Journal of Current Science and Technology, October-December 2025
Vol. 15 No. 4, Article 139
ISSN 2630-0656 (Online)

Cite this article: Chomchit, P., Aramkul, S., Somchit, Y., & Champrasert, P. (2025). Earthquake early warning using multi-channels echo state extreme learning machine. *Journal of Current Science and Technology*, *15*(4), Article 139. https://doi.org/10.59796/jcst.V15N4.2025.139

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Earthquake Early Warning Using Multi-Channels Echo State Extreme Learning Machine

Phiphat Chomchit[1], Somrawee Aramkul[2], Yuthapong Somchit[3], and Paskorn Champrasert[3,*]

[1]Data Science Consortium, Chiang Mai University, Chiang Mai 50200, Thailand
[2]Department of Computer, Faculty of Science and Technology, Chiang Mai Rajabhat University, Chiang Mai 50300, Thailand
[3]OASYS Research Group, Faculty of Engineer, Chiang Mai University, Chiang Mai 50200, Thailand

*Corresponding author; E-mail: paskorn.c@cmu.ac.th

Received 23 January 2025; Revised 30 June 2025; Accepted 3 July 2025; Published online 20 September 2025

Abstract

Predicting earthquake strong motions is crucial for mitigating seismic risks and enhancing the effectiveness of Earthquake Early Warning (EEW) systems. While conventional models are capable of high precision, they often require substantial computational resources, limiting their practicality for real-time applications. This study proposes the Multi Echo-State Extreme Learning Machine (Multi ES-ELM), an efficient and effective alternative for strong motion prediction. It compares the performance of Multi ES-ELM with two well-established models-Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)-using multi-channel time-series data. The CNN model achieved high performance with an accuracy of 94.65 ± 0.30, recall of 92.84 ± 2.36, precision of 87.34 ± 2.35, and F1-score of 89.95 ± 0.41. In contrast, the RNN model showed significant variability, with an accuracy of 84.83 ± 19.40, recall of 84.93 ± 13.34, precision of 74.18 ± 18.15, and F1-score of 77.80 ± 16.40. Notably, the Multi ES-ELM model demonstrated competitive accuracy (93.46 ± 0.22), high recall (96.50 ± 0.52), precision (81.53 ± 0.53), and F1-score (88.38 ± 0.37), while using significantly fewer resources-only 882 parameters and a model size of 0.003 MB. These results highlight Multi ES-ELM as a highly efficient and reliable model for real-time EEW, overcoming the computational challenges of traditional approaches. Its performance and resource efficiency underscore its potential for practical implementation in seismic risk mitigation and for improving community resilience against seismic hazards.

Keywords: *earthquake early warning; echo state network; ground motion classification; neural network*

1. Introduction

Earthquake early warning (EEW) is crucial for mitigating and protecting against losses from ground shaking. Cremen & Galasso (2020) explain that EEW is a real-time system that enables people in risk areas to protect themselves before ground shaking occurs. It provides a warning and indicates the possible severity, giving people time to prepare and reducing losses. EEW systems are crucial for countries frequently experiencing earthquakes, as their warning messages can trigger automated responses in transportation services, factories, and facilities like power plants, gas pipelines, and dams, while also alerting individuals (Gasparini et al., 2007; Allen & Melgar, 2019). Various countries, such as Japan, Mexico, and Taiwan, have advanced EEW systems (Gasparini et al., 2007; Hsiao et al., 2009), while others are developing theirs (Gasparini et al., 2007; Allen & Melgar, 2019).

Although abundant earthquake data exist, most are from small events, making strong motions and large earthquakes rare. As a result, models must be updated to maintain predictive performance as new data become available. Norisugi et al. (2024) highlight that limited and non-diverse earthquake catalogs pose significant challenges for EEW systems operating in complex environments. Meng et al. (2023), and Lara et al. (2023) emphasize the need for real-time feature extraction from signal data and the requirement for smaller models that can run on compact, local computing devices. EEW models must function on small local machines, reducing reliance on cloud processing, enabling faster responses, and improving resilience. However, large predictive models can exceed the memory and processing capabilities of Central Processing Units (CPUs) and Graphics Processing Units (GPUs), preventing EEW deployment in areas with limited resources.

This research highlights the need for computationally efficient machine learning models for processing time-series data like seismic waves. The Multi-channel Echo State Extreme Learning Machine (Multi ES-ELM) is proposed, combining Echo State Networks (ESN) and Extreme Learning Machines (ELM) to optimize computational resource use while maintaining high earthquake prediction accuracy. Seismic monitoring data from telemetry stations are used as input, with the predicted Peak Ground Acceleration (PGA) as output. Gallicchio et al. (2017) demonstrate ESN's ability to extract features from time series data without training, while ELM rapidly learns from these features.

This study evaluates and compares CNN, RNN, and Multi ES-ELM for real-time EEW systems. Although statistical methods could be used, machine learning methods like CNN and RNN are superior at uncovering hidden patterns in complex, multi-channel time series data. Machine learning models can also adapt to new data without complete rebuilding, unlike conventional statistical approaches. The models are evaluated using accuracy, precision, recall, and F1-score, with performance stability assessed through the mean and standard deviation across multiple trials. This study further highlights Multi ES-ELM's advantages in training speed, memory efficiency, and overall performance, making it suitable for resource-constrained environments. Experimental results show that Multi ES-ELM performs comparably or better than CNN and RNN while consuming significantly fewer computational resources.

1.1 Earthquake Early Warning Systems

The EEW system detects seismic pulses during earthquakes and assesses potential damage (Chiang et al., 2022; Jozinović et al., 2020; Li et al., 2018). It plays a critical role in rapidly evaluating seismic activity and issuing warnings to affected individuals and organizations, enabling citizens to evacuate and services to suspend operations to mitigate damage (Gasparini et al., 2007). A key component of the EEW system is the seismometer, which continuously monitors ground motion in earthquake-prone regions. Data from seismometers are transmitted in real time to processing centers, where earthquake intensity and characteristics are evaluated. Once detected, warnings are disseminated through multiple channels, including the internet, SMS, sirens, and public broadcasts (Gasparini et al., 2007; Allen & Melgar, 2019), ensuring timely protective actions. Importantly, EEW systems do not predict earthquakes in advance. They detect seismic activity and estimate an earthquake's severity based on initial P-waves, which provide critical information about the more destructive surface waves (Chiang et al., 2022; Jozinović et al., 2020; Li et al., 2018). This ability to issue early warnings highlights their vital role in mitigating earthquake impacts.

