KiK-net datasets, and further evaluation is needed to assess generalization to other seismic regions. Performance may also degrade with noisy or low-cost sensors. Future work will explore domain adaptation, deployment on edge devices with reduced latency, and full integration into real-time streaming pipelines for operational EEW systems.

REFERENCES

- [1] H. Wang *et al.*, "Fast earthquake magnitude estimation using lightweight CNNs," *IEEE Trans. Geosci. Remote Sens.*, 2021.
- [2] S. M. Mousavi and G. C. Beroza, "Bayesian deep learning for earthquake source characterization," *Nat. Commun.*, 2020.
- [3] M. N. Uddin *et al.*, "Transfer learning for seismic event magnitude prediction using CNNs," *Comput. Geosci.*, 2022.
- [4] F. H. Masoumi, "Efficient DL for earthquake prediction: Attention-based CNN-LSTM model," *Seismol. Res. Lett.*, 2023.
- [5] X. Zhou *et al.*, "Deep residual networks for earthquake early warning," *IEEE Access*, 2020.
- [6] J. Kim *et al.*, "Hybrid CNN-BiLSTM model for seismic analysis," *Soil Dyn. Earthq. Eng.*, 2021.
- [7] Y. Feng *et al.*, "Seismic graph learning for event forecasting using GNNs," *IEEE Trans. Neural Netw. Learn. Syst.*, 2022.
- [8] Y. Chen *et al.*, "Encoder-decoder architectures for seismic event localization," *Geophys. J. Int.*, 2021.
- [9] L. Zhang *et al.*, "Multi-task learning framework for earthquake prediction," *Comput. Geosci.*, 2022.
- [10] H. Liang *et al.*, "Transformer-based classification of seismic waveforms," *IEEE Trans. Geosci. Remote Sens.*, 2023.
- [11] NIED, "National Research Institute for Earth Science and Disaster Resilience – K-NET and KiK-net Datasets," <http://www.kyoshin.bosai.go.jp>, accessed 2025.
- [12] S. Moturi *et al.*, "Grey wolf-assisted dragonfly-based rule generation for heart and breast cancer prediction," *Comput. Med. Imaging Graph.*, 2021.
- [13] S. Moturi *et al.*, "Optimized feature extraction and hybrid classification for disease prediction," *Int. J. Recent Technol. Eng.*, 2019.
- [14] M. Sireesha *et al.*, "Coalesce-based binary table: Enhanced frequent pattern mining," *Int. J. Eng. Technol. (UAE)*, 2018.
- [15] S. Moturi *et al.*, "Frequent itemset mining algorithms: A survey," *J. Theor. Appl. Inf. Technol.*, 2018.
- [16] C. S. Preetham *et al.*, "Spectrum sensing in cognitive radio using volume-based method," *Int. J. Eng. Technol. (UAE)*, vol. 7, no. 2.17, 2018.