# Using LSTM Neural Networks for Onsite Earthquake Early Warning

Chia-Yu Wang[1], Ting-Chung Huang[*1], and Yih-Min Wu[1,2,3]

## Abstract

**Onsite earthquake early warning (EEW) systems determine possible destructive *S* waves solely from initial *P* waves and issue alarms before heavy shaking begins. Onsite EEW plays a crucial role in filling in the blank of the blind zone near the epicenter, which often suffers the most from disastrous ground shaking. Previous studies suggest that the peak *P*-wave displacement amplitude ($P_d$) may serve as a possible indicator of destructive earthquakes. However, the attempt to use a single indicator with fixed thresholds suffers from inevitable errors because the diversity in travel paths and site effects for different stations introduces complex nonlinearities. In addition, the short warning time poses a threat to the validity of EEW. To conquer the aforementioned problems, this study presents a deep learning approach employing long short-term memory (LSTM) neural networks, which can produce a highly nonlinear neural network and derive an alert probability at every time step. The proposed LSTM neural network is then tested with two major earthquake events and one moderate earthquake event that occurred recently in Taiwan, yielding the results of a missed alarm rate of 0% and a false alarm rate of 2.01%. This study demonstrates promising outcomes in both missed alarms and false alarms reduction. Moreover, the proposed model provides an adequate warning time for emergency response.**

**Cite this article as** 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.

## Introduction

Destructive earthquake events have continuously caused severe loss of human lives and property. With studies showing the probable inherent unpredictability of earthquakes (Geller, 1997), reliable short-term earthquake prediction remains impractical (Kanamori *et al.*, 1997). The urgent need for seismic hazard mitigation thus demands an alternative approach. These circumstances encourage the development of earthquake early warning (EEW) systems.

An EEW system delivers ground-shaking alerts after an earthquake has nucleated. These alerts may provide crucial warning time for hazard mitigation procedures to be carried out, making it possible to avert human casualties and economic losses. In this section, the history of the development of EEW systems are reviewed. Next, we provide a detailed examination of the conceptual difference between regional EEW and onsite EEW.

In Taiwan, the idea for EEW system implementation came after an $M_w = 7.4$ (from the U.S. Geological Survey database) earthquake in 1986. Although the epicenter was located offshore of Hualien, the most severe damage occurred in Taipei, approximately 120 km away (Wu *et al.*, 1999). In 1994, the Central Weather Bureau (CWB) of Taiwan started the operation of a prototype EEW system around Hualien. The station signals were processed in real time, and the results were transmitted back to the CWB data center in Taipei. The idea was to utilize the velocity difference between the destructive *S* wave and transmission speed to provide a warning in the urbanized Taipei area. However, due to the station density and distribution, early warnings had an average error of 22 km in epicenter location and 0.7 units in magnitude (Wu *et al.*, 1999). Starting in 2001, a new system using a virtual subnetwork approach (Wu and Teng, 2002) became functional with the implementation of a new real-time strong-motion network operated by the CWB, resulting in a more accurate estimation of the source parameters and providing faster warnings. In the past decade, efforts to develop EEW systems have increased worldwide. Currently, several EEW systems are operational around the world. The present state of these systems can be divided into three categories. Public alerts distributed through broadcasts or cellphones have been achieved in Japan, Taiwan, South Korea, and Mexico. Limited alerts distributed to select users, such as schools, gas companies, and

1. Department of Geosciences, National Taiwan University, Taipei, Taiwan, https://orcid.org/0000-0002-8133-9879 (C-YW); https://orcid.org/0000-0002-8039-952X (T-CH); https://orcid.org/0000-0003-4542-1741 (Y-MW); 2. NTU Research Center for Future Earth, National Taiwan University, Taipei, Taiwan; 3. Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan

*Corresponding author: tingchunghuang@gmail.com

railway systems, function in India, Romania, Turkey, and the United States. Finally, EEW systems in Italy, Switzerland, Chile, and several other regions are in testing and development stages (Allen and Melgar, 2019).

The peak $P$-wave displacement amplitude ($P_d$) was proposed by Wu and Kanamori (2005) to serve as a possible indicator for onsite EEW. $P_d$ is obtained by double integration of the strong-motion acceleration signals, followed by application of a 0.075 Hz Butterworth high-pass filter to avoid the low frequency drifts caused by numerical integration. A specific time window of 3 s from the initial $P$ waves was suggested by previous studies for a compromise between reliable ground-motion estimation and early warning time (Wu and Kanamori, 2005; Wu *et al.*, 2007).

The permitted time window (PTW) and the threshold value are two important parameters in the $P_d$ approach. The PTW determines the length of the initial $P$ wave used for estimating the final peak ground shaking. The choice of the PTW is often a compromise between warning lead time and alert accuracy. If a longer PTW is chosen, $P_d$ will have access to more ground-motion information during the rupture, resulting in a more accurate estimation. However, the trade-off of a longer PTW is that the warning time will decrease. On the other hand, a shorter PTW will provide a longer warning time at the cost of estimation accuracy.

The threshold value, which is the other important parameter in the $P_d$ method, controls the size of the type one and type two errors. In other words, the threshold value determines the possibility of false alarms and missed alarms. A high threshold will cause fewer false alarms, but it will also lead to more missed alarms, whereas a low threshold will behave oppositely. Hsieh *et al.* (2015) examined the effect of the two parameters in the $P_d$ method using 1186 strong-motion records from four earthquake events. The PTW is tested over a range of 1–10 s, and the threshold is tested over a range of 0.1–0.6 cm with a 0.05 cm interval. The study suggested choosing a 3 s PTW and 0.35 cm threshold value to optimize the $P_d$ performance in Taiwan, resulting in an average lead time of 2.92 s and a successful detection rate of 90.91%.

In recent years, the rapid development of machine learning has opened up many opportunities to innovate solutions for existing problems. In the EEW, there are many machine learning applications. For example, Li *et al.* (2018) used a combination of generative adversarial networks and random forest to distinguish $P$ waves from local impulsive noise, Mousavi and Beroza (2019) used both convolutional and recurrent networks to estimate magnitude, and Saad *et al.* (2021) used a deep convolutional neural network (CNN) to estimate earthquake parameters, including magnitude, origin time, depth, and location.

In our in-progress exploration studies, we revisited the onsite EEW problem with the technique of CNNs. We reported a successful model with much improved false alarm and missed alarm rates compared with the optimized $P_d$ method (Liu, 2019). However, one disadvantage of CNNs is that the PTW must be fixed. Unfortunately, this high-success-rate model can sometimes result in a reduction in the warning lead time. To account for the lead time, we propose a new framework to map earthquake ground motion into time series using recurrent neural networks (RNNs). The RNN triggers an alarm whenever it experiences enough cumulative shaking from the past, meaning it no longer has to wait for a fixed PTW to estimate. Freeing up the PTW increases the lead time in many cases without compromising too much accuracy.

In this article, we employ the long short-term memory (LSTM) method, a well-known branch of RNNs, to solve the onsite early warning problem. The key points of this work are as follows:

- We describe in detail the proposed LSTM model for the EEW problem (see the Methodology section).
- We show the test results of the LSTM model against three moderate-to-large independent events (see the Results section).
- We compare the effectiveness of the LSTM method with that of the $P_d$ method. (see the Comparison with the $P_d$ Method section).

## Data

The training data used in this study were obtained from two seismic networks. The first source is the Taiwan $P$-alert network operated by National Taiwan University. The second source is the Japan strong-motion seismograph network (K-NET), which provides data to increase the training data size and complement large-magnitude events that are not included in the Taiwan $P$-alert network.

The $P$-alert seismic network consists of a large number of low-cost microelectromechanical system accelerometers (Wu *et al.*, 2013), resulting in an ultrahigh density network in Taiwan. As of 2021, there are more than 700 operational stations running in the $P$-alert network.

K-NET is a nationwide network spread across Japan that has been operational since 1996. More than 1000 stations are uniformly distributed, with an average distance of 20 km. The strong-motion signals are transmitted to the National Research Institute for Earth Science and Disaster Resilience (NIED) data center in real time and are made available for the general public.

