Attention-Based Neural Network for Onsite Peak Ground Velocity Earthquake Early Warning

Ting-Chung Huang*[1], Tzu-Ling Liu[1], Benjamin Ming Yang[1], and Yih-Min Wu[1,2]

Abstract

To improve on-site earthquake early warning for peak ground velocity (PGV), we leverage a machine learning approach. We propose a novel attention-based transformer architecture to address this challenging problem. A series of comparisons with other methods, including the traditional peak *P*-wave displacement amplitude approach and long short-term memory neural networks, is conducted. In addition, we demonstrate that the influence of building effects can be mitigated by incorporating station corrections to peak values in the seismograms as additional features during training. Finally, we discuss how the shape of the label can serve as a proxy to indicate the reliability of PGV determination within the first few seconds after the arrival time.

Cite this article as Huang, T.-C., T.-L. Liu, B. Ming Yang, and Y.-M. Wu (2025). Attention-Based Neural Network for Onsite Peak Ground Velocity Earthquake Early Warning, *Seismol. Res. Lett.* **XX**, 1–16, doi: 10.1785/0220240496.

Introduction

Earthquake early warning (EEW) relies on the speed difference between *P* waves and *S* waves ($V_P > V_S$) to deliver warnings of destructive earthquakes before significant ground motion occurs. In practice, to maximize warning time and enhance effectiveness, EEW systems must determine whether an earthquake is destructive within the first few seconds (Wu and Kanamori, 2005). This quasi-deterministic approach, which involves assessing the earthquake's destructiveness based on the initial behavior of ground motion, has been widely adopted. The parameters used for estimation may include magnitude (Wu and Zhao, 2006), peak ground acceleration (PGA), or peak ground velocity (PGV) (Chandrakumar *et al.*, 2024).

There are two types of EEW systems: regional EEW and onsite EEW. Regional EEW uses ground-motion data from stations near the hypocenter to estimate key earthquake parameters, including the earthquake's location and magnitude. After determining these parameters, the regional EEW takes advantage of the fact that internet transmission speeds are much faster than the speeds of *P* waves and *S* waves ($V_{\text{internet}} \gg V_P > V_S$), enabling it to deliver warnings to regions before the arrival of the largest seismic wave.

Regional EEW systems are highly reliable, with both low false alarm rates and low missed alarm rates. There are several operating regional EEW around the world, for example, in Japan (Kamigaichi *et al.*, 2009) and Mexico (Vaiciulyte *et al.*, 2024). However, regional EEW systems have a significant limitation: they require seismic data from multiple stations, and collecting these data take time. In addition, processing the data to estimate parameters and broadcasting warnings also requires time. As a result, issuing a regional warning typically takes ~10–20 s, depending on the event's coverage (Wu *et al.*, 2025). This delay creates what is known as the "blind zone," where regions close to the epicenter cannot receive warnings in time (Wald, 2020). The absence of warnings in the blind zone is particularly critical because these areas often experience the most severe damage.

To address the challenge of providing timely warnings within the blind zone, onsite EEW systems have been proposed and developed (Allen *et al.*, 2009; Satriano *et al.*, 2011; Wu and Mittal, 2021). These systems aim to deliver faster alerts by focusing on real-time seismic data from single stations, enabling them to function effectively in the most vulnerable areas.

The onsite EEW system uses seismic information from a single station to estimate the intensity of an upcoming wave. If the estimated wave exceeds a predefined threshold, the system directly issues a warning from that station.

The key advantages of the onsite EEW system are its speed and independence. Because the warning is issued solely based on ground motion observed at the station, the system avoids delays caused by transmitting ground-motion data and warnings back and forth. In addition, the onsite system can continue to function even during internet outages.

However, these advantages come with certain trade-offs. The primary disadvantage of the onsite EEW system is its reduced accuracy. This limitation arises from the reliance

1. Department of Geosciences, National Taiwan University, Taipei, Taiwan, https://orcid.org/0000-0002-8039-952X (T-CH); https://orcid.org/0000-0003-2954-1997 (BMY); https://orcid.org/0000-0003-4542-1741 (Y-MW); 2. Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan

*Corresponding author: tingchunghuang@gmail.com

on data from only one seismic station. Because biases in seismic records between stations cannot be averaged out, the uncertainty in ground-motion estimation is relatively high when using data from a single station.

Over the past decade, there have been significant advancements in EEW systems worldwide. Currently, operating EEW systems can be classified into three categories: (1) public alerts distributed through broadcasts or cellphones, (2) limited alerts provided to select users, and (3) systems still in the testing and development stage.

Among the three categories of EEW, Taiwan falls into the first category. The regional EEW system is operated by the official agency, the Central Weather Administration (CWA) of Taiwan. Taiwan is located in a seismically active region, experiencing ~3000 felt earthquakes per year. To mitigate earthquake damage, the development of EEW systems in Taiwan has been thriving.

The primary provider of the regional EEW system in Taiwan is the CWA. For the onsite EEW system, the *P*-alert seismic network plays a significant role. The *P*-alert network consists of a large number of low-cost microelectromechanical system (MEMS) accelerometers (Mittal *et al.*, 2021; Wu and Mittal, 2021). As of 2024, ~800 *P*-alert stations have been installed. Because the primary focus of the *P*-alert network is to provide onsite EEW to the public, the stations are deliberately located in public facilities where people gather. To minimize interference from human activity and noise, most *P*-alert stations are installed inside buildings. However, this introduces additional complexities because building effects must be addressed alongside the effects of varying geological settings.

The *P*-alert network provides onsite EEW using peak ground-motion displacement (P_d). Previous studies (Wu *et al.*, 2003, 2004; Wu and Kanamori, 2005; Wu and Zhao, 2006) have shown that the P_d measured during the first 3 s after the seismic wave's arrival correlates with the earthquake magnitude. The computational requirements for determining P_d are minimal, enabling the calculation to be performed efficiently by the station's microchip. Although the P_d indicator offers a reasonable warning time and acceptable warning accuracy (Hsieh *et al.*, 2015), its overall performance is insufficient to meet the demands of current-generation advancements.

Significant changes in the landscape of EEW systems have occurred following the implementation of the *P*-alert network. First, the intensity scale used in Taiwan has been revised by the CWA. Details about this revision were disclosed in a press release (in Chinese), and discussions regarding the new intensity scale can be found in several studies (Yang *et al.*, 2021; Liu and Wu, 2024). The updated scale incorporates both PGA and PGV to determine seismic intensity. For smaller ground motions, the revised scale primarily uses PGA, whereas for larger ground motions, it relies on PGV. This approach allows the revised scale to assign higher intensity values based on PGV, reflecting the observed correlation between PGV and building damage identified in previous studies (Yamaguchi and Yamazaki, 2001; Mittal *et al.*, 2021). Second, the computational capabilities of the *P*-alert stations are improving significantly. The next generation of *P*-alert devices is expected to feature computational power comparable to that of the fourth-generation Raspberry Pi. This substantial enhancement in processing power will enable more complex data processing and even allow the deployment of trained machine learning (ML) models.

As a response to the major changes in the landscape of onsite EEW, we aim to address the following questions in this study:

1. Because the revised intensity scale relies on PGV to identify higher intensities, can we develop an onsite EEW method for PGV that outperforms the traditional P_d method (Hsieh *et al.*, 2015)?
2. Given the upcoming upgrades in computational hardware and recent advancements in the application of ML in seismology (Li *et al.*, 2018; Mousavi and Beroza, 2020; Zhang *et al.*, 2021), can we use similar ML techniques to address the onsite EEW problem?
3. Considering the inhomogeneity between stations due to varying geological settings and installation conditions (Huang and Wu, 2023a), can we identify suitable ML features that effectively represent these differences between stations?

In this article, we use the attention mechanism (Vaswani *et al.*, 2017) to solve the onsite EEW of PGV problem. The attention method is a ML method that determines the importance of certain parts of the sequence of data. The advantage of the attention mechanism lies in its ability to identify the most crucial part of the seismic record and focus on it to improve estimation of the incoming peak velocity. We manage to answer all the aforementioned questions. The key points of our work are as follows:

- We describe in detail the proposed attention-based model for the onsite PGV EEW problem (the Methodology section).
- We demonstrate the effectiveness of including the station corrections as a feature in the training (the Results section).
- We show the test results of the proposed attention-based model against several independent medium to large events (the Results section).
- We compare the accuracy of the warning of the attention-based model with that of the other two methods, including the P_d method and the long short-term memory (LSTM) model (see the Results section).
- We discuss the shape of the optimized label can reflect the credibility of PGV determination (see the Testing with various shapes of label section).

Data

Figure 1 illustrates the dataset used in this study, including events in Taiwan, events in Japan, stations of *P*-alert network

