Artificial Intelligence in Geosciences 4 (2023) 39–46

Contents lists available at ScienceDirect

Artificial Intelligence in Geosciences

journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences

Toward earthquake early warning: A convolutional neural network for repaid earthquake magnitude estimation

Fanchun Meng, Tao Ren *, Zhenxian Liu, Zhida Zhong

Software College of Northeastern University, Shen Yang, China

A R T I C L E I N F O

Keywords:
Massive tectonic earthquakes
Earthquake early warning
Magnitude classification
DenseBlock
Multi-head attention

A B S T R A C T

Earthquake early warning (EEW) is one of the important tools to reduce the hazard of earthquakes. In contemporary seismology, EEW is typically transformed into a fast classification of earthquake magnitude, i.e., large magnitude earthquakes that require warning are in the positive category and vice versa in the negative category. However, the current standard information signal processing routines for magnitude fast classification are time-consuming and vulnerable to data imbalance. Therefore, in this study, Deep Learning (DL) algorithms are introduced to assist with EEW. For the three-component seismic waveform record of 7 s obtained from the China Earthquake Network Center (CENC), this paper proposes a DL model (EEWMagNet), which accomplishes the extraction of spatial and temporal features through DenseBlock with Bottleneck and Multi-Head Attention. Extensive experiments on Chinese field data demonstrate that the proposed model performs well in the fast classification of magnitude. Moreover, the comparison experiments demonstrate that the epicenter distance information is indispensable, and the normalization has a negative effect on the model to capture accurate amplitude information.

1. Introduction

Influenced by two major seismic zones in the world, the Asia-Europe Seismic Zone and the Pacific Rim Seismic Zone, China is an earthquake-prone country. EEW has been proven effective in reducing earthquake hazards (Kamigaichi et al., 2009; Doi 2011; Suárez et al. 2009). EEW includes source location warning, intensity warning, and magnitude warning. The magnitude warning is not required to obtain an exact magnitude, but to determine whether the magnitude is greater than a threshold.

In recent years, many scholars in the seismology have been conducting research in the EEW. The excellent period method, proposed by Nakamura (1988) and developed by Allen and Kanamori (2003), is an empirical method for estimating the magnitude based on data from 3 to 4 s after the P-wave. However, after conducting experimental tests, many seismologists have found that this method can indicate the magnitude to some extent, but its dispersion is still large. Kanamori (2005) used the wavelet analysis method to perform multi-scale time--frequency decomposition of seismic data, thus improving the excellent period method to some extent. However, the method still suffers from a relatively large degree of dispersion. Lancieri and Zollo (2007) proposed a magnitude warning method, the peak ground motion displacement Pd method, which uses the displacement maximum Pd at 2S after the P wave and the displacement maximum Sd at 1S after the S-wave to predict the magnitude. This method can indicate the general pattern of earthquake magnitude, but it still has the problem of relatively large dispersion. As can be seen from the magnitude warning methods described above, Although the current methods can roughly calculate the magnitude of seismic events, they generally suffer from a large dispersion, which leads to low accuracy of the results. Even if there are methods with high accuracy, they suffer from a lack of generalizability.

With the accumulation of earthquake data, DL techniques are widely adopted in the EEW (Kong et al., 2019). Hu and Zhang (2020) designed and trained a DL model for magnitude range prediction in EEW, which shows a decrease in the overall error and variance of the prediction results compared to the traditional methods. Lin et al. (2021) construct a CNN network magnitude prediction model using 3-s seismic waveform data based on CNN networks, which transformed the magnitude determination problem in EEW into a magnitude classification problem. Experimental results show that more than 80% of the errors are within ± 0.3. Furthermore, Münchmeyer et al. propose a DL model called TEAM-LM based on the CNN and Transformer model in 2021

* Corresponding author. No.195, Chuangxin Road, Hunnan District, Shenyang, 110169, China.
E-mail address: chinarentao@163.com (T. Ren).

https://doi.org/10.1016/j.aiig.2023.03.001
Received 20 December 2022; Received in revised form 19 March 2023; Accepted 19 March 2023
Available online 30 March 2023

(Münchmeyer et al., 2021). TEAM-LM is a multi-task model that can simultaneously estimate the magnitude and location. Since the TEAM-LM model relies on 10 s of data for magnitude estimation, it is also applied to EEW.

This study proposes a DL model EEWMagNet to address the problems of low generalizability and low accuracy in previous studies, which can distinguish whether the seismic event is greater than magnitude 4 by using single-station seismic waveform data within 7 s after the P-wave. EEWMagNet consists of a convolutional layer for feature extraction, a Transition layer for dimensionality reduction and a Multi-Head Attention layer for time-series feature extraction. In addition, we incorporate the epicentral distance feature in the network. After extensive experiments, EEWMagNet is able to perform the EEW task excellently, and the accuracy of classification reaches over 90%.

2. Data

2.1. Data selection

The data adopted in this study are three-component waveform data with a sampling frequency of 100 HZ, recorded by CENC from 1104 stations in China during the period 2009–2017. The distribution of stations and data is shown in Fig. 1. As seismic waves propagate, the energy attenuation is accelerated as the propagation distance increases, which leads to a decrease in the quality of waveform characteristics (Jing et al. 2005). Therefore, the seismic waveform with epicenter distance shorter than 200 km are selected. The waveforms may have broken traces and overlaps due to network communication anomalies and equipment failures, so obspy (Beyreuther et al., 2010) is adopted to refilter the data in this study.