The events are selected for analysis based on their potential damaging power and their need for onsite warning. All of the events occurred between 2011 and 2019. For the $P$-alert network, events with local magnitudes greater than five located within latitude 21° N ~ 27° N and longitude 118° E ~ 123° E are selected. A total of 167 $P$-alert events with 10,202 records are included in the dataset. Figure 1a,b shows the $P$-alert events used in this project. For K-NET, events with local magnitudes greater than five located within latitude 25° N ~ 50° N and longitude 125° E ~ 150° E are selected. An

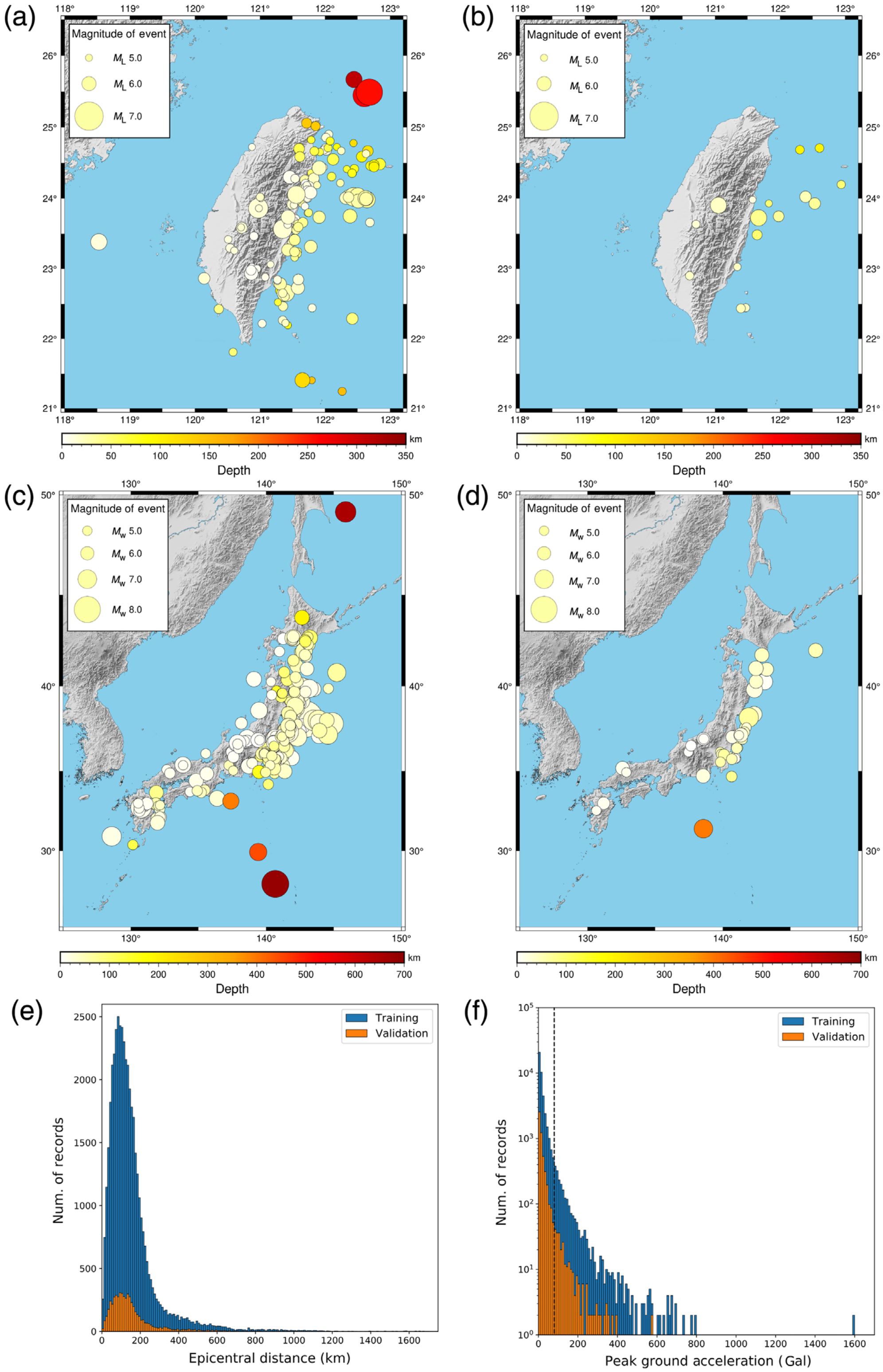

**Figure 1.** Illustration of the data used in this project. (a) Training events from the *P*-alert network. (b) Validation events from the *P*-alert network. (c) Training events from K-NET. (d) Validation events from K-NET. (e) Distribution of records with respect to epicentral distance. (f) Distribution of records with respect to peak ground acceleration. The color version of this figure is available only in the electronic edition.

additional recorded station number criterion is set to exclude some small offshore events that were unlikely to impact inland Japan. A total of 369 K-NET events with 38,113 records are included in the dataset. Figure 1c,d shows the K-NET events used in this project. Figure 1e shows the statistical distribution of the total records against epicentral distance. Figure 1f shows the statistical distribution of the total records against the peak ground acceleration (PGA).

# Methodology

## LSTM network

The LSTM network (Hochreiter and Schmidhuber, 1997; Gers *et al.*, 2000, 2005) is a distinct kind of RNN (Williams and Zipser, 1989; Werbos, 1990) that is capable of forecasting anomalies based on long-term and short-term trends. LSTM achieves this by employing both the cell state to represent the current state and multiple gates to screen the information flow with time. LSTM updates the warning state when acquiring new information each time step (0.01 s in our EEW problem). Inspired by the use of the short-term average/long-term average (STA/LTA) method for seismic detection, we consider that the LSTM, which also learns from long-term and short-term trends, can be used with the time series of seismic waveforms to identify damaging earthquakes.

## Data preprocessing: feature extraction

Figure 2 summarizes the preprocessing procedures. The feature generation process starts with raw records from the *P*-alert and K-NET stations,

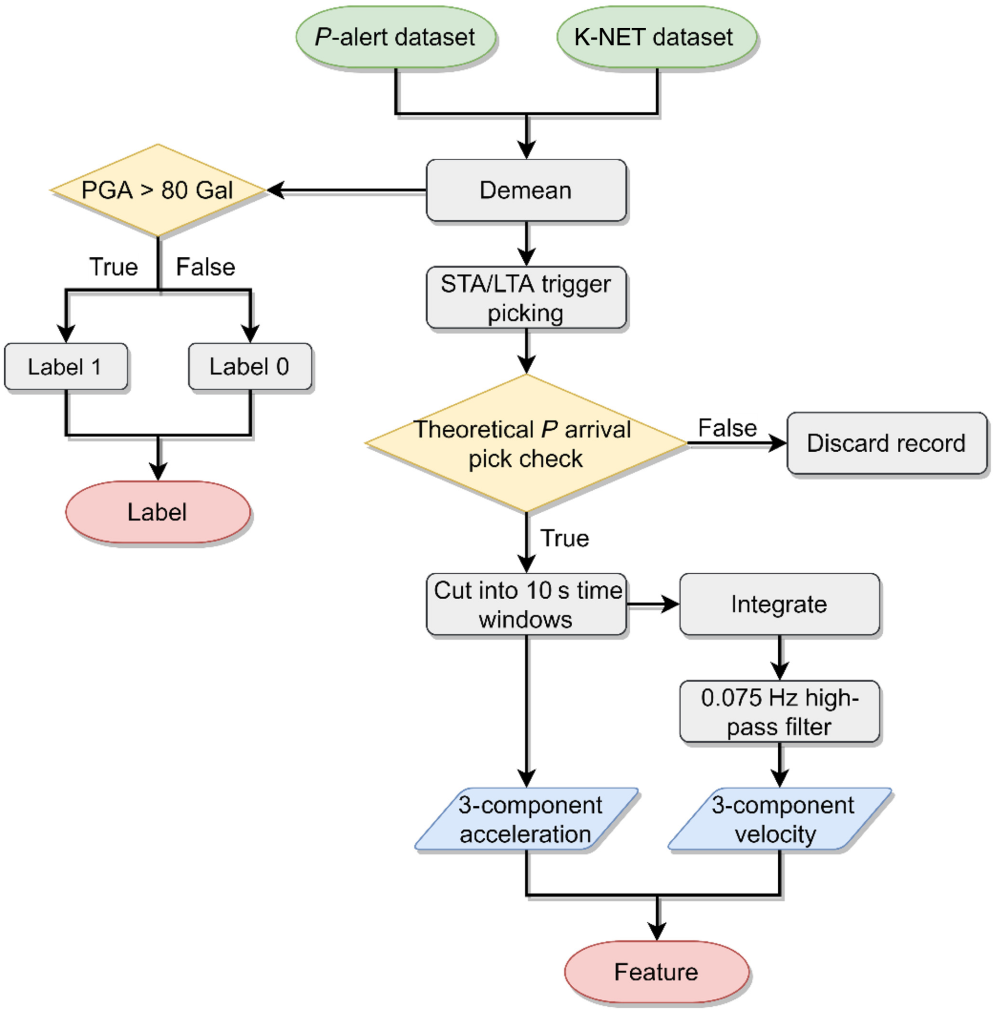


**Figure 2.** Flowchart of the training data preprocessing. The right side of the flowchart illustrates how the features are generated. The left side of the flowchart illustrates how the label is determined. The color version of this figure is available only in the electronic edition.

both with a 100 Hz sampling rate. The raw records are demeaned before determining *P* arrivals with an STA/LTA picker. The results of the *P*-arrival determination are checked with the 1D velocity model (Preliminary Reference Earth Model [PREM]) and another automatic picker (Huang and Wu, 2019). These checks control the data quality and ensure that the picks do not severely deviate. The records are discarded if the picking time difference between the methods exceeds one second. The feature data consist of 10 s time windows starting 1 s before the pick and continuing to 9 s after the pick. There are a total of six channels included in the input features, including three components of accelerations and three velocity components processed with a 0.075 Hz Butterworth high-pass filter. Although the feature consists of 10 s time windows, the machine learning platform, Keras in our case, does not employ it all at once. Instead, it reads from our code to know that it is meant for the RNN architecture. Then it treats the feature data as time series and uses the feature incrementally for training, validation, and testing.