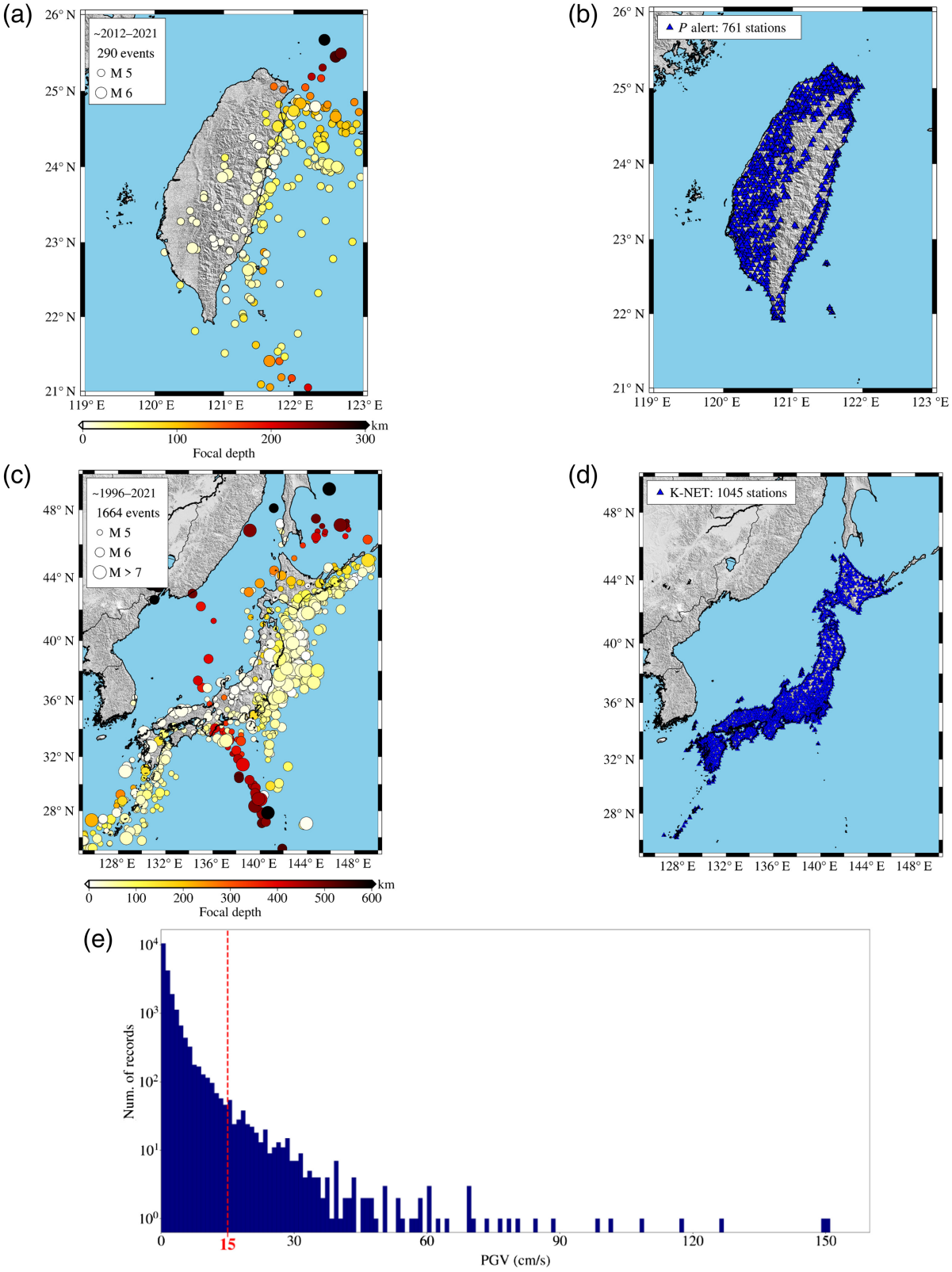

Figure 1. Illustrations of dataset used in this study. (a) 290 Taiwan events recorded by the *P*-Alert network. (b) 761 stations used in the *P*-alert network. (c) 1664 Japan events recorded by the Kyoshin-net (K-NET) network. (d) 1045 stations used in the K-NET network. (e) Histogram of the peak ground velocity (PGV) for all records in the dataset. The color version of this figure is available only in the electronic edition.

in Taiwan, stations of Kyoshin net (K-NET) network in Japan, and the PGV histogram of all the events combined.

P-alert dataset

The *P*-alert network consists of low-cost MEMS sensors designed to record real-time ground motion in Taiwan. Currently, there are ~800 operational stations, which enables the network to achieve ultra-high density. Each *P*-alert station is equipped with a three-component MEMS accelerometer capable of providing 16-bit readings at a sampling rate of 100 Hz. The amplitude range that a *P*-alert station can record spans from −2g to +2g.

For this project, we collected all events with $M_L > 5.0$ recorded by the *P*-alert network between 2012 and 2021. A total of 290 medium to large events were identified, all located within the coordinates 21°–26° N and 119°–123° E. Notice that the choice of $M_L > 5.0$ for our dataset is based on the observation that records with higher PGV are extremely rare in smaller magnitude events.

K-NET dataset

Because of the rarity of higher PGV records, we also collected K-NET data from Japan with $M_{JMA} > 5.0$ between 1996 and 2021 as part of a data enhancement and transfer learning backup plan. A total of 1664 medium to large events were included. A similar transfer learning plan has been implemented and proven successful for the PGA-based onsite EEW problem in a previous study (Wang *et al.*, 2022). Details of the performance of transfer learning using K-NET data are provided in the Discussion section, in which we compare the performance of the original training dataset (*P*-alert records only) with that of the enhanced training dataset (*P*-alert records combined with K-NET records).

Methodology

Data preprocessing

The primary goal of onsite EEW is to address the insufficient warning time within the blind zone. Therefore, we selected records from within the blind zone for training. Generally speaking, the area within 50–60 km of the epicenter is considered the blind zone for regional EEW systems in Taiwan. Note that it is a rough estimation, and the exact region of blind zone depends on several parameters, including depth, V_S, alert calculation time, broadcasting time, and more. In this study, we focused exclusively on records with an epicentral distance <70 km.

Following the data selection based on epicentral distance, a series of data quality checks was implemented. Records with missing or abnormal data (e.g., records with large noisy spikes) were excluded. After preliminary data cleaning, we proceeded to pick arrival times. We chose to use the local extrema scalogram (LES) picker (Huang and Wu, 2019) because of its robustness in noisy environments. After obtaining the preliminary *P* arrivals, we used the SNR_d method (Huang and Wu, 2023b), which calculates the signal-to-noise ratio based on the logarithm of the summation of vertical displacement squared after the *P* arrival ($\sum_{t=0}^{3} d^2[t]$) divided by the summation of vertical displacement squared before the *P* arrival ($\sum_{t=-3}^{0} d^2[t]$). Records with excessively low SNR ($SNR_d < 1$) were discarded because they were contaminated by noise. Finally, we manually verified the quality of the arrival times in the remaining records by comparing the performance of the LES picker with the traditional short-term average (STA)/long-term average (LTA) picker. A 10 s time window from the curated records was selected as a training feature. In addition, velocity and displacement values derived from the acceleration records were included as candidates for training features. Figure 2 summarizes the preprocessing procedures described earlier. The three components of velocity are derived from the three components of acceleration by integration in time and a high-pass Butterworth filter of 0.075 Hz. The three components of displacement are derived from the three components of velocity by integration in time and a high-pass Butterworth filter of 0.075 Hz.

Station corrections of initial peak ground motion

To account for the differences in ground motion at *P*-alert stations due to overall station effects, including site effects, topographic effects, and building effects, we incorporated station corrections of initial peak ground motion (Huang and Wu, 2023a) as candidates for training features. Station corrections of initial peak ground motion are statistical parameters estimated using the iterative regression (IR) method (Huang and Wu, 2021). IR treats the station corrections as missing data. IR alternatively uses a seismic dataset to perform two types of estimation, including (1) given the station corrections, estimation of attenuation model, and (2) given the attenuation model, estimation of station corrections. After several iterations, both estimations will converge to true values. The convergence of station corrections to their true values is guaranteed by the corresponding statistical theory. In practice, convergence is typically achieved within a few iterations of estimation. In this project, we included three types of station corrections as candidates for training features: the station correction of initial peak acceleration (SC_a), the station correction of initial peak velocity (SC_v), and the station correction of initial peak displacement (SC_d).

Data labeling

Data labeling is crucial in supervised learning because it provides ground truth for the given features, enabling the machine to learn. In this project, the label for onsite EEW is associated with the decision to issue a warning. In other words, the label reflects the legitimacy of sending out a warning. In practice, we consult the revised intensity scale by the CWA and set the judging criterion at PGV = 15 cm/s, which distinguishes milder ground motions from severe ground motions. Records with

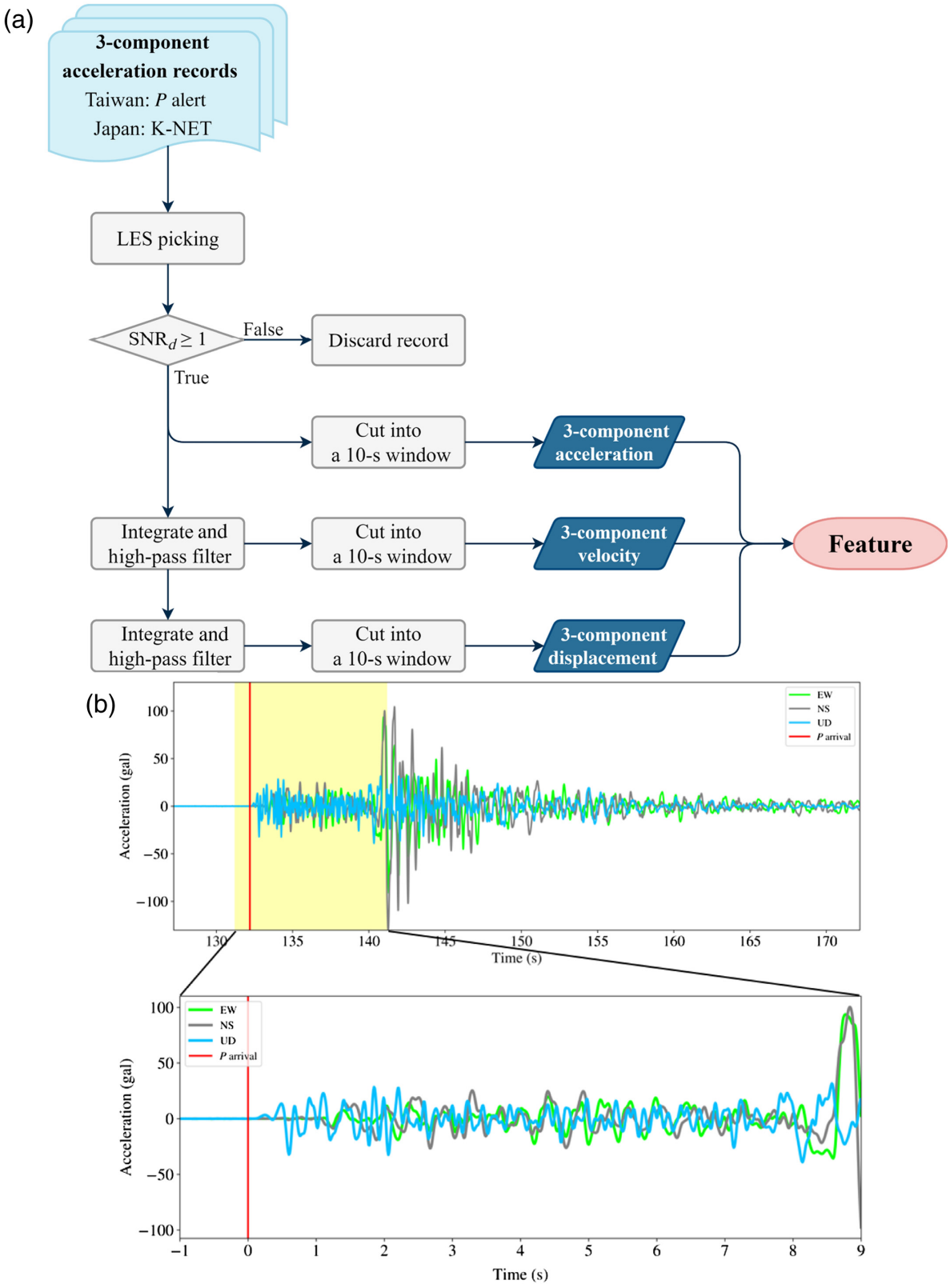

Figure 2. The preprocessing procedure and its output feature. (a) The preprocessing procedure used in this study. (b) An example of the features used for training, showing a 10 s window from one acceleration channel. EW, east–west; LES, local extrema scalogram; NS, north–south; UD, up–down. The color version of this figure is available only in the electronic edition.

