Early Peak Ground Acceleration Prediction for On-Site Earthquake Early Warning Using LSTM Neural Network

T. Y. Hsu[1,2] and A. Pratomo[1]

[1] Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan.

[2] National Center for Research on Earthquake Engineering, Taiwan.

Corresponding author: T. Y. Hsu (tyhsu@mail.ntust.edu.tw)

Key Points:

- An on-site earthquake early warning technique is proposed to predict PGA using long short-term memory neural network.
- The predicted PGA values were quite promising but were generally overestimated.
- The predicted PGA of the Chi-Chi earthquake was more accurate than those predicted using a support vector regression approach.

Abstract

On-site earthquake early warning techniques, which issue alerts based on seismic waves measured at a single station, are promising, and have performed quite successfully during some damaging earthquakes. Conventionally, most existing techniques extract several P-wave features from the first few seconds of seismic waves after the trigger to predict the intensity or destructiveness of an incoming earthquake. This type of technique neglects the behavior of temporal varying features within P waves. In other words, the characteristics of data sequences are not considered. In this study, a long short-term memory (LSTM) neural network, which was capable of learning order dependence in seismic waves, was employed to predict the PGA of the coming earthquake. A dense LSTM architecture was proposed and a large data set of earthquakes was used to train the LSTM model. The general performance of the LSTM model indicated that the predicted PGA values were quite promising but were generally overestimated. However, the predicted PGA of the Chi-Chi earthquake data set, whose fault rupture was complex and long, using the proposed LSTM model was more accurate than the PGA predicted in a previous study using a support vector regression approach. In addition, an alternative alert criterion, which issues alerts when the predicted PGA exceeds the threshold in successive time windows, is presented, and the performance of the proposed LSTM model when different PGA thresholds are considered is also discussed.

Plain Language Summary

On-site earthquake early warning techniques have successfully issued alerts during some damaging earthquakes. Conventionally, most existing techniques draw out features from the first few seconds of seismic waves after the trigger to predict the intensity or destructiveness of an incoming earthquake. This type of technique neglects the behavior of temporal varying features within seismic waves. In this study, a long short-term memory (LSTM) neural network, which was capable of learning order dependence in seismic waves, was employed to predict the peak ground acceleration of the coming earthquake. A dense LSTM architecture was proposed and a large data set of earthquakes was used to train the LSTM model. The general performance of the LSTM model indicated that the predicted PGA values were quite promising but were generally overestimated. However, the predicted PGA of the Chi-Chi earthquake data set, whose fault rupture was complex and long, using the proposed LSTM model was more accurate than the PGA predicted in a previous study using a support vector regression approach.

1 Introduction

After an earthquake rupture occurs, it is possible for an earthquake early warning (EEW) system to send alerts to earn several seconds to tens of seconds of lead time before destructive seismic waves strike. Regional EEW techniques estimate the magnitude and location of an earthquake based on seismic waves measured at several stations, while on-site EEW techniques estimate the earthquake's intensity or identify destructive earthquakes based on seismic waves measured at only a single station. Successful EEW alerts were issued during several large earthquakes using both regional (e.g., Fujinawa & Noda 2013; Cuéllar et al. 2014; Yamada et al.

2014; Kodera et al., 2016) and on-site techniques (e.g., Hsu et al. 2016; Hsu et al. 2018; Wu et al. 2019). On-site techniques could attain a longer lead time at the region close to the epicenter during several large earthquakes (e.g., Hsu et al. 2018; Wu et al. 2019). Nowadays, a number of EEW systems issue alerts based on hybrid techniques (e.g., Kodera et al. 2018; Kodera et al. 2019; Hsu et al. 2021). The status of these EEW techniques and systems have been addressed by several recent review papers (e.g., Allen & Melgar 2019; Cremen and Galasso 2020; Wald 2020).

In order to predict the source characteristics, e.g., magnitude, distance from the epicenter, and seismic intensity (peak ground acceleration [PGA] and peak ground velocity [PGV]), of a coming earthquake at the stage of initial seismic waves, artificial intelligence techniques have recently been employed, especially for on-site EEW. Often, several P-wave features are extracted from the initial waveforms after the trigger and inputted into the prediction model constructed using artificial intelligence techniques for on-site EEW. For instance, the integration of the absolute value of acceleration, velocity, and displacement time histories, which describe the envelope of the waveforms in a simplistic way, is used as the input in multilayer feedforward neural networks (Bose et al. 2012). Similarly, effective predominant period, cumulative absolute velocity, integral of the squared velocity, peak acceleration, peak velocity, and peak displacement are used as the input in the prediction model constructed using the support vector regression (SVR) technique (Hsu et al. 2013). In this model, only some important P-wave features are used, so other important underlying information in the waveforms may be ignored. Hsu et al. (2021) attempted to exploit convolutional neural networks to automatically extract useful features from previously measured P-wave data without losing too much information in the seismic waveforms and successfully predict the PGA of coming earthquakes.

However, different prediction models are required to be constructed using the above-mentioned approaches based on artificial intelligence techniques if different time-window lengths of P waves are employed after the trigger. For instance, the typical length of these approaches is 3 s. In addition, temporal variation within P waves is also neglected in the above-mentioned approaches, and only lumped information is considered. In other words, the characteristics of data sequences are not considered in these approaches.

Long short-term memory (LSTM) is a class of recurrent neural networks that provides a connection between previous inputs and current inputs. The network stores information selectively from previous time steps, which then acts as an additional input in the current time step. The LSTM model can capture the dependencies between time steps and has been proven to be very useful for some sequential data problems such as speech recognition (Senior et al., 2020), prediction of stock prices (Murtaza et al. 2017), and short-term traffic forecasting (Zhao et al., 2017). Wang et al. (2020) developed an LSTM neural network for on-site EEW. The LSTM model was trained to predict whether or not a large earthquake (with a PGA of more than 80 Gal) was imminent. In other words, the LSTM model only predicted the class of earthquake. However, PGA values are quite useful for different application purposes. For issuing alerts to the general public, 25 Gal is often employed as the threshold in Taiwan (Hsu et al. 2018). The machines in semiconductor fabrication plants are sensitive to small earthquakes and could be shut down by the emergency stop due to registering small vibration signals, e.g., with a PGA of several Gal. In contrast, high-speed trains may have different emergency responses to a PGA of 40, 80, or 120 Gal. The nuclear power plants in Taiwan are designed to resist earthquakes with a PGA of several hundred Gal. Therefore, in this study, we constructed an LSTM model to predict

PGA values. The methodology of the proposed LSTM approach and the earthquake data used are explained in the second section. The results are discussed in the third section, and conclusions are drawn in the final section.

2 Methodology and Earthquake Data

2.1 A Brief Description of the LSTM Neural Network

Each LSTM unit, shown in Figure 1, contained an internal cell, a forget gate, an input gate, and an output gate (Hochreiter and Schmidhuber 1997). The internal cell stored values over time intervals via a self-recurrent connection to the previous time step, and value updating was controlled by the gates. The algorithm of an LSTM unit was defined as

$$f_t = \sigma\left(w_f[H_{t-1}, X_t] + b_f\right) \quad (1),$$

$$i_t = \sigma(w_i[H_{t-1}, X_t] + b_i) \quad (2),$$

$$\tilde{C}_t = tanh(w_c[H_{t-1}, X_t] + b_c) \quad (3),$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4),$$

$$o_t = \sigma(w_o[H_{t-1}, X_t] + b_o) \quad (5), \text{ and}$$

$$H_t = o_t \times tanh(C_t) \quad (6).$$

The forget gate took the current time step input (X_t) and the hidden state (H_{t-1}) from the previous time step as concatenated input, multiplied the forget gate weight (w_f), added the forget gate bias (b_f), and applied the activation function. The input gate was provided with a concatenated input (H_{t-1},X_t) with its weights (w_i) and biases (b_i) along with the candidate cell state (C_t). The cell state was updated with the forget gate (f_t) by applying both the previous cell state (C_{t-1}) additional to the input gate (i_t) and the candidate cell state ($\tilde{C}_t$). The input gate determined whether or not the information in the current time step was memorized and stored in the cell state. The output gate (o_t) managed the output to update the cell state with its specific weights (w_o) and biases (b_o) using concatenated input similar to that of the input gate.

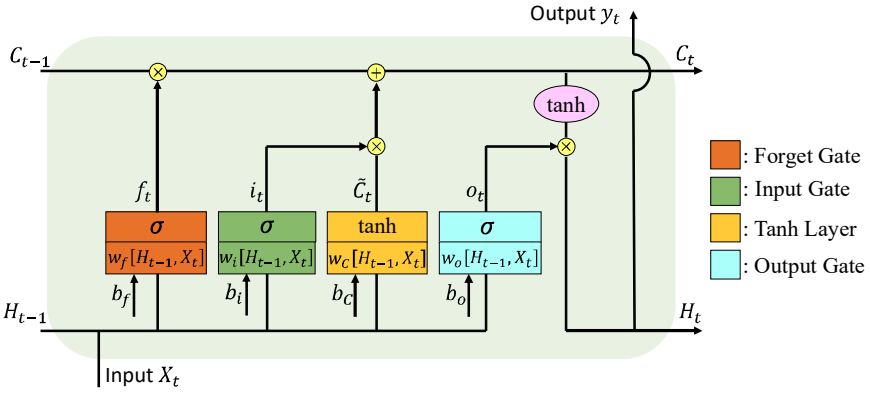

Figure 1. The LSTM unit.

