IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 7, 2019, accepted February 21, 2019, date of publication February 28, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902224

Precursory Pattern Based Feature Extraction Techniques for Earthquake Prediction

LEI ZHANG[1], LANGCHUN SI[2], HAIPENG YANG[2], YUANZHI HU[2], AND JIANFENG QIU[1]
[1]Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Anhui 230039, China
[2]School of Computer Science and Technology, Anhui University, Anhui 230039, China

Corresponding author: Jianfeng Qiu (qiujianf@ahu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61876184 and Grant 61502001, in part by the Key Program of Natural Science Project of Educational Commission of Anhui Province under Grant KJ2017A013, and in part by the Academic and Technology Leader Imported Project of Anhui University under Grant J01006057.

ABSTRACT Earthquake prediction is an important and complex task in the real world. Although many data mining-based methods have been proposed to solve this problem, the prediction accuracy is still far from satisfactory due to the deficiency of feature extraction techniques. To this end, in this paper, we propose a precursory pattern-based feature extraction method to enhance the performance of earthquake prediction. Especially, the raw seismic data is firstly divided into fixed day time periods, and the magnitude of the largest earthquake in each fixed time period is labeled as the main shock. The precursory pattern is a part of the seismic sequence before the main shock, on which the existing mathematical statistic features can be directly generated as seismic indicators. Based on these precursory pattern-based features, a simple yet effective classification and regression tree algorithm is adopted to predict the label of the main shock in a predefined future time period. The experimental results on two historical earthquake records of the *Changding-Garzê* and *Wudu-Mabian* seismic zones of China demonstrate the effectiveness of the proposed precursory pattern-based features with the selected CART algorithm for earthquake prediction.

INDEX TERMS Earthquake prediction, pattern discovery, time series, precursory pattern, CART.

I. INTRODUCTION

Earthquake is one of the devastating events in natural hazards that causes great casualties and property damage every day in the world since that it is hard to predict. With the increasing amount of earthquake datasets collected, many researchers try to solve the task of predicting the earthquake in the future time. Earthquake prediction is to estimate the time, location and magnitude of the future earthquake, which is one of the theoretical foundation of geophysics, geology, computer science and so on [1]–[4]. With the development of data mining techniques, a large number of scholars have devoted to discover the earthquake patterns from seismic time series based on various feature extraction methods and achieved some success [5], [6].

Since Gutenberg and Richter [7] designed seismic indicators based on mathematical statistical methods, e.g. earthquake magnitude, earthquake energy, earthquake acceleration, *b*-value and so on, a lot of researchers have proposed different feature extraction methods to obtain indicators for earthquake prediction. One category is based on the fixed number of seismic events. For example, Nuannin *et al.* [8] applied the sliding time and space windows containing a fixed number of seismic events to obtain earthquake indicators. Based on this method, Florido *et al.* [6] considered a fixed number of seismic events before main earthquake as the precursory pattern to extract features, which is useful for analyzing the trend of earthquakes. However, this feature extraction method cannot detect the range of earthquake magnitude of main shock well. The work in [5], similarly, extracted features from the fixed length seismic sequences before main earthquake, which can estimate the magnitude of earthquakes. However, the methods mentioned above cannot infer the effective time range of earthquake prediction results. Thus, another feature extraction method based on the fixed length of time before main shock is proposed to make earthquake prediction. Specifically, the historical earthquake records for a given region are divided into a number of predefined equal time periods such as one month or 15 days in [9]. The advantage of this method is that the

The associate editor coordinating the review of this manuscript and approving it for publication was Mohsin Jamil.

representative training samples can be obtained, which is critical for the learning of earthquake prediction models. But the fixed time window cannot make full use of the events before the main shock in current time period and lead to unsatisfactory prediction results.

To this end, in this paper, we propose a precursory pattern based feature extraction method for earthquake prediction, which can predict both the magnitude range of future earthquakes and obtain the effective time range of prediction results. In this study, earthquake precursor refers to a part of seismic records before the main shock, which is represented as the precursory pattern of earthquake. In order to obtain the representative learning samples, the raw seismic data is firstly divided into a set of fixed day time periods and the magnitude of the largest earthquake of each time period called main shock is as the label of the fixed period according to [9]. Then the sequence composed of the last $w(w > 0)$ events in the last time period before the current time and the events before the main shock in current time period is treated as earthquake precursory pattern. And the seismic indicators based on the obtained precursory patterns with a selected classification and regression tree algorithm named CART [10] can lead to satisfactory earthquake prediction results on two real-world seismic datasets of *Changding-Garzê* and *Wudu-Mabian* zones. In summary, the contributions of this paper can be summarized as follows:

- We propose a precursory pattern based feature extraction method to better capture the characteristics of earthquake, thus can be used to enhance earthquake prediction. Noting that the proposed method can not only obtain the seismic features, but also be used to estimate the effective time period of prediction results.
- We evaluate the effectiveness of the proposed feature extraction method with CART algorithm on two real-world seismic datasets of the *Changding-Garzê* and *Wudu-Mabian* zones in China. The experimental results show the superior performance of our method over the comparison algorithms, which indicates that the proposed precursory pattern based feature extraction method with selected CART is a promising method for accurate earthquake prediction.

The rest of this paper is structured as follows, we first give the formulation of earthquake prediction problem and the related work in Section II. In Section III, we present the proposed precursory pattern based feature extraction method and the adopted CART algorithm, in which the precursory pattern is described in detail. We describe the seismic dataset, evaluation metrics, comparison algorithms and the experimental results to demonstrate the effectiveness of the precursory pattern based feature extraction method with CART algorithm in Section IV and the conclusions drawn from this study are finally shown in Section V.

II. PROBLEM FORMULATION AND RELATED WORK

In this section, some preliminaries about earthquake prediction are firstly described, and then the related work about earth prediction with feature extraction techniques is introduced.

A. PROBLEM FORMULATION

In order to predict the magnitude range of future earthquakes as well as obtain the effective time range of prediction results, we formulate the seismic sequence based problem as follows. Specifically, assuming that the sequence of historical seismic records is denoted as Φ:

$$\begin{aligned} \Phi &= \{E_i | i = 1, 2, 3, \ldots, m\} \\ E_i &= \langle f_{i1}, f_{i2}, \ldots, f_{ij}, \ldots, f_{ik} \rangle \end{aligned} \tag{1}$$

where E_i is the i-th earthquake event and m is the total number of earthquakes. Besides, f_{ij} is the j-th seismic feature of event E_i and k is the total number of features. Usually, each record E_i contains at least two seismic features, i.e. earthquake magnitude M_s and time t. Similar to the work of [9], the historical earthquake sequence Φ can be divided into a number of fixed pre-defined N days. The earthquake with the largest magnitude that occurred during each time period is called main shock, denoted as E_j^m $(1 \leq j \leq n)$, where n the total number of main shocks in Φ.

Based on the above notations, the problem of earthquake prediction is formally defined as:

Definition 1 (Earthquake Prediction): Given the historical earthquake sequence Φ and the pre-defined N days, suppose Φ can be divided into a set of N days-periods with the size of n and E_j^m is the main shock with the largest magnitude that occurred during each time period $(1 \leq j \leq n)$, the task of earthquake prediction is to predict the magnitude of the main shock in the future N days-period based on last N days-period sequence data before the current time.

In the above task, one of key challenges is the feature extraction technique. In other words, how to extract effective features before main shocks is the key factor for accurate earthquake prediction. For example, Florido *et al.* [6] proposed to use a set of fixed length of events (denoted as SL) before main shocks to generate seismic indicators. To be specific, suppose there are w earthquakes chosen before main shock, then SL can be defined as:

$$\begin{aligned} SL &= \{SL_j | j = 1, 2, 3, \ldots, n\} \\ SL_j &= \langle E_{p-1}, E_{p-2}, \ldots, E_{p-w} | p > w \rangle \end{aligned} \tag{2}$$

where n is the number of SL_j and p is the serial number of E_j^m in Φ. Thus, Φ is transformed into sets of SL_j and E_j^m. As shown in Figure 1, there are 99 events in Φ and SL with the size n can be obtained when $w = 4$. In this example, the 4 events before each main shock is considered as precursory pattern, which can be used to generate seismic indicators. Suppose we can get the seismic indicators for the precursory pattern before each main shock. Then SL can be split into two parts: the first one is the training data set $\langle SL_i, E_i^m \rangle$ for $i \in [1, t]$, and the other one is the testing data set $\langle SL_j, E_j^m \rangle$ for $j \in [t+1, n]$. Based on training data set, various prediction models are constructed to make earthquake

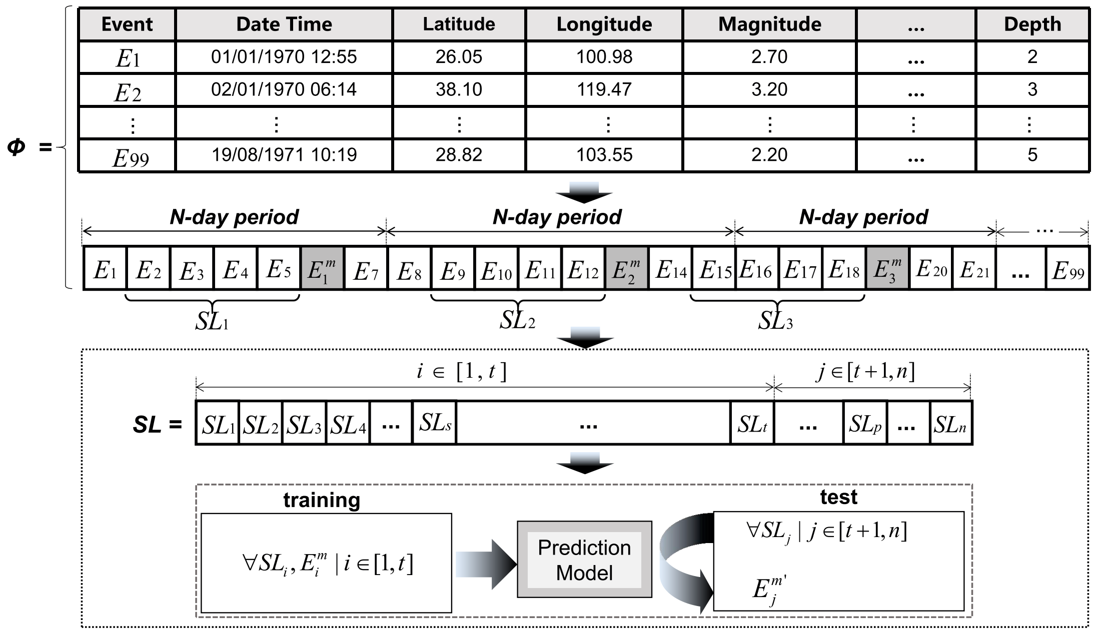


FIGURE 1. **The seismic sequence based framework of earthquake prediction process for $w = 4$.**

prediction such as ANN [5], [11], RBF [12], BP [13] and so on. Noting that the difference between $E_j^{m'}$ and E_j^m can be used to evaluate the earthquake prediction models, where $E_j^{m'}$ is the prediction result of SL_j and E_j^m is the real event.

In Section III, we will propose a new precursory pattern, based on which effective seismic indicators can be generated. Based on these precursory pattern based features, a simple yet effective CART algorithm can be used to make accurate earthquake prediction.

B. RELATED WORK

In this section, the related work on earthquake prediction is presented, in which the feature extraction methods are described in detail.

In recent years, many data mining methods have been proposed to make earthquake prediction based on various feature extraction methods since the seismic indicators based on statistical methodology are firstly proposed by [7]. In order to detect the precursors for large earthquakes, Nuannin *et al.* [8] used the sliding time-windows containing a constant number of events to examine the spatial distribution of *b*-value. Particularly, observed variations in *b* reveals a precursory potential which could be used in medium-term (months, years) earthquake prediction. However, this method is designed for large earthquakes (magnitude greater than 7) and does not provide the specific precursory pattern for earthquakes.

Florido *et al.* [6] proposed to enhance earthquake prediction by detecting the precursory patterns, which is an improvement on the basis of [8]. Specifically, the data are first grouped into the set of five chronologically ordered earthquakes according to [8], and then the sighed variation on the *b*-values in the time interval for the five earthquakes is used to discover the precursory patterns of earthquakes with magnitude larger than a constant. The results showed that the precursory patterns are useful for earthquakes with magnitude larger than 4.4. Unfortunately, these precursor signals extracted from the earthquakes with magnitude exceeding a fixed threshold cannot obtain the specific magnitude range of earthquakes.

To solve the above problem, the task of earthquake prediction can be converted into the classification problem of earthquake magnitude [5], [13], [14]. For example, Narayanakumar and Raja [13] extracted seismic features of a fixed number of events before main shock to make earthquake prediction with BP neural network technique. The historical earthquake catalog with Richter magnitudes between 3.0 and 8.6 is divided into fifteen groups where each group is comprised of earthquakes of magnitude in a 0.4 Richter range, and the category tag is considered as predicted target. The results show that this method can provide better accuracy for medium-large earthquakes, but still cannot achieve satisfactory prediction results for large earthquake. Besides, this feature extraction method cannot estimate the effective time range of the predicted earthquakes.

To address the problem of magnitude range prediction and the effective time range of prediction results, Adeli and Panakkat [9] proposed a novel feature extraction method. To be specific, the historical records for a given region are divided into fixed day time periods such as 15 days or one month, and the earthquake with the largest magnitude in each time period is labeled as main shock. And then the author predicted the magnitude range of the largest earthquake in the following pre-defined day time periods. Thus, the problem of the effective time range of prediction results is well handled. Moreover, the author claimed that this feature extraction method with selected probabilistic neural network(PNN) can provide good prediction results for earthquakes with magnitude between 4.5 and 6.0 Richter. Similarly, Mirrashid [15] investigated the prediction of earthquakes with magnitude 5.5 or more based on the adaptive neuro-fuzzy inference system (ANFIS), and the experimental results validated that ANFIS can obtain the best results in terms of precision comparing to the baseline algorithms. Asencio *et al.* [16] proposed to use a clustering method for seismogenic zones partitioning, and then use different machine learning techniques to make earthquake prediction, including KNN, ANN, NB, C4.5 decision trees and SVM, which can build reliable and general earthquake prediction systems.

In order to improve the precision of earthquake prediction, many researchers used feature selection techniques to eliminate redundant features, for example, Mart *et al.* [17] adopted the information gain of each seismic indicator for feature selection. Later, Asim *et al.* [18] designed a hybrid embedded feature selection method, which can be used to make accurate earthquake prediction. In addition, Hamze-Ziabari and Bakhshpoori [4] recently proposed an efficient bagging ensemble model of M5' and CART algorithms to predict ground motion parameters such as Peak Ground Acceleration, Peak Ground Velocity, and Peak Ground Displacement. These parameters are well known to characterize an earthquake, which are very helpful for seismic analysis of structures and risk assessment.

Noting that the method in [6] is useful for discovering the earthquake precursory patterns, but it cannot provide the effective time range of earthquake prediction results. Besides, the feature extraction method proposed in [9] can obtain the specific magnitude range of prediction results and provide the reference time range of the predicted earthquakes, but it does not consider the earthquakes before main shock in the current time window. Moreover, traditional data mining methods exposed their limitations on mining data with complex nonlinear correlation and the lack of data. Different from the above works, this paper proposed a precursory pattern based feature extraction method with the selected CART algorithm to enhance earthquake prediction, which can not only solve the shortcomings of [9] and [6], but also reduce the influence of redundant features.

III. METHODOLOGY

In this section, we first introduce the proposed precursory pattern based feature extraction methodology, and then briefly describe the CART algorithm applied to seismic prediction.

A. PRECURSORY PATTERN BASED FEATURES EXTRACTION

The method proposed by [6] is useful for detecting potential earthquake precursory patterns, and the feature extraction method proposed by [9] can estimate the time range of the prediction results. In order to predict both the magnitude range of future earthquakes as well as obtain the effective time range of prediction results, we propose a precursory pattern based features extraction method by combining the advantages of these two methods.

To be specific, the sequence of historical seismic records Φ is firstly divided into N-day time period according to [9]. The magnitude range of the largest earthquake in the corresponding time period is the prediction target (i.e. main shock), labeled as E_i^m. Different from the work [6] that a set of fixed length of events before main shock is defined as the precursory pattern, in this paper, the precursory pattern is the pattern sequence before the main shock that consists two part, i.e. the internal pattern sequence PI_i and the external pattern sequence PE_i^w. Here, w is the length of PE_i. The internal pattern sequence PI_i is the sequence of events before main shock in the current time period i, thus, the length of PI_i may be different for different time periods since that the position of E_i^m in each time period may be different. The external pattern sequence PE_i^w is the sequence of events with the fixed size w in the last time period before the current time period. The precursory pattern combining both PI_i and PE_i^w can generate effective seismic indicators. The union of PE and PI, denoted as SL, is used to calculate seismic indicators. Formally, the definition of SL is given in the following:

$$\begin{aligned} SL &= \{SL_i | i = 1, 2, \ldots, n\}, \\ SL_i &= \langle PE_i \cup PI_i \rangle \end{aligned} \tag{3}$$

Figure 2 gives an example of generated precursory patterns when w is set to 2. Based on the obtained SL, the mathematical statistics based earthquake features can be generated as the seismic indicators. There are eight indicators that can be selected as the features. In the following, we will give the formal definition for each indicator.

The first indicator is time ΔT, which is the time span over n events. And ΔT is the time span of seismic records in SL. The equation for ΔT is given by below:

$$\Delta T_i = t_{p-1} - t_{p-n}, \quad s.t.\ p > n, \tag{4}$$

where p is the index of E_i^m in Φ and n represents the number of elements in SL_i.

The second seismic indicator considered is the mean magnitude of earthquakes in SL_i and it is related to the magnitude of the earthquake that occurred before the main shock.

$$\overline{M}_i = \frac{1}{n} \sum_{SL_i} M, \tag{5}$$

where M is the magnitude of earthquake in SL_i.

The third seismic indicator is dE, that is, the rate of square root of seismic energy released during ΔT_i determining the magnitude of earthquake. Seismic crustal plates drift caused by motion-class from the kinetic energy of the plate movement and volcanic eruption energy from the center of the earth are the two main sources of seismic energy. Thus the energy released obtained through Eq. (6).

$$dE_i^{1/2} = \sqrt{\sum_{SL_i} M^{4.8+1.5M}}. \tag{6}$$

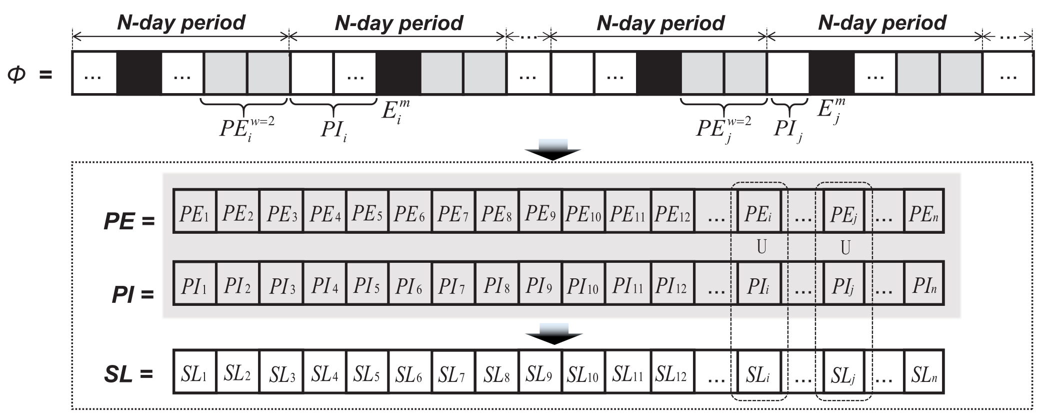


FIGURE 2. Precursory patterns are generated with $w = 2$, where E_i^m is the main shock of i-th N-day time period.

TABLE 1. Sample training datasets for 10 time periods between 26th October 1973 and 14th March 1974 for the *Changding-Garzê* seismic zone showing eight input indicators computed for *SL* and the corresponding input class based on the magnitude of the largest earthquake that occurred during that time period.