The upper panel of Figure 3 demonstrates the operation of the STA/LTA picker on one *P*-alert trace. The red line denotes the time when the STA/LTA exceeds the threshold. The lower panel of Figure 3 demonstrates a magnified waveform from the same trace on three channels. The red line denotes the pick time, and the shaded area shows the range of the 10 s time window used as a feature input.

## Data preprocessing: label determination

The label criterion is based on the Gal, which follows a previous study (Hsieh *et al.*, 2015). The label criterion is in accordance with the definition of the CWB intensity scale and its definition of serious shaking. Records with PGA of the whole event greater than 80 Gal are labeled as 1 (alert), and records with PGA of the whole event less than 80 Gal are labeled as 0 (no alert). Figure 4 shows both labels in the time series format used in the LSTM machine learning. The time series of label 1 consists of a step function that rises from zero to one at the *P* arrival time. The time series of label 0 remains zero in probability over the whole range of the feature time window.

## Training and validating

Throughout the training process in machine learning, model overfitting is always a daunting issue to conquer. Overfitting denotes the phenomenon that a trained model fits nicely with the training data but fails to generalize to the new data. In other words, overfitting occurs when the performance of the training data is much higher than the performance of the test data.

Overall, we do not observe severe overfitting in this work, just some minor rebound overfitting cases, which coresponds to overfitting after model convergence is achieved. It usually happens after many epochs of training. We adopt some common practices of machine learning to monitor and prevent

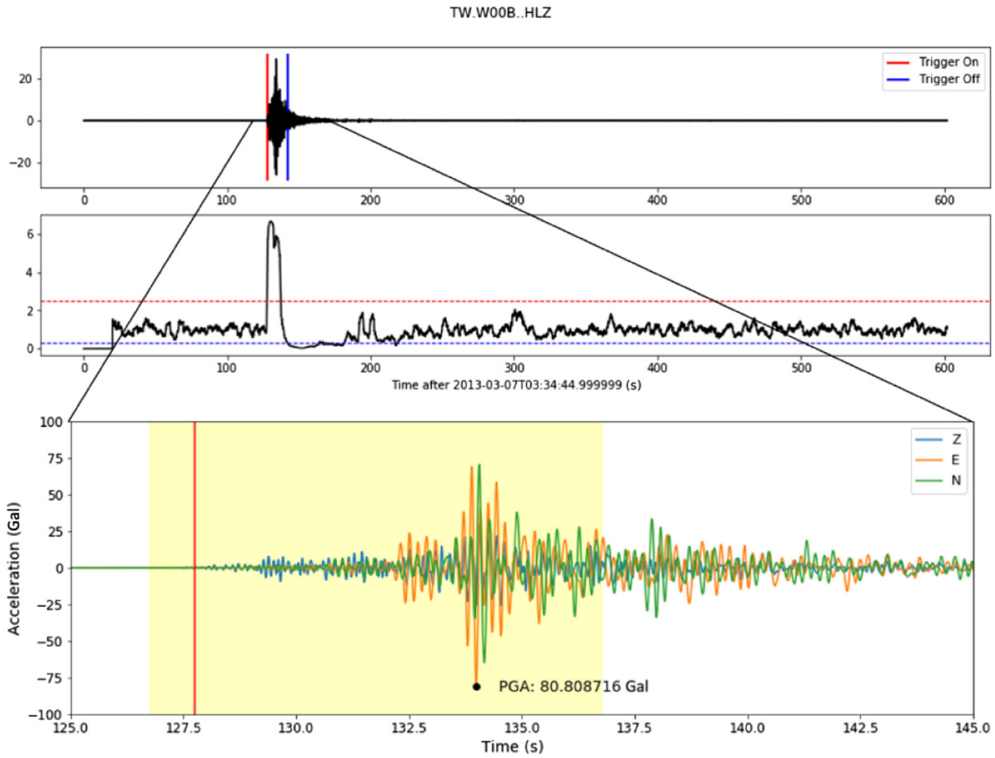


**Figure 3.** Demonstration of the short-term average/long-term average (STA/LTA) picker and the 10 s window feature. The red line denotes the STA/LTA pick. The shaded area denotes the 10 s window as the input feature. The color version of this figure is available only in the electronic edition.

overfitting. First, we use the validation data to monitor the learning process. The validation dataset comes from the training dataset, but it is independent of the learning process. In practice, the whole training dataset with 536 events is divided into 479 events (43,969 records) for training and 57 events (5346 records) for validation. Table 1 shows the details of the training data and validation data. By grouping events into these two bins, we are able to avoid possible data leakage between the training and validation datasets. Second, an early training stop is used to prevent rebound overfitting. The early training stop uses minimal validation loss as an indicator to determine the best epoch (iteration of training process) at which to stop. In practice, we consider using the early stop if rebound overfitting happens after 70–80 epochs.

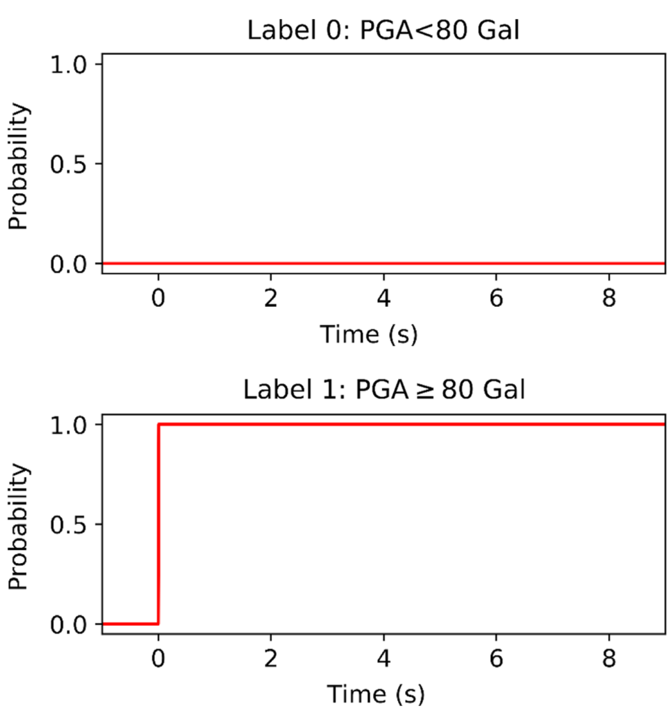


**Figure 4.** Ground-truth time series labeling for long short-term memory (LSTM) machine learning. Label 0 is a constant function with a value of 0. Label 1 is a Heaviside step function with a jump at the *P* arrival. The color version of this figure is available only in the electronic edition.

# Results

## Optimized hyperparameter search

The deep learning model architecture can be determined through various hyperparameters. We perform a grid search of these hyperparameters to derive the optimal model for the onsite EEW problem. There are four types of hyperparameters in our LSTM models.

The first two hyperparameters are the number of hidden LSTM layers (depth) and the number of units in one layer (width). Although the universal approximation theorem (Cybenko, 1989; Leshno *et al.*, 1993) states that neural networks with a single hidden layer can approximate any continuous function, the width of such a shallow network would need to grow exponentially according to the input dimension, which

TABLE 1
**Training and Validation Dataset**

| | *P*-Alert | K-NET | Total |
|---|---|---|---|
| Training events | 151 | 328 | 479 |
| Training records | 10,157 | 33,812 | 43,969 |
| Validation events | 16 | 41 | 57 |
| Validation records | 1,045 | 4,301 | 5,346 |