PGV >15 cm/s are labeled as 1 (100% credibility to issue a warning), and the remaining records are labeled as 0 (0% credibility to issue a warning). It is important to note that we treat the records as time series, and because a warning can be issued at any time after the *P*-wave arrival, time introduces an additional degree of complexity to the labeling. For example, consider a record with PGV > 15 cm/s. Its label should also be represented as a time series. The beginning and ending of the label are straightforward: before the arrival of the *P* wave, the credibility is 0 because no information is available. After the *P*-wave arrival, the credibility is 1. An interesting question is how the label transitions from 0 to 1. To maximize lead time, we start with a step function–shaped label that jumps at the *P*-wave arrival, similar to the approach used in the case of PGA-based EEW. In the Discussion section, we compare the performance of different label functions.

Note that theoretically, it is possible to have records exceeding 15 cm/s that are outside the 70 km epicentral distance region. One may be concerned that the training dataset does not include these records because of the 70 km criterion. However, records exceeding 15 cm/s are fairly rare. In practice, we do not observe any of such records yet in the *P*-alert catalog. Therefore, we think our dataset captures most of the large PGV records.

Data imbalance and data augmentation

One of the inevitable difficulties in applying ML to the EEW problem is data imbalance, a very common challenge in ML tasks (Sun *et al.*, 2009; Fan *et al.*, 2020; Thabtah *et al.*, 2020). Specifically, if all available records are included in the training dataset, the predictions will be heavily biased toward "no warning," reflecting the rarity of large PGV records. To address this data imbalance, we use data augmentation techniques to "create" additional large PGV records. In this work, we adopt three common data augmentation methods (Fan *et al.*, 2020): (1) adding background noise to the signal, (2) shifting the arrival time to simulate the uncertainty in determining the *P*-wave arrival, and (3) flipping the signal's polarity. By combining these three data augmentation techniques, we successfully increased the number of large PGV records by eightfold.

ML methods for onsite PGV problem

In this work, we have adopted two types of ML architectures in solving the onsite PGV problem. The first one is the LSTM architecture (Sutskever *et al.*, 2014), which was used to solve the onsite PGA problem in the previous study (Wang *et al.*, 2022). The second one is an attention-based architecture designed to predict the time series based on the "context" of the time series (Vaswani *et al.*, 2017).

The LSTM model. The LSTM architecture (Hochreiter and Schmidhuber, 1997; Gers *et al.*, 2002) consists of two sets of memories: short-term memory and long-term memory. The architecture can forecast anomalies from the sequential data based on comparing long-term trends with short-term trends. In the context of using the LSTM architecture to solve the onsite EEW problem, we follow closely with our previous study by Wang *et al.* (2022), in which we proposed an LSTM model for onsite EEW.

Conceptually, the LSTM model for onsite EEW operates similarly to the STA/LTA picker in that it compares information in the long-term memory with information in the short-term memory to determine specific targets. In the case of the STA/LTA picker, the target is to distinguish seismic signals from noise, which can be achieved using a simple average of the time series. However, if the target is to identify the incoming intensity of a seismic wave, the required function would be far more complex than a simple average. This complexity makes it appealing to use a neural network to derive such a function. The LSTM model is a viable candidate because of its proven effectiveness in predicting onsite PGA. In this project, we adapt a modified LSTM model, originally designed for the onsite PGA problem, to fit our onsite PGV problem.

The attention-based model. The attention-based model (transformer layers) is one of the latest and most popular architectures developed for solving natural language processing (NLP) tasks. It can learn context by tracking specific relationships within sequential data (Luong *et al.*, 2015). The attention-based model has achieved success in many otherwise challenging tasks, including language translation (Cho *et al.*, 2014; Bahdanau *et al.*, 2015), sound-to-text conversion, and various NLP applications such as chatbots. Among all applications involving the attention-based model, Chat Generative Pre-trained Transformer (ChatGPT) is perhaps the most well-known example (Brown *et al.*, 2020). ChatGPT is a powerful chatbot capable of interpreting commands, understanding their context and meaning, and generating appropriate feedback as an output.

The ability to learn context is enabled by the mechanism of attention, which comprises an encoder and a decoder. In our project, we use a simplified version of the attention mechanism (Wu *et al.*, 2020) to enable the architecture to identify which parts of the seismic signal are significant for predicting subsequent PGV. The attention model can be viewed as a generalization of the LSTM model, extending the short-term memory of the current signal to specific, significant portions of the signal. We anticipate performance enhancements with the addition of the attention mechanism.

Figure 3 illustrates the attention-based model adopted in our study. The architecture was originally designed for time-series prediction (Wu *et al.*, 2020). In our implementation, we feed the sequential seismic signal (the x_1 to x_t in Fig. 3) into both the encoder and decoder parts of the architecture. The output is the probability of the PGV surpassing the 15 cm/s threshold (the y_1 to y_t in Fig. 3).

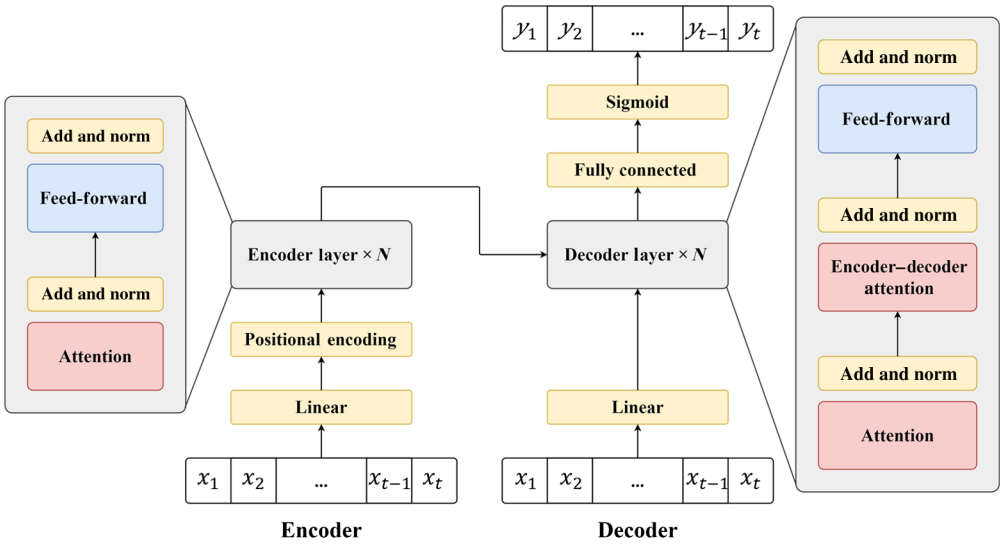

Figure 3. The proposed attention-based architecture. One encoder is on the left side, and one decoder is on the right side. The inputs are sequential seismic data $x_1 \ldots x_t$, in which the subscript denote the timesteps. The outputs are warning probabilities at every time steps $y_1 \ldots y_t$. The color version of this figure is available only in the electronic edition.

To reduce the computational complexity, we design our attention-based model to update every second, not every sampling period. For example, at $t = 2$, the model already knows the records from −1 to 1. Then we further input the records from 1 to 2. Now the model knows updated records from −1 to 2. Finally, the model will output a probability of the alert at $t = 2$. The prompt for calculating alert probability is once every second.

Results

Model score based on performance metrics

We use two metrics to evaluate performance. The first metric measures how good the alert classification is. In ML, the common used metrics are the precision and the recall, which depict the effect of type I and type II errors. They are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2)$$

in which TP denotes the true positive cases, FP denotes the false positive cases, and FN denotes the false-negative cases. To have a combined measure of the two types of errors, we used the F1-score. The F1-score is the harmonic mean of precision and recall, and it is defined as

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3)$$

The second metric assesses the effectiveness of the warnings. An effective warning should be issued early enough to provide the recipient with sufficient time for proper preparation. The effectiveness is commonly depicted through the lead time of the warnings, defined as the time difference between when the velocity exceeds the threshold and when the warning is issued. The lead time (T_{lead}) is defined as

$$T_{\text{lead}} = T_{V \geq 15\text{cm/s}} - T_{\text{alert}}, \quad (4)$$

in which $T_{V \geq 15\text{cm/s}}$ denotes the time when the velocity first exceeds 15 cm/s and T_{alert} denotes the time when the model issues alert.

Our goal is to identify a model that generates both accurate warnings and warnings issued as quickly as possible. To achieve this, we introduce a combined model score that incorporates both metrics (Wang *et al.*, 2022). The model score is defined as

$$\text{Model score} = 0.5 \times \text{PR(F1)} + 0.5 \times \text{PR(mean lead time)}, \quad (5)$$

in which the two metrics are normalized to a range of 0–100 based on their performance through a population of models with different hyperparameters and settings. The normalization is done by using percentile rank (PR), which is commonly used for student performance assessment in education (Crawford and Garthwaite, 2009). The 0.5 and 0.5 coefficients are assigned to give equal importance to both metrics. The highest attainable model score is 100, corresponding to the best possible F1-score and mean lead time. We acknowledge that there are other ways to combine different metrics into one single overall metric. Another possible approach is to use *t*-scores. However, we opted for the current design of the model score instead of *t*-scores because of the lack of normality in the distributions of both metrics.

