Earth and Space Science

RESEARCH ARTICLE
10.1029/2023EA003363

Data-Knowledge Driven Hybrid Deep Learning for Earthquake Early Warning

J. Zhu[1,2], S. Li[1,2], and J. Song[1,2]

[1]Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin, China, [2]Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Management, Harbin, China

Key Points:

- A Data-knowledge driven Hybrid deep Learning network (DHLnet) is proposed for earthquake early warning (EEW)
- The DHLnet mainly consists of the knowledge embedding, convolutional neural network and graph convolutional network
- We demonstrate that DHLnet outperforms knowledge-driven EEW methods and data-driven deep learning models

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Song,
jdsong@iem.ac.cn

Citation:

Zhu, J., Li, S., & Song, J. (2024). Data-knowledge driven hybrid deep learning for earthquake early warning. *Earth and Space Science*, *11*, e2023EA003363. https://doi.org/10.1029/2023EA003363

Received 18 OCT 2023
Accepted 12 JUL 2024

Author Contributions:

Conceptualization: J. Zhu, S. Li, J. Song
Formal analysis: J. Zhu, S. Li, J. Song
Funding acquisition: J. Song
Investigation: J. Zhu, S. Li, J. Song
Methodology: J. Zhu, J. Song
Software: J. Zhu
Supervision: J. Zhu, S. Li
Validation: J. Zhu
Visualization: J. Zhu
Writing – original draft: J. Zhu

Abstract Earthquake early warning (EEW) is of great significance in mitigating seismic disasters. Traditional EEW algorithms, which are knowledge-driven approaches, rely on seismologists' analysis. The limited intensity measures were extracted by seismologists from P-wave signals. And there is considerable uncertainty for predicting epicentral distance, magnitude, peak ground acceleration (PGA), and peak ground velocity (PGV). Currently, data-driven deep learning methods with the strong learning abilities do not consider knowledge information from seismologists in EEW; thus, there is unexplored potential in enhancing the performance of deep learning models for EEW. Here, we construct the Data-knowledge driven Hybrid deep Learning network (DHLnet) for EEW using the waveform input, knowledge embedding, convolutional neural network and graph convolutional network, aiming to integrate knowledge information from knowledge-driven methods and the strong learning ability of data-driven deep learning methods, that is, improving the performance of EEW. For the same test data set, compared with knowledge-driven methods and data-driven deep learning models, we demonstrate that DHLnet enhances the timeliness and robustness in predicting the epicentral distance, magnitude, PGA, and PGV during 10 s time window following the arrival of P-wave. Furthermore, to validate the generalization and robustness of the DHLnet in EEW, we applied the trained DHLnet to an independent data set, within first few seconds after an earthquake occurs, DHLnet can provide robust magnitude estimation, epicentral distance estimation and high alarm accuracy. The potential of the proposed network is to enhance the performance of EEW systems and provides new insights into the exploration of deep learning methods for EEW domain.

Plain Language Summary Earthquake early warning (EEW) relies heavily on crucial parameters like epicentral distance, magnitude, and peak ground motion (peak ground acceleration [PGA] and velocity [PGV]). To quickly and accurately determine these parameters, traditional EEW algorithms, which are knowledge-driven approaches, rely on seismologists' analysis based on earthquake rupture physics, and establish empirical EEW parameter prediction equations. Currently, data-driven deep learning models with strong learning ability in EEW are mainly used to extract features from raw seismic waveforms, and do not take existing knowledge information from traditional EEW algorithms into account. Therefore, there is underutilized potential in improving the generalization, reliability and interpretability of deep learning model for EEW. Here, a Data-knowledge driven Hybrid deep Learning network (DHLnet) for EEW is proposed and demonstrate that compared with knowledge-driven methods and data-driven deep learning models, DHLnet has better performance on predicting epicentral distance, magnitude, PGA and PGV.

1. Introduction

In recent years, the development of earthquake early warning (EEW) has proven to be a highly efficacious approach to mitigate earthquake disasters (Allen & Melgar, 2019; Allen & Stogaitis, 2022). EEW information is released prior to the arrival of destructive seismic waves in the surrounding areas of the epicenter, which includes magnitude, ground motion intensity, and other relevant parameters. This advance notice provides businesses and the public with crucial seconds to tens of seconds to take necessary earthquake emergency measures, thereby reducing potential casualties and property losses (Allen & Kanamori, 2003; Cremen et al., 2022; Strauss & Allen, 2016). EEW systems also offer safety assurance services for major infrastructure projects. By receiving alert messages from the EEW system, critical facilities such as city gas and power supply systems, nuclear power plants, reservoirs and dams, large substations, oil and gas pipelines, and high-speed railways can automatically activate emergency response systems, such as braking and shutdown, in accordance with established procedures.

This helps mitigate both direct earthquake disasters and secondary hazards (McBride et al., 2022; Papadopoulos et al., 2023). Quickly and robustly predicting EEW parameters, including epicentral distance, magnitude, peak ground acceleration (PGA) and velocity (PGV), is essential for disseminating warning information and determining potential earthquake damage (Allen et al., 2009; Caruso et al., 2017; Colombelli et al., 2015; Jozinović et al., 2022; Odaka et al., 2003; Wald et al., 1999; Zollo et al., 2023).

Traditional EEW algorithms, which are knowledge-driven approaches, mainly rely on seismologists to analyze historical earthquake events from the perspective of earthquake rupture physics, identify the intensity measures (IMs) related to EEW parameters based on P-wave signals, and establish empirical EEW parameter prediction equations (Colombelli et al., 2014; Festa et al., 2008; Kanamori, 2005; Odaka et al., 2003; Olson & Allen, 2005; Wu & Zhao, 2006), which have provided a technical foundation for the existing EEW systems in many countries and regions (Alcik et al., 2009; Allen & Melgar, 2019; Kohler et al., 2020; Peng et al., 2020). Odaka et al. (2003) analyzed seismic data from Japan's Kyoshin Net (K-NET) and found that the slope of the beginning acceleration of P-wave decreases as the distance from the epicenter increases. They established the epicentral distance prediction equation according to the correlation with the slope of the beginning seismic acceleration. Wu and Zhao (2006) analyzed seismic data collected in southern California and found that the peak displacement (P_d) of seismic wave contains information about the early stages of the seismic rupture process (Colombelli et al., 2014). Meanwhile, P_d is used to estimate magnitude, attenuates with increasing distance. According to a study conducted by Festa et al. (2008), it was demonstrated that the squared velocity integral (IV2) encompasses valuable data concerning the energy emitted by the rupture as it progresses along the fault plane. Consequently, this technique offers a direct means of gaining profound understanding regarding the physical attributes of the fracture. Moreover, they determined that at 3 s time window of P-wave, there is a certain correlation between IV2 and the earthquake magnitude. Furthermore, the peak ground motion is predicted by establishing the prediction equation based on IV2 (Brondi et al., 2015). Wu and Kanamori (2005) examined seismic activities occurring in Taiwan and concluded that the P_d observed during the early stage of the P-wave contains valuable insights regarding both the magnitude of the earthquake and the intensity of the resulting ground motion. However, the limited IMs extracted by seismologists from P-wave are insufficient to estimate EEW parameters, and the predicted EEW parameters are prone to substantial uncertainty (Brondi et al., 2015; Festa et al., 2008; Kanamori, 2005; Kong & Zhao, 2012; Odaka et al., 2003; Olson & Allen, 2005). Meanwhile, improving the timeliness and accuracy for predicting EEW parameters remains a problem that need to be explored by seismologists.

With the advancement of computer science, artificial intelligence technologies have been extensively applied in various domains. Deep learning methods have also been introduced into seismology (Bergen et al., 2019; Jiao & Alavi, 2020; Kong et al., 2019). Currently, data-driven deep learning models with strong learning ability are mainly used to extract features from raw seismic waveforms for source parameter characterization and earthquake monitoring. Meanwhile, most studies focused on earthquakes with magnitudes below 5, and the waveform length used as input for deep learning models is usually longer and contains S-wave information (Chakraborty et al., 2022; Mousavi & Beroza, 2020a, 2020b, 2022; Münchmeyer et al., 2021a; Saad et al., 2022; Zhang et al., 2021, 2022). Some researchers proposed a convolutional neural network (CNN) for estimating epicentral distance, and the model input was 1-min seismic waveforms, which include S-wave information (Mousavi & Beroza, 2020b). Mousavi and Beroza (2020a) proposed a magnitude estimation network according to CNN and recurrent neural network (RNN). The input to the model was a 30 s seismic waveform that contained S-wave information. Chakraborty et al. (2022) fine-tuned the magnitude estimation model proposed by Mousavi and Beroza (2020a) and used 2-s P-wave signals as the model input; this model was applied mainly for EEW magnitude estimation of earthquake events with magnitudes below 5, and the magnitude estimation results were somewhat uncertain. Some researchers proposed magnitude estimation model to predict the magnitude of earthquakes with magnitudes below 4.5 based on the vision transformer network and raw seismic waveforms (Saad et al., 2022). The applicability of data-driven deep learning models to predict EEW parameters require further investigation. Prior knowledge information based on earthquake rupture physics shows that some IMs from P-wave signal are related to EEW parameters (Colombelli et al., 2014; Kanamori, 2005; Olson & Allen, 2005). Notably, existing purely data-driven deep learning methods do not incorporate knowledge information in EEW (Bergen et al., 2019; Chakraborty et al., 2022; Münchmeyer et al., 2021a; van den Ende & Ampuero, 2020; Zhang et al., 2021, 2022); thus, there is underutilized potential in improving the generalization, robustness and interpretability of deep learning models for EEW. Moreover, for EEW, some studies have attempted to directly use expert-defined IMs as inputs for machine learning and deep learning models to predict

the magnitude, PGA and PGV (Song et al., 2022, 2023; Zhu et al., 2021, 2022a, 2022b). Nevertheless, the disregard for latent attributes that seismologists have not derived from P-waves and the capability of deep learning models to directly derive profound attributes from unprocessed data are overlooked by this methodology. Additionally, in practical EEW scenarios, these models depend on the epicentral distance information obtained from the current EEW system (Song et al., 2022, 2023; Zhu et al., 2021, 2022a).

Here, we propose a Data-knowledge driven Hybrid deep Learning network (DHLnet) for EEW using the waveform input, knowledge embedding, CNN and graph convolutional network (GCN), aiming to integrate knowledge information from knowledge-driven EEW methods and the strong learning ability of data-driven deep learning methods. The proposed DHLnet can predict epicentral distance, magnitude, PGA and PGV using single-station seismic data in EEW. We demonstrate that DHLnet models have better performance than knowledge-driven EEW methods and purely data-driven deep learning models. Meanwhile, DHLnet can improve the timeliness and robustness for predicting the epicentral distance, magnitude, PGA and PGV in EEW.