1.2 Earthquake Early Warning Techniques

EEW systems are designed to detect hazardous earthquakes and issue warnings. Two conventional algorithms predict earthquake severity: source-based algorithms, which use seismic source information such as hypocenter, epicenter, and magnitude, and ground motion-based algorithms, which analyze real-time ground shaking data like acceleration and velocity (Chandrakumar et al., 2022). Recent studies propose using P-waves directly to predict peak ground acceleration (PGA), bypassing the need for detailed source information (Chiang et al., 2022; Yanwei et al., 2021; Gasparini et al., 2007). Since P-waves travel faster than other seismic waves, early analysis enables rapid assessment and timely warnings (Chiang et al., 2022; Allen & Melgar, 2019), helping to reduce earthquake impacts.

Conventional EEW models relied on physical and statistical methods, typically issuing warnings when ground motion exceeded thresholds (Chiang et al., 2022; Meilano et al., 2022). Although effective, they often produced high false alarm rates and lacked adaptability to new seismic data (Gasparini et al., 2007). Manual calibration further limited their responsiveness. These limitations have driven the shift toward AI-based models, which offer greater precision, lower false alarm rates, and better adaptability to the complexities of real-world seismic activity (Chiang et al., 2022).

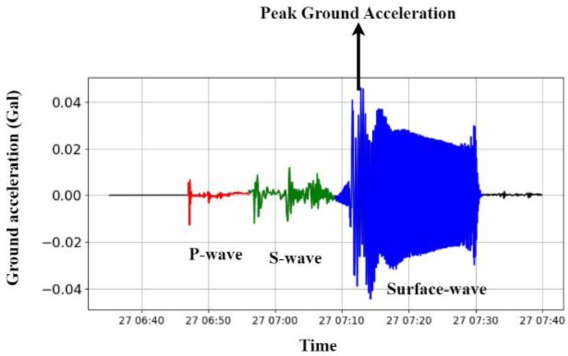


Figure 1 P-wave, S-wave, Surface-wave, and PGA

1.3 Seismic Waves and Peak Ground Acceleration

Understanding the distinction between body and surface waves is essential for assessing earthquake dynamics. Body waves, which travel through the Earth's interior, are the first detected during an earthquake. As shown in Figure 1, P-waves arrive first, followed by S-waves. P-waves, being the fastest, are crucial for early earthquake prediction. In contrast, surface waves travel along the ground and are slower than body waves (Mousavi & Beroza, 2020).

Despite being slower, surface waves are the primary cause of earthquake damage. The highest amplitude of the surface waves is named peak ground acceleration (PGA), which has a measuring unit as Gal ($1Gal = 1cm/sec^2$) (Chiang et al., 2022). In the case of a PGA higher than 80 Gal, the human can feel the ground shaking. So, the P-wave can predict the dangers of the S-wave and surface waves that travel slower. People are given a window of opportunity for evacuation due to the time lag between P-waves and surface waves. Peak Ground Acceleration (PGA) is a critical parameter in seismic engineering, influencing the design and assessment of structures in earthquake-prone areas. This review synthesizes recent studies on PGA, focusing on prediction methods, attenuation characteristics, and regional analyses.

1.4 Deep Learning in Earthquake Early Warning

In recent years, deep learning methods have emerged as powerful tools for enhancing the EEW system. Unlike traditional models, deep learning algorithms do not require extensive expert intervention for modeling, relying instead on data-driven learning. Researchers have explored various deep learning architectures, including artificial neural networks, RNN, and CNN, to improve earthquake monitoring and warning accuracy (Chiang et al., 2022; Jozinović et al., 2020; Li et al., 2018; Mousavi & Beroza, 2020). These advanced methods have shown promise in capturing the complex patterns of seismic data, leading to more accurate and timely warnings.

Various countries have researched and applied deep learning models to the EEW system (Gasparini et al., 2007). The global interest in this approach highlights its potential to revolutionize seismic activity monitoring and early warnings. By leveraging the power of deep learning, researchers aim to create a more responsive and reliable EEW system that can better protect lives and infrastructure in earthquake-prone regions. However, implementing deep learning models in the EEW system presents significant challenges, primarily due to their high computational demands. Real-time seismic data processing requires substantial computational resources, making it costly to train and deploy these models effectively (Wu et al., 2021). The need for rapid data processing in an emergency context further complicates deploying these resource-intensive models, particularly in regions with limited access to high-performance computing.

Additionally, the EEW system must frequently retrain models to incorporate new seismic data from newly installed monitoring stations and adapt to the specific seismic characteristics of different regions. The complex data necessitate models that can efficiently handle multiple time series data channels and quickly retrain to maintain accuracy and reliability. The ability to continuously update the models ensures that the EEW systems remain effective as new data become available. Thus, an EEW system that uses deep learning models incurs immense computational costs for training and classification (Li et al., 2018). These costs are a significant barrier to widespread implementation, particularly in resource-constrained environments. However, despite these challenges, the potential benefits of deep learning in improving the accuracy and timeliness of earthquake warnings make it a promising area of ongoing research and development.

1.5 ESN and ELM in Earthquake Early Warning

The Echo State Network (ESN) model is a machine learning model that must learn from time series data like an RNN, but ESN has a fixed-weight recurrent part (Cucchi et al., 2022). Due to the fixed recurrent unit weights, the ESN can avoid the backpropagation process, which leads to vanishing gradient problems and consumes significant time to train. Therefore, the ESN has efficient learning and fast training for time series data (Bianchi et al., 2020).

ESNs utilize a fixed random network to project input signals into a high-dimensional space, effectively capturing temporal dynamics. This fixed random network possesses memory properties, as the spectral radius of its weight matrix plays a critical role in determining the network's memory capacity, directly influencing its performance in temporal tasks (Soltani et al., 2023). A significant parameter in ESN is the spectral radius, which sets the network's memory characteristics. The fixed random network size is another parameter determined for optimizing the ESN model. The crucial ability of the fixed random network in the ESN model is signal denoising and feature extraction from input signal data (Sun et al., 2024). However, the simple ESN extracts the single time series data. Consequently, the conventional ESN cannot be used for multiple time series data like RNNs can.

The Extreme Learning Machine (ELM) is a single hidden layer neural network. The structure of ELM consists of two parts. The first part comprises feature extractors, including random weights and nonlinear functions such as the sigmoid. The first part tries to map feature data into a nonlinear space. Also, this part avoids the training process. The second part is a linear layer. This part map features from the nonlinear space to labeled data. This part uses a pseudo-inverse method to optimize the weights of the linear layer. This method requires only one brief training phase. Thus, ELM training is extremely fast, as reflected in its name, "Extreme Learning Machine".