Feature selection

Feature selection is a common practice in ML. The primary goal is to identify the features most relevant to the target variables. This process enhances model performance by eliminating irrelevant and redundant features. In this project, we adopt an exhaustive search method from the wrapper approach to determine the optimal combination of features. The dataset used in the feature selection are confined to the training dataset and the validation dataset.

We examine combinations of six types of features, including the following: three components of acceleration (A), three

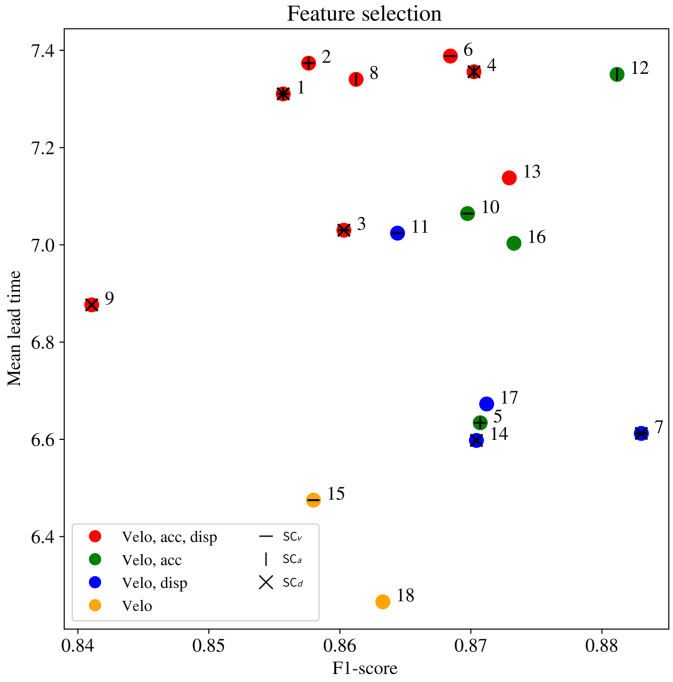

Figure 4. Distribution of performance using different subsets of features. The *x*-axis represents the F1-score, and the *y*-axis represents the lead time. A feature set with better performance corresponds to points located toward the upper-right corner of the figure. SC_a, the station correction of initial peak acceleration; SC_d, the station correction of initial peak displacement; SC_v, the station correction of initial peak velocity. The color version of this figure is available only in the electronic edition.

components of velocity (V), three components of displacement (D), station correction for initial peak acceleration (SC_a), station correction for initial peak velocity (SC_v), and station correction for initial peak displacement (SC_d).

Figure 4 presents the results of all the feature combinations considered. The figure indicates that the best F1-score performance is achieved when the model includes velocity, displacement, SC_v, and SC_d. Conversely, the best lead-time performance is observed when the model incorporates acceleration, velocity, displacement, and SC_v. Finally, the highest model score performance occurs when the model includes acceleration, velocity, and SC_a. The optimal feature combination for model score appears reasonable for several reasons: (1) it includes acceleration, which captures raw ground-motion data; (2) it incorporates velocity, a derivative of acceleration, which has also been used as a sensitivity modulator in previous studies; and (3) it includes station corrections to address the heterogeneity among *P*-alert stations. Notably, the selection of SC_a station correction reflects its compatibility with acceleration ground-motion data.

Hyperparameter selection

We use the same model score to select the hyperparameters for the architectures, including both the LSTM architecture and the attention-based architecture. The datasets used in the hyperparameter selection are confined to the training dataset and the validation dataset.

For attention-based architectures, we explore three aspects of hyperparameters: the dimension of the hidden layer, the scaling factor of the feedforward neural networks, and the number of layers in the encoder and decoder. Generally, the dimension of the hidden layer determines the resolution of the learned features; models with higher dimensions typically capture more details. Similarly, the scaling factor of the feedforward neural networks and the number of encoder and decoder layers influence the model's complexity. Although increased complexity often enhances performance, it also raises the risk of overfitting. After finding the model with the highest model score, we determined that the optimal hyperparameters are as follows:

1. Dimension of the hidden layer: 20
2. Scaling factor of the feedforward neural networks: 40
3. Number of layers in the encoder and decoder: 1

For LSTM architectures, we investigate three hyperparameter aspects: the number of hidden layers (N), the scaling ratio of dimensions between hidden layers (U), and the width factor of the final layers (i). The output dimension of the LSTM hidden layers is calculated as

$$\dim = U^{N+i}. \tag{6}$$

After evaluation, we identified the optimal hyperparameters as follows:

1. Number of hidden layers (N): 4
2. Scaling ratio of dimensions between layers (U): 2
3. Width factor of the final layers (i): 0

The performance of testing events

To ensure a fair performance assessment, we selected a testing dataset that is independent of both the training and validation datasets. The testing dataset comprises four medium to large events. Three of these events occurred on the eastern side of Taiwan, and one event took place on the western side. Notably, all the selected events have larger magnitudes because, otherwise, they would not produce ground motions with PGV > 15 cm/s.

We have selected four independent earthquake events in the testing dataset. These earthquakes represent four active seismic locations in Taiwan. The first event is the 2016 Meinong earthquake, which had a local magnitude of 6.6 and a PGA >400 gal. This earthquake caused the collapse of one building and damaged several others. The death toll was 117, making it the most destructive earthquake in Taiwan since the 1999 Jiji earthquake. The second event is the 2022 Hualien earthquake, with a local magnitude of 6.7. This earthquake caused damage to several

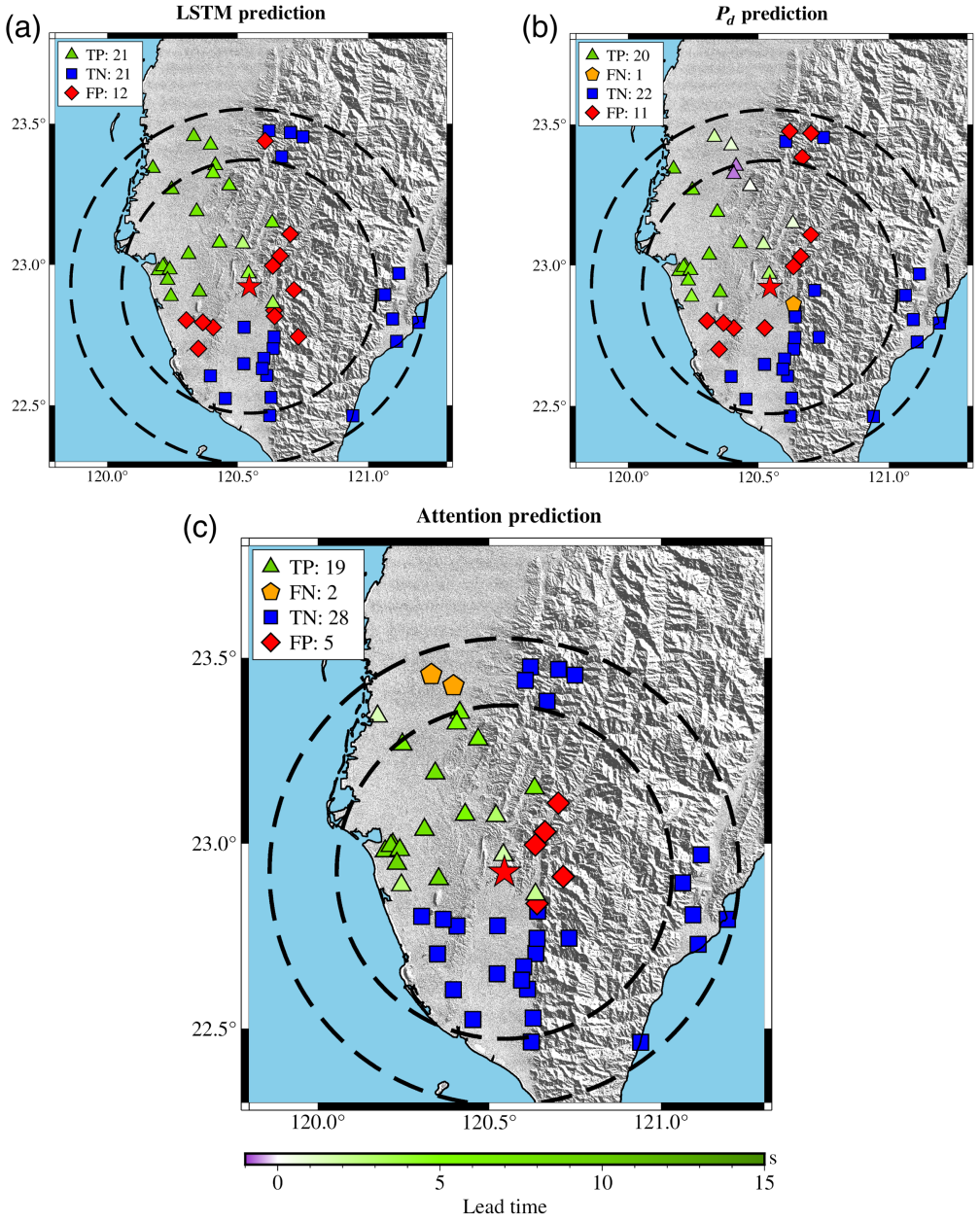

Figure 5. Earthquake early warning (EEW) performance for the 2016 Meinong earthquake. The triangles represent the true positive (TP) stations, with color fillings indicating the lead time. The squares represent the true negative (TN) stations. The diamonds represent the false positive (FP) stations. The pentagons represent the false negative (FN) stations. The two dashed circles around the epicenter (star) illustrate the 50 and 70 km radii around the epicenter. (a) The results of the long short-term memory (LSTM) model. (b) The results of the P_d method. (c) The results of the proposed attention-based model. The color version of this figure is available only in the electronic edition.

buildings, bridge failures, and gas leaks. The third event is the 2022 Guanshan earthquake, which had a local magnitude of 6.6. The fourth and final event is the 2022 Chishang earthquake, with a local magnitude of 6.8. The 2022 Guanshan earthquake is considered the largest foreshock of the 2022 Chishang earthquake. It is important to note that the last three events occurred outside the temporal scope of the training dataset, ensuring there is no data leakage. The first event was deliberately chosen to include a representative earthquake from western Taiwan. We combine all the records in all the testing events into one whole testing dataset, without extra weighting imposed.

Figure 5 illustrates the warning results for the 2016 Meinong earthquake. The triangle labeled TP represents stations with TP warnings, meaning that the PGV exceeds 15 cm/s, and the system successfully issues a warning. The color filling of the triangle indicates the lead time for each station. The pentagon labeled FN represents stations with FN warnings (missed alarms), in which the PGV exceeds 15 cm but the system fails to issue a warning. The square labeled TN represents stations with true negative warnings, indicating that the PGV does not exceed 15 cm and the system correctly does not issue a warning. The diamond labeled FP represents stations with FP warnings (false alarms), in which the PGV does not exceed 15 cm, yet the system still issues a warning. The F1-scores for the three models—attention-based model, LSTM model, and P_d method—are 0.844, 0.778, and 0.769, respectively. The mean lead times (arithmetic mean of the TP cases) for the three models are 5.95, 7.03, and 4.18 s, respectively. Notably, we examined all five FP cases in the attention-based model and found that they all exhibited PGA values exceeding 80 gal. These FP cases are considered challenging to classify because of the ambiguity between PGV-based and PGA-based intensity measures.

Figure 6 presents the warning results for the 2022 Hualien earthquake. The F1-scores for the three models—attention-based model, LSTM model, and P_d method—are 0.727, 0.586, and 0.692, respectively. The mean lead times (arithmetic mean of the TP cases) for the three models are 6.78, 7.12, and 5.62 s,

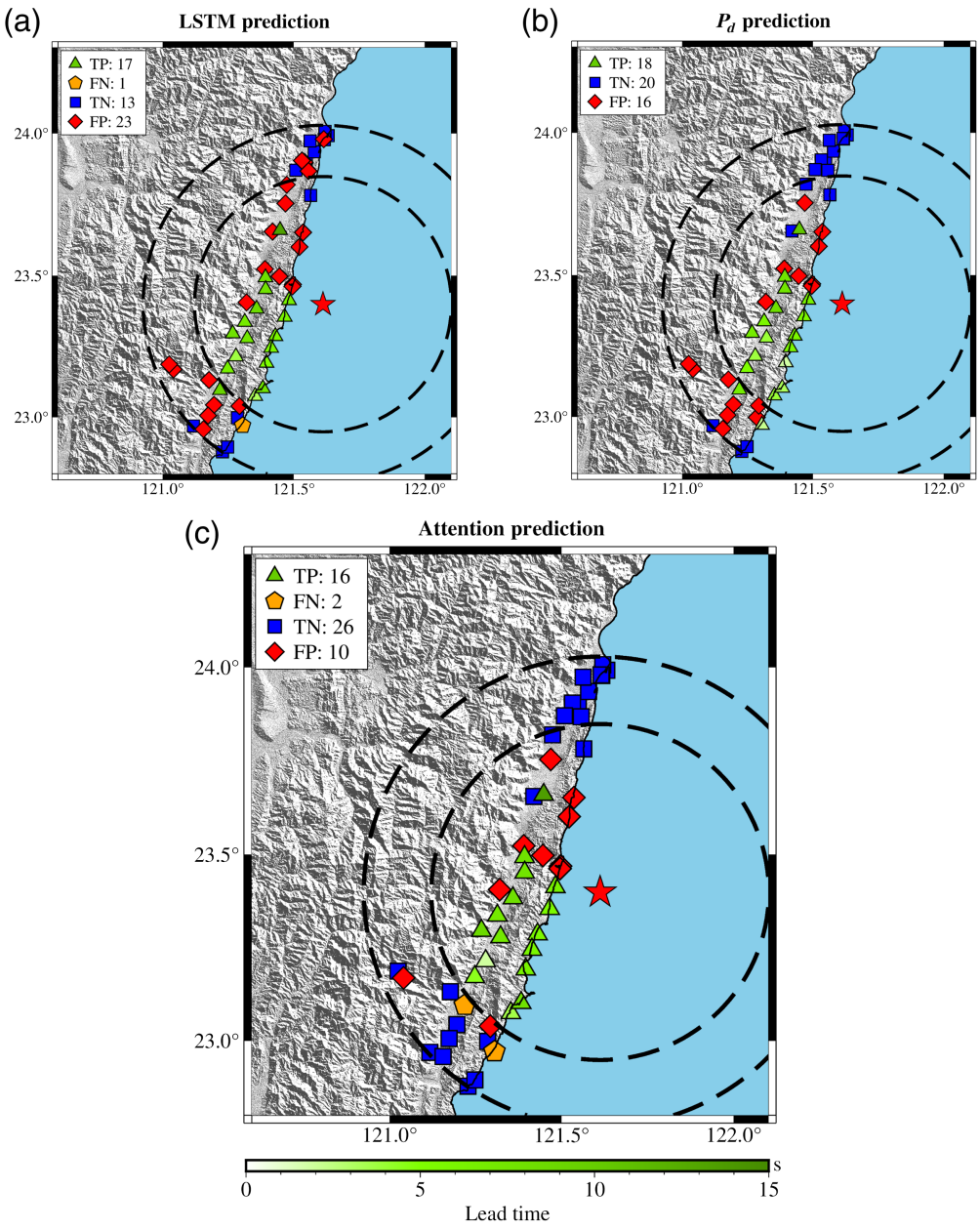

Figure 6. EEW performance for the 2022 Hualien earthquake. The triangles represent the TP stations, with color fillings indicating the lead time. The squares represent the TN stations. The diamonds represent the FP stations. The pentagons represent the FN stations. The two dashed circles around the epicenter (star) illustrate the 50 and 70 km radii around the epicenter. (a) The results of the LSTM model. (b) The results of the P_d method. (c) The results of the proposed attention-based model. The color version of this figure is available only in the electronic edition.

respectively. Among the 10 FP cases, 9 have PGA values ranging from ~100 to 400 gal.

Figure 7 illustrates the warning results for the 2022 Guanshan earthquake. The F1-scores of the three models—attention-based model, LSTM model, and P_d method—are 0.561, 0.469, and 0.511, respectively. The mean lead times (arithmetic mean of the TP cases) for the three models are 5.11, 5.33, and 5.42 s, respectively. We observe more TP cases in the south of the epicenter and more FP cases in the north of the epicenter. We think one possible explanation for the observation could be related to the empirical southward directivity seen in the evolution of the ShakeMap. The directivity effect (Doppler effect) enhances the shaking for records associated with incoming rupture. These larger initial signals are making themself more distinguishable from the noise, and it is thus easier to perform classification. This results in more true cases, including TPs. On the other hand, the records associated with outgoing rupture have smaller shaking. Those records are less distinguishable from the noise and thus more difficult to identify possible large PGVs.

Figure 8 presents the warning results for the 2022 Chishang earthquake. The F1-scores of the three models—attention-based model, LSTM model, and P_d method—are 0.756, 0.643, and 0.704, respectively. The mean lead times (arithmetic mean of the TP cases) for the three models are 8.26, 8.84, and 7.42 s, respectively. A close examination of the FN case for the attention-based model within 50 km reveals that this case is a "late bloomer." It takes 16.78 s to reach the 15 cm/s threshold after the *P*-wave arrival. Because our protocol only employs a 10 s window during training, this specific case highlights a limitation of the current approach. We believe this issue can be mitigated by incorporating longer training windows.

Table 1 presents the overall performance of the three methods across all four testing events. The overall F1-scores of the three models—attention-based model, LSTM model, and P_d method—are 0.694, 0.594, and 0.645, respectively. The overall mean lead times for the three models are 6.35, 6.97, and 5.68 s, respectively. Figure 9 illustrates the distributions of lead times. In summary, we find the improvement in the F1-score of the attention-based model to be significant. This model notably reduces the number