2. Materials and Methods

2.1. Data

Here, we obtained strong ground motion data from 2007 to 2016 recorded by Kyoshin Net (K-NET). These records were acquired from the National Research Institute for Earth Science and Disaster Prevention, Japan, via their website. Our focus was on collecting data related to inland earthquake events in Japan with magnitudes ranging from $7.9 \geq M_{JMA} \geq 3$ and focal depths less than 10 km (Huang et al., 2015). Then, we employed the algorithm proposed by Allen (1978) to automatically detect the arrival of P-waves. This detection was based on unfiltered vertical acceleration records. For our study, we considered only the data where the P-wave arrival was successfully detected. Any data lacking this detection were deemed invalid and subsequently removed from our analysis. Overall, our data set encompasses a total of 54,201 valid three-component strong ground motion records. These records correspond to 1,750 distinct earthquakes. Meanwhile, we randomly divide the data set into training and test data sets based on earthquake events. The training data set includes 1,380 earthquake events, consisting of 42,162 three-component strong motion records (approximately 80% of the total number of strong motion records) (see Table S1); The test data set includes 370 earthquake events, which are composed of 12,039 three-component strong motion records (approximately 20% of the total strong motion records) (see Table S2). To provide a visual representation, Figures 1a and 1b display the spatial distribution of the epicenters and K-NET stations in the training data set, respectively; Figures 1c and 1d respectively show the spatial distribution of the epicenters and K-NET stations in the test data set.

We integrate the acceleration records of different components to obtain the velocity records of different components. Afterward, we acquire displacement records for each component by integrating the velocity records. Next, we employ a high-pass Butterworth filter having four poles and a cutoff frequency of 0.075 Hz to filter the acceleration, velocity, and displacement records (Peng et al., 2017). In this study, we employed three-component acceleration records, three-component velocity records, and three-component displacement records as input data for the DHLnet. The distributions of the magnitude, PGA, and PGV at various epicentral distances are depicted in Figures 2a–2c for both the training and test data sets.

2.2. Knowledge Base

Seismologists have identified many important intensity measures denoted as IMs in P-wave signals through the statistical analysis of many historical earthquake events from the perspective of earthquake rupture physics (Fayaz & Galasso, 2022). These IMs are meaningful for predicting the EEW parameters. This has also provided a technical foundation for the existing EEW systems in many countries and regions (Carranza et al., 2013; Fayaz & Galasso, 2024; Hsiao et al., 2009; Kamigaichi et al., 2009; Kohler et al., 2020; Peng et al., 2020). As these IMs were defined by seismologists in previous research, in this paper, these IMs extracted by seismologists from the perspective of earthquake rupture physics are knowledge information, and the knowledge information is consolidated into a knowledge base.

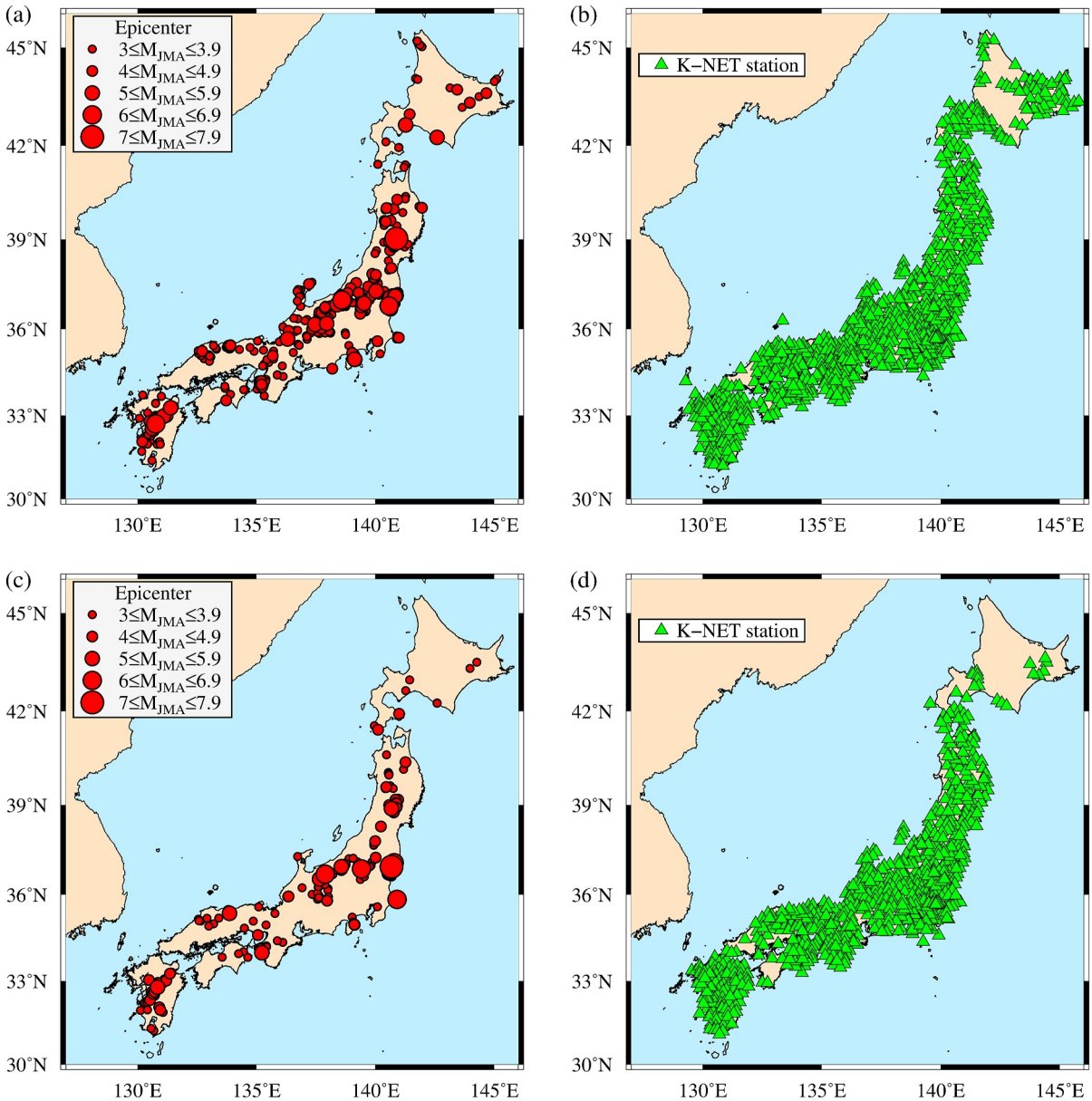

Figure 1. Map of earthquake events. (a) Distribution of epicenters of events for the training data set. (b) Distribution of K-NET stations for the training data set. (c) Distribution of epicenters of events for the test data set. (d) Distribution of K-NET stations for the test data set. Different magnitudes of earthquake events are represented by red solid circles of varying sizes.

2.2.1. Knowledge Base for Epicentral Distance Estimation

We divide the knowledge information used for predicting epicentral distance into three categories: IMs correspond to the slope of the initial P-wave amplitude, IMs correspond to the ground motion amplitude, and IMs correspond to the ground motion frequency. There are 8 IMs correspond to the slope of the initial P-wave amplitude: the B value, C value, $I_{A\text{-}C}$ value, CAV_C value, $IV2_C$ value, AAS_C value, AVS_C value, and ADS_C value. Some researchers proposed the B-Δ method and defined the B value (Odaka et al., 2003). Based on acceleration records, Odaka et al. (2003) presented a straightforward equation represented as $Bt \bullet \exp(-At)$, whereby they fitted this equation to the initial segment of the logarithmic waveform's envelope. They determined the value of B using the method of least squares. Their findings unveiled an inverse relationship between the logarithm of B and the logarithm of Δ. Here, Δ denotes the epicentral distance, while the B value signifies the slope of the initial P-wave acceleration. Yamamoto et al. (2012) developed a C-Δ method and defined the C value. This method is a

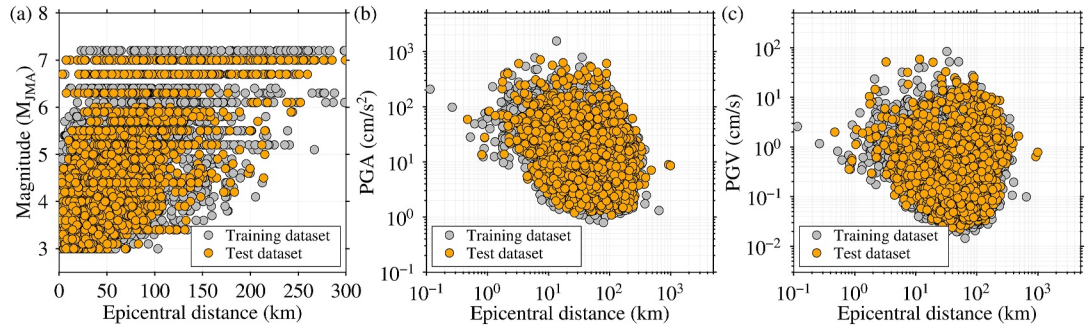

Figure 2. The data sets for training and test are characterized by their distributions. (a) Different epicentral distances exhibit varying magnitudes distributions. (b) PGA distributions differ for distinct epicentral distances. (c) Various epicentral distances show distinct PGV distributions. The training data set is represented by gray circles, while the test data set is represented by orange circles.

simplified version of the B-Δ method, where the Bt•exp(−At) function is simplified to a Ct form. Their research demonstrated that the C-Δ method maintains accuracy while shortening the time window after P-wave arrival. According to Yamamoto et al. (2012), we considered ground motion parameters that rapidly increase in the initial phase (Arias, 1970; Festa et al., 2008; Reed & Kassawara, 1990; Song et al., 2023), including the Arias intensity (I_A), absolute acceleration summary (AAS), absolute velocity summary (AVS), and absolute displacement summary (ADS), cumulative absolute velocity (CAV), squared velocity integral (IV2). Then, we separately fit the Ct form of the function to the initial portion of the time history for each ground motion parameter, thereby obtaining the growth rate of each ground motion parameter in the initial stage. We defined the initial slope of the I_A time history as the $I_{A\text{-}C}$ value, the initial slope of the CAV time history as the CAV_C value, the initial slope of the IV2 time history as the $IV2_C$ value, the initial slope of the AAS time history as the AAS_C value, the initial slope of the AVS time history as the AVS_C value, and the initial slope of the ADS time history as the ADS_C value. There are three IMs correspond to the ground motion amplitude (Kong & Zhao, 2012; Lockman & Allen, 2005): the peak displacement (P_d), peak acceleration (P_a) and peak velocity (P_v). Additionally, there are three IMs correspond to the ground motion frequency (Böse, 2006; Huang et al., 2015; Kanamori, 2005; Kong & Zhao, 2012; Lockman & Allen, 2005): product parameter (TP), average period (τ_c), and peak ratio (T_{va}).

