Measurement 104 (2017) 180–191

Contents lists available at ScienceDirect

Measurement

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

Probability distribution pattern analysis and its application in the Acute Hypotensive Episodes prediction

Dazhi Jiang [a], Chenfeng Peng [b], Yifei Chen [a], Zhun Fan [c,d], Akhil Garg [e,*]

[a] Department of Computer Science, Shantou University, 515063, China
[b] Interactive Entertainment Group, Tencent Technology (Shenzhen) Company Ltd., 518057, China
[c] Department of Electronic Engineering, Shantou University, 515063, China
[d] Guangdong Provincial Key Laboratory of Digital Signal and Image Processing, Shantou University, 515063, China
[e] Department of Mechatronic Engineering, Shantou University, 515063, China

ARTICLE INFO

Article history:
Received 19 January 2017
Received in revised form 28 February 2017
Accepted 16 March 2017
Available online 18 March 2017

Keywords:
Acute Hypotensive Episodes
Probability Distribution Pattern Analysis
Time series
Prediction

ABSTRACT

At present many hospitals have to deal with the patient's care and nursing for Acute Hypotensive Episodes (AHE) occurring in Intensive Care Units. AHE can cause fainting or shock suddenly, leading to irreversibility organ damage, and even death. Therefore, forecasting of occurrence of AHE is of practical value. However, the prediction of clinical AHE largely depends on the doctors' experience, which cannot guarantee the high rate of success. It is thus very meaningful for the clinical care to use appropriate methods to predict the AHE with an automatic and reliable method. In this study, a Probability Distribution Patterns Analysis (PDPA) method is presented to solve the time series prediction problem of AHE. In the first phase, the features are extracted from the PDPA in the global and integral time series, and the partial local time series in the fixed time window. In the second phase, the proposed algorithm combining Genetic Algorithm (GA) and Support Vector Machine (SVM), namely GA-SVM is adopted to select the vital features for the effective classification. In order to demonstrate the generality of our method, we also conduct experiments on a classical time series problem-Control Chart Patterns (CCPs) multi-class time series, which is a benchmark problem in the process control. For CCPs problem, the experimental results demonstrate that the proposed method outperforms several traditional methods. The obtained accuracy is 98.65%, which is superior to listed previous works using the same CCPs model. For AHE classification and forecasting, the methodology is applied in two data sets, a small data set (37 records) and a big one (2892 records). The test accuracy of 89.19%, sensitivity of 91.67%, specificity of 88% in the small data set, and a test accuracy of 80.76%, sensitivity of 78.19%, specificity of 81.51% in the big data set are achieved with the classification model.

1. Introduction

Depend on the urgency of the symptom, the hypotensive episode can be divided into two classes, Chronic Hypotensive Episode (CHE) and Acute Hypotensive Episode (AHE). CHE is defined as Systolic Arterial Blood Pressure (SABP) lower than 90 mmHg, and Diastolic Arterial Blood Pressure (DABP) lower than 60 mmHg. On the contrary, the AHE is defined as any period of 30 min or more during which at least 90% of the mean arterial pressure (MAP) measurements are at or below 60 mmHg. For the normal human blood pressure, the SABP is between 90 and 140 mmHg, and DABP is between 60 and 90 mmHg, and the MAP is between 70 and 105 mmHg [1].

The AHE is a series and common postoperative complication in the Intensive Care Units (ICU), which can damage the patient's organs and eventually, cause the patients to decease (If not promptly and proper treated in time). Thus, the early detection of AHE by computer becomes vitally important for clinical therapy and intervention in AHE. As provided in PhysioNet [2], the episodes of AHE can be predicted with reasonably high specificity (in 2009, the PhysioNet held a research competition of AHE prediction, which let the AHE automated prediction become a worldwide research topic). In this challenge, the data used to AHE prediction is blood pressure, which is a typical complex and non-linear time series data. Compared with other problems, the AHE prediction based on the time series is more complicated, because the human

* Corresponding author.
E-mail address: akhil@stu.edu.cn (A. Garg).

http://dx.doi.org/10.1016/j.measurement.2017.03.030
0263-2241/

body produces the blood pressure data quickly, heavily and dynamically.

Many studies have been done for building automatic prediction AHE system based on the machine learning and data mining methods. In these methods, neural network was frequently used to create the classifier for AHE prediction. For example, Henriques et al. developed a multi-models neural network for the forecast of the AHE [3]. In 2011, Teresa Rocha et al. [4] used the similar method to solve the AHE problem in a big data sets (the total number of samples is 2344, the proposed method obtained sensitivity 82.8%, and specificity 78.4%). Furthermore, Zhou et al. used a Chebyshev neural network to solve the challenge dataset (110 samples), and the result shows that the proposed method performs better than other solutions in AHE prediction [5]. In the thesis of M Ghassemi, the multi-variate neural network is the best method for AHE prediction which is tested in 1168 samples data sets. For each sample, the vital values are MAP, respiration rate, oxygenation level and heart rate. The yielded accuracy is 84% [6].

For other methods, Hoseinnia et al. presented a hybrid approach for predicting AHE, which is based on the wavelet transform and neural network. The wavelet transform is used to decompose the MAP, and the neural network is used to forecast the wavelet approximation coefficients [7]. Arasteh et al. presented to use Empirical Mode Decomposition (EMD) method to solve the AHE prediction. EMD is used to decompose the MAP time series into several Intrinsic Mode Functions (IMFs), and then some statistical features are extracted from the IMFs. Finally, the SVM is applied to the classification [8]. Lehman et al. [9] use a combination of Gaussian Mixture Model based clustering and K-Nearest Neighbours Classier on 227 patient records from the MIMIC II database using both Heart Rate and MAP measurements. They achieve an accuracy of 70% with 74% sensitivity and 60% specicity. Lee et al. [10], who use 1311 patient records using heart rate, MAP and clinical data from the MIMIC II database. Their algorithm uses 102 statistical, wavelet-based and clinical features and predicts the event 1 h in advance with 86% accuracy.

In our research group, Sun et al. predict the AHE with the Particle Swarm Optimizer (PSO) and K-means method. The method based on PSO and K-means is to extract the vital features of MAP data, then the SVM method is used to create the classifier model. The experiment is verified on the 2863 samples, and the best accuracy is 81.2% [11]. Jiang et al. used the EMD method to calculate patient's MAP data. Then, for features, the bandwidth of amplitude modulation, the frequency modulation and the power of IMFs were extracted. The total features are 5. A Multiple Genetic Programming (Multi-GP) is presented for the classification of the AHE [12,13]. Furthermore, in 2016, Jiang et al. transform the IMFs data into probabilistic distribution, then the statistic features, such as Peak, Mode, Skewness, Kurtosis and Shannon Entropy are extracted in different IMFs probabilistic distribution. Multi-GP was still used to create the classifier models, and the best achieved training and testing accuracy are 82.92% and 79.93% respectively [14].

According to the previous research, the features selection and classification method are two critical factors for improving the prediction performance of AHE. As a natural extension of previous work [14], this paper presents a generic methodology for time series prediction, which extracts features base on statistics and Probability Distribution Patterns Analysis (PDPA). Furthermore, in order to describe the dynamic distribution patterns in time series, the features are not only extracted from the global and integral time series, but also from the local and partial time series in the fixed time window. After features extraction, a method, which is combined with Genetic Algorithm and Support Vector Machine (GA-SVM), is presented to create classifier and select features simultaneously. All the features were extracted based on statistic so the process is very forthright and fast. This characteristic is especially important when the method is adopted in a real-time monitoring system of medical device. In addition, this method has a general applicability on the time series prediction and classification problem. In the experimental verification, for CCPs problem [15–17], the obtained accuracy is 98.65%, which is superior to listed previous works used the same CCPs model. For AHE classification and forecasting, the methodology is applied in the two data sets, the test accuracy of 89.19% in the small data set and 80.76% in the big data set are achieved from the classification model.

The rest of the paper is organized as follows: in Section 2, the experimental data sets and the prediction of AHE problem are briefly introduced. Section 3 describes the methodology for feature extraction and selection. Section 4 is the experiments verification, includes typical multi-class time series problem, small data sets of AHE and big data sets of AHE. Section 5 are conclusions and future research topic.