2.2 Design of the Input and Output of the LSTM Model

Before we introduce the neural network architecture, the design of the input and output of the proposed LSTM model is explained in this section. The goal of the proposed LSTM approach was to predict the PGA of the coming earthquake, as soon as possible after the arrival of the P wave, based on the measured seismic waves. Hence, the target output of the LSTM model, shown in Figure 2, was designated as a step function with the amplitude of the PGA:

$$\boldsymbol{h}(t) = \begin{cases} 0 & , t_i \leq t < t_p \\ PGA & , t_p \leq t \leq t_e \end{cases} \quad (7),$$

$$t_e = t_p + t_w\,; \qquad t_w = 0.5,1.0,1.5,2.0,2.5,3.0 \quad (8), \text{ and}$$

$$t_i = 4 - t_w \quad (9),$$

where t_p was the P-wave arrival time determined based on the short-term average over long-term average (STA/LTA) algorithm; t_i and t_e were start time and end time, respectively; and t_w was the length of the time-window between t_p and t_e.

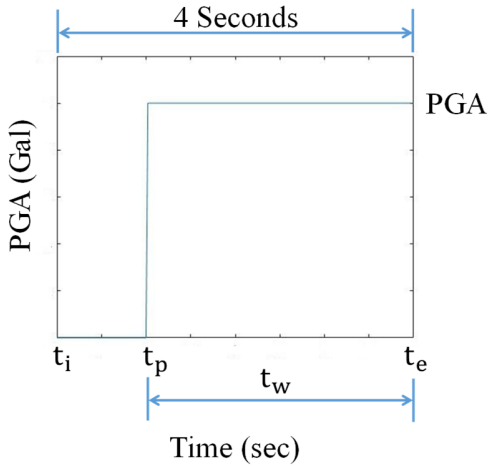

Figure 2. The schematic diagram of the target output of the LSTM model.

In this study, six t_w lengths with an interval of 0.5 s were employed; hence, corresponding six truncated lengths with 4 s of "windowed" inputs were also employed. For instance, when $t_w = 0.5$, only seismic waves with a time-window length of 0.5 s after the trigger were employed in the inputs, while those with a time-window length of 1.0 s after the trigger were employed when $t_w = 1.0$, and so on. An example of the input and output targets of one typical earthquake is shown in Figure 3. The windowed input when $t_w = 0.5$ is illustrated in

Figure 3(a); Figure 3(a) to 3(c) show three acceleration time histories and three velocity time histories for different time-window lengths. The velocity time histories were obtained from integration of the acceleration time histories. The corresponding target output time histories are shown in Figure 3(d) to 3(f).

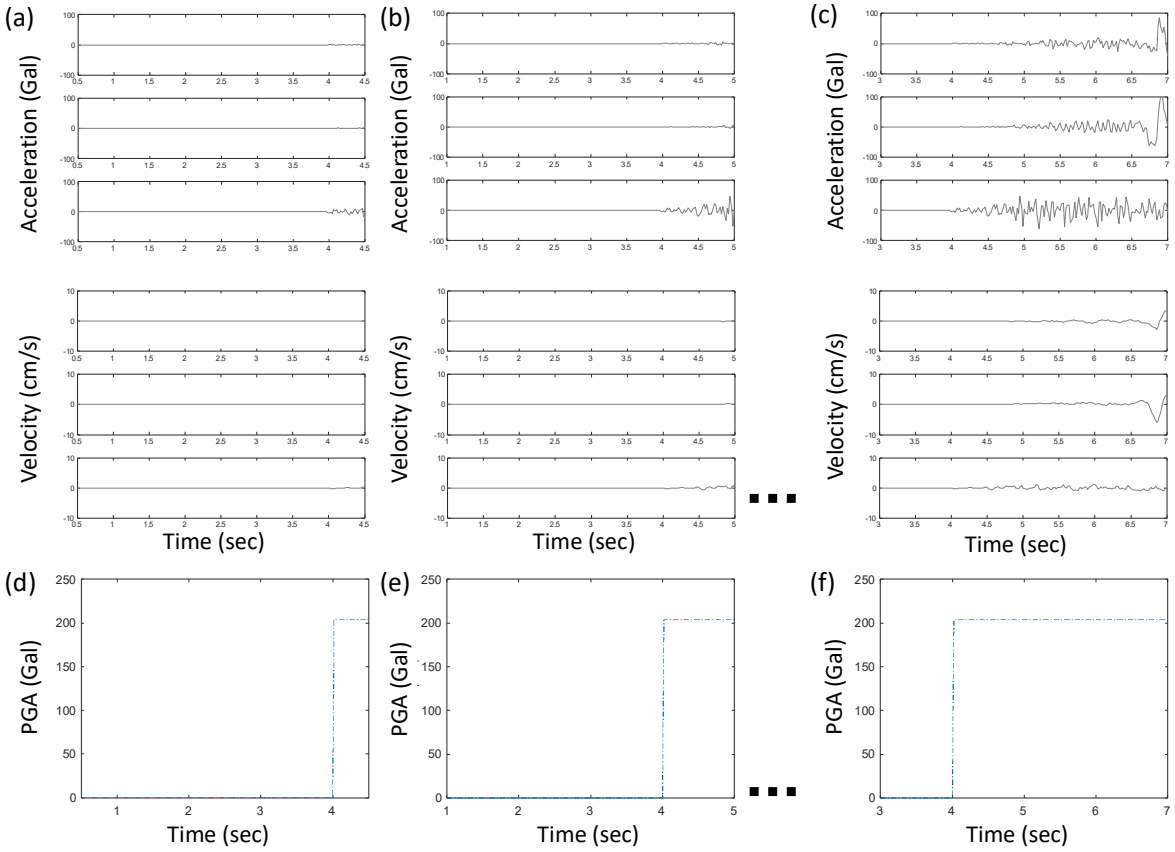

Figure 3. A typical example of the input/output of the LSTM model. (a) input of $t_w = 0.5$ s; (b) input of $t_w = 1.0$ s; (c) input of $t_w = 3.0$ s; (d) target output of $t_w = 0.5$ s; (e) target output of $t_w = 1.0$ s; and (f) target output of $t_w = 3.0$ s.

To save computational effort, the sampling rate of the input and output was reduced. Since the energy of seismic waves tended to be below 25 Hz, the original seismic waves were down-sampled from 200 Hz to 50 Hz using an 8^{th} order Chebyshev Type I infinite impulse response lowpass filter with a cutoff frequency of 40 Hz. Hence, the number of samples of each seismic wave with a window length of 4 s, more precisely, 4.02 s, was 201. As a result, the dimensions of the input and output of the LSTM model were 201 × 6 and 201 × 1, respectively. Both the input and output data were normalized using the RobustScaler normalization algorithm which removed the median and scaled the input and output data according to their interquartile range.

2.3 Proposed LSTM Architecture

Because a previous study showed that the results of PGA prediction or large earthquake detection using conventional LSTM architecture were not acceptable (Wang et al. 2020), a dense LSTM architecture was employed in this study for PGA prediction. The dense LSTM architecture contained several LSTM layers, with each layer consisting of a parallel LSTM series, as shown in Figure 4. After the final dense LSTM layer, two fully connected layers were employed to generate the output data. The first layer consisted of a parallel ReLU (Rectified Linear Unit) series, which was a series of ReLU activation functions, while the second layer was a series of linear activation functions. The final LSTM model is shown in Figure 4. The number of LSTM layers and the number of parallel ReLU series in the first fully connected layer were determined based on a grid-search approach. The first hyper-parameter in the grid search was the number of LSTM layers, N, and the number of LSTM series in each layer was determined automatically as 2^{N+2} in descending order. The number of LSTM layers considered in this study were [4, 5, 6]. The other hyper-parameter in the grid search was the number of ReLU series in the first fully connected layer. The number of ReLU series in the first fully connected layer considered in this study were [8, 16, 32].

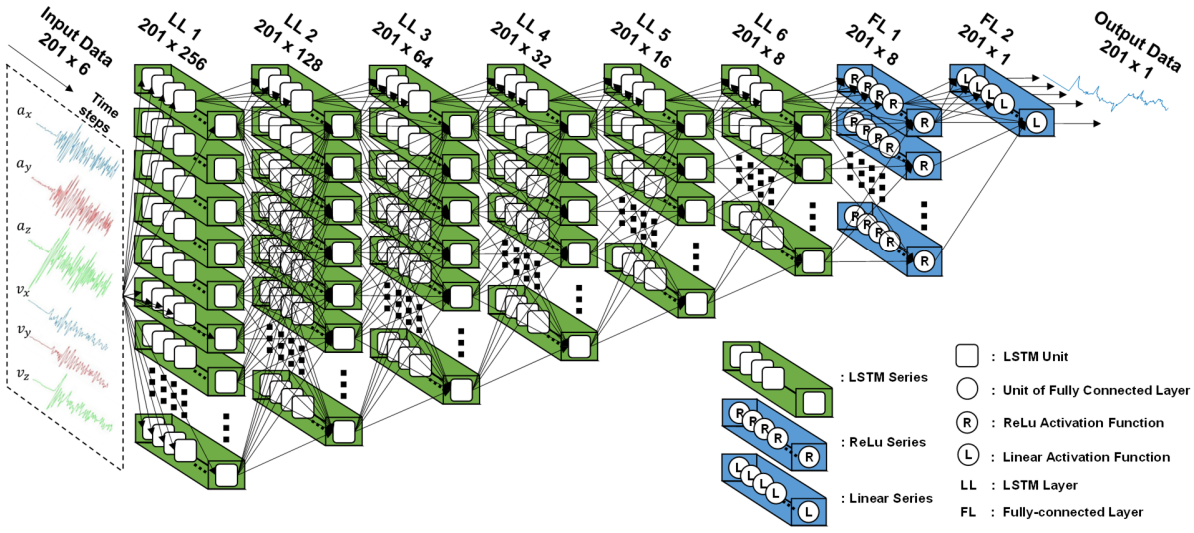

Figure 4. Proposed architecture of the LSTM model.