2.2.2. Knowledge Base for Magnitude Estimation

We divide the knowledge information used for magnitude estimation into four categories: IMs correspond to the amplitude, IMs correspond to the energy, distance feature, and IMs correspond to the frequency. There are three IMs correspond to the amplitude: P_d (Wu & Zhao, 2006), P_v, and P_a (Zollo et al., 2023). Some researchers have put forth the P_d and discovered that the P_d value post the P-wave arrival carries crucial details regarding the initial phases of the earthquake rupture procedure (Colombelli et al., 2014; Wu & Zhao, 2006). Additionally, they observed that P_d tends to diminish with growing distance as it functions in relation to magnitude. Several investigations have also employed P_v and P_a values following the P-wave arrival to estimate the magnitude in EEW (Zollo et al., 2023). There are six IMs correspond to the energy: I_A (Arias, 1970), IV2 (Festa et al., 2008), CAV (Reed & Kassawara, 1990), AAS, AVS, and ADS (Song et al., 2023). According to a study conducted by Festa et al. (2008), it was discovered that the calculation of IV2 value using P-wave signals contains energy data emitted by the progressing rupture along the fault plane. Consequently, this enables us to gain direct understanding about the physical properties of the fracture. Moreover, there is a certain correlation between IV2 and the magnitude. Considering that IMs correspond to both amplitude and energy diminish as distance increases in relation to magnitude (Festa et al., 2008; Zollo et al., 2006), this work considers the epicentral distance as a distance feature. There are three IMs correspond to the frequency: τ_c (Böse, 2006), T_{va} (Kanamori, 2005), and TP (Huang et al., 2015).

2.2.3. Knowledge Base for Peak Ground Motion Prediction

We divide the knowledge information used for peak ground motion (PGA and PGV) prediction into three categories: IMs correspond to the amplitude, IMs correspond to the energy, and distance feature. There are three IMs correspond to the amplitude (Wu & Kanamori, 2005): P_d, P_v, and Pa. There are six IMs correspond to the energy:

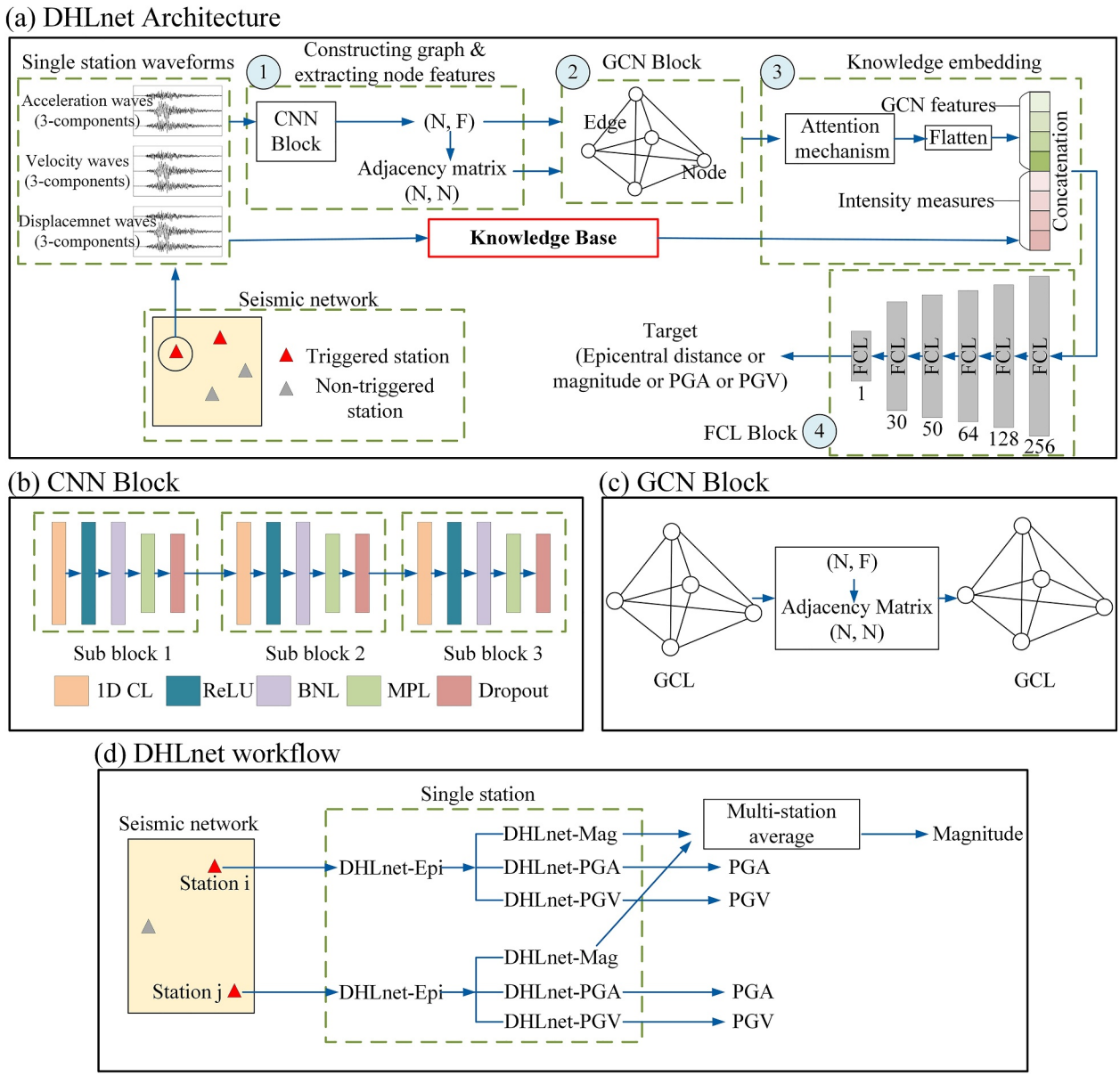

Figure 3. DHLnet for predicting EEW parameter. (a) DHLnet architecture. (b) Detailed structure of the CNN block. (c) Detailed structure of the GCN block. (d) Workflow of the DHLnet when an earthquake occurs.

IV2 (Brondi et al., 2015), CAV (Reed & Kassawara, 1990), I_A (Arias, 1970), AAS, AVS, and ADS (Song et al., 2023). Similarly, the epicentral distance is considered a distance feature.

Due to the detailed description of relevant IMs in the corresponding references, we do not elaborate on each IM here. The calculation formulas and introduction for the relevant IMs are provided in Text S1 in Supporting Information S1. The Text S1 in Supporting Information S1 includes the calculation formulas and introduction to the relevant IMs.

2.3. DHLnet Architecture and Training Procedure

The DHLnet architecture proposed in this work consists of four main parts: graph construction and node feature extraction, graph convolutional network (GCN) block, knowledge embedding, and fully connected layer (FCL) block. These parts run in sequence, and we describe the detailed architecture of DHLnet in Figure 3a. When convolutional neural networks (CNNs) process data, only local relationships in the data are considered, and global

structural information is neglected (LeCun et al., 2015). The multiscale information can be fused by inputting the features extracted from the raw data by the CNN into the GCN. During the graph convolution operation, the GCN aggregates node information, enabling the fusion of features from both local and global nodes; this approach can better capture the relationships and global information between nodes and enhance the feature expression capability of the model (Kipf & Welling, 2016).

The input data were fed into the CNN block to generate nodes and node features for the graph. The CNN block consists of three subblocks (Figure 3b), each of which is made up of a 1D convolutional layer (CL) (Kiranyaz et al., 2015), batch normalization layer (BNL) (Ioffe & Szegedy, 2015), max pooling layer (MPL) (Nagi et al., 2011), rectified linear unit (ReLU) activation function layer (Nair & Hinton, 2010), and dropout layer (Srivastava et al., 2014). The processing of time series data involves the utilization of the 1D CL. The output of the 1D CL is the local features obtained from the sliding window convolution with a one-dimensional convolutional kernel (filter) in different time windows. The output is represented as (timesteps, filters), where the timesteps represent the number of time steps contained in each sample and the filters represent the number of feature maps output by the 1D CL. These feature maps are representations of the extracted features by the CL. The number of feature maps in each 1D CL is 32, 64, and 128. The kernel size for each 1D CL is 4, with a stride of 2. The pooling size for each MPL is 2, with a stride of 2. The dropout layer has a drop rate of 0.2. Meanwhile, based on the seismic signals recorded by a single station, we construct an input data structure of size (L, 9), where 9 represents the three-component acceleration waveforms, three-component velocity waveforms, and three-component displacement waveforms, and L represents the length of the input data. These input data are then fed into the first part, the CNN block, which generates the node feature matrix (NFM). Table S3 in Supporting Information S1 presents the detailed settings of the CNN block. After a series of operations, including the 1D CL, ReLU activation layer, BNL, 1D MPL, and dropout layer, the CNN block outputs an NFM with dimensions (N, F). Here, the graph contains N nodes and each node consists of F node features. To construct a complete graph, in addition to the NFM, we also need to generate edges, that is, the so-called adjacency matrix. The connections among the graph nodes can be depicted using an adjacency matrix. In order to generate this matrix for the given nodes, we follow a procedure involving normalization, centralization, and correlation calculation using the NFM.

$$\text{mean} = \frac{1}{N}\sum_{i=1}^{N} F_i \tag{1}$$

$$\text{std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - \text{mean})^2} \tag{2}$$

$$x_{\text{centered}} = M_F^N - \text{mean} \tag{3}$$

$$\text{Cov}_{\text{mat}} = \frac{1}{N} x_{\text{centered}}^T x_{\text{centered}} \tag{4}$$

$$\text{Cor}_{\text{mat}} = \frac{\text{Cov}_{\text{mat}}}{\text{std} \bullet \text{std}^T} \tag{5}$$

where N denotes the number of nodes, F_i denotes the feature vector of the i-th node, M_F^N represents the NFM (N, F), Cor_{mat} represents the correlation matrix. Then, a Laplacian spectral transformation is applied to the correlation matrix to obtain the Laplacian spectral transformation matrix, which is used as the adjacency matrix (N, N) to obtain the low-frequency node feature information and global graph structure information and provide useful feature representations for subsequent processing and tasks with the GCN.

The NFM and the adjacency matrix generated in the CNN block are fed into the GCN block. The GCN block consists of two graph convolutional layers (GCLs) (Figure 3c). In the GCL, information aggregation is achieved through the multiplication of the adjacency matrix and the NFM. This allows each node to interact and aggregate its features with those of its neighbor nodes, thereby capturing the structural information among the nodes to output updated node representations. The information aggregation process in the graph convolution is formulated as follows:

$$H = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}XW\right) \tag{6}$$

$$\hat{A} = A + I \tag{7}$$

$$\hat{D} = \sum \hat{A}_{ij} \tag{8}$$

where X is the NFM (N, F), H is the updated NFM, A denotes the adjacency matrix (N, N), I represents the identity matrix, W represents the parameter matrix, and σ denotes an activation function. After the first GCL outputs the updated NFM, the correlations between the updated nodes are analyzed in the same way as in the first part, and an updated adjacency matrix is obtained. Then, the updated NFM and adjacency matrix are fed into the second GCL. The number of channels for each GCL is 64 and 32. Additionally, each GCL utilizes the ReLU activation function. Table S4 in Supporting Information S1 presents the detailed settings of the GCN block.