2. Data sets

As previous work, the data for the AHE experiment was collected from the Multi-Intelligent Monitoring in Intensive Care (MIMIC) II [18]. The data, where ware got from MIMICII database, regards a patient as a unit, and records the patients' vital signs, such as systolic arterial blood pressure (SABP), and diastolic arterial blood pressure (DABP). The SABP (Fig. 1 red[1] box) and DABP (Fig. 1 green circle) are the maximum pressure and the minimum pressure respectively. In this experiment, we focus on the mean arterial pressure (MAP), which is actually a combination of the SABP and DABP, and calculated as follows:

$$MAP = DABP + \frac{SABP - DABP}{3}$$

According to the above calculation method, the ABP data could be transformed into the MAP data as follows (see Fig. 2):

For AHE prediction, the validation set consists of two datasets, which are a small dataset and a big one. The small dataset was obtained from PhysioNet 2009 challenge [19], and the big dataset is downloaded from MIMICII [18]. In both datasets, instant T_0 is a marked stamp for the prediction. For the small dataset, in the training set, the data records contain all the data before and after instant T_0. In the testing set, the records are truncated at T_0 for the purpose of performance testing. Which means in the training set, every record contains the data obtained from 2 h before T_0 and 1 h after T_0. In the testing sets, the data is only collected from 2 h before T_0. In the small dataset, because some data are missing, only 48 records are selected as the training set, and 37 records as the testing set (two classification problems, AHE and NO_AHE problem, AHE means the patient will suffer in during the forecast widow and NO_AHE means no AHE symptom). The big dataset contains 2892 records. 600 records are selected randomly as the training set, which contains 300 AHE records and 300 NO_AHE records. The remaining is the testing set, which has 2292 records.

3. Methodology

Normally, the pattern in the systems and processes can be conveyed in the form of probabilistic distribution functions (PDFs) [20]. It inspired us to extract data PAP from time series to gain an insight into the underlying distribution pattern in the global and integral time series (such as the 2 h data before T_0 for AHE problem) and local distribution pattern in the local and partial time series in the fixed time window.

[1] For interpretation of color in Figs. 1 and 10, the reader is referred to the web version of this article.

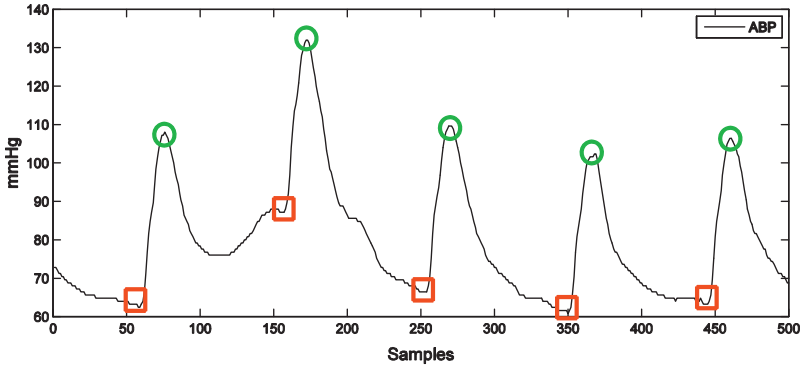

Fig. 1. ABP, SABP and DABP.

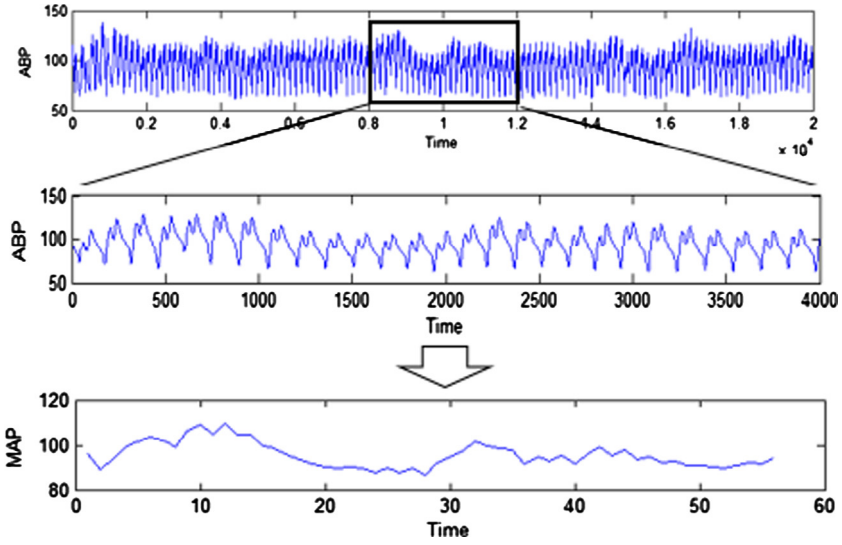

Fig. 2. Transform ABP to MAP data.

Actually, probability distribution, global distribution pattern and local distribution pattern are not new concepts for time series. For example, E. Bautista-Thompsom and S. Santos De la Cruz designed histogram and shape similarity to compare two time series. The histogram shows the statistical distribution of point to point difference between two time series, and the shape similarity between two series is quantified based on descriptive statistical features from histograms of Euclidean distances [21]. For AHE prediction, TCT Ho and X Chen allocated the MAP data into bines to form a histogram for analysis, and then the separation of NO_AHE and AHE can be deduced according to the histogram [22]. Vaibhav Awandekar and A.N. Cheeran presented to calculate probability distributions of MAP time series, then the Bhattacharyya distances is adopted to measure the similarity of two discrete or continuous probability distributions [23]. Alexander Waldin [24] envisions a Dynamic Bayesian Network (DBN) system that can use varying amounts of a patient's blood pressure history and that can predict at varying times in the future. This method cut the whole time series into sub-segments, and for each sub-segment, three kinds of aggregate features, including Moments, Trends, and Differentials, are abstracted for each time step in DBN. According to the above analysis, it can be roughly categorized that the shape similarity, Euclidean distances and Bhattacharyya distances are the global distribution pattern; Moments, Trends, and Differentials are the local distribution pattern.

In the research of AHE forecast model, the most important thing is to extract useful features to distinguish AHE occurrence or NO_AHE occurrence from the original MAP data. Three approaches are proposed for feature extraction in this paper, namely probability distribution, global distribution patterns and local distribution patterns. The extraction procedure consists of two steps. The first step gets the shape of data distribution. The second step extracts features from the global and local distribution pattern of time series data. For global distribution pattern, the statistical analysis methods are used to get features, which include skewness, kurtosis, peak, mode, mean, standard deviation (STD), median, mean absolute deviation (MAD) and Shannon Entropy. For local distribution patterns, an average method based on fixed-length sliding window is adopted to get ten features eventually, which will be discussed in detail later.

However, when the extract procedures is finished, maybe not all of features are useful for specific problem. Therefore, the feature selection procedure is necessary. A combination of genetic algorithm (GA) [25] and support vector machine (SVM) [26] (GA-SVM) are used to select useful features for specific applications.

3.1. Features extraction

As mentioned before, two hours MAP data before T_0 is used to predict AHE. The data we used is real clinical data, which has pos-

sibly been polluted by noise. Ideally, the noise should be removed before features extraction. In medicine, the MAP above 140 mmHg belongs to hypertension, and less than 35 mmHg means the patient is in coma or death state. In order to achieve better analysis and forecasting, we only need MAP in the range of 35–140 mmHg. A simple moving average method, which uses the average of ten data before the current point to replace this current point, is used to remove noise.

A. *Distribution pattern*

Firstly, do some data preprocessing on Time series $S = \{s_1 s_2 \dots s_L\}$), let $s'_i = [s_i]$, where $[\cdot]$ is rounded up. Then we get a new series $S' = \{s'_1 s'_2 \dots s'_L\}$. Next we employ the usual histogram method to make a statistic on S'. C_i is the number of occurrences of a certain MAP value.

$$P_i = \frac{C_i}{\sum_{i \in M} C_i}$$

Here we give a definition of P_i, which is calculated by the above-mentioned formula. Obviously, P_i is the probability that particular MAP value appeared in S'. Furthermore, M is a set of all MAP value in S'. Draw a line of P_i, $i \in M$. Then the the data distribution pattern was obtianed. The above-mentioned description could be illustrated with the following figure (Fig. 3). Two samples (NO_AHE and AHE respectively) are used to explain the differences of the probability distribution in different condition.

