ECG Pattern Classification Based on Generic Feature Extraction

HEE-SOO PARK, SOO-MIN WOO, YANG-SOO KIM, BUB-JOO KANG, SANG-WOO BAN*
Department of Information and Communication Engineering
Dongguk University
707 Seokjang-Dong, Gyeongju, Gyeongbuk 780-714
REPUBLIC OF KOREA
park_hs_2000@hanmail.net, smin2000@naver.com, togu@nate.com, bjkang@dongguk.ac.kr, swban@dongguk.ac.kr http://wwwk.dongguk.ac.kr/prof/20060234
*Corresponding author

Abstract: In this paper, we propose a new ECG pattern classification model based on a generic feature extraction method. The proposed classifier is applied for indicating supraventricular arrhythmia in order to verify the performance of the proposed approach. A generic approach based on a histogram of 1st derivative of signals is applied for feature extraction. Principal component analysis (PCA) is considered for both reducing dimension of features and extracting more plausible features from the extracted features. A simple $k$-means algorithm works for ECG signal classification in feature space for discriminating abnormal ECG beats caused by supraventricular arrhythmia from normal ECG ones.

Key-Words: ECG signal processing, Generic feature extraction, PCA, $K$-means algorithm, Supraventricular arrhythmia

1 Introduction
A ubiquitous health monitoring system is one of important issues in a modern society for caring old people with increasing populations as well as for providing a proper emergency treatment for saving a human life in dangerous situations. Among many different bio-signals, the ECG is the widely used signal in clinical practice [1], which is measuring the electrical activity of the heart. Most of the algorithms processing ECG signals are focused on indicating a specific feature for diagnosis, which works well in many applications [1-4]. However, in some cases, it is very hard to find an essential characteristic feature for indicating an occurrence of abnormal symptom. Therefore, it is needed to develop a generic feature extraction method for monitoring any emergency situation by detecting abnormal ECG signals having different characteristic features compared with ECG signals obtained during normal states.

In this paper, we proposed an autonomous ECG pattern classification model with a new feature extraction approach with generic characteristics. In order to extract generic features of ECG signals, the histogram of 1st derivative of input signals are considered in conjunction with a principal component analysis (PCA). Moreover, for classifying the extracted features into normal status and abnormal status, a simple $k$-means algorithm is applied. For verifying the performance of the proposed model, the proposed model was applied to discrimination of ECG signals caused by the supraventricular arrhythmia symptoms from normal ECG signals. The ECG signals used in the experiments are obtained from MIT-BIH database [5].

Supraventricular arrhythmia is a type of arrhythmias that cause the heart to pump blood less effectively [6]. Supraventricular arrhythmias occur in two upper chambers of the heart called the atrium [6]. Arrhythmias cause nearly 250,000 deaths each year [6]. As many as 2 million Americans are living with atrial fibrillation, a type of arrhythmia, which is a very common long term arrhythmia [6]. A normal heart beats between 60 and 100 times a minute. However, in atrial fibrillation, the atria (upper lobes of the heart) beat 400 to 600 times per minute [6]. Therefore, the statistical features of ECG signals caused by supraventricular arrhythmia, such as the histogram of 1st derivative signals, may show different characteristics compared with ECG signals caused by normal heart.

Section 2 describes the proposed ECG classification model including the proposed feature extraction algorithm in detail. The experimental results of the proposed model are shown in Section 3. Conclusion and discussion will follow in Section 4.
2 Proposed ECG classification model

Fig. 1 shows a procedure of the proposed ECG classification model in this paper, which mainly consists of two parts. One is a generic feature extraction part from the input signals, and the other is a classification part implemented by a \( k \)-means algorithm.

![Diagram](image)

**Fig. 1.** A process of the proposed ECG classification model.

### 2.1 Generic feature extraction

As shown in Fig. 1, 1\textsuperscript{st} derivative of input signals is applied to extract primitive features of the ECG, which is calculated by Eq. (1):

\[
dS(t) = S(t+\Delta t) - S(t) \tag{1}
\]

where \( S(t) \) and \( \Delta t \) are an amplitude of raw input ECG signal at time \( t \) and a sampling time, respectively. The 1\textsuperscript{st} derivative of input signals can generically represent simple patterns of time-series signals. Therefore it is natural that the 1\textsuperscript{st} derivative of input signals can be used for describing general features of target signals.

For generating more generic feature representation, a histogram approach is applied for 1\textsuperscript{st} derivative signals. For generalization, before obtaining the histogram, normalization of signal amplitudes is processed by Eq. (2).

\[
nS(t) = \frac{dS(t)}{\max \{dS(t) \mid \text{for } \forall t\}} \tag{2}
\]

Then quantization of the normalized 1\textsuperscript{st} derivative signals follows before obtaining the histogram of the normalized 1\textsuperscript{st} derivate signals in order to make every histogram features have the same dimension. The quantized signals of the normalized 1\textsuperscript{st} derivative signals are obtained by Eq. (3):

\[
qS(t) = \frac{\max nS - nS(t)}{\max nS - \min nS} \times 100 \mod n\_levels \tag{3}
\]

where \( \max nS \) and \( \min nS \) are a maximum value and a minimum value among the amplitudes of the total normalized signals, respectively. And \( n\_levels \) in Eq. (3) is the number of quantized levels and the ‘mod’ operator calculates the quotient after division, which is used as a quantized index. From the quantized signals, the histogram, \( hS(i) \), is obtained by Eq.(4).

\[
hS(i) = \# \text{ of } qS(t) \text{, where } qS(t) = i \tag{4}
\]