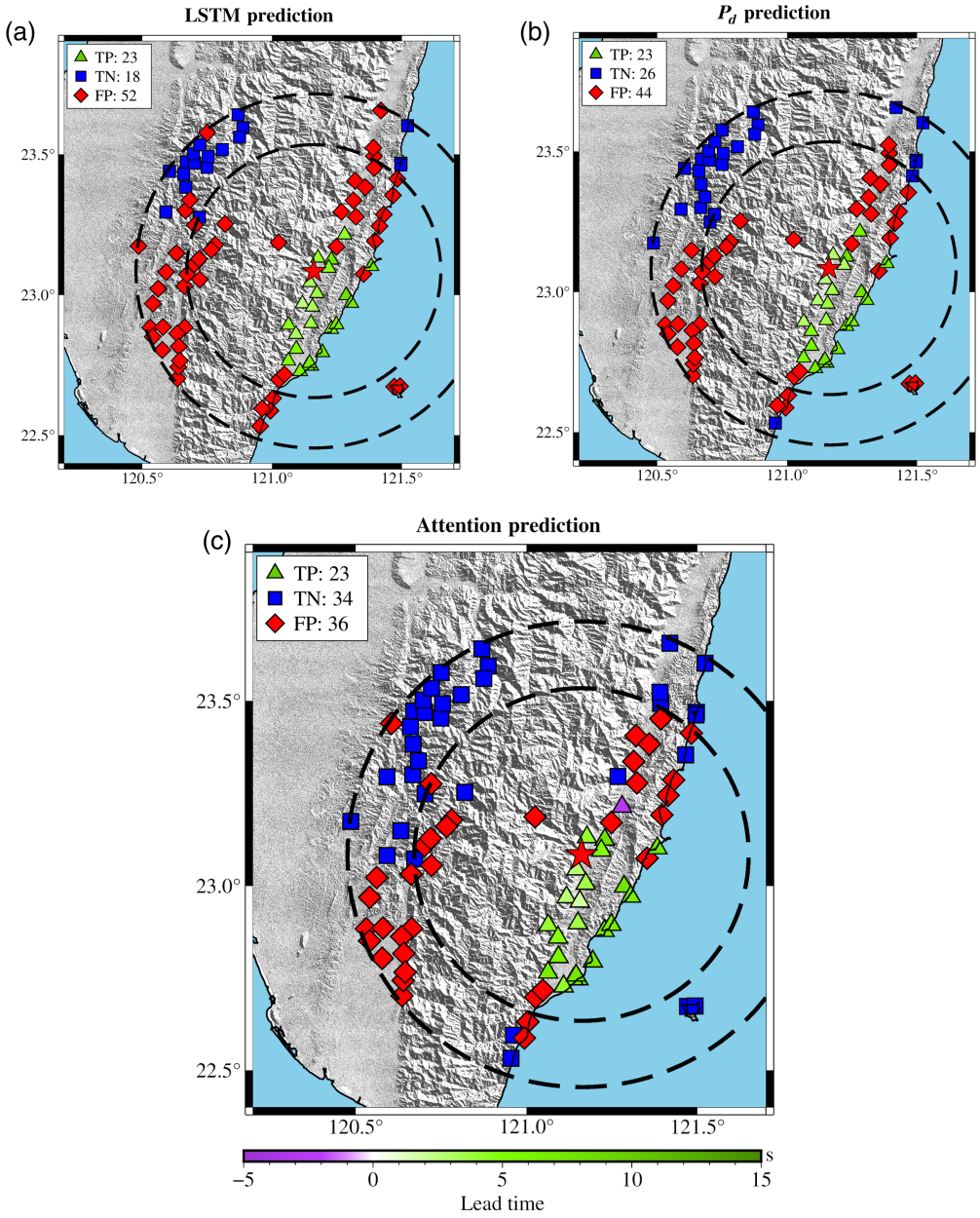

Figure 7. EEW performance for the 2022 Guanshan earthquake. The triangles represent the TP stations, with color fillings indicating the lead time. The squares represent the TN stations. The diamonds represent the FP stations. The pentagons represent the FN stations. The two dashed circles around the epicenter (star) illustrate the 50 and 70 km radii around the epicenter. (a) The results of the LSTM model. (b) The results of the P_d method. (c) The results of the proposed attention-based model. The color version of this figure is available only in the electronic edition.

of false alarms. Regarding lead time, both the LSTM model and the attention-based model perform well. The LSTM model slightly outperforms the attention-based model in lead time. However, the larger lead time and higher number of FP cases suggest that the LSTM model is more easily triggered.

Discussion

In this section, we discuss extra tests and their implications found in the proposed attention-based model.

Testing with different probability threshold

The output of the attention-based model is the probability of a large PGV, represented as a number between 0 and 1. By default, we set the probability threshold for issuing a warning to 0.5. When the probability of a large PGV exceeds 50% (0.5), the system issues a warning. This probability threshold is a metaparameter that can be adjusted by the user. Table 2 presents the performance results for different probability thresholds.

The results highlight the trade-off between the F1-score and lead time. Setting the probability threshold to higher values requires the model to have greater confidence before issuing a warning. Consequently, this increases the confirmation time but improves warning accuracy. In our metrics, this trade-off is evident: because the probability threshold increases, the F1-score improves while the lead time decreases. Conversely, lowering the probability threshold makes the model more sensitive to changes in ground motion. This reduces the confirmation time but compromises warning accuracy.

In practice, the optimal probability threshold can often be determined by identifying the "corner" of the trade-off curve—a common approach in seismic inversion. In our case, the corner of the trade-off curve is close to 0.6.

Testing with different composition subset of training datasets

Because of the scarcity of large PGV records in the *P*-alert network, we use the concept of transfer learning, building on the experience of a previous study, and incorporate large PGV records from events in Japan. This approach may raise concerns about the validity of using events from Japan. To address this,

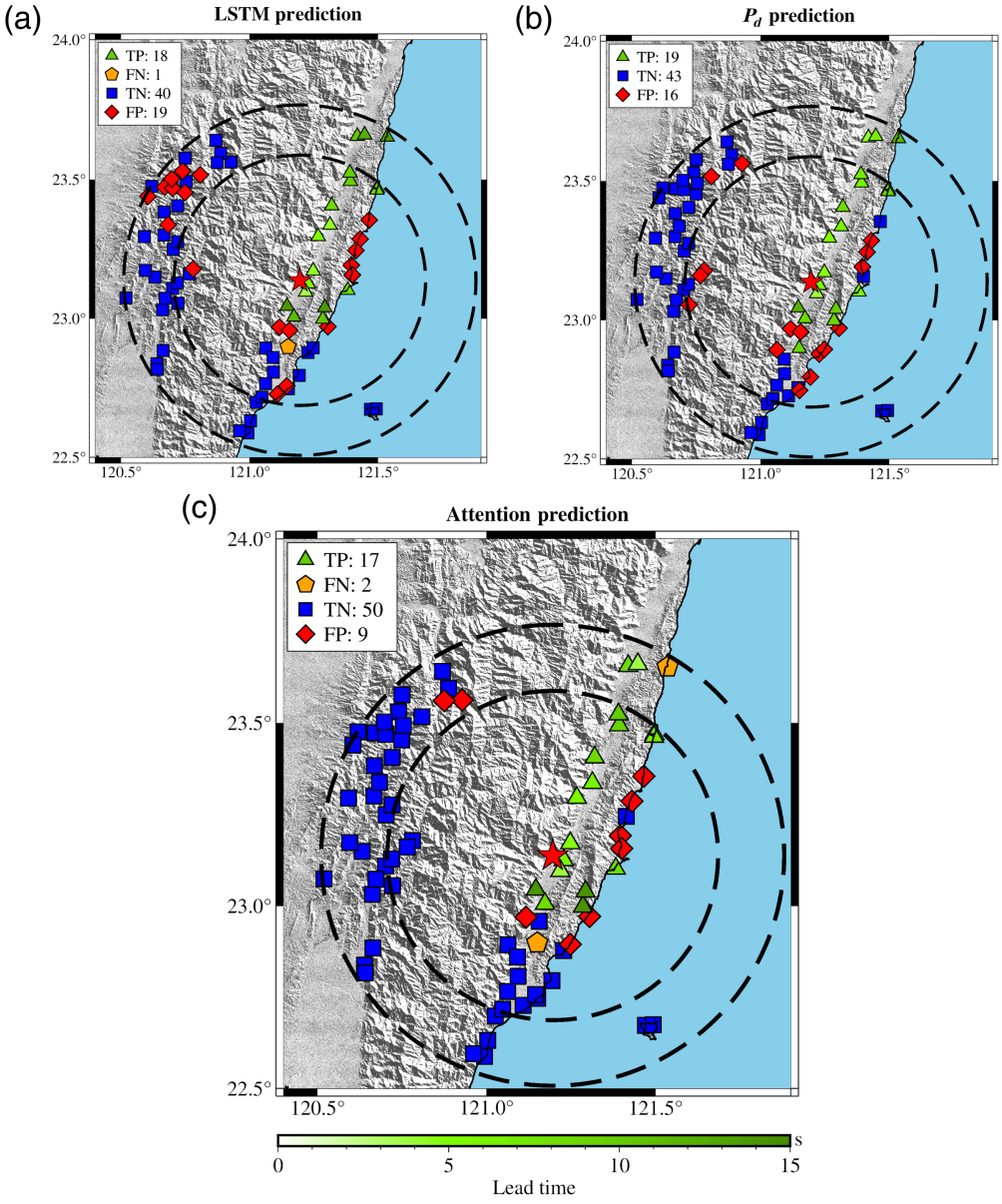

Figure 8. EEW performance for the 2022 Chishang earthquake. The triangles represent the TP stations, with color fillings indicating the lead time. The squares represent the TN stations. The diamonds represent the FP stations. The pentagons represent the FN stations. The two dashed circles around the epicenter (star) illustrate the 50 and 70 km radii around the epicenter. (a) The results of the LSTM model. (b) The results of the P_d method. (c) The results of the proposed attention-based model. The color version of this figure is available only in the electronic edition.

we conducted a sanity test by training the model using only events from Taiwan.

The results indicate that while the accuracy of the warnings increases, the lead time decreases. Specifically, the F1-score improves from 0.694 to 0.779, and the lead time reduces from 6.35 to 4.96 s. In terms of overall model performance, the model using transfer learning outperforms the model trained solely on events in Taiwan.

By comparing these results, we draw the following conclusions:

1. The optimal dataset for training would consist of all events in Taiwan with a sufficient number of large PGV records. However, the transfer learning strategy incorporating events from Japan proves to be a viable and effective alternative given the current limitations.
2. The improvement in the F1-score highlights differences in seismic records between Japan and Taiwan. Because all testing events are from Taiwan, a dataset with similar conditions allows the model to better capture the nuances of seismic records.
3. The reduction in lead time underscores the significant impact of a lack of large PGV records in the training dataset on the speed of recognition.

Testing with various shapes of label

In our proposed attention-based model and the LSTM model, we adopt the step-function label approach used in a previous study, in which the label reaches a probability of 1 at the *P*-wave arrival for all true records. Although this labeling choice pushes the limit of lead time, it raises concerns that the model may require additional time to make reliable decisions.

To investigate the effect of label shapes on performance, we trained the attention-based model using different label shapes and conducted comparisons. Figure 10 illustrates three alternative label shapes. The first label follows a sigmoid function. The second label adopts a similar pattern but saturates earlier. The third label follows a linear function.

