JGR Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH001005

Key Points:

- The proposed model predicts peak ground velocity using multi-station velocity waveforms and physically meaningful seismic features
- Incorporating low-frequency waveforms and engineered features significantly reduces systematic underestimation of strong ground motion
- TT-SAM achieves 97.8% of predictions within ±1 intensity level on the test data set, demonstrating reliable accuracy

Correspondence to:
C.-H. Chan,
chchan@ncu.edu.tw

Citation:
Chen, Y.-H., Chan, C.-H., Chang, C.-C., & Ma, K.-F. (2026). A deep learning framework for peak ground velocity prediction using multi-station velocity waveforms: The Taiwan Transformer Shaking Alert Model (TT-SAM). *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2025JH001005. https://doi.org/10.1029/2025JH001005

Received 10 SEP 2025
Accepted 13 APR 2026

Author Contributions:
Conceptualization: Chung-Han Chan
Data curation: Yu-Heng Chen
Formal analysis: Yu-Heng Chen, Chieh-Chen Chang
Funding acquisition: Chung-Han Chan, Kuo-Fong Ma
Investigation: Chung-Han Chan, Chieh-Chen Chang
Methodology: Yu-Heng Chen, Chung-Han Chan, Chieh-Chen Chang
Project administration: Chung-Han Chan, Kuo-Fong Ma

A Deep Learning Framework for Peak Ground Velocity Prediction Using Multi-Station Velocity Waveforms: The Taiwan Transformer Shaking Alert Model (TT-SAM)

Yu-Heng Chen[1,2], Chung-Han Chan[1,2], Chieh-Chen Chang[1,2], and Kuo-Fong Ma[2,3]

[1]Department of Earth Sciences, National Central University, Taoyuan, Taiwan, [2]Earthquake-Disaster & Risk Evaluation and Management (E-DREaM) Center National Central University, Taoyuan, Taiwan, [3]Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan

Abstract This study proposes a deep-learning–based regional earthquake early warning model, the Taiwan Transformer Shaking Alert Model (TT-SAM). The model adopts peak ground velocity (PGV) as its primary ground shaking prediction unit, aiming to better reflect actual structural damage and thereby enhance the practical utility and accuracy of the warning system. Unlike most deep learning earthquake warning approaches that rely solely on raw waveforms, TT-SAM integrates velocity waveforms, low-frequency velocity waveforms, and physically meaningful seismic features to enhance the model's ability to capture strong ground-motion characteristics. A rolling-update prediction strategy enables preliminary forecasts within 3 s of P-wave arrival, with continuous updates as additional waveform data become available. To simulate conditions of incomplete waveform acquisition, a time-masking mechanism is applied during training to improve generalization. Furthermore, magnitude oversampling and bias-to-close-station strategies are employed to alleviate underprediction arising from scarce high-intensity samples in the training data. When evaluated on the Taiwan seismic events data set in 2016, TT-SAM achieved 97.8% of its intensity-level predictions within ±1 level, demonstrating outstanding predictive capability and stability. By directly predicting PGV and aligning model outputs with official intensity-based warning criteria, TT-SAM offers a robust and practical framework for next-generation EEW systems. The proposed approach demonstrates the value of combining deep learning with physics-guided feature engineering to enhance both the accuracy and reliability of real-time seismic warning.

Plain Language Summary Earthquake Early Warning (EEW) systems provide crucial seconds of advance notice before strong shaking arrives, allowing people to take protective action. We developed a new deep learning model for Taiwan, the Taiwan Transformer Shaking Alert Model (TT-SAM), to improve these warnings. Unlike traditional systems that focus on acceleration, TT-SAM predicts ground velocity, which is more directly linked to building damage. By combining velocity waveforms with physical features such as displacement from multiple stations, the model captures key shaking patterns more effectively. Tests on the 2016 Meinong earthquake sequence show that TT-SAM achieved 97.8% accuracy, outperforming conventional methods and offering more reliable warnings for disaster prevention. This work demonstrates how combining artificial intelligence with physical understanding of earthquakes can lead to more accurate and reliable early warning systems, helping to reduce earthquake risk and improve public safety.

1. Introduction

Earthquake Early Warning (EEW) systems can issue alerts prior to the arrival of destructive seismic waves, providing response time to relevant stakeholders and playing a critical role in seismic disaster mitigation (Allen & Kanamori, 2003; Allen & Melgar, 2019). Traditional EEW methods can be categorized into source-based approaches (Böse et al., 2012; Chung et al., 2019) and propagation-based approaches (Hoshiba, 2013). Notable operational implementations include the ShakeAlert system in the United States (Kohler et al., 2020; McBride et al., 2022). Source-based models estimate source parameters and employ predictive models to infer ground motion intensity at target locations. The advantage of this method lies in its rapid estimation, which can provide longer warning lead times (defined here as the interval between alert issuance and the arrival of Peak Ground Velocity (PGV) at a target station). However, for large earthquakes with $M \geq 7$, the estimation of source parameters is susceptible to saturation effects, leading to underestimation of earthquake magnitude and consequently affecting the accuracy of intensity predictions (Allen & Melgar, 2019). In contrast, propagation-based

Resources: Chung-Han Chan, Kuo-Fong Ma
Software: Yu-Heng Chen, Chieh-Chen Chang
Supervision: Chung-Han Chan, Kuo-Fong Ma
Validation: Yu-Heng Chen, Chung-Han Chan
Visualization: Yu-Heng Chen
Writing – original draft: Yu-Heng Chen, Chung-Han Chan, Chieh-Chen Chang
Writing – review & editing: Yu-Heng Chen, Chung-Han Chan, Kuo-Fong Ma

models analyze current ground motions using propagation physics to predict intensity changes over short-term intervals (on the order of tens of seconds), yielding more accurate forecasts (Hoshiba, 2013; Lara et al., 2023). However, because these methods rely on waveforms from nearby stations, the available warning time is relatively short (Kodera et al., 2018), potentially creating blind zones near the epicenter with insufficient lead time before severe ground shaking (Meier et al., 2020).

In recent years, advances in machine learning and deep learning have led to numerous studies integrating these techniques into seismology, with applications such as earthquake early warning systems (e.g., Lara et al., 2023; J. T. Lin et al., 2021; Münchmeyer et al., 2021; Saad et al., 2024; Zhu et al., 2022). Pretrained deep learning models can effectively extract features from seismic waveforms, enabling EEW systems to maintain high accuracy while providing longer warning times.

Based on the aforementioned advantages of deep learning, this study draws on the Transformer Earthquake Alert Model (TEAM; Münchmeyer et al., 2021) to develop a deep learning–based EEW model. TEAM is a regional EEW model that adopts an end-to-end deep learning architecture. Unlike traditional methods that depend on source parameter estimation, it processes raw seismic waveforms and triggers station data directly to predict ground motion at target stations. Owing to its end-to-end nature and extensive pretraining on seismic waveforms, this approach enhances the accuracy and operational efficiency of EEW systems, providing a more reliable technological foundation for real-time ground motion prediction. While the original TEAM model predicts Peak Ground Acceleration (PGA), we specifically target Peak Ground Velocity (PGV). This choice is motivated by the strong correlation between PGV and the destructive energy of seismic waves. Previous studies have demonstrated that PGV serves as a robust proxy for structural damage and seismic loss, making it a critical indicator for assessing impact potential (Baltay & Hanks, 2014; Wald et al., 1999; Wu et al., 2004). Consequently, PGV is widely utilized in operational macroseismic intensity scales to determine damage levels. Accordingly, we train our deep learning model to predict PGV directly. This design choice ensures that the model focuses on features directly relevant to structural impact, thereby enhancing the practical utility of the issued alerts.

In this study, we develop a deep learning–based earthquake early warning model built upon the Transformer Earthquake Alert Model (TEAM) framework to directly predict peak ground velocity (PGV) from seismic waveforms. Using a comprehensive regional data set, we systematically evaluate the model's predictive performance and demonstrate its ability to capture the spatial distribution of ground motion in real time.

2. Data

To develop the earthquake early warning model, this study trains on seismic records from the Taiwan Strong Motion Instrumentation Program (TSMIP; Liu et al., 1999). TSMIP features a high-density network of strong-motion stations (Figure 1), providing an extensive set of high-quality records suitable for the large training data sets required by deep learning models. For data selection, events with $M_L \geq 5.5$ during 1999–2008 and $M_L \geq 3.5$ during 2009–2019 were included, yielding 1,259 earthquakes, 952 stations, and a total of 51,902 records. Furthermore, each selected event includes at least one station recording a Central Weather Administration (also known as "CWA") intensity ≥4 (the CWA issues warnings for $M_L > 5.0$ with intensity ≥4 or $M_L > 6.5$ with intensity ≥3; the corresponding intensity levels follow the CWA intensity scale listed in Table 1).

The data set is split into training, validation, and test sets: all 2016 events form the test set, while the remaining data are randomly divided into 80% for training and 20% for validation. The training and test sets exhibit similar distributions in magnitude, depth, and spatial coverage (as shown in Figures 2a–2c, respectively), ensuring that features learned during training generalize effectively to the test data. Station metadata include longitude, latitude, elevation, and V_{s30} (the average shear-wave velocity in the top 30 m) (Figure 1). The data set comprises standard velocity waveforms, low-frequency velocity components, and three physical features—cumulative vertical absolute velocity (CVAV; Reed & Kassawara, 1990), peak displacement (Pd; Wu & Kanamori, 2005), and the frequency-related TP (P. L. Huang et al., 2015). These three physical features are also represented as time-series data and can be input into the model synchronously with the waveform data.

The inclusion of low-frequency velocity components is based on the findings of Y. Y. Lin et al. (2018), who identified two sets of P- and S-wave arrivals near the epicenter of the 2016 M_L6.6 Meinong earthquake—corresponding to the foreshock and the mainshock—separated by approximately 1.8–5.0 s. After applying a 0.33 Hz low-pass filter, the smaller foreshock signals are removed, leaving only the mainshock's

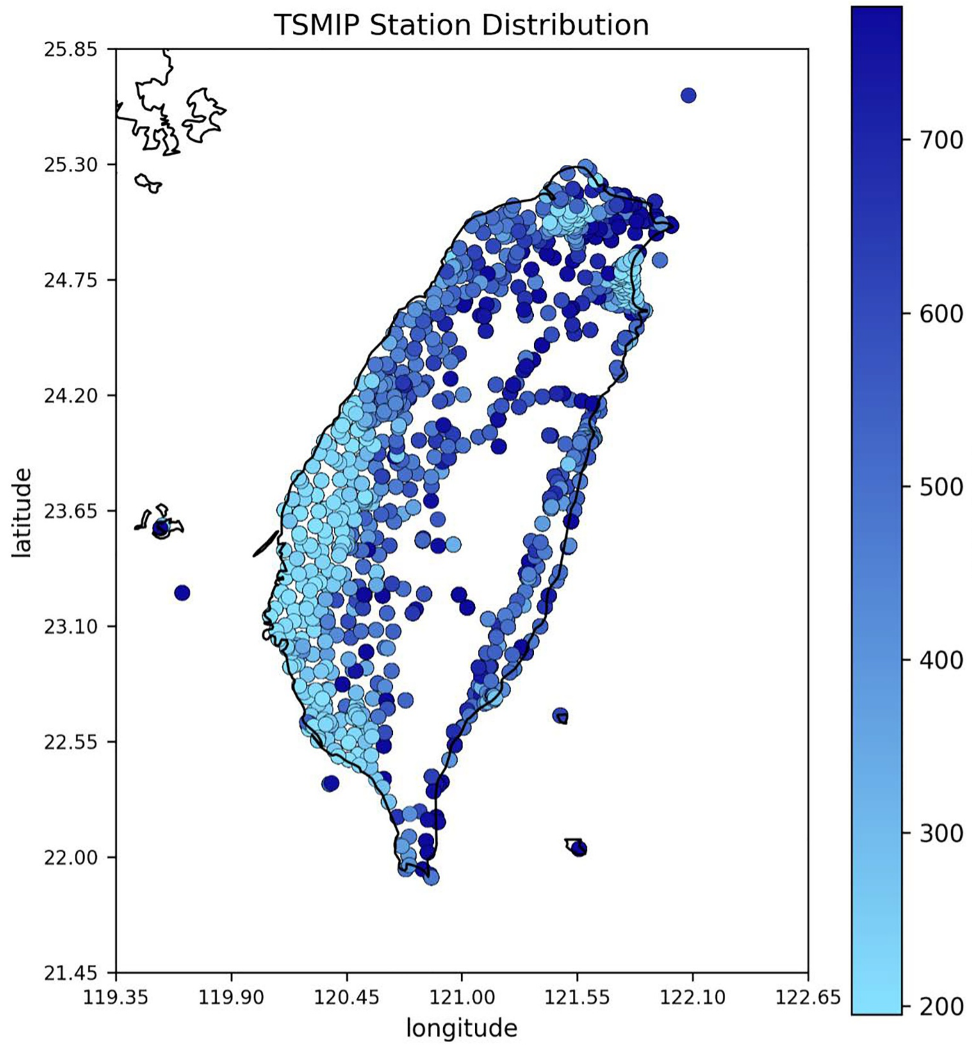

Figure 1. Spatial distribution of stations; station colors represent the V_{s30} values (the average shear-wave velocity in the top 30 m).

pronounced waveform characteristics. Hence, for larger earthquakes, low-frequency filtering effectively suppresses nearby smaller events that occur close in time to the mainshock, reducing misclassification.