The GCN block aggregates and updates the node information. The GCN block is followed by an attention mechanism layer (Vaswani et al., 2017), which enhances the attention to important information in the node features output by the GCN block and further improves the accuracy of the model. The output of the attention mechanism layer is then flattened to obtain a one-dimensional feature vector, referred to as the GCN features. According to the knowledge base and the seismic waveforms as input, the IMs are extracted as one-dimensional vectors. Then, the IMs as knowledge information are embedded into the network through concatenating the GCN features and IMs. Next, the concatenation of GCN features and IMs is fed into the FCL block of DHLnet architecture. The FCL block consists of six FCLs with 256, 128, 64, 50, 30, and 1 neuron. All FCLs, except for the last FCL, utilize the ReLU activation function, while the last FCL uses a linear activation function. Table S5 in Supporting Information S1 presents the detailed settings of the FCL block. The learning process of neural networks consists of forward and backward propagation processes (LeCun et al., 2015). During the forward propagation process, the IMs embedding influences the network's prediction results. The gradient information is then backpropagated to optimize the network parameters of the CNN and GCN blocks. Importantly, this process enhances the network interpretability and enables to the network to effectively use knowledge information.

We used the mean squared error (MSE) loss function for 300 epochs with a batch size of 256 to train the DHLnet. Throughout the process of training, the optimization of DHLnet was carried out employing the Adam optimizer, commencing with an initial learning rate of 0.001 (Kingma & Ba, 2014). In order to form a validation set, we randomly assigned a fraction of 10% from the training data, using the command "validation split = 0.1." To optimize the learning rate, we employed "ReduceLROnPlateau" to monitor the validation loss. In case the loss of validation failed to decrease consecutively for five epochs, a multiplication factor of 0.1 was applied to the initial learning rate. The learning rate was determined to have a minimum value of 0.1×10^{-7}. Furthermore, the employment of "EarlyStopping" technique was adopted to supervise the validation loss and avert overfitting, as mentioned in the study of Prechelt (2012). If, even after 10 epochs, the validation loss showed no signs of decrease, the training process was promptly terminated.

In this study, based on the DHLnet architecture, we separately trained DHLnet models for predicting epicentral distance (DHLnet-Epi model), magnitude (DHLnet-Mag model), PGA (DHLnet-PGA model), PGV (DHLnet-PGV model). Besides, to assess the impact of varying time window durations following the initial P-wave detection on the prediction of EEW parameters within the DHLnet models, we trained ten DHLnet-Epi models, ten DHLnet-Mag models, ten DHLnet-PGA models, and ten DHLnet-PGV models with different time window lengths between 1 and 10 s (interval of 1 s) after P-wave arrival. Figures S1–S4 in Supporting Information S1 display the loss curves for the DHLnet-Epi, DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models, respectively. Once the triggered stations record the P-wave signal, the DHLnet-Epi model is employed to predict the epicentral distance. Subsequently, we use the epicentral distance predicted by the DHLnet-Epi model as the knowledge embedding for the DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models. Moreover, by averaging the predicted magnitudes for multiple triggered stations (Text S2 in Supporting Information S1), the early warning magnitude at that moment can be obtained using the DHLnet. And after an earthquake occurs, the DHLnet workflow is shown in Figure 3d.

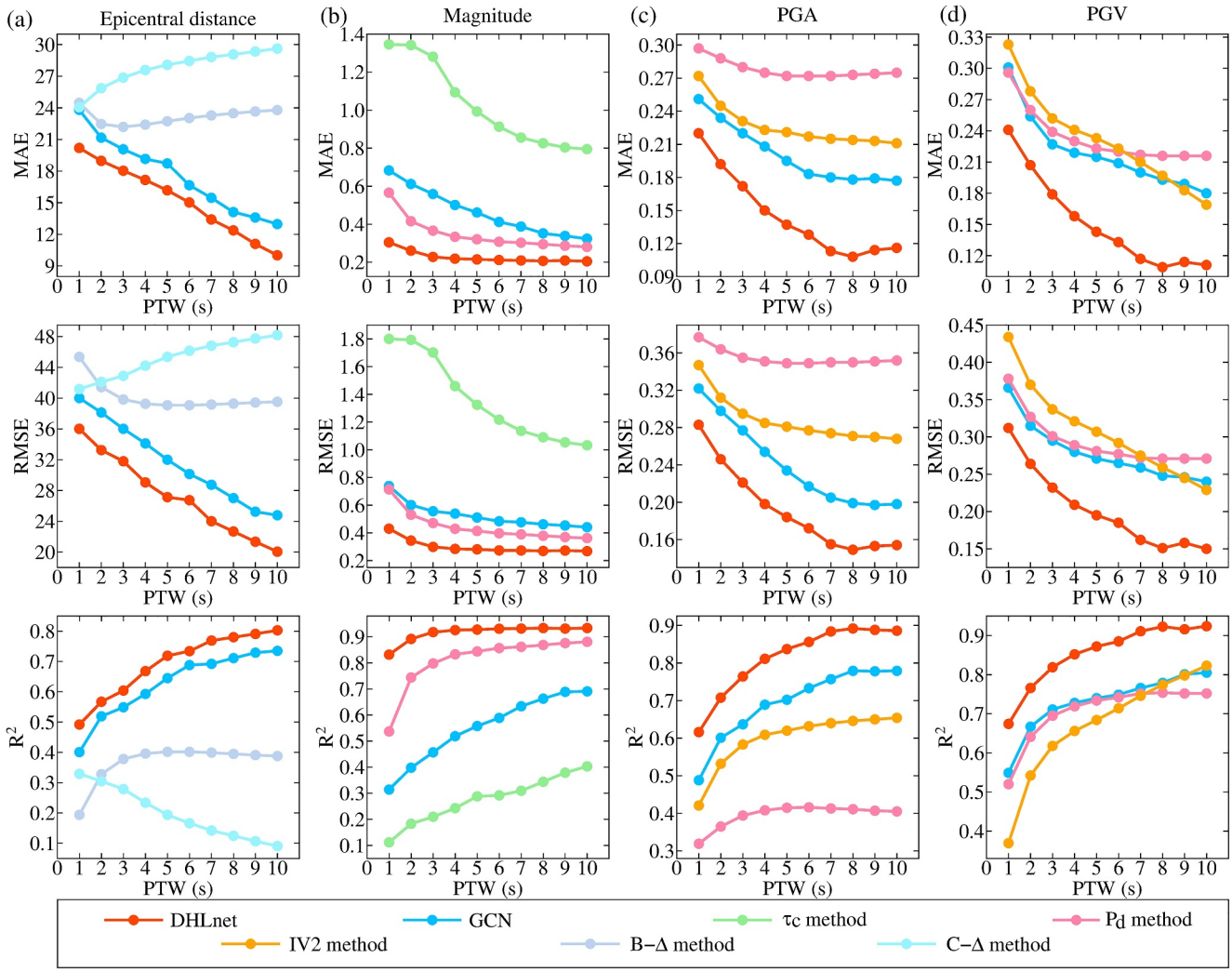

Figure 4. The results of predicting EEW parameter based on the test data set. (a) The capability of DHLnet and the baseline models in predicting epicentral distance. (b) The capability of DHLnet and the baseline models in estimating magnitude. (c) The capability of DHLnet and the baseline models in predicting $\log_{10}$(PGA) prediction. (d) The capability of DHLnet and the baseline models in predicting $\log_{10}$(PGV) prediction.

3. Results

3.1. Performance of DHLnet

In order to analyze the performance of the DHLnet for predicting epicentral distance, magnitude, PGA and PGV in EEW, using the same test data set, we compared the DHLnet with baseline models (Figure 4), which include knowledge-driven traditional EEW methods and purely data-driven deep learning models. Meanwhile, for a fair comparison, we trained the baseline models using the same training data set as the DHLnet. In this study, the knowledge-driven traditional EEW methods include the B-Δ method for predicting epicentral distance (Odaka et al., 2003), C-Δ method for predicting epicentral distance (Yamamoto et al., 2012), τ_c method for estimating magnitude (Kanamori, 2005), P_d method for predicting magnitude and peak ground motion (PGA and PGV) (Wu & Kanamori, 2005; Wu & Zhao, 2006), IV2 method for predicting peak ground motion (PGA and PGV) (Brondi et al., 2015). The purely data-driven deep learning model is GCN model for predicting epicentral distance, magnitude, PGA and PGV. The GCN model is the network architecture after removing the knowledge base from the DHLnet. And the input of the FCL block in GCN architecture is only the GCN features without the IMs embedding. Meanwhile, GCN models were trained using the same training procedures as the DHLnet. Detailed descriptions of the baseline models are provided in Text S3, Tables S6–S13 in Supporting Information S1.

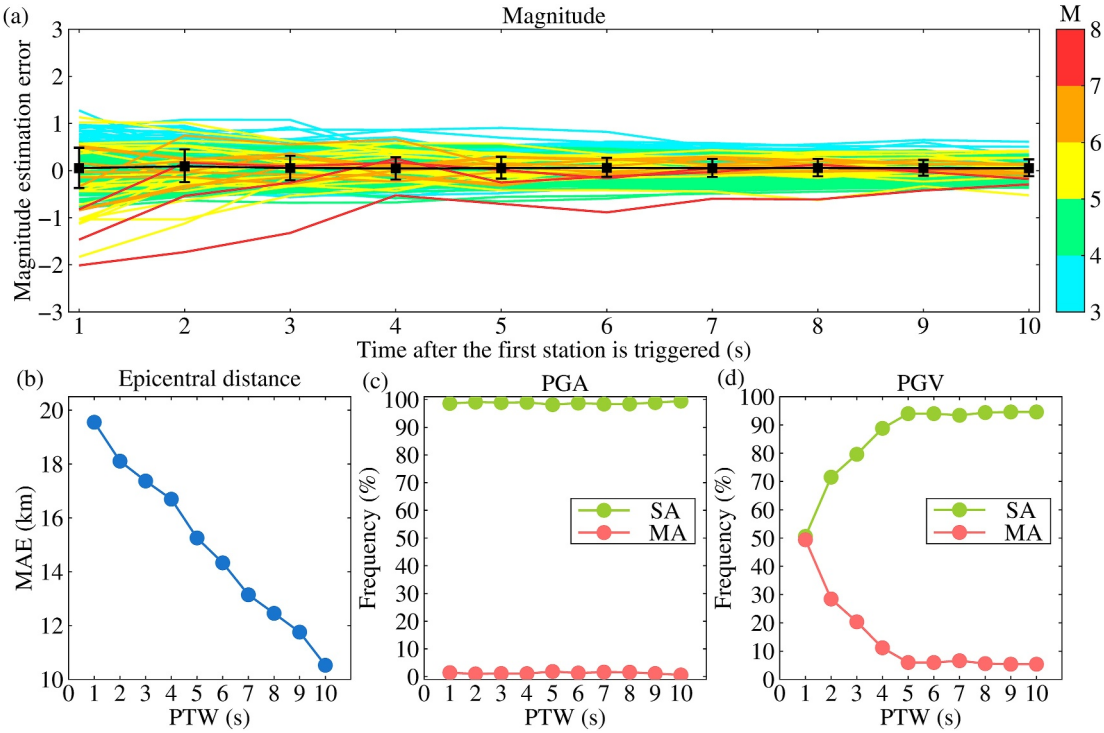

Figure 5. Performance of the DHLnet on the independent data set. (a) The temporal evolution of the error in magnitude estimation from the moment the initial station is activated. (b) The MAE for predicting epicentral distance. (c) Alarm performance based on predicting PGA. (d) Alarm performance based on predicting PGV.