Since earthquake data with ML $\geq$ 4 are much less than those with ML $<$ 4 in the data recorded by CENC. In order to reduce the impact of data imbalance, this study downsamples the other earthquake magnitude ranges by the amount of data above magnitude 4, so that each magnitude range has an equal amount of data. The dataset is divided into a training set and a test set according to a ratio close to 8:2. The distributions of the training and test sets are shown in Table 1.

Research on EEW has shown that if the earthquake warning time is 3 s, the casualty rate can be reduced by 14%; if the time is 10 s, the casualty rate can be reduced by 39%; and if the time is 60 s, the casualty rate can be reduced by 95% (Chen et al., 2008). Therefore, this experiment is conducted to identify the most cost-effective input length by trying different input data lengths within 10 s separately, adopting Accuracy, Precision, Recall, and F1 Score as indicators in the process. As shown in Fig. 2.

The test results are shown in Fig. 2. When the input length is 7 s or less, the model performance increases significantly with the increase of input length, taking the accuracy as an example, from 81% to 91%. The increase in data length allows the model to extract more features that contribute to the magnitude estimation. However, with the continuous increase of input length, the performance of the model fluctuates and even declines. This is because too long data fails to provide more key features and instead introduces more invalid data such as noise. In addition, when the input length is only 7 s, the highest recall, i.e., the least number of misclassifications, is achieved with an accuracy rate higher than 90%, achieving the best balance of time and effectiveness. Therefore, this study decided to choose 7 s as the most cost-effective input length for the EEWMagNet model. The data are intercepted as shown in Fig. 3.

Table 1
Distributions of Dataset. Waveform distributions for the imbalanced dataset provided by CENC are included, as well as distributions after downsampling.

Magnitude range	Waveforms in CENC dataset	Waveforms in downsampled dataset	
		training set	test set
1–2	214963	3012	1005
2–3	91206	3012	1005
3–4	14080	3012	1005
≥4	4017	3012	1005

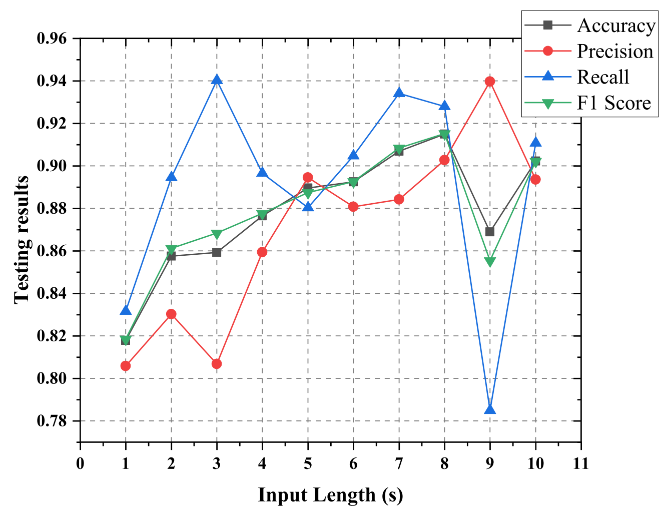
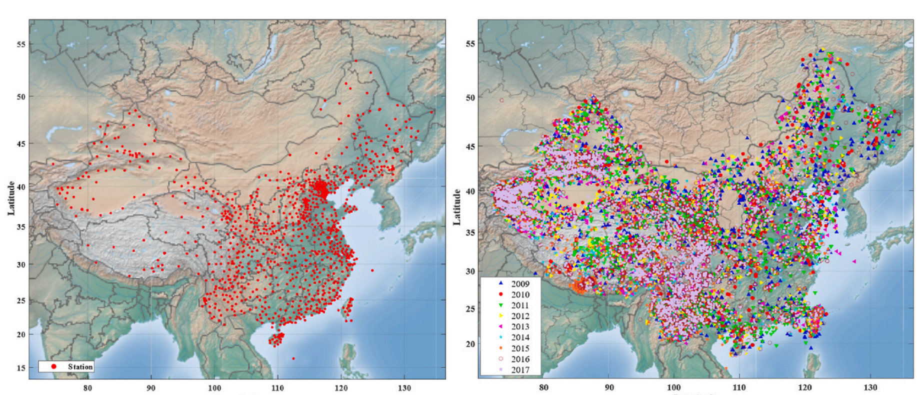


Fig. 2. Testing results of different input lengths.



(a) Distribution of stations

(b) Data distribution, where the different colors and symbols represent the different years

Fig. 1. Spatial distribution of the network of seismic stations in China.