B. *Feature extraction*

For the feature extraction of data distribution pattern, the skewness and kurtosis are often extracted as features. Here, we extract more features from distribution pattern of global time series. In addition, we extract features from local distribution patterns of a sequence of intervals represented by a sliding time window. The features extraction in the distribution pattern of global time series has been described in the previous work [14]. This work is a natural derivation of the previous work, and the rational of the features extraction in the global and local distribution pattern is that the distribution pattern of global time series will miss the time domain information. Therefore, in order to retain some time domain information we divide the whole time series data into a sequence of time windows and get the distribution information in every time windows. The procedure of the whole features extraction is illustrated in Fig. 4 and 5.

(1) *Global features*

Peak: Peak value is the max point in the MAP probability distribution.

$$Peak = max\{Fre_{x_1}, Fre_{x_2}, \dots, Fre_{x_n}\}$$

Fre_{x_i} is the frequency value of MAP value x_i, and n is the number of different MAP values.

Mode: With the Peak value, the Mode value can be obtained. The Mode value is the most frequently value appeared in each MAP time series.

Skewness: Skewness is a measure of the asymmetry of the data around the sample mean.

$$skewness = \frac{E(x-\mu)^3}{\sigma^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^3}$$

where μ is the mean of $X = (x_1, x_2, \dots, x_n)$, and σ is the standard deviation of X.

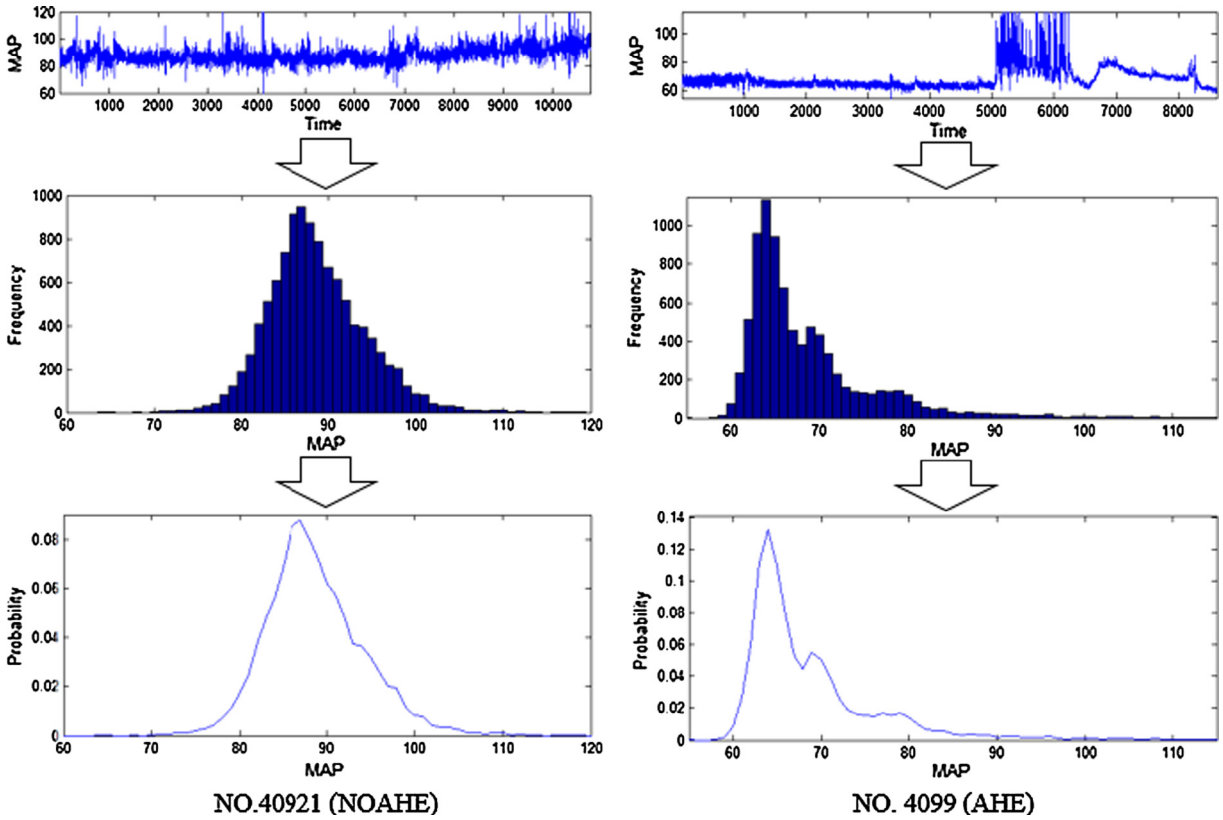

Fig. 3. The Steps for transforming MAP to probability distribution.

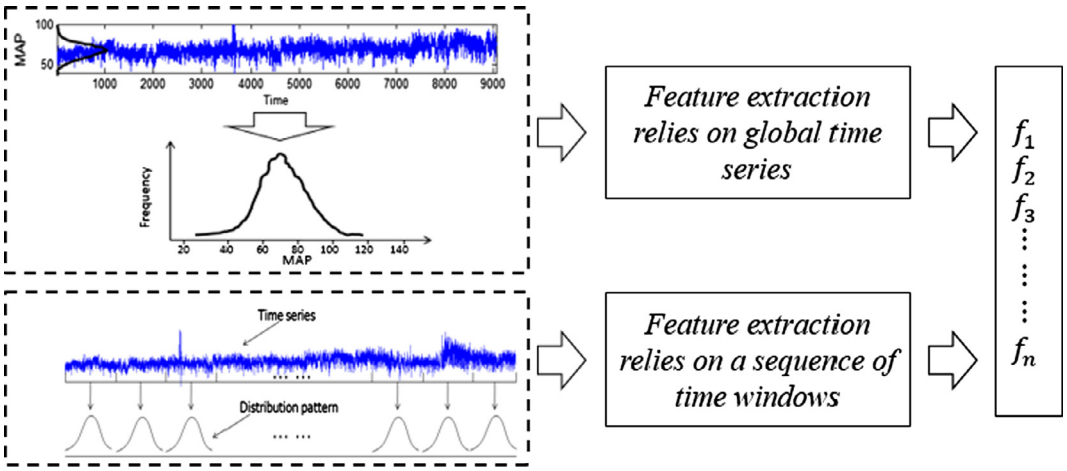

Fig. 4. Feature extraction: global distribution pattern and local distribution pattern.

