AIMS Geosciences, 11(3): 577–599.
DOI: 10.3934/geosci.2025025
Received: 26 February 2025
Revised: 21 May 2025
Accepted: 30 May 2025
Published: 09 July 2025

https://www.aimspress.com/journal/geosciences

Research article

Peak ground acceleration prediction and seismic intensity estimation based on ensemble framework for earthquake early warning in Taiwan

Shieh-Kung Huang* and Kuan-Hao Tseng

Department of Civil Engineering, National Chung Hsing University, Taiwan

* **Correspondence:** Email: skhuang@nchu.edu.tw.

Abstract: Among various natural disasters, earthquakes can produce tremendous life loss and extensive damage to public and private property, particularly in densely populated areas. To facilitate immediate protective actions, earthquake early warning (EEW) was developed to provide a lead time of a few seconds to minutes before impending ground motions at specific sites. Conspicuously, enhancing the prediction of peak ground acceleration (PGA) and the estimations of seismic intensities is crucial for EEW systems to be protective. Numerous methods have been employed to predict PGA and estimate intensities, and we aimed to achieve these objectives through the rapid advancements in machine learning (ML) techniques. Instead of using time histories directly, the site parameters and the P-wave features were treated as inputs and trained using supervised learning approaches, including neural networks and decision trees, to generate accurate predictions, thereby creating a more efficient ML model. Apparently, the performance of these models depends on the training inputs, the ML models/algorithms, and the validation processes; thus, they were compared and discussed in this study. Consequently, the importance of the site parameters was demonstrated, including the high-frequency attenuation rate, the shear wave velocity, and the horizontal depth associated with significant shear wave velocities. In addition to the conventional ML models, an ensemble framework combining neural networks and decision trees was proposed to capitalize on the advantageous and disadvantageous characteristics of the individual models. The effectiveness of the proposed ensemble framework in predicting PGA and estimating seismic intensities was evaluated using performance metrics. The enhancement after integrating the site parameters and the proposed ensemble framework was

thoroughly demonstrated via the accuracy, recall, precision, F1 score, and false alarm rate (FAR). This comparison also highlights the reliability of applying the proposed ML model in EEW systems, and the recent advances of ML make them especially suitable for emergency response and decision-making.

Keywords: earthquake early warning; on-site; site parameter; P-wave feature, peak ground acceleration; seismic intensity; ensemble model

1. Introduction

Earthquakes rank among the most devastating natural disasters, frequently resulting in significant destruction, especially in areas with high population density. The aftermath of such events can lead to considerable life loss and extensive damage to public and private property. A notable instance occurred on March 11, 2011, when a magnitude 9.0 earthquake struck 231 miles northeast of Tokyo, Japan, producing nearly 16,000 fatalities in direct and indirect ways [1]. This seismic event also triggered a tsunami with waves reaching 30 feet, which severely impacted several nuclear reactors in the vicinity.

Due to several inherent challenges associated with the nature of seismic events, earthquake prediction remains unreliable nowadays. First, the fundamental characteristics of seismic events frustrate the development of short-term forecasts for strong earthquakes, despite the availability of abundant observational data and advanced equipment [2]. Likewise, the identification of reliable precursors also makes short-term predictions extremely difficult, as the seismic processes are highly nonlinear and the precursors must be distinguishable from background noise generated by other seismic activities. In other words, earthquakes are inherently (or effectively) unpredictable due to their highly sensitive nonlinear dependence on the initial conditions [3]. Consequently, researchers have generally concluded that predictions of seismic events are unsuccessful, although recent advancements in AI-based methods have shown some promise [4].

In contrast, earthquake early warning (EEW) focuses on detecting seismic waves as they travel through the Earth after a seismic event has begun. EEW systems are real-time monitoring installations designed to provide a lead time of a few seconds to minutes for impending ground motions at specific sites, enabling immediate protective actions in the short term, such as stopping trains or shutting down utilities. Unlike predictions, which attempt to forecast when and where an earthquake will occur, an inherently uncertain process, EEW systems react to actual seismic activity, making them more reliable [5]. They benefit from the differences in the propagation speeds of seismic waves; the typical velocity of P-waves is around 6 km/s, while S-waves travel at approximately 3 km/s. Two types of EEW systems are currently in operation worldwide [6,7]. The first type is a regional (or front-detection or network-based) system, where seismometers installed in the earthquake source area provide early warnings and transmit information to distant locations before the seismic waves arrive [8]. The second type is an on-site system, which determines earthquake information from the initial portion of P-waves recorded by on-site seismometers and predicts the seismic intensity of subsequent S-waves for emergency response [9]. Generally, the first one is more reliable, while the second one is fast enough to provide useful warnings to sites, even at very short distances, where early warnings are most critical [10].

Typically, EEW and its systems adhere to the "Goldilocks principle" [11,12]. When the sites are too distant from the epicenters, the warnings may be more accurate and offer a longer lead time; however, the seismic intensity is weaker and presents minimal risk. Despite the effectiveness of regional systems in providing accurate warnings, they usually have a significant late alert zone (also known as a blind zone) surrounding the epicenter due to the very limited time available for preparation [13]. Lead time refers to the interval between the issuance of a warning and the arrival of the S-wave; if the S-wave arrives before the warning is issued, there is no opportunity to prepare for the seismic event. Admittedly, the lead times offered by on-site systems are considerably longer, particularly in areas close to the epicenter, since no joint detection has been required. However, the trade-off comes with the accurate prediction because of the limited information [5,12]. Therefore, enhancing the lead time and prediction of the on-site EEW systems is crucial for improvement.

Predicting ground motions of subsequent S-waves, such as peak ground acceleration (PGA) and seismic intensities, is crucial for EEW systems in emergency response [14]. Accurate information related to PGA and seismic intensities helps the assessment of seismic damage, the effectiveness of lead time and emergency response, and the continuous measures for seismic impacts. In the past decade, numerous studies have demonstrated that these predictions align well with observed seismic intensities [15–19]. Furthermore, recent advancements in machine learning (ML) and artificial intelligence (AI) have introduced innovative approaches for PGA and estimating seismic intensities in the context of nonlinear seismic processes. These advancements ensure that EEW systems remain effective under varying complex conditions. Examples of these techniques include the use of long short-term memory (LSTM) neural networks, convolutional neural networks (CNNs), deep neural networks (DNNs), gradient boosting regressors, and more [20–23]. However, utilizing time histories as inputs hinders the efficiency of the ML models.

We aim to predict PGA and estimate intensities through the rapid advancements in ML techniques. Instead of using time histories directly, we exploit the site parameters and the P-wave features, enabling a more efficient ML model. The remainder of the paper is organized as follows: In the first section, the dataset collected from Taiwan's seismic array is introduced. Next, the preprocessing steps and the selected parameters utilized as training inputs are described. The validation methods and performance metrics are outlined in the following section to facilitate the training process for the proposed ML model. Subsequently, the preliminary results are presented to explore the advantages of using the site parameters, accompanied by several case studies employing conventional ML models. Finally, the results of the proposed method are shown, and relevant conclusions are duly drawn.

2. Description of the dataset

Predicting PGA is essential for EEW systems, as it is closely linked to seismic damage assessment. Enhancement in prediction accuracy through advanced modeling techniques not only provides critical lead time for emergency responses but also improves the overall effectiveness and timeliness of EEW systems. These improvements play a vital role in the preliminary assessment of seismic damage, ultimately contributing to reduced risks associated with earthquakes and bolstering public safety in vulnerable regions. To achieve these objectives, numerous methods have been developed to correlate features extracted from P-waves with subsequent PGA values during seismic events, particularly by

leveraging the rapid advancements in ML techniques. In this study, various models are trained using supervised learning approaches to estimate PGA, and they are compared to discuss the performance. Additionally, a reliable earthquake dataset from Taiwan, along with several key earthquake features, is introduced for predicting seismic intensities. The satisfactory accuracy achieved by different models demonstrates the potential of utilizing EEW systems for effective PGA prediction.

2.1. Earthquake data

Taiwan is situated at the convergence boundary of the Eurasian Plate and the Philippine Sea Plate, making the region particularly renowned for its frequent earthquakes and high seismic activity. Due to the significant seismic hazards present, several seismic arrays have been established across the island. The dataset used for training various models employing supervised learning approaches in this study was collected from the Taiwan Strong Motion Instrumentation Program (TSMIP), which is operated by the Central Weather Bureau (CWB) of Taiwan [24]. This program, initiated in 1992, aims to capture highly qualified recordings of earthquake waveforms during seismic events. Over the past three decades, more than 700 free-field strong motion stations instrumented with reliable seismographs have been continuously operational.