We employed the mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) as evaluation metrics to assess the performance of the models. Figure 4 presents the results of the DHLnet and baseline models for predicting epicentral distance, magnitude, PGA and PGV within a 10 s time window following the arrival of the P-wave. By observing Figure 4, it becomes evident that (a) during the time window of 1–10 s after the P-wave arrival (PTW), compared to the knowledge-driven traditional EEW methods and purely data-driven deep learning model, DHLnet has smaller MAE and RMSE values, as well as larger R^2 value for epicentral distance estimation, magnitude estimation, and PGA and PGV prediction; (b) The DHLnet with 1 s PTW outperform the baseline models, which require longer PTWs to achieve similar performance. (c) The DHLnet with shorter PTWs achieve similar performance to the baseline models with 10 s PTW. These findings demonstrate that DHLnet has better performance than knowledge-driven EEW methods and purely data-driven deep learning models. Meanwhile, DHLnet enhances the timeliness and robustness of predicting epicentral distance, magnitude, PGA and PGV in EEW, which also indicates that DHLnet can integrate the knowledge information from knowledge-driven EEW methods and the strong learning ability from data-driven deep learning methods. In other words, knowledge information embedding can effectively improve the deep learning model (GCN model) performance in EEW. Meanwhile, data-knowledge driven hybrid deep learning architecture is significant and promising for EEW systems.

3.2. Application on an Independent Data Set

To validate the generalization and robustness of the data-knowledge driven hybrid deep learning network for EEW, we applied the trained DHLnet to an independent data set, which was distinct from both the training and test data sets. This independent data set consists of 497 earthquake events with $7.9 \geq M_{JMA} \geq 3$ that occurred in Japan between 2017 and 2020 (Table S14). Figure S5 in Supporting Information S1 illustrates the spatial arrangement of epicenters and K-NET stations. Meanwhile, the distributions of the magnitude, PGA, and PGV at various epicentral distances are depicted in Figures S6a–S6c in Supporting Information S1 for the independent data set. After each earthquake event occurs, according to the multistation average method (Text S2 in Supporting Information S1), the DHLnet was used to estimate the magnitude within the initial 10 s period following the first station detects the P-wave. Figure 5a shows that within the initial 10 s period following the first station detects the

P-wave, the magnitude estimation error is mainly distributed within ±1 magnitude units. Additionally, it is observed that the magnitude estimation error gradually diminishes as time progresses after the first triggered station. This observation indicates that although the DHLnet is based on single-station magnitude estimation, it still yields robust magnitude estimations when combined with multistation averaging for EEW. According to Figure 5b, it can be observed that as time elapses after the arrival of the P-wave, there is a notable decline in the MAE related to the prediction of the epicentral distance. Specifically, within a 10 s time window subsequent to the occurrence of the P-wave, the MAE for the epicentral distance prediction equates to 10.5 km.

Considering that the PGA and PGV are mainly used for alerts in on-site EEW (Colombelli et al., 2015; Wang et al., 2022), we analyzed the alarm performance based on the DHLnet. In our study, we employed a moderate-to-strong intensity level denoted as I_{mm} = VI (PGA = 9.2 cm/s^2, PGV = 8.1 cm/s). This specific intensity level is associated with the transition from moderate to strong perceived shaking and potential damage, as elaborated by Wald et al. (1999). The threshold values for I_{mm} = VI were set at 9.2 cm/s^2 for the PGA and 8.1 cm/s for the PGV. Moreover, in accordance with previous studies, in the alarm performance evaluation for the EEW system, we deemed a one-grade error (I_{mm}) as an acceptable level for users, and the corresponding allowed alarms were determined accordingly (Ahn et al., 2023; Hsu & Pratomo, 2022). For the stations with observed PGA and PGV values exceeding the thresholds, based on our observations in Figures 5c and 5d, it can be noted that (a) the DHLnet accurately predicts the PGA, with an approximately 100% successful alarms (SAs) within 1 s of the P-wave arrival. There is a very low percentage of missed alarms (MAs). (b) Similarly, in the case of the predicted PGV by the DHLnet, the percentage of SAs and MAs stands at around 50% during the same 1 s duration following the P-wave arrival. Nevertheless, as the time window following the P-wave arrival (PTW) expands, the rate of SAs gradually increases while the incidence of MAs diminishes. In addition, analyzing the occurrence of false alarms (FAs) for stations with observed PGA and PGV values below the thresholds is also crucial for EEW systems, Figures S7a and S7b in Supporting Information S1 demonstrate that for both the predicted PGA and PGV, the percentage of successful no alarms (SNAs) is close to 100%, and the percentage of FAs is close to 0%. These findings indicate that the data-knowledge driven hybrid deep learning network enable good alerting performance in EEW systems. Moreover, in practical EEW applications, selecting appropriate thresholds according to different regions and user requirements are crucial (Münchmeyer et al., 2021b). Further in-depth analysis is needed in future EEW research to address these aspects.

4. Discussion

Purely data-driven deep learning models use deep neural network to extract features from raw seismic waveform data and use them for target data prediction. Jozinović et al. (2020) used raw seismic waveform data as input for CNN and predicted the earthquake ground shaking intensity. Meanwhile, Bloemheuvel et al. (2023) used seismic waveform data as input to a graph neural network, which outputs maximum intensity measurements of ground shaking. In addition, some studies have also attempted to combine domain knowledge with machine learning (Fayaz & Galasso, 2024). Fayaz and Galasso (2022) proposed a deep learning network framework for predicting acceleration response spectra, using site feature vectors and seven intensity measures (IMs) calculated from ground motion record at 3 s after P-wave arrival as inputs. Unlike Jozinović et al. (2020) and Bloemheuvel et al. (2023), which only use seismic waveform data as input for deep neural network and predict maximum intensity measures of ground shaking; Unlike Fayaz and Galasso (2022), which only use domain knowledge as input for machine learning methods and predict acceleration response spectra, this approach may overlook important domain knowledge that has not been obtained from seismic waveforms. Here, we propose a Data-knowledge driven Hybrid deep Learning network (DHLnet) for EEW using the waveform input, domain knowledge embedding, CNN and GCN, aiming to integrate knowledge information from knowledge-driven EEW methods and the strong learning ability of data-driven deep learning methods. The proposed DHLnet can predict epicentral distance, magnitude, PGA and PGV using single-station in EEW. In this study, for the same test data set, we find that DHLnet models improve the accuracy of epicentral distance, magnitude, PGA and PGV prediction. Moreover, it is essential in the future to explore which IMs play significant roles in enhancing the model performance and whether expanding the information in the knowledge base can further improve model performance. The proposed data-knowledge driven hybrid deep learning architecture is promising for EEW systems, providing new insights into the exploration of deep learning methods for EEW domain. Meanwhile, we expect that data-knowledge driven hybrid deep learning has potential to improve the performance of earthquake focal mechanism prediction (Kuang et al., 2021), earthquake detection and phase picking (Mousavi et al., 2020).

Table 1
The Impact of the Epicentral Distance on the Performance of the DHLnet-Mag Model

DHLnet-Mag model	MAE	R^2	RMSE
With distance feature	0.228	0.918	0.299
Without distance feature	0.409	0.736	0.539

Additionally, in this study, we applied the proposed method to an independent data set composed of earthquake events that occurred in Japan from 2017 to 2020, which was distinct from both the training and test data sets. For the independent data set, based on the DHLnet models, we analyzed the epicentral distance estimation within 10 s after P-wave arrival and the magnitude estimation within 10 s after the first station detects the P-wave. We found that at 1 s after the first station detects the P-wave, the magnitude estimation error based on the DHLnet model was mainly distributed within the range of ±1 magnitude units; at 10 s after the first station detects the P-wave, the magnitude estimation error based on the DHLnet model was mainly distributed within the range of ±0.5 magnitude units. Additionally, within 10 s after the P-wave arrival, the mean absolute error (MAE) of the estimated epicentral distance based on the DHLnet model is 10.5 km. Meanwhile, for the independent data set, we analyzed the threshold alarm performance of the DHLnet model based on the prediction results of on-site PGA and on-site PGV within 10 s after P-wave arrival. For moderate-to-strong intensity level denoted as I_{mm} = VI (PGA = 9.2 cm/s^2, PGV = 8.1 cm/s), within 10 s after P-wave arrival, based on the predicted PGA and PGV of the DHLnet models, the percentage of successful alarms and successful no alarms is close to 100%, and the percentage of missed alarms and false alarms is close to 0%. These results also indicate that the DHLnet model has a certain degree of generalization ability on the independent data set composed of earthquake events that occurred in Japan from 2017 to 2020. Additionally, transfer learning can help accelerate and optimize the learning efficiency and ability of pretrained models on new target domains and tasks (Jozinović et al., 2022). In future research, we will further explore the application of transfer learning to the methods proposed in this study, improve the interpolation and extrapolation capabilities of the proposed model, and enhance its generalization ability, so that the model can be applied to more regions and countries.

To further analyze the performance of the DHLnet models with different epicentral distances and signal-to-noise ratio (SNR) ranges, Tables S15 and S16 in Supporting Information S1, respectively show the results of different methods for epicentral distance estimation with different epicentral distances and SNR ranges for the same test data set, at 3 s after the arrival of P-wave. In this work, according to the previous research, the SNR is defined as the ratio of the peak velocity within 5 s after the arrival of the P-wave to the peak velocity within 5 s before the arrival of the P-wave (Carranza et al., 2015). From Tables S15 and S16 in Supporting Information S1, it can be found that compared with the knowledge-driven EEW methods (B-Δ method and C-Δ method) and purely data-driven deep learning model (GCN model), the DHLnet model has smaller MAE and RMSE, as well as larger R^2, for epicentral distance estimation in different ranges of epicentral distance and SNR. This also indicates that the estimated epicentral distance of the DHLnet model is less affected by the epicentral distance and SNR. Meanwhile, for the same test data set, at 3 s after P-wave arrival, Tables S17 and S18 in Supporting Information S1, respectively show the results of different methods for magnitude estimation with different epicentral distances and SNR ranges. It can be found from Tables S17 and S18 in Supporting Information S1 that compared with the knowledge-driven EEW methods (τ_c method and P_d method) and purely data-driven deep learning model (GCN model), the DHLnet model has smaller MAE and RMSE, as well as larger R^2, for magnitude estimation in different ranges of epicentral distance and SNR. This also indicates that the estimated magnitude of the DHLnet model is less affected by the epicentral distance and SNR. Additionally, for the same test data set, at 3 s after P-wave arrival, Tables S19 and S20 in Supporting Information S1, respectively show the results of different methods for on-site PGA prediction with different epicentral distances and SNR ranges. It can be seen from Table S19 and S20 in Supporting Information S1 that compared with the knowledge-driven EEW methods (P_d method and IV2 method) and purely data-driven deep learning model (GCN model), the DHLnet model has smaller MAE and RMSE, as well as larger R^2, for on-site PGA prediction in different ranges of epicentral distance and SNR. This also indicates that the predicted PGA of the DHLnet model is less affected by the epicentral distance and SNR. Besides, for the same test data set, at 3 s after P-wave arrival, Tables S21 and S22 in Supporting Information S1, respectively show the results of different methods for on-site PGV prediction with different epicentral distances and SNR ranges. We can see from Tables S21 and S22 in Supporting Information S1 that compared with the knowledge-driven EEW methods (P_d method and IV2 method) and purely data-driven deep learning model (GCN model), the DHLnet model has smaller MAE and RMSE, as well as larger R^2, for on-site PGV prediction in

Table 2
The Impact of the Epicentral Distance on the Performance of the DHLnet-PGA Model

DHLnet-PGA model	MAE	R^2	RMSE
With distance feature	0.172	0.764	0.221
Without distance feature	0.194	0.710	0.246

Table 3
The Impact of the Epicentral Distance on the Performance of the DHLnet-PGV Model

DHLnet-PGV model	MAE	R^2	RMSE
With distance feature	0.179	0.819	0.232
Without distance feature	0.198	0.776	0.258

different ranges of epicentral distance and SNR. This also indicates that the predicted PGV of the DHLnet model is less affected by the epicentral distance and SNR.