2.4 Earthquake Data Sets

The Central Weather Bureau (CWB) of Taiwan collects high-quality data on ground motions at several hundred seismic stations around Taiwan via the Taiwan Strong Motion Instrument Program (TSMIP). The TSMIP data set of 91,128 earthquakes during the period from July 29, 1992, to December 31, 2006, was employed in this study. Approximately half of the TSMIP data, 40,000 earthquakes in total, for each PGA range defined by the CWB was used for training (32,000 earthquakes) and validation (8,000 earthquakes) of the proposed LSTM model, as listed in Table 1. The P-wave arrival time was determined based on the STA/LTA algorithm.

In addition, two data sets of the most damaging recent earthquake events in Taiwan were employed to test the performance of the proposed LSTM approach. The 2016 Meinong earthquake event (Mw = 6.53) and the 2018 Hualien earthquake event (Mw = 6.2) in Taiwan resulted in 117 deaths and 17 deaths, respectively.

Table 1. The number of earthquakes of different ranges in the TSMIP data set.

	PGA (Gal)							
Data set	0.8~2.5	2.5~8	8~25	25~80	80~250	250~800	800~	Total
Whole TSMIP	117	31,192	43,334	13,366	2,843	268	8	91,128
Training	46	12,477	12,882	5,347	1,138	108	2	32,000
Validation	13	3,119	3,220	1,336	284	26	2	8,000

2.5 Classification Metrics and Lead Time

In this study, the PGA threshold for general application was assumed to be 25 Gal. In the field of machine learning, the confusion matrix of false positives (FPs), false negatives (FNs), true positives (TPs), and true negatives (TNs) allows performance to be easily visualized, as shown in Figure 5(a). Note that if a correct alert is issued after the observed PGA reaches the threshold, i.e., negative lead time, it will be classified as an FN. Lead time, t_{lead}, is defined as the time of issuing an alert minus the time when the observed PGA reaches the threshold. The classification metrics of precision, recall, and F1 score are calculated as

$$precision = \frac{TP}{TP+FP} \quad (10),$$

$$recall = \frac{TP}{TP+FN} \quad (11), \text{ and}$$

$$F1\ score = \frac{2\times precision\times recall}{(precision+recall)} \quad (12).$$

For community earthquake early warning, the false alert ratio (FAR) and missed alert ratio (MAR) are more straightforward to observe than alert performance, and can be calculated as

$$FAR = \frac{FP}{TP+FP} \quad (13) \text{ and}$$

$$MAR = \frac{FN}{TP+FN} \quad (14).$$

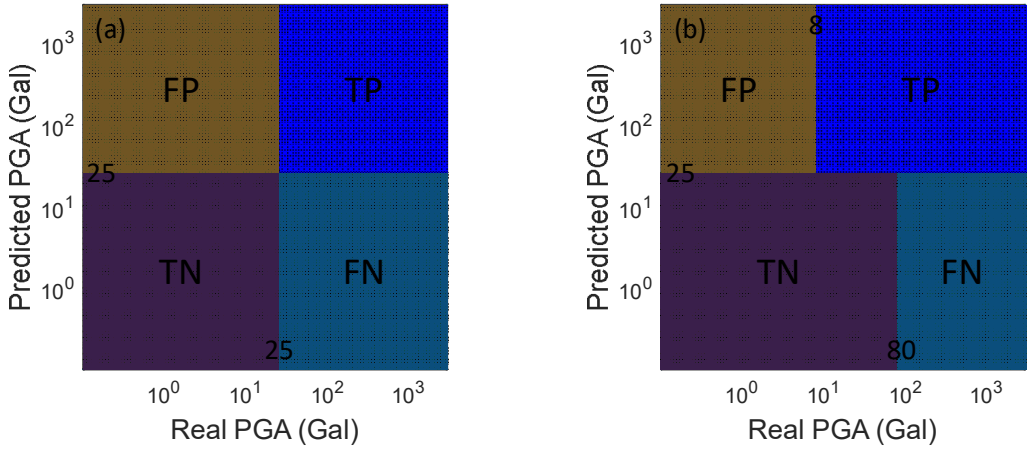

Figure 5. The confusion matrix when the threshold is designated as 25 Gal (a) without tolerance, and (b) with tolerance.

3 Results and Discussion

3.1 Training and Validation

When large deviations exist in the data, e.g., the PGA of large earthquakes can be almost 1,000 times those of small ones, the root mean squared logarithmic error (RMSLE) is often employed to estimate the loss in prediction accuracy, which is defined as

$$E = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(log\left(y_j^p + 1\right) - log\left(y_j^r + 1\right)\right)^2} \quad (15),$$

where y_j^r and y_j^p are the actual and predicted PGA of the j[th] earthquake, respectively, and N is the total number of earthquakes. The parameters in the LSTM model were updated based on the Nadam optimizer with a learning rate = 0.0001 and batch size = 32 during the training process. The training process stopped when validation loss kept increasing over three successive epochs; the loss history during a typical training process of the LSTM is shown in Figure 6. The LSTM model was trained using Python language on an Intel® Core™ i7-6700 CPU 3.40 GHz computer.

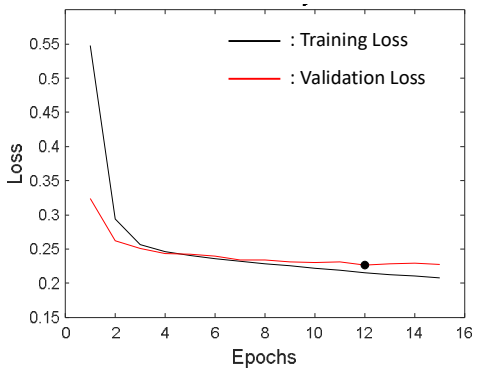

Figure 6. Loss history during a typical training process.

3.2 Grid Search Results

In this study, a grid search approach was employed to determine the two hyper-parameters (the number of LSTM layers and the number of ReLU series in the first fully connected layer) in the LSTM architecture. Both PGA prediction accuracy and lead-time length were considered in the evaluation of the performance of the LSTM architecture. Specifically, the model score, S_M, of the LSTM architecture was calculated using the Min-Max normalized F1 score, $\tilde{F}_1$, and the Min-Max normalized average lead-time, $\tilde{t}_{lead}$, (without tolerance) as

$$S_M = \tilde{F}_1 + \tilde{t}_{lead} \tag{16}$$

Table 2 shows the F1 score, average lead time, and model score of the nine combinations of the two hyper-parameters. The highest model score was achieved by the 7th LSTM model. Thus, the selected LSTM model consisted of six hidden LSTM layers and eight ReLU series in the first fully connected layer, which was used in the following analyses.

Table 2. The nine combinations of the two hyper-parameters of the LSTM models. The F1 score, average lead time, and model score of these models are also given.

Model No.	Number of LSTM layers	ReLU series	Average lead time (s)	F1 Score (w/o tolerance)	Model score
1	4	8	6.512	0.757	0.623
2	4	16	6.539	0.730	0.270
3	4	32	6.537	0.764	0.837
4	5	8	6.520	0.783	1.091
5	5	16	6.572	0.788	1.340
6	5	32	6.773	0.728	1.000
7	6	8	6.670	0.785	1.613
8	6	16	6.468	0.778	0.834
9	6	32	6.673	0.734	0.769

3.3 Typical Predicted PGA Time Histories

In this section, the raw output of the predicted PGA time history using the proposed LSTM approach is discussed. The predicted PGA time histories of a TP case using different input windows are shown in Figure 7. When $t_w = 0.5$ s, the observed PGA time history and the predicted PGA time history of the LSTM model are shown as the thin dashed–dotted gray line and the thick color line, respectively, in Figure 7(a). The target PGA is shown as the step function in a brown dashed–dotted line style, where the trigger time is indicated by the location of a "jump" in the step function. Because the predicted PGA exceeds 25 Gal at approximately 4.04 s (shown as the dotted salmon-colored vertical line), the alert is issued in this window, i.e., at 4.5 s, as indicated by the dashed–dotted, teal-colored vertical line (at the very end of this figure). When $t_w = 1.0$ s (Figure 7(b)), since the predicted PGA already exceeded the threshold in the previous window when $t_w = 0.5$ s, the alert has already been issued and the alert time remains at 4.5 s. The measured accelerations exceed the threshold approximately at 4.88 s (shown as the dashed purple line); hence, the lead time of this TP case is 4.88 - 4.5 = 0.38 s. Note that an alert is classified as a TP only when it is issued before the threshold is reached, i.e., the lead time is positive. If a correct alert is issued but the lead time is zero or negative, the alert is classified as a false negative (FN). The observed PGA time history and the predicted PGA time history of the LSTM model when $t_w = 1.5$, 2.0, 2.5, and 3.0 s are shown in Figures 7(c) to 7(f). Figure 7(a) to 7(f) can be combined together to produce Figure 8(a) for conciseness of this paper.