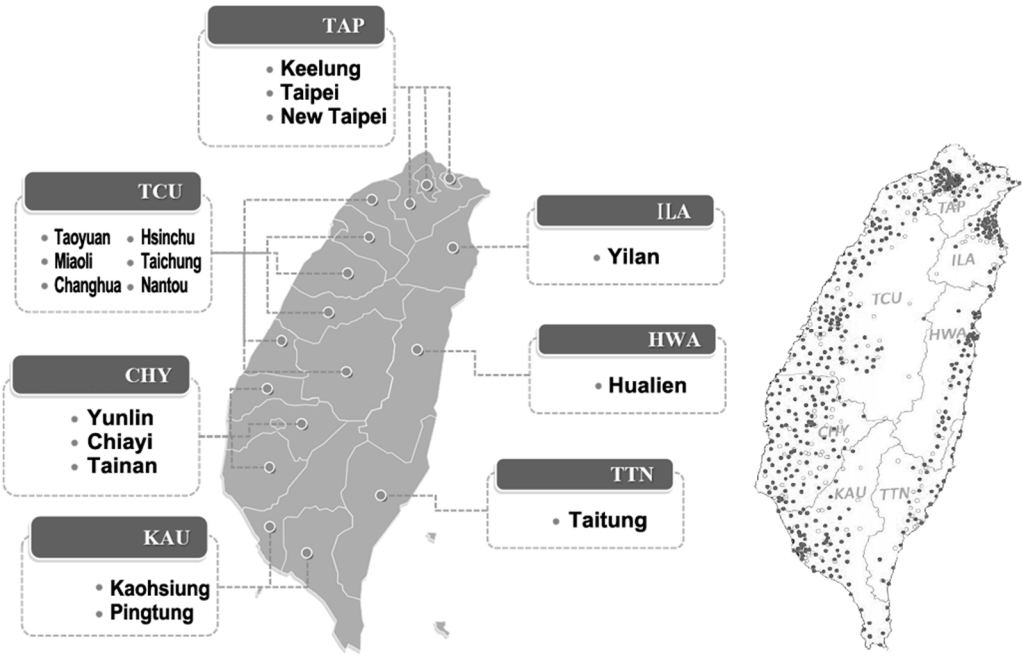


Figure 1. Seven areas and seismograph distribution for the TSMIP seismic array and the earthquake datasets.

The earthquake data utilized in this study is sourced from the TSMIP seismic array, which divides Taiwan into seven regions: KAU, TCU, TAP, TTN, HWA, CHY, and ILA, as illustrated in Figure 1. To ensure qualified data collection, the waveforms captured by the seismic array are sampled at rates of either 200 or 250 samples per second, with a resolution of 16 bits or higher. Furthermore, ±2000 gal (cm/s^2) is a typical dynamic range for the seismographs in the network. The datasets for each region

comprise approximately 3,300 earthquakes that occurred around Taiwan over a 15-year period. The first recording was collected on July 29, 1992, while the last recording in this dataset was collected on December 31, 2006.

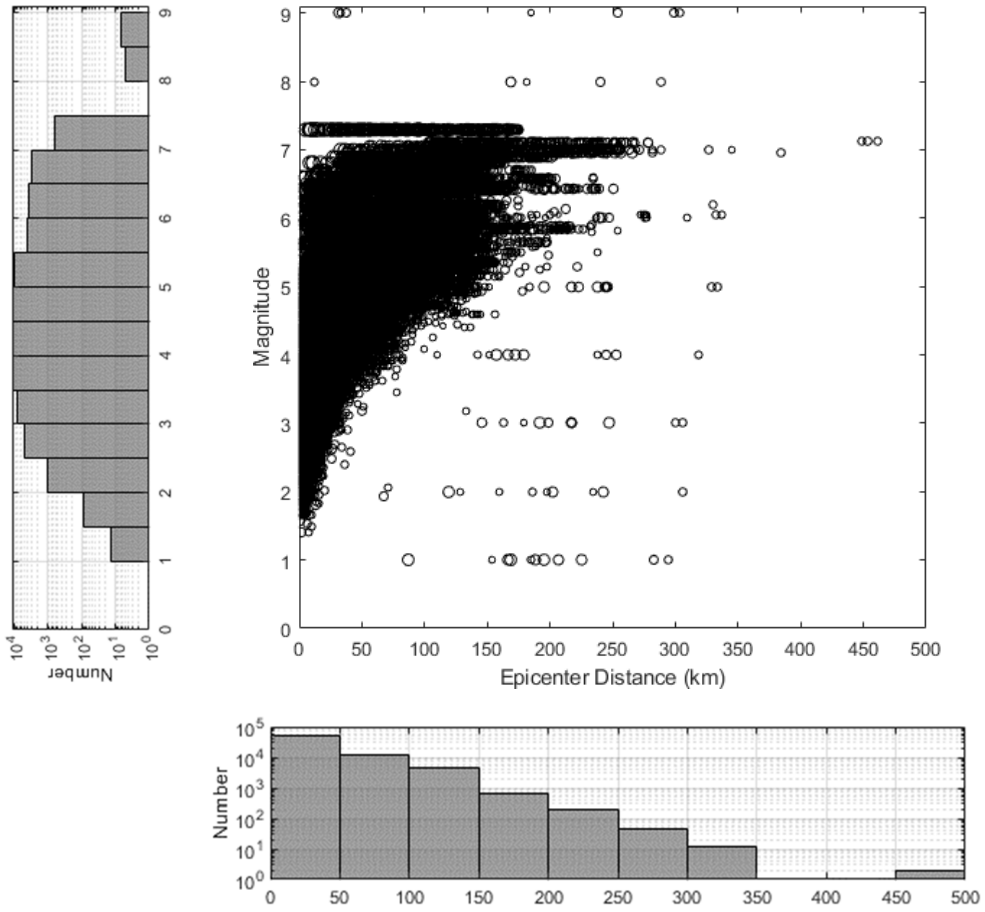


Figure 2. Distribution and histogram of magnitudes and epicenter distances included in the dataset; the size of circles indicates the PGA on a logarithmic scale.

To ensure a comprehensive dataset for analysis, we utilized approximately 15 years of recordings from the TSMIP, spanning July 29, 1992 to December 31, 2006. The original database contained 105,385 recordings; however, those that were clearly distorted or exhibited dropouts or glitches were excluded. Additionally, recordings that were either misidentified or insufficiently documented were removed from the dataset. Recordings shorter than ten seconds in duration were eliminated because we intended to predict seismic intensities through the P-wave. Ultimately, a total of 70,947 recordings were retained for evaluating the performance of the different models in this study. The earthquakes included in this collection feature magnitudes ranging from 1.0 to 9.0 and epicenter distances from 0.085 to 461.72 km, with data lengths varying between 13 and 400 seconds. The distribution and histogram of magnitudes and epicenter distances are shown in Figure 2.

Since the TSMIP employs strong motion accelerographs, the recordings were adjusted to create zero-mean signals and were integrated to have velocity and displacement, respectively. A four-order 0.25 Hz high-pass filter was then applied to eliminate any offsets and trends following the final

integration. Furthermore, the Short-Term Average/Long-Term Average (STA/LTA) algorithm was utilized as an automatic P-wave picker to determine the P-wave arrival time (also referred to as triggering time) from the vertical acceleration recordings [25].

2.2. Training inputs

The concepts, methods, and backgrounds of EEW were reviewed, and the estimations of seismic intensities through P-waves was summarized in the previous study [6]. To predict PGA before a significant shock, the site parameters and the P-wave features can be effectively utilized for EEW systems. Notably, considering that earthquake preparation and generation take very complex processes, multiple parameters/features need to be considered to accurately predict seismic intensities.

The site parameters are essential for accurately representing site effects during a seismic event; key parameters include the high-frequency attenuation rate, the shear wave velocity, and the horizontal depth associated with significant shear wave velocities. The high-frequency decay rate, commonly referred to as k, is a critical site parameter in seismology that quantifies the attenuation of seismic waves at high frequencies. It indicates the rate at which high-frequency seismic energy diminishes as it propagates through various media and can be calculated from the slope of the Fourier spectrum in a linear logarithmic space for specific frequency ranges [26,27]. Another essential site parameter is the shear wave velocity, usually denoted as V_S, which plays a crucial role in site response analysis, site classification, and seismic loss estimation. The average shear wave velocity over the top 30 meters of the subsurface profile, V_{S30}, is commonly exploited to represent the overall site conditions for seismic waves traveling through a given location [28,29]. Additionally, the horizontal depth associated with significant shear wave velocities has been introduced to address the limitations of V_{S30}, especially in areas with substantial sediment thickness. Specific velocities, such as 1.0, 1.5, or 2.5 km/s, have been incorporated into ground motion prediction equations developed during the Next Generation of Attenuation (NGA) projects, and in this study, the depth to the horizons with shear wave velocity exceeding 1.0 km/s, so-called $Z_{1.0}$, was used [30,31]. The site parameters in this study were provided by the Engineering Geological Database for TSMIP (EGDT), which was created by the National Center for Research on Earthquake Engineering (NCREE), Taiwan (https://egdt.ncree.org.tw/News_eng.htm). Those parameters were measured and derived according to the demand of the SSHAC Level 3 project (http://sshac.ncree.org.tw/) in Taiwan.

In addition to the site parameters, the P-wave features were introduced to represent the characteristics of seismic waves. According to the book by Kramer and Stewart, the earthquake features briefly include frequency content, energy, amplitude, duration, and other characteristics [32]. Some features may focus on one specific characteristic, while others may coincide with several characteristics; therefore, a variety of features were investigated in this study to enhance the overall effectiveness and timeliness of EEW systems.

The first P-wave feature selected in this study pertains to the frequency content of an earthquake. The predominant period of seismic waves, τ, serves as a great indicator for characterizing the earthquakes. It can be determined by applying the Fourier transform and picking the peak on the spectrum; however, for practical purposes, the average period, τ_a, proposed by Kanamori, was used as an alternative to the predominant period [8,33]. It can be calculated as

$$\tau_a = 2\pi \sqrt{\frac{\int_0^t x^2(t)\,dt}{\int_0^t \dot{x}^2(t)\,dt}} \tag{1}$$

where x is the ground displacement. It has been proved that the average period, τ_a, is highly correlated to the earthquake magnitude and intensity from the attenuation relationships studied in Taiwan.

