Gondwana Research 123 (2023) 164–173

Contents lists available at ScienceDirect

Gondwana Research

journal homepage: www.elsevier.com/locate/gr

Deep learning for magnitude prediction in earthquake early warning

Yanwei Wang [a], Xiaojun Li [b], Zifa Wang [c,d,*], Juan Liu [a,d]

[a] Guangxi Key Laboratory of Geomechanics and Geotechnical Engineering, Guilin University of Technology, Guilin 541004, China
[b] Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, 100124 Beijing, China
[c] CEAKJ ADPRHexa, Inc, Shaoguan 512000, China
[d] Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China

ARTICLE INFO

Article history:
Received 11 November 2021
Revised 22 May 2022
Accepted 16 June 2022
Available online 21 June 2022

Keywords:
Magnitude
Deep learning
Earthquake early warning
P-wave
Ground motion records

ABSTRACT

Fast and accurate magnitude prediction is the key to the success of earthquake early warning (EEW). However, it is difficult to significantly improve the performance of magnitude prediction by empirically defined characteristic parameters. In this study, we have proposed a new approach (EEWNet) based on deep learning to predict magnitude for EEW. The initial few seconds of P-wave recorded by a single station without any preprocessing is used as the input to EEWNet, and the maximum displacement for the whole record is predicted and by which the magnitude is calculated. A large number of borehole underground strong motion records are used to train, validate and test the proposed EEWNet, and the predicted results are compared against those by empirical peak displacement P_d method. The comparison demonstrates that EEWNet produces better and quicker results than those by P_d, and EEWNet can predict magnitude between 4.0 and 5.9 as early as the first 0.5 s P-wave arrives. EEWNet is therefore expected to significantly enhance the accuracy and speed of magnitude estimation in practical regional EEW systems.

1. Introduction

Earthquake early warning (EEW) has become an effective emerging technology in reducing the impact and damage of earthquakes. There have been 9 countries or regions around the world that have operating EEW systems, and another 13 countries or regions are currently testing EEW systems (Cremen & Galasso, 2020). EEW sends warning signals to target areas before the arrival of destructive seismic waves, and the pre-warning time can be as short as a few seconds, which is especially important in reducing the impact and damage of earthquakes (Strauss & Allen, 2016). The regional EEW system performs early warning by determining earthquake parameters, such as the time, location, and magnitude of the earthquake, and the onsite EEW system can directly predict Peak Ground Acceleration or Intensity for early warning (Hsu et al., 2018). Currently, a regional EEW system is commonly used because of its ability to provide more reliable earthquake early warning services (Allen & Melgar, 2019; Cremen & Galasso, 2020). The magnitude of the earthquake is a key component in the warning signal, and fast and accurate magnitude estimation is an important subject worth extensive study for a regional EEW system.

Most magnitude prediction algorithms rely on the relationship between magnitude and empirically defined parameters based on the initial few seconds of P-wave (typically 3 s) recorded by a single station, such as the predominant period τ_p (Nakamura,1988), maximum predominant period τ_p^{max} (Allen & Kanamori, 2003), average period τ_c (Kanamori, 2005), peak displacement P_d (Wu & Zhao, 2006) and summation of velocity squares IV^2 (Festa et al., 2008). The successful application of these parameters verified that the initial P-wave contains necessary information on estimating the magnitude (Kanamori, 2005). Based on this understanding, new parameters are continuously proposed in recent years, such as the log-average period τ_{log} (Ziv, 2014), the departure time from P-wave similar growth T_{dp} (Noda & Ellsworth, 2016), wavelet transform period λ_{log} (Atefi et al., 2017) and combining parameter of amplitude and period $S_{d\tau}$ (Wang et al., 2021a). Among these approaches, the one based on P_d and hypocentral distance is regarded as the best for estimating magnitude (Kuyuk & Allen, 2013; Nazeri et al., 2017; Kohler et al., 2018; Leyton et al., 2018; Wang et al., 2020c), and the approach based on P_d is also applied in many EEW systems (Kamigaichi et al., 2009; Hsiao et al., 2009; Hoshiba et al., 2011; Chen et al., 2015; Zhang et al., 2016; Chung et al., 2019; Kohler et al., 2020; Chamoli et al., 2021). A common issue in these parameters is that they are empirically defined, which could only use partial information in the initial seismic wave, thus limiting their magnitude estimation performance.

* Corresponding author at: CEAKJ ADPRHexa, Inc, Shaoguan 512000, China
E-mail address: zifa@iem.ac.cn (Z. Wang).

https://doi.org/10.1016/j.gr.2022.06.009
1342-937X/

As observed in other domains in the areas of image, speech recognition and natural language processing, when deep learning methods have been used to replace the manual features selection, the above-mentioned shortcoming is addressed and significant progress has been made (Lecun et al., 2015). The achievements by deep learning in those areas have attracted the attention of seismologists, and the deep learning approach has been applied to identify wave phases with much better accuracy than by human eyes (Perol et al.,2018; Ross et al., 2018; Zhu & Beroza, 2019; Li et al., 2018; Wang et al., 2021b). Deep learning is an advanced machine learning approach, and the biggest advantage of this approach is that it can automatically learn features from data instead of manual processing (Lecun et al., 2015). Deep learning has also been applied in estimation of magnitude in the last two years to directly predict magnitude based on seismic waves (Lomax et al., 2019; Mousavi & Beroza, 2020; van den Ende & Ampuero, 2020; Kuang et al., 2021; Münchmeyer et al., 2021; Zhang, et al., 2021). They can be categorized into three types of deep learning for magnitude prediction based on the difference of data input. The first is to estimate earthquake magnitude based on the input from a single station (SS-input-DP) (Lomax et al., 2019; Mousavi & Beroza, 2020), the second is to estimate the magnitude based on multiple stations (12 stations or more) (MS-input-DP) (van den Ende & Ampuero, 2020; Kuang et al., 2021; Münchmeyer et al., 2021; Zhang, et al., 2021), and the third is to use multiple parameters (up to 12 parameters such as P_d, τ_c, IV^2) from a single station as the input (Zhu et al., 2021). Among these three approaches, SS-input-DP and MS-input-DP automatically extract features from the data without manually picking features. MS-input-DP has realized real time estimation of magnitude when there is only 0.5 s P-wave data with better results than those by traditional methods based on empirical parameters using data from a single station (Münchmeyer et al., 2021), but there are still substantial challenges in applying this approach in a practical EEW system. Since speed of magnitude estimate is vital for an EEW system, it often uses the data from the first 1–4 stations to estimate magnitude (Nakamura et al., 2011; Hoshiba et al., 2011; Zhang et al., 2016; Cochran et al., 2017), but MS-input-DP needs data from multiple stations, which delays the information delivery process (Münchmeyer et al., 2021; Zhang, et al., 2021). Furthermore, MS-Input-DP couples the data from multiple stations to estimate magnitude, and if there is disturbance or noise in the data from any station, the impact on the estimation is unknown while magnitude estimated by a single station can be used to check against each other to eliminate erroneous estimate. To satisfy the efficiency and accuracy requirement in an EEW system, the magnitude estimate approach has to be able to simultaneously estimate magnitude using data from either a single station or multiple stations. The proposed SS-input-DP uses at least 30 s of data or the whole record to estimate the magnitude (Lomax et al., 2019; Mousavi & Beroza, 2020), and it cannot satisfy the time efficiency requirement of an EEW system. Therefore, there is a need to develop an approach that only uses the first few seconds of data which satisfies both the time efficiency and accuracy requirements.