For deep learning model training, recent studies have not only used waveforms but have also transformed them into additional physical features (e.g., Zhu et al., 2022). This feature engineering approach—a common practice in machine learning—extracts or transforms key characteristics from raw waveforms to enhance model performance. In designing the input feature set for TT-SAM, we considered various standard EEW parameters,

Table 1
The Central Weather Administration (CWA) Intensity Scale Used in Taiwan, Showing the Correspondence Between Intensity Levels and PGA/PGV Ranges

Intensity	0	1	2	3	4	5−	5+	6−	6+	7
PGA (cm/s^2)	<0.8	0.8–2.5	2.5–8.0	8.0–25	25–80	80–140	140–250	250–440	440–800	>800
PGV (cm/s)	<0.2	0.2–0.7	0.7–1.9	1.9–5.7	5.7–15	15–30	30–50	50–80	80–140	>140

Note. PGA is used to define intensities below level 4, whereas PGV is used for level 5 and above. This table is provided for reference to interpret the intensity-based warning criteria used in this study. Color shading highlights the transition from PGA-based intensity levels (0–4) to PGV-based levels (5 and above).

10.1029/2025JH001005

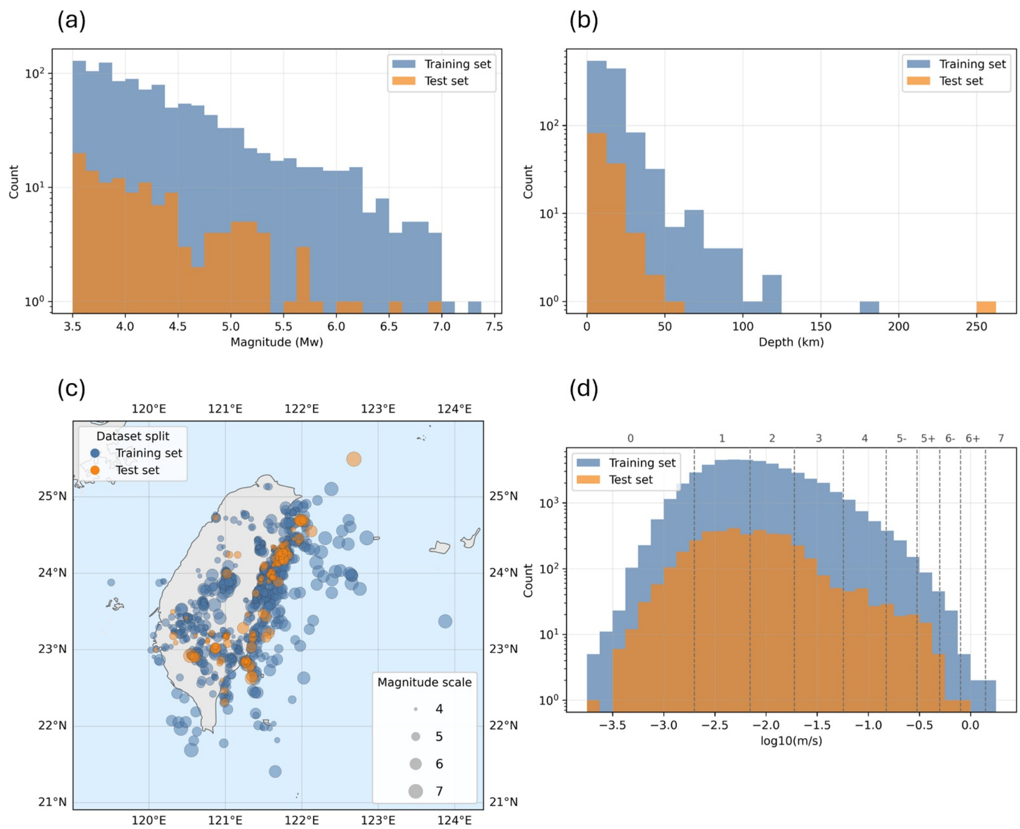

Figure 2. Distribution of seismic events recorded by TSMIP from 1999 to 2019, showing the distributions of (a) magnitude, (b) depth, (c) spatial locations, and (d) peak ground velocity (PGV), with training data indicated in blue and test data in orange.

including squared-velocity metrics (e.g., IV2) and frequency parameters (e.g., τ_c). However, directly including all available parameters can lead to information redundancy and multicollinearity. Therefore, we adopted a physics-based selection strategy to prioritize complementarity and numerical stability. Specifically, while squared metrics like IV2 are widely used, they are mathematically dominated by peak amplitudes, making them physically redundant with Pd—the standard EEW amplitude proxy we selected. Furthermore, squared metrics significantly expand the dynamic range of inputs, which can introduce numerical instability and hinder optimization efficiency in deep learning models compared to linear metrics (Goodfellow et al., 2016; LeCun et al., 2012). Consequently, we excluded IV2 to ensure a well-conditioned optimization landscape. Instead, to capture the shaking energy without such redundancy, we selected CVAV. As a linear integration metric, CVAV aligns directly with the physical dimensionality of our prediction target (PGV) and provides unique duration-related information distinct from Pd. Finally, instead of using τ_c in isolation, we incorporated it through the composite parameter TP ($Pd \times \tau_c$). This parameter integrates frequency content to improve magnitude estimation, helping the model distinguish large earthquakes at a distance from small earthquakes nearby. Together, these features efficiently cover amplitude, energy, and frequency without unnecessary complexity.

Traditional EEW systems often predict ground motion through source parameter estimation and Ground-Motion Models (GMMs), which account for seismic wave attenuation and site effects—both critical factors influencing intensity estimates (D. Y. Chen et al., 2012; Hsiao et al., 2011). In contrast, the original TEAM model (Münchmeyer et al., 2021) uses only acceleration waveforms as input and does not explicitly incorporate site effects. Therefore, this study additionally includes V_{s30} as a station feature. We utilize V_{s30} as a widely accessible proxy for first-order site characterization, although we acknowledge that it may not fully capture complex site effects in all regions (Sahakian et al., 2018; Trugman & Shearer, 2018). Typically, a low V_{s30} value implies soft soil conditions prone to amplification, whereas a high V_{s30} value suggests stiffer ground (Kuo et al., 2012).

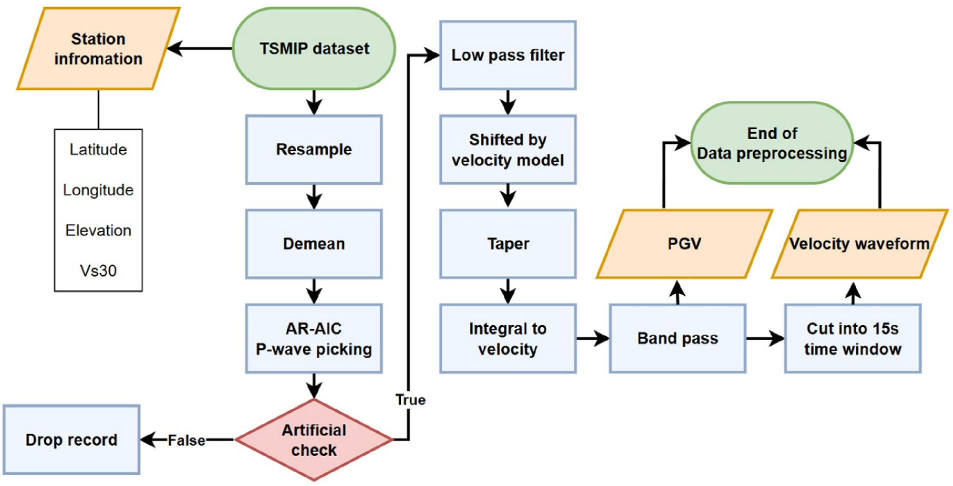

Figure 3. Data preprocessing workflow for this study. The process begins with acquisition of TSMIP data (green capsule), followed by waveform preprocessing (blue rectangle) and manual waveform screening (red diamond), and concludes with output of the preprocessing results (orange parallelogram). This workflow yields waveform data, station parameters (including longitude, latitude, elevation, and V_{s30}), and the PGV recorded at each station.

For data preprocessing (Figure 3), an AR-AIC automatic picker (Akazawa, 2004) is used to detect P-wave arrivals, which are then verified manually. Records are discarded if the automatic and manual picks differ by more than 1 s, and any waveforms exhibiting damage or significant anomalies are also removed to ensure data quality. Next, the three-component waveforms are combined using the Euclidean norm to compute the PGV for each record. To further refine P-wave arrival times, a three-dimensional velocity model from H. H. Huang et al. (2014) is applied, ensuring that waveform data accurately correspond to the correct seismic events. Finally, based on the first trigger station's P-wave pick, 5 s of pre-event background noise and 15 s of seismic waveform are extracted as the input time window for model training.

3. Methodology

Our model is built upon the foundation of the TEAM model (Münchmeyer et al., 2021). However, TEAM was designed using seismic data from Japan and Italy, so directly applying it to Taiwanese data may lead to limitations in accuracy and applicability. Therefore, TT-SAM applies slight architectural modifications and additionally incorporates additional station metadata and physical features to enhance the model's predictive performance for Taiwan. Moreover, the original TEAM model was implemented in TensorFlow 1.14 with CUDA 10.0, which is no longer supported by current GPUs and leads to compatibility issues. To address this, we adapted the model into the PyTorch 1.10.2 framework to ensure compatibility with modern hardware.

3.1. Model Architecture

This study proposes the Taiwan Transformer Shaking Alert Model (TT-SAM), designed to address Taiwan's strong-motion characteristics and earthquake early warning requirements. Its architecture comprises three main components (Figure 4), described below.