ELM training is simple and fast. Many studies have recently switched the training process from conventional methods like stochastic gradient descent, which consumes a lot of resources, to an ELM training process. Additionally, some ELM variants replace the feature extractors with other architectures, such as CNN or RNN. The ESN model has also been applied in seismic exploration, as discussed by (Carvalho et al., 2018). One challenge in this context is that seismic waves contain primary and multiple reflections, complicating seismic analysis. To address this, Carvalho et al. (2018) employed nonlinear models that integrate ESN and ELM to improve the attenuation of various reflections. The results demonstrate the potential of ESN and ELM models for filtering noisy signal data. Carvalho et al., (2018) emphasize that ESN and ELM can effectively balance processing capability and computational resource consumption.

1.6 Metrics

The Earthquake Early Warning systems are the systems (EEWS) that classify whether a seismic wave can generate strong motion or not. Therefore, this research requires metrics for binary classification problems (Chiang et al., 2022). The most commonly used metric for binary classification is accuracy. However, this metric is unsuitable for imbalanced data such as the severity scale of ground motions. Therefore, the proposed model must be evaluated using other metrics such as accuracy, precision, recall, and F1-score, as described in Equations 1, 2, 3, and 4, respectively.

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100. \tag{1}$$

$$\text{precision} = \frac{TP}{TP+FP} \times 100. \tag{2}$$

$$\text{recall} = \frac{TP}{TP+FN} \times 100. \tag{3}$$

$$\text{f1score} = 2 \times \frac{\text{precision} \times \text{truepositiverate}}{\text{precision} + \text{truepositiverate}} \times 100. \tag{4}$$

where $TP, TN, FP,$ and FN are true positive, true negative, false positive, and false negative, respectively.

2. Objectives

The EEW system requires a model to classify the PGA from seismic waves. Furthermore, this model should exhibit critical attributes such as efficient training capabilities, real-time processing, rapid classification, compatibility with multi-channel time series data. This study aims to develop a predictive model for PGA classification using seismic wave data, with improved efficiency in terms of retraining time and memory utilization compared to conventional methods. The proposed model combines elements of the ELM model, replacing the random linear layer with an ESN for feature extraction. This study, which investigates the prediction of strong earthquake motions, a critical aspect of enhancing early warning systems and reducing seismic risks, compares three models: Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and the innovative

Multi Echo-State Extreme Learning Machine (Multi ES-ELM).

3. Methodology

This section is divided into four parts that provide an overview of the data structure, the Multi ES-ELM model structure, the feature extraction method, and the training method. Part 3.1 explains the input data and the labels used to train the proposed model. Part 3.2 describes the structure of the Multi ES-ELM model, which combines ESN and ELM. Part 3.3 provides an overview of the ESN structure and shows how to use the ESN to extract features from seismic waves. Part 3.4 shows how to fit the extracted features to labels using the Extreme Learning Machine.

3.1 Input Data and Labels

The seismic station records ground motions that contain both P-waves and S-waves. The recorded ground motion is time series data that consisting three channels. Figure 2 shows the input data, which is the ground acceleration of the initial P-wave (500 samples), consisting of three axes: NS , EW, and Z (Chiang et al., 2022; Jozinović et al., 2020). Consider the observation signal S_t at time t, which is denoted as $[NS_t, EW_t, Z_t]^\top$. The input to the Multi ES-ELM model is multivariate time series data, represented as a matrix for the i-th input sample.

Figure 2 describes the three motion channels as NS, EW and Z corresponding to north-south, east-west, and up-down directions of ground movement, respectively. The NS and EW channels detect the horizontal displacement of the ground along the north-south and the east-west axis, accordingly. The Z channel records the ground's vertical motion (Chiang et al., 2022). These three channels of motion data are mathematically represented as a matrix for the i-th recorded signal, as shown in Equation 5. The matrix structure is as follows:

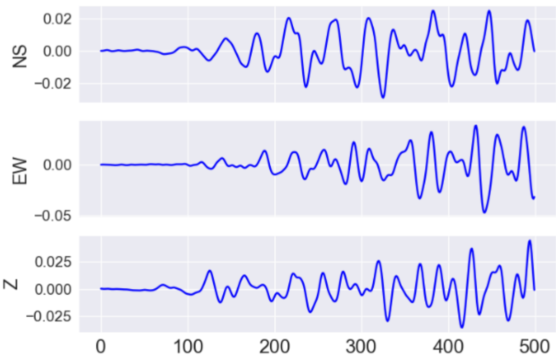


Figure 2 NS, EW, and Z channels of Initial P-waves

$$\mathcal{S}_i = \begin{bmatrix} NS_1 & NS_2 & \cdots & NS_T \\ EW_1 & EW_2 & \cdots & EW_T \\ Z_1 & Z_2 & \cdots & Z_T \end{bmatrix}_i \in \mathbb{R}^{3\times T} \qquad (5)$$

Here, T represents the number of samples in the initial P-wave, and N represents the total number of signal data points in the dataset. To compile the complete dataset, the signals from all seismic recordings are combined into a tensor X, defined as Equation 6.

$$X = \begin{bmatrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \vdots \\ \mathcal{S}_N \end{bmatrix} \in \mathbb{R}^{N\times 3\times T} \qquad (6)$$

In this tensor, N is the total number of signal recordings, and each $\mathcal{S}_i$ corresponds to the ground motion data from a specific seismic event. The label y_n of the data $\mathcal{S}_i$ is the seismic signal's peak ground acceleration (PGA). The classification of each seismic event is based on its PGA value. The labels are encoded using a one-hot encoding scheme. If the PGA of the signal $\mathcal{S}_i$ is less than *80Gal*, then the label is set to $y_i = [1,0]$ (Not warning). Otherwise, if the PGA exceeds this threshold, the label is set to y_i=[0, 1] (Warning) (Chiang et al., 2022). The label matrix for the entire dataset can be represented as Equation 7.

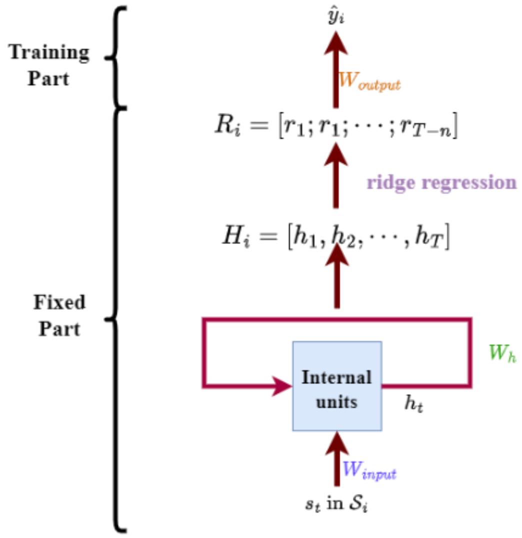
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^{N\times 2} \qquad (7)$$

This matrix Y categorizes seismic events, enabling practical training of machine learning models to detect and predict earthquake warnings.