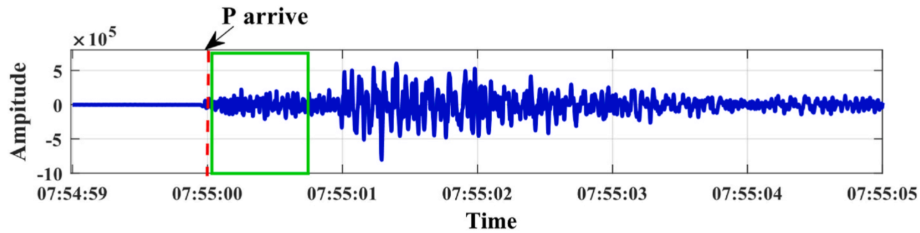


Fig. 3. The experimental intercept in this study. The y-axis represents the amplitude of the waveform, and the x-axis represents the sampling point with the frequency of 100HZ. The first arrival of the P-wave is marked with a red line and the data in the green rectangular box is intercepted for the experiment. The waveform in the figure is from the seismic event recorded at 100HZ in in March 2017 by Sichuan Network.

2.2. Data augmentation

In this study, the filtered dataset lacks sufficient data to maintain the performance of DL. Given that DL is data-driven, the lack of sufficient data may lead to model overfitting and low generalizability. Therefore, this study adopts a data augmentation to expand the data of ML $\geq$ 4. Data augmentation consists of various techniques to generate new training samples based on the collected datasets, expanding the size and variety of the training samples. The data augmentation technique is proven to prevent model from overfitting and dramatically improve the results by making small adjustments to the data (Shorten and Khoshgoftaar 2019). In this study, slip and superimposed noise are adopted to generate new data.

First, this study performs a small-scale sliding of the waveform data, i.e., moving the data interception frame forward by 25 or 50 samples, as shown in Fig. 4 (a). The sliding operation not only generates new data, but also enables the model to be less sensitive to location information (Zhang et al., 2022). In order to ensure that the new data are sufficiently different from the original data, and to increase the number of available data. Different decibel noise data from the same station are superimposed on the sliding data in this study (Dong et al., 2021). As shown in Fig. 4 (b). The distribution of the data after augmentation is shown in Table 2.

Table 2
Data distribution after data augmentation.

Dataset	Magnitude range (ML)	waveforms in the Dataset
Training set	$\geq$4	9036
	$<$4	9036
Test set	$\geq$4	1005
	$<$4	3015

3. Analysis and development of the method

The data available for EEW are short in length and low in information, therefore, adequate feature extraction and feature fusion become the key to achieving accurate warning. In this study, DenseBlock (Huang et al., 2017) is adopted as the main part of the feature extraction structure. The calculation process of DenseBlock is shown in equation (1). Where X_L is the output of the Lth layer and H denotes the nonlinear transformation function of the Lth layer. Although DenseBlock is able to fuse the shallower features with the deeper ones, overcoming the problem that traditional DL ignores the shallow features in the forward propagation process. However, dimensionality of input of DenseBlock increases as network layers deepens. Therefore, this study reduces the computational cost of the feature extraction part by adding Bottleneck module. The final DenseBlock structure with Bottleneck adopted in this study is shown in Fig. 5.

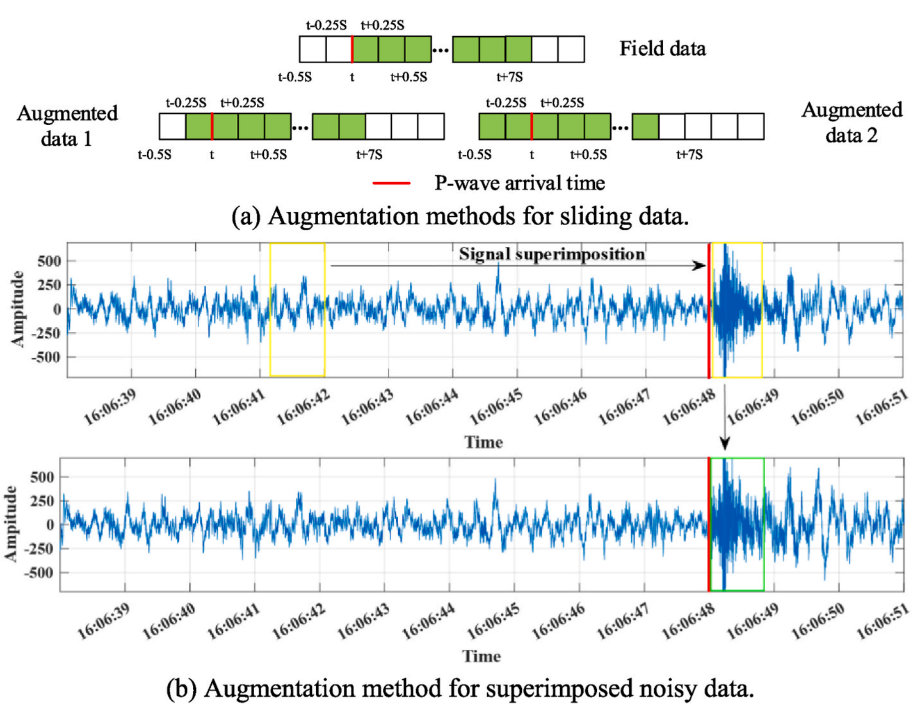


Fig. 4. An illustration of the data augmentation method. (a) represents the sliding augmentation method, where the green box represents the selected data, and the original data is slid to obtain the new data. (b) represents the superimposed noise augmentation method, where the yellow box in the figure represents the pre-selected noise and event waveforms, as well as the green box represents the new data generated by superimposing the pre-selected data. The waveform in the figure is from the seismic event recorded at 100 HZ in February 2015 by Xinjiang Network.