3.1.1. Feature Extraction

At the initial stage, a convolutional neural network (CNN; O'Shea & Nash, 2015) is employed to extract waveform features. The CNN consists of normalization, convolutional layers, pooling layers, and fully connected layers. The input data include three-component velocity waveforms recorded at the trigger station, low-frequency velocity waveforms, and physical features, with a time window spanning from 5 s before P-wave arrival at the first station to 15 s after, yielding 4,000 data points at a 200 Hz sampling rate. Additionally, station spatial information is incorporated via positional embeddings (Vaswani et al., 2017), which convert geographic coordinates into vector representations. The resulting embedding vectors are fused with the CNN-extracted feature

10.1029/2025JH001005

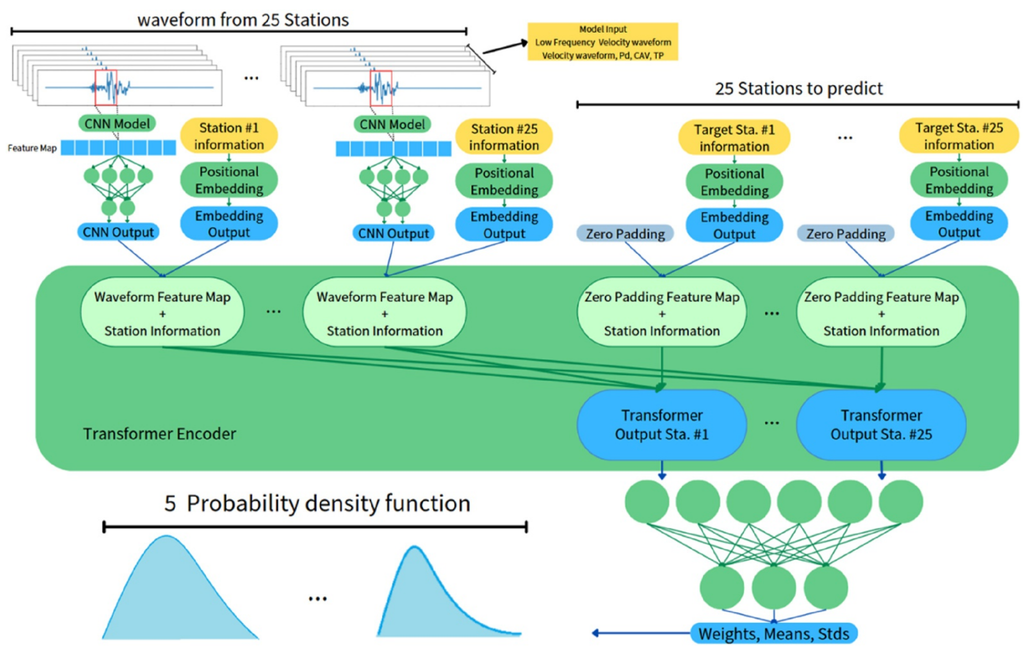

Figure 4. Architecture of the proposed TT-SAM model, with module functions indicated by color: the main model structure in green, input data in yellow, and model outputs in blue. The model input consists of waveform data and station metadata from 25 trigger stations; if fewer than 25 stations are available, the remaining inputs are zero-padded. The input waveforms are first processed by a CNN for feature extraction, producing a feature map. Simultaneously, station metadata (e.g., coordinates and V_{s30}) are transformed via positional embeddings and fused with the CNN output. The combined features are fed into a Transformer Encoder, which uses self-attention to model relationships between input and target stations. Finally, the Transformer Encoder output is used to parameterize a mixture of five Gaussian distributions representing the predictive probability density of PGV.

matrix through element-wise addition, combining temporal features with spatial context for subsequent model processing.

3.1.2. Station Relationship Modeling

We introduce a Transformer Encoder based on the self-attention mechanism (Vaswani et al., 2017) to model spatial and site-effect relationships among stations. Originally applied in natural language processing (Chowdhary, 2020), the Transformer architecture now underpins modern large language models (Achiam et al., 2023) and excels at capturing interactions within sequential data. In this study, the Transformer Encoder receives the feature extraction output and computes inter-station relationships via self-attention. Through this module, the model considers not only the spatial distances of seismic wave propagation but also incorporates site effects and geological conditions, thereby enhancing its ability to recognize spatial variability.

3.1.3. PGV Estimation

After integrating information from both the target and trigger stations, a mixture density network (MDN; Bishop, 1994) generates five Gaussian kernels for each target station and assigns a weight to each component. The PGV estimate is computed as the weighted sum of the means of these components, producing a single point prediction corresponding to the expected value of the mixture distribution. This formulation not only provides a PGV estimate but also quantifies prediction uncertainty. In general, a narrower mixture distribution indicates higher model confidence in the predicted PGV.

However, our analysis indicates that the predicted probability densities tend to be overly concentrated, suggesting that the current mixture distributions are not yet sufficiently calibrated to support robust probabilistic evaluation metrics such as exceedance probabilities or Brier scores. For this reason, the present study focuses primarily on the mean prediction of the mixture distribution when evaluating system performance.

10.1029/2025JH001005

3.2. Training Procedure

During training, the data are divided into input stations and target stations. The model input consists of the seismic waveforms and station features from the first 25 stations triggered by the P-wave; the model must then predict the PGV values at the target stations based on this input.

To stabilize the data distribution and improve learning efficiency, the target PGV values are log-transformed prior to training to compress their range and mitigate the impact of extreme values. The model is trained using the negative log-likelihood loss (Yao et al., 2020) and optimized with the Adam optimizer (Kingma & Ba, 2014). To prevent overfitting, early stopping is employed: training halts automatically if the validation loss does not decrease for 15 consecutive epochs, and the model weights from the epoch with the lowest validation error are retained.

To simulate incomplete waveforms in practical applications, a random time-masking mechanism has been introduced during training. For each training sample, a random interval between 3 and 15 s after the first P-wave arrival is masked by replacing that segment of the waveform with zeros. This design forces the model to rely on early earthquake signals and enables continuous forecasting at different time points during an event. By learning feature variations across different waveform lengths, the model achieves rolling-update predictions: as more data arrive with time, forecasts are automatically updated, allowing the warning system to provide progressively refined estimates and enhancing its dynamic responsiveness and utility.

Additionally, to bolster robustness under station outages, a random station-masking strategy replaces the waveforms and features of a random subset of input stations with zeros. This approach prevents over-reliance on specific stations and simulates real-world station outages, thereby improving model generalization.

Previous studies have reported systematic underestimation of high-intensity events in deep learning–based intensity prediction (e.g., Böse et al., 2012; Münchmeyer et al., 2021), likely due to the scarcity of high-intensity samples in the training data. To improve the model's performance on high-intensity events, two sample enhancement strategies—magnitude oversampling and Bias-to-Close-Station—are employed, as detailed below.

3.2.1. Magnitude Oversampling

For earthquakes with magnitude ≥4.0, samples are replicated according to their magnitude using a heuristic oversampling scheme designed to mitigate the strong imbalance toward smaller events. Larger-magnitude earthquakes are therefore more frequently sampled during training.

$$\text{Repeat time} = \text{int}\left(\alpha^{\text{mag}-1} - 1\right), \text{mag} \geq 4$$

where α is the oversampling factor (set to 1.25) and mag is the event magnitude. Larger events are replicated more frequently, increasing their presence during training.

3.2.2. Bias-to-Close-Station

If more than 25 stations are triggered by a single event—indicating wide amplitude propagation—the nearest 25 stations (most likely to record high intensities) are repeatedly sampled across training batches when the triggered station count exceeds 25, thereby increasing the frequency of high-intensity samples (Figure 5).

Together, these strategies effectively increase the proportion of intensity ≥4 samples in the data set, helping to mitigate systematic underestimation of high-intensity events. However, to counter overfitting due to sample replication, random time and station masking are employed to enhance sample diversity and improve model generalization.

4. Models and Results

This chapter compares the performance of the three TT-SAM model variants in terms of intensity prediction and lead time. First, a baseline model using only velocity waveforms is evaluated on the 2016 Meinong earthquake to assess its predictive performance; next, low-frequency waveforms obtained via a 0.33 Hz low-pass filter are added to evaluate the improvement in high-intensity event detection provided by long-period signals; and finally,

10.1029/2025JH001005

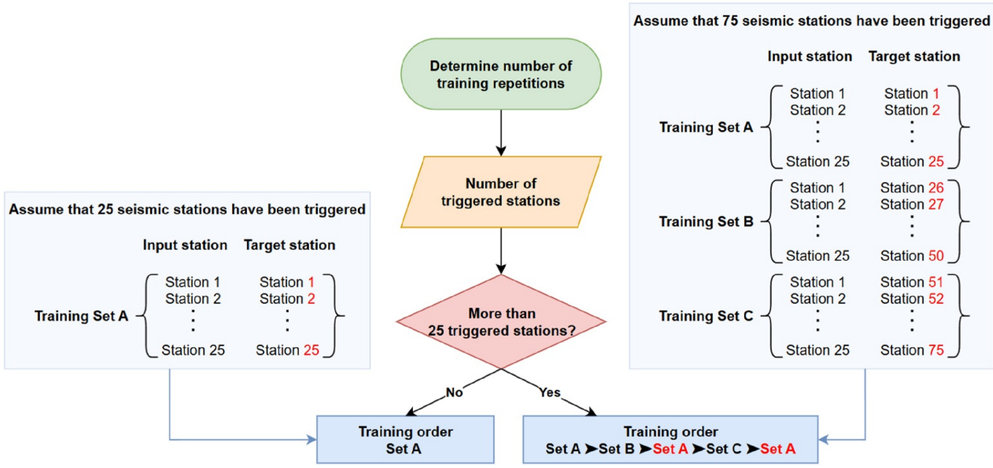

Figure 5. Illustration of the Bias-to-Close-Station strategy. Stations are sorted by P-wave arrival time and grouped into sets of 25, each set serving as input and target stations, and if more than 25 stations are triggered by a single event, the first group (Set A in the figure, i.e., the 25 stations with the earliest P-wave arrivals) is reused during training, thereby increasing its appearance frequency in training to boost the proportion of high-intensity samples.

physically meaningful features are integrated to investigate how feature engineering affects warning precision and recall. To validate the performance of the EEW system, the average lead time is back-tested on two earthquakes with magnitude greater than 6.0.

4.1. Velocity Waveform

To validate model performance, this study uses the 2016 events as the test data set. Figure 6 summarizes the intensity prediction performance of the three model variants using confusion matrices and predicted-versus-observed scatter plots. The forecast is generated 7 s after P-wave arrival at the station with the highest accuracy, and the predicted values are compared against observed intensities (CWA intensity). Model performance is evaluated using multiple evaluation metrics (Table 2). The results show that 45.2% of predicted intensities exactly match the observations, and 87.6% fall within ±1 intensity unit of the actual values (Figure 6a). Notably, the model exhibits systematic underestimation for events with observed intensity above 3 (Figures 6a and 6b). In the intensity range of 3–5+, some samples are underestimated as low as intensity 1; these cases predominantly originate from the 2016 M_L6.6 Meinong earthquake. The underestimation may stem from the multi-fault rupture characteristics of the Meinong earthquake (Y. Y. Lin et al., 2018), where the main energy release is not concentrated at the event onset but distributed in later waveform segments, causing the model to miss key intensity features during the early phase.

Additionally, the TT-SAM model allows users to set different PGV thresholds for warning triggers based on the application context. An event is considered warranting a warning when the predicted PGV exceeds the specified threshold. Given varying intensity tolerance requirements across applications, the model's threshold can be flexibly adjusted to suit user-specific needs. According to the warning criteria of the Public Warning System, an alert is issued for events with estimated magnitude >5.0 and predicted intensity ≥4, or magnitude >6.5 and intensity ≥3. Using intensity 3 as the warning threshold, the model achieves a precision of 70.7% but a recall of only 5.5% (Figure 7a). These results indicate that although the model is relatively accurate when issuing warnings, it fails to predict most actual warning-worthy events, reflecting very limited coverage and low sensitivity.

Therefore, using only velocity waveforms as input can maintain high precision but poses a serious risk of missed warnings during strong events, which limits its practical applicability in EEW systems.