3.2 Model Architecture

The Multi ES-ELM model consists of two parts. Figure 3 illustrates the training process of the Multi ES-ELM, which is divided into a fixed part and a training part. Figure 4 explains how to construct and use the ESN, which serves as the fixed part of the Multi ES-ELM model. Figure 3 shows how ESN trains and predicts data. The ESN model consists of two components: a fixed part and a training part. The input of ESN is signal data $\mathcal{S}_i$, which includes three channels of seismic waves channels, and the output is $\hat{y}_i$, representing the earthquake warning.

The first part is the ESN layer, which has fixed weight parameters and does not need a training process. The fixed weight of ESN transforms the input signal X into a hidden matrix H. The Multi ES-ELM embeds matrix H and returns an embedded vector R via ridge regression. The second part, W_{output}, is the linear layer used to predict the label from feature data extracted by the first part (Cucchi et al., 2022; Bianchi et al., 2020). This second part is trained using the pseudo-inverse method, as described in subsection 4.5.



$$h_t = tanh(W_{input} s_t + W_h h_{t-1})$$

$$r_t = (h_t h_t^{\top})^{-1} h_t h_{t+n}^{\top}$$

Figure 3 The fixed part and the training part in the ESN model

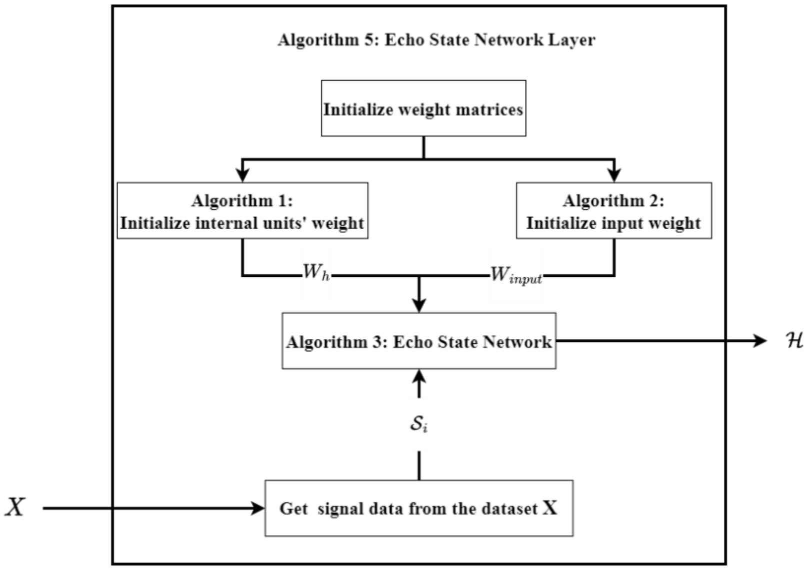


Figure 4 The Echo State Network layer algorithm

3.3 Feature Extraction

Algorithm 1 Internal Units Weight Initialization

Input: Number of internal units λ, spectral radius $r = 0.99$

Output: Initialized internal weights $W_h \in \mathbb{R}^{\lambda\times\lambda}$

```
W_h ← [w_ij] ∈ ℝ^{λ×λ},  w_ij ~ 𝒩(0,1)
E ← eigenvalue(W_h)
e_max ← max(abs(E))
W_h ← (W_h × r)/abs(e_max)
return W_h
```

Algorithm 2 Input Weight Initialization

Input: Number of internal units λ, number of channels data $C = 3$

Output: Initialized input weights $W_{input} \in \mathbb{R}^{\lambda\times C}$

```
W_input ← [w_ij] ∈ ℝ^{λ×C}, w ~ U(0,1)
for i ← 1 to λ do
    for j ← 1 to C do
        if w_ij > 0.5 then
            W_input[i, j] ← 1
        else
            W_input[i, j] ← 0
return W_input
```

Let the matrix W_h be composed of internal recurrent units. To avoid training the encoding of input time series data, the parameters W_h need to follow the conditions of the echo state property. There are many methods to set the matrix W_h to echo state. This research uses the techniques that normalize W_h by its largest absolute eigenvalue. The methods to initialize W_{input} and W_h that are shown in Algorithm 1, 2 respectively (Bianchi et al., 2020; Jaeger, 2002).

The fixed-part ESN model does not need to be trained but can extract the feature used to train the supervised model (Cucchi et al., 2022; Bianchi et al., 2020; Jaeger, 2002). ESN is used to model sequential data and is governed by the RNN states equation (Bianchi et al., 2020) as represented in Equation 8.

$$h_t = \tanh\left(W_{input} s_t + W_h h_{t-1}\right) \tag{8}$$

$H = [h_1, \dots, h_T]^\top$ is the sequence of hidden states h_t generated over time and represents the encoding of signal point s_t in input signal data S. The sequence of hidden states H can be calculated by Algorithm 3.

Algorithm 3 Echo State Network

Input: Signal $\mathcal{S} \in \mathbb{R}^{C\times T}$, input weight $W_{input} \in \mathbb{R}^{\lambda\times C}$, internal weight $W_h \in \mathbb{R}^{\lambda\times\lambda}$

Output: Hidden matrix $H = [h_1; h_2; , \cdots; h_T]$

```
H ← ∅
h_0 ← 0_{λ×1}
for t ← 1 to T do
    s_t ← 𝒮[:, t] ∈ ℝ^{C×1}
    h_t ← tanh(W_input · s_t + W_h · h_{t−1}) ∈ ℝ^{λ×1}
    H[t] ← h_t
    h_{t−1} ← h_t
return H ∈ ℝ^{λ×T}
```

Algorithm 4 Echo State Network Layer

Input: Signal dataset $X = [\mathcal{S}_1; \mathcal{S}_2; \cdots; \mathcal{S}_N] \in \mathbb{R}^{N\times C\times T}$

Output: Set of hidden a matrix $\mathcal{H} = [H_1; H_2; , \cdots; H_N]$

```
ℋ ← ∅
W_h ← Algorithm 2 (λ)
W_input ← Algorithm 3 (λ, C)
for i ← 1 to N do
    ℋ[i] ← Algorithm 4 (X[i], W_input, W_h)
return ℋ ∈ ℝ^{N×λ×T}
```