The performance results reveal a trade-off tendency between the F1-score and lead time. Delaying the time at which the probability reaches 1 by 3 s generally increases the F1-score while reducing the lead time. Among the three label shapes, the linear

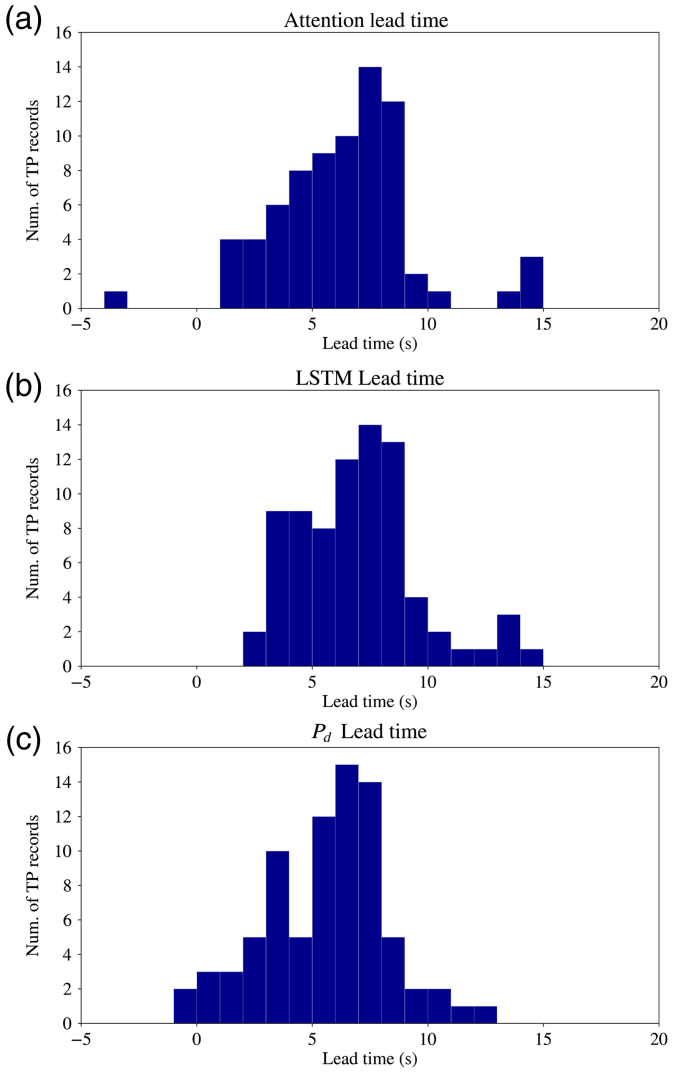

Figure 9. Lead time histogram of all three methods. (a) Results from the proposed attention-based model. (b) Results from the LSTM model. (c) Results from the P_d method. The color version of this figure is available only in the electronic edition.

label achieves the highest F1-score (0.764) but also the shortest lead time (4.35 s). The second label (early saturation) has the longest lead time (5.67 s) but a comparatively lower F1-score. In terms of overall model performance, the best label is the first label (sigmoid-shaped increasing function), which achieves a balanced F1-score (0.72) and lead time (5.24 s). However, none of the alternative labels surpass the original step-function label in terms of model score.

Based on these results, we offer the following observations:

1. The original step-function-at-arrival label is validated as an effective choice within the context of our model because it achieves a balanced trade-off between the F1-score and lead time.
2. Just as model performance can reveal the best feature sets, identifying the optimal label shape may also enhance our understanding of how the credibility of EEW systems evolves over time because the label is a sequential of probabilities for warning. For instance, one could investigate the ideal label shape to maximize the F1-score while maintaining fixed values at both endpoints (0.5 at 0 s and 1.0 at 3 s).
3. Among the three alternative labels, we think the linear label corresponds more closely to real-world scenarios of credibility changes over time, owing to its superior F1-score.

TABLE 1
Performance Results of All Four Testing Events

	Attention	**LSTM**	P_d
TP	75	79	80
FN	6	2	1
TN	138	92	111
FP	60	106	87
Miss alarm rate	7.4%	2.5 %	1.2 %
False alarm rate	30.3%	53.5%	43.9%
F1-score	0.694	0.594	0.645
Mean lead time	6.35 ± 2.96	6.97 ± 2.64	5.68 ± 2.63

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

TABLE 2
Performance Results with Different Threshold

Threshold	**F1-Score**	**Lead Time**
0.1	0.576	7.36 ± 2.69
0.2	0.619	7.16 ± 2.54
0.3	0.653	6.87 ± 2.62
0.4	0.679	6.76 ± 2.69
0.5	0.694	6.35 ± 2.96
0.6	0.706	6.04 ± 3.15
0.7	0.731	5.39 ± 3.42
0.8	0.762	5.06 ± 3.04
0.9	0.763	3.66 ± 2.96

Testing on adding noise and changing epicenter criterion in training dataset

Finally, we tested two possible modulations of the training dataset selection. The first involves expanding the dataset with additional noise records. The performance results show a marginal increase in the F1-score (0.696) and a moderate increase in lead time (6.78 s). Overall, we observed an improvement in the model score.

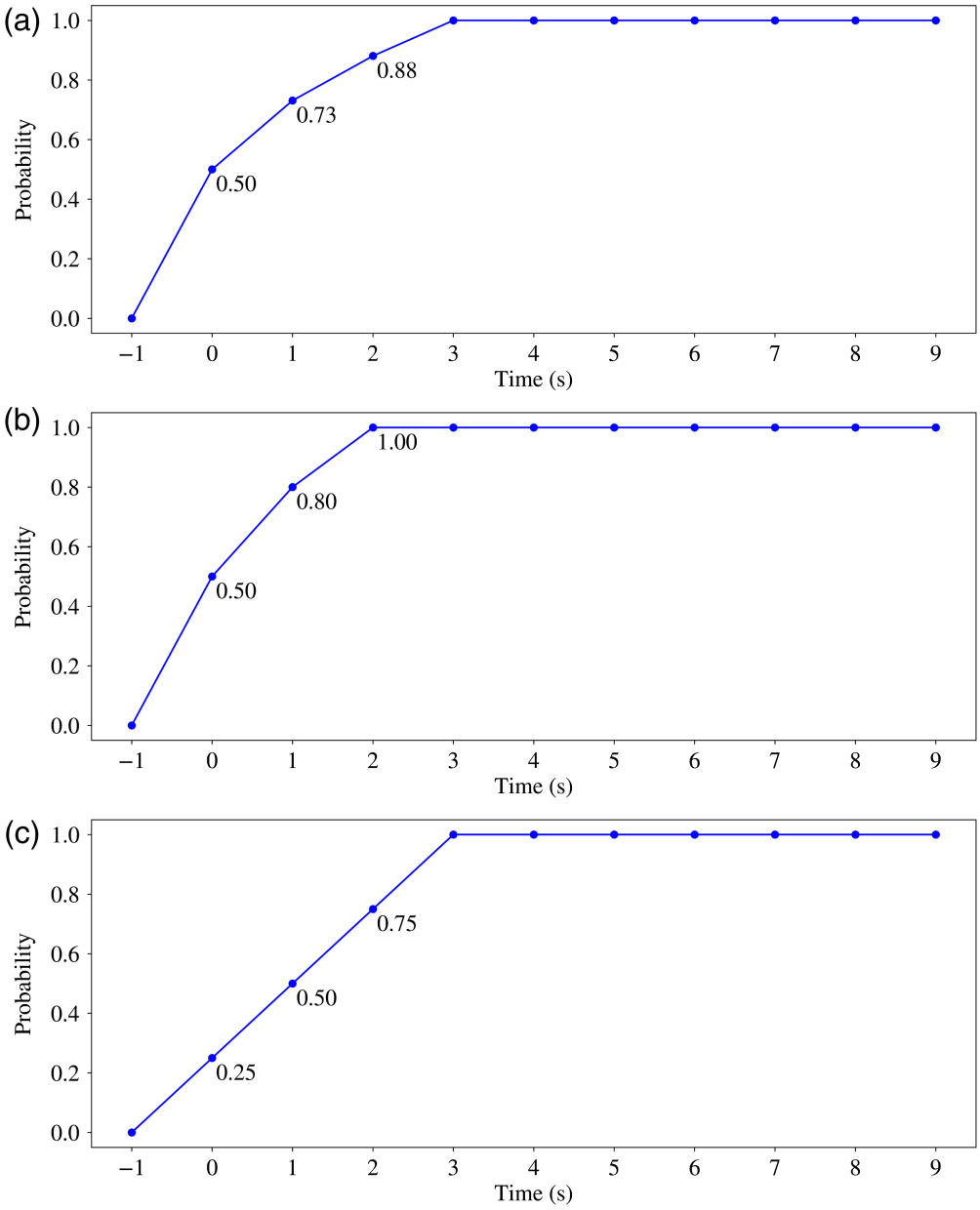

Figure 10. (a–c) Various shapes of labels used in the discussion. The color version of this figure is available only in the electronic edition.

The second approach involves increasing the dataset by relaxing the record selection criterion for the epicenter. Specifically, the new dataset includes records that were previously excluded for being too far from the epicenter. The performance results show an increase in the F1-score (0.712) but a notable decrease in lead time (5.74 s).

In summary, we believe that enlarging the current training dataset with noise records from the *P*-alert stations can be beneficial. This approach may help the model better capture detailed features hidden within the records. However, adding excessive noise records to the training dataset could exacerbate the existing data imbalance problem.

Performance difference between onsite PGV and onsite PGA

Compared with the previous study on onsite PGA ML, we observed that the overall performance is lower for onsite PGV. We believe there are two main reasons for this performance difference. The primary reason is the lack of TP records for large PGV. Specifically, there are relatively few records exceeding 15 cm/s in the *P*-alert dataset, which likely impacts the classification performance. The second reason relates to the nature of velocity evolution over time. We have identified a significantly higher number of "late bloomer"—records that start with mild motion but gradually exhibit more substantial velocity at later times. This phenomenon can occasionally confuse the trained ML model. Interestingly, the "late bloomers" are less observed in the PGA dataset.

Limitations

In the testing events, we only use the records with normal behavior. These are records without missing data and abnormal data. The proposed model cannot properly issue alerts with abnormal records. The proposed model might give unpredictable results.

Another limitation lies in the bias in the training dataset regarding the types of seismic sources the events have. Because of the rarity of the large PGV records, the corresponding events are also limited. The trained model could potentially have worse alert performance if the records come from events with very different source parameters, including source mechanisms, rupture dynamics, source depths, and so forth.

Computational specification

The ML training was conducted on a server equipped with an NVIDIA Tesla T4 Tensor Core graphic processing units. The Tesla T4 features 2560 CUDA cores and 16 GB of RAM.

Subsequently, we present some benchmarks for ML performance using the Tesla T4. For a one-layer attention mechanism architecture, the training time is ∼1.5 hr, with each epoch taking about 100–110 s. In contrast, for a four-layer LSTM architecture, the training time is significantly shorter at around 20 min, with each epoch taking ∼6–8 s.

Conclusions

In this project, we adopt an attention-based neural network to address the onsite PGV EEW problem. The attention-based architecture demonstrates its feasibility in delivering fast and reliable early warnings based on the eventual PGV. Our findings indicate a noticeable improvement in the accuracy of onsite PGV predictions using the attention-based model. The mean lead time is slightly better than the LSTM model. In addition, the feature selection results highlight the importance of station corrections in accounting for the inhomogeneity of stations.

Furthermore, we have explored the optimized label for achieving the best performance. We think the optimized label illustrates how the credibility of EEW grows over time, potentially providing an alternative approach to assessing the validity of EEW systems.