In this paper we constructed a deep learning model (EEWNet) based on CNN (Convolutional Neural Network) that can estimate magnitude using initial less than 3 s P-wave from a single station. CNN is a multi-layer feedforward neural network based on convolutional computation, which can automatically learn features from data. In the proposed EEWNet approach, the initial P-wave is used as the input to estimate the peak displacement for the whole record which is then used to calculate the magnitude. A large number of borehole underground records from Japan were used to evaluate the accuracy of the proposed approach, and the result was compared with that by the widely used P_d approach where the parameter is empirically defined. The comparison demonstrates that EEWNet has better accuracy and time efficiency than P_d.

2. Data and methods

2.1. Data sets

In this study, we used a data set that consists of 30,756 accelerograms recorded by 688 strong motion borehole sensors of the Kiban Kyoshin network (KiK-net) (National Research Institute for Earth Science and Disaster Resilience, 2019), with magnitude between 4 and 9 (3,621 earthquakes) and hypocentral distance between 25 km and 200 km and at least 3 s of wave. The data set includes records from offshore earthquakes. All accelerograms were resampled at 100 Hz. Two selection criteria were used to guarantee the quality of these accelerograms for accuracy of magnitude estimation (Kuyuk, et al., 2014). One was the peak ground acceleration over 2 Gal (Cheng et al., 2014), and the other was the signal to noise ratio (SNR) of the records is greater than 10 (Ziv, 2014; Perol et al., 2018). We manually picked the P-wave arrival time to ensure accuracy. Each vertical accelerogram was labeled for the logarithm of the maximum displacement in the horizontal components. The data sets were then divided into subsets of training, validation and testing according to time sequence. This data division method not only implements the concept of using past data to predict future trends, but also avoids the issue of the records from the same event being included in different data sets. The training and validation data sets were used for training and validating of EEWNet. The test data set was used to assess the performance of the trained EEWNet. The three data sets contain the following information: 2,147 events ($4 \leq M \leq 9$) from 10/1997 to 12/2011 were used as the training data set (Fig. 1(a)), 701 events ($4 \leq M \leq 7$) from 01/2012 to 12/2014 were used as the validation data set (Fig. 1(b)), and 773 events ($4.0 \leq M \leq 7.4$) from 01/2015 to 03/2019 were used as the test data set (Fig. 1(c)).

2.2. EEWNet

The architecture of EEWNet is developed based on one-dimensional CNN, which is used to deal with the regression of a sequential data set. The architecture of EEWNet is composed of an input, multiple hidden layers, a fully connected layer, and an output, as shown in Fig. 2. The input is the initial vertical component of P-wave, and the output is the logarithm of the maximum horizontal compound displacement for the whole record. EEWNet was trained by the training data set through optimization of the loss function defined as mean square error of output values. The hidden layer consists of convolution operation, activation function and pooling operation for automatic feature extraction from the input. In the hidden layer, the process of feature extraction is as follows: firstly, the features are filtered from the input by the convolution operation, then the activation function adds nonlinear characteristics to the features, and finally the pooling operation (subsampling) retains the valuable features to complete the feature extraction. By concatenating multiple hidden layers, iterative feature extraction is realized, and the most relevant features to the output are finally obtained from the input. To avoid overfitting Dropout randomly discards some features (Hinton et al., 2012; Srivastava et al., 2014). The fully connected layer performs regression calculation on the features to obtain the output. There are many hyperparameters in hidden layer that need to be set manually, such as filter size, stride, and padding (Supplementary Fig. S1 for detailed explanation). These hyperparameters determine the performance of the neural network, however, the optimal hyperpa-

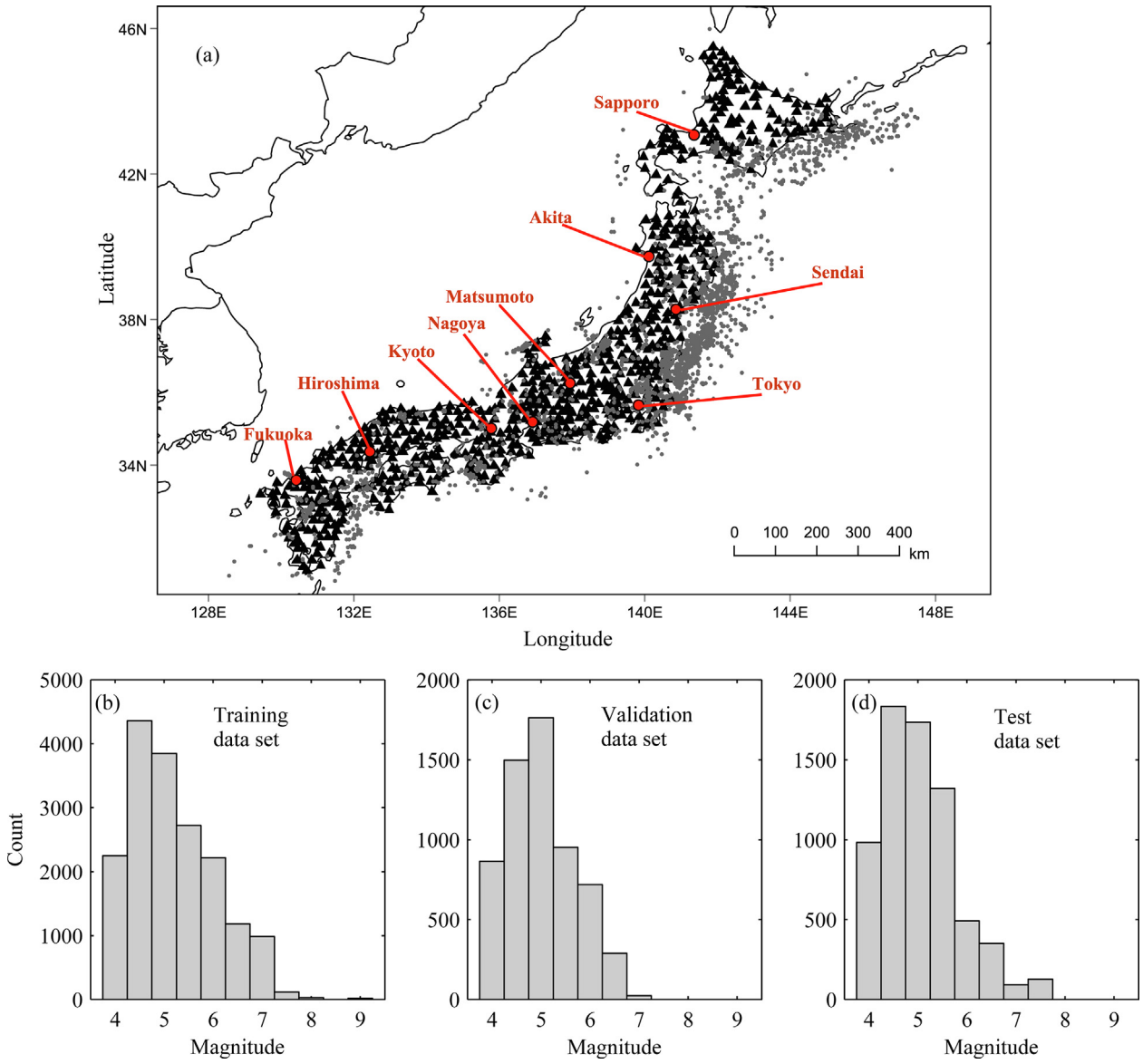

Fig. 1. Training, validation, and test data sets. (a) The distribution of earthquakes (grey points) and stations (black triangles). (b) Training data set consists of 17,717 accelerograms. (c) Validation data set contains 6,106 accelerograms. (d) Test data set has 6,933 accelerograms.