The second and third P-wave features pertain to the integrated energy released during a seismic event. For instance, the cumulative absolute velocity (V_a) is proposed by the Electric Power Research Institute (EPRI) as an alternative intensity measure [34]. It can be calculated by integrating the area under the absolute acceleration, and its mathematical expression is as follows:

$$V_a = \int_0^t \left|\ddot{x}(t)\right| dt \tag{2}$$

The cumulative absolute velocity is considered to be more indicative of structural damage compared to other intensity measures, so it can be used for seismic hazard assessments [35]. Another example is the integral of squared velocity (V_I), which is correlated with the early-radiated energy produced by the advancing rupture on the fault plane [36]. It can be calculated by integrating the squared velocity as

$$V_I = \int_0^t \dot{x}^2(t)\,dt \tag{3}$$

The integral of squared velocity gives direct insights into the physics of fractures and has numerous applications for EEW and seismic damage assessment [16,37].

The last three P-wave features pertaining to the peak amplitudes are most commonly used to describe an earthquake. These features include the peak P-wave acceleration (A_P), the peak P-wave velocity (V_P), and the peak P-wave displacement (D_P) as

$$D_P = \max_t \left|x(t)\right| \tag{4}$$

$$V_P = \max_t \left|\dot{x}(t)\right| \tag{5}$$

$$A_P = \max_t \left|\ddot{x}(t)\right| \tag{6}$$

The forces produced by a seismic event are mainly inertial forces, which makes the acceleration collected by a seismograph a crucial parameter. Commonly, velocity and displacement are obtained by integrating the recorded acceleration. Larger acceleration is often correlated with greater damage under an earthquake. In contrast, velocity and displacement are associated with lower-frequency components, which can be more detrimental to long-period structures, such as tall buildings. This distinction underscores the importance of understanding both acceleration and its derived measures in assessing the impacts of earthquakes on various types of infrastructure.

3. Intensity and PGA prediction

Leveraging supervised learning approaches to correlate features extracted from P-waves with subsequent PGA values during seismic events can gradually improve the overall effectiveness and timeliness of EEW systems. The earthquake data was first utilized to train various models, the model's hyperparameters were then tuned through optimizers, and the well-trained models were used to generate accurate predictions. The development of these models relied on the training inputs (the site parameters and the P-wave features), the ML models/algorithms, the optimization techniques, and the validation processes. The performance for predicting PGA and estimating seismic intensities was also crucial in the evaluation of various models. To clarify these key elements, the validation methods and performance metrics are described in this section, and the preliminary results related to the training inputs are demonstrated. Through these discussions, the accuracy, efficacy, generalization, and reliability of EEW systems can be established.

3.1. Validation methods

Cross-validation is an important part when assessing the reliability of supervised learning in ML, and it is typically employed to evaluate how well a model can be generalized to an independent dataset. This process involves partitioning the data into subsets, training the model on some subsets, and validating it on others. The primary methods of cross-validation include K-fold, stratified K-fold, leave-one-out (LOO), leave-p-out (LPO), blocked, and Monte Carlo (repeated random subsampling) cross-validation.

Each of these methods has its own strengths and weaknesses, making them suitable for different types of data and modeling scenarios. For example, K-fold cross-validation is advantageous because it utilizes all data points for both training and validation, leading to a more reliable estimate of model performance. However, the stratified K-fold cross-validation can help in achieving better generalization by preserving class distribution across folds. Admittedly, the choice of cross-validation methods depends on the specific requirements of the analysis and the characteristics of the dataset being used. In this study, the Monte Carlo cross-validation was adopted due to its ease of implementation. Also known as repeated random subsampling cross-validation, this method involves randomly splitting the original dataset into training and validation subsets. Unlike K-fold cross-validation, the number of cases in Monte Carlo cross-validation can be arbitrarily assigned by the user because of its random nature. In this study, the ratio between the training and validation subsets was set at 80:20. The results of 20 times of Monte Carlo cross-validation were performed and averaged to assess the generalization of the well-trained models.

To avoid overfitting, an independent subset was introduced to verify that any increase in accuracy on the training subset actually yielded an increase in accuracy on the other subsets within the training models. In other words, if the accuracy on the training subset increased while the accuracy on the testing subset remained the same or decreased, the models were terminated early to preserve generalization. Although this method presented an inevitable trade-off similar to the issue of curve fitting, where a higher number of degrees of freedom in the approximating function enabled greater flexibility and better fitting to the inputs, consideration must be given to the potential impact of noise

in the data. Therefore, smoother and more generalized models were preferred, even if they exhibited slight inaccuracies. Moreover, in each iteration, 10% of the training subset was used as the testing subset in this study.

3.2. Performance metrics

Performance metrics are another part with great importance in ML. These metrics help quantify how well a model performs on specific tasks, enabling users to refine their models for improved accuracy and reliability. As a category of supervised learning approaches, the performance metrics can be roughly divided into classification metrics and regression metrics, each tailored to different types of prediction tasks. In this study, the predictions of PGA could be effectively evaluated using the regression metrics, such as root mean square error (RMSE), among others.

However, the well-trained models were designed not only to predict PGA but also to estimate seismic intensities, which were usually represented by several integers. This offered a more accurate representation of the effects that earthquakes have on individuals, structures, and the environment, thereby enhancing disaster response efforts by government agencies. Although CWB introduced a modified intensity scale on January 1, 2020, a more complex formula was used in the new scale, which included peak ground velocity starting at intensity level 5. In contrast, the old scale was computed solely based on PGA; therefore, we adopted the old scale, as shown in Table 1. For the estimations of seismic intensities, the classification metrics were used, and common metrics included accuracy, recall, precision, F1 score, false positive rate (FPR), and so on.

The accuracy is the ratio of correct predictions to total predictions, calculated as

$$\text{Acc.} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively. The recall is sensitivity, which indicates how well the model identifies positive instances

$$\text{Rec.} = \frac{\text{TP}}{\text{TP} + \text{TN}} \tag{8}$$

The precision measures the accuracy of positive predictions

$$\text{Pre.} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

The F1 score is the harmonic mean of precision and recall, providing a balance between the two

$$\text{F1} = 2 \times \frac{\text{Pre.} \times \text{Rec.}}{\text{Pre.} + \text{Rec.}} \tag{10}$$

Last but not least, considering a practical scenario, the FPR represents the ratio of the number that is falsely alarmed against the total number of wrong predictions in percentage:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{11}$$

It is also referred to as a false alarm rate (FAR), and a lower FPR value indicates that the system usability level is higher.

Table 1. Old seismic scale and the corresponding acceleration used before January 1, 2020.

Seismic Intensities	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
		I	II	III	IV	V	VI	VII
PGA values (gal)	< 0.8	0.8 > < 2.5	2.5 > < 8.0	8.0 > < 25	25 > < 80	80 > < 250	250 > < 400	> 400
Numbers in the dataset	0	16	19982	35436	12432	2793	200	88

3.3. Preliminary results of the neural network

A total of 20 cross-validations were performed to ensure generalization. For the well-trained model, the overall performance is shown in Figure 3. For predicting PGA, Figure 3(a) and Figure 3(b) present comparative diagrams for the training and validation subsets, respectively. It is important to note that only one cross-validation is displayed here due to the high similarity observed across the multiple times. The predictions and observations demonstrate a strong positive correlation throughout the dataset, indicating that the predictions are proportional to the observations, with a standard deviation of errors less than 30 gal. The regions highlighted by the red lines indicate 7 distinct intensity levels, and the predictions and observations are approximately distributed within these regions. Similar observations have been reported by researchers utilizing support vector machines (SVM) and neural networks [16,38]. Consequently, acceptable predictions are defined as having a one-level difference (±1) on the seismic intensity scale of Taiwan. As a result, the well-trained model achieves an accuracy of 98.9% for the training and validation subsets, as shown in Figure 3(c) and Figure 3(d). Noteworthy, the well-trained models used only the P-wave features generally ended with an accuracy of around 98.6%, indicating the improvement of using the site parameters. However, predictions for seismic intensities below level 1 tend to lean towards the non-conservative side, with the predicted PGA being lower than the observation. Fortunately, this phenomenon is mitigated at higher intensity levels and will not significantly impact early warning for seismic events.

In this study, the site parameters and P-wave features were designated as training inputs; however, the necessity of these inputs must be verified. Considering that other studies have confirmed the significance of P-wave features [16,38], the site parameters were removed one by one to investigate their importance, and we compared the results with the case where all site parameters were included as training inputs. Figure 3(c) and Figure 3(d) present a comparison of the accuracy for estimating intensities. All accuracies reach 99%, making it difficult to determine which site parameter is most important for estimation.

To further investigate the necessity of the site parameters, Figure 4 shows performance metrics beyond accuracy. The models perform worst at intensity level 7 in terms of recall and at intensity level 1 in terms of precision. The F1 score also illustrates this by providing a balance between them. The

results primarily come from the lack of training data and primarily stem from the lack of training data; the data numbers for intensity levels 1 and 7 are only 13 and 71, respectively, whereas the minimum data number for intensity levels 2 to 6 is 160. Upon examining the 4 cases, no significant differences were found, indicating that the importance of the three site parameters is generally equal. The performance of the case utilizing all parameters is typically higher than that of the other cases across intensity levels 2 to 7. Although the case without V_{S30} (represented by the yellow bar) outperforms the other cases across performance metrics at intensity level 1, this superiority is negligible, as this level represents the worst overall performance for a neural network model. In summary, all site parameters are recommended for predicting PGA and estimating intensities via neural networks, and performance is notably strong for intensity levels 2 to 6. From a precision perspective, the estimations of seismic intensities remain valid at intensity level 7.