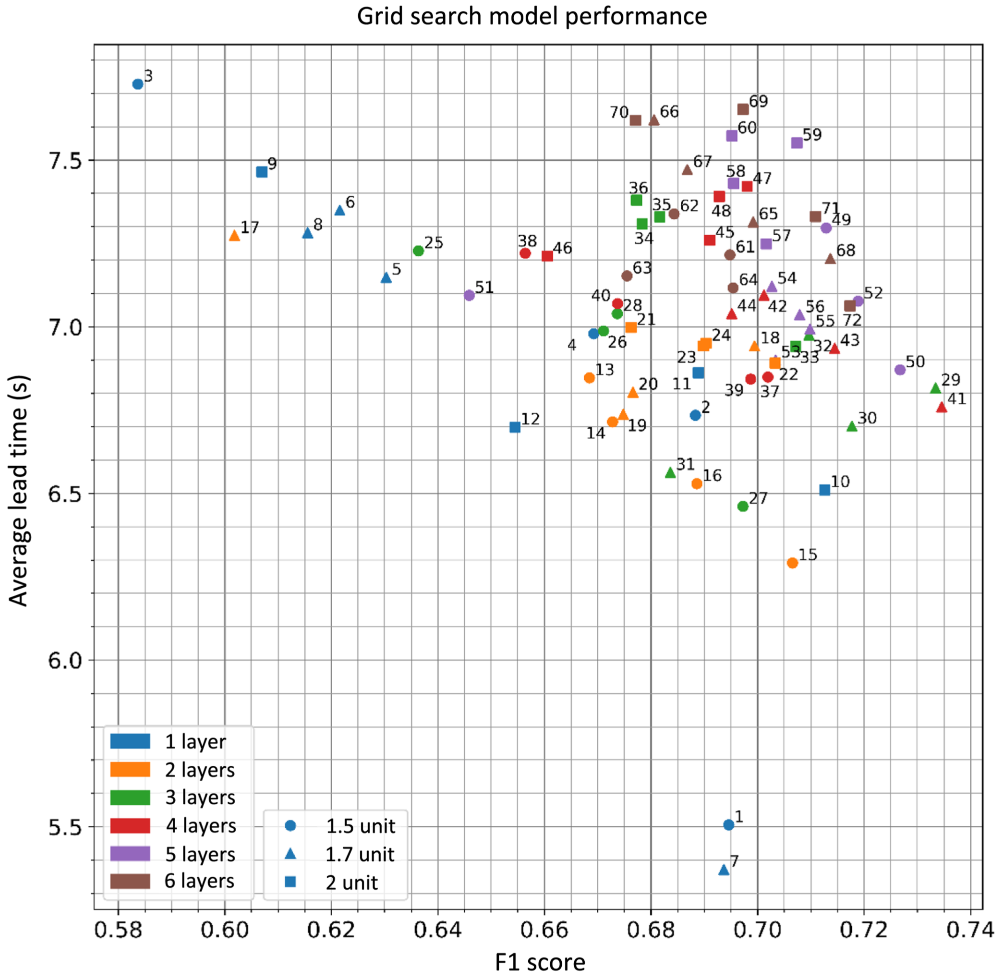


**Figure 5.** Hyperparameter grid search. The color indicates the number of LSTM layers. The shape indicates different base units of the layer. The best model is number 59 near the upper-right corner. The color version of this figure is available only in the electronic edition.

makes it computationally impractical. A study has shown that adding layers to extend the depth of neural networks is often computationally effective and has more expressive power (Lu *et al.*, 2017) . Inspired by their analysis on the different neurons at different layers, we tested models with one to six layers with different numbers of units. We set the number of units at each layer as $[U^{n+2}]$, in which $U$ is the base units, and $n$ is the $n$th LSTM layer starting from the output layer. There are three base units tested in this study: 1.5, 1.7, and 2.0. In the unit setting expression, the floor function guarantees the setting numbers of units are all positive integers.

Batch size is a hyperparameter that controls the number of training samples propagated through the network during a single iteration. During the training process, the gradient of the loss function is calculated on every feedforward pass and determines how the weights and biases in the network are to be adjusted. Using a subset (batch) of the whole training dataset, the loss gradient of the batch can be efficiently calculated, which represents an estimation of the loss gradient of the whole dataset. Large batch sizes contain more entries, and the batch loss gradients are closer to the overall loss gradient. In addition, this approach has the advantage of better parallelizing the computing process with GPUs. Small batch sizes often lead to a less accurate estimation of the overall loss gradient. However, the small-batch loss gradient can result in a more robust model because the large deviation escapes local minima or saddle points more easily. In this work, batch sizes of 256 and 512 are tested.

The final hyperparameter tested in this study is the optimization method. The most common optimization method is stochastic gradient descent (SGD), a stochastic approximation of the gradient descent method. Various methods based on SGD have been developed over the years. In this study, we tested two of them, RMSprop and Nadam. RMSprop controls the learning rate by an exponential decay average of all past squared gradients. Nadam combines RMSprop with a Nesterov accelerated gradient to provide faster convergence to the loss minimum.

The grid-search model performance is evaluated using the validation dataset and is compared on two metrics. The F1 score,

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \tag{1}$$

which is the harmonic mean of precision and recall, provides a comprehensive indicator of how well the model performs as a classifier. The other metric is the average warning lead time:

$$t_{\text{lead}} = t_{\text{PTA}>80} - t_{\text{alert}}. \tag{2}$$

It denotes a measure of how much early warning time the model provides. The two metrics are normalized to one and combined as the final score:

$$\text{score} = 50 \times \overline{F1} + 50 \times \overline{t_{\text{lead}}}, \tag{3}$$

in which the overline denotes the normalized quantities. In this study, we assign equal weight to both the alert accuracy and

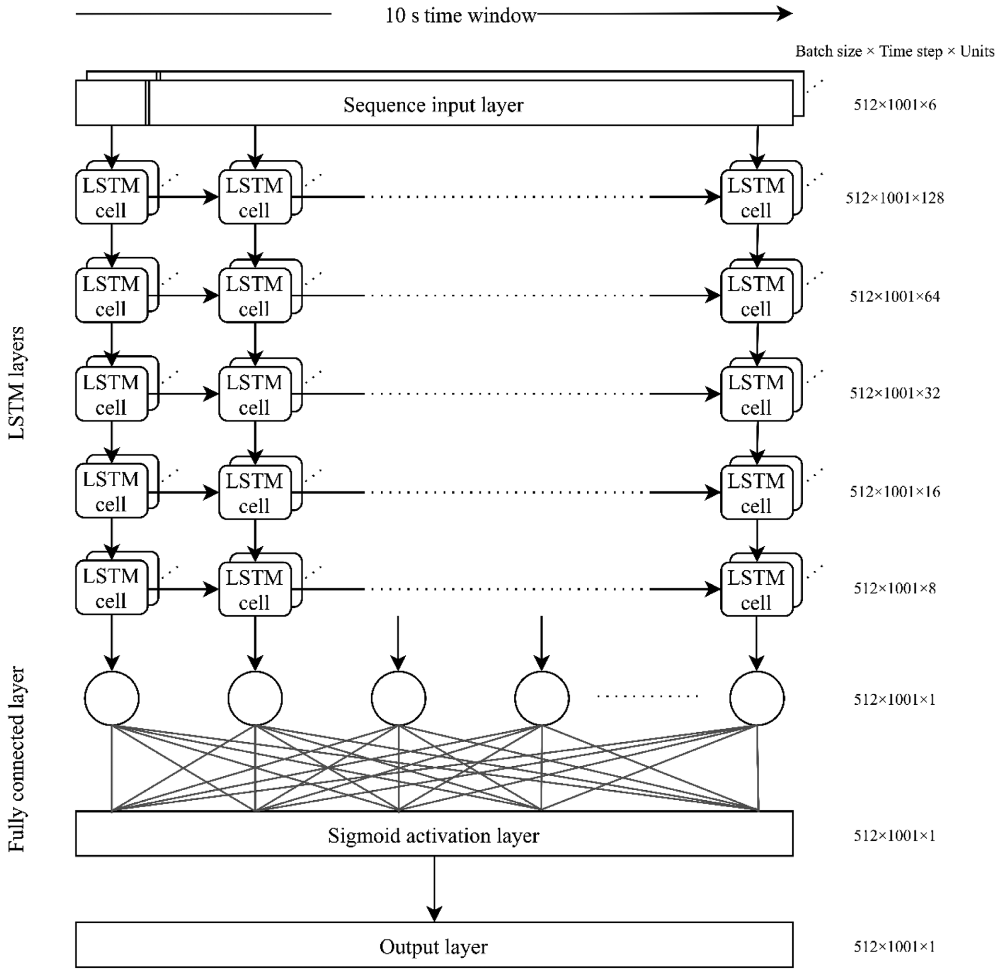


**Figure 6.** Model architecture of the best model.

warning time in the final score, meaning that the importance of both metrics is treated as equal.

Figure 5 shows the F1 score, lead time, and final score based on the validation dataset. There are 72 models shown in the grid search. The name of the model denotes the specific hyperparameters used, including layers (L), units (U), batch size (B), and optimization method (O). The final optimized model is L5U2B512Onadam, with the highest model score of 87.24. Figure 6 shows the architecture of the final optimized model. Figure 7 shows the training and validating curves of the training process of the final opitmized model. Although there is no sign of overfitting in this specific architecture, there are a fraction of models that shows a sign of rebound overfitting after many epochs. For those models, we follow the common practice of early stopping to avoid overtraining.

## LSTM model performance on the test data

Because of the selection bias from using the validation dataset as an index to choose the best performing model, a stand-alone test dataset is needed in addition to the training data and validation data to derive the truly unbiased performance of the model.

The unbiased test dataset for this study consists of three seismic events, two of which caused casualties and economic loss in Taiwan. The first event was the 2016 Meinong earthquake ($M_w = 6.4$), with a total of 328 records available for testing. The second event was the 2018 Haulien earthquake ($M_w = 6.4$), with 512 records available. The third event was a recent moderate earthquake in eastern Taiwan in May 2021 ($M_w = 4.9$), with 164 records available. We include this event to demonstrate that our model can also operate on smaller earthquakes. The total count of available records is 1004, in which 112 records are label 1 (PGA > 80 Gal) and 892 are label 0 (PGA < 80 Gal).