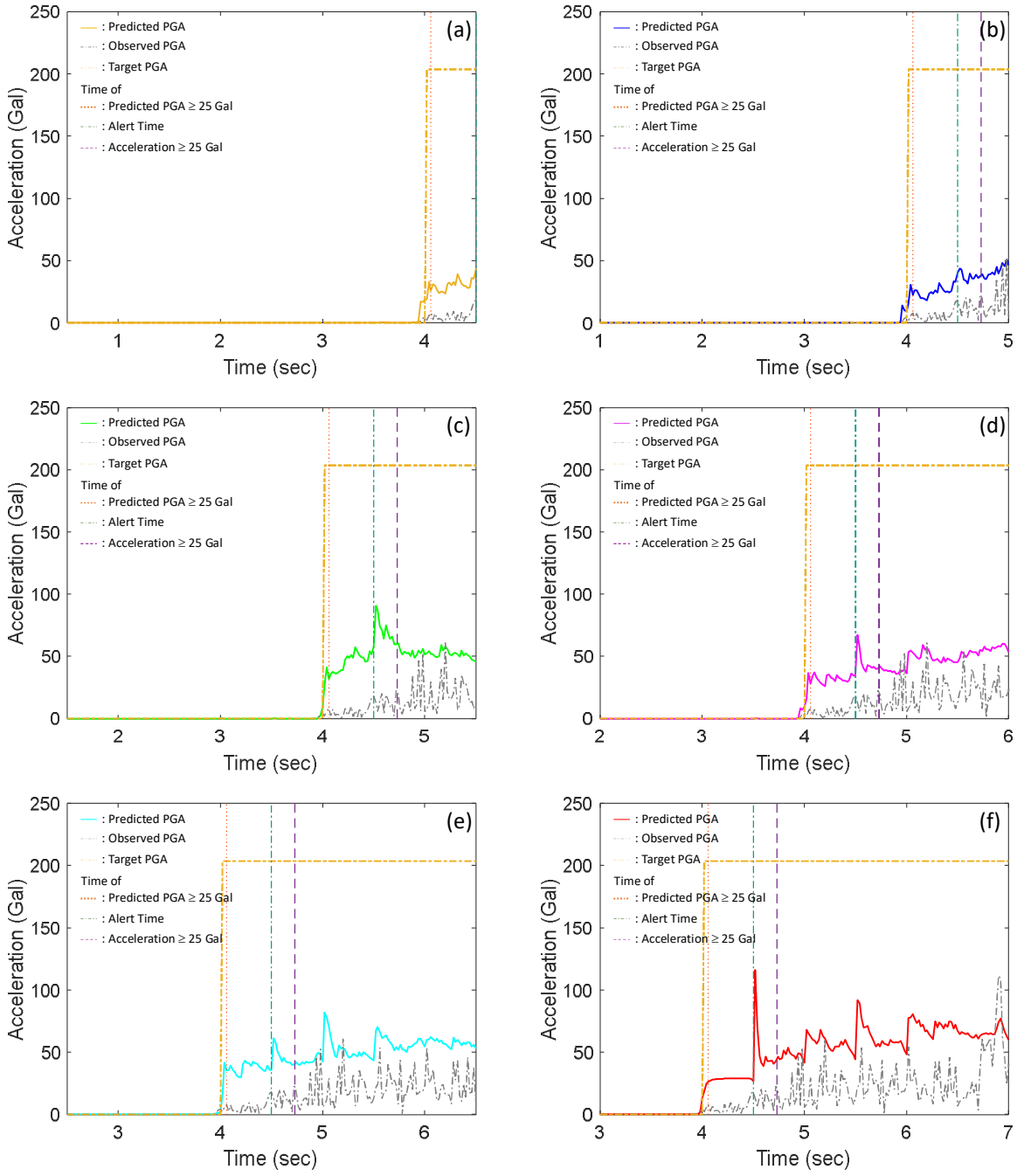

Figure 7. The observed PGA time histories, predicted PGA time histories, and target PGA time histories of a typical true positive (TP) case using (a) t_w = 0.5 s; (b) t_w = 1.0 s; (c) t_w = 1.5 s; (d) t_w = 2.0 s; (e) t_w = 2.5 s; and (f) t_w = 3.0 s.

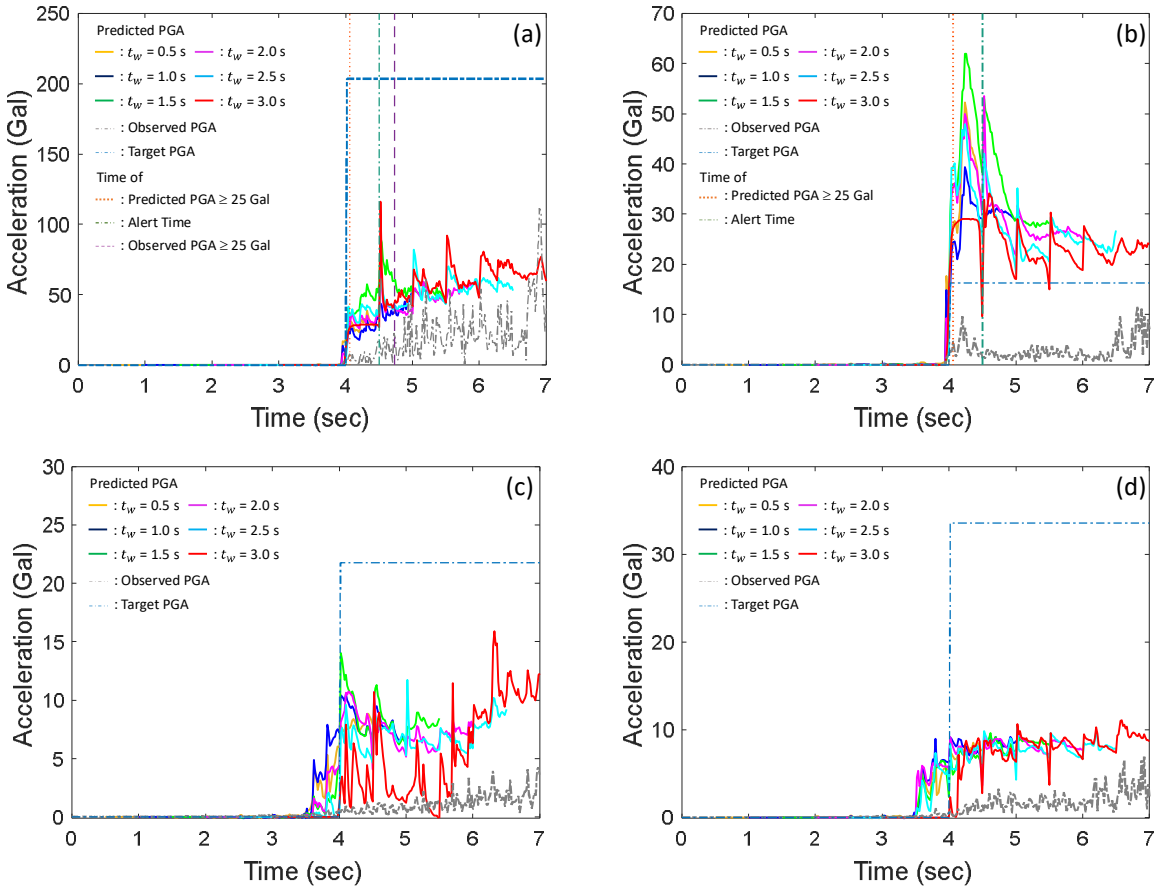

Figure 8. Observed PGA time histories, predicted PGA time histories of different windows, and target PGA time histories of a typical (a) TP, (b) FP, (c) TN, and (d) FN case.

The results of a typical FP case using different input windows are shown in Figure 8(b). Because the predicted PGA exceeds 25 Gal at approximately 4.02 s (shown as the dotted salmon-colored vertical line), the alert is issued in this window, i.e., at 4.5 s (shown as the dashed–dotted, teal-colored vertical line). However, the actual PGA is only about 16 Gal, as indicated by the dashed–dotted cobalt blue step function of the target PGA; hence, this is an FP case if no tolerance of the error of the predicted PGA is allowed. The results of a typical true negative (TN) case using different input windows are shown in Figure 8(c). Although the time histories of the predicted PGA at each window are quite different, the maximum predicted PGA of all windows are approximately below 15 Gal, hence no alert will be issued in this case. The actual PGA is only about 21.7 Gal as indicated by the step function of the target PGA; hence, no alert is necessary in this case. The results of a typical FN case using different input windows are shown in Figure 8(d). The actual PGA is approximately 33.6 Gal as indicated by the step function of the target PGA. However, the maximum predicted PGA of all the six windows is only about 11.1 Gal; hence, no alert is issued in this case.