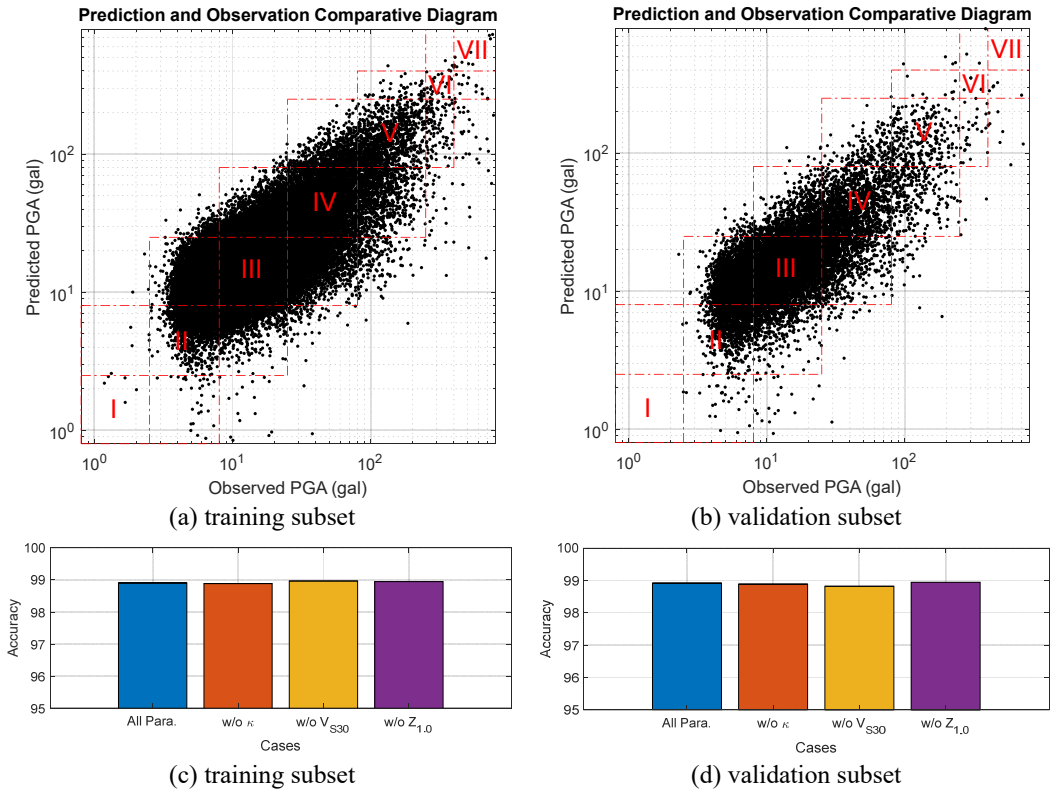
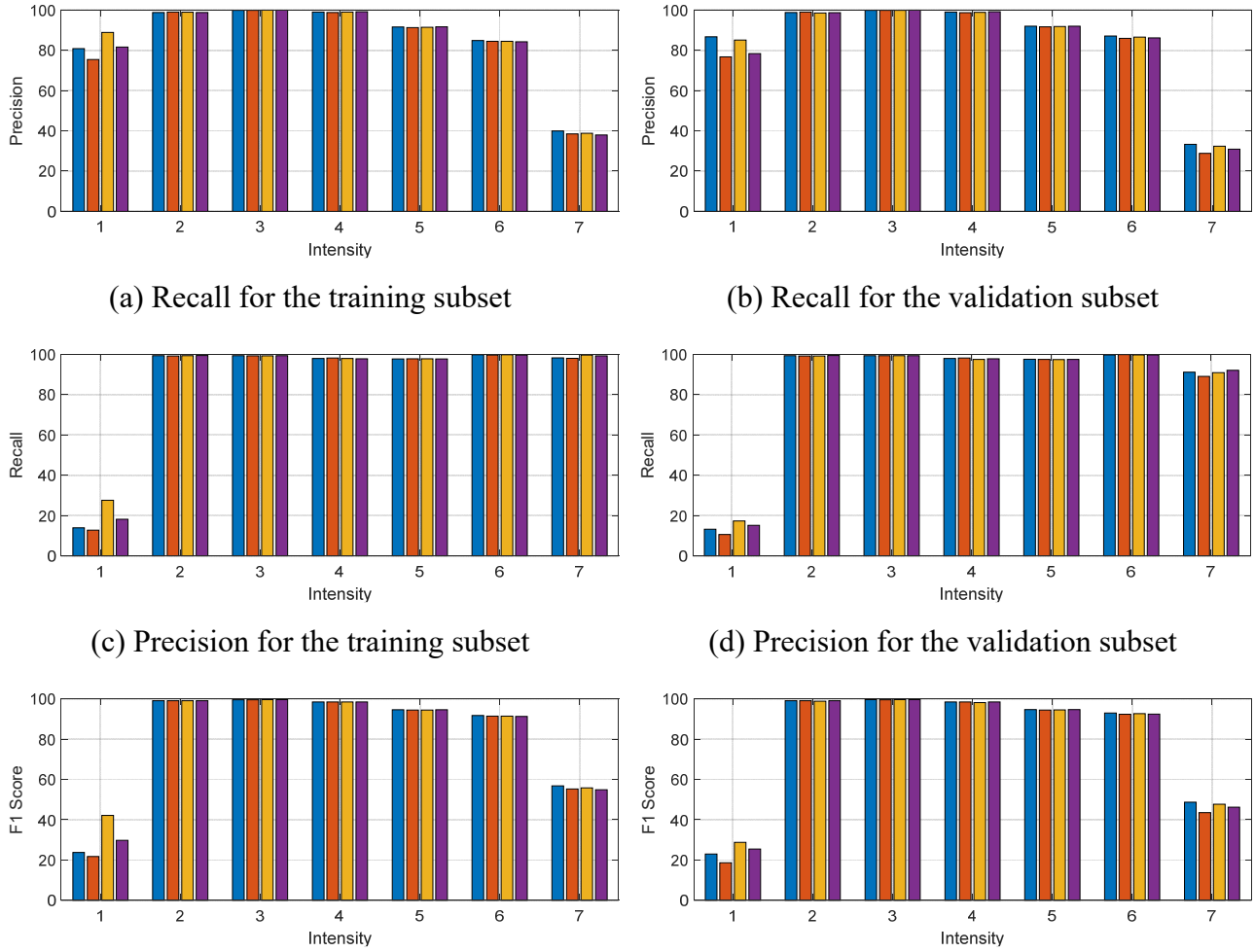


Figure 3. Predictions, observations, and accuracy of well-trained neural network models using the training and validation subsets; the red annotations indicate the seismic intensities.



(a) Recall for the training subset

(b) Recall for the validation subset

(c) Precision for the training subset

(d) Precision for the validation subset

(e) F1 score for the training subset

(f) F1 score for the validation subset

Figure 4. Other performance metrics for the training and validation subsets; the blue, orange, yellow, and purple bars represent the cases of all parameters, w/o k, w/o V_{S30}, and w/o $Z_{1.0}$, respectively.

4. Proposed ensemble framework

Through well-trained models utilizing neural networks, site parameters and earthquake features extracted from P-waves can be employed to predict PGA, as well as estimate intensities, during seismic events. However, neural networks represent only one type of supervised learning approach for correlating P-wave features with subsequent PGA; other candidates include SVM, decision trees, random forests (RF), and others. Validation methods and performance metrics were again exploited in this section to evaluate and investigate the effectiveness of various supervised learning approaches in predicting PGA and estimating seismic intensities. The results from different models are presented and discussed. Additionally, an ensemble model was proposed to address the advantages and disadvantages of the individual models. The accuracy of the ensemble model was also reported to demonstrate the reliability of applying the proposed method in EEW systems.

4.1. Results of ML models

As concluded in the previous section, all site parameters are recommended for predicting PGA and estimating intensities via neural networks. These parameters/features are also considered in various ML models. For the well-trained model, the overall performance is illustrated in Figure 5 and Figure 6. To be specific, Figure 5 demonstrates the accuracy of the well-trained ML models using the training and validation subsets, while Figure 6 presents the recall, precision, and F1 score for each intensity.