Time period	Input vector								Class
	ΔT	b	$\overline{M}$	η	$dE^{1/2}(e+07)$	ΔM	c	fre	
26/10/1973 - 08/11/1973	0	2.1815	3.3000	0.0000	0.0167	0.00	0.0000	1	**class 1**
09/11/1973 - 22/11/1973	0	-2.1616	3.7000	0.0000	0.0723	0.00	0.0000	1	**class 3**
23/11/1973 - 06/12/1973	6	1.0882	3.1000	0.0100	0.0603	0.20	0.0323	3	**class 3**
07/12/1973 - 20/12/1973	0	1.4521	3.2000	0.0000	0.0082	0.00	0.0000	1	**class 1**
21/12/1973 - 03/01/1974	20	5.4916	3.4200	0.1270	0.0514	0.90	0.1042	5	**class 2**
04/01/1974 - 17/01/1974	5	-12.6806	3.5333	0.1733	0.2387	0.80	0.1178	3	**class 4**
18/01/1974 - 31/01/1974	0	-0.2412	5.3000	0.0000	1.1147	0.00	0.0000	1	**class 5**
01/02/1974 - 14/02/1974	0	-474.4195	3.5000	0.0000	0.0716	0.00	0.0000	1	**class 3**
15/02/1974 - 28/02/1974	4	2.4947	3.3250	0.0892	0.0315	0.70	0.0898	4	**class 2**
01/03/1974 - 14/03/1974	10	0.8702	3.0000	0.0000	0.0148	0.00	0.0000	3	**class 1**

The fourth seismic indicator is the well-known b-value, which is the proportional coefficient in relationship between magnitude and frequency. According to the work [7] and [19], the equation of b-value is given in Eq. (7).

$$b_i = \frac{\log e}{\overline{M_{all}} - \frac{1}{n}\sum_{SL_i} M}. \tag{7}$$

The fifth seismic indicator is the mean square deviation about the regression line based on Gutenberg-Richter inverse power law for the events in SL, which is denoted as η. This indicator is related to the magnitude change of the earthquake in SL. The equation of η is given in Eq. (8).

$$\eta_i = \sqrt{\frac{1}{n}\sum_{SL_i}(M - \frac{1}{n}\sum_{SL_i} M)^2}. \tag{8}$$

The sixth seismic indicator is the maximum difference of seismic magnitude in SL, which is taken as a seismic indicator obtained from Eq. (9).

$$\Delta M_i = M_i^{max} - M_i^{min}, \quad where\ M_i^{max},\ M_i^{min} \in SL_i. \tag{9}$$

The seventh seismic indicator is c, that is the coefficient of variation of earthquake magnitudes. And it is defined by Eq. (10).

$$c_i = \frac{\sqrt{\frac{1}{n}\sum_{SL_i} M^2 - (\frac{1}{n}\sum_{SL_i} M)^2}}{\frac{1}{n}\sum_{SL_i} M} \tag{10}$$

The last seismic indicator is the frequency of earthquake. Compared with main shock, the magnitude of foreshock is often large enough, and the frequency of earthquake is increased before the main shock according to Jones [20]. Thus, the seismic frequency before main shock is considered as a seismic indicator, and defined as Eq. (11):

$$fre_i = \#num(SL_i). \tag{11}$$

Table 1 shows the sample of the eight indicators computed for SL with $N = 14$ and the corresponding input class for the ten time periods between 26th October 1973 and 14th March 1974 for the *Changding-Garzê* earthquake dataset used in our experiments.

B. CLASSIFICATION MODEL

Classification and regression tree algorithm(CART) [10] is an important data mining method [21]. CART is a binary recursive partitioning procedure capable of splitting the scalar attributes and processing the continuous and discrete attributes. Besides, CART is nonparametric and it does not require variables to be selected in advance because this decision tree can do feature selection work automatically. In addition, CART algorithm can generate understandable rules and its calculation cost is relatively small. Thus, this method has been applied in many fields, such as financial [22] and medical [23]. The building process of CART includes two main steps: the construction of CART and pruning.

1) THE CONSTRUCTION OF CART

In general, a decision tree contains a root node, several internal nodes and leaf nodes. The leaf node is the decision result, and other nodes are based on the attribute test. Therefore, the key of the decision tree algorithm is to select the partitioning attribute based on data purity, and there are many methods to measure the purity of data, such as information gain, Gini index, χ^2 statistics and so on. Classification and regression tree(CART) algorithm [10] is based on Gini index. Specifically, the Gini index of the dataset D is given by Eq. (12).

$$Gini(D) = 1 - \sum_{k=1}^{|y|} p_k^2, \tag{12}$$

where p_k is the proportion of the k-th class in dataset D, and $|y|$ is the number of samples. The Gini index reflects the probability that two samples randomly selected from the dataset D with different category labels. Thus, the purity of the dataset D with a small Gini index $Gini(D)$ is high.

In the binary partitioning process of the CART algorithm, the purity of one attribute exceeds the predefined threshold and is divided into the left subtree, otherwise it is divided into the right subtree. And the Gini index of the attribute a is defined as Eq. (13).

$$Gini\_index(D, a) = \sum_{v=1}^{V} \frac{|D^v|}{|D|} Gini(D^v), \tag{13}$$

where V is the number of nodes when a is used to divide dataset D, and D^v is the sample of the attribute a on the v-th branch node. Therefore, the attribute with the smallest Gini index in candidate attribute set A is regard as the optimal partition attribute a_*. The optimal attribute a_* is calculated as follows.

$$a_* = \arg\min_{a \in A} Gini\_index(D, a). \tag{14}$$

2) PRUNING

Pruning is the main means for the decision tree to deal with overfitting. In the learning process of decision tree, the partition nodes are frequently generated on order to classify training samples more correctly. Meanwhile, the excessive branch of the decision tree leads to overfitting. Thus, it is essential to pruning the decision tree in order to avoid overfitting problem. The pruning strategies of the decision tree includes pre-pruning and post-pruning. The pre-pruning refers to evaluate each node before partitioning in the process of decision tree construction. If current node partitioning does not improve the performance of decision tree, the current node is marked as leaf node. The post-pruning means that each non-leaf node partitioning is evaluated after the whole decision tree is built. And the CART algorithm adopts the second pruning strategy.

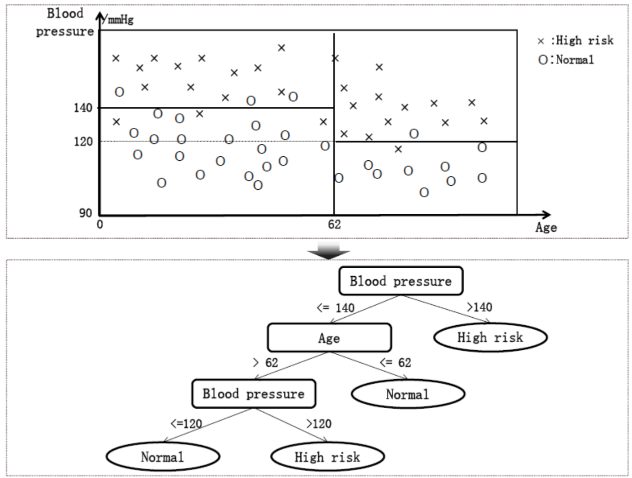


FIGURE 3. This picture gives a simple example of applying the CART algorithm to infer the health status of the body based on blood pressure and age: normal or high risk.