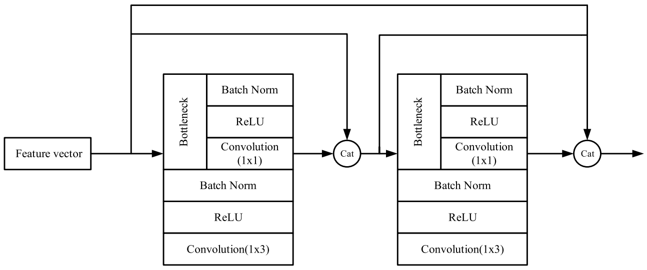


Fig. 5. An illustration of the Bottleneck module.

$$xL = H([x0, x1, \ldots, xL-1]) \tag{1}$$

As can be seen in Fig. 5, the Bottleneck layer consists of the convolutional layer with kernel size 1 × 1, the BatchNorm layer (BN), and the ReLU layer. The role of the convolution layer is to reduce the dimensionality of the feature map. The role of the BN layer is to normalize the data, thus speeding up the convergence of the model and avoid gradient explosion and overfitting (Ioffe and Christian, 2015). The ReLU is a widely adopted nonlinear activation function.

Given that the sampling frequency of the data adopted in this study is 100 HZ, the size of the 7 s three-component waveform data has reached 3 × 700. In order to reduce the computation of the EEWMagNet and to alleviate overfitting, a Transition module is added among the DenseBlocks. The Transition module can fuse and compress the feature maps output from the DenseBlock. In this study, the channel fusion compression ratio of the Transition module is set as 0.5, i.e., the number of output channels is half of the input.

Although DenseBlock is effective in extracting features, it is difficult to make features correlated in the time dimension. Therefore, extracting the temporal dimensional correlation of feature vectors is crucial to improve the effectiveness of EEWMagNet network. Recurrent Neural Networks (RNN) networks are usually adopted for processing temporal data, but the problem of short-term memory in RNN networks makes it difficult to model longer feature sequences temporally. Therefore, the Multi-Head Attention (Vaswani et al., 2017) structure is introduced, which fully considers the correlation of each feature with all other features. The computation process of this structure consists of matrix multiplication, so it can realize parallelized computation, while the RNN model can only be computed serially. the computation process of Multi-Head Attention is shown in equations (2)–(6).

$$Q_i = X \times w_q^i \tag{2}$$

$$K_i = X \times w_k^i \tag{3}$$

$$V_i = X \times w_v^i \tag{4}$$

$$o_i = \text{softmax}\left(\frac{Q_i k_i^T}{\sqrt{d_k}}\right) \times V_i \tag{5}$$

$$O = \text{Concat}(O_1, O_2, \ldots, O_i) w_o \tag{6}$$

i represents the computation process of the ith Head, where X is the input feature sequence. W_q, W_k and W_v are three weight matrices, and each Head has its own separate W_q, W_k and W_v weight matrices. Once the output value O_i of each Head is calculated, the Concat function in equation (6) combines all O_i values and later multiplies them with another weight matrix W_o to achieve different levels of attention for each Head, and finally obtains the result O. In this study Head is set as 8.

In this study, the EEWMagNet model is designed and implemented for EEW. As shown in Fig. 6, the EEWMagNet model mainly consists of four parts: a denseblock layer with a bottleneck, a transition layer, a Multi-Head Attention layer, and an epicenter distance fusion layer. The number of convolutional layers in the four DenseBlock layers is set as 6, 12, 24, and 16, and the number of convolutional kernels in each convolutional layer is set as 32. Based on the number of convolutional layers and the number of convolutional kernels, the number of feature channels output by the four DenseBlock modules in the EEWMagNet network are 256, 512, 1024, and 1024. In addition, the epicenter distance contributes to the magnitude classification task, so the epicenter distance is combined with the output of the average pooling layer after the flattening operation in this study. Finally, the EEWMagNet uses two fully connected layers to perform classification calculations on the feature vectors of the final output of the model.

4. Experimentations

This section conducts experiments to verify the effectiveness of the EEWMagNet model for the magnitude warning classification task. Subsections 4.1 and 4.2 we verify the effect of seismic data normalization and epicenter distance features on the classification of magnitude warnings. Subsection 4.3 demonstrates the effect of the Multi-Head Attention structure on the classification of magnitude warnings.

The environment configuration for the experiments is shown in Table 3.

The problem studied in this section is a binary classification problem,

Table 3
Environment configuration.