rameter settings can only be determined empirically after repeated trials with the training and validation data sets (trial and error). Take the setting of keep probability of dropout as an example, the keep probability = 1 means no dropout and the low keep probability indicates many discarded features. In trial calculation, the keep probability of dropout was set to 0.1, 0.3, 0.5 and 0.8 respectively. For each value, the neural network was trained by the training data set, then the validation data set was used to validate the performance of the trained neural network, and finally the value with minimum loss function value of validation data set was selected (Supplementary Fig. S2). After many times of training and validation as explained for the case of dropout keep probability, the other hyperparameters of the EEWNet were finally determined. The architecture of the EEWNet is determined by the input length and is explained as follows. For 2^N or fewer samples of input, the total number of hidden layers is N, and each hidden layer is composed of a standard convolution operation consisting of 2^{L+3} filters and a pooling operation. L is the serial number of hidden layers which can be 1, 2, 3, . . ., N. For each filter, the kernel size is 2, stride is 1 and padding is 'same'. The activation function of the convolutional layer uses rectified linear unit (ReLU) (Nair & Hinton, 2010). For each pooling operation, the method is'max pooling', size is 2 and stride is 2. The keep probability of dropout behind the last pooling operation uses 0.5. The size of fully connected layer is $2^{N+3} \times 1$. The Adam stochastic optimization algorithm (Kingma & Ba, 2015) is used for the optimization with a learning rate of 1e − 4. The batch size and epochs can be adjusted according to computer memory and computational convergence, respectively. Here each batch consists of 512 accelerograms and epochs are 200. The architecture of EEWNet has four characteristics: the number of layer and filter size are adaptive to input length; the length of input time series data can be less than 2^N, but 2^N is recommended; each feature map in the last hidden layer is a value (1×1 dimension); and there are no normalization and regularization operations for the input. EEWNet repeats convolution and pooling operations to achieve desired performance. For this study, EEWNet was

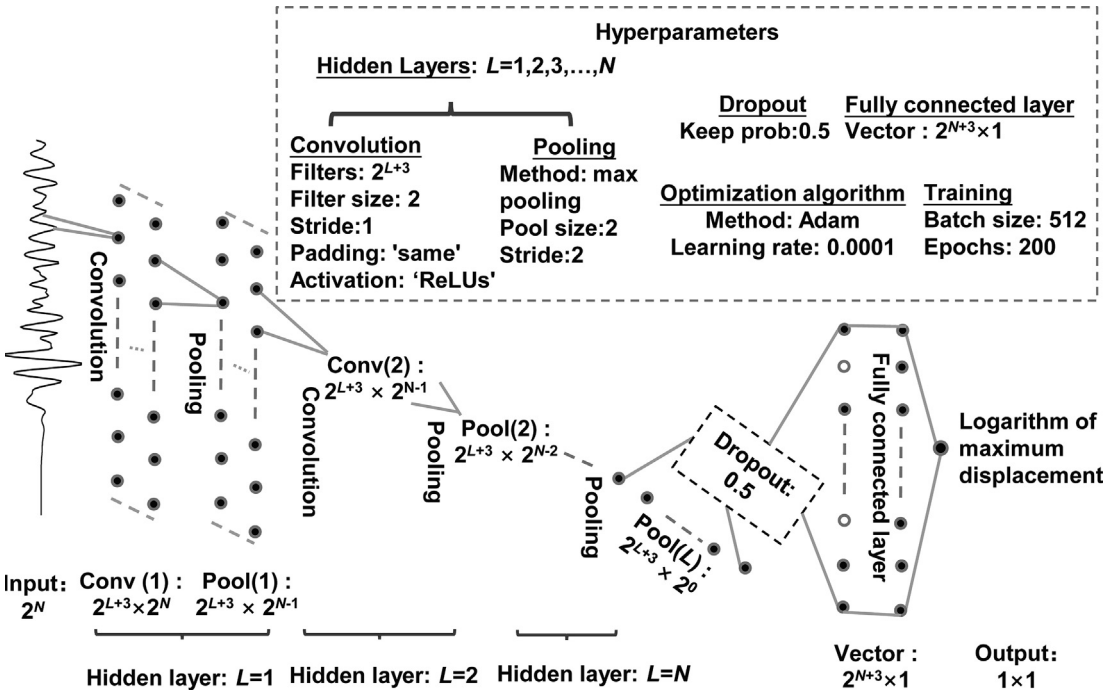

Fig. 2. The architecture and hyperparameters of EEWNet.

programmed with TensorFlow GPU 2.1, and the source code examples of EEWNet with 256 samples and 512 samples are shown in Supplementary Fig. S3 and Supplementary Fig. S4 as examples.

2.3. Magnitude prediction

According to the Richter magnitude scale (Richter, 1935) and magnitude of KiK-net given by Japan Meteorological Agency (JMA) (Doi, 2014), there is an empirical formula to estimate magnitude using surface deformation amplitude and hypocentral distance as defined in the following equation.

$$\log_{10}(A) = a \times M + b \times \log_{10}(R) + c \quad (1)$$

where A is the maximum horizontal compound displacement (μm) for the whole record ($A = \sqrt{A_N^2 + A_E^2}$, A_N and A_E are the maximum displacement amplitude of north–south component and east–west component, respectively), M is magnitude, R is hypocentral distance, a, b and c are fitting coefficients. In EEW, magnitude is predicted by using a formula like equation (1), but A is calculated using the initial few seconds of vertical P-wave instead of using the complete record, as in using P_d (Wu & Zhao, 2006). In this study, initial few seconds of vertical accelerogram is used as the input to EEWNet to predict $\log_{10}(A)$, and then magnitude can be calculated via equation (1).

3. Results

3.1. Predicting magnitude with initial 3s of P-wave

To evaluate the performance of EEWNet, P_d was also used to predict magnitude for comparison. The reason for choosing P_d for comparison is that regional EEW systems in several countries or regions use P_d to estimate the magnitude, such as Japan (Kamigaichi et al., 2009; Hoshiba et al., 2011), the United States (Chung et al., 2019; Kohler et al., 2020), northern India (Chamoli et al., 2021), China's Fujian (Zhang et al., 2016), and Taiwan (Hsiao et al., 2009; Chen et al., 2015). In addition, P_d approach is more accurate than other traditional parametric methods for estimating the magnitude (Kuyuk & Allen, 2013; Nazeri et al., 2017; Kohler et al., 2018; Leyton et al., 2018; Wang et al., 2020c). Since the introduction of the theoretical analysis of point source model (Kanamori, 2005), several studies (Wu & Zhao, 2006; Ziv, 2014; Huang et al., 2015; Wang et al., 2020c) and regional EEW systems (Chen et al., 2015; Zhang et al., 2016; Armando et al., 2018; Chamoli et al., 2021) had shown that the magnitude estimation can start from the first 3 s P-wave, therefore the first 3 s (300 samples) P-waves were used for training, validation, and testing of EEWNet to derive results to compare with those by P_d. The architecture of EEWNet was determined by N = 9 (the input length of EEWNet is 2^9). The fitting coefficients a, b and c of $\log_{10}(P_d)$ and $\log_{10}(A)$ were estimated based on the training and validation data sets (Fig. 3). From the empirical coefficient of the source distance (Fig. 3), it can be inferred that the estimated magnitude of EEWNet is less affected by the hypocentral distance deviation than that of P_d, because the estimated magnitude error of EEWNet caused by hypocentral distance error (ΔR) is $0.7589^*\log_{10}[(R + \Delta R)/R]$, while that for P_d is $1.0023^*\log_{10}[(R + \Delta R)/R]$.