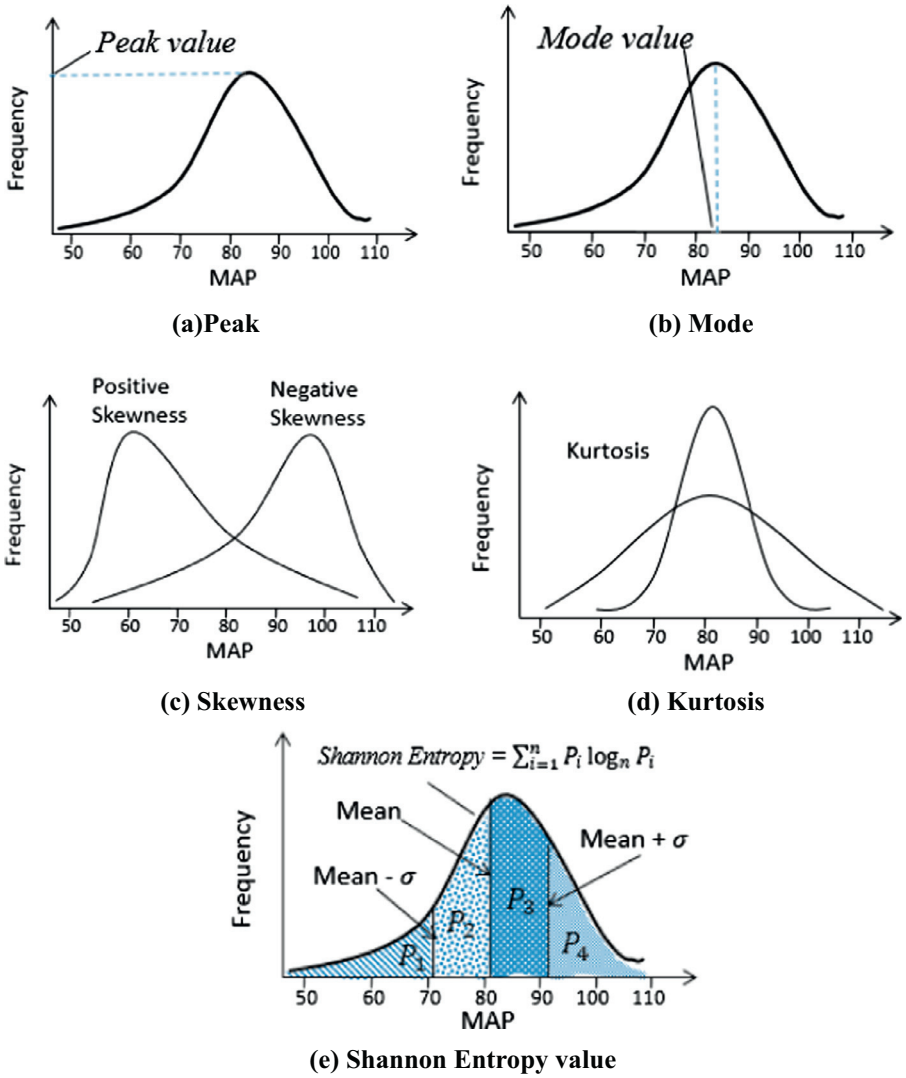

Fig. 5. Peak, Mode, Skewness, Kurtosis, Shannon Entropy value in data distribution pattern.

Kurtosis: We measure the Kurtosis of the MAP probability distribution, and which is calculated as follows.

$$kurtosis = \frac{E(x-\mu)^4}{\sigma^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

Shannon Entropy: Shannon Entropy is a complexity measurement method of the system. This paper use the Shannon Entropy as the index of the complex of the MAP time series.

$$Shannon\ Entropy = \sum_{i=1}^{m} P_i \log_m P_i$$

where m is the fixed size of for separation in the abscissa axis of MAP probability distribution. In the AHE prediction, $m = 4$. The separation are separated by the three points, which are $Mean - \mu$, $Mean$ and $Mean + \mu$ respectively. Then, P_i is the size in each separation.

Besides the features mentioned above, several classical features are selected for time series analysis, such as the mean, standard deviation (std), median and median absolute deviation (MAD).

(2) *Local Varying Features*

In the extraction procedure of local varying features, a whole time series $S = \{s_1 s_2 \dots s_L\}$ is divided into M time windows. For each time window, the feature extraction procedure is the same as to the procedure on global time series. For example, if we focus on feature f_g, then it will be extracted from each time window, so we will get a feature sequence $F_g = \{f_{g1}, f_{g2}, \dots, f_{gM}\}$ from the M time windows. Then, the local varying features could be obtained based on the sequenceF_g.

Definition: For a pair of neighbors $\{f_{g(i)}, f_{g(i+1)}\}$ in F_g, we define $drift = f_{g(i+1)} - f_{g(i)}$. As a result, we can get a series of *drift* $D = \{drift_1, drift_2, \dots, drift_{M-1}\}$.

In this way, several local varying features will be extracted from D which contains largest positive drift (*LPD*), largest negative drift (*LND*), largest absolute drift (*LAD*), summation drift (*SD*) and average absolute drift (*AAD*) (see Fig. 6).

LPD: $LPD = \max\{drift_1, drift_2, \dots, drift_{M-1}\}$, which captures the largest positive jump of a feature as the time goes on. If all drift in D are negative, then LPD = 0.

LND: $LND = \max\{-drift_1, -drift_2, \dots, -drift_{M-1}\}$, which captures the largest negative mutation of a feature as the time goes on. If all drift in D are positive, then LND = 0.

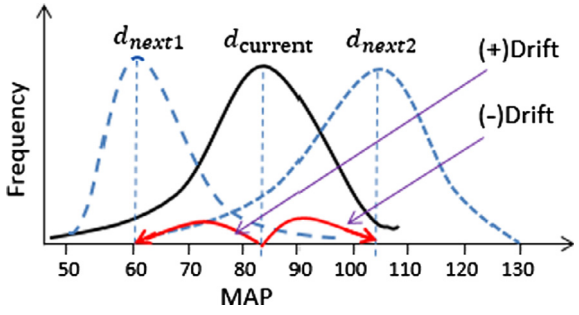

Fig. 6. Change of data distribution pattern.

LAD: $LAD = \max\{|drift_1|, |drift_2|, \dots, |drift_{M-1}|\}$, which captures the largest jump of a feature as the time goes on. The jump direction can be positive or negative.

SD: $SD = \sum_{i=1}^{M-1} drift_i$, which describes the trend of a feature as the time goes on. If $SD > 0$, the overall trend of time series towards to the positive direction, and vice versa.

AAD: $AAD = \sum_{i=1}^{M-1} |drift_i|/(M-1)$, which describe the Instability Grade of a feature as the time goes on.

Remarkably, more features can be extracted from the data distribution pattern. Taking into account the amount of computation, only 9 features in global distribution pattern and 10 local varying features (*LPD*, *LND*, *LAD*, *SD*, *AAD* in Mode and Mean) are extracted for problem solving in this paper. Meanwhile, maybe only a proportion of them are useful for a specific application, so in the next section, a method will be adopted to select the features needed for the specific application.

3.2. GA-SVM for features selection

After features extraction, the features selection is a vital part in the pattern recognition system. For the problem we solved, the proper set of features should be constructed. However, the irrelevant and redundant features will affect the prediction accuracy. What's more, for the different problems, the proper features set may be different for each other. So, it is vital to select the feature set automatically that is closely related with the occurrence of AHE. Here, a method that combines Genetic Algorithm (GA) and Support Vector Machine (SVM), named GA-SVM approach was used to select the feature set.

Genetic Algorithm (GA), is a self-adaptive and self-learning search algorithm which can be used to solve complicated optimization problems automatically. Support Vector Machine (SVM) is an algorithm based on the principle of minimizing structural risk that contains high generalization ability. In this paper, GA is combined with SVM to find the optimal feature set for the specific given problems. Furthermore, the K-Fold validation [27] based on SVM is adopted to get the average classification accuracy (Fig. 7).

(1) Chromosome coding and initial population

Binary coding is adopted in designing the chromosome. A chromosome $X = \{x_1, x_2, \dots, x_n\}$ $x_i \in \{0, 1\}$ represents a feature set $F = \{f_1, f_2, \dots, f_n\}$. $x_i = 1$ indicates the corresponding feature f_i is selected and $x_i = 0$ indicates f_i is unselected. The initial population consists of P_{size} individuals generated randomly.

(2) Fitness Calculation

- Get the Training set

In Fig. 8, M is the number of records in the original data set. In each record, features are selected if the corresponding bit value is '1'.

