IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 1, 2025

EARTHQUAKE PREDICTION USING ATTENTION MECHANISM IN DEEP LEARNING

Kiran Onapakala
Integration software Developer
Pacific Source health plans
Portland, Oregon, USA
kiran.onapakala@pacificsource.com

ABSTRACT
Earthquakes are one of the natural phenomena which have incessantly caused break and loss of human life in olden times. Earthquake prediction is an important aspect of any society's plans and can boost public preparedness and decrease damage to a great extent. Due to the stochastic character of earthquakes and the challenge of achieving an efficient and dependable model for earthquake prediction, efforts have been insufficient thus far, and new methods are required to solve this problem. This paper proposes a novel prediction method based on attention mechanism using Deep learning which can predict the number and maximum magnitude of earthquakes in each area of the region. This model focuses on effective earthquake characteristics and produces more accurate predictions. Firstly, pre-processing on earthquake data set is applied. Secondly, to effectively use spatial information and reduce dimensions of input data, the deep learning algorithm is used to capture the spatial dependencies between earthquake data. Thirdly, RNN is employed to capture the temporal dependencies. Fourthly, the Attention Mechanism layer is introduced to highlight its important features to achieve better prediction performance. The results show that the proposed method has better performance and generalize ability than other prediction methods.

1. Introduction

Earthquakes are one of the most devastating natural disasters in the world, which occur without an explicit warning and may cause serious injuries or loss of human lives. One of effective solutions for reducing earthquakes loss is the earth quake prediction, which aims to use the known earthquake data to specify three elements, namely when, where and the magnitude of the future earthquake. Therefore, effective earthquake prediction can reduce the earthquake damage to a large extent, which is of great significance to the country and society, and there has been an increasing interest and academic research on predicting seismic events.

In summary, the contributions of this paper can be summa rized as follows: • We argue that the feature extraction methods used in previous earthquake prediction methods obtain explicit features by geologists and implicit features by deep learning methods individually, and lack a general model that can combine the advantages of both explicit features and implicit features. • Wepropose anovel deep learning model named DLEP for earthquake prediction. In DLEP, the explicit features and implicit features are combined effectively by a suggested attention-based strategy. Furthermore, a dynamic loss function is also designed for dealing with the category imbalance problem of seismic data. • We evaluate the effectiveness of our model DLEP com paring to state-of-the-art baselines, and the experimental results on eight datasets with different characteristics demonstrate the promising performance of the proposed DLEP, which indicates that the idea of fusing both ex plicit features and implicit features is an effective solution for accurate earthquake prediction.

2. Related work

In the last few decades, many researchers regarded earth quake prediction as a purely geological and physical problem. They tried to discover more effective features and earthquake precursors to predict the future earthquake with the development of physics and geology . For example, Zhang et al [1] recently proposed a precursory pattern-based feature extraction method for earthquake prediction, where the eight mathematical statistic features can be generated as seismic indicators (i.e. the time, mean magnitude, seismic root of seismic energy, b-value, mean square deviation, maximum difference, and coefficient of variation). Compared with dif ferent models, their experimental results on two historical earthquake records demonstrated the effectiveness of their precursory pattern based features with the selected CART algorithm for earth quake prediction. Unfortunately, the performance of these methods is usually limited by the characteristics of seismic zones. For example, the work in [2] predicted the earthquake events in Chile with the magnitude larger than 4.4, while the work in [3] only adopted two zones in China. For other

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

seismic data with differ ent properties, previous methods often need some adjustment or even modify the prediction algorithm. To sum up, these seismic indicators (explicit features) designed by humans have strong interpretability from the theoretical system. However, they may fail to fully utilize information contained in seismic sequences. For this purpose, people hope to discover the plentiful features hiding in seismic data.

Dataset

In this paper, we adopt eight popular seismic zones with different characteristics as our datasets to test the performance of the comparison algorithms.Specifically, the eight zones are Sichuan Province,Xinjiang Province, Qinghai- Tibetv plateau,Shandong- Jiangsu Province,Japan, the Philippines,Chicagoand LosAngeles. Table I gives the main characteristics of eight seismic datasets, including the number of instances,corresponding longitude and latitude of different regions,belonged countries,the range of earthquake categories and the number of instances in each category.For each dataset, we manually divide the magnitude range into five labels. It is noted that the dividing threshold for each label is slightly different for each dataset,with the aim to get the balanced number of instances for each label. Basedon this,we can regard the prediction of earthquake magnitude range as a classification problem[4].