The results of test data set were used for comparison. $\log_{10}(A_{pre})$ predicted by EEWNet is linear with the true $\log_{10}(A_{true})$ (correlation coefficient is 0.83) (Fig. 4 (a)), and 77.12% of its relative errors is within ± 20% (Fig. 4 (b)). Magnitude prediction errors of P_d and EEWNet with 3 s P-wave are shown in Fig. 5 (magnitude was calculated by the predicted $\log_{10}(A_{pre})$ using equation (1)). EEWNet's prediction error has smaller mean μ (mean for P_d is 0.28 while −0.02 for EEWNet) and standard deviation σ than those by P_d (σ for P_d is 0.67, while it for EEWNet is 0.43) (Fig. 5 (a) and 5 (b)). The average absolute error of EEWNet is significantly smaller than that by P_d (Fig. 5(e)). In addition, according to Fig. 5 (e), Fig. 5 (a) and Fig. 5 (b), for magnitude between 6.7 and 7.4, P_d's estimate has a bias toward smaller estimation of the magnitude, while the bias of EEWNet is not obvious. This bias also appears in the results

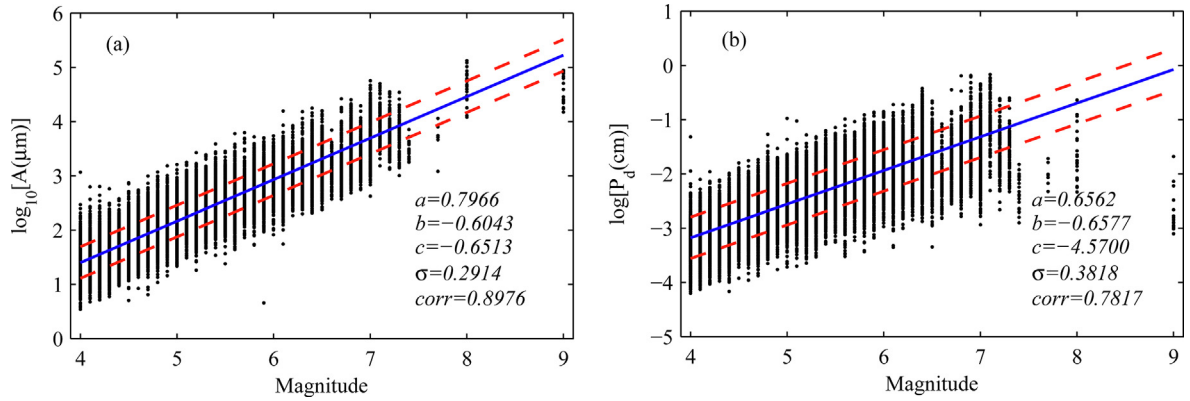

Fig. 3. Empirical equations of A and P_d. *A* was calculated by the horizontal components for the whole record of training and validation data sets. P_d was calculated by the initial 3 s of the vertical component in the training and validation data sets. Each black point indicates a record. The blue solid line is the best fit of empirical equation (1) and the red dashed lines are mean ± σ (standard deviation). *a*, *b* and *c* are the fitting coefficients of equation (1). *corr* is the correlation coefficient between parameter and magnitude. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

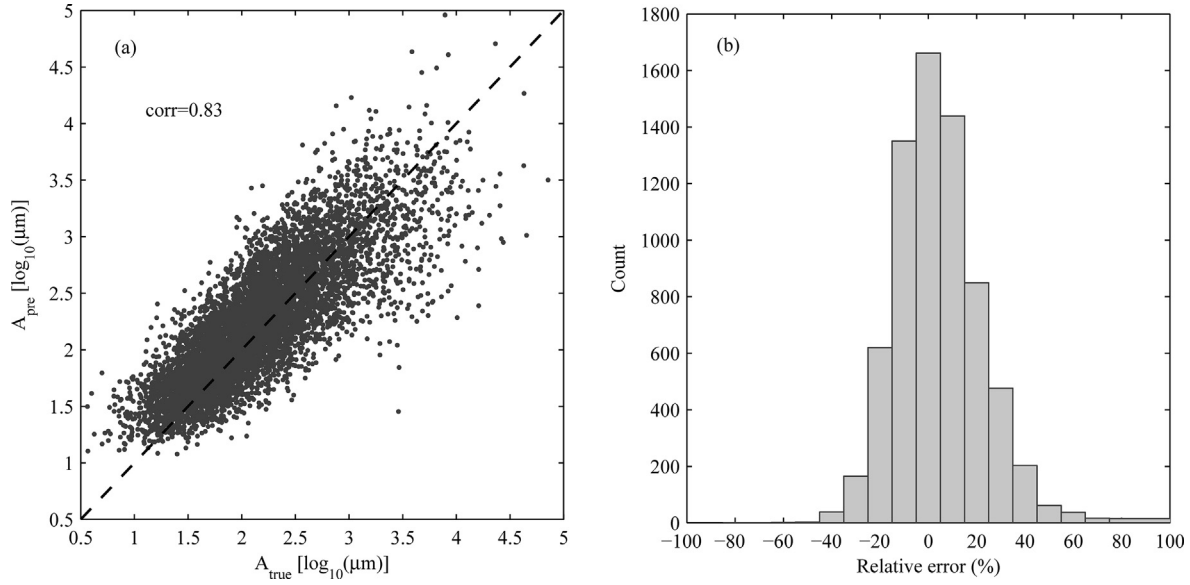

Fig. 4. Predicted maximum displacement versus true maximum displacement (a) and distribution of relative errors (b). Each black point indicates a record. A_{pre} is the maximum displacement predicted by EEWNet using initial 3 s P-wave. The dashed line is a 1: 1 reference line. A_{true} is the true maximum horizontal displacement for the whole record. *corr* is the correlation coefficient between predicted maximum displacement (A_{pre}) and true maximum displacement (A_{true}).

from many other similar studies, and it is considered to be the effect of magnitude saturation (Kanamori 2005, Wu & Zhao, 2006; Zollo et al., 2006). The distribution of P_d's prediction error is not uniform at different hypocentral distances (Fig. 5(c)), the estimated magnitude of P_d is large when the hypocentral distance is small, and the estimated magnitude of P_d is small when the hypocentral distance is large, while the distribution of EEWNet's prediction error is more uniform at different hypocentral distances (Fig. 5(d)).

3.2. Predicting magnitude with initial 0.5 s – 2.0 s of P-wave

In EEW systems, magnitude estimation must be performed quickly after the P-wave arrival, and the estimate must be continuously updated. To evaluate the timeliness of EEWNet with different lengths of initial P-wave, the first 0.5 s (50 samples), 1.0 s (100 samples), 1.5 s (150 samples) and 2.0 s (200 samples) of P-wave were used as the inputs to EEWNet, whose architecture was determined by input length of 64, 128, 256, and 256, respectively.

The prediction errors of the test data set as a function of magnitude with different length of P-wave were shown in Fig. 6. When the duration of P-wave was 0.5 s (Fig. 6(a)), EEWNet's estimation error for magnitude between 4.0 and 5.5 was relatively small (most of them were between (-1, 1)), but the estimate for magnitude between 5.6 and 7.4 was smaller than the actual one (magnitude saturation). The magnitude saturation is gradually improved with increasing duration of P-wave. When the duration of P-wave was 2.0 s (Fig. 6(f)), EEWNet's magnitude saturation was only visible for magnitude between 6.7 and 7.4. In addition, the standard deviation σ of the estimation error for EEWNet decreased from 0.54 to 0.43 with increasing duration (0.5 s – 2 s) of P-wave, and it was always smaller than that by P_d when its duration of P-wave was 3 s. For the initial 0.5 s – 2 s P-wave, the distribution of EEWNet's prediction error is uniform for different hypocentral distances (Fig. 6 (c), (d), (g), and (h)).