Designation	Version
GPU	NVIDIA RTX 3090 24G
CPU	Intel Xeon Silver 4210 * 2
RAM	64G
Hard disk	4T
Programming language	Python 3.8
DL Framework	Pytorch 1.10
Operating system	Ubuntu 18.04

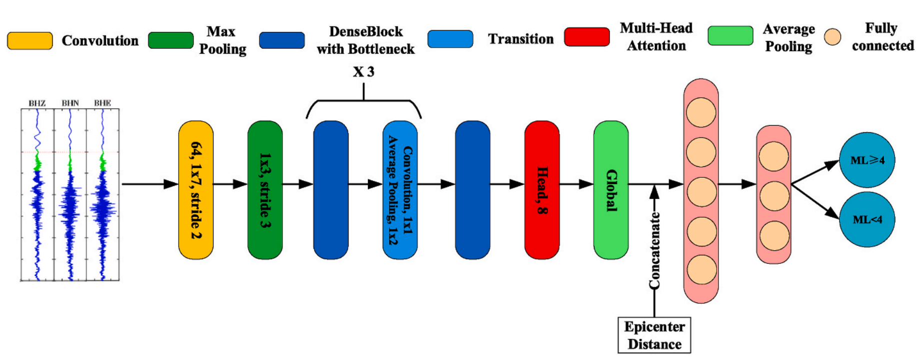


Fig. 6. Structure of EEWMagNet.

so four evaluation metrics, Accuracy, Precision, Recall, and F1 Score, are used in this section. F1 Score is a comprehensive metric used in statistics to measure binary classification and is a balanced average of accuracy and recall. For EEW, it is important to make accurate judgments about large-magnitude events, so these evaluation parameters have the concept of positive and negative samples. In this section, a positive sample is defined as a seismic event of magnitude greater than or equal to 4, and a negative sample is defined as a seismic event of magnitude less than 4. The formulae for the evaluation parameters are shown in equations 7–10.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

The optimizer used in this section is Adam and the initial value of the learning rate is set as 0.0005. To reduce the oscillation of the model during training, the learning rate is reduced by 10% after every 5 training rounds, and the change in the learning rate is shown in Fig. 7. The overall number of training rounds in this chapter is 80 and the batch size is set as 64. Finally, during the training of the model, the optimized model from the trained rounds is saved for testing and analysis of the test dataset.

4.1. Normalization of seismic data

This subsection discusses the impact of normalization on the classification task of EEW. Normalization is a common pre-processing method in DL, which can speed up the convergence of the model and ensure stability during training. However, when normalization is applied to seismic data, the maximum and minimum values of the amplitudes are adjusted.

Fig. 8 shows the change curves of Accuracy, Precision, Recall, and F1 Score during the training process for the two data pre-processing used in this experiment. Without normalization, the method converges much faster and outperforms the normalized counterpart in the early stages. Meanwhile, when normalized data is given, the method shows much steeper trend and exceed the accuracy of unnormalized case at around 40 epochs. However, the exceedance in later stages should be interpreted with caution – instead of increased performance, the model is likely be overfitting.

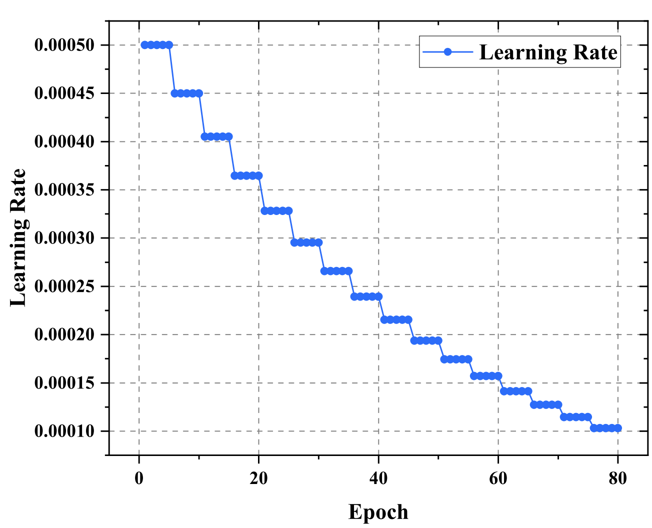


Fig. 7. Learning rate change curve.

The results are shown in Table 4. The method without normalization outperforms the normalized counterpart in all four metrics, i.e., accuracy, precision, recall, and F1 score. The test model used is the best test model for all training rounds. After removing the normalization preprocessing, the classification accuracy of the model improves from 83.18% to 90.23%. In addition, without normalization preprocessing, the test and training results are closer, which demonstrates that normalization preprocessing may lead to overfitting of the model.

4.2. Epicenter distance characteristics

This subsection verifies the impact of epicenter distance features on the classification task of EEW. With a high density of seismic monitoring stations deployed, the results of seismic localization can be obtained within a few seconds after the P-wave, so the results of seismic localization can be used to calculate the epicenter distance when performing EEW classification. Therefore, the EEWMagNet network, which uses the epicenter distance feature, can still be applied to EEW.

Fig. 9 shows the curves of Accuracy, Precision, Recall, and F1 Score of the two models used in this experiment during the training process. It can be seen that the training convergence speed and the effect on the training set of the network with the introduction of the epicenter distance feature are better than those of the network without the epicenter distance.

The results are shown in Table 5. The network with the epicenter distance feature outperformed the network without the epicenter distance feature in all four metrics. The classification accuracy of the model increased from 76.95% to 90.23% with the use of the epicenter distance feature. Seismic waves attenuate during propagation, resulting in degradation of waveform quality. The compensation should be calculated based on the epicenter distance during the magnitude estimation. Therefore, the performance is significantly improved after fusing the epicentral distance feature.