Strong-motion records of the three test events are preprocessed similarly to the training data, as depicted in Figure 2, but with a 20 s time window starting 1 s before the STA/LTA picker and continuing for 19 s after. A longer time window is chosen to accommodate the real scenario of the ground motion of the event. In practical use cases, even if the first few seconds of the *P* waves fail to trigger the LSTM model, the following *S* waves still ensures that an alert is issued at the cost of warning time. This is made possible by the cumulative time series property of the LSTM neural networks.

This study interprets the EEW problem as a sequential binary classification problem. The classification results can thus be divided into four possible outcomes: true positives, true negatives, false positives, and false negatives. True positives denote the records that should be alerted and are successfully predicted by the LSTM model. True negatives are records that should not be alerted and the model also classifies as no alerts. False positives are records that should not be alerted but are misclassified by the model and issued as alert, also called false alarms. False negatives are records that should alerts but were not alerted by the model, also called missed alarms. To measure the rate of the two types of error, two measures are introduced. The precision is the fraction of true positives among all

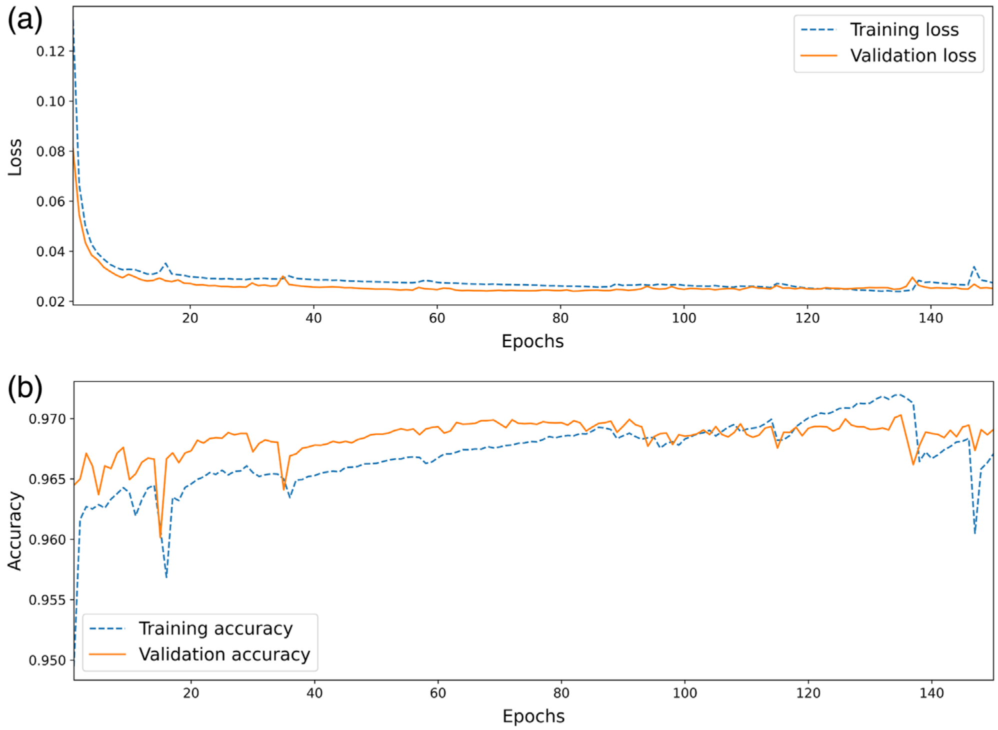


**Figure 7.** Illustration of training process of the final optimized model. (a) Training and validating loss curve. (b) Training and validating accuracy curve. The color version of this figure is available only in the electronic edition.

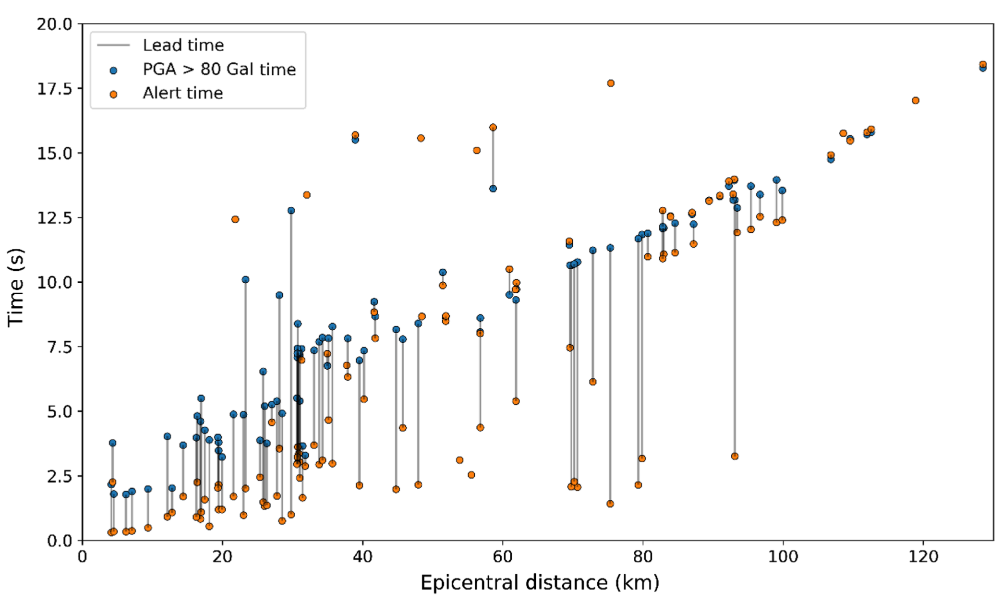


**Figure 8.** Lead time of the 100 true positive records. The time starts at the *P* arrival time on each record. The color version of this figure is available only in the electronic edition.

positive predictions made by the classifier, and the recall is the fraction of true positives among the actual positives.

The classification performance of the LSTM model gives a result of 18 false alarms and 0 missed alarms out of the 1004 total records, providing a precision rate of 86.15% and a recall rate of 100%. Within the 18 false positives, 10 falsely alarmed records had a PGA in the range of 60–80 Gal. Such close differences in PGA are difficult for the LSTM model to classify and are reasonably misinterpreted.

In addition to the classification performance, another important aspect of EEW systems is the early warning time provided. The earlier the warning arrives, the more opportunities there are for hazard mitigation procedures to be carried out. In this study, the warning lead time is defined as the time difference between the time at exceedance of 80 Gal for PGA and the model-provided alert time. The LSTM model lead time has an average of 2.64 s with a standard deviation of 2.79 s. The lead times of individual stations are shown in Figure 8, in which the stations within an epicentral distance of 50 km (regional warning blind zone) show good results in warning lead times, ranging from 2 to 5 s for most of them. For stations farther from the epicenter, several sites outside the range of 50 km fail to provide warnings within the duration of the initial *P* waves. This may be due to either the window length (10 s) used for training (not enough representative data for longer time windows) or simply the features' (three-component acceleration and velocity) attenuation over the longer travel path.

A true positive LSTM model output is presented in Figure 9a. The upper panel shows the recorded three-component acceleration signals, and the lower panel shows the probability output produced by the LSTM neural network at the corresponding time steps. The black-dashed line shows when