We believe that our prototype attention-based model can contribute to the future development of onsite EEW systems. In addition, our project raises several novel interesting questions for future exploration. For instance, what is the most suitable architecture for the PGV EEW problem? What are the common characteristics of optimized labels? How should we define the best model? What makes the most reasonable model score?

Data and Resources

The strong-motion waveform records from the *P*-alert networks can be downloaded at http://palert.earth.sinica.edu.tw/db/. The strong-motion waveform records from the Kyoshin net (K-NET) networks can be downloaded at National Research Institute for Earth Science and Disaster Resilience website https://www.kyoshin.bosai.go.jp/kyoshin, doi: 10.17598/NIED.0004. All websites were last accessed in June 2025.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

This work was supported by the Ministry of Science and Technology (MOST) of Taiwan under MOST 106-2116-M-002-019-MY3 and MOST 109-2116-M-002-030-MY3. T. C. H. thanks Lily Wong for her hospitality during the making of the article. T. C. H. initiated the project, developed the machine learning (ML) model, and wrote the article. T. L. L. conducted the data analysis, trained and tested the ML, and developed the code for processing. T. C. H. and T. L. L. share the first authorship of this article. B. M. Y. trained and tested the ML and supported the hardware of the server. Y. M. W. supervised this work and advised. All authors discussed the training results and contributed to the final article.

References

Allen, R. M., P. Gasparini, O. Kamigaichi, and M. Bose (2009). The status of earthquake early warning around the world: An introductory overview, *Seismol. Res. Lett.* **80,** 682–693, doi: 10.1785/gssrl.80.5.682.

Bahdanau, D., K. Cho, and Y. Bengio (2015). Neural machine translation by jointly learning to align and translate, *Proc. of the 3rd International Conf. on Learning Representations (ICLR)*, San Diego, California, 7–9 May 2015.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.* (2020). Language models are few-shot learners, *Advances in Neural Information Processing Systems*, Vol. 33, 1877–1901.

Chandrakumar, C., M. L. Tan, C. Holden, M. Stephens, A. Punchihewa, and R. Prasanna (2024). Estimating S-wave amplitude for earthquake early warning in New Zealand: Leveraging the first 3 seconds of P-wave, *Earth Sci. Inf.* **17**, 4527–4554, doi: 10.1007/s12145-024-01403-6.

Cho, K., B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation, 1724–1734.

Crawford, J. R., and P. H. Garthwaite (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks, *Clin. Neuropsychol.* **23,** no. 2, 193–204, doi: 10.1080/13854040801968450.

Fan, J., C. Sun, C. Chen, X. Jiang, X. Liu, X. Zhao, L. Meng, C. Dai, and W. Chen (2020). EEG data augmentation: Towards class imbalance problem in sleep staging tasks, *J. Neural Eng.* **17,** 056017, doi: 10.1088/1741-2552/abb5be.

Gers, F. A., N. N. Schraudolph, and J. Schmidhuber (2002). Learning precise timing with lstm recurrent networks, *J. Machine Learn. Res.* **3,** no. 1, 115–143, doi: 10.1162/153244303768966139.

Hochreiter, S., and J. Schmidhuber (1997). Long short-term memory, *Neural Comput.* **9,** no. 8, 1735–1780, doi: 10.1162/neco.1997.9.8.1735.

Hsieh, C. Y., W. A. Chao, and Y. M. Wu (2015). An examination of the threshold-based earthquake early warning approach using a low-cost seismic network, *Seismol. Res. Lett.* **86**, 1664–1667, doi: 10.1785/0220150073.

Huang, T.-C., and Y.-M. Wu (2019). A robust algorithm for automatic P-wave arrival-time picking based on the local extrema scalogram, *Bull. Seismol. Soc. Am.* **109,** no. 1, 413–423, doi: 10.1785/0120180127.

Huang, T.-C., and Y.-M. Wu (2021). Revisiting M_L determination in Taiwan based on the expectation-maximization algorithm, *J. Seismol.* **25,** 1077–1087, doi: 10.1007/s10950-021-10013-4.

Huang, T.-C., and Y.-M. Wu (2023a). Improving earthquake early warning initial peak ground motion magnitude estimation with station corrections: A case study using the P-alert network in Taiwan, *J. Earthq. Eng.* **28,** no. 6, 1532–1551, doi: 10.1080/13632469.2023.2245498.

Huang, T.-C., and Y.-M. Wu (2023b). Revisiting the initial peak P-wave displacement and the ground motion characteristic period with signal-to-noise ratios: A case study using a low-cost sensor network in Taiwan, *J. Earthq. Eng.* **27,** no. 16, 4694–4704, doi: 10.1080/13632469.2023.2190416.

Kamigaichi, O., M. Saito, K. Doi, T. Matsumori, S. Tsukada, K. Takeda, T. Shimoyama, K. Nakamura, M. Kiyomoto, and Y.

Watanabe (2009). Earthquake early warning in Japan: Warning the general public and future prospects, *Seismol. Res. Lett.* **80,** no. 5, 717–726, doi: 10.1785/gssrl.80.5.717.

Li, Z., M. A. Meier, E. Hauksson, Z. Zhan, and J. Andrews (2018). machine learning seismic wave discrimination: Application to earthquake early warning, *Geophys. Res. Lett.* **45**, 4773–4779, doi: 10.1029/2018GL077870.

Liu, P.-H., and J.-H. Wu (2024). New methodology for assessing seismic ground vulnerability based on microtremor and Taiwan seismic intensity scale, *J. Earthq. Eng.* doi: 10.1080/13632469.2024.2415992.

Luong, M. T., H. Pham, and C. D. Manning (2015). Effective approaches to attention-based neural machine translation, *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, 1412–1421.

Mittal, H., B. M. Yang, T. L. Tseng, and Y. M. Wu (2021). Importance of real-time PGV in terms of lead-time and shakemaps: Results using 2018 M_L 6.2 and 2019 M_L 6.3 Hualien, Taiwan earthquakes, *J. Asian Earth Sci.* **220,** 104936, doi: 10.1016/j.jseaes.2021.104936.

Mousavi, S. M., and G. C. Beroza (2020). A machine-learning approach for earthquake magnitude estimation, *Geophys. Res. Lett.* **47,** e2019GL085976, doi: 10.1029/2019GL085976.

Satriano, C., Y. M. Wu, A. Zollo, and H. Kanamori (2011). Earthquake early warning: Concepts, methods and physical grounds, *Soil Dynam. Earthq. Eng.* **31,** no. 2, 687–719, doi: 10.1016/j.soildyn.2010.07.007.

Sun, Y., A. K. Wong, and M. S. Kamel (2009). Classification of imbalanced data: A review, *Int. J. Pattern Recognit. Artif. Intell.* **23,** no. 4, 687–719, doi: 10.1142/S0218001409007326.

Sutskever, I., O. Vinyals, and Q. V. Le (2014). Sequence to sequence learning with neural networks, *Advances in Neural Information Processing System*, Vol. 2, 3104–3112.

Thabtah, F., S. Hammoud, F. Kamalov, and A. Gonsalves (2020). Data imbalance in classification: Experimental evaluation, *Inform. Sci.* **513**, 429–441, doi: 10.1016/j.ins.2019.11.004.

Vaiciulyte, S., D. A. Novelo-Casanova, and A. L. Husker (2024). Demystifying response to eew in Mexico: Socio-technical motivations in protective action, *Saf. Sci.* **174**, 106469, doi: 10.1016/j.ssci.2024.106469.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017). Attention is all you need, *Advances in Neural Information Processing Systems*, Vol. 30, 5998–6008.

Wald, D. J. (2020). Practical limitations of earthquake early warning, *Earthq. Spectra* **36,** no. 3, 1412–1447, doi: 10.1177/8755293020911388.

Wang, C.-Y., T.-C. Huang, and Y.-M. Wu (2022). Using LSTM neural networks for onsite earthquake early warning, *Seismol. Res. Lett.* **93,** 814–826, doi: 10.1785/0220210197.

Wu, N., B. Green, X. Ben, and S. O'Banion (2020). Deep transformer models for time series forecasting: The influenza prevalence case, *Proc. of the 37th International Conf. on Machine Learning*, Vienna, Austria, 13–18 July 2020.

Wu, Y. M., and H. Kanamori (2005). Rapid assessment of damage potential of earthquakes in Taiwan from the beginning of P waves, *Bull. Seismol. Soc. Am.* **95,** no. 3, 1181–1185, doi: 10.1785/0120040193.

Wu, Y. M., and H. Mittal (2021). A review on the development of earthquake warning system using low-cost sensors in Taiwan, *Sensors* **21,** 7649, doi: 10.3390/s21227649.

Wu, Y.-M., and L. Zhao (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning, *Geophys. Res. Lett.* **33,** no. 16, L16312, doi: 10.1029/2006GL026871.

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

Wu, Y.-M., Y.-H. Lin, B. M. Yang, and S.-S. Ke (2025). Performance of the P-alert real-time shakemaps system and onsite warning during the 2025 M_L 6.4 Dapu earthquake, *Terr. Atmos. Ocean. Sci.* **36,** no. 3, doi: 10.1007/s44195-025-00086-w.

Wu, Y. M., T. L. Teng, T. C. Shin, and N. C. Hsiao (2003). Relationship between peak ground acceleration, peak ground velocity, and intensity in Taiwan, *Bull. Seismol. Soc. Am.* **93,** no. 1, 386–396, doi: 10.1785/0120020097.

Yamaguchi, N., and F. Yamazaki (2001). Estimation of strong motion distribution in the 1995 Kobe earthquake based on building damage data, *Earthq. Eng. Struct. Dynam.* **30,** no. 6, 787–801, doi: 10.1002/eqe.33.

Yang, B. M., H. Mittal, and Y.-M. Wu (2021). Real-time production of PGA, PGV, intensity, and Sa Shakemaps using dense MEMS-based sensors in Taiwan, *Sensors* **21,** no. 3, 943, doi: 10.3390/s21030943.

Zhang, X., M. Zhang, and X. Tian (2021). Real-time earthquake early warning with deep learning: Application to the 2016 M 6.0 Central Apennines, Italy earthquake, *Geophys. Res. Lett.* **48,** 2020GL089394, doi: 10.1029/2020GL089394.

Manuscript received 30 December 2024
Published online 1 August 2025