- SVM_K-Fold Cross validation

Cross-validation (CV) is a model validation technique. It is a statistical analysis method which is used to verify the performance of the classifier algorithm. In K-Fold Cross validation, the original samples are randomly partitioned into k equal subset samples. Then the $k-1$ subset samples are used to train the model, and the last subset samples are used to test the model. After repeat

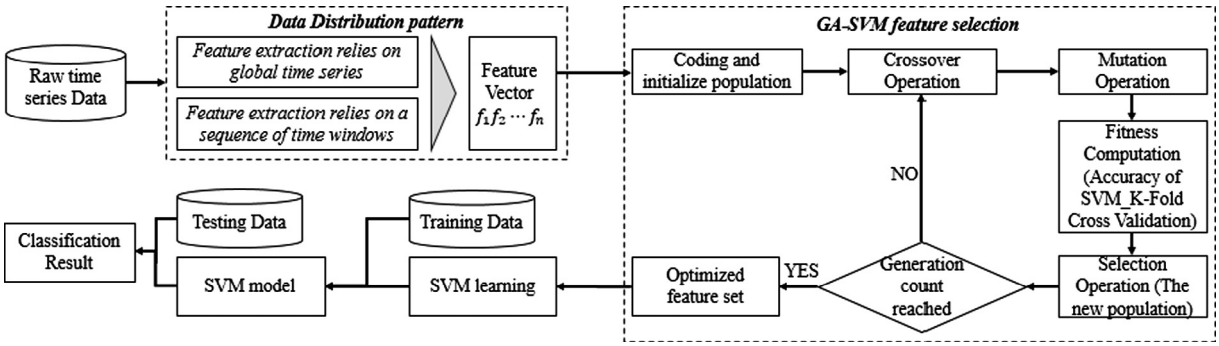

Fig. 7. The general process of feature selection and classification.

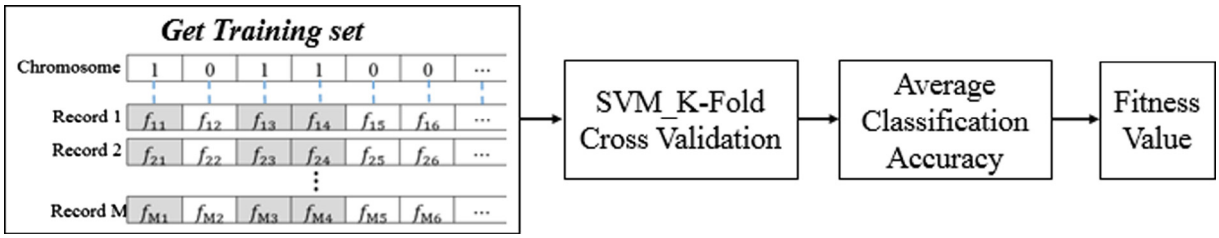

Fig. 8. The process of fitness calculation.

the cross validation process k times, the average of result from k fold is the estimation of average accuracy.

- Fitness Function

The individual fitness value is calculated based on the average accuracy of SVM_K-Fold.

The function is described as follows:

$$F(X) = F(x_1 x_2 \ldots x_n) = Acc(x_1 x_2 \ldots x_n) - \lambda \sum_{i=1}^{n} x_i$$

where X is individual, x_i is the$i$th bit. $Acc(X)$ is the average classification accuracy of SVM_K-Fold on X. $\sum_{i=1}^{n} x_i$ is the quantity of selected features. λ is a parameter which can be used to adjust the length of features selected and set by the user. For example, if a smaller value to λ is set, relatively high accuracy may be achieved with more feature selected. On the contrary, if a larger value to λ is set, a feature subset with fewer features may be obtained.

4. Experimental results

This section consists of two parts. Part one is an experimental verification for solving a multi-class classification problem using the proposed methodology. Part two attempts to use the proposed method to solve the binary classification problem for AHE prediction.

4.1 Multi-class time series

Control Chart Patterns (CCPs) is a typical and classical multi-class time series, which can be used to test the performance of the time series classification methods.

(A) *Data set*

The CCPs have six patterns, which are Downward Trend (A), Cyclic (B), Normal (C), Upward Shift (D), Upward Trend (E) and (F) Downward Shift. Fig. 9 shows the six different pattern time series and each pattern are consisted with ten examples.

However, it is difficult and time consuming to obtain. Therefore, the mathematical method, such as GARH (Generalized Autoregressive Conditional Heteroskedasticity Model) [28] has been widely used to generate the controls chart patterns. The mathematical equations are utilized in this work can be found in [29] and are shown as follows.

(A) Downward Trend: $y(t) = \mu + r(t)\sigma - gt$
(B) Cyclic: $y(t) = \mu + r(t)\sigma + a\sin(2\pi t/T)$
(C) Normal: $y(t) = \mu + r(t)\sigma$
(D) Upward Shift: $y(t) = \mu + r(t)\sigma + ks$
(E) Upward Trend: $y(t) = \mu + r(t)\sigma + gt$
(F) Downward Shift: $y(t) = \mu + r(t)\sigma - ks$

Where $y(t)$ is time series value; μ is mean value; $\boldsymbol{r(.)}$ is normally distributed random number; **t** is time; σ is standard deviation; **a** is amplitude of cyclic variations; **g** is magnitude of gradient trend; **k** is determines shift position; **s** is shift magnitude; **T** is period of cycle.

This dataset, which are obtained from Alcock and Manolopoulos [30], contains 600 examples of control charts synthetically.

(B) *Result*

In the experiment we select 70 records randomly from each class as the training set, and the remaining records are used as the testing set. The experimental result is as follows in Fig. 10.

After GA-SVM running for 200 generations, 100 relatively good individuals are obtained (blue dot in Fig. 10, each individual is a subset of features), which are drawn in the Fig. 10. The ultimate aim of feature selection is to find a feature subset which contains as few features as possible and obtains as high classification accuracy as possible. Here, the Pareto front [31] is adopted to analyze the feature subsets in order to get the minimal feature set which is significantly related to the problem and get the maximal classification accuracy. In Fig. 10, the red line is Pareto front. The dots on

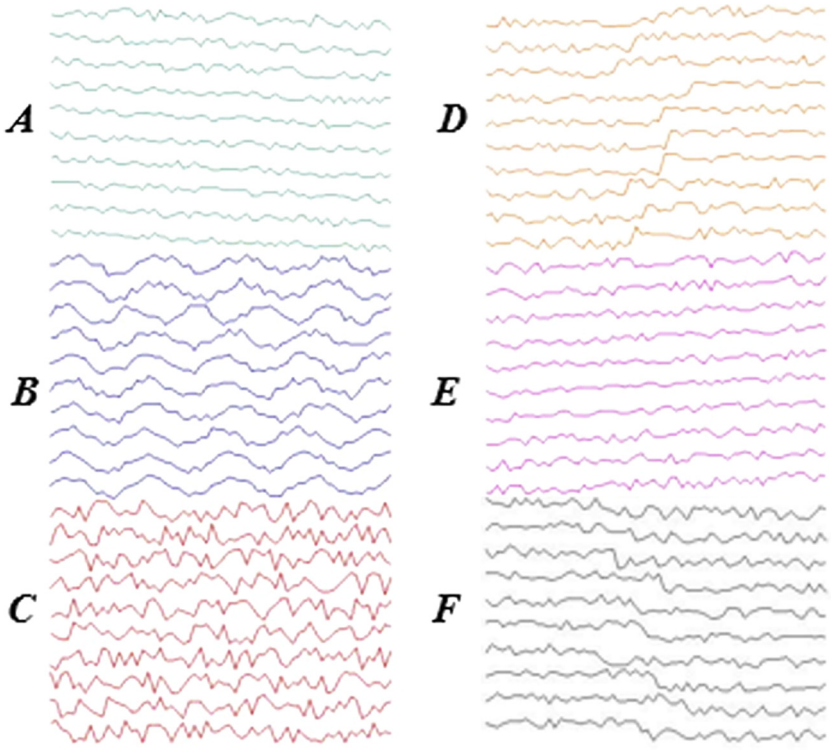

Fig. 9. Example of six control chart patterns.

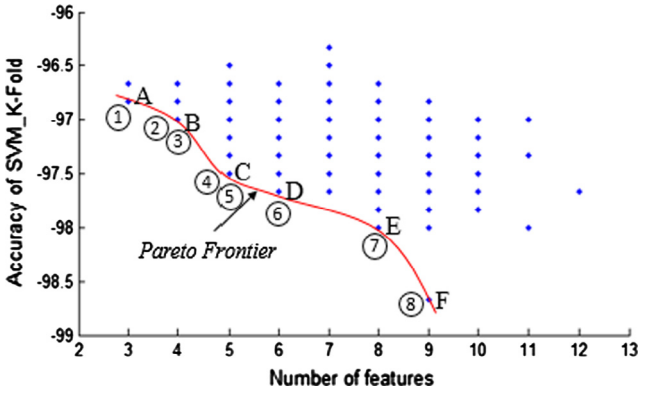

Fig. 10. Pareto Frontier of feature set of the CCPs.

the line are non-dominated trade-offs in the space of feature dimension and classification accuracy.