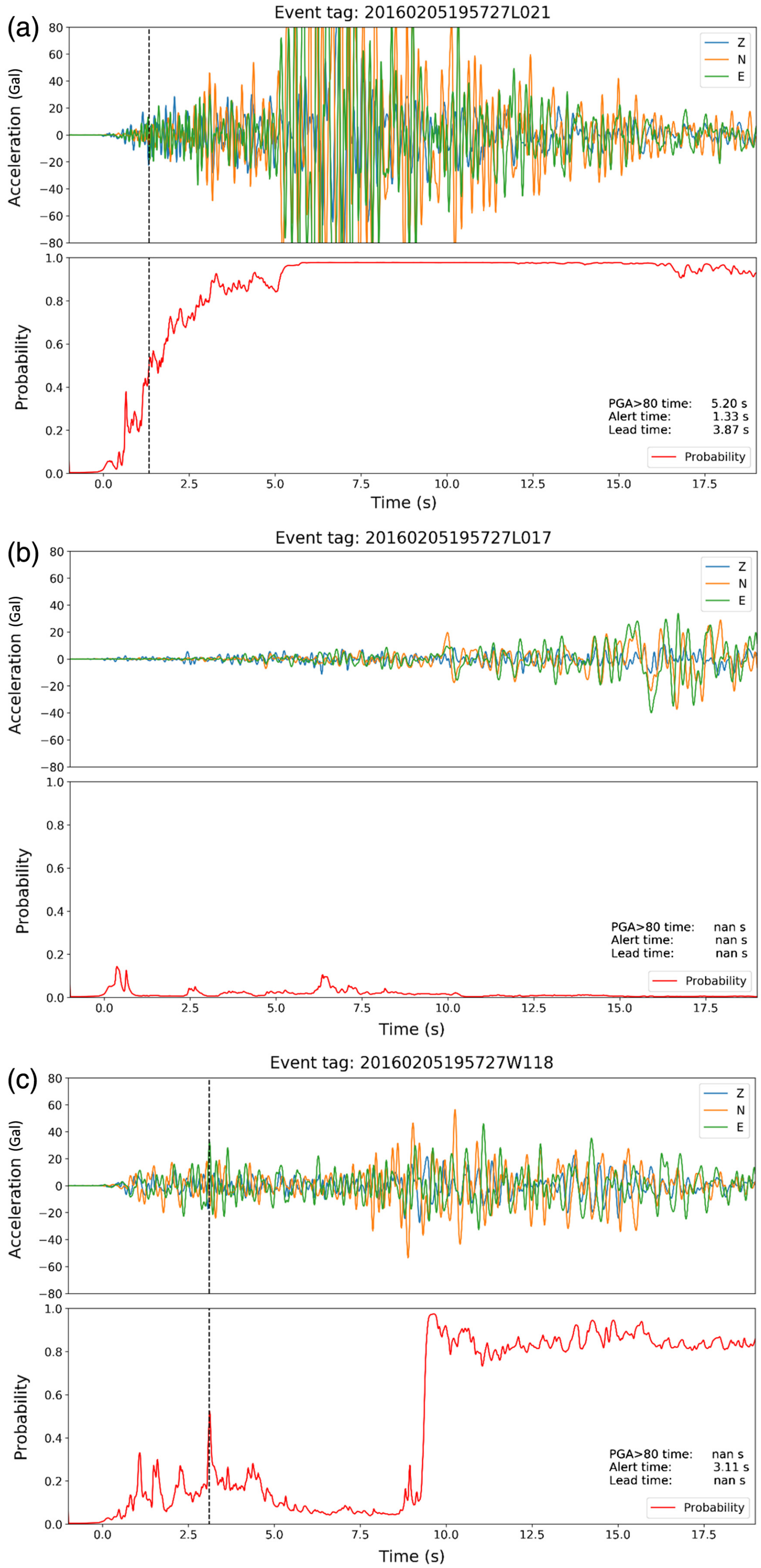


**Figure 9.** Model classification demonstration. (a) True positive record. (b) True negative record. (c) False positive record. The color version of this figure is available only in the electronic edition.

the output probability exceeds the 0.5 threshold and triggers a warning. EEW in the displayed station successfully alerted at the 1.33 s time mark, giving a lead time of up to 3.87 s.

Figure 9b shows an example of a true negative record. During the 20 s time span, the model output probability of severe ground shaking (PGA > 80 Gal) remains low and fails to exceed the probability threshold of 0.5.

A false positive record is displayed in Figure 9c, in which the model output probability exceeds 0.5 at the 3.11 s mark. However, the absolute ground acceleration did not surpass 80 Gal. This alert is thus considered a false alarm.

## Discussion

### Comparison with the $P_d$ method

A brief introduction of the two trigger settings on the *P*-alert device is provided before comparing the EEW performance of the LSTM approach with that of the $P_d$ approach. The *P*-alert device has two triggers for setting off an EEW. The first trigger is the vertical component $P_d$ set at a threshold of 0.35 cm within the PTW of 3 s. These parameter settings were selected according to a previous study (Hsieh *et al.*, 2015). The second trigger is the exceedance level of 80 Gal from the recorded acceleration signal (Wu *et al.*, 2016). In the following comparisons, the $P_d$ approach issues alerts by considering only the $P_d$ trigger.

For the 2016 Meinong earthquake, the LSTM approach shown in Figure 10a successfully alerted sites within the blind zone and covered a large area of stations the entire way up to Taichung, 150 km from

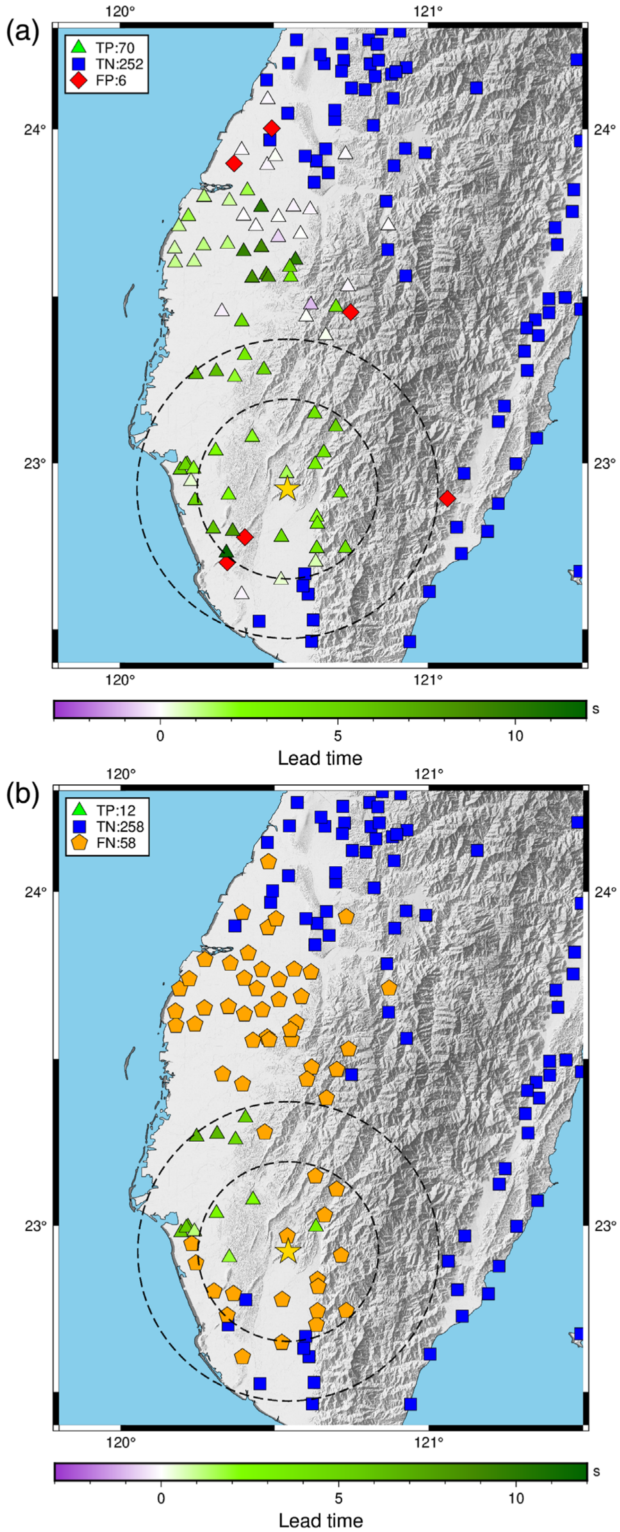


**Figure 10.** Earthquake early warning (EEW) performance for the 2016 Meinong earthquake. The triangles represent the true positive stations. The squares represent the true negative stations. The diamonds represent the false positive stations. The pentagons represent the false negative stations. The two dashed circles illustrate the 30 and 50 km radii around the epicenter. (a) The results of the LSTM method. (b) The results of the $P_d$ method. The color version of this figure is available only in the electronic edition.

the epicenter. Despite several false alerts, the LSTM model shows great performance in this case. The $P_d$ approach shown in Figure 10b successfully covers 11 stations within the 50 km blind zone but misses alerting a dozen stations distributed to the southwest, southeast, and northeast of the epicenter. This result may be caused by the rupture directivity of the 2016 Meinong earthquake, in which the event ruptured toward the northwest.

For the 2018 Hualien earthquake, as shown in Figure 11, both the LSTM and $P_d$ methods successfully issued warnings in the blind zone area; although the LSTM once again provided EEWs to farther stations than $P_d$, these areas could be covered by regional EEWs.

For the 2021 eastern Taiwan earthquake, as shown in Figure 12, the $P_d$ method did not issue any warning in the blind zone. On the other hand, the LSTM method successfully issued warnings. There are still several false-positive warnings because the eventual PGAs did not exceed 80 Gal, but came close to it (many of them 60 to 78 Gal).

The warning lead time differences between the two methods are displayed in Figure 13. For the 2016 Meinong event, the $P_d$ method provided longer lead times for most of the stations, whereas the LSTM approach had a longer lead time at most of the stations for the 2018 Hualien event. For the 2021 events, only LSTM method provides lead time. Although the $P_d$ method may achieve a faster alert for some of the stations, the LSTM model is capable of providing many more correct alerts both within and outside the 50 km blind zone. From these results, it may be natural to consider using the LSTM model as an additional trigger on $P$-alert devices, along with the $P_d$ trigger and PGA trigger (negative lead times can be eliminated to zero). Through these multiple triggers, the possibility of providing longer lead times can be maximized, with the different methods complementing each other. Figure 14 displays the lead time distributions for both the LSTM and $P_d$ approaches, showing a much larger number of alerts provided by the LSTM model.

## Design choices and technicalities

Throughout the implementation of LSTM for onsite EEWs, we made some design choices. The following choices may be useful for future machine learning projects.