4.2. Velocity Waveform With Low-Frequency Component

In addition to velocity waveforms, low-frequency velocity waveforms processed with a 0.33 Hz low-pass filter are incorporated as inputs to enhance the model's sensitivity to long-period signals during seismic events. The evaluation method is the same as in Section 4.1. Results show that the proportion of predictions exactly matching

10.1029/2025JH001005

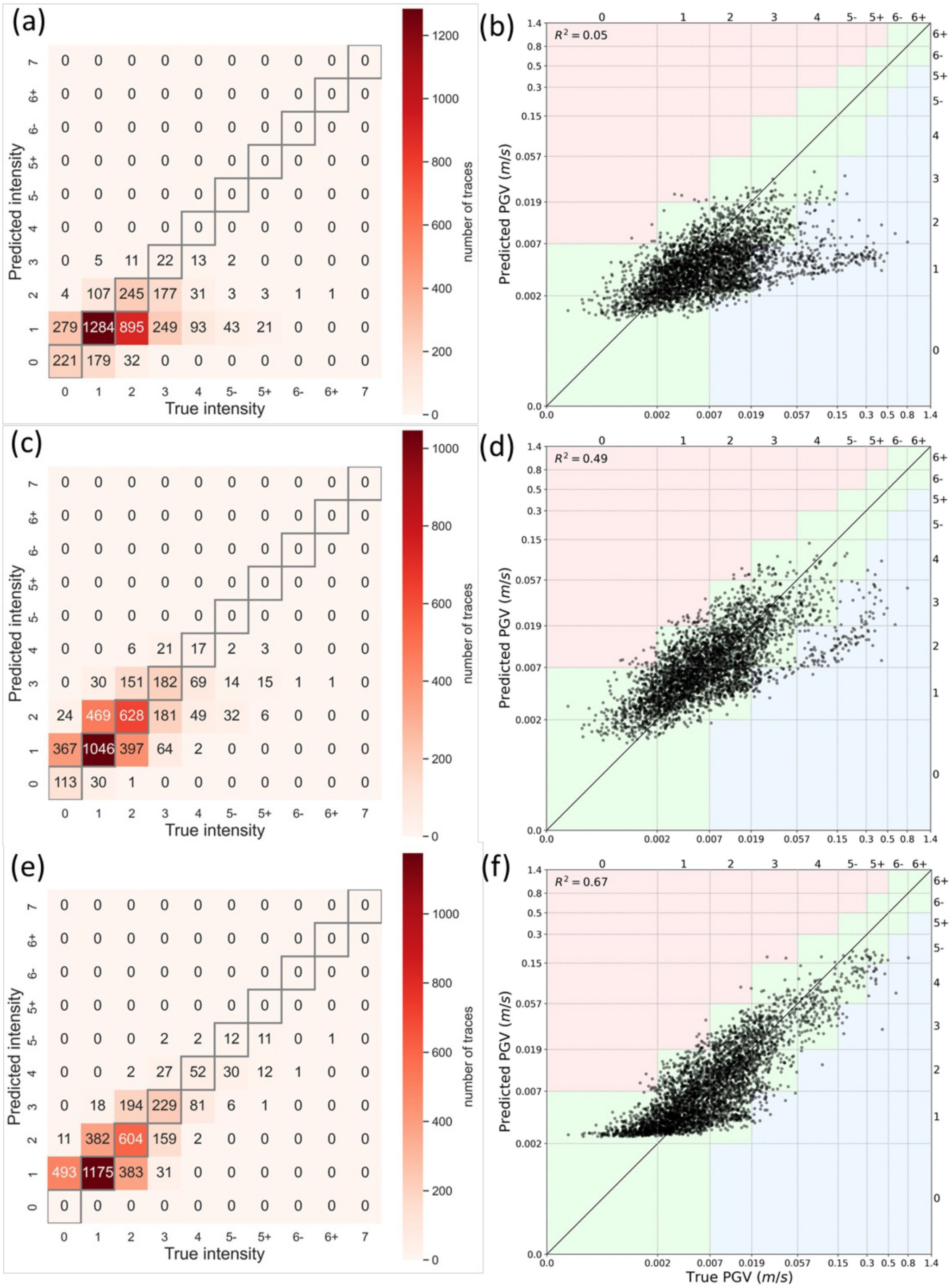

Figure 6. Overall performance of the three models on the intensity prediction task. (a), (c), and (e) are confusion matrices of predicted versus observed intensities, with each cell showing the number of samples for the corresponding intensity pairing. (b), (d), and (f) are scatter plots of predicted versus observed intensities, with the diagonal representing the ideal prediction line and R^2 indicating their correlation. Panels (a) and (b) correspond to the model using only velocity waveforms, which shows overall underestimation with $R^2 = 0.05$. Panels (c) and (d) correspond to the model with added low-frequency velocity waveforms, which improves accuracy but still underestimates high intensities, yielding $R^2 = 0.49$. Panels (e) and (f) correspond to the model further incorporating Pd, CVAV, and TP features, which significantly increases predictions within ±1 intensity, achieves the highest overall accuracy, and raises R^2 to 0.67.

Table 2
Performance Comparison of Different Models Using Multiple Evaluation Metrics

	R^2	RMSE	MAE	F1-score	Precision	Recall
Velocity	0.048	0.497	0.363	0.104	**0.698**	0.056
Low Frequency	0.493	0.363	0.278	0.556	0.635	0.495
Full Model	**0.669**	**0.293**	**0.236**	**0.699**	0.685	**0.712**
Without V_{s30}	0.591	0.326	0.249	0.655	0.747	0.583

Note. Regression performance is assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), while classification performance is evaluated using F1-score, precision, and recall. The best performance for each evaluation metric is shown in bold.

the observed intensity increased from 45.2% in the previous model (Section 4.1) to 50.7%, and the percentage of predictions within ±1 intensity also rose from 87.6% (Figure 6a) to 93.7% (Figure 6c).

Further analysis of high-intensity events reveals that the addition of low-frequency velocity waveforms helps alleviate the systematic underestimation observed in the original model for events above intensity 3. Data points previously underestimated as intensity 1 in the initial model are predicted closer to their true intensity after incorporating low-frequency components, indicating a modest improvement in the model's performance on high-intensity samples, although underestimation persists overall (Figure 6d). Including low-frequency components reduces misclassification caused by small initial waveform amplitudes, thereby improving strong-motion prediction performance. Moreover, using intensity 3 as the warning threshold, the model achieves a precision of 63.3% and a substantially increased recall of 49.2% (Figure 7b), demonstrating superior predictive capability compared to the velocity-only model.

Overall, the second model yields a significant improvement in recall, albeit with slightly lower precision than the first model, but offers greater stability. Analyzing trends across different PGV thresholds reveals that precision increases once the threshold exceeds 4 cm/s. This pattern indicates that at higher thresholds, the model adopts a more conservative prediction approach, reducing the number of warnings and thus enhancing prediction precision, but at the potential cost of increased missed warnings, highlighting the trade-off between false alarms and missed detections.

4.3. Incorporation of Physical Features

Building on the velocity and low-frequency velocity waveforms, we systematically evaluated several waveform-derived physical features to identify those suitable for real-time earthquake early warning. Among the tested parameters, the model incorporating cumulative vertical absolute displacement (CVAD) achieved moderate classification capability (F1-score = 0.565). However, because CVAD requires double integration of acceleration, its value cannot be reliably obtained in real time and therefore does not meet the operational requirements of an EEW system. In contrast, cumulative vertical absolute acceleration (CVAA) shows very limited predictive capability (F1-score = 0.109). The cumulative vertical absolute velocity (CVAV), while simpler to compute and

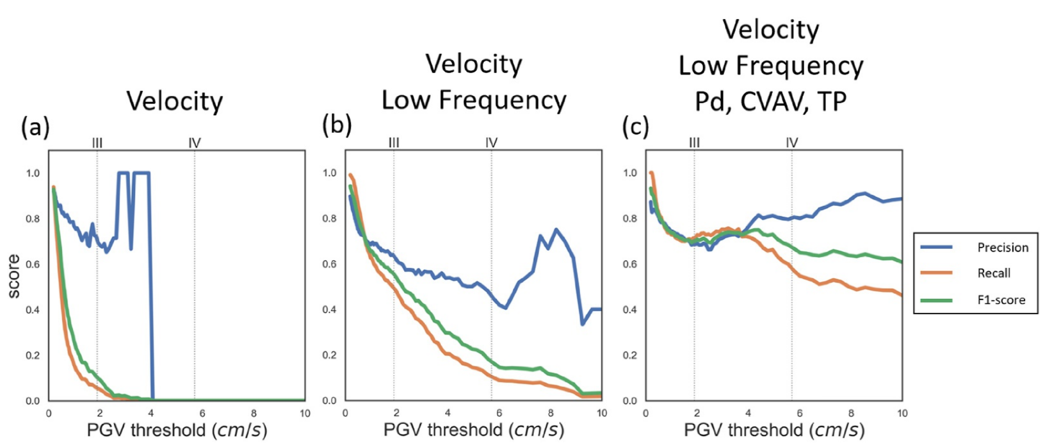

Figure 7. The warning performance of the three models (indicated in the figure) at their respective optimal forecast times: 7, 7, and 13 s. Blue, orange, and green curves represent precision, recall, and F1-score trends, respectively, as the PGV threshold varies. Dashed vertical lines indicate the PGV thresholds corresponding to intensity levels 3 and 4. Overall, F1-score and recall decrease as the threshold increases and improve with later forecast times; the third model demonstrates the most stable performance across all thresholds and times. Precision, recall, and F1-score are defined with respect to the issued warning decisions based on the selected PGV thresholds.

10.1029/2025JH001005

physically consistent with the PGV prediction target, provides substantially improved performance compared to CVAA (F1-score = 0.348), making it more suitable for real-time implementation.

We also examined energy-related features such as the integral of velocity squared (IV2). Although a model using IV2 alone performs well (F1-score = 0.753), experimental results indicate that incorporating IV2 together with other physical features leads to degraded overall performance, likely due to feature redundancy and numerical dominance. Considering both predictive performance and operational feasibility, IV2 was therefore excluded from the final feature set.

Based on these analyses, our model further incorporates three waveform-derived physical features—Pd, CVAV, and TP—which together provide complementary information on amplitude, energy accumulation, and frequency content while remaining compatible with real-time EEW constraints. These features serve as transformed representations of the original waveforms and enhance the model's ability to capture the physical characteristics of earthquake rupture and the resulting ground motion intensity.

The training and evaluation procedures remain identical to those previously described, with 2016 events as the test set and forecasts made 13 s after P-wave arrival at the first triggered station. Results indicate that the proportion of exact matches between predicted and observed intensities increases to 52.8%, and the proportion within ±1 intensity rises to 97.8% (Figure 6e), the highest overall accuracy among all models tested. Further inspection of high-intensity events shows that nearly all samples fall within ±1 intensity, significantly mitigating the systematic underestimation observed in events above intensity 3 by the initial model. The correlation between predicted and observed intensities improves substantially across model configurations, as quantified by the coefficient of determination (R^2), defined as

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

where y_i denotes the observed (true) PGV, $\hat{y}_i$ represents the model-predicted PGV, and $\bar{y}$ is the mean of the observed PGV. As shown in Figures 6b, 6d, and 6f, R^2 increases from 0.05 in the baseline model using only raw waveforms to 0.49 after incorporating low-frequency components, and further to 0.67 in the proposed model, indicating a progressively stronger agreement between predicted and observed intensities and demonstrating the superior predictive capability of the final model.

Moreover, under the 13-s input condition and using intensity 3 as the warning threshold, the model achieves a precision of 68.6% and a recall of 71.4% (Figure 7c), outperforming the other two variants. At higher PGV thresholds corresponding to greater intensities (e.g., an intensity of 4 corresponds to 5.7 cm/s), precision increases to 80.2%, while recall remains at 58.0% (Figure 7c), indicating that the model can enhance accuracy without severely compromising coverage. Overall, the third model strikes a favorable balance between precision and recall, effectively managing the trade-off between false alarms and missed detections, and demonstrating strong practical applicability. The F1-score (Figure 7c) further confirms this, reaching 0.70 and 0.67 at intensity 3 and 4 thresholds, respectively—the highest among the three models. The F1-score, as the harmonic mean of precision and recall, serves as a comprehensive metric for overall warning performance, reflecting the balance between false alarm control and alert coverage.

To assess potential systematic bias, residual diagnostics were performed with respect to earthquake magnitude and source-to-station distance (Figure 8). The results indicate no clear magnitude-dependent bias within the magnitude range represented in the training data, although increased variability is observed for larger events due to the limited number of samples (Figure 8a). In addition, the residual distribution remains centered near zero across epicentral distances, suggesting that the model captures the general distance attenuation behavior of ground motion. The larger dispersion observed at greater distances is consistent with increased path and site variability (Figure 8b).

By introducing transformed physical features such as Pd, CVAV, and TP, the EEW model leverages both temporal waveform patterns and source physics, further enhancing its discrimination of intensity levels. These results validate the potential of feature engineering in EEW, and future work may explore additional physical metrics to strengthen model generalization across diverse source types. Overall analysis reveals that stepwise

10.1029/2025JH001005

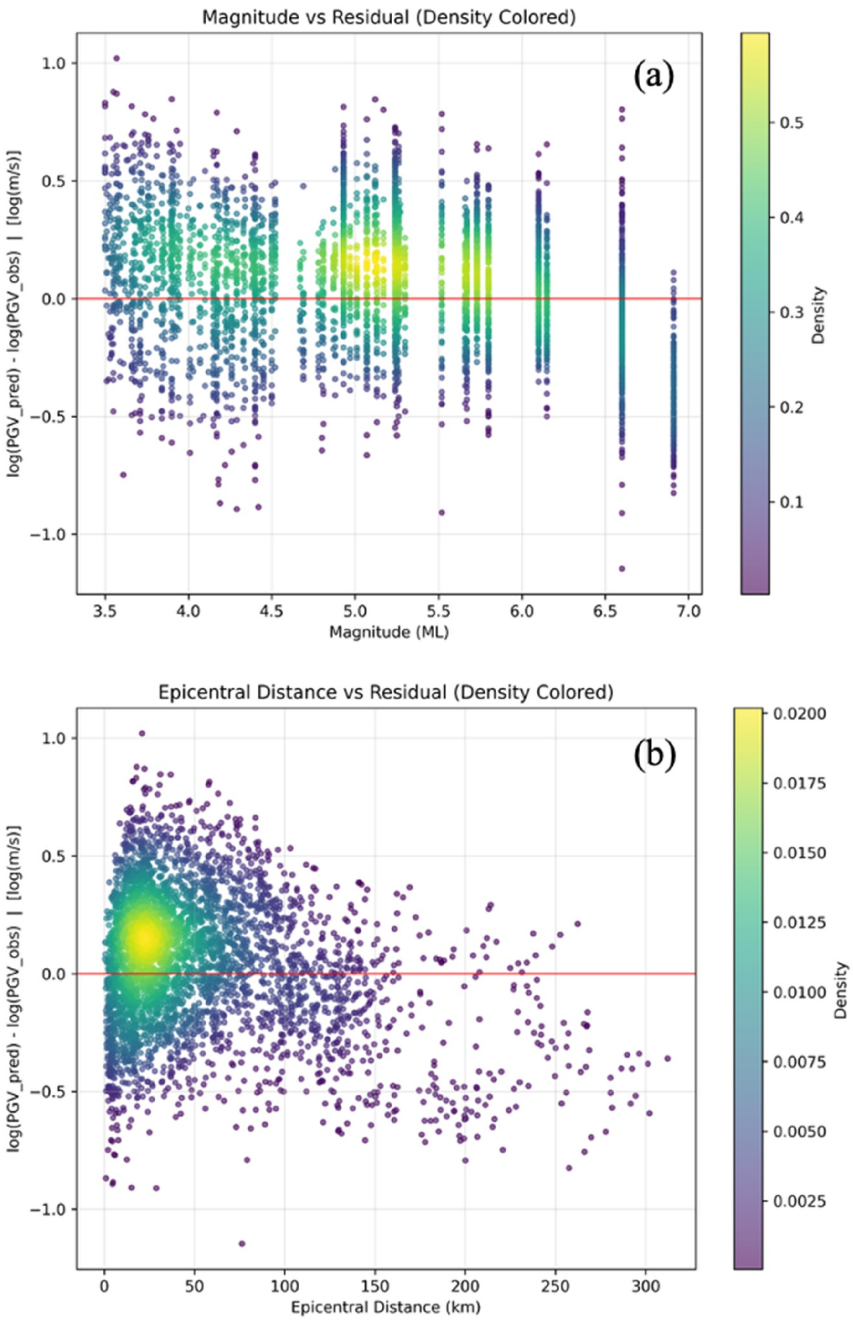

Figure 8. Relationship between prediction residuals and (a) earthquake magnitude (ML) and (b) epicentral distance. Residuals are defined as $\log(PGV_{pred}) - \log(PGV_{obs})$. Each point represents a station-level observation and is colored by the local data density estimated using kernel density estimation. The red horizontal line denotes zero residual, indicating unbiased prediction.

enhancement of input features yields clear improvements in predictive performance. While the initial model substantially reduces false alarms, its very low recall indicates inadequate detection of warning-worthy events. As low-frequency signals and multiple physical features are incorporated, both accuracy and coverage improve, with the third model achieving the best balance and the most stable overall performance. Moreover, as the warning lead time increases and more waveform data are received, both the F1-score and recall continue to rise (Figure 9). Practical deployment of EEW systems must seek an optimal balance between warning timeliness and prediction accuracy.

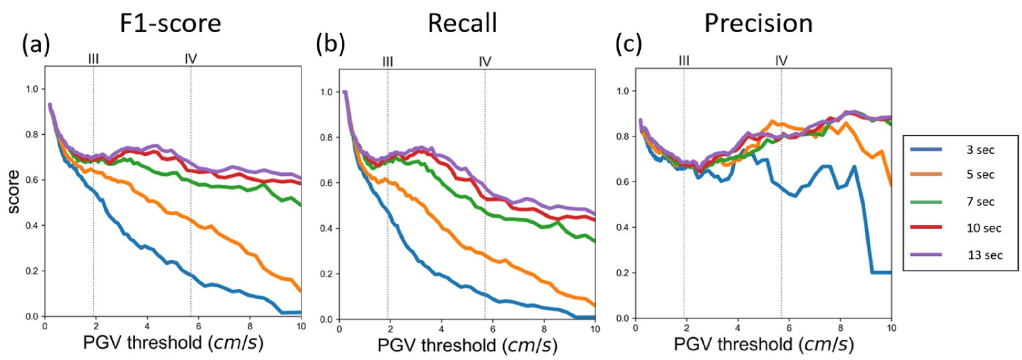

Figure 9. Warning performance of the model incorporating low-frequency waveforms and three physical features (Pd, CVAV, TP) at different lead times. The figure displays the model's (a) F1-score, (b) recall, and (c) precision curves at various lead times. The colored curves show that as lead time increases and more waveform data are available, both F1-score and recall steadily improve, indicating a positive impact of lead time on model performance.

4.4. Lead Time

In an EEW system, lead time is defined as the interval between the issuance of an alert and the arrival of peak ground velocity (PGV) at a station; the longer the lead time, the more time users have to take emergency actions. We analyze the model's performance in terms of lead time using two 2016 earthquakes with magnitudes greater than 6.0.

For alert issuance, we employ a rolling-update prediction strategy to improve efficiency. As mentioned earlier, balancing warning speed and accuracy is crucial in EEW: waiting for more data can improve precision but may deprive near-source areas of critical reaction time. Therefore, the rolling-update strategy allows the model to make rapid initial forecasts and issue alerts as soon as any station's predicted ground motion exceeds the threshold, without waiting for full convergence. As additional waveform data are received and processed, the model incrementally updates its forecasts and issues alerts to stations that have not yet been warned, thereby improving recall over time (Figure 9).

In the retrospective case of the Meinong earthquake, the model generates an initial forecast 3 s after the P-wave arrival, yielding an average lead time of 14.21 s (Figure 10a). However, only 85 stations receive a warning at that time. In this case, the rolling-update strategy proves effective: at 13 s after P-wave arrival, the model provides an average lead time of 11.29 s—slightly reduced due to the later forecast—but the number of warned stations increases to 203, balancing lead time and coverage.

In the case of the Taitung offshore earthquake, the model issues an initial forecast 3 s after the P-wave arrival, yielding an average lead time of 10.11 s and warning 32 stations (Figure 10b). However, due to the smaller affected area, most regions had already experienced PGV by 13 s after the P-wave arrival, resulting in blind zones and reducing the number of warned stations to 10. This underscores that issuing alerts in early stages of an earthquake can effectively avoid sacrificing lead time for accuracy.

5. Discussion and Conclusions

5.1. Impact of Low-Frequency Waveforms and Other Physical Features

In previous EEW studies, most classical methods rely on a single physical feature to estimate intensity or shaking severity (e.g., Brondi et al., 2015; Wu et al., 2016). In contrast, some machine learning approaches rely solely on raw acceleration or velocity waveforms as inputs, such as the TEAM model (Münchmeyer et al., 2021), thereby overlooking other representative seismic features and frequency content. In the initial phase of this research, we attempted to train the model using only velocity waveforms; however, the training results demonstrated clearly insufficient predictive performance (Table 2). Analysis of the model's prediction errors revealed that the primary issue was systematic underestimation of high-intensity events (Figure 6), a phenomenon likely related to the complex rupture characteristics of large earthquakes, for example, the M_L 6.6 Meinong earthquake.

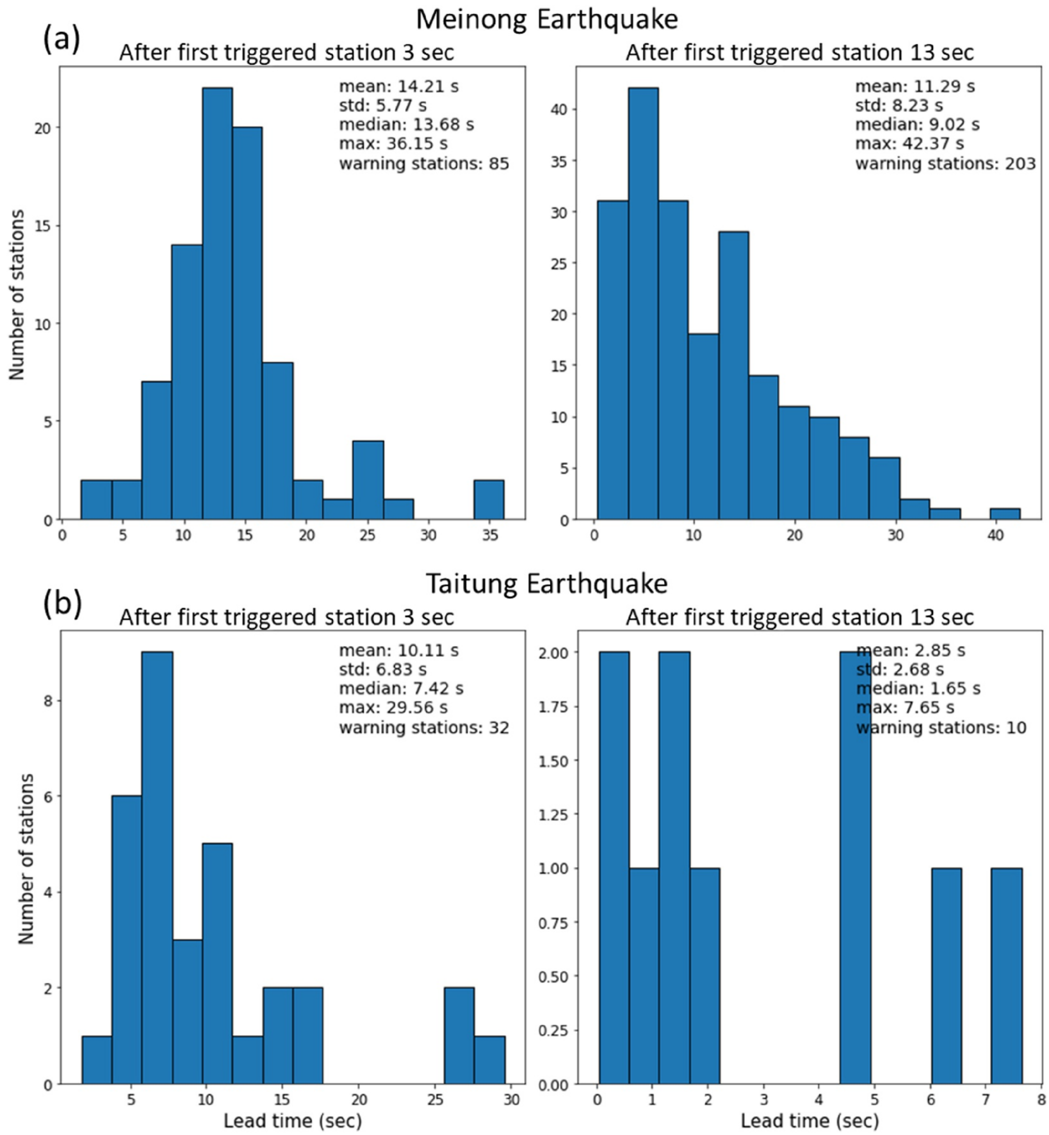

Figure 10. Statistical distributions of lead time generated by the model at 3 and 13 s after P-wave arrival for (a) the Meinong earthquake and (b) the Taitung offshore earthquake. Histograms display the lead times obtained at different stations, where lead time is defined as the interval between the model's alert issuance and the arrival of PGV at that station.

Therefore, this study incorporates low-frequency velocity waveforms as inputs; these components effectively filter out smaller earthquakes while preserving the characteristics of larger events, thus reducing misclassification (Y. Y. Lin et al., 2018). Additionally, we introduce three physically meaningful features—peak displacement, cumulative vertical absolute velocity, and the frequency-related TP—as part of a feature engineering pipeline to extract vibration patterns and rupture modes that are difficult to learn directly from raw waveforms. Experimental results show that after adding low-frequency waveforms and physical features, the model exhibits significant improvements in recall, F1-score, and R^2, with predictive accuracy markedly superior to the velocity-only model (Table 2 and Figure 7).

Taking the Meinong earthquake in the test data set as an example, the initial model exhibited clear underestimation for this event (Figure 11a), the introduction of low-frequency waveforms led to improvements (Figure 11b), and the incorporation of multiple physical features ultimately produced predictions whose overall distribution most closely matched the observed intensities (Figure 11c), confirming the utility of these physical features in recognizing strong-motion events.

10.1029/2025JH001005

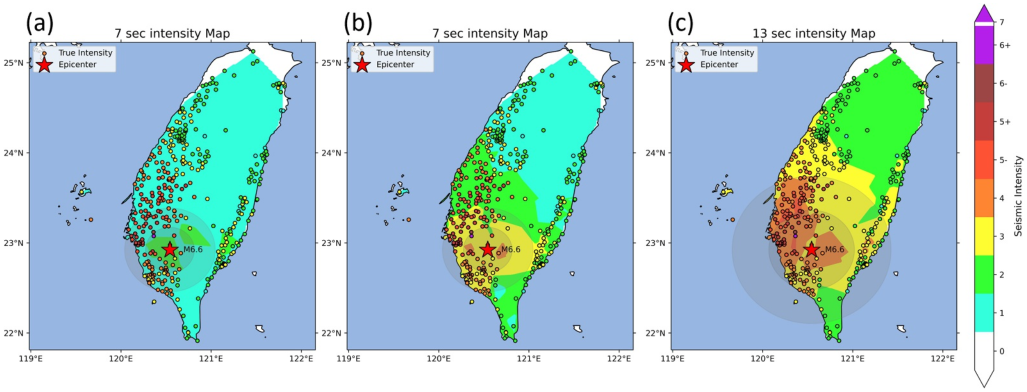

Figure 11. Comparison of intensity prediction results for the M_L 6.6 Meinong earthquake using three models: (a) velocity waveforms only; (b) velocity waveforms combined with low-frequency velocity components; and (c) integration of three physical features (Pd, CVAV, and TP). Dots denote station locations, with colors indicating the observed intensities at each station, while the background color represents the model-predicted intensity distribution. The velocity-only model exhibits clear underestimation. Incorporating low-frequency components yields a more realistic spatial intensity pattern but still underestimates shaking levels. The final model, which integrates multiple physical features, captures the full range of observed intensities more effectively and achieves the best overall predictive performance.

5.2. Impact of Training Time Window Duration on Results

To simulate ground motion prediction using seismic waveforms during the early stages of an earthquake, we implemented a time-masking mechanism during training. For each training sample, a random time between 3 and 15 s after the P-wave arrival is selected, and all waveform data beyond that point are set to zero. This strategy enhances the model's ability to forecast with limited early data and improves generalization. However, experiments reveal a potential bias: because the masking point is uniformly random, the model predominantly sees earlier segments of the waveform during training and learns less from later portions. For example, if the mask is applied at 7 s, the model only reads data up to 7 s post–P-wave, with all subsequent samples being zero (Figure 12). Over time, this leads the model to overemphasize early features and fail to learn from the complete signal.

This effect is also observed in the second model (using both velocity and low-frequency waveforms). The results show that predictions at 7 s post–P-wave outperform those at 10 s (Figure 13a) despite the latter having more complete data, indicating insufficient learning of later features. To address this issue, the third model extends the input window from 5 s before to 15 s after P-wave arrival (instead of 10 s after) (Figure 12c), thereby strengthening the model's ability to learn later-phase features. Although the third model also incorporates three physical features, experimental results show a steady improvement in performance over time, peaking at 13 s post–P-wave. Unlike the second model's performance drop at 10 s, the third model does not decline until 15 s (Figure 13b). This trend indirectly confirms the masking-induced bias. Extending the input window not only compensates for the masked segments but also broadens the forecasting horizon. Combined with the rolling-update strategy, this provides more continuous and complete prediction updates. This aligns with the continuous alert update concept seen in other modern frameworks, such as the Bayesian approach proposed by Chatterjee et al. (2022). Just as their method progressively refines predictions using event-specific residuals, our model leverages accumulating waveform data to reduce uncertainty and stabilize intensity estimates in real-time.

However, longer inputs require balancing computational efficiency and real-time applicability. While longer inputs can boost accuracy, they also increase model complexity and computational cost, and excessively delayed alerts may become impractical. Therefore, adopting a 15-s post–P-wave training window represents a pragmatic compromise between accuracy and efficiency.

5.3. Impact of Data Imbalance on Model Performance

A fundamental challenge in observation-based EEW is the strong imbalance in available training data, with small-to-moderate earthquakes occurring far more frequently than rare high-intensity events. This imbalance is inherent to natural seismicity and nearly affects all data-driven EEW studies. In the Taiwan data set used here, the majority

10.1029/2025JH001005

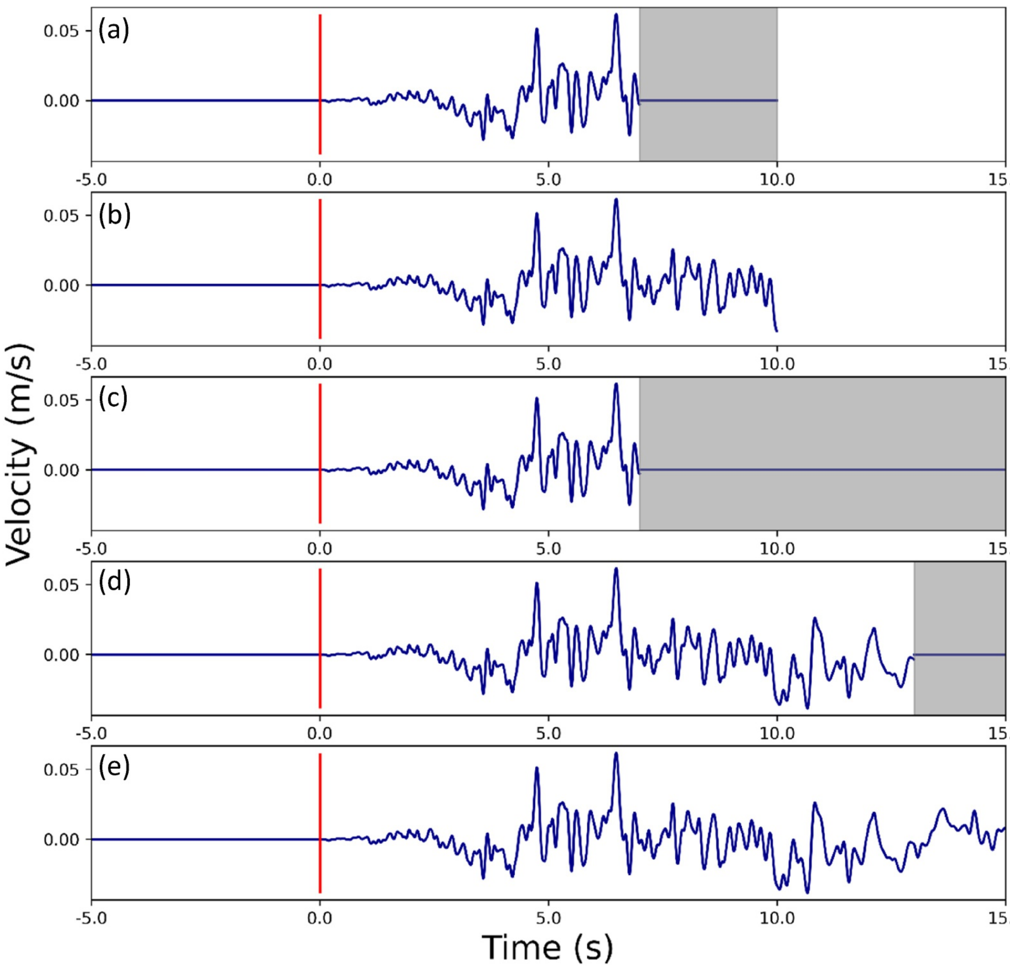

Figure 12. A single velocity waveform record from the 2016 Meinong earthquake. Panel (a) shows the waveform with masking applied beginning 7 s after the P-wave arrival, while panel (b) shows the same time window without masking. The bottom row extends the input window to 15 s after the P-wave arrival: panel (c) applies masking from 7 s, panel (d) from 13 s, and panel (e) shows the waveform without masking. The red line marks the P-wave arrival time, and the gray shaded region indicates the masked interval, where the waveform is replaced with a zero matrix.

of records correspond to lower intensity levels, while observations associated with large PGV—typically produced by large-magnitude earthquakes at short epicentral distances—are comparatively scarce (Figure 2d). Such imbalance increases the risk that data-driven models preferentially optimize performance for dominant low-intensity classes, potentially leading to systematic underestimation of strong ground motion.

To mitigate this effect, TT-SAM incorporates multiple complementary strategies during training. First, a magnitude oversampling scheme is applied to increase the effective sampling frequency of larger-magnitude events to counterbalance their natural scarcity. Second, the Bias-to-Close-Station strategy preferentially resamples the earliest triggered stations for events with widespread station activation, thereby increasing the representation of high-intensity near-source observations (Figure 5). These approaches explicitly target the underrepresentation of damage-relevant shaking without artificially altering the physical distribution of waveforms.

In addition, random time masking and station masking are employed to enhance sample diversity and reduce overfitting caused by repeated exposure to the same high-intensity records. By forcing the model to learn from incomplete waveforms and partial station sets, these mechanisms help prevent memorization of specific events and promote more robust feature extraction across varying observational conditions.

The effectiveness of these mitigation strategies is supported by independent testing on the 2016 earthquake sequence, which is excluded entirely from training. As shown in Figures 6 and 7, the model maintains robust

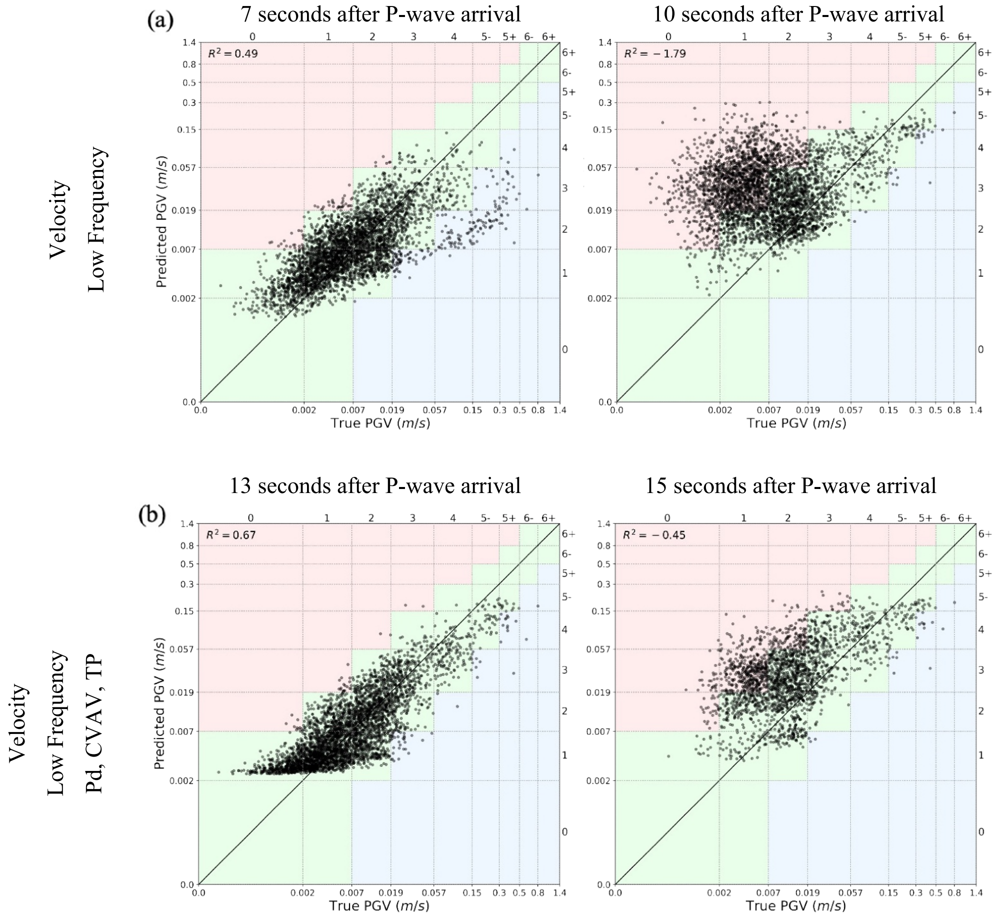

Figure 13. Comparison of PGV prediction accuracy at different lead times. (a) The model input includes raw and low-frequency velocity waveforms; the plot shows prediction results at 7 and 10 s after P-wave arrival. At 7 s, R^2 is 0.49, indicating a moderate level of predictive accuracy, but by 10 s it drops to −1.79, revealing poor learning of later waveform features and a decline in correlation between predictions and observations. (b) When three physical features (Pd, CVAV, TP) are additionally included, R^2 improves markedly, and the model can accurately predict using later-stage waveform information. However, R^2 declines again at 15 s, confirming our hypothesis that the model's ability to learn late-stage features remains limited.

performance for events not used during training, achieving 97.8% of intensity predictions within ±1 level. Notably, the model also performs well for observations with large PGV values, indicating that the learned relationships are not limited to small or moderate earthquakes at long epicentral distances. This suggests that the combination of imbalance-aware sampling and physics-guided feature design substantially reduces bias against rare, high-intensity events.

Despite these measures, we acknowledge that data imbalance cannot be fully eliminated in real-world EEW data sets, particularly for extreme events. More aggressive imbalance-aware techniques—such as focal loss (T. Y. Lin et al., 2017) or class-weighted objectives—may further enhance sensitivity to rare classes but can introduce trade-offs between recall stability and false-alarm rates in an operational setting. Given the decision-critical nature of EEW, this study prioritizes robust generalization and stable warning performance over optimizing a single class. Future work will explore alternative loss formulations and hybrid sampling strategies to further improve performance for extreme low-frequency events while preserving operational reliability.

5.4. Challenges in Cross-Regional Generalization

To critically assess model robustness, we applied our model to a case study using seismic events observed by different seismic networks. We applied our model to predict ground motion for the M_w 9.0 mainshock and the

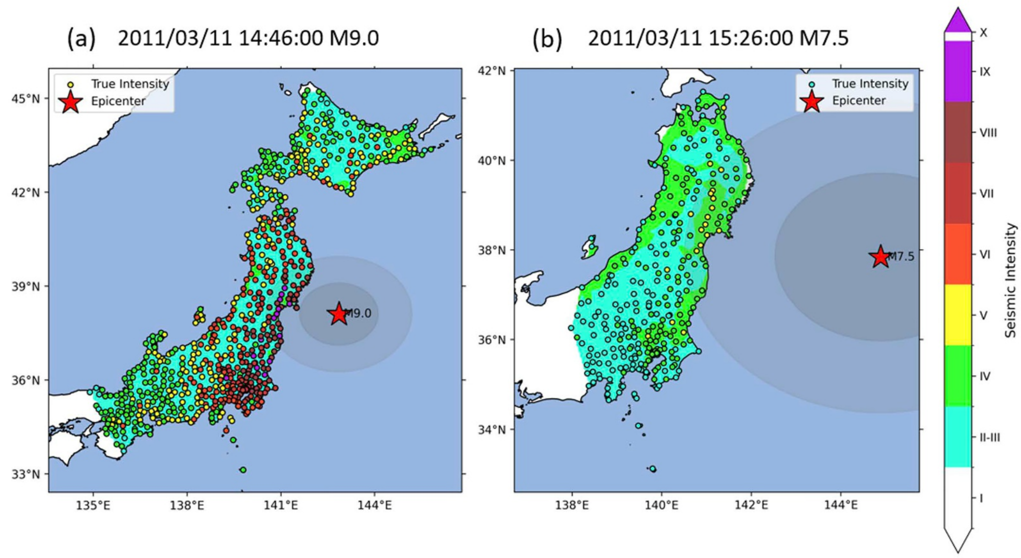

Figure 14. Case study illustrating model performance in an unseen tectonic setting (Japan). (a) Predicted intensity map for the 2011 Tohoku-Oki mainshock (M9.0) at 13 s after the initial P-wave trigger, showing a pronounced underestimation of ground motion amplitudes. (b) Predicted intensity map for the M7.7 aftershock (15:26 JST). While amplitude predictions are systematically biased low, the model maintains the relative spatial structure of shaking, including the concentration of stronger motion in the southeastern offshore region.

major M_w 7.7 aftershock of the 2011 Tohoku-Oki earthquake sequence. For the mainshock (Figure 14a), intensity predictions based on waveforms truncated at 13 s after the initial P-wave arrival exhibit substantial underestimation across the affected region. For the major aftershock (Figure 14b), while the model manages to capture the relative spatial distribution of shaking—particularly in the southeastern region of Honshu where the predicted pattern aligns somewhat with observations—the absolute intensity levels remain systematically suppressed compared to the ground truth. This indicates that while the model retains some capacity for spatial feature extraction, it suffers from a global amplitude scaling bias when applied to the target domain, especially for high intensities.

Despite high precision in Taiwan, our model systematically underestimated ground motion amplitudes in the Japanese data set. This highlights the challenge of domain shift, consistent with Münchmeyer et al. (2022), who demonstrated that regional variations in noise and network characteristics limit cross-dataset transferability for phase picking. Our findings suggest that these generalization barriers extend to regression tasks in earthquake early warning.

A likely contributing factor driving this underestimation is seismic network geometry. Taiwan's dense network ensures that the fixed number of earliest-triggered stations are typically clustered near the epicenter. In contrast, Japan's larger inter-station spacing means that later inputs in the sequence often originate from significantly greater distances with naturally weaker amplitudes. Trained on dense-network data, the model misinterprets this distance-induced attenuation as a weaker seismic source, leading to the observed systematic underestimation.

In addition to highlighting cross-regional generalization challenges, the Japan case study can be interpreted as a realistic robustness test under low-density seismic network conditions. Compared to Taiwan's dense strong-motion network, the Japanese station configuration features larger inter-station spacing and fewer near-source observations, resembling scenarios with sparse networks or effective station loss. In this context, the observed amplitude underestimation reflects the combined effects of domain shift and reduced station density. Notably, despite this bias, the model preserves the relative spatial pattern of shaking, indicating that the time- and station-masking strategies employed during training provide partial robustness in extracting spatial features under data-limited conditions. These results underscore both the resilience and the limitations of the proposed approach when applied to networks with substantially different geometries.

Furthermore, differences in tectonic setting and site conditions exacerbate this bias. While V_{s30} serves as a coarse proxy for site effects, the model likely implicitly encodes Taiwan-specific path-dependent attenuation patterns during training. Applying these learned representations to Japan's distinct tectonic setting constitutes an extrapolation beyond the training distribution, potentially resulting in the observed predictive bias.

5.5. Real-Time Performance and Operational Feasibility of PGV-Based Earthquake Early Warning

For an earthquake early warning (EEW) system to be operationally viable, the prediction model must satisfy strict real-time constraints, including low inference latency, modest hardware requirements, and stable memory usage. In this study, we evaluate the computational performance of the proposed PGV-based model to assess its suitability for real-time EEW deployment.

All performance benchmarks were conducted using the independent test data set on a workstation equipped with an Intel Core i9-14900KF CPU (24 cores, 32 threads), an NVIDIA GeForce RTX 5080 GPU with 16 GB VRAM, and 64 GB of system memory. Under this configuration, the model demonstrates highly efficient inference performance. The average inference time per event is 9.43 ms, while the maximum observed latency is 296 ms. Importantly, the 99th percentile inference time is only 12.17 ms, indicating that nearly all events can be processed within a few tens of milliseconds even under unfavorable conditions.

Memory consumption during inference is minimal. The peak GPU memory usage is 112.63 MB, which is negligible relative to the available GPU resources and allows concurrent execution with other operational EEW processes. The model contains 7,084,747 trainable parameters, corresponding to a compact model size of 27.03 MB, facilitating rapid loading and deployment across different computing environments.

These results demonstrate that the proposed PGV-based model meets the stringent latency and resource constraints required for operational EEW systems. The combination of low inference time, small memory footprint, and moderate model size ensures that the system can deliver timely ground-motion predictions within the critical early-warning window.

While PGA has been widely adopted in traditional EEW systems, its primary advantage lies in rapid detectability rather than direct relevance to damage or response decisions. In contrast, PGV has long been recognized in seismological and engineering practice as a more robust proxy for structural damage, seismic loss, and macroseismic intensity, particularly under moderate-to-strong shaking conditions (Baltay & Hanks, 2014; Wald et al., 1999). The practical motivation for adopting PGV as the prediction target in TT-SAM is therefore rooted not only in physical considerations but, more importantly, in operational verification and real-world warning workflows.

In Taiwan, the operational earthquake intensity scale issued by the CWA explicitly transitions from PGA-based definitions at low intensity levels (≤4) to PGV-based thresholds for intensity levels ≥5, which correspond to damage-relevant shaking and public warning issuance (Wu et al., 2004). As a result, PGV serves as the key ground-motion metric that directly governs warning decisions, emergency response activation, and post-event impact assessment. By predicting PGV rather than PGA, TT-SAM aligns its output variable with the quantity that is already embedded in Taiwan's operational EEW and public warning systems.

Importantly, the evaluation strategy adopted in this study is designed to reflect this operational context. Model performance is not assessed solely using framework-based regression metrics such as R^2 or loss functions, but instead through decision-oriented measures—precision, recall, F1-score, warning coverage, and lead time (Table 2, Figures 7 and 9)—defined with respect to intensity-based warning thresholds. These metrics directly quantify the system's ability to issue timely and reliable warnings under realistic operational criteria, balancing false alarms against missed detections. The improved recall and F1-score achieved by the PGV-based TT-SAM model therefore translate into tangible gains in warning effectiveness, rather than abstract improvements in model accuracy.

Furthermore, the use of observed CWA intensity levels as the reference for validation constitutes an external and operationally meaningful benchmark. Because these intensity products are generated independently by the national warning authority and used in real-time decision-making, consistency between predicted PGV-derived intensities and observed intensity distributions provides a form of external validation beyond internal model diagnostics. This distinction is critical because the demonstrated advantage of PGV in this study arises from its closer correspondence with operational intensity products and warning outcomes, rather than from properties intrinsic to the deep learning framework itself.

While PGA-based systems may retain advantages for rapid detection and near-source alerting, particularly for low-intensity shaking, PGV-based prediction offers superior stability for damage-relevant warning decisions at regional scales. The results presented here suggest that PGV-oriented EEW frameworks are better suited for end-to-end warning pipelines that extend from early detection to actionable public alerts. Accordingly, the practical superiority of PGV in this study should be understood not as a universal replacement of PGA but as an operationally justified choice for warning systems focused on impact-relevant ground motion and reliable decision support.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The earthquake catalog and waveform data used in this study are available from the Taiwan Geophysical Database Management System (GDMS; Central Weather Administration, Taiwan). V_{s30} data were obtained from the Engineering Geological Database for TSMIP (EGDT; National Center for Research on Earthquake Engineering, Taiwan) and supplemented with the nationwide V_{s30} data set provided by Lee and Tsai (2008). The source code used in this study is archived in Chen (2026a). The data set used in this work is also publicly available (Chen, 2026b). These archived versions correspond to those used to generate the results reported in this study and include the processed waveform data and associated metadata required to reproduce the experiments.

Acknowledgments
We would like to thank Editors Daniel O'Malley and John Rundle, as well as the three anonymous reviewers, for their constructive and insightful comments, which have significantly improved the quality of this manuscript. *Funding*: This research was supported by the National Science and Technology Council of Taiwan under Grants MOST 112-2116-M-008-015-MY2, MOST 114-2116-M-008 021-, and MOST 114-2124-M-865-001-. Additional financial support was provided by the Earthquake Disaster & Risk Evaluation and Management Center (E-DREaM) through the Featured Areas Research Center Program, within the framework of the Higher Education Sprout Project funded by the Ministry of Education in Taiwan.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., et al. (2023). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*. https://doi.org/10.48550/arXiv.2303.08774

Akazawa, T. (2004). A technique for automatic detection of onset time of P-and S-phases in strong motion records. In *13th World Conference on Earthquake Engineering* (Vol. 786, p. 786).

Allen, R. M., & Kanamori, H. (2003). The potential for earthquake early warning in *southern California*. *Science*, *300*(5620), 786–789. https://doi.org/10.1126/science.1080912

Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, *47*(1), 361–388. https://doi.org/10.1146/annurev-earth-053018-060457

Baltay, A. S., & Hanks, T. C. (2014). Understanding the magnitude dependence of PGA and PGV in NGA-West 2 data. *Bulletin of the Seismological Society of America*, *104*(6), 2851–2865. https://doi.org/10.1785/0120130283

Bishop, C. M. (1994). *Mixture density networks. (NCRG; No. 4288)*. Aston University.

Böse, M., Heaton, T. H., & Hauksson, E. (2012). Real-time finite fault rupture detector (FinDer) for large earthquakes. *Geophysical Journal International*, *191*(2), 803–812. https://doi.org/10.1111/j.1365-246X.2012.05657.x

Brondi, P., Picozzi, M., Emolo, A., Zollo, A., & Mucciarelli, M. (2015). Predicting the macroseismic intensity from early radiated P wave energy for on-site earthquake early warning in Italy. *Journal of Geophysical Research: Solid Earth*, *120*(10), 7174–7189. https://doi.org/10.1002/2015JB012367

Chatterjee, A., Igonin, N., & Trugman, D. T. (2022). A real-time and data-driven ground-motion prediction framework for earthquake early warning. *Bulletin of the Seismological Society of America*, *113*(2), 676–689. https://doi.org/10.1785/0120220180

Chen, D. Y., Lin, T. L., Wu, Y. M., & Hsiao, N. C. (2012). Testing a P-wave earthquake early warning system by simulating the 1999 Chi-Chi, Taiwan, Mw 7.6 earthquake. *Seismological Research Letters*, *83*(1), 103–108. https://doi.org/10.1785/gssrl.83.1.103

Chen, Y. H. (2026a). TT-SAM: A multi-station deep learning framework for probabilistic ground-motion prediction [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.18923222

Chen, Y. H. (2026b). Demo dataset for TT-SAM study [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.18885307

Chowdhary, K. R. (2020). Natural language processing. In *Fundamentals of artificial intelligence* (pp. 603–649). https://doi.org/10.1007/978-81-322-3972-7_19

Chung, A. I., Henson, I., & Allen, R. M. (2019). Optimizing earthquake early warning performance: ElarmS-3. *Seismological Research Letters*, *90*(2A), 727–743. https://doi.org/10.1785/0220180192

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hoshiba, M. (2013). Real-time prediction of ground motion by Kirchhoff-Fresnel boundary integral equation method: Extended front detection method for earthquake early warning. *Journal of Geophysical Research: Solid Earth*, *118*(3), 1038–1050. https://doi.org/10.1002/jgrb.50119

Hsiao, N. C., Wu, Y. M., Zhao, L., Chen, D. Y., Huang, W. T., Kuo, K. H., et al. (2011). A new prototype system for earthquake early warning in Taiwan. *Soil Dynamics and Earthquake Engineering*, *31*(2), 201–208. https://doi.org/10.1016/j.soildyn.2010.01.008

Huang, H. H., Wu, Y. M., Song, X., Chang, C. H., Lee, S. J., Chang, T. M., & Hsieh, H. H. (2014). Joint Vp and Vs tomography of Taiwan: Implications for subduction-collision orogeny. *Earth and Planetary Science Letters*, *392*, 177–191. https://doi.org/10.1016/j.epsl.2014.02.026

Huang, P. L., Lin, T. L., & Wu, Y. M. (2015). Application of $\tau c \times Pd$ in earthquake early warning. *Geophysical Research Letters*, *42*(5), 1403–1410. https://doi.org/10.1002/2014GL063020

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980

10.1029/2025JH001005

Kodera, Y., Yamada, Y., Hirano, K., Tamaribuchi, K., Adachi, S., Hayashimoto, N., et al. (2018). The propagation of local undamped motion (PLUM) method: A simple and robust seismic wavefield estimation approach for earthquake early warning. *Bulletin of the Seismological Society of America*, *108*(2), 983–1003. https://doi.org/10.1785/0120170085

Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., et al. (2020). Earthquake early warning ShakeAlert 2.0: Public rollout. *Seismological Research Letters*, *91*(3), 1763–1775. https://doi.org/10.1785/0220190245

Kuo, C. H., Wen, K. L., Hsieh, H. H., Lin, C. M., Chang, T. M., & Kuo, K. W. (2012). Site classification and V_{s30} estimation of free-field TSMIP stations using the logging data of EGDT. *Engineering Geology*, *129*, 68–75. https://doi.org/10.1016/j.enggeo.2012.01.013

Lara, P., Bletery, Q., Ampuero, J. P., Inza, A., & Tavera, H. (2023). Earthquake early warning starting from 3 s of records on a single station with machine learning. *Journal of Geophysical Research: Solid Earth*, *128*(11), e2023JB026575. https://doi.org/10.1029/2023JB026575

LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (2012). Efficient BackProp. In G. Montavon, G. B. Orr, & K. R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 9–48). Springer. https://doi.org/10.1007/978-3-642-35289-8_3

Lee, C. T., & Tsai, B. R. (2008). Mapping V_{s30} in Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences*, *19*(6), 671–682. https://doi.org/10.3319/TAO.2008.19.6.671(PT)

Lin, J. T., Melgar, D., Thomas, A. M., & Searcy, J. (2021). Early warning for great earthquakes from characterization of crustal deformation patterns with deep learning. *Journal of Geophysical Research: Solid Earth*, *126*(10), e2021JB022703. https://doi.org/10.1029/2021JB022703

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2980–2988). https://doi.org/10.1109/ICCV.2017.324

Lin, Y. Y., Yeh, T. Y., Ma, K. F., Song, T. R. A., Lee, S. J., Huang, B. S., & Wu, Y. M. (2018). Source characteristics of the 2016 Meinong (ML 6.6), Taiwan, earthquake, revealed from dense seismic arrays: Double sources and pulse-like velocity ground motion. *Bulletin of the Seismological Society of America*, *108*(1), 188–199. https://doi.org/10.1785/0120170169

Liu, K. S., Shin, T. C., & Tsai, Y. B. (1999). A free-field strong motion network in Taiwan: TSMIP. *Terrestrial, Atmospheric and Oceanic Sciences*, *10*(2), 377–396. https://doi.org/10.3319/TAO.1999.10.2.377(T)

McBride, S. K., Smith, H., Morgoch, M., Sumy, D. F., Jenkins, M., Peek, L., et al. (2022). Evidence-based guidelines for protective actions and earthquake early warning systems. *Geophysics*, *87*(1), WA63–WA75. https://doi.org/10.1190/geo2021-0222.1

Meier, M. A., Kodera, Y., Böse, M., Chung, A., Hoshiba, M., Cochran, E., et al. (2020). How often can earthquake early warning systems alert sites with high-intensity ground motion? *Journal of Geophysical Research: Solid Earth*, *125*(2), e2019JB017718. https://doi.org/10.1029/2019JB017718

Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). The transformer earthquake alerting model: A new versatile approach to earthquake early warning. *Geophysical Journal International*, *225*(1), 646–656. https://doi.org/10.1093/gji/ggaa609

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., et al. (2022). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, *127*(1), e2021JB023499. https://doi.org/10.1029/2021JB023499

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*. https://doi.org/10.48550/arXiv.1511.08458

Reed, J. W., & Kassawara, R. P. (1990). A criterion for determining exceedance of the operating basis earthquake. *Nuclear Engineering and Design*, *123*(2–3), 387–396. https://doi.org/10.1016/0029-5493(90)90259-Z

Saad, O. M., Helmy, I., Mohammed, M., Savvaidis, A., Chatterjee, A., & Chen, Y. (2024). Deep learning peak ground acceleration prediction using single-station waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–13. https://doi.org/10.1109/TGRS.2024.3367725

Sahakian, V., Baltay, A., Hanks, T., Buehler, J., Vernon, F., Kilb, D., & Abrahamson, N. (2018). Decomposing leftovers: Event, path, and site residuals for a small-magnitude Anza region GMPE. *Bulletin of the Seismological Society of America*, *108*(5A), 2478–2492. https://doi.org/10.1785/0120170376

Trugman, D. T., & Shearer, P. M. (2018). Strong correlation between stress drop and peak ground acceleration for recent *M* 1–4 earthquakes in the San Francisco Bay Area. *Bulletin of the Seismological Society of America*, *108*(2), 929–945. https://doi.org/10.1785/0120170245

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30, pp. 5998–6008). Curran Associates, Inc. Retrieved from https://papers.nips.cc/paper/7181-attention-is-all-you-need

Wald, D. J., Quitoriano, V., Heaton, T. H., & Kanamori, H. (1999). Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California. *Earthquake Spectra*, *15*(3), 557–564. https://doi.org/10.1193/1.1586058

Wu, Y. M., Hsiao, N. C., & Teng, T. L. (2004). Relationships between strong ground motion peak values and seismic loss during the 1999 Chi-Chi, Taiwan earthquake. *Natural Hazards*, *32*(3), 357–373. https://doi.org/10.1023/B:NHAZ.0000035550.36929.d0

Wu, Y. M., & Kanamori, H. (2005). Rapid assessment of damage potential of earthquakes in Taiwan from the beginning of P waves. *Bulletin of the Seismological Society of America*, *95*(3), 1181–1185. https://doi.org/10.1785/0120040193

Wu, Y. M., Liang, W. T., Mittal, H., Chao, W. A., Lin, C. H., Huang, B. S., & Lin, C. M. (2016). Performance of a low-cost earthquake early warning system (P-alert) during the 2016 ML 6.4 Meinong (Taiwan) earthquake. *Seismological Research Letters*, *87*(5), 1050–1059. https://doi.org/10.1785/0220160058

Yao, H., Zhu, D. L., Jiang, B., & Yu, P. (2020). Negative log likelihood ratio loss for deep neural network classification. In *Proceedings of the Future Technologies Conference (FTC) 2019: Volume 1* (pp. 276–282). Springer International Publishing. https://doi.org/10.1007/978-3-030-32520-6_22

Zhu, J., Li, S., & Song, J. (2022). Hybrid deep-learning network for rapid on-site peak ground velocity prediction. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–12. https://doi.org/10.1109/TGRS.2022.3230829