To demonstrate the decision principle of CART algorithm more clearly, an example of using the CART algorithm to infer the health status of body is presented in Figure 3. As shown in this figure, if the current blood pressure is higher than 140 mmHg, the patient is at high risk. Moreover, the current blood pressure below 140 mmHg does not indicate that the patient is in a normal state when the age is over 62 years old. This example shows that the CART algorithm can not only generate clear decision rules, but also the importance of attributes can be obtained based on the position in the decision tree. Thus, in this study, we apply this method to earthquake prediction problem. The main procedures of CART algorithm can be found in Algorithm 1.

IV. RESULTS

In this section, we first give the description of earthquake dataset. Then we present comparison algorithms and the evaluation metrics. Finally, we compare the prediction performance of different algorithms through experimental results.

A. DATASET

In this study, we adopt earthquake datasets from two seismic zones of China. One is the *Changding-Garzê* seismic zone, which is an area between geographic coordinating 29 N and 34 N north latitude and 98 E and 103 E east longitude. And the range of historical records from 5th January 1970 to 26th June 2015. The other is the *Wudu-Mabian* seismic zone, which is an area between 28 N and 34 N north latitude and 103 E and 105 E east longitude. And the range of historical records from 20th January 1970 to 2nd August 2015.

For the two datasets, the earthquakes with magnitude greater than 3.0 in both regions are selected. In this paper, the pre-defined equal time period is set to 14 days. For each time period, the magnitude of the largest earthquake

Algorithm 1 CART Algorithm

```
Input   : The training set
          D = {(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)};
          The attribute set A = {a_1, a_2, ..., a_d};
Process: TreeGenerate(D, a)
Generate node;
if Only one class C in D then
    node ← C; return
end
if A = ∅ or The sample of D has the same value in A then
    Label node as leaf node;
    And mark A as the class with the most samples in D; return
end
a_* = arg min_{a∈A} Gini_index(D, a);
for each a_*^v in a_* do
    Generate a branch for node;
    Let D^v denote a subset of D that take a value of a_*^v on a_*;
    if D^v = ∅ then
        Label the branch node as the leaf node;
        And mark A as the class with the most samples in D; return
    end
    else
        TreeGenerate(D^v, A {a_*}) marked as branch node;
    end
end
Output:
          A decision tree with node as the root node;
```

TABLE 2. The statistics for the data *Changding-Garzê* seismic zone of China.

Class	Magnitude range	Number of events (instances)
Class 1	< 4.0	995
Class 2	$4.0 \sim 4.5$	74
Class 3	$4.5 \sim 5.0$	31
Class 4	$5.0 \sim 5.5$	17
Class 5	> 5.5	26

(i.e. main shock) can be obtained, which can be divided into five classes: the first class consists of earthquakes with magnitude less than 4.0 Richter, the second class is between 4.0 Richter and 4.5 Richter, the third class is between 4.5 Richter and 5.0 Richter, the fourth class is between 5.0 Richter and 5.5 Richter, and the last class is larger than 6.0 Richter. Table 2 and Table 3 present the number of events available for each class on the two datasets, respectively.

For the two datasets, we re-arrange the obtained events in random order (i.e. randomly selected the training and testing data) and use 5 fold cross-validation to evaluate the prediction results.

TABLE 3. The statistics for the data *Wudu-Mabian* seismic zone of China.

Class	Magnitude range	Number of events (instances)
Class 1	< 4.0	923
Class 2	$4.0 \sim 4.5$	134
Class 3	$4.5 \sim 5.0$	65
Class 4	$5.0 \sim 5.5$	30
Class 5	> 5.5	21

B. COMPARISON ALGORITHMS

In this paper, we compare the proposed precursory pattern based feature extraction method with two baseline methods. The first one is the method used in 2016 [13] called **2016N**, which extracts seismic indicators before the earthquakes with the magnitude within a specific range. The other baseline method extracts seismic features by using the earthquakes in last N-day time period before the main shock in current time period in 2009 [9] called **2009A**.

Besides, based on these feature extraction techniques, there are five earthquake prediction models chosen to compare with the selected CART algorithm, including multi-class SVM [24], BP artificial neural network used in [13], the probabilistic neural network (PNN) used in [9], the adaptive neuro-fuzzy inference system (ANFIS) used in [15] and the artificial neural network (ANN) used in [16]. In addition, to further show the reason why we select CART as the classification method for earthquake prediction, we also give the comparison results between CART and genetic algorithm (GA) [25] with SVM, BP, PNN, ANFIS, ANN as the basic classifiers.

C. EVALUATION METRICS

In this paper, two well-known metrics are used to evaluate the effectiveness of the proposed method and the baselines.

The first metric is *Accuracy* [26], which is usually used to evaluate the classification effect of various classifiers. The definition of accuracy is given in Eq. (15):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{15}$$

where TP is the number of times that upcoming earthquake has been correctly predicted. TN is the number of times that neither an earthquake prediction model has triggered an alarm nor an earthquake has occurred. FP is the number of times that an earthquake prediction model has triggered an alarm but no earthquake has occurred. FN is the number of times that a classifier has not triggered an alarm but did not occur.

The other one is the extension of AUC [27] to multi-class problems ($MAUC$) [28], which is very important to multi-class cost-sensitive learning and imbalanced learning problems. Thus, $MAUC$ is also used to evaluate the proposed method. And the equation of $MAUC$ is given as follows.

$$MAUC = \frac{2}{c \times (c-1)} \sum_{i<j} \frac{A_{ij} + A_{ji}}{2}, \tag{16}$$

TABLE 4. The average *Accuracy* and *MAUC* of the six classification methods based on the precursory pattern based method and other feature extraction methods on the dataset of *Changding-Garzê* seismic zone in 5-fold cross validation.

Feature Extraction		*Accuracy*						*MAUC*					
		PNN	**BP**	**SVM**	**ANFIS**	**ANN**	**CART**	**PNN**	**BP**	**SVM**	**ANFIS**	**ANN**	**CART**
Baselines	2016N	0.8704	0.8702	0.8625	0.8310	0.8632	0.8433	0.5139	0.5163	0.5184	0.5989	0.5163	0.6362
	2009A	0.8705	0.8518	0.8696	0.8040	0.8167	0.8171	0.5037	0.5165	0.5035	0.5052	0.5288	0.5048
The proposed method	$w = 1$	0.8714	0.8711	0.8697	0.8425	0.8509	0.9283	0.4655	0.4940	0.4851	0.5865	0.4950	0.7890
	$\boldsymbol{w = 2}$	0.8714	0.8693	0.8688	0.8565	0.8693	**0.9326**	0.4655	0.4971	0.5035	0.5741	0.4844	**0.8084**
	$w = 3$	0.8714	0.8605	0.8670	0.8443	0.8658	0.9099	0.4655	0.4946	0.4801	0.6099	0.4929	0.7446
	$w = 4$	0.8714	0.8614	0.8688	0.8627	0.8623	0.9003	0.4655	0.4926	0.4792	0.6093	0.4930	0.6957
	$w = 5$	0.8714	0.8649	0.8696	0.8434	0.8658	0.8968	0.4655	0.4910	0.4751	0.5828	0.4930	0.7046