Average absolute error was used to further compare the accuracy of EEWNet for different durations of P-wave against that by P_d with duration of 3 s. As shown in Fig. 7, the average absolute

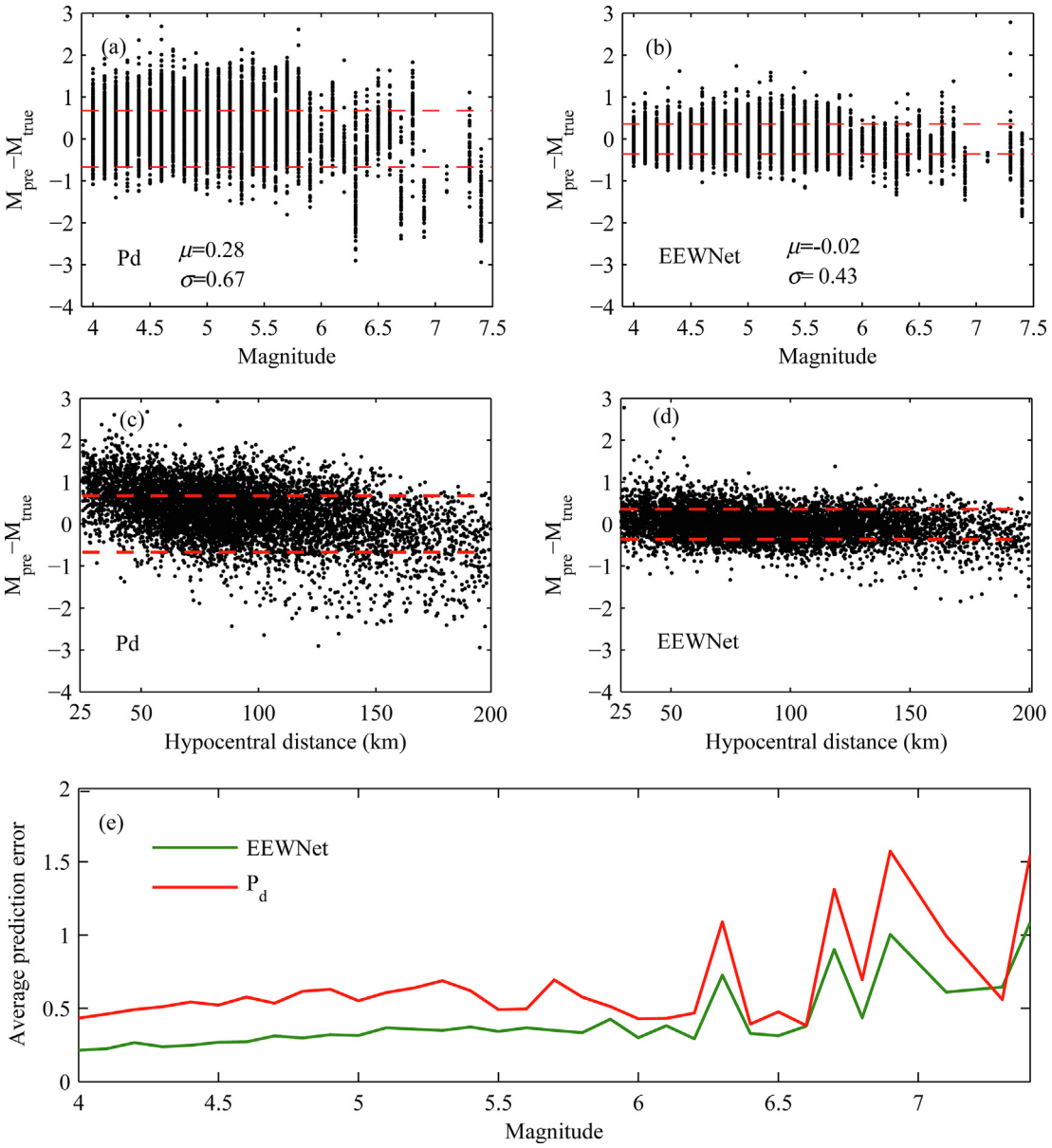

Fig. 5. Magnitude prediction errors of P_d and EEWNet using initial 3 s P-wave. (a) Error as a function of magnitude for P_d. (b) Error as a function of magnitude for EEWNet. (c) Error as a function of hypocentral distance for P_d. (d) Error as a function of hypocentral distance for EEWNet. (e) Average absolute error for P_d and EEWNet. M_{pre} is predicted magnitude and M_{true} is the actual magnitude. μ and σ are mean and standard deviation of errors. The black point indicates the error of each prediction, and the red dashed lines are for 0 ± σ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

error by EEWNet decreases with increasing P-wave duration. For magnitude between 4.0 and 5.9, the average absolute error of EEWNet with P-wave duration of 0.5 s is smaller than the error by P_d with P-wave duration of 3 s. For magnitude between 4.0 and 7.1, the average absolute error of EEWNet with P-wave duration of 2 s is smaller than it by P_d with P-wave duration of 3 s.

4. Discussion

In this study, we have proposed a deep learning model EEWNet to improve magnitude prediction of EEW. A large number of borehole underground records from the Japanese KiK-net were used to compare the magnitude estimated by P_d and EEWNet. When 3 s of P-wave were used, the standard deviation and mean of prediction error, and the average absolute error of EEWNet were all smaller than the corresponding values by P_d, demonstrating the better prediction accuracy of EEWNet, and the effect of hypocentral distance on the estimated magnitude of EEWNet is much smaller than that of P_d. In addition, when the initial P-wave duration is less than 3 s, the EEWNet result still has better accuracy than that by P_d using 3 s of P-wave. For example, for magnitude between 4 and 5.9, EEWNet using 0.5 s of P-wave still produced a better estimation than that by P_d using 3 s of P-wave. For magnitude between 4.0 and 7.1, the prediction error of EEWNet using 2 s of P-wave was smaller than that by P_d with a P-wave duration of 3 s. This demonstrates that EEWNet has a better time efficiency, that is, it can predict magnitude 1 s – 2.5 s ahead of the approach by P_d (calculated on a graphics card NVIDIA Quadro P4000, EEWNet predicts the peak displacement in about 0.0004 s). The fundamental reason why