Support Vector Machines (SVM)

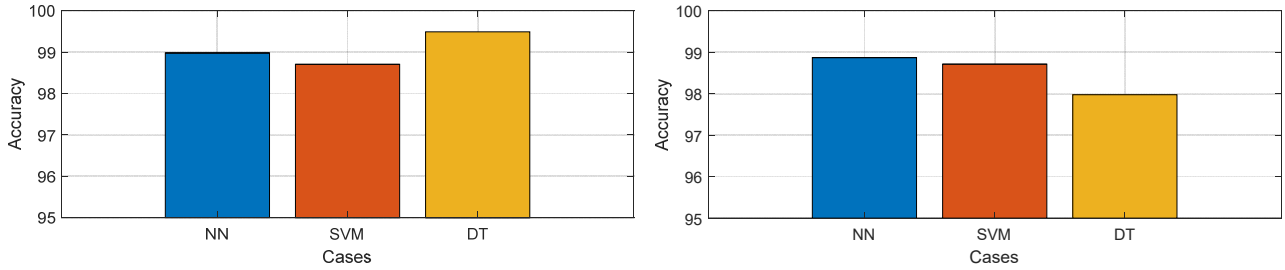
One result following the neural networks is learning using SVM [16]. SVM is a powerful and widely used supervised learning approach primarily used for classification tasks, although it can also be adapted for regression. Developed by Vladimir Vapnik and his colleagues in the 1990s, the primary concept behind SVM is to find the optimal boundary (hyperplane) that maximizes the margin between two classes [39,40]. The margin is the distance between the hyperplane and the closest data points from each class, known as support vectors, which are critical to defining the optimal hyperplane. The support vectors help make the model robust to changes in the data, meaning the model will perform well on unseen data. The advantages of SVM include its effectiveness in high-dimensional spaces, robustness to overfitting, and versatility; on the contrary, it has disadvantages like the expensive computation and the requirement of careful tuning.

The accuracy resulting from SVM demonstrates similar performance across ML models, showing a comparable capability to that of neural networks. In terms of other performance metrics, both SVM and neural networks exhibit distinct strengths; one model possesses better outcomes at low intensity levels, while the other shows slight superiority at high intensity levels. Generally, it can be concluded that the performance of these two ML models is equal or similar.

Decision Trees

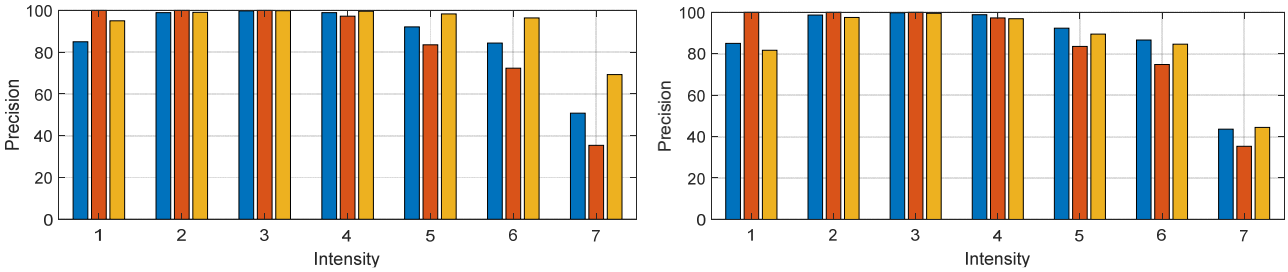
The last result is learning using decision trees. A decision tree is also a popular supervised learning approach used for both classification and regression tasks. The primary concept behind decision trees is to make decisions by following a series of simple rules based on feature values, working by recursively splitting the dataset into subsets based on feature values. It creates a tree-like model where each internal node represents a decision based on a feature, each branch represents an outcome of that decision, and each leaf node represents a final prediction or class [41]. The training process for decision trees is to choose the best feature to split the data at each node, with the goal of reducing uncertainty or disorder. The advantages of decision trees include their straightforward interpretation, scaling-free inputs, and nonlinear relationships; on the contrary, it has disadvantages like the overfitting problem, instability, and bias toward features with more levels.

The accuracy achieved by decision trees is highest in the training subset but lowest in the validation subset, indicating a tendency toward overfitting. When examining other performance metrics, the trends for decision trees differ significantly from those of SVM and neural networks. For instance, decision trees have exclusive performance in the training subset regarding recall, especially at high intensity levels; however, their performance is notably poor across all levels in the validation subset. Conversely, in terms of precision, decision trees outperform SVM and neural networks at low intensity levels, although this superiority does not persist at high intensity levels.

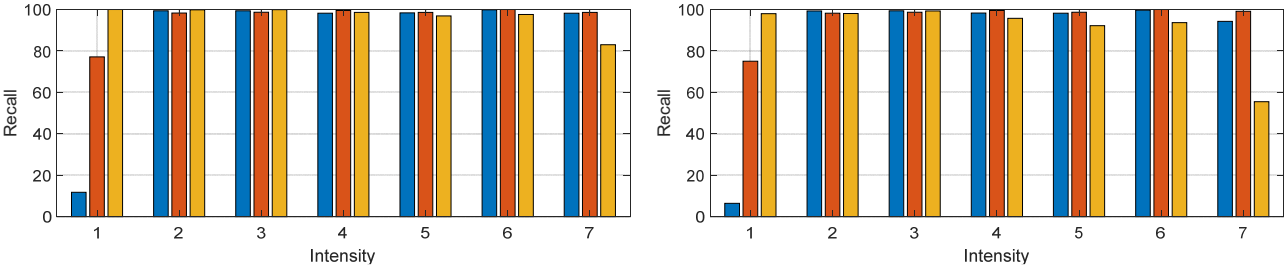


(a) Training subset (b) Validation subset

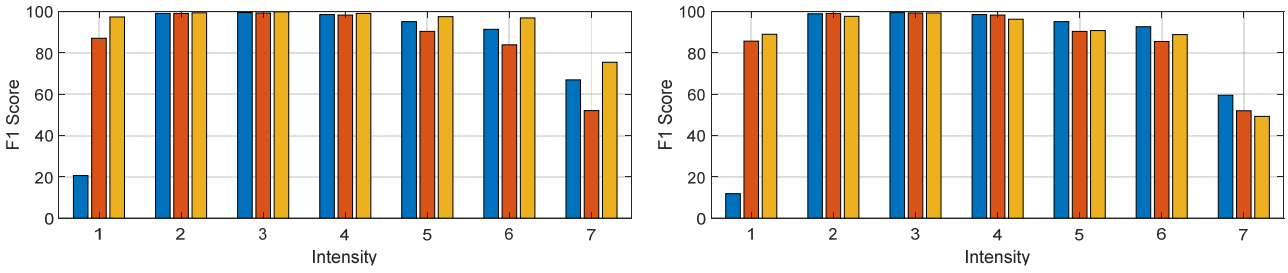
Figure 5. Accuracy of well-trained ML models using the training and validation subsets.



(a) Recall for the training subset (b) Recall for the validation subset



(c) Precision for the training subset (d) Precision for the validation subset



(e) F1 score for the training subset (f) F1 score for the validation subset

Figure 6. Other performance metrics for the training and validation subsets; the blue, orange, and yellow bars represent the models of neural network, SVM, and decision tree, respectively.

4.2. Network architecture

Even when incorporated with the testing subset, decision trees still reveal a tendency toward overfitting. Noteworthy, the validation subset is more similar to the real application since the parameters/features from the next seismic events eventually differ from the inputs in the training subset. Nevertheless, SVM and neural networks are more robust and have better generalization, although their performance does not surpass that of decision tree models in many respects. Compared to the models from SVM and neural networks, decision trees yield better outcomes at low intensity levels, especially at intensity level 1, since the range of low intensity levels is much narrower, as shown in Table 1. SVM and neural networks exhibit slight superiority at high intensity levels, particularly for neural networks. Thus, an ensemble model that combines the well-trained neural network and decision tree models can provide advantageous estimations.

As illustrated in Figure 7, a neural network is initially trained to classify inputs into low or high intensity levels. Based on this classification, a second neural network was trained to estimate high intensity levels using the same inputs due to the excellent performance demonstrated on the validation subset. Furthermore, a decision tree was trained to estimate low intensity levels, addressing the subpar performance exhibited by the neural network. This hierarchical architecture ensured that each model capitalized on its strengths while mitigating its weaknesses.

Subsequently, the threshold between low and high intensity levels were determined through engineering methods to enhance the ensemble model. In this study, 8, 25, and 80 gals were selected as candidate thresholds, while 40 gal was also included, as most transportation systems utilize 40 gal as an alarm threshold for decision-making during seismic events [9,42,43]. Among the results, the model using 40 gal outperforms the others and thus has been selected for further analysis. It is important to note that 20 models were trained for the first neural network to jointly classify inputs into low or high intensity levels in this study. This approach yields improved performance, achieving approximately 95% accuracy in distinguishing whether the estimation is greater than or less than 40 gal.