We focus on the last generation of GA every time. After GA-SVM approach, the Pareto frontier set contains 6 dots A, B, C, D, E, F. Actually, dots B and C both have two feature sets, which are ID2, 3 and ID4, 5 respectively (see Table 1). ID2 and ID3 are two feature sets which have the same feature number (4 features) and same K-Fold accuracy (97%), and so are ID4 and ID5 (5 features and 97.5%). Finally, totally 8 subset of features are obtained after GA-SVM approach (ID1 ∼ ID8).

Fig. 11 shows the frequency of each feature 8 Pareto frontier features set. According to Fig. 11, we find that two features, mean (15th) and MAD (18th) are always appear in the features set, which means that these two features are significantly related to CCPs classification problem. However, only two features are not enough

Table 1
Performance compared with different Pareto frontier feature set for CCPs problem.

ID	Feature set	K-Fold (%)	Data set	Accuracy (%)
1	*15,17,18*	96.8333	Training set	98.3333
			Testing set	95.5556
2	*13,14,15, 18*	97	Training set	98.5714
			Testing set	94.4444
3	*11,13,15, 18*	97	Training set	98.8095
			Testing set	93.3333
4	*1,13,15,17, 18*	97.5	Training set	98.5714
			Testing set	93.8889
5	*1,3,13,15, 18*	97.5	Training set	98.3333
			Testing set	95
6	*12,15,16,17, 18, 19*	97.6667	Training set	98.3333
			Testing set	93.8889
7	*1,3,10,13,14,15, 18*	98	Training set	99.5238
			Testing set	97.7778
8	*1,3,9,11,14,15,17,18,19*	98.6667	Training set	99.5238
			Testing set	97.7778

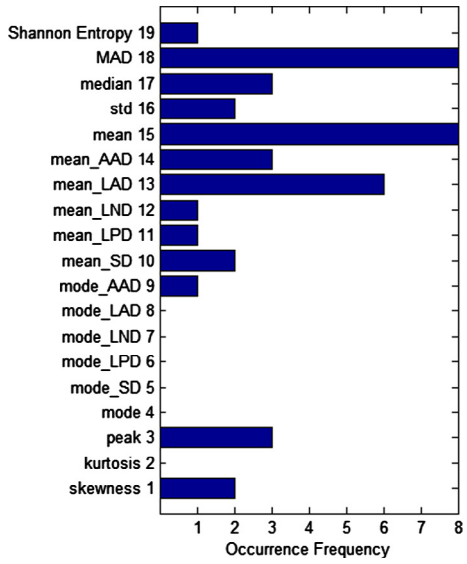

Fig. 11. Frequency of the Pareto frontier feature set of CCPs.

for problem solving, and then the features extracted from the probability distribution pattern analysis approach can help get a better result.

The result obtained by previous work and our work are list in Table 2, which shows that our method gets the best average accuracy among them.

4.2 MAP time series

(A) *The small data set*

Independent experiments were conducted 20 times in order to get a more credible result of the AHE classification. The experiments' parameter setting is given in Table 3.

We focus on the last generation of GA every time, and get a statistics result as shown in Fig. 12.

In Fig. 12, the point B has two feature sets, ID 2 and 3, and the point C has three feature sets, ID 4, 5 and 6. These feature sets are listed in Table 4.

We therefore can make a histogram on the Pareto frontier set to check the frequency of every feature selected by the feature sets. The result is displayed in Fig. 13.

Table 2
Comparison among previous work and the proposed method.

	Method	Accuracy (%)
1	Neural Network based classifier [32]	90.47
2	Evolutionary Computation (SARG) based classifier [33]	98.00
3	Multilayer Perceptron Classifier with SAX Preprocessing [30]	94.78
4	Time-lag Classifier with SAX Preprocessing [30]	98.50
5	Our proposed Method	98.65

Table 3
Parameters setting.

Parameter	Description	Value
P_{size}	Size of population	100
λ	Concise weight	0.1
C	Punishment coefficient	1
δ	Width of Gaussian kernels	$1/n$
w	Window size	L/20
k	K-Fold validation	3

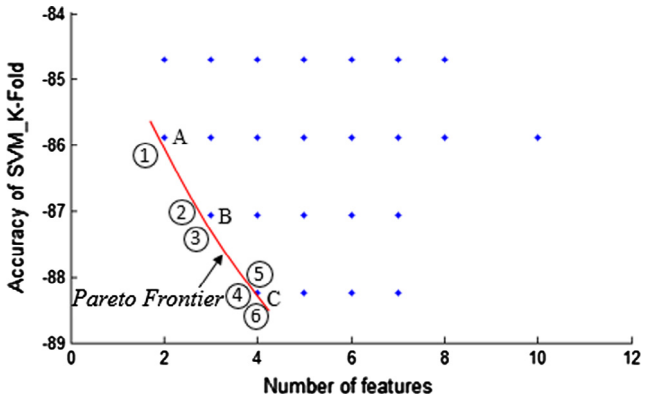

Fig. 12. Pareto Frontier of feature subsets of the small data set.

From Fig. 13 and Table 4, the features mode and MAD are most frequently present together in the Pareto frontier. According to this, we suggest that the combination of mode and MAD is an important factor for AHE detection. Coincidentally, the two features both describe the centralized trend of the data. That means that there exist a difference between AHE and NO_AHE on the centralized trend of MAP. In Table 4, the feature set of ID 1 only has two features, which are Mode and MAD, and its K-Fold result is 85.8823%. It can be seen that the result of ID 1 still has considerable space to improve. After combining with some other PDPA features defined in Fig. 13 (i.e. skewness and Shanon Entropy), the feature set ID 6 gets a higher accuracy, which indicates that the PDPA features may have a positive role in distinguishing between AHE and NO_AHE in a more detailed level. In order to get robust prediction effect on big unknown complex data, the feature set we finally selected contains all the features appearing in the Pareto frontier set. In Table 5, all 8 features are used to classify the AHE and NO_AHE data, and obtain a classification accuracy is 100% in the training set and 89.1892% in the testing set (see Figs. 14 and 15).

(B) *The big data set*

From Table 6, we find that the features selected for big data set are different from those for small data set. One reason for this may be that the small data is the MAP transformed from ABP signals sampled at 125 Hz, but the big data is the MAP signal measured directly on blood pressure sampled at 1 Hz. In this experiment, the feature 9 mode_AAD and 15 mean appear most frequently in the Pareto frontier. From the definition of AHE, it is easier to interpret that mean is a valuable feature - a lower mean value of MAP implies that AHE more likely occurs. The feature mode_AAD is the average absolute drift of mode as the time goes on, which describes the instability level of the time series data.

The feature sets listed in Table 7 are selected from the Pareto frontier. In Table 7, all 7 features are used to classify the ANE and NO_AHE data and obtain a classification accuracy of 85% in the training set and 80.7592% in the testing set.

According to Tables 5 and 7, we can find that the feature sets selected for the small set and the big set are very different. Besides the different methods of data collection, we think that the different data scales may be another important factor that contributes to the difference of features sets selected and used for prediction. To verify this, we therefore make a cross experiment with the results shown from Tables 8–11.

According to the result, we can find that in small data set, the result of Sensitivity, Specificity and Accuracy for training set are declined 4.55%, 13.64% and 9.09% respectively. Meanwhile, the Sensitivity, Specificity and Accuracy for testing set are declined

Table 4
Performance compared with different Pareto frontier feature set for Small data set.