4.3. Multi-Head Attention

This experiment is based on the EEWMagNet network to verify the effect of the Multi-Head Attention structure on the classification of EEW. Two networks were constructed for comparison experiments, i.e., the full EEWMagNet network and the EEWMagNet network without the Multi-Head Attention structure.

Fig. 10 shows the change curves of Accuracy, Precision, Recall and F1 Score of the two networks during the training process. It can be seen that there is little difference in the training convergence speed of the two models on the four metrics, but the complete EEWMagNet network outperforms the EEWMagNet network without the Multi-Head Attention structure on the training set. This suggests that the Multi-Head Attention structure enhances the ability of the EEWMagNet network to fit seismic data.

The results of the test set data are shown in Table 6. The EEWMagNet network performs better in the three metrics of Accuracy, Recall and F1 Score compared to the EEWMagNet network without the Multi-Head Attention structure, except for a slight decrease in the performance of the Precision metric. Particularly, the improvement in the Accuracy metric was about 2.15%. The experimental results show that the incorporation of the Multi-Head Attention structure is beneficial in improving the effectiveness of the model for EEW classification tasks. This is because the Multi-Head Attention structure can fully learn the sequence correlation specific to seismic data, thus making the EEWMagNet network more suitable for processing seismic data.

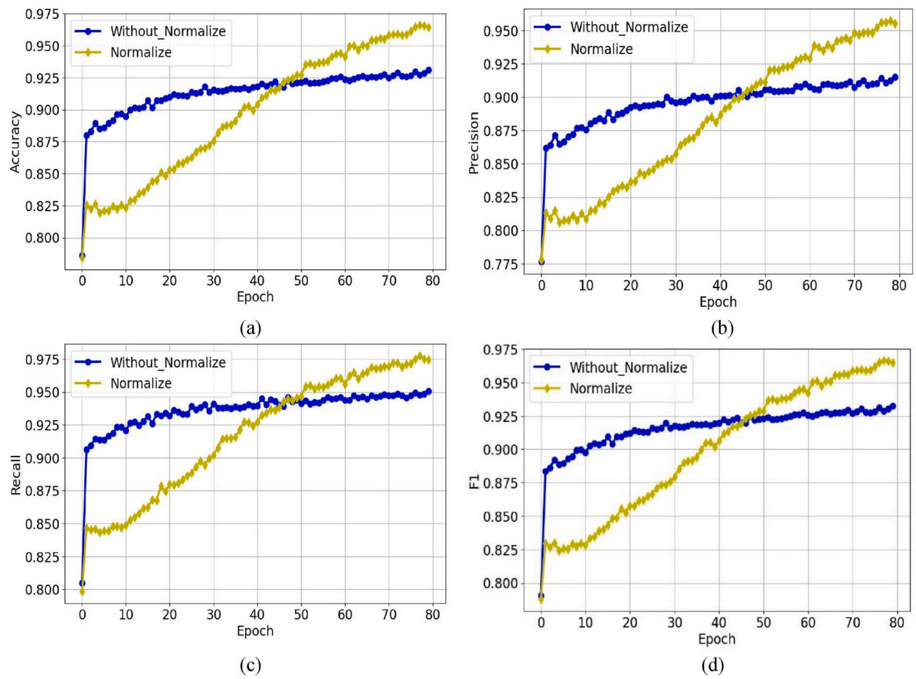


Fig. 8. Training process of Experiment 1 ((a) Accuracy comparison chart (b) Precision comparison chart (c)Recall comparison chart(d) F1 Score comparison chart).

Table 4
Testing results on the effect of normalization.

Method	Accuracy	Precision	Recall	F1 Score
Without_Normalize	**0.9023**	**0.8935**	**0.9108**	**0.9021**
Normalize	0.8318	0.8189	0.8469	0.8326

5. Conclusions

In this study, we used the EEWMagNet model to classify earthquake magnitudes and accomplish earthquake early warning. EEWMagNet utilizes convolutional neural networks for feature extraction from single-station three-component waveform data, and also incorporates epicentral distance data to enhance the network's ability to resolve features.

Table 5
Test results on the effect of epicenter distance.

Model	Accuracy	Precision	Recall	F1 Score
EEWMagNet	**0.9023**	**0.8935**	**0.9108**	**0.9021**
Without_Epic_Dis_EEWMagNet	0.7695	0.763	0.7738	0.7684