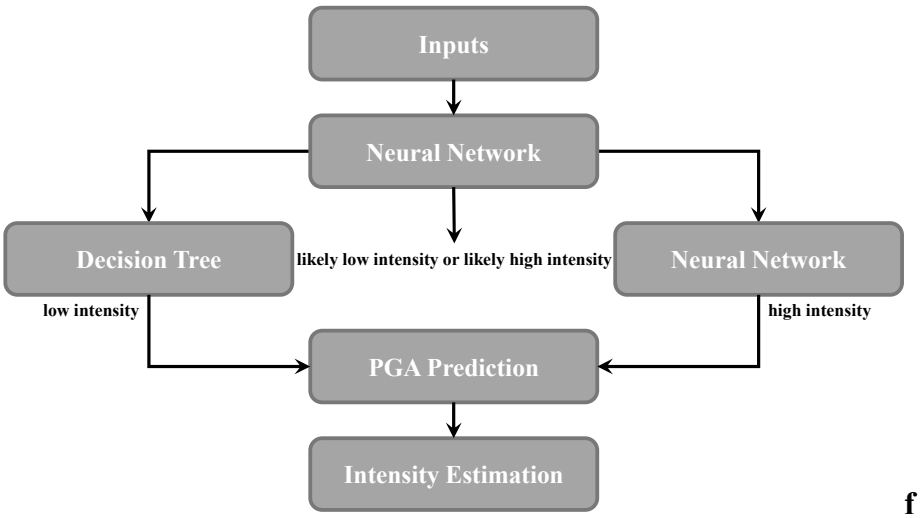
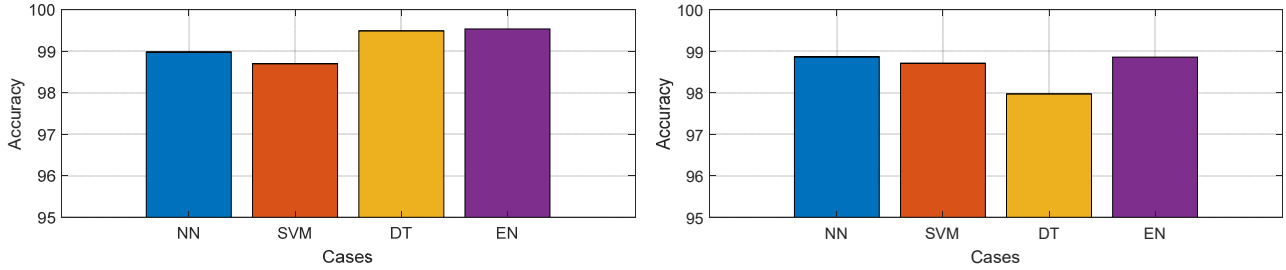


Figure 7. Flowchart and network architecture for ensemble models.

4.3. Results of the ensemble framework

Again, a total of 20 cross-validation were conducted to ensure generalization, and the performance of the ensemble model compared to other models is shown in Figure 8, Figure 9, and Figure 10. Regarding accuracy, the ensemble model achieves 99.5% and 98.9% for the training and validation subsets, respectively; the values catch up with those from decision trees in the training subset and from neural networks in the validation subset, as shown in Figure 8(a) and Figure 8(b). The observation generally confirms that decision trees produce better outcomes at low intensity levels, while neural networks demonstrate slight superiority at high intensity levels, and the hierarchical architecture ensures that each model leverages its strengths while minimizing its weaknesses.



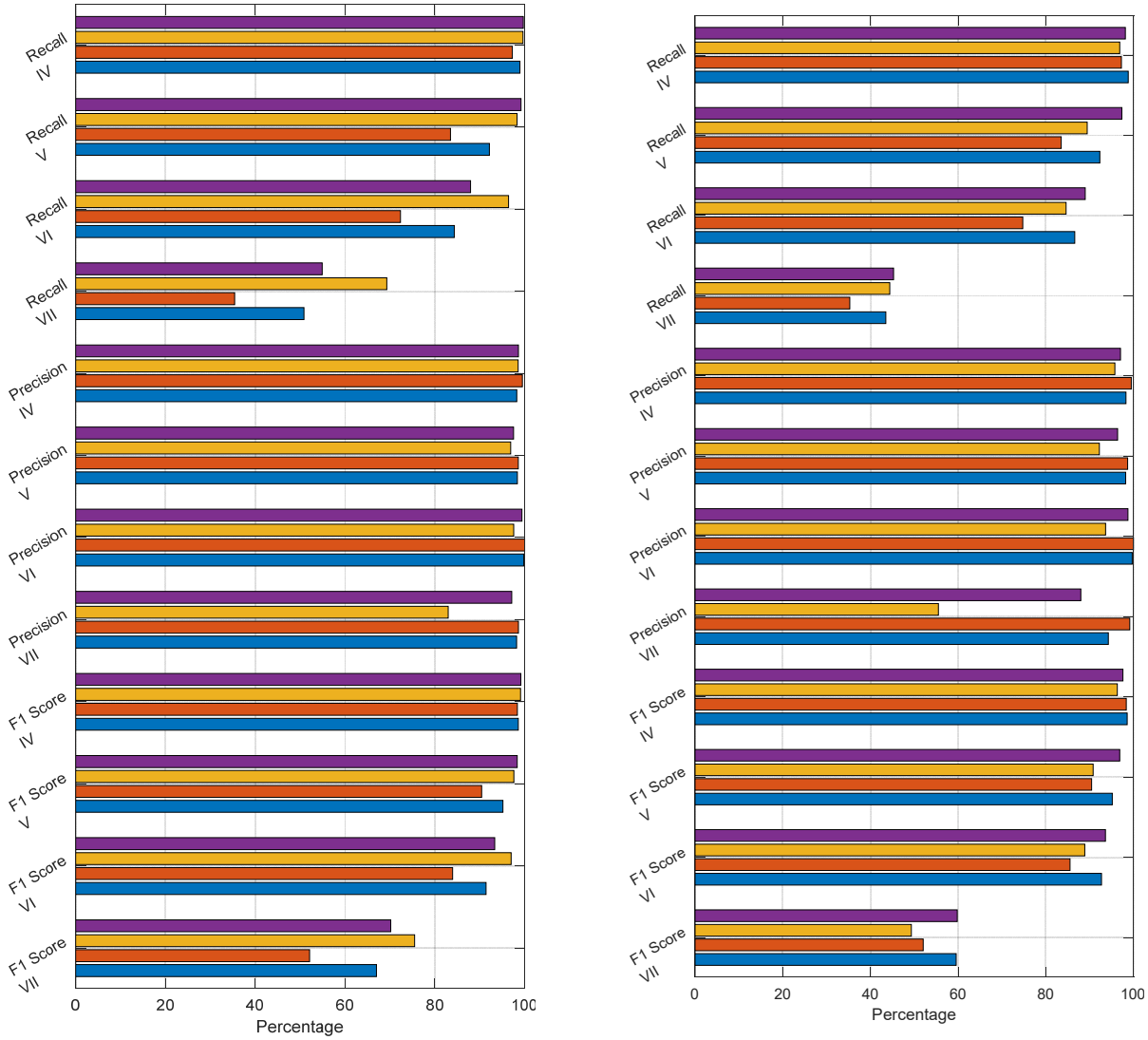
(a) Training subset (b) Validation subset

Figure 8. Accuracy of the ensemble model using the training and validation subsets.

To further investigate the benefits of using the ensemble model, the performance metrics beyond accuracy are presented in Figure 9; however, only intensity levels from 4 to 7 are illustrated due to space limitations. The ensemble model performs exceptionally well compared to neural networks, SVM, and decision trees. Some bars representing decision trees (yellow) are the highest in Figure 9(a) because of the overfitting. For example, the recall at intensity levels 6 and 7 (8.5% and 14.4%), the F1 score at intensity levels 6 and 7 (3.7% and 5.3%). The corresponding bars in Figure 9(b) are significantly lower, and as a result, the performance of the ensemble model closely resembles that of neural networks. For instance, the purple bars consistently exceed the yellow bars in the recall across all intensity levels, with the greatest difference being approximately 8% (at intensity level 5) in this figure. In terms of the F1 score, the purple bars are longer than the yellow bars, except at intensity level 1, where the largest difference exceeds 10% (at intensity level 7). The superiority is over 30% because decision trees provide the worst performance at intensity level 7. On the other hand, the ensemble model also surmounts the well-trained neural network model. In the training subset, the performance of the ensemble model consistently surpasses the neural networks in both the recall and F1 score across all intensity levels. Even though there are some inferiorities in the precision, the maximum transcendence is only about 1% in this figure. For the validation subset, the performance of the ensemble model and the neural networks is comparable; the largest differences are approximately 5%, 2%, and 6% for the recall, precision, and F1 score, respectively. This indicates that the ensemble model achieves a level of generalization similar to that of neural networks.

After discussing the estimations of seismic intensities, Figure 10 presents the predictions of PGA across the dataset alongside the observations. Again, only one cross-validation is displayed here due