3.4 General Performance of the TSMIP Data Set

The accuracy of the predicted PGA of the entire TSMIP data set of 91,119 earthquakes is discussed here to understand the general performance of the LSTM model in PGA prediction. The predicted PGA is a time history of each window, hence the data set for predicting the PGA of 91,119 earthquakes using six different windows is quite substantial. Because an alert will be issued when any predicted PGA within any window exceeds the threshold, the maximum predicted PGA within all the windows is a critical value affecting performance in issuing alerts and will be discussed here. Figure 9 illustrates the comparison between the maximum predicted PGA and the actual PGA of all 91,119 earthquakes. The RMSLE value of the maximum predicted PGA is 0.804. It is evident that the maximum predicted PGA tends to be larger than the actual one, which is reasonable because the maximum value of the predicted PGA within all the windows is considered.

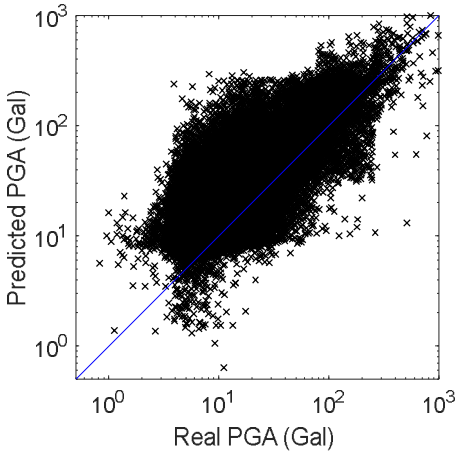

Figure 9. (a) Lead time distribution of the TSMIP data set; and (b) close view of the distribution of lead times shorter than 2 s.

The performance of the alerts issued based on the PGA prediction results of the LSTM model is discussed next. The confusion matrix of FP, FN, TP, and TN when no tolerance of the predicted PGA is allowed and the PGA threshold for issuing an alert is designated as 25 Gal (Figure 5(a)). The performance metrics based on the confusion matrix, such as the number of TPs, FPs, FNs, TNs, and the values of the F1 score, precision, and recall are shown in Table 3. Although recall is quite high (95.17%), precision and F1 score are quite low, only 32.4% and 48.3%, respectively. This is because the predicted PGA tends to be overestimated, hence the number of FPs is quite large, i.e., 32,720, compared to the number of TPs, i.e., 15,684. In other words, the false alert rate (FAR), at 67.6%, is quite high and the missed alert rate (MAR), at 4.90%, is quite low.

As proposed by Meier (2017), a tolerance range can be used when evaluating the classification performance of an EEW system. A ± 1 level of the CWB intensity scale based on PGA, as shown in Figure 5(b), is employed here, and the performance metrics of tolerance are also summarized in Table 3. While tolerance is allowed, only 17 and 7,119 cases of FNs and

FPs, respectively, remain and precision, recall, and F1-score increase to 85.3%, 100.0%, and 92.1%, respectively. Hence, the overall potential alert performance using the proposed LSTM model appears to be quite promising in general if tolerance is allowed. The false alert rate (FAR), at only 14.7%, is quite acceptable, and the missed alert rate (MAR), at 0.03%, is almost zero.

Table 3. The performance metrics of the TSMIP, Chi-Chi earthquake, Meinong earthquake, and Hualien earthquake data sets using the proposed LSTM model.

	TSMIP		Chi-Chi		Meinong		Hualien	
Metric	w/o Tolerance	w/ Tolerance	w/o Tolerance	w/ Tolerance	w/o Tolerance	w/ Tolerance	w/o Tolerance	w/ Tolerance
RMSLE	0.804		0.941		1.141		0.882	
TP	15,684	41,295	310	317	186	268	55	133
FP	32,720	7,119	7	-	106	24	109	31
FN	801	17	59	22	71	24	25	8
TN	41,924	42,708	1	38	160	207	275	292
Precision (%)	32.4	85.3	97.8	100	63.7	91.7	32.7	80.9
Recall (%)	95.1	100	84.0	93.5	72.4	91.1	66.3	92.9
F1 score (%)	48.3	92.1	90.4	96.6	67.8	91.4	43.8	86.5
FAR (%)	67.6	14.7	2.21	0.00	36.3	8.30	67.3	19.1
MAR (%)	4.90	0.03	16.0	6.49	27.6	8.90	33.7	7.10
Lead time (s)	7.46		13.66		6.04		4.29	

The distribution of lead times of the cases when both the actual and predicted PGA is ≥ 25 Gal is shown in Figure 10. The lead times of these cases are between -1 s and 26 s, and the average lead time is 7.46 s. The lead times of only five cases are negative, and these cases are classified as FN, as shown in Figure 10(b).

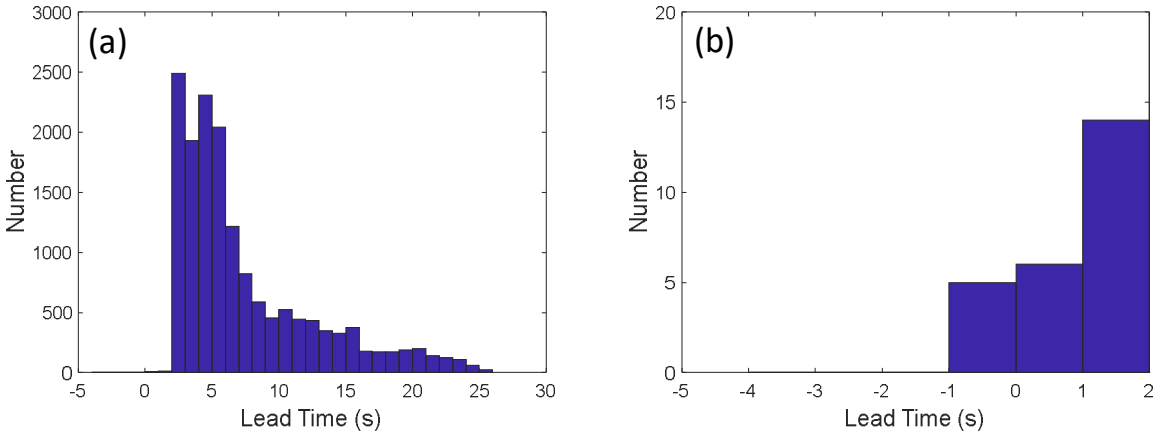

Figure 10. (a) Lead time distribution of the TSMIP data set; and (b) close view of the distribution of lead times shorter than 2 s.

3.5 Performance of the Chi-Chi Earthquake Data Set

The Chi-Chi earthquake data set is also included in the TSMIP data set of 91,119 earthquakes. The fault rupture of the Chi-Chi earthquake has been reported as long and complex (Ma et al. 2001). In addition, several studies have indicated that the magnitude or intensity estimation based on the short beginning of vibration signals (usually several seconds) is not reliable for such large earthquakes, which are usually generated by a long and complex rupture (e.g., Wu & Zhao, 2006; Rydelek et al., 2007; Hoshiba et al., 2011; Kodera, 2019). As in the case of the Chi-Chi earthquake, underestimation of PGA is usually observed when the prediction model is established using SVR (Hsu et al. 2013), a multilayer feedforward network (Hsu et al. 2020), or a convolutional neural network (Hsu et al. 2021). Hence, we also assessed the PGA prediction results of the Chi-Chi earthquake data set using the proposed LSTM approach in this study, as shown in Figure 11(a) and Table 3. The maximum PGA of the same windows, i.e., 0.5 s, 1 s, 1.5 s, 2 s, 2.5 s, and 3 s, after the trigger, predicted using the SVR approach, is shown in Figure 11(b) for comparison (Hsu et al. 2013). It is evident that the PGA predicted using the SVR approach is substantially underestimated, while that predicted using the LSTM approach is much closer to the actual PGA, although the predicted PGA of some earthquakes is still underestimated. The number of TPs, FPs, FNs, TNs, and the F1 score, precision, recall, FAR, and MAR using the SVR approach are shown in Table 4. Tables 3 and 4 show that while the FARs of both LSTM and SVR approaches are very low, the MARs of the latter approach are much higher than those of the former. In addition, the average lead time of the proposed LSTM approach is much longer than that of the SVR approach. Based on these results, the proposed LSTM approach may have an advantage over the other approach in more accurately predicting the PGA of earthquakes with a long and complex rupture process, but further study is necessary to support this finding.

Table 4. The performance metrics of the Chi-Chi earthquake data set using the SVR approach.