ID	Feature set	K-Fold (%)	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	*4, 18*	85.8823	Training set	100	96	97.9167
			Testing set	91.67	88	89.1892
2	*2, 4, 18*	87.0588	Training set	100	100	100
			Testing set	83.33	88	86.4865
3	*1, 4, 18*	87.0588	Training set	100	96	97.9167
			Testing set	91.67	92	91.8919
4	*3, 4, 18, 19*	88.2352	Training set	100	92	95.8333
			Testing set	91.67	84	86.4865
5	*3, 4, 14, 17*	88.2352	Training set	100	100	100
			Testing set	91.67	88	89.1892
6	*1, 4, 18, 19*	88.2352	Training set	100	96	97.9167
			Testing set	91.67	92	91.8919

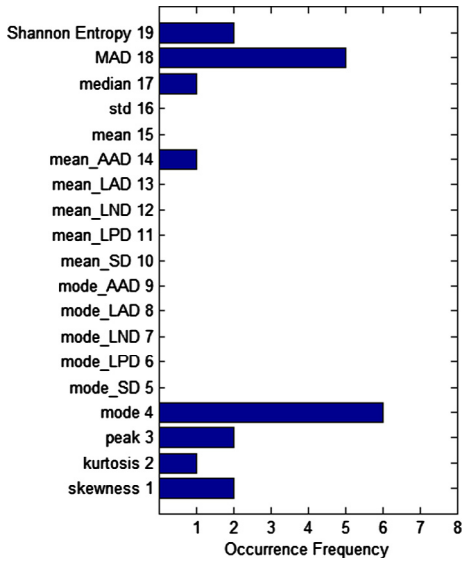

Fig. 13. Frequency of the Pareto frontier feature set of the Small data set.

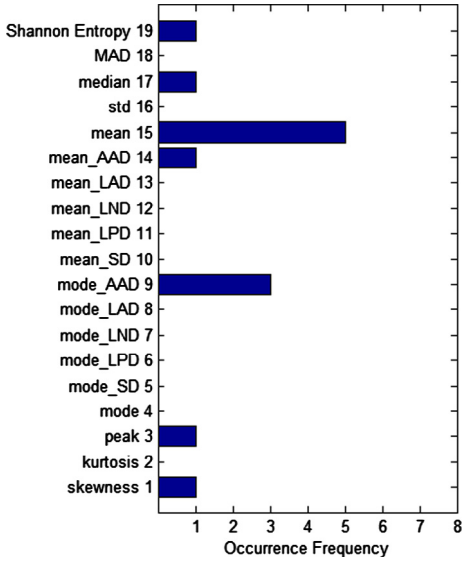

Fig. 15. Frequency of the Pareto frontier feature set of the big data set.

Table 5
Sensitivity, specificity and accuracy obtained with the hybrid feature set.

Feature set	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1, 2, 3, 4, 14, 17, 18, 19	Training set	100	100	100
	Testing set	91.67	88	89.1892

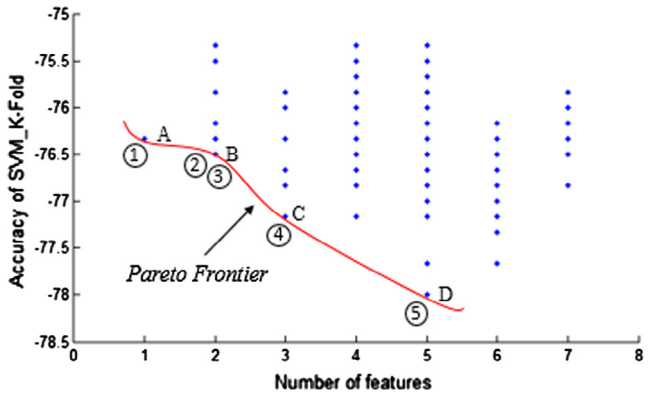

Fig. 14. Pareto Frontier of feature subsets of the big data set.

10.01%, 29.41% and 22.22% respectively. However, in big data set, the results of Sensitivity, Specificity and Accuracy for training set are improved 7.37%, 0.77% and 4.14% respectively. Meanwhile, the Sensitivity, Specificity and Accuracy for testing set are declined 8.01%, 2.7% and 3.81% respectively.

It can be seen that the total accuracy is reduced for two different features sets. Because the small data set is incomplete, so the feature sets selected based on the small data set cannot fit more data very well. However, when the feature sets selected based on the big data set were used to test the small data set, the test result is still worse similarly.

Feature sets that be selected based on big data were used to test the small data.

5. Conclusion

As a typical medical time series data, MAP signals for AHE classification are analyzed in this study. A PDPA approach is proposed to solve this time series prediction problem. Then the features are extracted from the PDPA in the global and integral time series, and the partial local time series in the fixed time window. Through the GA-SVM selection process, the features extracted from distribution pattern and valuable to classify the time series are selected and used for predication of AHE. The methodology is applied in the three datasets, including the CCPs, the small set of 10th PhysioNet Challenge, and the big sets obtained from MIMICII. The result

Table 6
Performance compared with different Pareto frontier feature set for big data set.

ID	Feature set	K-Fold (%)	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	*15*	76.3333	Training set	72.67	82	77.3333
			Testing set	72.39	81.45	79.4066
2	*15, 17*	76.5	Training set	85	82	83.5
			Testing set	76.04	78.57	76.6143
3	*9, 15*	76.5	Training set	78.33	84.67	81.5
			Testing set	73.94	80.78	79.2321
4	*1, 9, 15*	77.1666	Training set	81	86	83.5
			Testing set	75.29	81.34	79.9738
5	*3, 9, 14, 15, 19*	78	Training set	77.67	84.67	81.1667
			Testing set	77.61	80.16	79.5812

Table 7
Sensitivity, specificity and accuracy obtained with the hybrid feature set.

Feature set	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1, 3, 9, 14, 15, 17, 19	Training set	83.67	86.33	85
	Testing set	78.19	81.51	80.7592

shows that the proposed approach achieves good performance not only in predicting AHE, but also in solving a typical multi-class time series predication problem.

The methodology presented in this paper is a generic technology which has the potential to be applied in different time series problems. With the transforming the original series into the probabilistic distribution, the feature extraction is used to obtain the intrinsic pattern of time series in the aspect of global features and local varying features. The feature extraction method is very different from the traditional time series processing strategy. One of the future research could focus on feature extraction method. In this paper, only nine features are extracted from the PDPA in the global and integral time series, and ten features from the global and integral time series. However, more features can be obtained according to the requirements of the problem solving. It is hard to say the features extraction method presented in this paper is the preferable method for the CPPs and AHE. In many cases, the most difficult thing for the time series problem is the proper features extraction. Furthermore, it is worth to know the difference and effectiveness of the features presented in this paper. In this context, the evolutionary approach of genetic programming (GP)

Table 8
Performance comparison with different Pareto frontier feature sets which are obtained by the small data set but used for predication in the big data set.

ID	Feature set	K-Fold (%)	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	*4, 18*	71.6667	Training set	84	79	81.5
			Testing set	76.83	75.31	75.6545
2	*2, 4, 18*	72	Training set	89.33	83.00	86.1667
			Testing set	77.22	74.30	74.9564
3	*1, 4, 18*	72.5	Training set	82.67	81.67	82.1667
			Testing set	76.25	75.93	76.0035
4	*3, 4, 18, 19*	71.5	Training set	82.00	75.33	78.6667
			Testing set	76.83	75.37	75.6981
5	*3, 4, 14, 17*	72.1667	Training set	83.67	82.67	83.1667
			Testing set	75.29	78.24	77.5742
6	*1, 4, 18, 19*	72.1667	Training set	80.33	81.67	81
			Testing set	75.29	77.23	76.7888

Table 9
Sensitivity, specificity and accuracy obtained with the hybrid feature set which are obtained by the small data set but used for predication in the big data set.

Feature set	K-Fold (%)	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1, 2, 3, 4, 14, 17, 18, 19	72.8333	Training set	90.33	87	88.6667
		Testing set	72.39	79.37	77.7923

Table 10
Performance comparison with different Pareto frontier feature sets which are obtained by the small data set but used for predication in the big data set.