The ridge embedding of H of the ESN model is the representation of latent space (H) by predicting the next n time steps latent state (h_{t+n}) from the recent latent state (h_t) with auto-regressive models (Bianchi et al., 2020) as represented in Equation 9.

$$r_t = (h_t h_t^\top)^{-1} h_t h_{t+n}^\top \tag{9}$$

Ridge embedding of hidden matrix H_i of input signal data $\mathcal{S}_i$ is defined by

$$R_i = [r_1; r_2; \cdots; r_{T-n}]_i^\top \tag{10}$$

The Algorithm 4 is the Echo State Network layer that can extract the feature from muti-channel time series data like seismic wave data. The Algorithm 5 transforms hidden matrix H_i to ridge embedding R_i. Also, for each ridge embedding R_i of signal $\mathcal{S}_i$ are stacked into the list of ridges embedding $\mathcal{R}$ as represented in Equation 11.

$$\mathcal{R} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix} \tag{11}$$

Algorithm 5 Ridge Embedding

Input: Hidden spaces $H = [h_1; h_2; , \cdots; h_T]$, step size n

Output: Embedded vector R

1: $R \leftarrow \varnothing$
2: **for** $t \leftarrow 1$ **to** $T - n$ **do**
3: $\quad r \leftarrow (h_t h_t^\top)^{-1} h_t h_{t+n}^\top \in \mathbb{R}^{\lambda\times\lambda}$
4: $\quad R[i] \leftarrow flatten(r) \in \mathbb{R}^{1\times(\lambda\times\lambda)}$
5: **return** $R \in \mathbb{R}^{(T-n)\times(\lambda^2)}$

3.4 Training

Conventional artificial neural networks that use forward-backward algorithms require a significant amount of training time. The Extreme Learning Machine (ELM) is a single-layer feed-forward network (SLFN). The input weights of this model are fixed (not necessary to train), and the pseudo-inverse matrix optimizes the output weights of the model. The ELM runs faster than classical ANNs and can maintain high accuracy (Salam et al., 2021). The efficiency of ELM is not only due to its rapid training process but also to its ability to achieve accuracy comparable to classical artificial neural networks (ANNs) (Salam et al., 2021). This speed and reliability make ELM a promising candidate for real-time applications, such as the EEW system, where rapid processing is essential.

The Multi ES-ELM model leverages the pseudo-inverse method from the ELM for its training process. The training method in the Multi ES-ELM model is designed to optimize the output weight $\widehat{\mathcal{B}}$ using the hidden layer representation G, which is derived from the input data matrix A and the weight matrix W as depicted in Equation 12.

$$G\mathcal{B} = V \tag{12}$$

where

$$G = \begin{bmatrix} g(a_1 \cdot w_1 + b_1) & \cdots & g(a_1 \cdot w_l + b_l) \\ \vdots & \ddots & \vdots \\ g(a_N \cdot w_1 + b_N) & \cdots & g(a_N \cdot w_l + b_l) \end{bmatrix}$$

$$\mathcal{B} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \quad and \quad V = \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

$g,\ A = [a_1, \dots a_N]^\top,\ b = [b_1, \dots, b_l]^\top, W = [w_1, \dots, w_l]^\top, V$ and $\mathcal{B}$ are a non-linear function, input data, bias, input weight, label data, and output weight respectively.

In Algorithm 6, the process of finding the optimal output weight $\widehat{\mathrm{B}}$ for the ELM model is outlined. The hidden layer G is a learnable representation that plays a crucial role in this process. The pseudo-inverse of the G matrix, denoted as $\mathrm{G}^\dagger$, is used to compute the optimal output weight $\widehat{\mathrm{B}}$. This method ensures that the ELM model can efficiently learn and adapt to new data (Huang et al., 2006).

Algorithm 6 Extreme Learning Machine (ELM) Training

Input: Input matrix $A \in \mathbb{R}^{N\times d}$, output matrix $V \in \mathbb{R}^{N\times k}$, activation function $g(\cdot)$

Output: Input weight $W \in \mathbb{R}^{d\times l}$, hidden bias $b \in \mathbb{R}^{l\times 1}$, output weight matrix: $\mathcal{B} \in \mathbb{R}^{l\times k}$

1: $W \leftarrow [w_{ij}] \in \mathbb{R}^{d\times l}, w_{ij} \sim U(-1,1)$
2: $b \leftarrow [b_{ij}] \in \mathbb{R}^{l\times 1}, b_{ij} \sim U(-1,1)$
3: $G \leftarrow g(A \cdot W + b^\top)$
4: $G^\dagger \leftarrow (G^\top G)^{-1} G^\top$
5: $\widehat{\mathcal{B}} \leftarrow G^\dagger V$
6: **return** $W, b, \widehat{\mathcal{B}}$

The Multi ES-ELM model, which builds upon the ELM framework, employs a similar approach to determine the optimal weight $\widehat{W}_{output}$. This method allows the model to maintain the advantages of ELM while enhancing its capability to handle the complexity of multi-channel time series data in earthquake early warning systems.

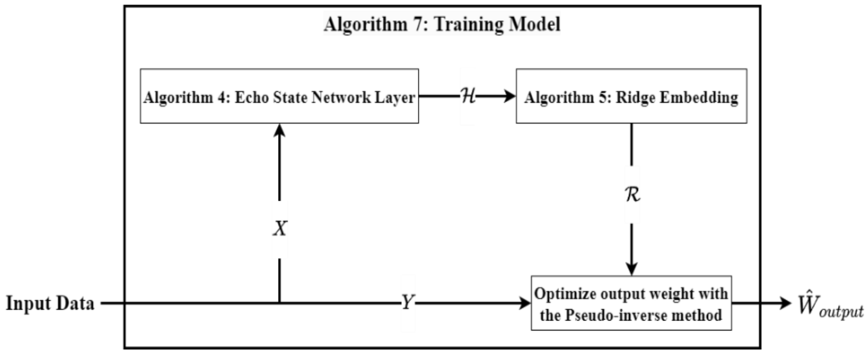


Figure 5 The multi-channel echo state extreme learning machine

The Multi ES-ELM model uses the list of ridges embedding $\mathcal{R}$ to be the learnable representation. The Algorithm 7 shows how to train the Multi ES-ELM model with the input signal data X and label data Y. The algorithm finds the optimal output weight from the list of embedding R.

The Algorithm 6 shows the pseudo-inverse method to calculate the optimal output weight $\widehat{\mathcal{B}}$ of ELM. Figure 5 explains how the Multi ES-ELM model extracts features from data like a list of the hidden matrix H and a list of ridges embedding R that is calculated by the Algorithms 4 and 5, the two features are calculated by the fixed part of the model.