Metric	Chi-Chi	
	w/o Tolerance	w/
RMSLE	2.005	
TP	80	80
FP	1	1
FN	267	109
TN	29	187
Precision (%)	98.7	98.7
Recall (%)	23.0	42.3
F1 score (%)	37.3	59.2
FAR (%)	3.33	0.53

MAR (%)	76.9	57.7
Average lead time (s)	5.70	

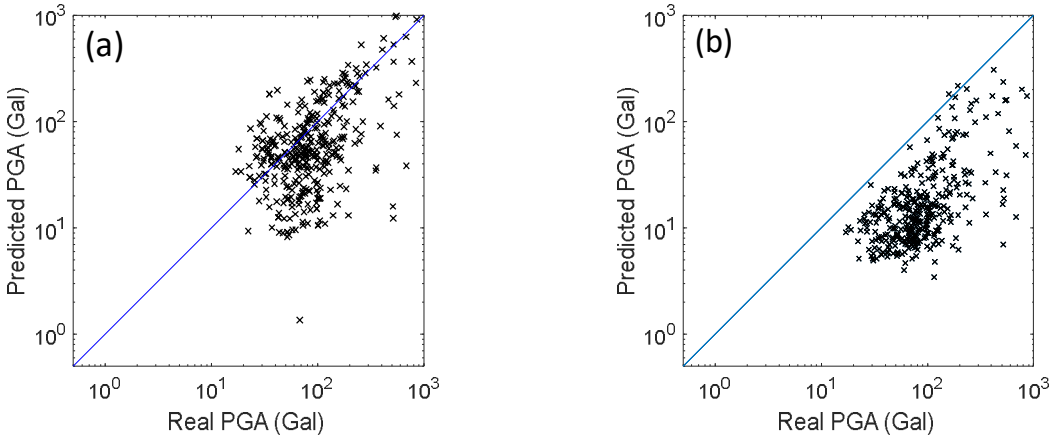

Figure 11. Comparison between the predicted PGA and actual PGA of the Chi-Chi data set using (a) the proposed LSTM approach, and (b) the SVR approach.

3.6 Performance of the Independent Data Sets

In addition to the TSMIP data set, the performance of the proposed LSTM approach was also studied using the independent data sets of the 2016 Meinong earthquakes and the 2018 Hualien earthquakes. The comparison between the predicted and actual PGA of these two data sets is shown in Figure 12. The distribution indicates that the independent data sets could also demonstrate the anticipated accuracy of PGA prediction. The RMSLE of the Meinong and Hualien data sets was 1.410 and 0.882, respectively. The performance metrics, based on the confusion matrices in Figure 5, are shown in Table 3. If no tolerance is allowed, the number of FPs is larger than the number of FNs, resulting in higher recall than precision of both data sets. If tolerance is allowed, precision, recall, and F1 score of these two data sets are quite high, between 80.9%–92.9%. Hence, the performance of alert classification seems quite promising in general if tolerance is allowed.

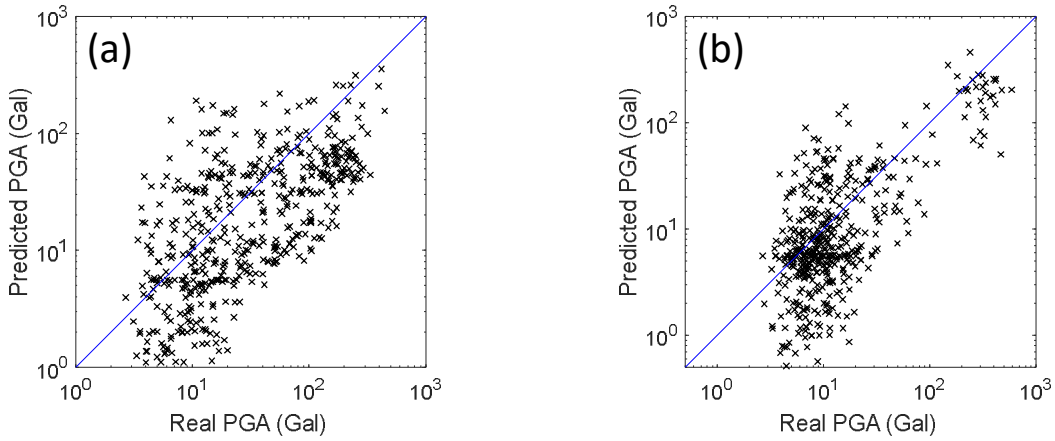

Figure 12. Comparison between the predicted PGA and actual PGA of the (a) Meinong data set, and (b) Hualien data set.

The lead times of these two data sets were between -2 s and 17 s, and the average lead times were 6.04 s and 4.29 s, as shown in Figure 13 and Table 3. Although the predicted PGA was larger than the threshold for all the cases with a negative lead time, these cases were classified as FN due to late alerts, whose locations are shown as inverted triangles in Figure 14. Note that it is not necessary for the stations with a negative lead time to be close to the epicenter. In fact, the lead time of only one of these cases was < -1; the observed PGA and the predicted PGA using different input windows of this case are shown in Figure 15(a). The amplitude of observed PGA time history was quite small and increased very slowly with time, which may make accurate prediction of PGA in time quite difficult. In all other cases with negative lead times, their lead times were actually between -0.5 and 0 s. A typical case with negative lead times is shown in Figure 15(b). Because the predicted PGA exceeded 25 Gal in the same time window with the actual PGA exceeding 25 Gal, the negative lead-time cases could become positive if smaller intervals are used (for example, when the time interval is changed to 0.1 s).

The locations of the classification results with tolerance for the Meinong earthquakes are shown in Figure 14(a). The background of PGA distribution may result from the northwestward rupture of the left–lateral strike-slip fault. Most of the alerts of the stations close to the epicenter are TP. In addition, most of the stations with FN alerts are located in the region approximately 80 km to 130 km north of the epicenter. This phenomenon could mainly be related to the seismic wave difference due to the rupture directivity of the fault. In addition, most of the FP alerts are located in east Taiwan. Besides the rupture directivity of the fault, another possible reason for this phenomenon could be the effect of the path from west to east through the mountainous area in central Taiwan. In the case of the Hualien earthquakes, the background of PGA distribution in Figure 14(b) indicates that the rupture was oriented toward the south, which was identical to the focal mechanism solutions. The majority of the alerts of stations near the epicenter were TP, except for those of several stations very close to the epicenter. Most of the FP stations were located north of the epicenter. This phenomenon may also be related to the seismic wave difference due to the rupture directivity of the fault.

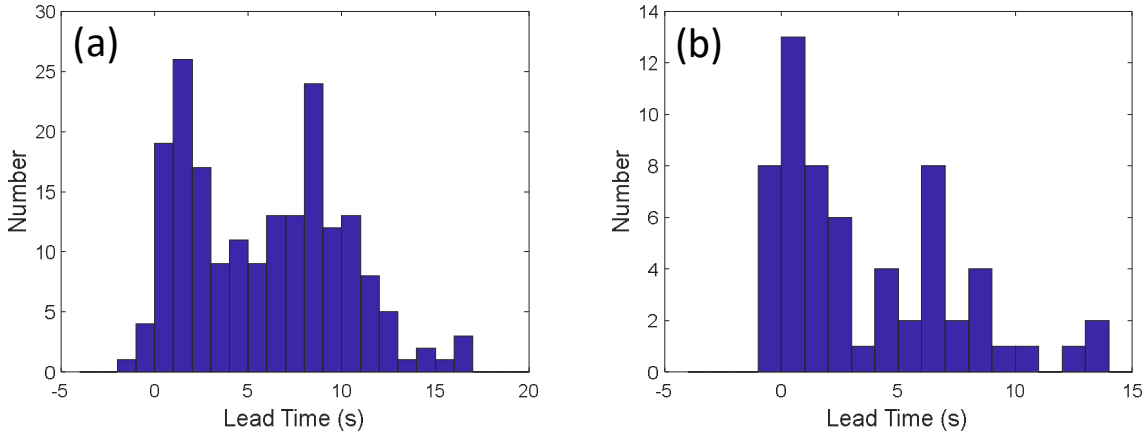

Figure 13. Lead time distribution of the (a) Meinong data set, and (b) Hualien data set.

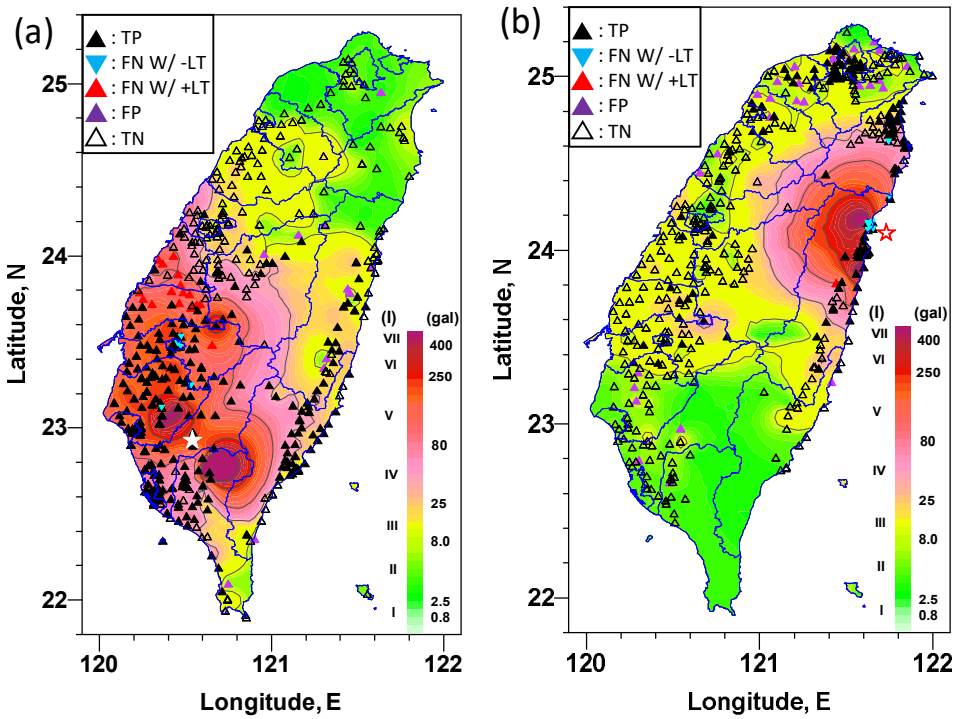

Figure 14. The locations of the stations of different classification conditions of the (a) Meinong data set, and (b) Hualien data set. The background indicates the distribution of observed PGA.