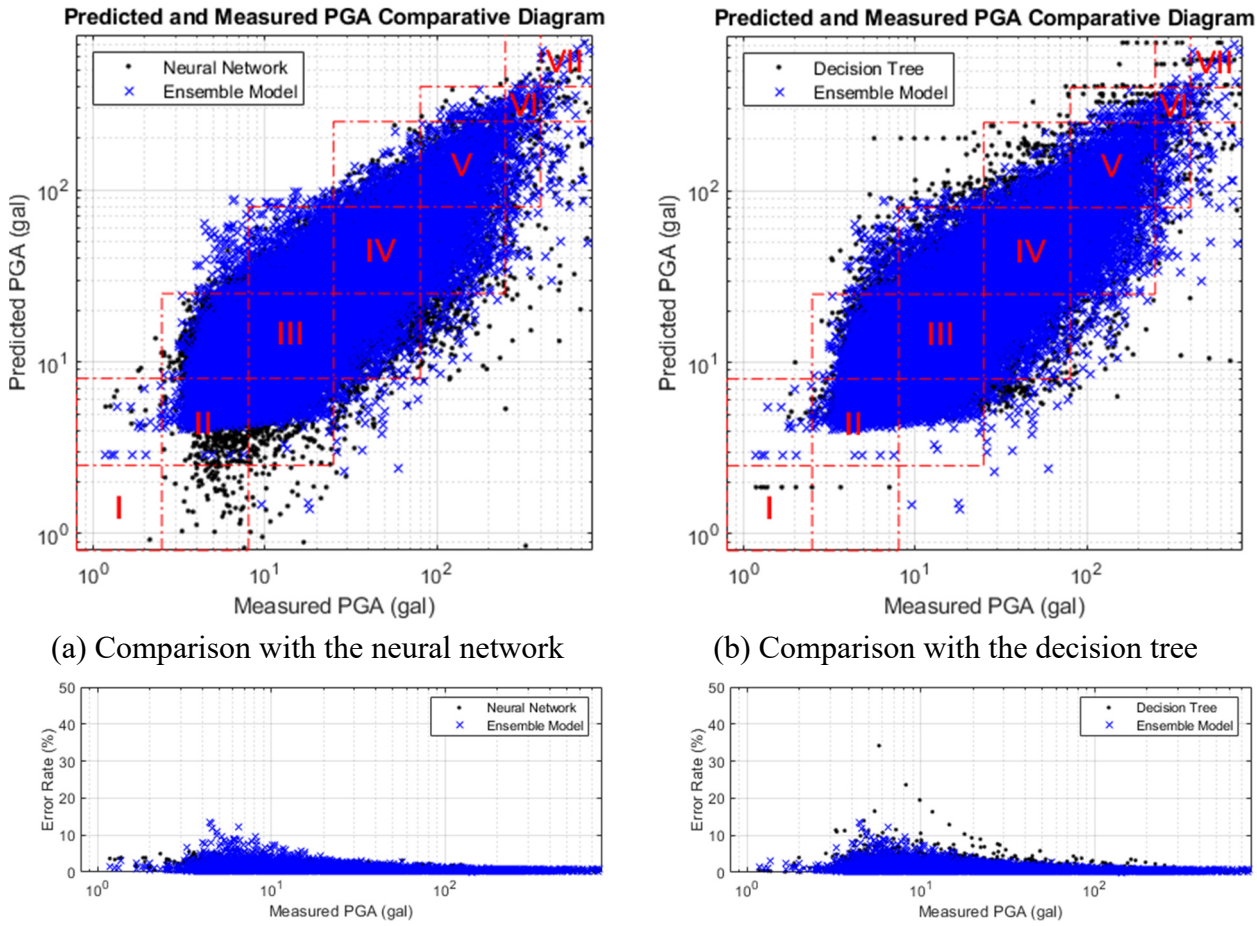
to the high similarity observed across multiple cross-validations. Each model demonstrates a strong positive correlation, indicating that the predictions are proportional to the observations, with a standard deviation of errors less than 30 gal, even when using decision trees and the ensemble model. Moreover, the error rates illustrated in Figure 10(c) and Figure 10(d) plainly exhibit that the predictions of the ensemble model sequentially resemble decision trees and neural networks at low and high intensity levels. To be specific, the superiority and inferiority of neural networks and decision trees are evident in this figure. The model based on neural networks exhibits significant deviations (far from the diagonal) at lower measured PGA, whereas the one based on decision trees shows many deviations at higher measured PGA. Conspicuously, the ensemble model capitalizes on the strengths of each model while minimizing its weaknesses.



(a) Training subset

(b) Validation subset

Figure 9. Results for the training and validation subsets (only intensity levels from 4 to 7); the blue, orange, yellow, and purple bars represent the models of neural network, SVM, decision tree, and ensemble model, respectively.



(a) Comparison with the neural network

(b) Comparison with the decision tree

(c) error rate compared with the neural network

(d) error rate compared with the decision tree

Figure 10. Predictions and observations of well-trained models throughout the dataset; the red annotations indicate the seismic intensities.

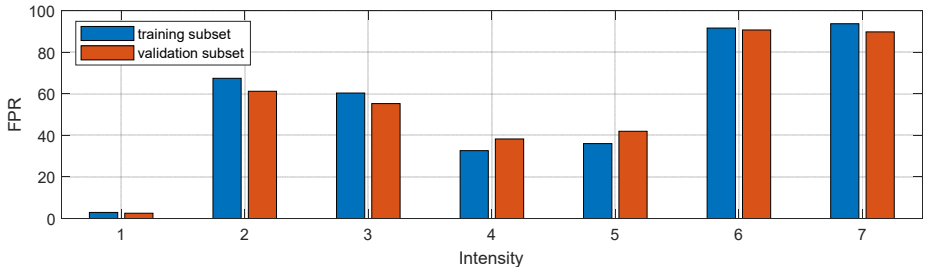


Figure 11. FPR of the ensemble model using the training and validation subsets.

The usability of EEW systems can be discussed in a more practical scenario to take advantage of FPR. Figure 11 demonstrates this performance metric, which results from the ensemble model using the training and validation subsets. Unfortunately, significant FPR can be observed throughout the seismic intensities. The minimum value occurs only at intensity level 1. For medium or medium-low intensity levels, FPR is approximately 50%, and for high intensity levels, FPR approaches 90%. However, considering the definition of FPR, the observation also indicates that the EEW systems

implemented with the ensemble model tend to estimate towards the conservative side, with the predicted PGA being higher than the observation. Certainly, conservative information is more advantageous as a warning system for a destructive disaster, although there must be some false alarms.

4.4. Discussions and limitations of this study

The EEW systems can roughly be divided into two sequential functions: One is an automatic P-wave picker to determine a P-wave arrival time, and the other is a seismic feature predictor to provide an informative warning message. Although the warning time is a combination of the two, it focuses more on the P-wave picker and, therefore, is not addressed in this study. According to references, effective warning can be done by on-site systems if the epicenter distance is larger than 20 km, and it is about 30 km for regional systems, considering the propagation speeds of seismic waves [6]. However, the late alert zone, which cannot effectively receive a warning, may vary from 30 to 50 km depending on the time consumed to analyze the event and issue the warning [5,8,44]. The distance takes about 5 to 8 seconds after the P-wave propagates from the epicenter. Thus, on-site prediction using three-second seismic information is very competitive. In fact, 3 seconds is typically considered after P-wave arrival in many studies, so the intensity estimation, followed by the issuing of a warning in near-source regions, may take at least 8 to 10 seconds for both systems [5,7,13,14].

As mentioned in Section 2.1, very short recordings were eliminated because we intended to predict seismic intensities through the P-wave. However, the recordings with a close P-wave and PGA arrival times were not excluded from the dataset to consider the real-time operations of an on-site system. Those represent data with very short epicenter distances, and the integrated velocity and displacement may consequently be deteriorated by the S-wave beginning within 3 seconds of the triggering time. Thus, the training inputs, P-wave features, can be contaminated because of erroneous integration. In fact, 11,810 of 70,947 recordings have epicenter distances shorter than 7.5 km, which is an approximate distance difference between P-waves and S-waves after 3 seconds of travel. This about 16.6% of all recordings. Moreover, 2,124 recordings have a narrow time window (less than 3 seconds between the P-wave and PGA arrival times). Over 90% of it is within magnitude 4.0 and 5.0, and this about 3.0% of all recordings. The above information indicates that, although some performance metrics at some intensity levels are not great, the discussion in Section 4 can fully reflect the results under the consideration of real-time operations.

5. Conclusion

Predicting PGAs and estimating intensities are essential for EEW systems, so the enhancements in accuracy not only provide critical lead time for emergency responses but also improve the overall effectiveness and timeliness of these systems. Numerous models based on ML techniques have been developed to achieve these objectives by correlating features extracted from P-waves with subsequent PGA values during seismic events. In this study, various models are trained using supervised learning approaches to estimate PGA, and the well-trained models are employed to generate accurate predictions. The development of these models relies on the training inputs (the site parameters and the P-wave features), as well as the ML models/algorithms, and the validation processes; thus, they are

compared and discussed in this study. The effectiveness of various models in predicting PGA and estimating seismic intensities is evaluated through performance metrics. Additionally, an ensemble model is proposed to leverage the strengths and mitigate the weaknesses of the individual models. The accuracy of the proposed ensemble framework is also reported to demonstrate the reliability of applying the proposed ML model in EEW systems. The conclusions drawn from these results can be summarized as follows:

- Besides the P-wave features, the site parameters have an advantageous effect on predicting PGA and estimating intensities (via neural networks), and performance is particularly robust across intensity levels.
- Despite some issues of overfitting, decision trees demonstrate exceptional performance at low intensity levels; conversely, SVM and neural networks outperform decision trees at high intensity levels, necessitating trade-offs between two ML models.
- By incorporating neural networks and decision trees, the ensemble model achieves exclusive results across all performance metrics; nevertheless, the training and validation subsets demonstrate a level of generalization comparable to that of neural networks.

Using the proposed ensemble framework, the accuracy of predicting PGAs and estimating intensities is demonstrated for the effectiveness and timeliness of EEW systems. The recent advances and effectiveness of ML and AI make them especially suitable for seismic damage assessment in earthquake engineering and give valuable information for emergency response and decision-making. However, the proposed model and conclusion are limited to Taiwan's datasets so far, although they are applicable to various seismic regions. Future research can be conducted with global datasets to address the adaptability of this ensemble framework.

Author contributions

Conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing – original draft, and writing – review & editing, S.-K.H; data curation, formal analysis, software, and validation, G.-H. Tseng. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

Shieh-Kung Huang is a Guest Editor for AIMS Geosciences and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

Acknowledgements

The study was supported by National Science and Technology Council under contract No. NSTC 113-2625-M-005-011-, Taiwan. The data used in the study was provided by Central Weather Bureau (CWB), and the National Center for Research on Earthquake Engineering (NCREE), Taiwan.