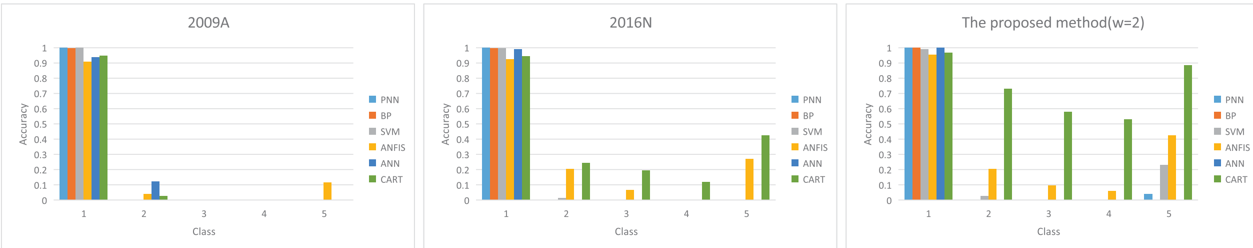


FIGURE 4. The *Accuracy* of each magnitude range (i.e. class) obtained by different prediction models with different feature extraction methods on *Changding-Garzê* seismic zone.

where A_{ij} is the *AUC* between class i calculated from the i-th column of a $n \times c$ matrix M (n is the number of instances and c is the number of classes). Note that for multi-class problems, A_{ij} may not equal to A_{ji}, and thus both of them need to be involved in the calculation of *MAUC*. For both *Accuracy* and *MAUC*, a larger value indicates a better prediction performance.

D. RESULTS

1) THE EFFECTIVENESS OF PROPOSED METHOD

a: RESULTS FOR CHANGDING-GARZÊ SEISMIC ZONE

To verify the effectiveness of the proposed feature extraction method with the selected CART classifier, Table 4 records the average *Accuracy* and *MAUC* of the six comparison classification methods on *Changding-Garzê* seismic zone based on three kinds of feature extraction techniques: 2016N [13], 2009A [9] and the proposed precursory pattern based technique with different values of w (from 1 to 5 at the step of 1).

From this table, we can observe that the method PNN obtains the best accuracy based on two baseline feature extraction techniques, but the *MAUC* of this algorithm is still far from satisfactory due to class imbalance problem. For *MAUC*, CART based on 2016N obtains the best *MAUC* value although its accuracy is lower than PNN. However, based on the proposed feature extraction method, CART algorithm obtains the best performance on both *Accuracy* and *MAUC*, that is, 93.26% and 80.84% respectively when w is set to 2. Moreover, for most baseline methods, their performance are improved, which can validate the effectiveness of the proposed feature extraction technique. For example, the *Accuracy* of ANFIS has improved from 83.10% to 85.65% even if the *MAUC* has a slight drop.

Furthermore, we can also find that the accuracy of some classifiers such as ANFIS and CART with the proposed precursory pattern based feature extraction method is firstly increased to the best and then decreased with the parameter w increasing from 1 to 5 at the interval of 1. For example, CART obtains the best performance with the proposed precursory pattern based features when $w = 2$. In addition, Figure 4 presents *Accuracy* of each magnitude range (i.e. class) obtained by different prediction models with the three feature extraction methods on *Changding-Garzê* seismic zone. As shown in this figure, it can be found that the accuracy of CART on each magnitude range is greatly improved by using the proposed precursory pattern based features.

In summary, the above experimental results can validate the superiority of the proposed feature extraction method comparing to the two baseline methods.

b: RESULTS FOR WUDU-MABAIN SEISMIC ZONE

In order to further validate the effectiveness of the proposed feature method with the selected CART algorithm, Table 5 gives the prediction results of the six classification methods on the *Wudu-Mabian* seismic zone based on the three feature extraction techniques. As can be observed from this table, based on the two baseline feature extraction techniques, PNN obtains the best *Accuracy* value 78.68% although its *MAUC* is only 51.86%, while ANFIS obtains the best *MAUC* value 63.35% although its *Accuracy* is only 74.31%, CART can get better tradeoff results with *Accuracy* 77.47% and *MACU* 61.68%. Based on the proposed feature extraction technique, among all the baselines, CART can get the best *Accuracy* value (i.e. 92.07%) when $w = 2$ and the best *MAUC* value (i.e. 88.61%) when $w = 1$. As shown in Figure 5,

TABLE 5. The average *Accuracy* and *MAUC* of the six classification methods based on the precursory pattern based method and other feature extraction methods on the dataset of *Wudu-Mabian* seismic zone in 5-fold cross validation.

Feature Extraction		*Accuracy*						*MAUC*					
		PNN	BP	SVM	ANFIS	ANN	CART	PNN	BP	SVM	ANFIS	ANN	CART
Baselines	2016N	0.7866	0.7863	0.7807	0.7431	0.7641	0.7747	0.5137	0.4902	0.5414	0.6335	0.5071	0.6168
	2009A	0.7868	0.7667	0.7843	0.7264	0.7487	0.7416	0.5186	0.4950	0.5187	0.5896	0.4881	0.5905
The proposed method	$w=1$	0.7869	0.7863	0.7945	0.8406	0.6547	0.9199	0.4883	0.4910	0.5225	0.7087	0.4831	**0.8861**
	$\boldsymbol{w=2}$	0.7869	0.7863	0.7971	0.8346	0.5274	**0.9207**	0.4883	0.4910	0.5224	0.6645	0.4808	0.8726
	$w=3$	0.7869	0.7863	0.7962	0.8354	0.5308	0.8892	0.4883	0.4910	0.5202	0.6919	0.4777	0.8382
	$w=4$	0.7869	0.7863	0.7962	0.8312	0.5359	0.8824	0.4883	0.4910	0.5315	0.6944	0.4788	0.8528
	$w=5$	0.7869	0.7863	0.7971	0.8141	0.5385	0.8593	0.4883	0.4910	0.5249	0.6605	0.4792	0.7914

the CART algorithm with the proposed feature extraction method achieves very high percentage of accuracy on each magnitude range.

Based on the above experiments, we can conclude that the proposed precursory pattern based features with the selected CART algorithm can greatly improve the performance of earthquake prediction in terms of both *Accuracy* and *MAUC*.

2) THE ROBUSTNESS OF THE SELECTED CART

One key advantage of CART algorithm is that the importance value of each feature can be obtained when the tree of CART has been built. In addition, the CART algorithm can automatically select the important feature to split the training data, thus CART is robust even features are redundant. This is the reason why the CART algorithm is adopted as the classifier in this paper for earthquake prediction. To further validate our analysis here, Table 6 and Table 7 give the importance values of different features obtained in CART in descending order based on datasets *Changding-Garzê* and *Wudu-Mabian* respectively (w is set to 2 in precursory pattern). As can be observed from these tables, the feature $dE^{1/2}$ plays the most important role in CART for earthquake prediction on both *Changding-Garzê* and *Wudu-Mabian* seismic zones.