In actual EEW, after an earthquake occurs, we do not know the actual on-site PGA and actual magnitude. We predict PGA based on the P-wave signal recorded by the station. Meanwhile, it is also important for the EEW system to have a low false alarm rate for the site where the recorded PGA < 0.01 g data is located. The false alarms of EEW system can also cause unnecessary concerns and economic losses to the public. Therefore, reliable prediction for ground motion record with PGA < 0.01 g is also very meaningful for EEW. Compared with ground motion data with PGA < 0.01 g, the site where the recorded ground data with PGA ≥ 0.01 g is located may have more severe seismic damage (Wald et al., 1999), so it is also important to pay attention to ground motion data with PGA ≥ 0.01 g. We retrained and retested the DHLnet models proposed in this study using ground data with PGA ≥ 0.01 g. Meanwhile, we also retrained and retested the baseline models (including knowledge-driven traditional EEW methods and purely data-driven deep learning models) using the same data set as the DHLnet models. Detailed descriptions of the baseline models are provided in Text S3, Tables S6–S13 in Supporting Information S1. For the same test data set (PGA ≥ 0.01 g), Table S23 in Supporting Information S1 shows the estimation results of epicentral distance at 3 s after the arrival of P-wave. B-Δ method and C-Δ method are the knowledge-driven traditional EEW methods used for epicentral distance estimation, while the GCN model is a purely data-driven deep learning model used for epicentral distance estimation, and the GCN model does not have knowledge embedding. The GCN model is the network architecture after removing the knowledge base from the DHLnet. And the input of the FCL block in GCN architecture is only the GCN features. From Table S23 in Supporting Information S1, it can be observed that compared with B-Δ method, C-Δ method and the GCN model, the DHLnet model has smaller MAE, RMSE, and larger R^2 for epicentral distance estimation. For the same test data set (PGA ≥ 0.01 g), at 3 s after the P-wave arrival, Table S24 in Supporting Information S1 shows the estimation results of magnitude. τ_c method and P_d method are the knowledge-driven traditional EEW methods used for magnitude estimation, while the GCN model is a purely data-driven deep learning model used for magnitude estimation. From Table S24 in Supporting Information S1, it can be observed that compared with τ_c method, P_d method and the GCN model, the DHLnet model has smaller MAE, RMSE, and larger R^2 for magnitude estimation. For the same test data set (PGA ≥ 0.01 g), at 3 s after the P-wave arrival, Table S25 in Supporting Information S1 shows the prediction results of on-site PGA. P_d method and IV2 method are the knowledge-driven traditional EEW methods used for on-site PGA prediction, while the GCN model is a purely data-driven deep learning model used for on-site PGA prediction. From Table S25 in Supporting Information S1, it can be observed that compared with P_d method, IV2 method and the GCN model, the DHLnet model has smaller MAE, RMSE, and larger R^2 for on-site PGA prediction. For the same test data set (PGA ≥ 0.01 g), at 3 s after the P-wave arrival, Table S26 in Supporting Information S1 shows the prediction results of on-site PGV. P_d method and IV2 method are the knowledge-driven traditional EEW methods used for on-site PGV prediction, while the GCN model is a purely data-driven deep learning model used for on-site PGV prediction. From Table S26 in Supporting Information S1, it can be observed that compared with P_d method, IV2 method and the GCN model, the DHLnet model has smaller MAE, RMSE, and larger R^2 for on-site PGV prediction.

LSTM (Long Short-Term Memory) recurrent neural networks (RNNs) have also been widely used in EEW (Hsu & Pratomo, 2022; Wang et al., 2022). LSTM mainly focuses on time dependence in sequences, and its global feature extraction ability is weak. The Data-knowledge driven Hybrid deep Learning network (DHLnet) proposed in this study combines CNN and GCN. DHLnet uses CNN block to extract local features from input seismic waveform data and obtains node feature matrix and adjacency matrix. Then, node information is aggregated through GCN block to achieve the fusion of local and global node features. Compared to LSTM, the DHLnet proposed in this study can better capture global information and enhance the model's feature expression ability. To analyze the performance of the DHLnet and LSTM proposed in this paper, based on the LSTM network architecture proposed by Hsu and Pratomo (2022), we retrained the LSTM model using the same training data set as the DHLnet. In addition, we also used the same input and knowledge embedding as the DHLnet model for the LSTM model. Meanwhile, the comparison with LSTM is essentially a comparison of different network architectures, rather than a comparison of methods. Detailed descriptions of the LSTM models are provided in Supporting Information Text S4 in Supporting Information S1. For the same test data set, we can find from Table

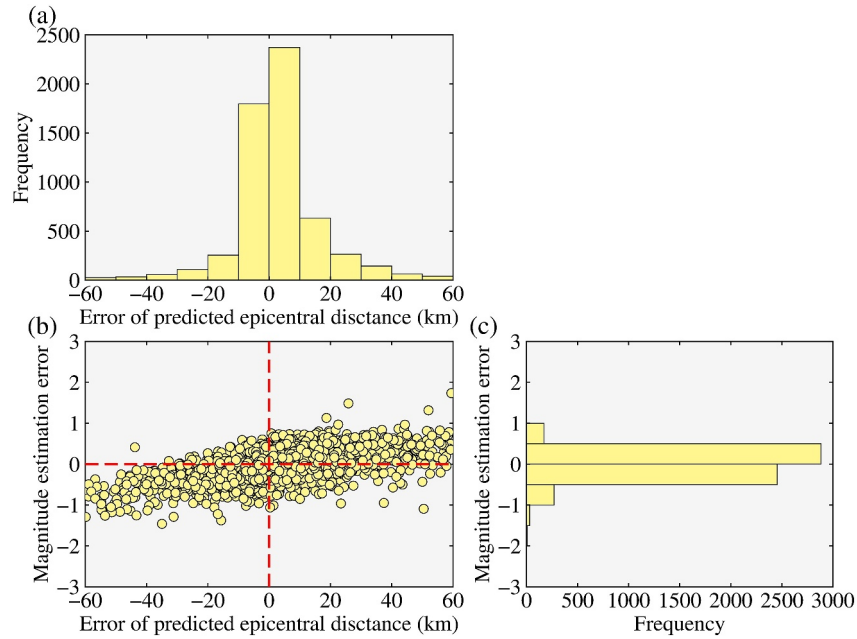

Figure 6. Impact of prediction errors in the epicentral distance of the DHLnet-Epi model on the prediction result of the DHLnet-Mag model. (a) Distribution of prediction errors in epicentral distance. (b) Relationship between prediction error in magnitude and epicentral distance prediction error. (c) Distribution of estimation errors in magnitude.

S27 in Supporting Information S1 that compared to the LSTM models, the DHLnet models have smaller MAE and RMSE, as well as larger R^2 for epicentral distance, magnitude, PGA, and PGV prediction, at 3 s after the arrival of P-wave. This result also indicates that the DHLnet model proposed in this study has better performance and advantages in earthquake early warning compared to the LSTM model.

Besides, although the occurrence frequency of large earthquakes ($M \geq 8$) is very low, their destructive potential is extremely severe. Additionally, Li (2022) showed that complex and irregular ruptures in large earthquakes are very rare, which to some extent limits the magnitude predictability of EEW systems. Considering the rarity of large earthquakes ($M \geq 8$), existing deep learning methods that rely on large training sets may not be well suited for predicting such events. Therefore, in this study, we focus on earthquakes with magnitudes below eight in both the training and test data sets and the independent data set. Our research demonstrates that the DHLnet exhibits excellent performance in predicting EEW parameters for earthquakes with magnitudes below eight in the data set utilized in this study.

In prior studies regarding the anticipation of EEW parameter, the significant information for estimating the magnitude and predicting the peak ground motion was typically the epicentral distance (Festa et al., 2008; Zollo et al., 2006). Therefore, in this work, the epicentral distance is used as the knowledge embedding for the DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models. Tables 1–3 show the impact of the epicentral distance on the performance of the DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models at the 3 s P-wave arrival based on the same test data set. Tables 1–3 demonstrate that the distance feature embedding improves the performance of the DHLnet for predicting magnitude, PGA and PGV in EEW to a certain extent, as indicated by the increased R^2 and decreased MAE and RMSE values.

Furthermore, in this study, in addition to distance feature (epicentral distance), the IMs embedded in the DHLnet model were calculated using ground motion records after P-wave arrival. In practical EEW applications, we used the epicentral distance predicted by the DHLnet-Epi model as the knowledge embedding for the DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models. However, the epicentral distance predicted by the DHLnet-Epi model may have some errors. For the independent data set, Figures 6–8 illustrate the influence of the DHLnet-Epi model's epicentral distance prediction errors on the prediction results of the DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models at 10 s PTW. Here, the error is defined as the disparity between the anticipated value and the actual value. It can be observed from Figures 6–8 that (a) there exists a positive correlation between the errors

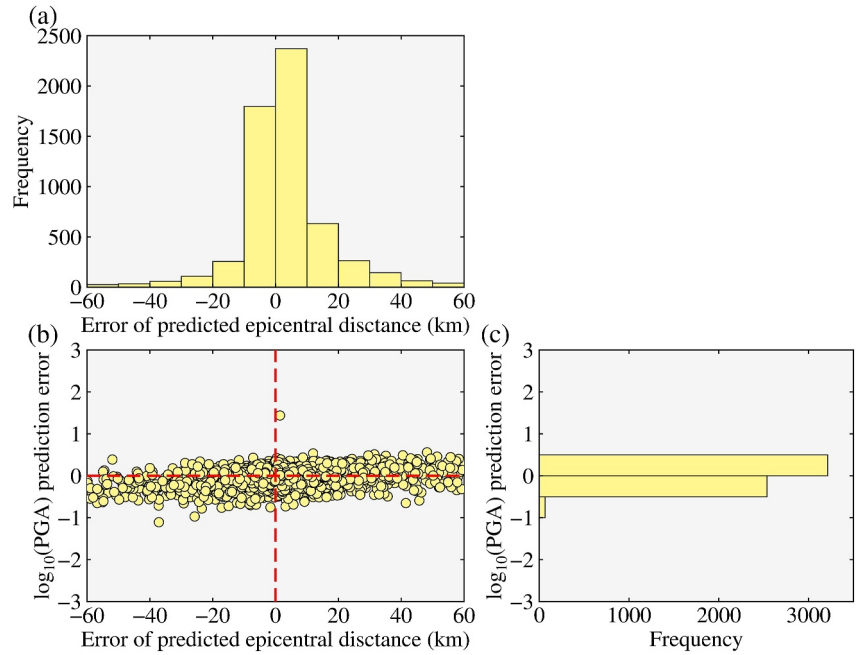

Figure 7. Impact of prediction errors in the epicentral distance of the DHLnet-Epi model on the prediction result of the DHLnet-PGA model. (a) Distribution of prediction errors in epicentral distance. (b) Relationship between prediction error in $\log_{10}$(PGA) and epicentral distance prediction error. (c) Distribution of prediction errors in $\log_{10}$(PGA).

in magnitude estimation, $\log_{10}$(PGA) prediction, and $\log_{10}$(PGV) prediction, and the errors in epicentral distance prediction. (b) While the errors in epicentral distance prediction do impact the performance of the DHLnet-Mag, DHLnet-PGA, and DHLnet-PGV models to some extent, the errors in magnitude estimation, $\log_{10}$(PGA) prediction, and $\log_{10}$(PGV) prediction are predominantly confined within the ±1 range.