We use filtered velocities as input. In our LSTM model, there are six channels of inputs. Three raw acceleration records and three filtered velocity records. These filtered velocities are derived from the accelerations. However, we found that the filtered velocities are quite useful and serve as stabilizers that effectively reduce false alarms in LSTM caused by reinforcement of previous fluctuations.

We use Japanese data to conquer unbalanced data issues. To train a model to classify binary outcomes, the true and false cases in the training sample should be equal or at least compatible. Otherwise, the machine learning algorithm will be heavily inclined to predict the outcome of the majority case

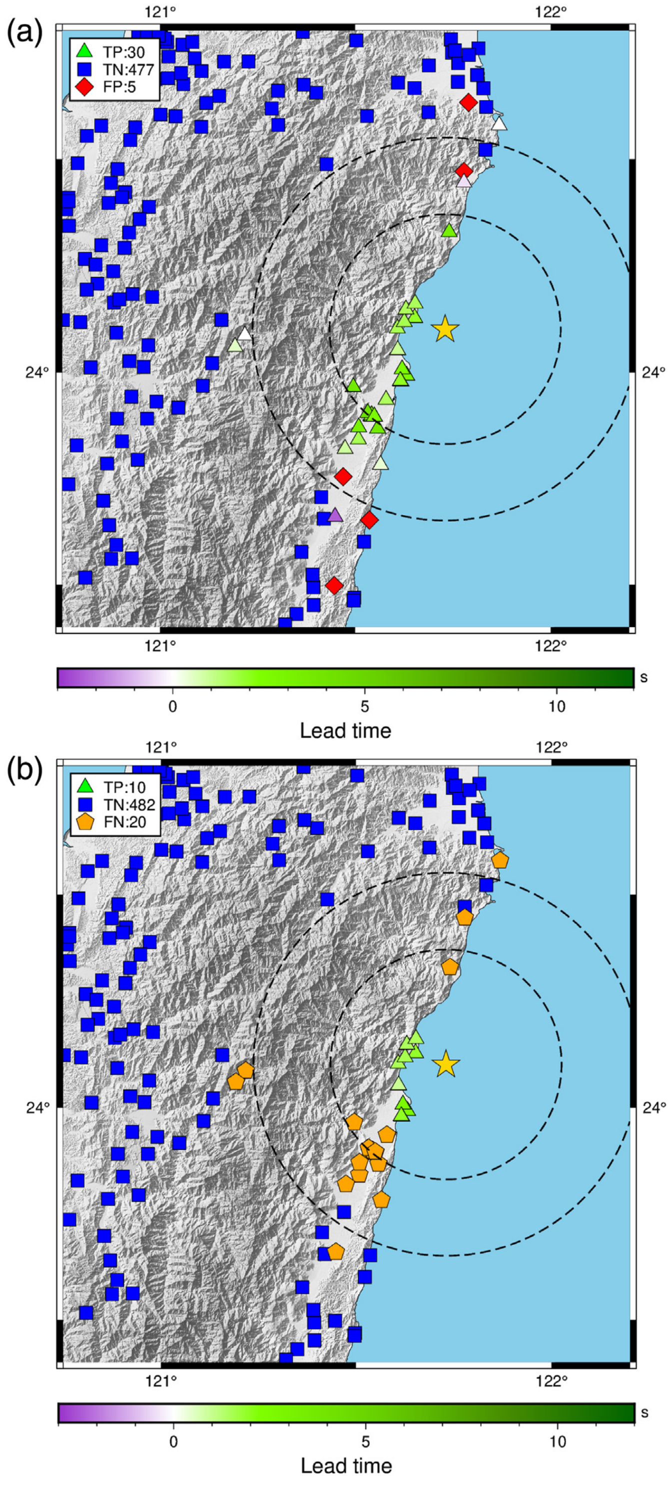


**Figure 11.** EEW performance for the 2018 Hualien earthquake. The triangles represent the true positive stations. The squares represent the true negative stations. The diamonds represent the false positive stations. The pentagons represent the false negative stations. The two dashed circles illustrate the 30 and 50 km radii around the epicenter. (a) The results of the LSTM method. (b) The results of the $P_d$ method. The color version of this figure is available only in the electronic edition.

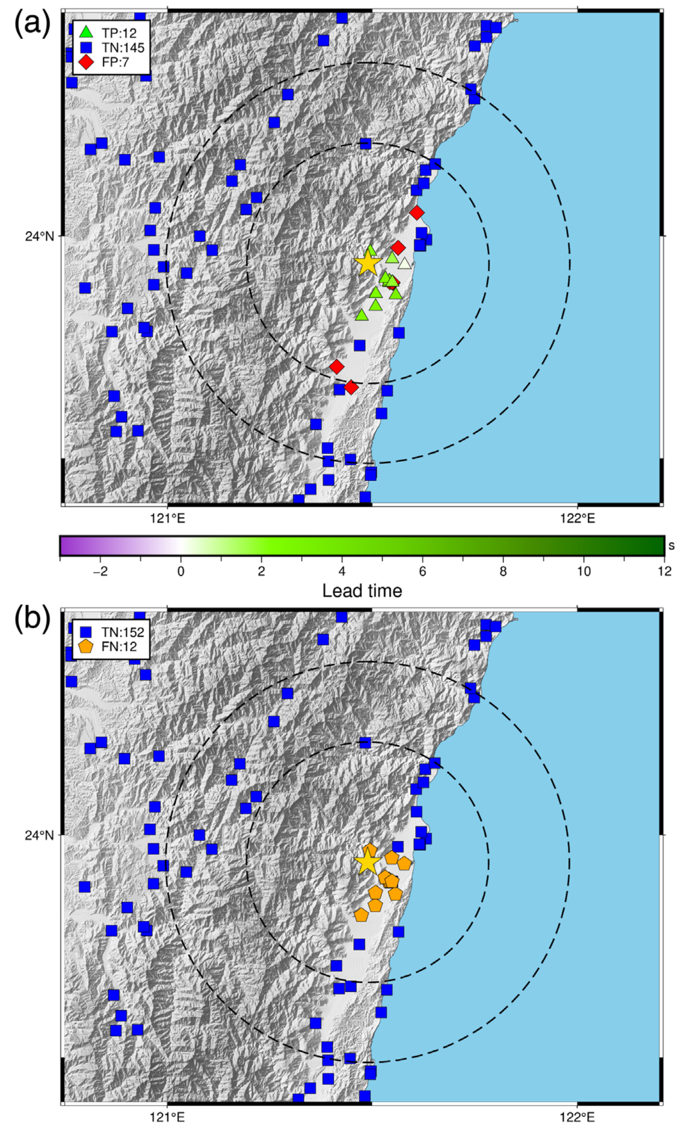


**Figure 12.** EEW performance for 2021 eastern Taiwan earthquake. The triangles represent the true positive stations. The squares represent the true negative stations. The diamonds represent the false positive stations. The pentagons represent the false negative stations. The two dashed circles illustrate the 30 and 50 km radii around the epicenter. (a) The results of the LSTM method. (b) The results of the $P_d$ method. The color version of this figure is available only in the electronic edition.

of the training data. This imbalance issue is especially difficult for EEW because the Gutenberg–Richter law states that there should be fewer large earthquakes. In practice, we decided to add more true cases into the training set using the Japanese seismic data. The final model works well with this setup and does not suffer from the inclination problem.

## Computational specification

The training part of the project is done on a desktop PC with Nvidia GeForce GTX-1060 GPU with 1280 CUDA cores and