TABLE 6. The ranking of the importance of eight seismic indicators based on CART algorithm in 5-fold cross validation on *Changding-Garzê* seismic zone ($w = 2$).

Ranking	1	2	3	4	5	6	7	8
Indicator	$dE^{1/2}$	$\overline{M}$	c	b	ΔT	Fre	ΔM	η
Importance	0.7703	0.1469	0.0515	0.0303	0.0010	0.0000	0.0000	0.0000

TABLE 7. The ranking of the importance of eight seismic indicators based on CART algorithm in 5-fold cross validation on *Wudu-Mabian* seismic zone ($w = 2$).

Ranking	1	2	3	4	5	6	7	8
Indicator	$dE^{1/2}$	$\overline{M}$	ΔM	c	Fre	b	ΔT	η
Importance	0.7692	0.1721	0.0217	0.0157	0.0095	0.0091	0.0027	0.0000

In order to further validate the robustness of the selected CART algorithm with redundant features, the genetic algorithm (GA) [25] based baselines (GA+SVM, GA+BP, GA+PNN, GA+ANFIS, GA+ANN) are compared with CART. Here, GA is used for feature selection, that is, to select best subset of features that can be used in basic classifiers to further improve the final performance. In GA, the fitness function is the measure *Accuracy*, the size of population is set to 100, the maximum generation is set to 10, the cross and the mutation probabilities are set to 1 and 0.5, respectively.

TABLE 8. Comparison results between CART and GA based baselines on *Changding-Garzê* seismic zone in 5-fold cross validation based on the proposed precursory pattern based features ($w = 2$ in precursory pattern).

Algorithms	Best feature subset	Acc.(noGA)	MAUC(noGA)
GA+PNN	$\{dE^{1/2}, c, \overline{M}, \Delta M, \Delta T\}$	0.8714(0.8714)	0.4655(0.4655)
GA+BP	$\{b, \Delta T, c, \eta\}$	0.8728(0.8693)	0.4952(0.4971)
GA+SVM	$\{dE^{1/2}, fre, \eta\}$	0.8898(0.8688)	0.5959(0.5035)
GA+ANFIS	$\{dE^{1/2}, \overline{M}, \eta, \Delta M\}$	0.8828(0.8565)	0.6347(0.5741)
GA+ANN	$\{\overline{M}, fre, b, c\}$	0.8763(0.8693)	0.5226(0.4844)
CART	The eight features	**0.9326**	**0.8084**

TABLE 9. Comparison results between CART and GA based baselines on *Wudu-Mabian* seismic zone in 5-fold cross validation based on the proposed precursory pattern based features ($w = 2$ in precursory pattern).

Algorithms	Best feature subset	Acc.(noGA)	MAUC(noGA)
GA+PNN	$\{b, c, \Delta M, \eta, \Delta T\}$	0.7869(0.7869)	0.4875(0.4883)
GA+BP	$\{b, \eta, \Delta M\}$	0.7872(0.7863)	0.4967(0.4910)
GA+SVM	$\{dE^{1/2}, \eta\}$	0.8457(0.7971)	0.7276(0.5224)
GA+ANFIS	$\{dE^{1/2}, \Delta M, \overline{M}\}$	0.8977(0.8346)	0.8290(0.6645)
GA+ANN	$\{fre\}$	0.8034(0.5274)	0.5584(0.4808)
CART	The eight features	**0.9207**	**0.8726**

Table 8 and Table 9 respectively present the comparison results between CART and GA based baselines on *Changding-Garzê* and *Wudu-Mabian* seismic zones in 5-fold cross validation ($w = 2$ in precursory pattern based features) in terms of *Accuracy* (Acc.) and *MAUC*, where the element in each black parenthesis is the performance without GA. From the two tables, we can find that using GA for feature selection for optimizing these basic classifiers (i.e. PNN, BP, SVM, ANFIS and ANN) can clearly improve the performance of these basic classifiers in terms of both *Accuracy* and *MAUC*. Among the GA-based baselines, GA+SVM and GA+ANFIS are the top-2 best ones, where the most important feature $dE^{1/2}$ obtained by CART is appeared in the best feature subset obtained by GA+SVM and GA+ANFIS on both two datasets, which can validate the effectiveness of CART for obtaining the importance of features. Secondly, although GA based baselines can improve the final performance, the CART still get the best performance in terms of both *Accuracy*

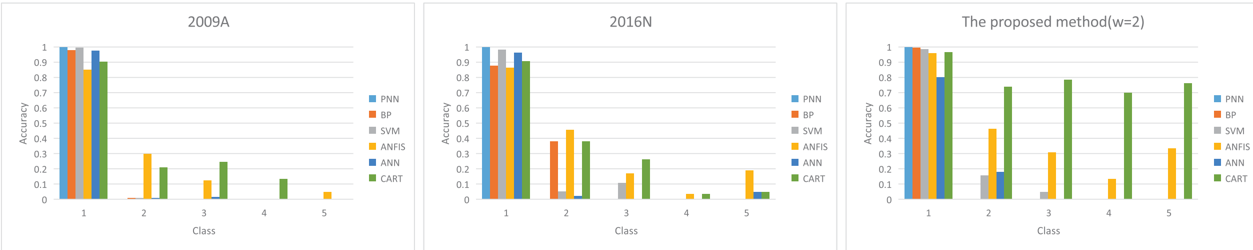


FIGURE 5. The *Accuracy* of each magnitude range (i.e. class) obtained by different prediction models with different feature extraction methods on *Wudu-Mabian* seismic zone.

and *MAUC*. The reason is the redundant features has little effect on CART algorithm since that CART can do feature selection automatically according to the information gain of each feature in the process of CART tree building.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a precursory pattern based feature extraction method with the selected CART approach for accurate earthquake prediction. To verify the effectiveness of the proposed method, many state-of-the-art baselines were compared on the two datasets, i.e. *Changding-Garzê* and *Wudu-Mabian* seismic zones of China. Noting that the prediction accuracy of the proposed method can reach 93.26% and 92.07% on the two datasets respectively, which is much better than the other baseline methods. To further analysis the robustness of the selected CART method, we also compared CART with GA based baselines. From the experimental results, the performance of all comparison algorithms has indeed improved by using GA for feature selection. However, CART with the proposed precursory pattern based feature extraction method still get the best performance since that CART can do feature selection automatically according to the information gain of each feature in the process of CART tree building. In summary, the proposed precursory pattern based feature extraction method with the selected CART model is a promising method for solving the task of earthquake prediction. In the future, more advanced earthquake prediction model should be designed to further improve the prediction performance.

REFERENCES

[1] C. H. Scholz, L. R. Sykes, and Y. P. Aggarwal, "Earthquake prediction: A physical basis," *Science*, vol. 181, no. 4102, pp. 803–810, 1973.