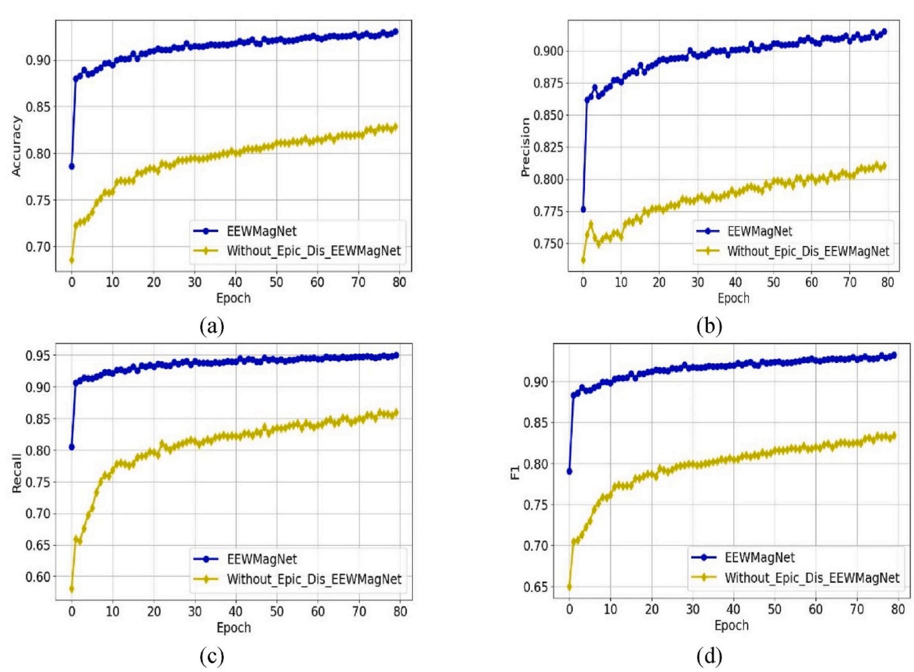


Fig. 9. Training process of Experiment 2 ((a) Accuracy comparison chart (b) Precision comparison chart (c)Recall comparison chart(d) F1 Score comparison chart).

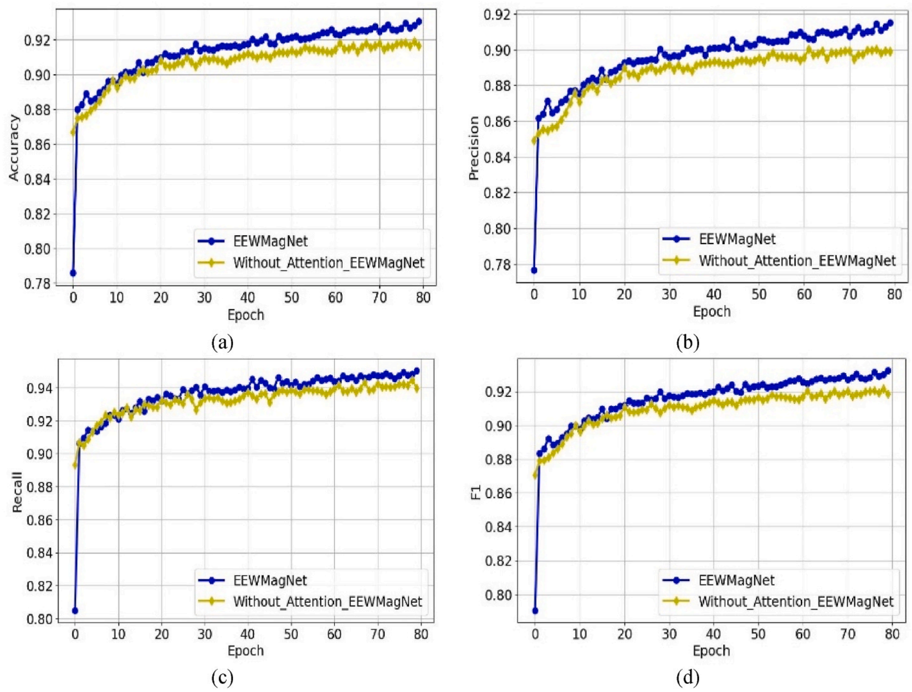


Fig. 10. Training process of Experiment 3 ((a) Accuracy comparison chart (b) Precision comparison chart (c)Recall comparison chart(d) F1 Score comparison chart.

Table 6
Test results on the effect of multi-head attention.

Model	Accuracy	Precision	Recall	F1 Score
EEWMagNet	**0.9023**	0.8935	**0.9108**	**0.9021**
Without_Attention_EEWMagNet	0.8808	**0.9022**	0.8509	0.8758

To ensure classification accuracy, we evaluated the results using various data lengths and model structures. Our findings indicate that EEWMagNet achieves over 90% accuracy and recall, even with only 7 s of data. The extraction of spatial and temporal dimensional features is indispensable in the design of DL models. DenseBlock with Bottleneck and Multi-Head Attention to extract features proved to improve the overall performance by about 3%. Moreover, we verified the impact of data normalization methods and epicentral distance features on performance. The normalization methods may result in the loss of amplitude information from waveforms, leading to an 8% reduction in model accuracy. In contrast, the fusion of epicentral distance features contributed to calculate waveform propagation loss and improved classification accuracy by 18%.

One of the biggest contributions of this study is the development of a deep learning network that can rapidly classify seismic magnitude. In contrast to traditional seismological computational methods, the proposed network, called EEWMagNet, exhibits greater accuracy while utilizing shorter waveform data. Furthermore, our study highlights the significance of epicenter distance information and emphasizes the deleterious effects of normalization on magnitude estimation.

Data and resources

The Chinese dataset is from CENC. Waveform data, metadata, or data products for STEAD can be downloaded from https://data.earthquake.cn/datashare/report.shtml?PAGEID=datasourcelist&dt=ff8080827e3317f4017e331e7ce 00002.

The proposed model is being prepared for deployment at CENC as one of the auxiliary tools for EEW.

Code availability section

The algorithm was developed based on Windows 10, pycharm, Python 3.6, and pytorch 1.10. The application may be downloaded from https://github.com/Fan-Chun- Meng/EEWMagNet. For enquiries contact chinafcmeng@163.com.