Table 1: Characteristic of eight seismic datasets

Region	Latitude	Longitude	Instances	Countries	Label-1	Label-2	Label-3	Label-4	Label-5	Instance numbers in each label
Sichuan	28-36N	98-106E	906	China	[3.0,4.0)	[4.0,4.5)	[4.5,5.0)	[5.0,5.5)	[5.5,max)	457 / 214 / 114 / 54 / 67
Xinjiang	35-50N	75-95E	1027	China	[3.0,4.0)	[4.0,4.5)	[4.5,5.0)	[5.0,5.5)	[5.5,max)	123 / 335 / 316 / 142 / 111
Qinghai-Tibet	26-39N	73-104E	1021	China	[3.0,4.0)	[4.0,4.5)	[4.5,5.0)	[5.0,5.5)	[5.5,max)	172 / 208 / 230 / 157 / 254
Shadong-Jiangsu	29-38N	114-124E	658	China	[3.0,4.0)	[4.0,4.5)	[4.5,5.0)	[5.0,5.5)	[5.5,max)	481 / 100 / 49 / 15 / 13
Japan	31-38N	136-143E	1096	Japan	[3.0,4.8)	[4.8,5.3)	[5.3,5.8)	[5.8,6.3)	[6.3,max)	136 / 416 / 317 / 139 / 88
Philippines	11-19N	115-124E	1052	Philippines	[3.0,4.8)	[4.8,5.3)	[5.3,5.8)	[5.8,6.1)	[6.1,max)	67 / 392 / 359 / 112 / 122
Chicago	38-47N	82-93W	610	USA	[3.0,3.2)	[3.2,3.5)	[3.5,4.0)	[4.0,4.5)	[4.5,max)	276 / 99 / 140 / 73 / 22
Los Angeles	30-40N	115-125W	1182	USA	[3.8,4.0)	[4.0,4.5)	[4.5,5.0)	[5.0,5.5)	[5.5,max)	417 / 409 / 196 / 118 / 42

3. Methodology

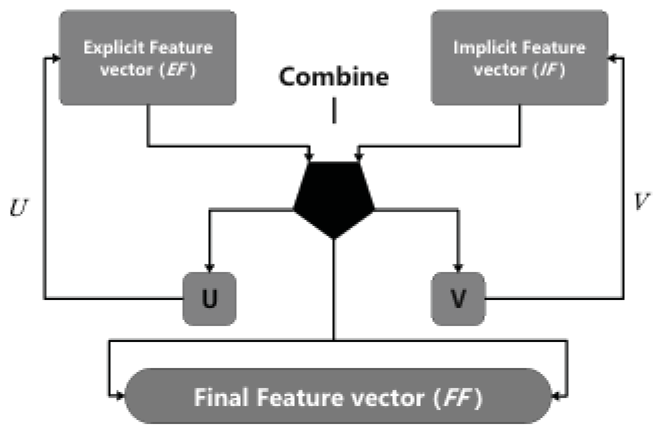
Fig. 1 gives the general framework of the proposed DLEP, which consists of four steps: data preprocessing, feature ex traction, feature fusion and prediction. In the first step of data preprocessing, we use the proposed segment method introduced in Section II-A to extract precursory patterns and training samples. In the second step of feature extraction, we adopt the eight mathematical statistics-based earthquake indicators [5] based on the obtained precursory patterns as the explicit feature vector, denoted as EF. The eight indicators are the time, mean magnitude, seismic root of seismic energy, b-value, mean square deviation, maximum difference, and coefficient of variation. In addition, we use CNN to extract implicit vector based on the obtained samples, denoted as IF. In the third step of feature fusion, we suggest an attention based strategy in Section III-B by using the parameter matrices U and V to weight the EF and IF respectively. Then, the fusion vector will be input into the full-connected layer to get the output. During the training phase, the category imbalance problem caused by data distribution tends to cause the model to converge to the local minimum, which is solved by the dynamic loss function proposed in Section III-C. Finally, the model outputs the magnitude range of main shock. More specifically, previous experiments have proved that the ReLU activation function is effective in the CNN, and the softmax is often adopted in the fully-connected layer as the activation function, thus we choose them in our model. To enhance the generalization performance of

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

our model, similar to the work in [6], we also adopt dropout layer and batch normalization layer in our model.