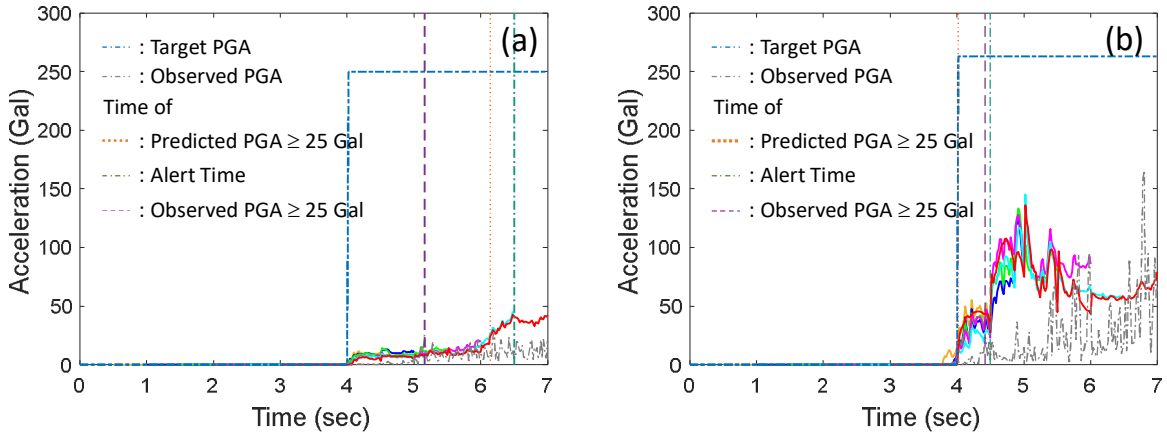

Figure 15. Observed PGA time histories, predicted PGA time histories of different windows, and target PGA time histories of (a) the only case of lead time < -1 s in the Meinong data set, and (b) a typical case of lead time < 0 s in the independent data sets.

3.7 Discussion of the Alert Criterion

From the general performance of the LSTM approach to the analysis of the TSMIP data set, it is evident that the predicted PGA tends to be overestimated, resulting in a higher FAR. The alert is issued once the predicted PGA is equal to or larger than the threshold in any of the six windows. In this section, an alternative algorithm for issuing an alert is proposed to reduce the FAR. This algorithm has two conditions to fulfill to issue an alert. The first is satisfied when the predicted PGA is equal to or larger than the threshold in two consecutive windows. The second is satisfied when the predicted PGA is equal to or larger than the threshold in the last window (t_w = 3.0 s). As shown in Table 5, the number of FP cases decreased from 32,720 to 27,024, while the FAR decreased from 14.7% to 11.5% when tolerance was allowed. However, the number of FN cases increased from 17 to 20, while the MAR increased from 0.03% to 0.04% when tolerance was allowed, and the average lead time also decreased from 7.46 s to 6.91 s. A trade-off was always present between FP cases and the FN cases, and also between prediction accuracy and lead time. If more missed alerts are not tolerable, then the original criterion to issue an alert should be employed.

Table 5. The performance metrics of the TSMIP data set using the proposed LSTM model with an alternative alert criterion.

Metric	TSMIP	
	w/o Tolerance	w/
TP	15,491	37,644
FP	27,024	4,871
FN	994	15

TN	47,620	48,599
Precision (%)	36.4	88.5
Recall (%)	94.0	100
F1 score (%)	52.5	93.9
FAR (%)	63.6	11.5
MAR (%)	6.03	0.04
Lead time (s)	6.91	

3.8 Performance of Different PGA Thresholds

As discussed in the Introduction, a threshold can be designated for different applications. In this study, two alternative thresholds were considered. The first was a relatively small threshold, which is required to issue a command to automatically park machines in semiconductor fabrication plants during the manufacturing process even when the intensity of the coming seismic vibration is small. By doing so, these machines can restart the manufacturing process after the earthquake much faster than when they are shut down by the emergency stop due to the measured vibration signals. The reduced time required to restart the manufacturing process is very beneficial to the semiconductor fabrication plants. Hence, the threshold of 8 Gal was considered in this study, and the general performance of the TSMIP data set was assessed.

The performance metrics when the PGA threshold for issuing an alert is designated as 8 Gal are shown in Table 6. If no tolerance is allowed, the number of FPs is quite large, i.e., 31,047, compared to the number of TPs, i.e., 59,778, resulting in the FAR being quite high at 34.2%. Note that the FAR is much lower than that (it is 67.6%) when the threshold is 25 Gal (Table 3). On the other hand, the number of FNs is only 41, resulting in the MAR being only 0.07%. The average lead time is 6.60 s. The number of FPs becomes only 112, which is very small compared to the number of TPs, i.e., 90,713, when a ±1 level of the CWB intensity scale is employed, as shown in Figure 16(a). In addition, only three FNs remain. Hence the FAR and MAR are very small, i.e., 0.12% and 0.01%, respectively. As a result, the performance of alert classification using 8 Gal as the threshold seems quite promising in general if tolerance is allowed.

The other threshold is relatively large, which could be useful for nuclear power plants because their design basis earthquake is much larger than that of ordinary structures. Hence, the threshold of 250 Gal is studied here, and the general performance of the TSMIP data set is assessed. The performance metrics when the PGA threshold for issuing an alert is designated as 250 Gal are also shown in Table 6. If no tolerance is allowed, the number of FPs is only 560, but it is relatively large compared to the number of TPs, i.e., 211, resulting in the FAR, at 72.6%, being quite high. On the other hand, the number of FNs is only 65, which is relatively small compared to the number of TPs, resulting in the MAR being only 23.2%. The average lead time is 6.32 s. When a ±1 level of the CWB intensity scale is employed here, as shown in Figure 16(b), the number of FPs is only 74, which is very small compared to the number of TPs, i.e., 697. In addition, the number of FNs becomes 0. Hence, the FAR and MAR are very small, i.e., 9.60% and 0.00%, respectively, lower than 14.7% and 0.03% when the threshold is 25 Gal

(Table 36). As a result, the performance of alert classification using 250 Gal as the threshold also seems quite acceptable in general if tolerance is allowed.

Table 6. The performance metrics of the TSMIP data set using the proposed LSTM model with different PGA thresholds.

Metric	Threshold			
	8 Gal		250 Gal	
	w/o Tolerance	w/ Tolerance	w/o Tolerance	w/ Tolerance
TP	59,778	90,713	211	697
FP	31,047	112	560	74
FN	41	3	65	-
TN	263	301	90,293	90,358
Precision (%)	65.8	99.9	27.4	90.4
Recall (%)	99.9	100	76.8	100
F1 score (%)	79.4	99.9	40.3	95.0
FAR (%)	34.2	0.12	72.6	9.60
MAR (%)	0.07	0.01	23.2	0.00
Lead time (s)	6.60		6.32	

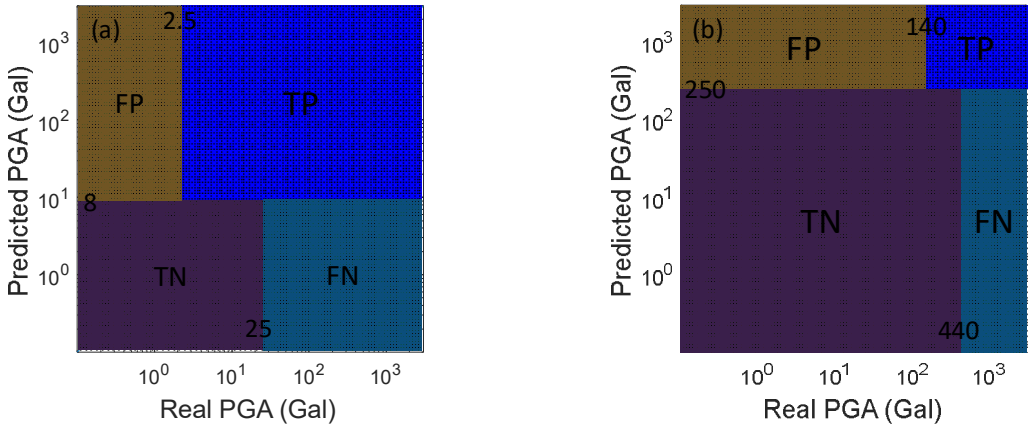

Figure 16. The confusion matrix with tolerance when the threshold is designated as (a) 8 Gal, and (b) 250 Gal.

4 Conclusion

In this study, a dense architecture was proposed for the LSTM neural network to predict the PGA of the incoming seismic wave at the same site. The input data of the LSTM model were the acceleration and velocity time histories of different window lengths of P-wave data after the trigger. The general performance of the proposed LSTM model using the TSMIP data set with 91,128 earthquakes was quite acceptable. However, the predicted PGA of the proposed LSTM model tended to be overestimated, hence a higher FAR (14.7%) and a lower MAR (0.03%) were obtained when a ±1 level of the CWB intensity scale was tolerated, and the average lead time was 7.46 s for the whole TSMIP data set.