The Multi ES-ELM's ridge embedding R can be compared to the ELM's hidden layer matrix G. Thus, the optimal output weight $\widehat{W}_{output}$. of Multi ES-ELM is calculated by Equation 13.

$$\widehat{W}_{output}=\mathrm{R}^{\dagger}\mathrm{Y}. \tag{13}$$

where $\mathrm{R}^{\dagger}=(\mathrm{R}^{\top}\mathrm{R})^{-1}\mathrm{R}^{\top}$.

Algorithm 7 Multi-channels Echo State Extreme Learning Machine Training

Input: Signal data $X \in \mathbb{R}^{N\times C\times T}$, label data $Y \in \mathbb{R}^{N\times M}$, step size n

Output: Optimal $\widehat{W}_{output}$

1: $\mathcal{R} \leftarrow \varnothing$
2: $\mathcal{H} \leftarrow$ Algorithm 5 $(X,\ n)$
2: **for** $i \leftarrow 1$ **to** N **do**
3: $\quad R \leftarrow$ Algorithm 6 $(\mathcal{H}[i])$
4: $\quad \mathcal{R}[i] \leftarrow flatten(R) \in \mathbb{R}^{1\times(T-n)\lambda^2}$
5: $\mathcal{R}^{\dagger} \leftarrow (\mathcal{R}^{\top}\mathcal{R})^{-1}\mathcal{R}^{\top} \in \mathbb{R}^{(T-n)\lambda^2\times N}$
6: $W_{output} \leftarrow \mathcal{R}^{\dagger}Y \in \mathbb{R}^{(T-n)\lambda^2\times M}$
7: **return** $\widehat{W}_{output}$

4. Results

The study divided into four parts. Part 4.1 explains the seismic data collection from the earthquake networks. Part 4.2 shows the resource consumption while model training to compare RNN, CNN, and Multi-ESELM performances. Part 4.3 illustrates the size of each model, showing a small device implementing capability. Part 4.4 describes the metrics to measure performance to detect large earthquakes that can harm people.

4.1 Data Collection

This study obtains data from two extensive earthquake networks: the Central Weather Bureau Seismic Network (CWBSN) and the Taiwan Strong Motion Instrumentation Program (TSMIP). The CWBSN dataset spans from 2014 to 2019 and includes data from 896 earthquake events. The TSMIP dataset covers a longer period, from 1991 to 2019, and includes data from 724 earthquake events. By integrating these comprehensive datasets, the Multi ES-ELM model is trained and tested on diverse seismic events, enhancing its robustness and accuracy in predicting PGA levels. The initial P-waves from three axes is used in the Multi ES-ELM model. This research utilizes datasets from the Taiwan earthquake station network as described in (Chiang et al., 2022). The datasets comprise N= 10,377 samples of ground motion data, which are divided into training data with $N_{training}$ = 6,900 samples and testing data with $N_{testing}$ = 3,477 samples, where T= 500 (representing 5 seconds after detecting the P-wave).

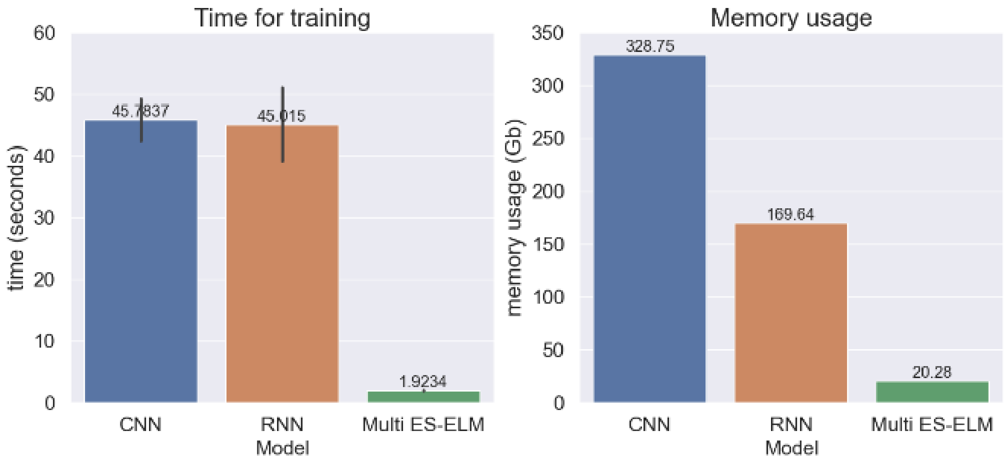


Figure 6 Time and Memory Usages

4.2 Resource Consumptions

Figure 6 compares the training time and memory usage of CNN, RNN, and Multi ES-ELM models. CNN and RNN show similar training times, around 45 seconds, while the Multi ES-ELM is significantly faster at 1.9234 seconds. In terms of memory usage, CNN and RNN consume about 328.75 GB and 169.64 GB, respectively, whereas the Multi ES-ELM requires only 20.28 GB. This efficiency is due to the ESN layer's fixed, randomly generated recurrent structure, eliminating the need for backpropagation. As a result, the Multi ES-ELM offers rapid training and low memory consumption, making it highly practical for real-time earthquake early warning systems.

4.3 Model's Size

In this study, the computational efficiency of the models is a critical factor in determining their suitability for real-time earthquake early warning systems. Table 1 presents a comparison of the number of parameters and the model size in megabytes (MB) for each model: CNN, RNN, and the proposed Multi ES-ELM.

Table 1 Model sizes

Metrics	CNN	RNN	Multi ES-ELM
number of model parameters	638850	17282	882
model size (Mb)	2.437	0.066	0.003

Table 1 shows that the CNN model, with 638,850 parameters and 2.437 MB storage, offers high predictive accuracy but has a large computational footprint, while the RNN model, with 17,282 parameters and 0.066 MB storage, is smaller but exhibits significant performance variability. In contrast, the proposed Multi ES-ELM model uses only 882 parameters and 0.003 MB storage, maintaining competitive accuracy, recall, precision, and F1-score. Its lightweight structure and stable performance make it ideal for real-time earthquake early warning systems, where efficiency and reliability are critical.

4.4 Performance Metric Results

This study conducts 40 trials to assess the performance and stability of the models ensuring a comprehensive evaluation. The mean of all trial results represents performance, while the standard deviation indicates stability. Table 2 summarizes the evaluation metrics for each model. Each cell shows the mean and standard deviation (SD) across 40 trials, offering insights into both average performance and consistency. These results help assess the models' robustness and reliability, ensuring a thorough evaluation process.

The CNN achieved high overall performance, with an accuracy and F1-score of 94.65 ± 0.30 and 89.95 ± 0.41, respectively, indicating stable results. However, its recall (92.84 ± 2.36) and precision (87.34 ± 2.35) showed slightly higher variability. The RNN performed worse, with lower metrics and larger standard deviations, highlighting instability. In contrast, the Multi ES-ELM achieved high performance with very low standard deviations, demonstrating stable results. Notably, its recall score outperformed the other models.

Table 2 Evaluation Metrics Results

Metrics	CNN	RNN	Multi ES-ELM
accuracy	94.65 ± 0.30	84.83 ± 19.40	93.46 ± 0.22
recall	92.84 ± 2.36	84.93 ± 13.34	96.50 ± 0.52
precision	87.34 ± 2.35	74.18 ± 18.15	81.53 ± 0.53
f1 score	89.95 ± 0.41	77.80 ± 16.40	88.38 ± 0.37

Note: Results are presented as mean ± standard deviation

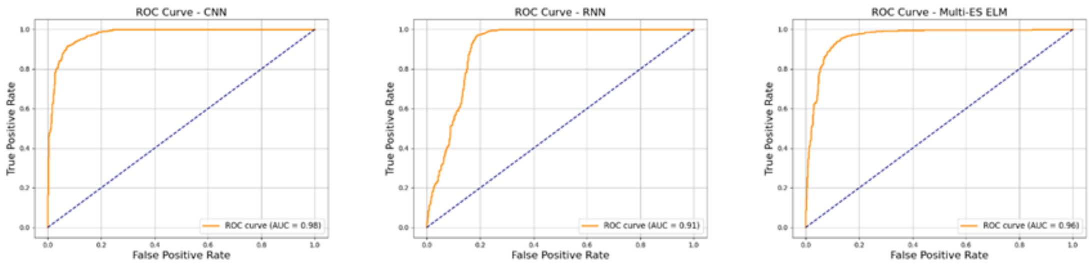


Figure 7 ROC Curves of CNN, RNN, and Multi ES-ELM Models for Strong Motion Prediction

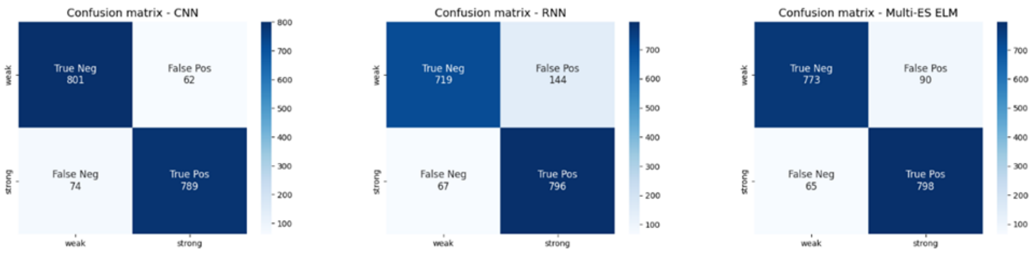


Figure 8 Confusion Matrices of CNN, RNN, and Multi ES-ELM Models for Strong Motion Classification

Figure 7 presents the ROC curves for the CNN, RNN, and Multi ES-ELM models, comparing their classification performance. The CNN achieved the highest AUC of 0.98, indicating an excellent discriminative ability. The Multi ES-ELM followed with a strong AUC of 0.96, whereas the RNN showed lower performance with an AUC of 0.91. These results highlight that both the CNN and Multi ES-ELM can effectively distinguish between classes, with the Multi ES-ELM maintaining competitive performance despite its lightweight structure.

Figure 8 shows the confusion matrices for the CNN, RNN, and Multi ES-ELM models. The CNN achieved balanced true positive (789) and true negative (801) rates, with relatively few false predictions. The RNN, despite achieving a high true positive count (796), suffered from a larger number of false positives (144). In contrast, the Multi ES-ELM maintained strong classification performance, achieving 798 true positives and 773 true negatives while keeping both false positives (90) and false negatives (65) lower than the RNN. These results confirm the Multi ES-ELM's ability to provide reliable predictions with fewer misclassifications.

5. Discussion

This study provides a comprehensive evaluation of the resource demands associated with each model by comparing training time, and memory consumption, model size, and parameter count. The insights gained from these comparisons can inform the selection of the most suitable model for earthquake early warning systems that require fast, accurate, and efficient processing in real-time environments. This section interprets and discusses the experimental results and potential applications of Multi ES-ELM model. Part 5.1 examines the resource consumption of each model, which may pose constraint to implementation in the EEW system. Part 5.2 addresses the model’s size, which affects the capability to be deployed in various devices. Part 5.3 analyzes the accuracy of the models in predicting strong motion.

5.1 Resource Efficiency

Efficient training time is critical for real-time systems, where models must quickly adapt to new data without compromising performance. The Multi ES-ELM outperforms other models in this aspect; as its ESN component extracts features from time series data without a training process, reducing processing time. In contrast, RNNs and CNNs require time-consuming feature extraction through gradient descent, which demands significantly more time and computational resources. Memory usage, another critical metric, reflects the computational resources required during training. Minimizing memory usage is crucial for real-time earthquake early warning systems operating in resource-constrained environments. The Multi ES-ELM, using the ELM method, predicts strong motions with a single short training session, consuming minimal computational resources. In contrast, RNNs and CNNs require significantly more memory due to repeated gradient descent iterations.

5.2 Model's Size

The number of parameters in a model refers to the total number of weights and biases that the model must learn during the training process. The number of parameters directly indicates the model's complexity, as more parameters generally enable the model to capture more intricate patterns in the data. However, a higher parameter count also increases the computational costs, requiring more memory and processing power. Additionally, it affects the model size, as larger models require more storage space and take longer to train and deploy. The Multi ES-ELM requires minimal memories to implement. Its compact size allows deployment on small devices, such as smartphones and embedded systems, making it well-suited for EEW applications. However, other models require substantially more memory to implement.

5.3 Metrics Analysis Results

The EEW systems aim to accurately predict strong ground motions, a task formulated as a binary classification problem. Misclassifications false positives (predicting strong motion when none occurs) and false negatives (failing to predict strong motion) can lead to significant losses, such as economic disruptions, reduced public trust, or threats to safety. Consequently, system performance is critical, requiring a model that balances high recall to minimize missed warnings and high precision to reduce false alarms. Metrics like F1-score and accuracy are essential for evaluating and improving the model to ensure reliable and effective warnings, minimizing both risks and unnecessary costs.