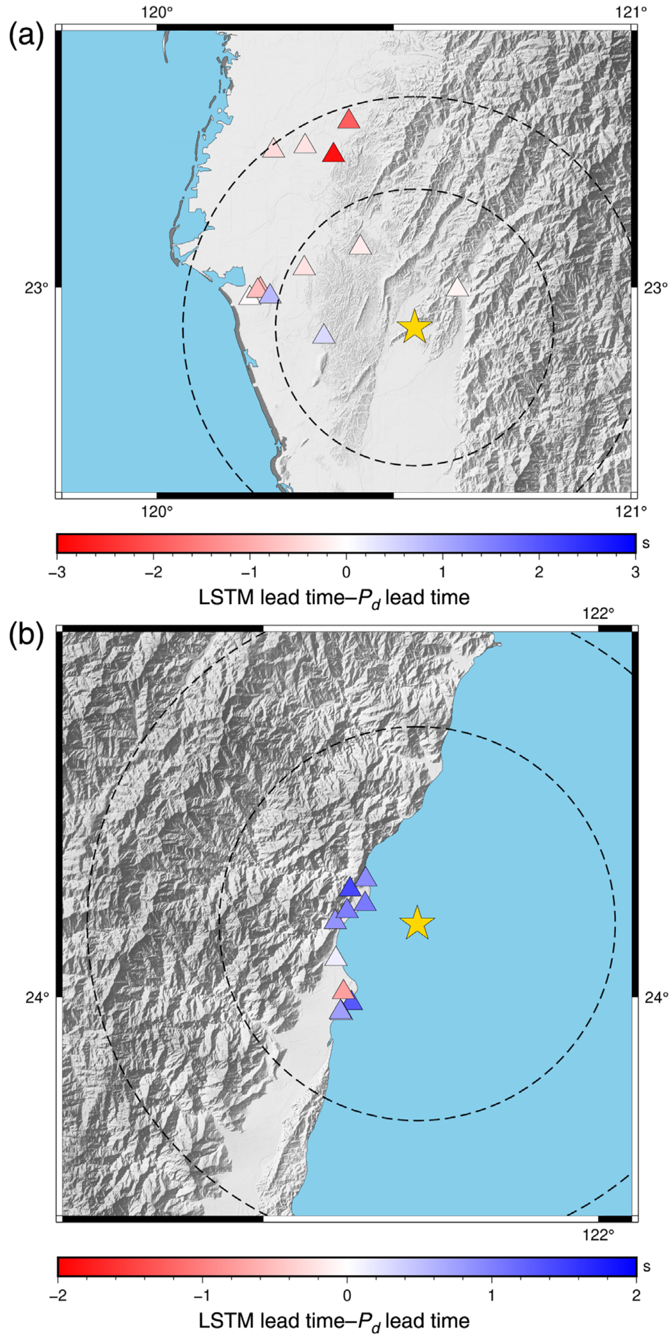


**Figure 13.** Lead time difference between the LSTM and $P_d$ methods. The two dashed circles illustrate the 30 and 50 km radii around the epicenter. (a) 2016 Meinong earthquake. (b) 2018 Hualien earthquake. The color version of this figure is available only in the electronic edition.

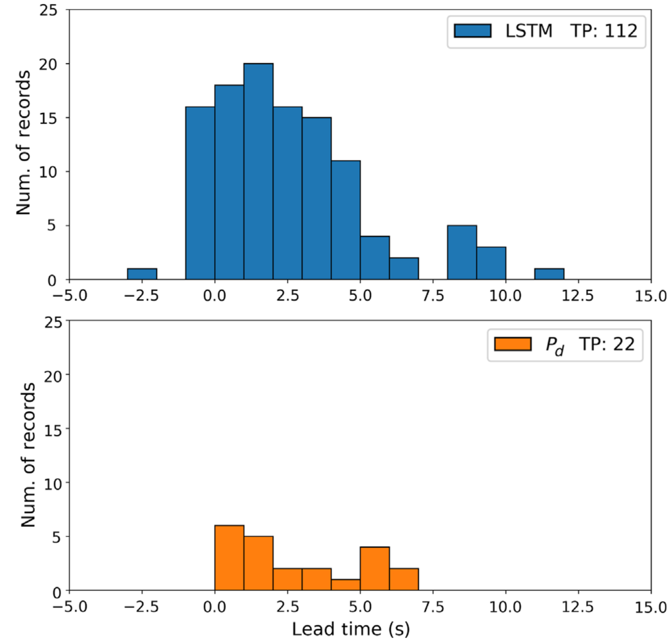


**Figure 14.** Lead time histogram of the LSTM and $P_d$ methods. The upper part shows the lead time distribution of the LSTM method. The lower part shows the lead time distribution of the $P_d$ method. The color version of this figure is available only in the electronic edition.

6 GB RAM. The training time for a typical five-layer LSTM model is about one hour, 25–30 s per epoch.

## Conclusions

In this project, we put the theoretical architecture of RNN deep learning into the context of a real-world onsite EEW problem. The LSTM model of RNN for onsite EEW demonstrates the ability to deliver fast and robust alerts to areas near the epicenter, which effectively mitigates human casualties and property losses. The positive results in terms of both accuracy and warning time using three out-of-training moderate-to-large earthquakes demonstrate the power of RNN-type deep learning. Currently, RNNs are the subject of intensive research, and many interesting variations have emerged over the years. We believe that employing state-of-art RNN techniques can further improve the performance of EEW in the future.

## Data and Resources

The strong-motion waveform records from the *P*-alert and K-NET networks can be downloaded at http://palert.earth.sinica.edu.tw/db/ and http://www.kyoshin.bosai.go.jp/, respectively (last accessed June 2021).

## Declaration of Competing Interests

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

## Authors Contributions

C.-Y. W. conducted the data analysis, trained and tested the machine learning model, and wrote this article. T.-C. H. initiated the project, developed the machine learning model, and wrote this article. Y.-M. W. supervised this work and advised. All authors discussed the training results and contributed to the final version of this article.

## Acknowledgments

This work, "Using LSTM Neural Networks for Onsite Earthquake Early Warning," 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. In additon, this work was also supported by the Research Center of Future Earth of National Taiwan University under Grant Number 107L901002 and financially supported by the NTU Research Center for Future Earth from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

## References

Allen, R. M., and D. Melgar (2019). Earthquake early warning: advances, scientific challenges, and societal needs, *Annu. Rev. Earth Planet. Sci.* **47,** no. 1, 361–388, doi: 10.1146/annurev-earth-053018-060457.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function, *Math. Control Signal Syst.* **2,** 303–314, doi: 10.1007/BF02551274.

Geller, R. J. (1997). Earthquake prediction: A critical review, *Geophys. J. Int.* **131,** no. 3, 425–450, doi: 10.1111/j.1365-246X.1997.tb06588.x.

Gers, F. A., J. Schmidhuber, and F. Cummins (2000). Learning to forget: Continual prediction with LSTM, *Neural Comput.* **12,** no. 10, 2451–2471, doi: 10.1162/089976600300015015.

Gers, F. A., N. N. Schraudolph, and J. Schmidhuber (2005). Learning precise timing with LSTM recurrent networks, *J. Mach. 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,** no. 6, 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.

Kanamori, H., E. Hauksson, and T. Heaton (1997). Real-time seismology and earthquake hazard mitigation, *Nature* **390,** 461–464, doi: 10.1038/37280.

Leshno, M., V. Y. Lin, A. Pinkus, and S. Schocken (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Netw.* **6,** no. 6, 861–867, doi: 10.1016/S0893-6080(05)80131-5.

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,** no. 10, 4773–4779, doi: 10.1029/2018GL077870.

Liu, C. N. (2019). Using low-cost seismometers and machine learning on onsite earthquake early warning, Master's dissertation, Geophysical Research Abstracts, Vol. 21, EGU General Assembly 2019.

Lu, Z., H. Pu, F. Wang, Z. Hu, and L. Wang (2017). The expressive power of neural networks: A view from the width, *NIPS 2017*, 6232–6240, doi: 10.5555/3295222.3295371.

Mousavi, S. M., and G. C. Beroza (2019). A machine-learning approach for earthquake magnitude estimation, *Geophys. Res. Lett.* **47,** no. 1, e2019GL085976 , doi: 10.1029/2019GL085976.

Saad, O. M., A. G. Hafez, and M. S. Soliman (2021). Deep learning approach for earthquake parameters classification in earthquake early warning system, *IEEE Geosci. Remote Sens. Lett.* **18,** no. 7, 1293–1297, doi: 10.1109/LGRS.2020.2998580.

Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it, *Proc. IEEE* **78,** no. 10, 1550–1560, doi: 10.1109/5.58337.

Williams, R. J., and D. Zipser (1989). A learning algorithm for continually running fully recurrent neural networks, *Neural Comput.* **1,** no. 2, 270–280, doi: 10.1162/neco.1989.1.2.270.

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 T. L. Teng (2002). A virtual subnetwork approach to earthquake early warning, *Bull. Seismol. Soc. Am.* **92,** no. 5, 2008–2018, doi: 10.1785/0120010217.

Wu, Y. M., J. K. Chang, T. C. Shin, N. C. Hsiao, W. H. K. Lee, and T. L. Teng (1999). Development of an integrated earthquake early warning system in Taiwan-case for the Hualien area earthquakes, *TAO* **10,** no. 4, 719–736, doi: 10.3319/TAO.1999.10.4.719(T).

Wu, Y. M., D. Y. Chen, T. L. Lin, C. Y. Hsieh, T. L. Chin, W. Y. Chang, W. S. Li, and S. H. Ker (2013). A high-density seismic network for earthquake early warning in Taiwan based on low cost sensors, *Seismol. Res. Lett.* **84,** no. 6, 1048–1054, doi: 10.1785/0220130085.

Wu, Y. M., H. Kanamori, R. M. Allen, and E. Hauksson (2007). Determination of earthquake early warning parameters, $\tau_c$ and $P_d$, for southern California, *Geophys. J. Int.* **170,** no. 2, 711–717, doi: 10.1111/j.1365-246X.2007.03430.x.

Wu, Y. M., W. T. Liang, H. Mital, W. A. Chao, C. H. Lin, B. S. Huang, and C. M. Lin (2016). Performance of a low-cost earthquake early warning system (P-alert) during the 2016 $M_L$ 6.4 Meinong (Taiwan) earthquake, *Seismol. Res. Lett.* **87,** no. 5, 1050–1059, doi: 10.1785/0220160058.

**Manuscript received 20 July 2021**
**Published online 5 January 2022**