[2] C. Y. King, "Gas geochemistry applied to earthquake prediction: An overview," *J. Geophys. Res., Solid Earth*, vol. 91, no. B12, pp. 12269–12281, 1986.

[3] Y. Liu, Y. Wang, Y. Li, B. Zhang, and G. Wu, "Earthquake prediction by RBF neural network ensemble," in *Proc. Int. Symp. Neural Netw. Adv. Neural Netw. (ISNN)*, Dalian, China, Aug. 2004, pp. 962–969.

[4] S. M. Hamze-Ziabari and T. Bakhshpoori, "Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5' and CART algorithms," *Appl. Soft Comput.*, vol. 68, pp. 147–161, Jul. 2018.

[5] G. Asencio-Cortés, F. Martínez-Álvarez, A. Troncoso, and A. Morales-Esteban, "Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks," *Neural Comput. Appl.*, vol. 28, no. 5, pp. 1043–1055, 2017.

[6] E. Florido, F. Martínez-Álvarez, A. Morales-Esteban, J. Reyes, and J. L. Aznarte-Mellado, "Detecting precursory patterns to enhance earthquake prediction in chile," *Comput. Geosci.*, vol. 76, pp. 112–120, Mar. 2015.

[7] B. Gutenberg and C. F. Richter, "Earthquake magnitude, intensity, energy, and acceleration: (Second paper)," *Bull. Seismological Soc. Amer.*, vol. 46, no. 2, pp. 105–145, 1956.

[8] P. Nuannin, O. Kulhanek, and L. Persson, "Spatial and temporal *b* value anomalies preceding the devastating off coast of NW Sumatra earthquake of December 26, 2004," *Geophys. Res. Lett.*, vol. 32, no. 11, pp. 339–357, 2005.

[9] H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," *Neural Netw.*, vol. 22, no. 7, pp. 1018–1024, 2009.

[10] L. Brieman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Cart: Classification and regression trees," *Encyclopedia Ecol.*, vol. 40, no. 3, pp. 582–588, 1984.

[11] F. Külahcı, M. Inceöz, M. Doğru, E. Aksoy, and O. Baykara, "Artificial neural network model for earthquake prediction with radon monitoring," *Appl. Radiat. Isot.*, vol. 67, no. 1, pp. 212–219, 2009.

[12] Y. Wang, Y. Chen, and J. Zhang, "The application of rbf neural network in earthquake prediction," in *Proc. Int. Conf. Genetic Evol. Comput.*, Oct. 2010, pp. 465–468.

[13] S. Narayanakumar and K. Raja, "A BP artificial neural network model for earthquake magnitude prediction in Himalayas, India," *Circuits Syst.*, vol. 7, no. 11, pp. 3456–3468, 2016.

[14] A. Panakkat and H. Adeli, "Neural network models for earthquake magnitude prediction using multiple seismicity indicators," *Int. J. Neural Syst.*, vol. 17, no. 1, pp. 13–33, 2008.

[15] M. Mirrashid, "Earthquake magnitude prediction by adaptive neuro-fuzzy inference system (ANFIS) based on fuzzy C-means algorithm," *Natural Hazards*, vol. 74, no. 3, pp. 1577–1593, 2014.

[16] G. Asencio-Cortés, S. Scitovski, R. Scitovski, and F. Martínez-Álvarez, "Temporal analysis of croatian seismogenic zones to improve earthquake magnitude prediction," *Earth Sci. Informat.*, vol. 10, no. 3, pp. 303–320, 2017.

[17] F. Martínez-Álvarez, J. Reyes, A. Morales-Esteban, and C. Rubio-Escudero, "Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula," *Knowl.-Based Syst.*, vol. 50, no. 3, pp. 198–210, 2013.

[18] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Earthquake prediction model using support vector regressor and hybrid neural networks," *PLoS One*, vol. 13, no. 7, 2018, Art. no. e0199004.

[19] T. Utsu, "A method for determining the value of 'b' in a formula log n= a-bM showing the magnitude-frequency relation for earthquakes," *Geophys. Bull. Hokkaido Univ.*, vol. 13, pp. 99–103, 1965.

[20] L. Jones and P. Molnar, "Frequency of foreshocks," *Nature*, vol. 262, no. 5570, pp. 677–679, 1976.

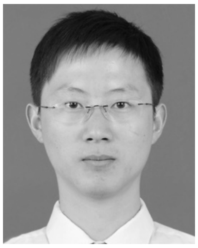
[21] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.

[22] B. Bai, J. Yen, and X. Yang, "False financial statements: Characteristics of China's listed companies and CART detecting approach," *Int. J. Inf. Technol. Decis. Making*, vol. 7, no. 2, pp. 339–359, 2008.

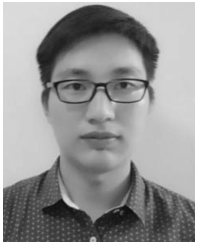
[23] J. L. Speiser, W. M. Lee, and C. J. Karvellas, "Predicting outcome on admission and post-admission for acetaminophen-induced acute liver failure using classification and regression tree models," *PLoS One*, vol. 10, no. 4, 2015, Art. no. e0122929.

[24] V. Franc and V. Hlavac, "Multi-class support vector machine," in *Proc. Int. Conf. Pattern Recognit.*, 2002, pp. 236–239.

[25] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 2, pp. 44–49, Mar. 1998.
[26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
[27] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.
[28] K. Tang, R. Wang, and T. Chen, "Towards maximizing the area under the roc curve for multi-class classification problems," in *Proc. AAAI Conf. Artif. Intell.*, 2011, pp. 483–488.



LEI ZHANG received the B.Sc. degree from Anhui Agriculture University, in 2007, and the Ph.D. degree from the University of Science and Technology of China, in 2014. He is currently an Associate Professor with the School of Computer Science and Technology, Anhui University, China. His main research interests include multi-objective optimization and applications, data mining, social network analysis, and pattern recommendation. He has published over 40 papers in refereed conferences and journals, such as the ACM SIGKDD, the ACM CIKM, the IEEE ICDM, the IEEE TCYB, the ACM TKDD, the IEEE CIM, and the Information Sciences. He is a recipient of the ACM CIKM'12 Best Student Paper Award.



LANGCHUN SI received the bachelor's degree from Anhui University, China. His main research interests include multi-objective optimization and applications, and data mining.

HAIPENG YANG received the bachelor's degree from Anhui University, China. His main research interests include multi-objective optimization and applications, and social network analysis.

YUANZHI HU received the bachelor's degree from Anhui University, China. His main research interests include data mining and social network analysis.

JIANFENG QIU received the B.Sc. degree from Anqing Normal University, in 2003, and the M.Sc. and Ph.D. degrees from Anhui University, China, in 2006 and 2014, respectively, where he is currently a Lecturer with the School of Computer Science and Technology. His main research interests include machine learning, imbalanced classification, multi-objective optimization, and complex networks.