ID	Feature set	K-Fold (%)	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	*15*	71.7647	Training set	95.65	88	91.6667
			Testing set	83.33	52	62.1622
2	*15, 17*	71.7647	Training set	95.65	88	91.6667
			Testing set	100	64	75.6757
3	*9, 15*	68.2353	Training set	91.30	92	91.6667
			Testing set	66.67	64	64.8649
4	*1, 9, 15*	69.4118	Training set	95.65	88	91.6667
			Testing set	66.67	68	67.5676
5	*3, 9, 14, 15, 19*	72.9412	Training set	95.65	88	91.6667
			Testing set	75	64	67.5676

Table 11
Sensitivity, specificity and accuracy obtained with the hybrid feature set which are obtained by the big data set but used for predication in the small data set.

Feature set	K-Fold (%)	Data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
1, 3, 9, 14, 15, 17, 19	67.0588	Training set	95.65	88	91.6667
		Testing set	83.33	68	72.973

has proven to be a powerful approach for solving symbolic regression and feature extraction problems [34–37]. A holistic framework which combines the GP in the present methodology can be formulated and the difference in analysis can be obtained.

Acknowledgement

The authors would like to thank anonymous reviewers for their very detailed and helpful review. This study was funded by National Natural Science Foundation of China (61502291), the Cultivation Project for Outstanding Young Teachers in Higher Education Institutions of Guangdong Province (YQ2015070), and the Characteristic Innovation Project in Higher Education Institutions of Guangdong Province (2015GXJK037, 2015KTSCX039). The authors also wish to acknowledge that this research has been supported by Shantou University Scientific Research Foundation (Grant No. NTF 16002).

References

[1] G.B. Moody, L.H. Lehman, Predicting acute hypotensive episodes: the 10th annual PhysioNet/computers in cardiology challenge, Comput. Cardiol. 36 (2009) 541–544.
[2] D. Gang, Z. Shi-Sheng, L. Yang, Time series prediction using wavelet process neural network, J. Chin. Phys. B 17 (2008) 6.
[3] J.H. Henriques, T.R. Rocha, Prediction of acute hypotensive episodes using neural network multi-models, in: Computers in Cardiology, IEEE, 2009, pp. 549–552.
[4] Teresa Rocha, Simao Paredes, Prediction of acute hypotensive episodes by means neural network multi-models, Comput. Biol. Med. 41 (2011) 881–890.
[5] Yuan Zhou, Qiang Zhu, Huifang Huang, Prediction of acute hypotensive episode in ICU using Chebyshev neural network, J. Softw. 8 (8) (2013) 1923–1931.
[6] Marzyeh Ghassemi, Methods and Models for Acute Hypotensive Episode Prediction MSc Thesis, University of Oxford, 2011.
[7] M. Hoseinnia, D. Sadr, A Hybrid Approach for Predicting Acute Hypotensive Episodes[C]//5th Symposium on Advances in Science Technology, 2011.
[8] A. Arasteh, A. Janghorbani, M.H. Moradi, Application of empirical mode decomposition in prediction of acute hypotension episodes[C]//Biomedical Engineering (ICBME), in: 2010 17th Iranian Conference of. IEEE, 2010, pp. 1–4.
[9] L.H. Lehman, M. Saeed, G.B. Moody, R.G. Mark, Similarity-based searching in multi-parameter time series databases, in: Computers in Cardiology, 2008, pp. 653–656. IEEE, 2008.
[10] Joon Lee and Roger, Roger G. Mark, An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care, Biomed. Eng. Online 9 (1) (2010) 62.
[11] Haojun Sun, Shukun Sun, Yunxia Wu, Meijuan Yan, Chengdian Zhang, A method for prediction of acute hypotensive episodes in icu via pso and k-means, in: Sixth International Symposium of Computational Intelligence and Design (ISCID), IEEE, 2013, pp.99–102.
[12] D. Jiang, L. Li, B. Hu, et al., An approach for prediction of acute hypotensive episodes via the Hilbert-Huang transform and multiple genetic programming classifier, Int. J. Distrib. Sens. Netw. 2015 (2015) 9.
[13] D. Jiang, B. Hu, Z. Wu, Predicting acute hypotensive episodes based on multi GP, in: International Symposium on Intelligence Computation and Applications, Springer Singapore, 2015, pp. 161-170.
[14] D. Jiang, B. Hu, Z. Wu, Prediction of acute hypotensive episodes using EMD, statistical method and multi GP, Soft. Comput. (2016) 1–10.
[15] D.C. Montgomery, Introduction to Statistical Quality Control, Wiley, New York, 1997.
[16] D.T. Pham, E. Oztemel, Control chart pattern recognition using neural networks, J. Syst. Eng. 2 (4) (1992) 256–262.
[17] Cheng-Xiang Yang, Yi-Fei Zhu, Using genetic algorithms for time series prediction, in: Proc. of 6th International Conference on Natural Computation - ICNC, Valencia, Spain, 2010, pp. 4405–4409.
[18] MIMICII, http://physionet.org/physiobank/database/mimicdb/.
[19] http://www.physionet.org/challenge/2009/.
[20] A.M. Kowalski, M.T. Martin, A. Plastino, G. Judge, On extracting probability distribution information from time series, Entropy 14 (2012) 1829–1841.
[21] E. Bautista-Thompsom, S. Santos De la Cruz, Shape similarity index for time series based on features of euclidean distances histograms, in: Proceedings of the 15th International Conference on Computing, 2006, pp. 60–64.
[22] T.C.T. Ho, X. Chen, Utilizing histogram to identify patients using pressors for acute hypotension, Comput. Cardiol. (September 2009) 787–800.
[23] Vaibhav Awandekar, A.N. Cheeran, Predicting acute hypotensive episode by Bhattacharyya distance, Int. J. Eng. Res. Appl. 3 (2) (2013) 370–372.
[24] Alexander Waldin, Learning Blood Pressure Behavior from Large Blood Pressure Waveform Repositories and Building Predictive Models, MIT Master Thesis, 2013.
[25] Mitchell Melanie, An Introduction to Genetic Algorithms, MIT press, 1999.
[26] Corinna Cortes, Vladimir Vapnik, Support-Vector Networks, Mach. Learn. 20 (1995) 273–297.
[27] Ron Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proc. Fourteenth Int. Joint Conference Artificial Intelligence 2 (12) (1995) 1137–1145.
[28] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Economet. 31 (1986) 307–327.
[29] K. Lavangnananda, P. Sawasdimongkol, Neural network classifier of time series: a case study of symbolic representation preprocessing for control chart patterns, in: Proc. of 8th International Conference on Natural Computation - ICNC, 2012, pp. 351–356.
[30] R.J. Alcock, Y. Manolopoulos, Time-series similarity queries employing a feature-based approach, in: Proc. of 7th Hellenic Conference on Informatics, 27–29 August, Ioannina, Greece, 1999, pp. 31–39.
[31] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization, in: Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001), 2001, pp. 95–100.
[32] K. Lavangnananda, A. Piyatumrong, Image processing approach to features extraction in classification of control chart patterns, in: Proc. of 2005 IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications (SMCia/05), 28–30 June, Espoo, Finland, 2005, pp. 85–90.
[33] K. Lavangnananda, C. Wongwattanakarn, Utilizing symbolic representation and evolutionary computation in classification of control chart patterns, in: Proc. of the 2007 IEEE Three-Rivers Workshop on Soft Computing in Industrial Applications (SMCia/07), 1–3 August, Passau, Germany, 2007, pp. 79–84.
[34] B.N. Panda, A. Garg, K. Shankhwar, Empirical investigation of environmental characteristic of 3-D additive manufacturing process based on slice thickness and part orientation, Measurement 86 (2016) 293–300.
[35] K. Abhishek, B.N. Panda, S. Datta, S.S. Mahapatra, Comparing predictability of genetic programming and ANFIS on drilling performance modeling for GFRP composites, Procedia Mater. Sci. 6 (2014) 544–550.
[36] B. Panda, A. Garg, Z. Jian, A. Heidarzadeh, L. Gao, Characterization of the tensile properties of friction stir welded aluminum alloy joints based on axial force, traverse speed, and rotational speed, Front. Mech. Eng. 11 (3) (2016) 289–298.
[37] A. Garg, B.N. Panda, D.Y. Zhao, K. Tai, Framework based on number of basis functions complexity measure in investigation of the power characteristics of direct methanol fuel cell, Chemomet. Intell. Lab. Syst. 155 (2016) 7–18.