The Multi ES-ELM excels in detecting hazardous seismic waves more effectively than conventional models. However, its focus on maximizing the detection of hazardous waves comes at the cost of reduced sensitivity, resulting in slightly lower performance in certain metrics compared to CNN models. While CNN achieves better precision and F1-score, the Multi ES-ELM outperforms it in recall and maintains competitive accuracy. The high recall of the Multi ES-ELM underscores its effectiveness in identifying actual earthquake events, a critical aspect for the reliability and timeliness of early warning systems.

6. Conclusion

CNN demonstrates solid overall performance with high accuracy, balanced precision, and recall. However, training requires significant computational resources. RNN shows potential but suffers from high variability and lower overall performance, likely due to its complex training requirements and sensitivity to hyperparameters. Multi ES-ELM offers balanced and consistent performance, and its resource efficiency is a significant advantage. The CNN model emerged as the most accurate, while the RNN model exhibited inconsistent results. Notably, with its competitive performance and resource efficiency, the Multi ES-ELM model not only presents a promising avenue for real-time earthquake prediction but also offers a practical solution for seismic risk mitigation strategies, thereby instilling confidence in its potential real-world applications.

In conclusion, while CNN provides the best overall performance in precision and F1-score, the proposed Multi ES-ELM stands out for its high recall and resource efficiency, making it a highly viable option for real-time earthquake prediction systems. Future work could focus on enhancing the precision of Multi ES-ELM and further optimizing its architecture for even better performance.

7. CRediT Statement

Phiphat Chomchit: Methodology, Software, Formal Analysis, Data curation, Writing original draft, Visualization.
Somrawee Aramkul: Validation, Writing, review & editing,Visualization
Yuthapong Somchit: Validation, Formal analysis, Writing, review & editing.

Paskorn Champrasert: Conceptualization, Methodology, Validation, Investigation, Resources, Writing, review & editing, Supervision, Project Administration.

8. References

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

Bianchi, F. M., Scardapane, S., Løkse, S., & Jenssen, R. (2020). Reservoir computing approaches for representation and classification of multivariate time series. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(5), 2169-2179. https://doi.org/10.1109/TNNLS.2020.3001377

Carvalho, H. S., Shams, F., Ferrari, R., & Boccato, L. (2018). Application of extreme learning machines and echo state networks to seismic multiple removal [Conference presentation]. *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil. https://doi.org/10.1109/IJCNN.2018.8489620

Chandrakumar, C., Prasanna, R., Stephens, M., & Tan, M. L. (2022). Earthquake early warning systems based on low-cost ground motion sensors: A systematic literature review. *Frontiers in Sensors*, *3*, Article 1020202. https://doi.org/10.3389/fsens.2022.1020202

Chiang, Y. J., Chin, T. L., & Chen, D. Y. (2022). Neural network-based strong motion prediction for on-site earthquake early warning. *Sensors*, *22*(3), Article 704. https://doi.org/10.3390/s22030704

Cremen, G., & Galasso, C. (2020). Earthquake early warning: Recent advances and perspectives. *Earth-Science Reviews*, *205*, Article 103184. https://doi.org/10.1016/j.earscirev.2020.103184

Cucchi, M., Abreu, S., Ciccone, G., Brunner, D., & Kleemann, H. (2022). Hands-on reservoir computing: A tutorial for practical implementation. *Neuromorphic Computing and Engineering*, *2*(3), Article 032002. https://doi.org/10.1088/2634-4386/ac7db7

Gallicchio, C., Micheli, A., & Pedrelli, L. (2017). Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, *268*, 87-99. https://doi.org/10.1016/j.neucom.2016.12.089

Gasparini, P., Manfredi, G., & Zschau, J. (Eds.). (2014). *Earthquake early warning systems*. Springer Science & Business Media.

Hsiao, N. C., Wu, Y. M., Shin, T. C., Zhao, L., & Teng, T. L. (2009). Development of earthquake early warning system in Taiwan. *Geophysical Research Letters*, *36*(5), Article L05002. https://doi.org/10.1029/2008GL036596

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1-3), 489-501. https://doi.org/10.1016/j.neucom.2005.12.126

Jaeger, H. (2002). Adaptive nonlinear system identification with echo state networks. *Advances in Neural Information Processing Systems*, *15*, 609-616.

Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2020). Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. *Geophysical Journal International*, *222*(2), 1379-1389. https://doi.org/10.1093/gji/ggaa233

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

Li, Z., Meier, M. A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*(10), 4773-4779. https://doi.org/10.1029/2018GL077870

Meilano, I., Susilo, S., & Sarsito, D. (2022). Rapid magnitude estimation of The August 5, 2018, Lombok earthquake using high-rate GNSS Data. *Geomate Journal*, *23*(98), 57-65. https://doi.org/10.21660/2022.98.3438

Meng, F., Ren, T., Liu, Z., & Zhong, Z. (2023). Toward earthquake early warning: A convolutional neural network for rapid earthquake magnitude estimation. *Artificial Intelligence in Geosciences*, *4*, 39-46. https://doi.org/10.1016/j.aiig.2023.03.001

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), Article e2019GL085976. https://doi.org/10.1029/2019GL085976

Norisugi, R., Kaneko, Y., & Rouet-Leduc, B. (2024). Machine learning predicts earthquakes in the continuum model of a rate-and-state fault with frictional heterogeneities. *Geophysical Research Letters*, *51*(9), Article e2024GL108655. https://doi.org/10.1029/2024GL108655

Salam, M. A., Ibrahim, L., & Abdelminaam, D. S. (2021). Earthquake prediction using hybrid machine learning techniques. *International Journal of Advanced Computer Science and Applications*, *12*(5), 654-665. https://doi.org/10.14569/IJACSA.2021.0120578

Soltani, R., Benmohamed, E., & Ltifi, H. (2023). Echo state network optimization: A systematic literature review. *Neural Processing Letters*, *55*(8), 10251-10285. https://doi.org/10.1007/s11063-023-11326-w

Sun, C., Song, M., Cai, D., Zhang, B., Hong, S., & Li, H. (2024). A systematic review of echo state networks from design to application. *IEEE Transactions on Artificial Intelligence*, *5*(1), 23-37. https://doi.org/10.1109/TAI.2022.3225780

Wu, A., Lee, J., Khan, I., & Kwon, Y.-W. (2021). Crowdquake+: Data-driven earthquake early warning via IoT and deep learning [Conference presentation]. *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA. https://doi.org/10.1109/BigData52589.2021.9671971

Yanwei, W., Xiaojun, L., Zifa, W., Jianping, S., & Enhe, B. (2021). Deep learning for P-wave arrival picking in earthquake early warning. *Earthquake Engineering and Engineering Vibration*, *20*(2), 391-402. https://doi.org/10.1007/s11803-021-2027-6