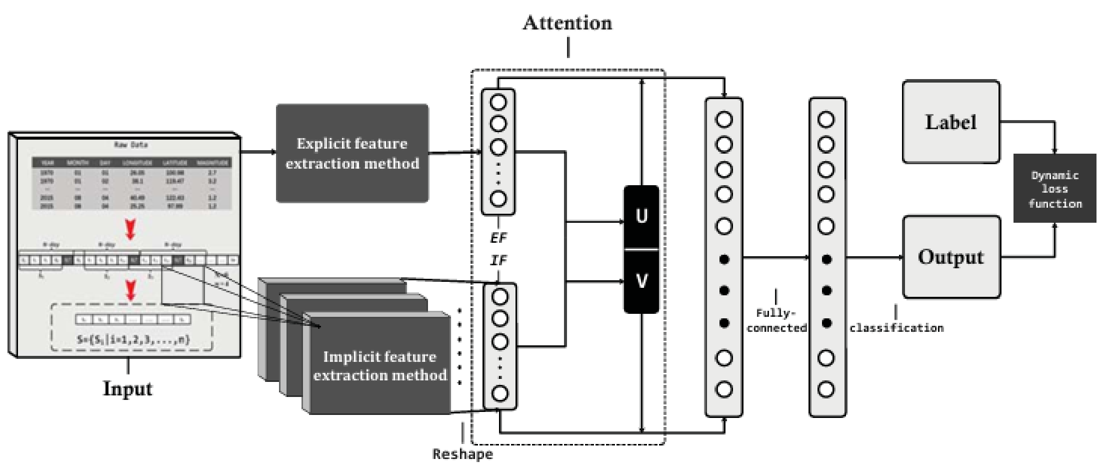


Fig 1. Architectural diagram

The input block contains the earthquake information in the per month, which are preprocessed through ZOH technique. The ZOH technique helps to reduce the effect of data's zero-value on the network training process and improve prediction performance. The feature extraction block is obtained by a one-dimensional CNN with nine layers, including four convolutional layers, four pooling layers, and one flatten layer. The filter size is wide selected in the first convolution layer of convolution, unlike the subsequent layers. When compared to small kernels, this structure is superior at damping high-frequency signals. Stacking several convolutional and pooling layers allows higher-level features to be extracted from the input, which helps represent the input data better. The Maxpooling layer is implemented after each convolution layer to reduce the dimensions and parameters within the network. In the feature extraction block, rectified linear unit (ReLU) is used as the activation function to avoid gradient vanishing or explosion problems while enhancing the convergence rate. Following each convolution layer, a batch normalization (BN) algorithm is employed as an effective regularization strategy. In addition to having a regularizing effect, it can reduce the shift of internal covariate, better the network's training performance, and increase the generalization capability of the network. BN is a feature normalization method in a layer-bylayer manner that is applied to accelerate the speed of the training process. Features in each layer are first normalized to the standard distribution and are then regulated to the ideal distributions.

4. Experiments

In this paper, mainland China has been chosen as the region of interest, situated in the southeast of the Eurasian plate. Mainland China is linked to the Siberia-Mongolia sub-plate, Philippine, and India plate; it is regarded as one of the most seismically active regions in the world. These earthquakes caused the death of more than 270,000 people, accounting for 54% of the overall death toll from natural catastrophes in mainland China. Therefore, reliable and effective predictions in this area can help to reduce the damage and casualties caused by earthquakes. One of the aims of the earthquake prediction problem is to predict and identify regions where major earthquakes occur. In order to analyze and more accurately predict the range of location of the next earthquake, mainland China is divided into several smaller regions. However, the lack of enough and appropriate data makes earthquake prediction and model training in small areas challenging. The study area was divided into nine small areas to address this challenge. The latitude range from 23 to 45 degrees, and the longitude range from 75 to 119 degrees; the

range of latitude and longitude is divided into three equal parts.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\overline{y}_i - y_i\right)^2}$$