Surprisingly, the accuracy of the predicted PGA of the Chi-Chi earthquake using the proposed LSTM model was quite acceptable, resulting in a zero FAR (0.00%) and a low MAR (6.49%) when a ±1 level of the CWB intensity scale was tolerated. Note that the predicted PGA of the Chi-Chi earthquake in previous studies was always seriously underestimated. For comparison, if the maximum predicted PGA for the same windows after the trigger is considered using the SVR approach, an FAR of 0.53% and a high MAR value of 57.7% with a shorter average lead time of 5.70 s are obtained. This illustrates one of the merits of the proposed LSTM approach.

Two independent data sets for damaging earthquakes that occurred recently in Taiwan were employed to validate the proposed LSTM approach. The results indicate that the anticipated accuracy of PGA prediction could also be demonstrated with the independent data sets. When tolerance was allowed, the performance of alert classification appeared quite promising in general because precision, recall, and F1-score were quite high, i.e., between 80.9%–92.9%. The lead times of these two data sets were between -2 s and 17 s, and the average lead times were 6.04 s and 4.29 s. Lead time can be improved if smaller intervals between time windows are employed, especially for the cases with negative lead times, but further studies are required to support this hypothesis.

Because the proposed LSTM approach tended to predict overestimated PGA in general, an alternative algorithm to issue alerts is proposed to reduce the FAR. In this alternative algorithm, an alert will be issued only if the PGA threshold is satisfied in two consecutive windows (except the last window). The performance of the alternative algorithm was tested with the TSMIP data set. When tolerance was allowed, although the FAR decreased from 14.7% to 11.5%, the MAR increased from 0.03% to 0.04% and the average lead time decreased from 7.46 s to 6.91 s.

The threshold can be designated for different applications. In this study, two alternative thresholds were studied. The performance of the different thresholds was tested with the TSMIP data set. If the PGA threshold for issuing an alert was designated as 8 Gal and tolerance was allowed, the FAR and MAR were very small, i.e., 0.12% and 0.01%, respectively, and the average lead time became somewhat shorter, i.e., 6.60 s. If the PGA threshold for issuing an alert was designated as 250 Gal and tolerance was allowed, the FAR and MAR were also very small, i.e., 9.60% and 0.00%, respectively, and the average lead time also became a little shorter, i.e. 6.32 s. Note that these FAR and MAR values were smaller than those, i.e., 14.7% and 0.03%, when the threshold was 25 Gal. However, more studies are required because the number of earthquakes with a PGA of < 2.5 Gal or >250 Gal is not large enough. Nevertheless, the results

still indicate possible application of the proposed LSTM approach to different designated thresholds.

Acknowledgments, Samples, and Data

The authors want to thank the Taiwan Strong Motion Instrument Program operated by Central Weather Bureau (CWB) in Taiwan for providing high-quality strong ground motions caused by earthquakes around Taiwan. The strong-motion accelerations of the Taiwan strong motion instrumentation program (TSMIP) can be downloaded from the geophysical database management system (GDMS) website (http://gdms.cwb.gov.tw/index.php, last accessed September 2021).

References

Allen, R.M. and Melgar, D. (2019), Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, 47, 361–388.

Böse, M., Heaton, T. and Hauksson, E. (2012), Rapid Estimation of Earthquake Source and Ground-Motion Parameters for Earthquake Early Warning Using Data from a Single Three-Component Broadband or Strong-Motion Sensor, *Bulletin of the Seismological Society of America*, 102(2), 738-750.

Cremen, G., and Galasso, C. (2020). Earthquake Early Warning: Recent Advances and Perspectives. Earth-Science Rev. 205 (February), 103184.

Cuéllar, A., Espinosa-Aranda, J.M., Suarez, R., Ibarrola, G., Uribe, A. (2014), The Mexican seismic alert system (SASMEX): its alert signals, broadcast results and performance during the M 7.4 Punta Maldonado earthquake of March 20th, 2012. In Early Warning for Geological Disasters, ed. F Wenzel, Z Zschau, pp. 71–87. Berlin: Springer-Verlag.

Fujinawa, Y. & Noda, Y. (2013), Japan's earthquake early warning system on 11 March 2011: Performance, shortcomings, and changes, *Earthquake Spectra*, 29(S1), S341–S368.

Hochreiter, S., Schmidhuber, J. (1997), Long short-term memory, *Neural Computation*, **9**(8): 1735–1780.

Hoshiba, M., & Iwakiri, K. (2011), Initial 30 seconds of the 2011 off the Pacific coast of Tohoku earthquake (Mw 9.0)—Amplitude and τc for magnitude estimation for earthquake early warning, *Earth, Planets and Space*, **63**, 553–557.

Hsu, T.Y., Huang, S.K., Chang, Y.W., Kuo, C.H., Lin, C.M., Chang, T.M., Wen, K.L., and Loh, C.H. (2013), Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan, *Soil Dynamics and Earthquake Engineering,* 49, 210–217.

Hsu, T.Y., Wang, H.H., Lin, P.Y., Lin, C.M., Kuo, C.H. and Wen, K.L. (2016), Performance of the NCREE's on-site warning system during the 5 February 2016 Mw 6.53 Meinong earthquake, *Geophysical Research Letters,* **43**, 8954–8959.

Hsu, T.Y., Lin, P.Y., Wang, H.H., Chiang, H.W., Chang, Y.W., and Kuo, C.H. (2018), Comparing the performance of the NEEWS earthquake early warning system against the CWB system during the 6 February 2018 Mw 6.2 Hualien earthquake, *Geophysical Research Letters,* **45**, 6001–6007.

Hsu, T.Y., Wu, R.T., Liang, C.W., Kuo, C.H. and Lin, C.M. (2020), Peak ground acceleration estimation using P-wave parameters and horizontal-to-vertical spectral ratios, *Terrestrial Atmospheric and Oceanic Sciences*, **31**(1), 1-8.

Hsu, T.Y., Kuo, C.H., Wang, H.H., Chang, Y.W., Lin, P.Y., Wen, K.L. (2021), The realization of an earthquake early warning system for schools and its performance during the 2019 ML 6.3 Hualien (Taiwan) earthquake, *Seismological Research Letters*, **92**(1), 342-351.

Hsu, T.Y., & Huang, C.W. (2021), Onsite Early Prediction of PGA Using CNN With Multi-Scale and Multi-Domain P-Waves as Input, *Front. Earth Sci.,* 9:626908.

Kodera, Y., Saitou, J., Hayashimoto, N., Adachi, S., Morimoto, M., Nishimae, Y., and Hoshiba, M. (2016), Earthquake early warning for the 2016 Kumamoto earthquake: performance evaluation of the current system and the next-generation methods of the Japan Meteorological Agency, *Earth, Planets and Space*, **68**, 202.

Kodera, Y., Yamada, Y., Hirano, K., Tamaribuchi, K., Adachi, S., Hayashimoto, N., Morimoto, M., Nakamura, M., and Hoshiba, M. (2018), The propagation of local undamped motion (PLUM) method: A simple and robust seismic wavefield estimation approach for earthquake early warning, *Bull. Seismol. Soc. Am.* **108**, 983–1003.

Kodera, Y. (2019). An earthquake early warning method based on Huygens principle: Robust ground motion prediction using various localized distance-attenuation models, *J. Geophys. Res.* **124**, 12981–12996.

Murtaza, R., Patel, H., Varma, S. (2017), Predicting Stock Prices Using LSTM. Int J Sci Res. **6**(4):1754–6.

Meier, M.A. (2017), How "good" are real-time ground motion predictions from Earthquake Early Warning systems?, *J. Geophys. Res. Solid Earth*, **122**, 5561–5577.

Ma, K.F., Mori, J., Lee, S.J., Yu, S.B. (2001), Spatial and temporal distribution of slip for the 1999 Chi-Chi, Taiwan, Earthquake. *Bull Seism Soc Am* **91**(5): 1069–1087.

Rydelek, P., Wu, C., Horiuchi, S. (2007), Comment on "Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records" by Aldo Zollo, Maria Lancieri, and Stefan Nielsen. *Geophysical Research Letters*, 34, L20302.

Sak, H., Senior, A.W., Beaufays, F. (2014), Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Google. http://research.google/pubs/pub43905.pdf

Wald, D.J. (2020), Practical limitations of earthquake early warning. *Earthq Spectra,* **36**(3):1412–1447.

Wang, C.Y., Huang, T.C., Wu, Y.M. (2020), A LSTM Neural Network for On-site Earthquake Early Warning, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-3696.

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, L16312, doi:10.1029/2006GL026871.

Wu, Y.M., Mittal, H., Huang, T.C., Yang, B.M., Jan, J.C. and Chen, S.K. (2019), Performance of a low-cost earthquake early warning system (P-Alert) and Shake Map production during the 2018 Mw 6.4 Hualien, Taiwan, earthquake. *Seismological Research Letters*, **90**(1), 19–29.

Yamada, M., Tamaribuchi, K., Wu, S. (2014), Faster and More Accurate Earthquake Early Warning System - Combination of Velocity and Acceleration-type seismometers, *Journal of Japan Association for Earthquake Engineering*, **14**(4),.21-34, (in Japanese).

Zhao, Z., Chen, W., Wu, X., Chen, P.C.Y., Liu, J. (2017), LSTM network: A deep learning approach for Short-term traffic forecast, *IET Intelligent Transport Systems*, **11**(2): 68–75.