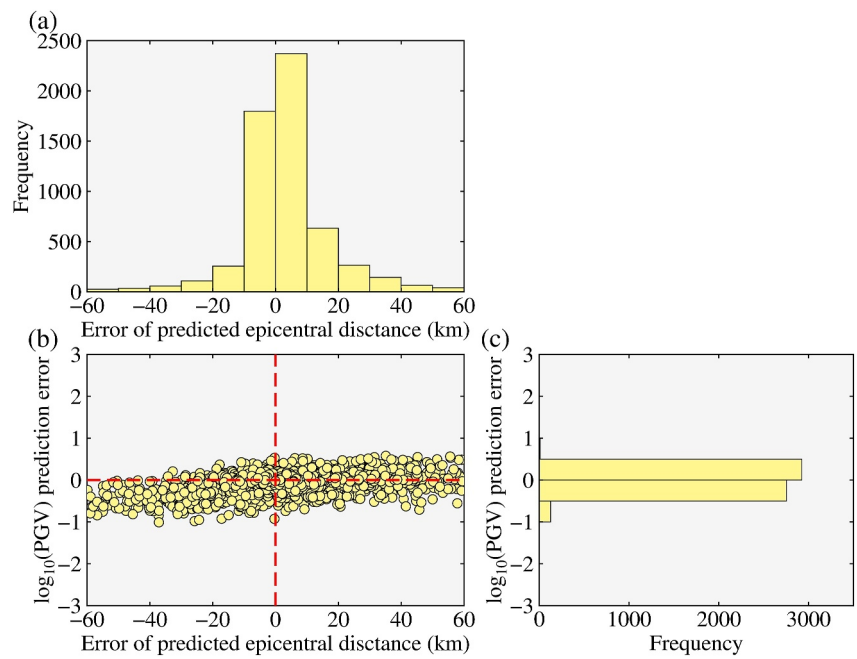

Figure 8. Impact of prediction errors in the epicentral distance of the DHLnet-Epi model on the prediction result of the DHLnet-PGV model. (a) Distribution of prediction errors in epicentral distance. (b) Relationship between prediction error in $\log_{10}$(PGV) and epicentral distance prediction error. (c) Distribution of prediction errors in $\log_{10}$(PGV).

In this work, the performance of the DHLnet model benefits to some extent from the complexity of the model. The more complex the model, the larger the computational resources and memory it uses. To ensure the efficiency and feasibility of the model in practical EEW systems with limited computational resources, the model needs to be lightweight. In other words, it means reducing the computational resources and memory size used by the model. Model compression or pruning technique is a technique that reduces model size by removing unimportant parameters or connections from the model. By using compression or pruning techniques, we can reduce the redundancy of the model, making it more suitable for running in resource limited environments and ensuring the overall efficiency of the model in actual EEW systems. In order to ensure the operational efficiency of the DHLnet model in actual EEW systems and its feasibility in environments with limited computational resources, in future research, we will further explore how to implement compression or pruning techniques on the DHLnet models, so that after using compression or pruning techniques, the required computing resources of the DHLnet model are significantly reduced and the performance of the model does not undergo significant changes.

5. Conclusions

By supplying advanced warning information to the general public prior to the arrival of destructive seismic waves at the target site, a dependable EEW system holds the potential to substantially mitigate earthquake disasters. The existing EEW systems mainly use knowledge-driven algorithms, which rely on seismologists to extract IMs from P-wave signals. However, the limited IMs extracted by seismologists from P-wave signals are insufficient to characterize the EEW parameters. Besides, data-driven deep learning methods with the strong learning abilities in EEW do not consider knowledge information from seismologists. Here, we develop a Data-knowledge driven Hybrid deep Learning network (DHLnet) for EEW, which can be used for predicting epicentral distance, magnitude, PGA and PGV based directly on single-station data. Meanwhile, we embed expert-defined IMs into the network. Through model training, we optimize the network parameters to enhance the model's interpretability and effectively apply knowledge information. Compared with knowledge-driven methods and data-driven deep learning models, we demonstrate that DHLnet improve the timeliness and robustness for predicting the epicentral distance, magnitude, PGA and PGV in EEW, which also indicates that DHLnet can integrate knowledge information from knowledge-driven EEW methods and the strong learning ability of deep learning methods. Besides, after an earthquake occurs, by combining the data collected by multiple triggered stations, the DHLnet can provide robust magnitude estimates within the first 10 s following the first triggered station. Meanwhile, for threshold-based warnings, according to the PGA and PGV values predicted by the DHLnet, we can obtain reliable alerts during the early stages of P-wave arrival. The proposed data-knowledge driven hybrid deep learning architecture is promising for EEW systems. It offers novel perspectives for exploring deep learning techniques within the EEW domain.

Data Availability Statement

The strong ground motion data utilized in this work can be obtained from the website (https://www.kyoshin.bosai.go.jp/kyoshin/quake/index_en.html) of National Research Institute for Earth Science and Disaster Prevention, Japan (NIED). And the data files (strong ground motion data) used in this paper are available at Aoi et al. (2011). The code/script (Zhu et al., 2023) used in the study for processing and analyzing the data can be available from https://doi.org/10.5281/zenodo.10033205, and Doi is 10.5281/zenodo.10033205.

Acknowledgments
We would like to thank Editor Paolo Diviacco, Editor's Assistant Charlene Chuquillanqui and two anonymous reviewers for their constructive comments and suggestions, which are very helpful for improving our manuscript. The National Natural Science Foundation of China (U2039209, 42304074, and 51408564), the Natural Science Foundation of Heilongjiang Province (LH2021E119), and the National Key Research and Development Program of China (2023YFF0725005, 2018YFC1504003) are acknowledged for their financial support of this research.

References

Ahn, J. K., Park, E., Kim, B., Hwang, E. H., & Hong, S. (2023). Stable operation process of earthquake early warning system based on machine learning: Trial test and management perspective. *Frontiers in Earth Science*, *11*, 1157742. https://doi.org/10.3389/feart.2023.1157742

Alcik, H., Ozel, O., Apaydin, N., & Erdik, M. (2009). A study on warning algorithms for Istanbul earthquake early warning system. *Geophysical Research Letters*, *36*(5), 3–5. https://doi.org/10.1029/2008gl036659

Allen, R. M., Brown, H., Hellweg, M., Khainovski, O., Lombard, P., & Neuhauser, D. (2009). Real-time earthquake detection and hazard assessment by ElarmS across California. *Geophysical Research Letters*, *36*(5), L00B08. https://doi.org/10.1029/2008gl036766

Allen, R. M., & Kanamori, H. (2003). The potential for earthquake early warning in Southern California. *Science*, *300*(5620), 786–789. https://doi.org/10.1126/science.1080912

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

Allen, R. M., & Stogaitis, M. (2022). Global growth of earthquake early warning. *Science*, *375*(6582), 717–718. https://doi.org/10.1126/science.abl5435

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532. https://doi.org/10.1785/bssa0680051521

Aoi, S., Kunugi, T., Nakamura, H., & Fujiwara, H. (2011). Deployment of new strong motion seismographs of K-NET and KiK-net [Dataset]. *Earthquake Data in Engineering Seismology: Predictive Models, Data Management and Networks*, 167–186. https://doi.org/10.17598/NIED.0004

Arias, A. (1970). A measure of earthquake intensity. In R. J. Hansen (Ed.), *Seismic design for nuclear power plants* (pp. 438–483). MIT Press.

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433), eaau0323. https://doi.org/10.1126/science.aau0323

Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., & Atzmueller, M. (2023). Graph neural networks for multivariate time series regression with application to seismic data. *International Journal of Data Science and Analytics*, *16*(3), 317–332. https://doi.org/10.1007/s41060-022-00349-6

Böse, M. (2006). *Earthquake early warning for Istanbul using artificial neural networks (Ph.D. thesis)*. University of Karlsruhe.

Brondi, P., Picozzi, M., Emolo, A., Zollo, A., & Mucciarelli, M. (2015). Predicting the macroseismic intensity from early radiated P wave energy for on-site earthquake early warning in Italy. *Journal of Geophysical Research: Solid Earth*, *120*(10), 7174–7189. https://doi.org/10.1002/2015jb012367

Carranza, M., Buforn, E., Colombelli, S., & Zollo, A. (2013). Earthquake early warning for southern Iberia: A P wave threshold-based approach. *Geophysical Research Letters*, *40*(17), 4588–4593. https://doi.org/10.1002/grl.50903

Carranza, M., Buforn, E., & Zollo, A. (2015). Testing the earthquake early-warning parameter correlations in the southern Iberian Peninsula. *Pure and Applied Geophysics*, *172*(9), 2435–2448. https://doi.org/10.1007/s00024-015-1061-6

Caruso, A., Colombelli, S., Elia, L., Picozzi, M., & Zollo, A. (2017). An on-site alert level early warning system for Italy. *Journal of Geophysical Research: Solid Earth*, *122*(3), 2106–2118. https://doi.org/10.1002/2016JB013403

Chakraborty, M., Fenner, D., Li, W., Faber, J., Zhou, K., Rümpker, G., et al. (2022). CREIME—A convolutional recurrent model for earthquake identification and magnitude estimation. *Journal of Geophysical Research: Solid Earth*, *127*(7), e2022JB024595. https://doi.org/10.1029/2022jb024595

Colombelli, S., Caruso, A., Zollo, A., Festa, G., & Kanamori, H. (2015). A P wave-based, on-site method for earthquake early warning. *Geophysical Research Letters*, *42*(5), 1390–1398. https://doi.org/10.1002/2014gl063002

Colombelli, S., Zollo, A., Festa, G., & Picozzi, M. (2014). Evidence for a difference in rupture initiation between small and large earthquakes. *Nature Communications*, *5*(1), 3958. https://doi.org/10.1038/ncomms4958

Cremen, G., Galasso, C., & Zuccolo, E. (2022). Investigating the potential effectiveness of earthquake early warning across Europe. *Nature Communications*, *13*(1), 639. https://doi.org/10.1038/s41467-021-27807-2

Fayaz, J., & Galasso, C. (2022). A deep neural network framework for real-time on-site estimation of acceleration response spectra of seismic ground motions. *Computer-Aided Civil and Infrastructure Engineering*, *38*(1), 87–103. https://doi.org/10.1111/mice.12830

Fayaz, J., & Galasso, C. (2024). Interpretability and spatial efficacy of deep-learning-based on-site early warning framework using explainable artificial intelligence and geographically weighted random forests. *Geoscience Frontiers*, *101839*. https://doi.org/10.1016/j.gsf.2024.101839

Festa, G., Zollo, A., & Lancieri, M. (2008). Earthquake magnitude estimation from early radiated energy. *Geophysical Research Letters*, *35*(22), L22307. https://doi.org/10.1029/2008gl035576

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), L00B02. https://doi.org/10.1029/2008gl036596

Hsu, T. Y., & Pratomo, A. (2022). Early peak ground acceleration prediction for on-site earthquake early warning using LSTM neural network. *Frontiers in Earth Science*, *10*, 911947. https://doi.org/10.3389/feart.2022.911947

Huang, P. L., Lin, T. L., & Wu, Y. M. (2015). Application of τc* Pd in earthquake early warning. *Geophysical Research Letters*, *42*(5), 1403–1410. https://doi.org/10.1002/2014gl063020

Ioffe, S., & Szegedy, C. (2015). Covariate shift adaptation and dataset shift decomposition in machine learning. In F. Bach & D. Blei (Eds.), *International conference on machine learning* (pp. 448–456). PMLR.