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

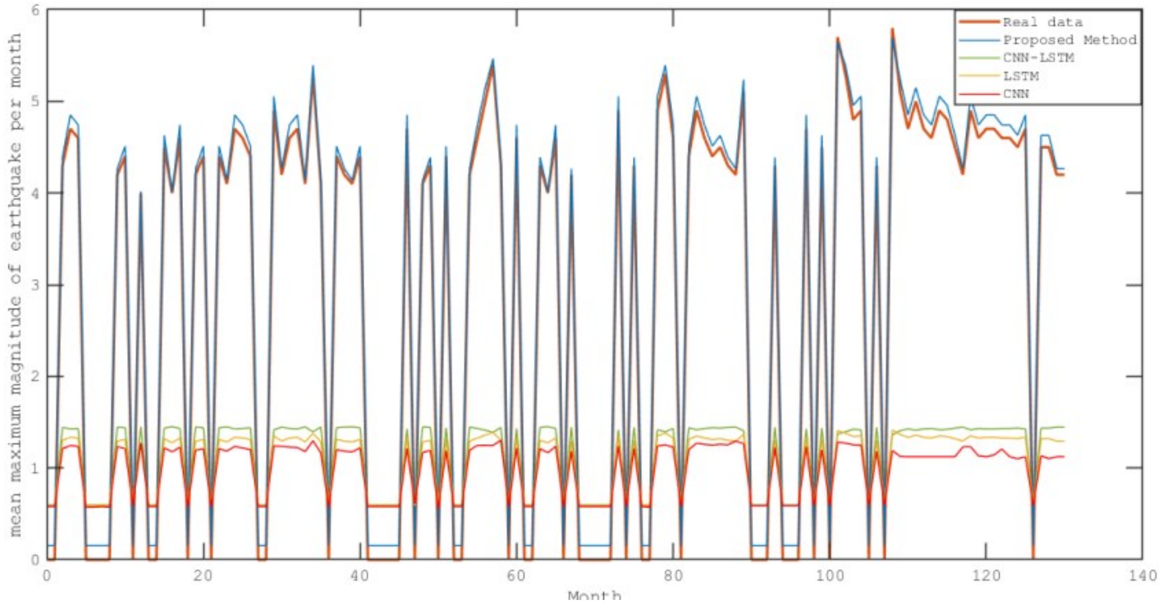


Fig. 9. The comparison of proposed model and deep learning models for maximum magnitude prediction in region 1

The proposed method is compared with CNN-BiLSTM, LSTM, CNN, RF, MLP, DT, and SVM to verify the efficiency, superiority, and generalization ability. Table III shows the results of the proposed method and comparison methods in predicting the number of earthquakes for nine different regions. It is clear from the outcomes that the proposed method consistently achieves the best prediction performance in all regions, with the lowest RMSE and MAE values and the highest $R2$ score. This shows that the proposed method is appropriately performed in predicting the number of earthquakes and its superior or competitive to other comparison methods. For example, in region 1, the other comparative models present high prediction errors, while the proposed model improvements prediction results with the RMSE value 0.24, the MAE value 0.018, and $R2$ value 0.956.

5. Conclusions

Due to the nonlinearity and complexity nature of earthquakes data, this paper provides a new CNN-BiLSTM-AM approach and a novel and efficient general framework for earthquake prediction in terms of number and maximum magnitude. The number and maximum magnitude of earthquakes that occurred in each month over the past 50 years are considered the model's input features, which makes the model can completely extract useful information from the historical data. A new data processing technique called ZOH is presented to better train the network and lessen the prediction difficulty. After data preprocessing, CNN is used to extract spatial characteristics. The features extracted by CNN are passed into BiLSTM. The BiLSTM is introduced to solve the data's long-term dependency, and the AM is used to highlight

the BiLSTM output features that have a high contribution to the prediction results. Finally, the output of the AM is sent to the fully connected layers to obtain the final result. Compared to other shallow machine learning and deep learning approaches, the simulation results in two case studies reveal that the proposed method has the best performance.

References

[1] A. S. Alarifi, N. S. Alarifi, and S. Al-Humidan, "Earthquakes magnitude predication using artificial neural network in northern red sea area," *Journal of King Saud University-Science*, vol. 24, no. 4, pp. 301–313, 2012.

[2] P. Cui, X.-Q. Chen, Y.-Y. Zhu, F.-H. Su, F.-Q. Wei, Y.-S. Han, H.-J. Liu, and J.-Q. Zhuang, "The wenchuan earthquake (may 12, 2008), sichuan province, china, and resulting geohazards," *Natural Hazards*, vol. 56, no. 1, pp. 19–36, 2011.

[3] N. Jain, D. Virmani, and A. Abraham, "Proficient 3-class classification model for confident overlap value based fuzzified aquatic information extracted tsunami prediction," *Intelligent Decision Technologies*, vol. 13,

no. 3, pp. 295–303, 2019.

[4] P. Cui, L.-z. Xiang, and Q. Zou, "Risk assessment of highways affected by debris flows in wenchuan earthquake area," *Journal of Mountain Science*, vol. 10, no. 2, pp. 173–189, 2013.

[5] R. Verdugo and J. Gonzalez, "Liquefaction-induced ground damages ´ during the 2010 chile earthquake," *Soil Dynamics and Earthquake Engineering*, vol. 79, pp. 280–295, 2015.

[6] J. D. Bray *et al.*, "Developing mitigation measures for the hazards associated with earthquake surface fault rupture," in *Workshop on seismic fault-induced failures—possible remedies for damage to urban*

facilities. University of Tokyo Press, pp. 55–79, 2001.

[7] R. Bilham, "The seismic future of cities," *Bulletin of earthquake engineering*, vol. 7, no. 4, pp. 839–887, 2009.

[8] N. N. Ambraseys and C. P. Melville, *A history of Persian earthquakes*. Cambridge university press, 2005.

[9] J. Jia, "Earthquake damages," in *Modern Earthquake Engineering*, pp. 413–431, Springer, 2017.

[10] G. A. Sobolev, "Methodology, results, and problems of forecasting

earthquakes," *Herald of the Russian Academy of Sciences*, vol. 85, no. 2, pp. 107–111, 2015.

[11] R. Kail, E. Burnaev, and A. Zaytsev, "Recurrent convolutional neural networks help to predict location of earthquakes," *IEEE Geoscience and Remote Sensing Letters*, 2021.