References

1. Shibahara S (2011) The 2011 Tohoku earthquake and devastating tsunami. *Tohoku J Exp Med* 223: 305–307. https://doi.org/10.1620/tjem.223.305
2. Koronovskii NV, Zakharov VS, Naimark AA (2021) The unpredictability of strong earthquakes: New understanding and solution of the problem. *Moscow Univ Geol Bull* 76: 366–373. https://doi.org/10.3103/S0145875221040074
3. Geller RJ (1997) Earthquake prediction: a critical review. *Geophys J Int* 131: 425–450. https://doi.org/10.1111/j.1365-246X.1997.tb06588.x
4. Bhardwaj A, Sam L, Martin-Torres FJ (2021) The challenges and possibilities of earthquake predictions using non-seismic precursors. *Eur Phys J Spec Top* 230: 367–380. https://doi.org/10.1140/epjst/e2020-000257-3
5. Kumar R, Mittal H, Sandeep, Sharma B (2022) Earthquake genesis and earthquake early warning systems: challenges and a way forward. *Surv Geophys* 43: 1143–1168. https://doi.org/10.1007/s10712-022-09710-7
6. Satriano C, Wu YM, Zollo A, Kanamori H (2011) Earthquake early warning: Concepts, methods and physical grounds. *Soil Dyn Earthq Eng* 31: 106–118. https://doi.org/10.1016/j.soildyn.2010.07.007
7. Allen RM, Melgar D (2019) Earthquake early warning: Advances, scientific challenges, and societal needs. *Annu Rev Earth Pl Sci* 47: 361–388. https://doi.org/10.1146/annurev-earth-053018-060457
8. Wu YM, Kanamori H (2005) Experiment on an onsite early warning method for the Taiwan early warning system. *B Seismol Soc Am* 95: 347–353. https://doi.org/10.1785/0120040097
9. Lin CCJ, Lin PY, ChangTM, et al. (2012) *Development of on-site earthquake early warning system for Taiwan*, 329–358.
10. Cremen G, Galasso C (2020) Earthquake early warning: Recent advances and perspectives. *Earth-Sci Rev* 205: 103184. https://doi.org/10.1016/j.earscirev.2020.103184
11. Cochran ES, Husker AL (2019) How low should we go when warning for earthquakes?. *Science*, 366: 957–958. https://doi.org/10.1126/science.aaz6601
12. Tan ML, Becker JS, Stock K, et al. (2022) Understanding the social aspects of earthquake early warning: A literature review. *Front Commun* 7: 939242. https://doi.org/10.3389/fcomm.2022.939242
13. Wald DJ (2020) Practical limitations of earthquake early warning. *Earthq Spectra* 36: 1412–1447. https://doi.org/10.1177/8755293020911388
14. Meier M. (2017) How "good" are real-time ground motion predictions from earthquake early warning systems?. *J Geophys Res: Solid Earth* 122: 5561–5577. https://doi.org/10.1002/2017JB014025

15. Böse M, Heaton TH, Hauksson E (2012) Real-time finite fault rupture detector (FinDer) for large earthquakes. *Geophys J Int* 191: 803–812. https://doi.org/10.1111/j.1365-246X.2012.05657.x
16. Hsu TY, Huang SK, Chang YW, et al. (2013) Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan. *Soil Dyn Earthq Eng* 49: 210–217. https://doi.org/10.1016/j.soildyn.2013.03.001
17. Colombelli S, Caruso A, Zollo A, et al. (2015) AP wave-based, on-site method for earthquake early warning. *Geophys Res Lett* 42: 1390–1398. https://doi.org/10.1002/2014GL063002
18. Hoshiba M, Aoki S (2015) Numerical shake prediction for earthquake early warning: Data assimilation, real-time shake mapping, and simulation of wave propagation. *Bull Seismol Soc Am* 105: 1324–1338. https://doi.org/10.1785/0120140280
19. Kodera Y, Saitou J, Hayashimoto N, et al. (2016) Earthquake Early Warning for the 2016 Kumamoto Earthquake: Performance Evaluation of the Current System and Simulations of the Next-Generation Methods of the Japan Meteorological Agency. *In AGU Fall Meeting Abstracts*, Vol. 2016, S32A-02. https://doi.org/10.1186/s40623-016-0567-1
20. Hsu TY, Pratomo A (2022) Early peak ground acceleration prediction for on-site earthquake early warning using LSTM neural network. *Front Earth Sci* 10: 911947. https://doi.org/10.3389/feart.2022.911947
21. Mohmoud MM, Keshk A, Ahmed H (2023) Prediction Model for Peak Ground Acceleration Using Deep Learning. *IJCI. International Journal of Computers and Information* 10: 175–183. https://doi.org/10.21608/ijci.2023.236143.1131
22. Liu Y, Zhao Q, Wang Y (2024) Peak ground acceleration prediction for on-site earthquake early warning with deep learning. *Sci Rep* 14: 5485. https://doi.org/10.1038/s41598-024-56004-6
23. Iaccarino AG, Cristofaro A, Picozzi M, et al. (2024) Real-time prediction of distance and PGA from P-wave features using Gradient Boosting Regressor for on-site earthquake early warning applications. *Geophys J Int* 236: 675–687. https://doi.org/10.1093/gji/ggad443
24. Central Weather Administration (CWA, Taiwan) (2012) Central Weather Administration Seismographic Network. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/T5
25. Allen RV (1978) Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am* 68: 1521–1532. https://doi.org/10.1785/BSSA0680051521
26. Yadav R, Kumar D, Chopra S (2018) The high frequency decay parameter κ (kappa) in the region of North East India. *Open J Earthquake Res* 7: 141–159. https://doi.org/10.4236/ojer.2018.72009
27. Stanko D, Markušić S, Korbar T, et al. (2020) Estimation of the high-frequency attenuation parameter kappa for the Zagreb (Croatia) seismic stations. *Appl Sci* 10: 8974. https://doi.org/10.3390/app10248974
28. Stolte A, Wentz R, Wotherspoon L (2024) Providing robust Vs30 estimates for seismic site classification in Aotearoa New Zealand. *SESOC J* 37: 58–65.
29. Liu W, Peng Y, Wang J (2023) Spatially correlated Vs30 estimation in the Beijing area. *Front Earth Sci* 10: 950582. https://doi.org/10.3389/feart.2022.950582
30. Kuo CH, Lin CM, Chang SC, et al. (2016) Progress Towards a Comprehensive Site Database for Taiwan Strong Motion Network. *In AGU Fall Meeting Abstracts*, 2016, S42B-03.

31. Kuo CH, Chen CT, Lin CM, et al. (2016) S-wave velocity structure and site effect parameters derived from microtremor arrays in the Western Plain of Taiwan. *J Asian Earth Sci* 128: 27–41. https://doi.org/10.1016/j.jseaes.2016.07.012
32. Kramer SL, Stewart JP (2024) *Geotechnical earthquake engineering*. CRC Press. https://doi.org/10.1201/9781003512011
33. Kanamori H (2005) Real-time seismology and earthquake damage mitigation. *Annu Rev Earth Planet Sci* 33: 195–214. https://doi.org/10.1146/annurev.earth.33.092203.122626
34. Reed JW, Kassawara RP (1990) A criterion for determining exceedance of the operating basis earthquake. *Nucl Eng Des* 123: 387–396. https://doi.org/10.1016/0029-5493(90)90259-Z
35. Wu MH, Wang JP, Chiang PE (2022) Cumulative absolute velocity (CAV) seismic hazard assessment for Taiwan. *J Earthq Eng* 26: 3440–3460. https://doi.org/10.1080/13632469.2020.1803161
36. Festa G, Zollo A, Lancieri M (2008) Earthquake magnitude estimation from early radiated energy. *Geophys Res Lett* 35. https://doi.org/10.1029/2008GL035576
37. Picozzi M, Bindi D, Spallarossa D, et al. (2018) A rapid response magnitude scale for timely assessment of the high frequency seismic radiation. *Sci Rep* 8: 8562. https://doi.org/10.1038/s41598-018-26938-9
38. Huang SK, Li CH, Shen ZP, et al. (2014) Predicting Peak Ground Acceleration Using Artificial Neuron Network with P Wave Features for On-site Earthquake Warning System in Taiwan. *In Proceedings of the 5th Asia Conference on Earthquake Engineering*, 16–18.
39. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20: 273–297. https://doi.org/10.1007/BF00994018
40. Ben-Hur A, Horn D, Siegelmann HT, et al. (2001) Support vector clustering. *J Mach Learn Res* 2: 125–137.
41. Breiman L (2017) Classification and regression trees. Routledge. https://doi.org/10.1201/9781315139470
42. Minson SE, Cochran ES, Wu S, et al. (2021). A framework for evaluating earthquake early warning for an infrastructure network: An idealized case study of a northern California rail system. *Front Earth Sci* 9: 620467. https://doi.org/10.3389/feart.2021.620467
43. Tan M, Hu Q, Wu Y, et al. (2024) Decision-making method for high-speed rail early warning system in complex earthquake situations. *Transp Safety Env* 6: tdad034. https://doi.org/10.1093/tse/tdad034
44. Wu YM, Mittal H, Chen DY, et al. (2021) Earthquake early warning systems in Taiwan: Current status. *J Geol Soc India* 97: 1525–1532. https://doi.org/10.1007/s12594-021-1909-6

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)