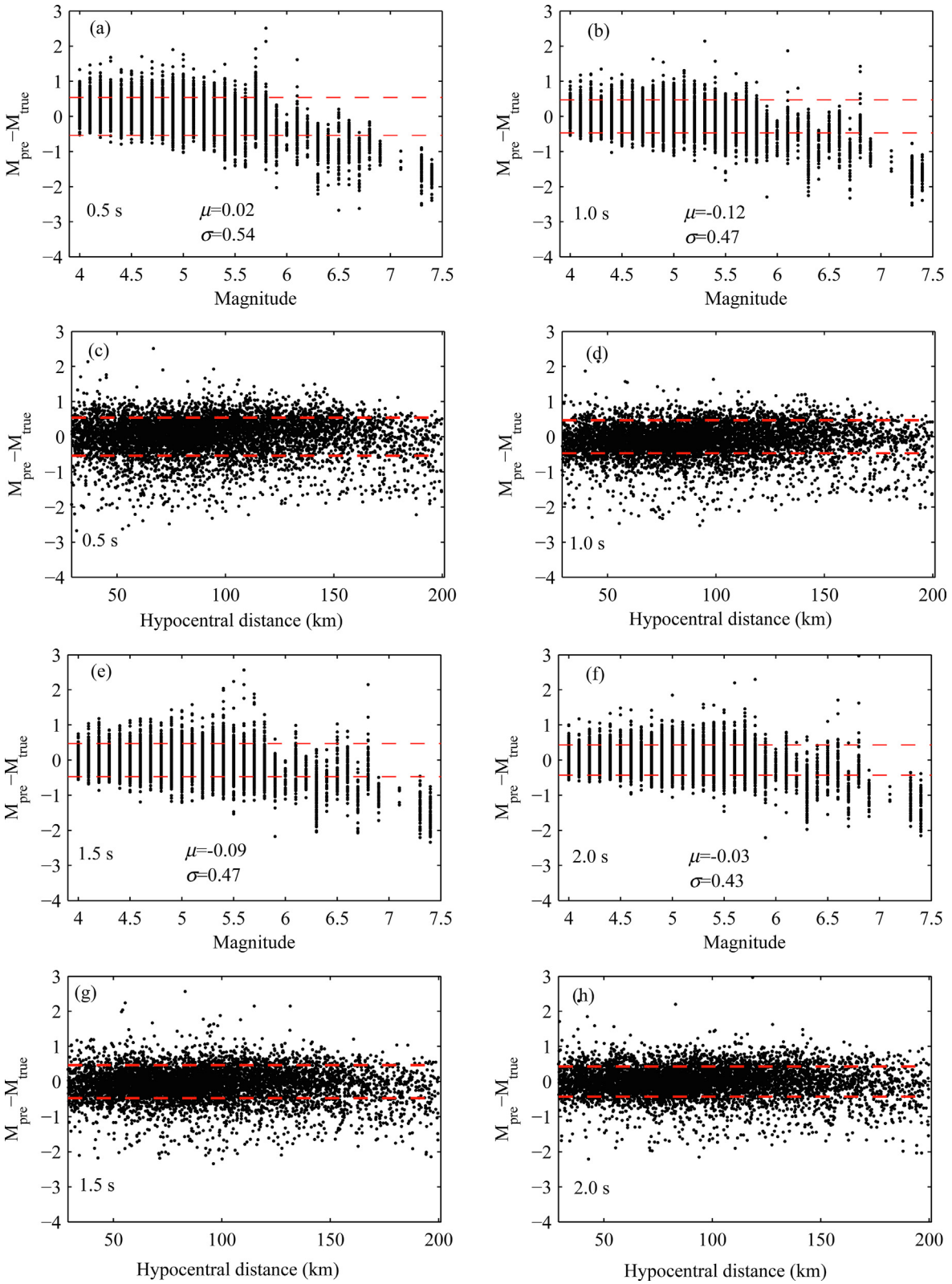

Fig. 6. Magnitude prediction errors of EEWNet with initial 0.5 s − 2 s P-wave. The black points indicate the errors of each prediction, and the red dashed lines are the 0 ± σ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

EEWNet has these advantages over P_d is that it retrieves more information relating to magnitude prediction than the information found based on human experience.

Comparing to the widely used approach of P_d in today's EEW systems (Kamigaichi et al., 2009; Hsiao et al., 2009; Zhang, et al., 2016; Kohler et al., 2018), EEWNet has demonstrated clear

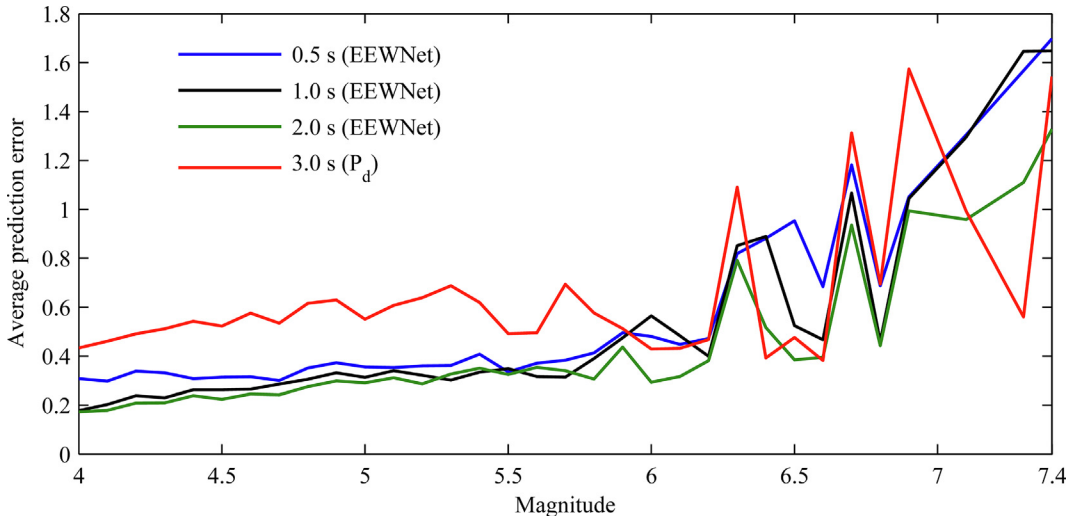

Fig. 7. Average absolute error for P_d and EEWNet with different length of P-wave.

advantages in accuracy and speed, and it can be expected that EEWNet will enhance the efficiency of the current EEW systems. First, as the accuracy of magnitude estimate increases, the impact area can be better predicted, and thus reducing false warnings of EEW systems. Second, with the speed improvement of magnitude estimate, more lead time can be given to service areas thus more preparation efforts can be organized.

Magnitude saturation is a big challenge in EEW, and it is also a much debated subject in the academic community. A few researchers believe that the short duration of P-wave used in the estimate contains limited information on seismic rupturing process, and the magnitude cannot be accurately predicted before the rupturing completion (Meier et al., 2016; Goldberg et al., 2019; Ide, 2019). Other researchers found that there is weak correlation between initial rupturing and the final rupture (Olson & Allen, 2005; Colombelli et al., 2014; Goldberg et al.,2018; Melgar & Hayes, 2019). In traditional EEW systems, longer duration of seismic waves is used to improve the magnitude estimation (Ozaki, 2011; Hoshiba et al, 2011; Colombelli et al., 2012; Goldberg et al., 2018; Trugman et al., 2019). Deep learning approach has been found by Münchmeyer et al. (2021) to reduce the magnitude saturation effect by increasing records with bigger magnitude in the training data set. In the test result of EEWNet, when the initial duration is increased to 3 s from 0.5 s, the magnitude saturation effect is significantly reduced. Therefore, adding records of large magnitude and increasing the duration of input seismic data can help improve the effect of reducing magnitude saturation in machine learning based approaches.

Comparing to the other two SS-input-DP methods proposed by Lomax et al. (2019) and Mousavi & Beroza (2020), EEWNet is different in the following four areas. 1) Those two methods require at least 30 s or the complete record as the input while EEWNet only needs the initial 0.5 s – 3 s of vertical P-wave as its input. 2) Although hypocentral distance is required by EEWNet, it is easily available in any EEW systems. EEWNet can also be applied in quick P-wave arrival picking (Wang et al., 2021b) and the arrival time from multiple locations can be used to estimate the hypocenter. 3) EEWNet can adjust its architecture based on the input length while the other two methods do not adjust architecture for different data length. The flexible adjustment suits well for an EEW system because data is constantly fed into the system with variable duration. Simpler architecture should be used in the initial periods when there are only limited arriving P-wave data points. 4) There is difference in the network model architecture although CNN is used for all. EEWNet's receptive field in the neuron in the fully connected layer uses the whole input (Luo et al., 2016) to better reflect the feature in the arriving waves. The above differences make EEWNet unique in estimating magnitude with better accuracy and has the potential to be improved for further application.

Comparing to the MS-input-DP method, TEAM-LM proposed by Münchmeyer et al. (2021) uses data from multiple stations while EEWNet uses data from a single station. In addition, trained EEWNet uses a fixed wave duration while trained TEAM-LM can use flexible wave duration, but this does not limit the use of EEWNet because EEWNet can be trained beforehand to accommodate different wave duration at pre-defined intervals.