Authorship contribution statement

Fanchun Meng: formal analysis, methodology, software, validation, visualization, writing-original draft. Tao Ren: Conceptualization, funding acquisition, resources, supervision, writing – review and edit. Zhenxian Liu: methodology. Zhida Zhong: Methodology, supervision.

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

This work is partially supported by Fundamental Research Funds for the Central Universities (N2217003), Joint Fund of Science & Technology Department of Liaoning Province, and State Key Laboratory of Robotics, China(2020-KF-12-11), National Natural Science Foundation of China (61902057,41774063), Science for Earthquake Resilience (XH21042).

References

Allen, Richard M., Kanamori, Hiroo, 2003. The potential for earthquake early warning in southern California. Science 300, 786–789.

Beyreuther, Moritz, Barsch, Robert, Krischer, Lion, Megies, Tobias, Behr, Yannik, Wassermann, Joachim, 2010. 'ObsPy: a Python toolbox for seismology. Seismol Res. Lett. 81, 530–533.

Chen, Yuntai, Zhang, Guomin, Sun, Qizheng, Sun, Shihong, Chen, Xinlian, Liu, Xihui, Du, Wei, Cao, Xuefeng, Li, Yizhuang, Feng, Yijun, Xiao, Chengye, 2008. 'Terminology of Protecting against and Mitigating Earthquake Disasters—Part 1: Basic Terms GB/T 18207.1-2008', *General Administration Of Quality Supervision,*

Inspection And Quarantine Of the People's Republic of China. Standardization Administration of the People's Republic of China, 2018, pp. 1–3.
Doi, Keiji, 2011. The operation and performance of earthquake early warnings by the Japan Meteorological Agency. Soil Dynam. Earthq. Eng. 31, 119–126.
Dong, Xintong, Wang, Hongzhou, Zhong, Tie, Yue, Li, 2021. 'Desert mixed seismic noise suppression by using multiple forward models and a supervised deep-learning method. Explor. Geophys. 52, 431–445.
Hu, Andong, Zhang, Haiming, 2020. 'Application of machine learning to magnitude estimation in earthquake emergency prediction system. Chin. J. Geophys. 63, 2617–2626.
Huang, Gao, Liu, Zhuang, Van Der Maaten, Laurens, Weinberger, Kilian Q., 2017. 'Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Ioffe, Sergey, Christian, Szegedy, 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift'. In: International conference on machine learning. pmlr, pp. 448–456.
Jing, Liping, Zhuo, Xuyang, Wang, Xiangjian, 2005. 'Effect of complex site on seismic wave propagation. Earthq. Eng. Eng. Vib. 25, 16 *(in Chinese)*.
Kamigaichi, Osamu, Saito, Makoto, Doi, Keiji, Matsumori, Toshiyuki, Tsukada, Shin'ya, Takeda, Kiyoshi, Shimoyama, Toshihiro, Nakamura, Kouji, Kiyomoto, Masashi, Watanabe, Yukihiro, 2009. 'Earthquake early warning in Japan: warning the general public and future prospects. Seismol Res. Lett. 80, 717.
Kanamori, Hiroo, 2005. 'Real-time seismology and earthquake damage mitigation. Annu. Rev. Earth Planet Sci. 33, 195–214.
Kong, Qingkai, Trugman, Daniel T., Ross, Zachary E., Bianco, Michael J., Meade, Brendan J., Gerstoft, Peter, 2019. 'Machine learning in seismology: turning data into insights. Seismol Res. Lett. 90, 3–14.
Lancieri, M., Zollo, A., 2007. A Bayesian Method for the Real Time Estimation of Magnitude From Early P-and S-wave Displacement Peaks. AGU Fall Meeting Abstracts.
Lin, Binhua, Xing, Jin, Kang, Lanchi, Yongxiang, Wei, Li, Jun, Zhang, Yanming, Chen, Huifang, Zhou, Shiwen, 2021. 'The research of earthquake magnitude determination based on Convolutional Neural Networks. Chin. J. Geophys. 64, 3600–3611.
Münchmeyer, Jannes, Bindi, Dino, Leser, Ulf, Tilmann, Frederik, 2021. 'Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. Geophys. J. Int. 226, 1086–1104.
Nakamura, Yutaka, 1988. On the urgent earthquake detection and alarm system (UrEDAS). In: Proc. Of the 9th World Conference on Earthquake Engineering, pp. 673–678.
Shorten, Connor, Khoshgoftaar, Taghi M., 2019. 'A survey on image data augmentation for deep learning. Journal of big data 6, 1–48.
Suárez, Gerardo, Novelo, David, Mansilla, Elizabeth, 2009. 'Performance evaluation of the seismic alert system (SAS) in Mexico City: a seismological and a social perspective. Seismol Res. Lett. 80, 707.
Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Łukasz, Polosukhin, Illia, 2017. 'Attention is all you need. Adv. Neural Inf. Process. Syst. 30.
Zhang, Shou, Ku, Bonhwa, Ko, Hanseok, 2022. 'Learnable maximum amplitude structure for earthquake event classification. Geosci. Rem. Sens. Lett. IEEE 19, 1–5.