Jiao, P., & Alavi, A. H. (2020). Artificial intelligence in seismology: Advent, performance and future trends. *Geoscience Frontiers*, *11*(3), 739–744. https://doi.org/10.1016/j.gsf.2019.10.004

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

Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2022). Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. *Geophysical Journal International*, *229*(1), 704–718. https://doi.org/10.1093/gji/ggab488

Kamigaichi, O., Saito, M., Doi, K., Matsumori, T., Tsukada, S., Takeda, K., et al. (2009). Earthquake early warning in Japan: Warning the general public and future prospects. *Seismological Research Letters*, *80*(5), 717–726. https://doi.org/10.1785/gssrl.80.5.717

Kanamori, H. (2005). Real-time seismology and earthquake damage mitigation. *Annual Review of Earth and Planetary Sciences*, *33*(1), 195–214. https://doi.org/10.1146/annurev.earth.33.092203.122626

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv:1412.6980.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv:1609.02907.

Kiranyaz, S., Ince, T., Hamila, R., & Gabbouj, M. (2015). Convolutional neural networks for patient-specific ECG classification. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 2608–2611). IEEE.

Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., et al. (2020). Earthquake early warning ShakeAlert 2.0: Public rollout. *Seismological Research Letters*, *91*(3), 1763–1775. https://doi.org/10.1785/0220190245

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Kong, Q., & Zhao, M. (2012). Evaluation of earthquake signal characteristics for early warning. *Earthquake Engineering and Engineering Vibration*, *11*(3), 435–443. https://doi.org/10.1007/s11803-012-0133-1

Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature Communications*, *12*(1), 1432. https://doi.org/10.1038/s41467-021-21670-x

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Z. (2022). A generic model of global earthquake rupture characteristics revealed by machine learning. *Geophysical Research Letters*, *49*(8), e2021GL096464. https://doi.org/10.1029/2021gl096464

Lockman, A. B., & Allen, R. M. (2005). Single-station earthquake characterization for early warning. *Bulletin of the Seismological Society of America*, *95*(6), 2029–2039. https://doi.org/10.1785/0120040241

McBride, S. K., Smith, H., Morgoch, M., Sumy, D., Jenkins, M., Peek, L., et al. (2022). Evidence-based guidelines for protective actions and earthquake early warning systems. *Geophysics*, *87*(1), WA77–WA102. https://doi.org/10.1190/geo2021-0222.1

Mousavi, S. M., & Beroza, G. C. (2020a). Bayesian-deep-learning estimation of earthquake location from single-station observations. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(11), 8211–8224. https://doi.org/10.1109/tgrs.2020.2988770
Mousavi, S. M., & Beroza, G. C. (2020b). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019gl085976
Mousavi, S. M., & Beroza, G. C. (2022). Deep-learning seismology. *Science*, *377*(6607), eabm4470. https://doi.org/10.1126/science.abm4470
Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. https://doi.org/10.1038/s41467-020-17591-w
Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021a). Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. *Geophysical Journal International*, *226*(2), 1086–1104. https://doi.org/10.1093/gji/ggab139
Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021b). The transformer earthquake alerting model: A new versatile approach to earthquake early warning. *Geophysical Journal International*, *225*(1), 646–656. https://doi.org/10.1093/gji/ggaa609
Nagi, J., Ducatelle, F., Di Caro, G. A., Ciresan, D., Meier, U., Giusti, A., et al. (2011). Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE international conference on signal and image processing applications (ICSIPA) Kuala Lumpur, Malaysia* (pp. 342–347). IEEE.
Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In J. Furnkranz & T. Joachims (Eds.), *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807–814). Omnipress.
Odaka, T., Ashiya, K., Tsukada, S. Y., Sato, S., Ohtake, K., & Nozaka, D. (2003). A new method of quickly estimating epicentral distance and magnitude from a single seismic record. *Bulletin of the Seismological Society of America*, *93*(1), 526–532. https://doi.org/10.1785/0120020008
Olson, E. L., & Allen, R. M. (2005). The deterministic nature of earthquake rupture. *Nature*, *438*(7065), 212–215. https://doi.org/10.1038/nature04214
Papadopoulos, A. N., Böse, M., Danciu, L., Clinton, J., & Wiemer, S. (2023). A framework to quantify the effectiveness of earthquake early warning in mitigating seismic risk. *Earthquake Spectra*, *39*(2), 938–961. https://doi.org/10.1177/87552930231153424
Peng, C., Ma, Q., Jiang, P., Huang, W., Yang, D., Peng, H., et al. (2020). Performance of a hybrid demonstration earthquake early warning system in the Sichuan–Yunnan border region. *Seismological Research Letters*, *91*(2A), 835–846. https://doi.org/10.1785/0220190101
Peng, C. Y., Yang, J. S., Zheng, Y., Zhu, X. Y., Xu, Z. Q., & Chen, Y. (2017). New τc regression relationship derived from all P wave time windows for rapid magnitude estimation. *Geophysical Research Letters*, *44*(4), 1724–1731. https://doi.org/10.1002/2016GL071672
Prechelt, L. (2012). Early stopping — But when? In G. Montavon, G. B. Orr, & K. R. Müller (Eds.), *Neural networks: Tricks of the trade* (2nd ed., pp. 53–67). Springer.
Reed, J. W., & Kassawara, R. P. (1990). A criterion for determining exceedance of the operating basis earthquake. *Nuclear Engineering and Design*, *123*(2–3), 387–396. https://doi.org/10.1016/0029-5493(90)90259-z
Saad, O. M., Chen, Y., Savvaidis, A., Fomel, S., & Chen, Y. (2022). Real-time earthquake detection and magnitude estimation using vision transformer. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023657. https://doi.org/10.1029/2021jb023657
Song, J., Zhu, J., & Li, S. (2023). MEANet: Magnitude estimation via physics-based features time series, an attention mechanism, and neural networks. *Geophysics*, *88*(1), V33–V43. https://doi.org/10.1190/geo2022-0196.1
Song, J., Zhu, J., Wang, Y., & Li, S. (2022). On-site alert-level earthquake early warning using machine-learning-based prediction equations. *Geophysical Journal International*, *231*(2), 786–800. https://doi.org/10.1093/gji/ggac220
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958.
Strauss, J. A., & Allen, R. M. (2016). Benefits and costs of earthquake early warning. *Seismological Research Letters*, *87*(3), 765–772. https://doi.org/10.1785/0220150149
van den Ende, M. P. A., & Ampuero, J. P. (2020). Automated seismic source characterization using deep graph neural networks. *Geophysical Research Letters*, *47*(17), e2020GL088690. https://doi.org/10.1029/2020gl088690
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems 30, NIPS'17* (Vol. 30, pp. 5998–6008).
Wald, D. J., Quitoriano, V., Heaton, T. H., & Kanamori, H. (1999). Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California. *Earthquake Spectra*, *15*(3), 557–564. https://doi.org/10.1193/1.1586058
Wang, C. Y., Huang, T. C., & Wu, Y. M. (2022). Using LSTM neural networks for onsite earthquake early warning. *Seismological Research Letters*, *93*(2A), 814–826. https://doi.org/10.1785/0220210197
Wu, Y. M., & Kanamori, H. (2005). Rapid assessment of damage potential of earthquakes in taiwan from the beginning of P waves. *Bulletin of the Seismological Society of America*, *95*(3), 1181–1185. https://doi.org/10.1785/0120040193
Wu, Y. M., & Zhao, L. (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters*, *33*(16), L16312. https://doi.org/10.1029/2006gl026871
Yamamoto, S., Noda, S., & Korenaga, M. (2012). An estimation method of epicentral distance based on characteristics of P-wave initial envelope. *RTRI Report*, *26*(9), 5–10.
Zhang, X., Reichard-Flynn, W., Zhang, M., Hirn, M., & Lin, Y. (2022). Spatiotemporal graph convolutional networks for earthquake source characterization. *Journal of Geophysical Research: Solid Earth*, *127*(11), e2022JB024401. https://doi.org/10.1029/2022jb024401
Zhang, X., Zhang, M., & Tian, X. (2021). Real-time earthquake early warning with deep learning: Application to the 2016 M 6.0 central Apennines, Italy earthquake. *Geophysical Research Letters*, *48*(5), 2020GL089394. https://doi.org/10.1029/2020gl089394
Zhu, J., Li, S., & Song, J. (2022a). Magnitude estimation for earthquake early warning with multiple parameter inputs and a support vector machine. *Seismological Research Letters*, *93*(1), 126–136. https://doi.org/10.1785/0220210144
Zhu, J., Li, S., & Song, J. (2022b). Hybrid deep-learning network for rapid on-site peak ground velocity prediction. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–12. https://doi.org/10.1109/tgrs.2022.3230829
Zhu, J., Li, S., & Song, J. (2023). Jingbaozhu1996/DHLnet: v1.0.0 (v1.0.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.10033205
Zhu, J., Li, S., Song, J., & Wang, Y. (2021). Magnitude estimation for earthquake early warning using a deep convolutional neural network. *Frontiers in Earth Science*, *9*, 653226. https://doi.org/10.3389/feart.2021.653226
Zollo, A., Colombelli, S., Caruso, A., & Elia, L. (2023). An evolutionary shaking-forecast-based earthquake early warning method. *Earth and Space Science*, *10*(4), e2022EA002657. https://doi.org/10.1029/2022ea002657
Zollo, A., Lancieri, M., & Nielsen, S. (2006). Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records. *Geophysical Research Letters*, *33*(23), L23312. https://doi.org/10.1029/2006gl027795