Even though EEWNet and other deep learning approaches can extract magnitude features directly from the seismic waves, deep learning application is still at its infancy stage and follow-up study is needed to improve the algorithm for accuracy and efficient implementation. First, the model design (architecture and its parameters) of deep learning is based on empirical experience, and this is unfavorable to improve the magnitude estimation, so further study needs to be carried out to optimize the model design. Second, deep learning adopts a black-box approach, and future study is needed to understand what features EEWNet extracts to find out whether they can be visualized and what physical meaning they represent. Third, the issue of more records of smaller events than those of larger ones needs further study to understand the impact of data set inhomogeneity on the training results. Fourth, using fully connected layers for regression calculation may not be optimal for EEWNet, and how to combine EEWNet and other machine learning methods to improve magnitude estimation, especially machine learning methods that have achieved remarkable results in the field of geotechnical engineering in the recent years (Wang et al., 2020a; Wang et al., 2020b; Zhang et al., 2020; Tang & Na, 2021). These studies are necessary for future improvement of deep learning in addressing many challenges in seismology and earthquake engineering.

5. Conclusions

We have proposed a new approach, EEWNet, based on convolutional neural network to predict magnitude for EEW. Unlike the magnitude estimation methods with human-defined characteristic parameters, EEWNet automatically extracts features from the initial P-wave to estimate earthquake. Extensive test and comparison

based on a large number of earthquake records demonstrate that EEWNet estimates the magnitude using the initial 0.5 s – 3 s of data better than the P_d approach using the first 3 s of data for magnitude between 4 and 7.4. This implies that EEWNet can predict earthquake magnitude 1 s – 2.5 s ahead of P_d with better accuracy and time efficiency. The result proves that deep learning can extract magnitude related features from the first seconds P-wave more effectively than humans. Deep learning has shown its great potential to be applied in EEW and other fields of seismology.

CRediT authorship contribution statement

Yanwei Wang: Conceptualization, Methodology, Software, Writing – original draft. **Xiaojun Li:** Writing – review & editing, Investigation. **Zifa Wang:** Conceptualization, Writing – original draft. **Juan Liu:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are extremely grateful to the anonymous reviewers for their careful review and thoughtful suggestions. We would like to thank KiK-net online database for the recorded data (last accessed in March 2019). This research has been supported by the National Natural Science Foundation of China (Grant No. 51968016 and No. 51978634) and the Guangxi Key Laboratory of Geomechanics and Geotechnical Engineering (Guikeneng-19Y-21-8).

Data availability statement

The data used is publicly available and can be downloaded from NIED KiK-net (https://www.kyoshin.bosai.go.jp/).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gr.2022.06.009.

References

Allen, R.M., Kanamori, H., 2003. The potential for earthquake early warning in Southern California. Science 300 (5620), 786–789.
Allen, R.M., Melgar, D., 2019. Earthquake early warning: Advances, scientific challenges, and societal needs. Annu. Rev. Earth Planet. Sci. 47 (1), 361–388.
Armando, C., Gerardo, S., Espinosa-Aranda, J.M., 2018. A Fast Earthquake Early Warning Algorithm Based on the First 3 s of the P-Wave Coda. Bull. Seismol. Soc. Am. 108 (4), 2068–2079.
Atefi, S., Heidari, R., Mirzaei, N., Siahkoohi, H.R., 2017. Rapid Estimation of Earthquake Magnitude by a New Wavelet-Based Proxy. Seismol. Res. Lett. 88 (6), 1527–1533.
Chamoli, B.P., Kumar, A., Chen, D.-Y., Gairola, A., Jakka, R.S., Pandey, B., Kumar, P., Rathore, G., 2021. A prototype earthquake early warning system for northern India. J. Earthqu. Eng. 25 (12), 2455–2473.
Cheng, M.H., Wu, S., Heaton, T.H., Beck, J.L., 2014. Earthquake early warning application to buildings. Eng. Struct. 60, 155–164.
Chen, D.Y., Hsiao, N.C., Wu, Y.M., 2015. The Earthworm based earthquake alarm reporting system in Taiwan. Bull. Seismol. Soc. Amer. 105 (2A), 568–579.
Chung, A.I., Henson, I., Allen, R.M., 2019. Optimizing earthquake early warning performance: ElarmS-3. Seismol. Res. Lett. 90 (2A), 727–743.
Cochran, E.S., Kohler, M.D., Given, D.D., Guiwits, S., Andrews, J., Meier, M.A., Ahmad, M., Henson, I., Harog, R., Smith, D., 2017. Earthquake Early Warning ShakeAlert System: Testing and Certification Platform. Seismol. Res. Lett. 89, 108–117.
Colombelli, S., Zollo, A., Festa, G., Kanamori, H., 2012. Early magnitude and potential damage zone estimates for the great Mw 9 Tohoku-Oki earthquake. Geophys. Res. Lett. 39 (22). L22306.
Colombelli, S., Zollo, A., Festa, G., Picozzi, M., 2014. Evidence for a difference in rupture initiation between small and large earthquakes. Nat. Commun. 5 (1), 1–5.
Cremen, G., Galasso, C., 2020. Earthquake early warning: Recent advances and perspectives. Earth-Sci. Rev. 205, 103184.
Doi, K., 2014. Seismic Network and Routine Data Processing-Japan Meteorological Agency. Summary of the Bulletin of the International Seismological Centre 47 (7–12), 25–42.
Festa, G., Zollo, A., Lancieri, M., 2008. Earthquake magnitude estimation from early radiated energy. Geophys. Res. Lett. 35, L22307.
Goldberg, D.E., Melgar, D., Bock, Y., 2019. Seismogeodetic P-wave amplitude: No evidence for strong determinism. Geophys. Res. Lett. 46 (20), 11118–11126.
Goldberg, D.E., Melgar, D., Bock, Y., Allen, R.M., 2018. Geodetic observations of weak determinism in rupture evolution of large earthquakes. J. Geophys. Res.-Solid Earth 123 (11), 9950–9962.
Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv, 1207.0580.
Hoshiba, M., Iwakiri, K., Hayashimoto, N., Shimoyama, T., Hirano, K., Yamada, Y., Ishigaki, Y., Kikuta, H., 2011. Outline of the 2011 off the pacific coast of Tohoku Earthquake (Mw 9.0) —earthquake early warning and observed seismic intensity—. Earth Planets Space 63 (7), 547–551.
Hsiao, N.C., Wu, Y.M., Shin, T.C., Zhao, L., Teng, T.L., 2009. Development of earthquake early warning system in Taiwan. Geophys. Res. Lett. 36 (5), L00B02.
Hsu, T.Y., Lin, P.Y., Wang, H.H., Chiang, H.W., Chang, Y.W., et al., 2018. Comparing the performance of the NEEWS earthquake early warning system against the CWB system during the 6 February 2018 Mw 6.2 Hualien earthquake. Geophys. Res. Lett. 45, 6001–6007.
Ide, S., 2019. Frequent observations of identical onsets of large and small earthquakes. Nature 573 (7772), 112–116.
Kamigaichi, O., Saito, M., Doi, K., Matsumori, T., Tsukada, S., Takeda, K., Shimoyama, T., Nakamura, K., Kiyomoto, M., Watanabe, Y., 2009. Earthquake early warning in Japan: Warning the general public and future prospects. Seismol. Res. Lett. 80 (5), 717–726.
Kanamori, H., 2005. Real-time seismology and earthquake damage mitigation. Annu. Rev. Earth Planet. Sci. 33 (1), 195–214.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv,1412.6980.
Kohler, M.D., Cochran, E.S., Given, D., Guiwits, S., Neuhauser, D., Henson, I., Hartog, R., Bodin, P., Kress, V., Thompson, S., Felizardo, C., Brody, J., Bhadha, R., Schwarz, S., 2018. Earthquake early warning ShakeAlert system: West coast wide production prototype. Seismol. Res. Lett. 89 (1), 99–107.
Kohler, M.D., Smith, D.E., Andrews, J., Chung, A.I., Hartog, R., Henson, I., Given, D.D., de Groot, R., Guiwits, S., 2020. Earthquake early warning ShakeAlert 2.0: Public rollout. Seismol. Res. Lett. 91 (3), 1763–1775.
Kuang, W., Yuan, C., Zhang, J., 2021. Network-Based Earthquake Magnitude Determination via Deep Learning. Seismol. Res. Lett. 92, 2245–2254.
Kuyuk, H.S., Allen, R.M., 2013. A global approach to provide magnitude estimates for earthquake early warning alerts. Geophys. Res. Lett. 40 (24), 6329–6333.
Kuyuk, H.S., Allen, R.M., Brown, H., Hellweg, M., Henson, I., Neuhauser, D., 2014. Designing a network-based earthquake early warning algorithm for California: ElarmS-2. Bull. Seismol. Soc. Amer. 104 (1), 162–173.
Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Leyton, F., Ruiz, S., Baez, J.C., Meneses, G., Madariaga, R., 2018. How fast can we reliably estimate the magnitude of subduction earthquakes? Geophys. Res. Lett. 45 (18), 9633–9641.
Li, Z., Meier, M.A., Hauksson, E., Zhan, Z., Andrews, J., 2018. Machine learning seismic wave discrimination: Application to earthquake early warning. Geophys. Res. Lett. 45 (10), 4773–4779.
Lomax, A., Michelini, A., Jozinovic, D., 2019. An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. Seismol. Res. Lett. 90 (2A), 517–529.
Luo, W., Li, Y., Urtasun, R., & Zemel, R., 2016. Understanding the effective receptive field in deep convolutional neural networks. In Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain, 4905-4913.
Meier, M.A., Heaton, T., Clinton, J., 2016. Evidence for universal earthquake rupture initiation behavior. Geophys. Res. Lett. 43 (15), 7991–7996.
Melgar, D., Hayes, G.P., 2019. Characterizing large earthquakes before rupture is complete. Sci. Adv. 5 (5), eaav2032.
Mousavi, S.M., & Beroza, G.C., 2020. A machine-learning approach for earthquake magnitude estimation. Geophys. Res. Lett. 47(1), e2019GL085976.
Münchmeyer, J., Bindi, D., Leser, U., Tilmann, F., 2021. Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. Geophys. J. Int. 226 (2), 1086–1104.
Nair, V., & Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10). Haifa, Israel, 807-814.
Nakamura, Y., 1988. On the urgent earthquake detection and alarm system (UrEDAS). Proceedings of 9th World Conference on Earthquake Engineering. Tokyo, Japan, 673-678.
Nakamura, Y., Saita, J., Sato, T., 2011. On an earthquake early warning system (eew) and its applications. Soil Dyn. Earthq. Eng. 31 (2), 127–136.
National Research Institute for Earth Science and Disaster Resilience, 2019. NIED K-NET, KiK-net, National Research Institute for Earth Science and Disaster Resilience. doi: 10.17598/NIED.0004

Nazeri, S., Shomali, Z.H., Colombelli, S., Elia, L., Zollo, A., 2017. Magnitude estimation based on integrated amplitude and frequency content of the initial P wave in earthquake early warning applied to Tehran. Iran. Bull. Seismol. Soc. Amer. 107 (3), 1432–1438.
Noda, S., Ellsworth, W.L., 2016. Scaling relation between earthquake magnitude and the departure time from P wave similar growth. Geophys. Res. Lett. 43 (17), 9053–9060.
Olson, E.L., Allen, R.M., 2005. The deterministic nature of earthquake rupture. Nature 438 (7065), 212–215.
Ozaki, T., 2011. Outline of the 2011 off the Pacific coast of Tohoku Earthquake (M w 9.0). Earth Planets Space 63 (7), 827–830.
Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4, (2) e1700578.
Richter, C.F., 1935. An instrumental earthquake magnitude scale. Bull. Seismol. Soc. Amer. 25 (1), 1–32.
Ross, Z.E., Meier, M.A., Hauksson, E., 2018. P-wave arrival picking and first-motion polarity determination with deep learning. J. Geophys. Res.-Solid Earth 123 (6), 5120–5129.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.
Strauss, J.A., Allen, R.M., 2016. Benefits and costs of earthquake early warning. Seismol. Res. Lett. 87 (3), 765–772.
Tang, L., Na, S.H., 2021. Comparison of machine learning methods for ground settlement prediction with different tunneling datasets. J. Rock Mech. Geotech. Eng. 13 (6), 1274–1289.
Trugman, D.T., Page, M.T., Minson, S.E., Cochran, E.S., 2019. Peak ground displacement saturates exactly when expected: Implications for earthquake early warning. J. Geophys. Res.-Solid Earth 124 (5), 4642–4653.
van den Ende, M.P., Ampuero, J.P., 2020. Automated seismic source characterization using Deep Graph Neural Networks. Geophys. Res. Lett. 47(17), e2020GL088690.
Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., Zhang, W., 2020a. Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. Bull. Eng. Geol. Environ. 79 (6), 2763–2775.
Wang, L., Wu, C., Tang, L., Zhang, W., Lacasse, S., Liu, H., Gao, L., 2020b. Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. Acta Geotech. 15 (11), 3135–3150.
Wang, Y., Li, X., Cao, Z., Lan, J., Liu, J., 2020c. Continuous estimation magnitude for earthquake early warning based on Kik-net borehole bedrock strong motions. Earthq. Eng. Eng. Dyn. 40 (04), 42–52. in Chinese.
Wang, Y., Li, X., Li, L., Wang, Z., Lan, J., 2021a. New Magnitude Proxy for Earthquake Early Warning Based on Initial Time Series and Frequency. Seismol. Res. Lett. 93, 216–225.
Wang, Y., Li, X., Wang, Z., Shi, J., Bao, E., 2021b. Deep learning for P-wave arrival picking in earthquake early warning. Earthq. Eng. Eng. Vib. 20 (2), 391–402.
Wu, Y.M., Zhao, L., 2006. Magnitude estimation using the first three seconds p-wave amplitude in earthquake early warning. Geophys. Res. Lett. 33, L16312.
Zhang, H., Jin, X., Wei, Y., Li, J., Kang, L., Wang, S., Huang, L., Yu, P., 2016. An earthquake early warning system in Fujian. China. Bull. Seismol. Soc. Amer. 106 (2), 755–765.
Zhang, W., Wu, C., Zhong, H., Li, Y., Wang, L., 2020. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. 12 (1), 469–477.
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. Geophys. Res. Lett. 48(5), 2020GL089394.
Zhu, J., Li, S., Song, J., Wang, Y., 2021. Magnitude Estimation for Earthquake Early Warning Using a Deep Convolutional Neural Network. Front. Earth Sci. 9, 653226.
Zhu, W., Beroza, G.C., 2019. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. Geophys. J. Int. 216 (1), 261–273.
Ziv, A., 2014. New frequency-based real-time magnitude proxy for earthquake early warning. Geophys. Res. Lett. 41 (20), 7035–7040.
Zollo, A., Lancieri, M., Nielsen, S., 2006. Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records. Geophys. Res. Lett. 33 (23), L23312.