In general, the histogram can represent statistical characteristics of extracted features, which makes the extracted features have more robustness against noise or disturbance of input signals than directly using the raw input signals. Moreover, in order to extract more important features from the histogram features as well as reduce dimensionality of the features, some eigenvectors with large eigen-values obtained by PCA are applied to the obtained histogram features. The number of eigen-vectors selected for transformation of features is decided by Eq. (5), which is the ratio between summation of the eigenvalues corresponding to the selected eigen-vectors and summation of all the eigenvalues. PCA is a well-known process for extracting essential features in a point of considering 2\textsuperscript{nd} order statistics and reducing dimensionality of a high-dimensional feature space.

\[
\text{arg max}_i \left\{ \frac{\sum_{z=1}^{i} \lambda_i}{\sum_{z=1}^{N} \lambda_i} > \theta \right\}, \quad \lambda_1 > \lambda_2 > \cdots > \lambda_N \tag{5}
\]

where \( N \) is a total number of eigenvalues and \( \Theta \) is a predefined threshold.

### 2.2 Classification of ECG signals

We applied a simple k-means algorithm in order to discriminate two-class signals such as normal ECG signal and abnormal one. The obtained features by the proposed generic feature extraction method are used as input of k-means algorithm. As initial code vectors, we used mean vectors obtained from training data of two-class signals. Finally two code-book vectors are obtained by the k-means algorithm, which works for representative vectors in decision making of the class of the input ECG signals.
3 Experimental results

In order to verify the proposed generic feature extraction method, we apply the proposed model to indicate supraventricular arrhythmia symptoms using some chosen ECG recordings of supraventricular arrhythmias in the MIT-BIH arrhythmia database [5]. Normal ECG data are obtained from Physiobank database [7]. The sampling rate of all the ECG signals is 125Hz.

Fig. 2 and Fig. 3 show raw ECG signals and 1st derivatives of the raw ECG signals for normal & abnormal signals, respectively.

![Fig. 2. Raw ECG signals for normal states (upper two graphs) and supraventricular arrhythmia (lower two graphs).](image1)

![Fig. 3. 1st derivatives for normal heart ECG signals (upper two graphs) and supraventricular arrhythmia ECG signals (lower two graphs).](image2)

Fig. 4 and Fig. 5 show normalized signals of the 1st derivatives shown in Fig. 3 for normal status (upper two graphs) and supraventricular arrhythmia (lower two graphs).

![Fig. 4. Normalized signals of the 1st derivates shown in Fig.3 for normal status (upper two graphs) and supraventricular arrhythmia (lower two graphs).](image3)

![Fig. 5. Histograms of the normalized 1st derivatives for normal heart ECG signals (upper two graphs) and supraventricular arrhythmia ECG signals (lower two graphs).](image4)

The number of levels, \( n_{\text{levels}} \) in Eq. (3), is 35 in this experiment. Therefore each ECG signal is represented by a 35-dimensions feature vector. As shown in Figs. 2-5, the discriminate characteristics become clearer in the course of the feature extraction process.

![Fig. 6 shows a projection result of the the histogram features onto two principal components obtained from PCA. In order to generate the inputs of the \( k \)-means algorithm, the proposed model considered 6 principal components selected by Eq. (5). Therefore every ECG signal is finally expressed by a 6-dimension vector for classification by the \( k \)-means algorithm. 48 ECG signals are used for training the \( k \)-means algorithm. Table 1 shows the performance of](image5)
the proposed classification model. 42 ECG signals are used for testing the proposed classification model. The proposed model correctly classify with 100% accuracy for training data and also shows plausible performance for testing data with 88.1% correct classification rate.

![Figure 6](image)

**Fig. 6.** Projection the histogram features onto two principal components obtained from PCA.

Table 1. Performance of the proposed classification model for test data

<table>
<thead>
<tr>
<th>Normal ECG</th>
<th>Supraventricular Arrhythmia ECG</th>
</tr>
</thead>
<tbody>
<tr>
<td># of total signals</td>
<td>15</td>
</tr>
<tr>
<td># of correctly classified signals</td>
<td>13</td>
</tr>
<tr>
<td>Performance</td>
<td>86.7%</td>
</tr>
</tbody>
</table>

### 4 Conclusion

A generic feature extraction method is proposed in this paper for dealing with arbitrary signal representation. For the generic feature extraction, a histogram of 1st derivative signals and PCA is applied. The proposed generic feature extraction method in conjunction with a simple k-means algorithm shows a plausible performance for ECG pattern classification problems for indicating supraventricular arrhythmia. The performance will be enhanced by considering more training data and more complex features.

As a further work, we need to more extensively verify the performance of the proposed model using larger database. Moreover, we are considering comparison of the performance with other approaches for verification of the proposed model. In addition, we are considering high-order statistical features of input signals such as 2nd momentum and 3rd momentum in order to more plausibly represent detail complex features containing in input signals. In this paper, we considered very simple classification algorithm because our main focus is a generic feature extraction that can successfully represent general signals. However, in order to enhance the classification performance, we need consider other classification methods such as SVM (support vector machine), AAMLP (auto-associative multilayer perceptron), LDA, etc. In addition, instead of PCA, we are also considering a kernel PCA for reflecting non-linearity of distribution of feature clusters in feature spaces in the course of a generic feature extraction.

Finally, the proposed ECG pattern classification model is aimed to be applied for a ubiquitous health monitoring system working in a wireless communication environment.

### Acknowledgements

This research was financially supported by the Ministry of Education, Science Technology (MEST) and Korea Industrial Technology Foundation (KOTEF) through the Human Resource Training Project for Regional Innovation.

### References:


