# Classification of ECG signals based on local fractal feature

**Wenjing Jiang[1] · Jian Wang[1,2,3]**

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Accurate and automatic analysis of electrocardiogram (ECG) signals plays a key role in the diagnosis of cardiovascular disease. This paper aims to investigate the performance of the multifractal detrending fluctuation analysis (MF-DFA) method based on a sliding window in ECG signal classification. We apply ECG signals to the detrended fluctuation analysis (DFA) method to obtain a group of local Hurst exponents and compose the DFA series. Afterwards, we use the MF-DFA method to get 6 generalized Hurst exponents of atrial premature beats (APB) and ECG signals with normal sinus rhythm (NSR), respectively. The 6 generalized Hurst exponents compose a feature vector and it is used in support vector machine (SVM) to examine the accuracy of ECG signal classification. The calculated results show that the MF-DFA method combined with a sliding window (SWD-MF-DFA) method performs well. Compared with directly using the MF-DFA method, with the same parameters, the accuracy of SWD-MF-DFA is higher, so as is the sensitivity and the specificity. As well, the proposed model can better analyze the local detailed features of time series and allows the application of MF-DFA in different parts of the time series, enabling the detection and analysis of time series variability.

**Keywords** Hurst exponent · Sliding window · MF-DFA

## 1 Introduction

In recent years, heart disease has consistently ranked as one of the leading causes of death worldwide. The mortality rate attributed to heart disease causes remains the highest globally,

✉ Jian Wang
003328@nuist.edu.cn

Wenjing Jiang
20211215023@nuist.edu.cn

[1] School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] Center for Applied Mathematics of Jiangsu Province, Nanjing University of Information Science and Technology, Nanjing 210044, China

[3] Jiangsu International Joint Laboratory on System Modeling and Data Analysis, Nanjing University of Information Science and Technology, Nanjing 210044, China

with related deaths accounting for one-third of all global fatalities each year. Consequently, urgent attention is required to address the prevention and timely diagnosis of heart disease.

Electrocardiogram (ECG) is one of the earliest biological signals studied and applied in clinical medicine. Meanwhile, ECG data is typically large-scale time-series data that requires reliable data storage and management. Cloud computing [1] provides an efficient solution for this, including features such as data backup, recovery, security, and access control, which has been successfully applied in [2]. As well, ECG is a crucial tool in the clinical diagnosis of cardiovascular disease. Consequently, improving the accuracy of ECG classification has become an attractive research focus [3–6]. Einthoven [7] invented the ECG machine in 1903 and it has been applied in clinical practice since then, making a significant contribution to the diagnosis, treatment and prevention of cardiovascular disease and arrhythmia. The automatic classification of ECG signals [8–11] has gained interest among academics due to the advancement of machine learning. Venkatesan et al. [12] utilized KNN to classify ECG signals by preprocessing the signals based on LMS algorithm, and the classifier exhibited excellent performance.

Sliding window algorithm is considered in this paper. Sliding windows are widely used in arrays or strings and perform well in the classification of time series. In time series of social indicators, Filho [13] combined the Detrended Fluctuation Analysis (DFA) with a sliding window to analyze the crime indicator fluctuations, which dynamically analyzed the related indicators like managing, monitoring, and controlling the information capacity of crime-ralated data. In biological series, for the precise detection of foot-to-ground contact phases, Jeon et al. [14] explored a model based on a sliding window and considered overlapping labels whose validation accuracy reached approximately 85%. Due to the excellent performance of sliding windows in time series, sliding windows have been taken into analyzing ECG signals [15–17]. Therefore, research of sliding windows is of great social significance.

Multifractal Detrended Fluctuation Analysis (MF-DFA) model is selected to extract features in this paper. MF-DFA is developed on the basis of DFA [18, 19], a scalar index calculation method proposed by Peng et al. in 1994 for analyzing long-range correlations in time series using the DNA mechanism. DFA determines the statistical autocorrelation of a signal based on Hurst exponents. However, traditional DFA only scales second-order statistical moments and assumes process being normally distributed. As well, DFA can be used to evaluate human ECG signals using the developed ECG device [20]. Subsequently, MF-DFA was proposed by Kantelhardt [21], enabling the extraction of scaled exponential and multifractal spectra of non-stationary time series. This allows for the efficient analysis of multifractal features in time series. Subsequently, MF-DFA has been widely used in various time series in different fields, such as meteorology, and financial mathematics [22, 23]. Morever, MF-DFA is also extensively used in practical applications, especially in classification. MF-DFA was adopted to extract features and performed excellently in classification [24]. Jiao et al. [25] used MF-DFA to identify chaotic features in three signal sets originating from states of Myocardial Infarction (MI), Arrhythmia, and a healthy state. In [26, 27], MF-DFA has satisfactory performance in classifying atrial premature beat people and normal sinus rhythm subjects. In addition to the overall study classification of ECG signals, a nearest neighbor classifier based on the local fractal dimension has been processed, demonstrating excellent adaptability [28].

In order to obtain more detailed information of ECG signals, the sliding window algorithm is adopted in this paper. Meanwhile, considering the multifractal nature of ECG signals, a classifier consisting of sliding windows and the MF-DFA method has been taken in to account.

The remainder of the paper is structured as follows. We introduce the methodology in Section 2. Section 3 intrduces some information of ECG signals. Section 4 exhibits the

experimental results. Section 5 presents the comparisonand discussion with some existing methods. Section 6 reports the conclusion.

## 2 Methodology

### 2.1 Sliding window algorithm

The sliding window algorithm is used to solve the sub-element problem of an array or alphabetic string. It can transform the nested loop problem into a single loop problem so that the time complexity can be decreased. The specific operation is that using fixed-size sublist and iterating over the whole list with the sublist. It is commonly used in processing sequences [29].

The specific steps are given as follows:

Step1: Fix the size of the sliding window;

Step2: Fix $q$;

Step3: Move the sliding window with an increment of 15 until the right of the series, and then calculate the local Hurst exponent $H$ in each sliding window respectively.

Step4: Compose a new sequence by using the calculated $H$ and then apply the new time series to the MF-DFA model to extract 6 generalized Hurst exponents.

Step5: Calculate the accuracy, sensitivity and specificity of the classification using support vector machine (SVM) based on 6 extracted generalized feature values.

Step6: Alter $q$, repeat the previous operations and compare the performances of the classification.

### 2.2 MF-DFA

Consider a time series $\{T\}_a$, $a = 1, 2, \ldots, W$. Construct a cumulative sequence as follows:

$$t(i) = \sum_{i=1}^{j}(T_i - \overline{T}), \ j = 1, ..., N, \tag{1}$$

where $\overline{T}$ is the mean of $\{T\}_a^W$.

Divide the new time series $t(i)$ into $W_s$ segments and $W_s = \lfloor \frac{W}{s} \rfloor$ where $s$ is the size of non-overlapping window and $W$ is the length of the series.

Here, in order to save more information, the segmentation operation is then performed again from the end to the begining. Therefore, $2W_s$ segments can be obtained. Then, for the $s$ points in each segment $\mathcal{I}$, $\mathcal{I} = 1, 2, \ldots, 2W_s$, applying least square method to polynomial fitting, we can obtain:

$$y_{\mathcal{I}}(\alpha) = b_1\alpha^k + b_2\alpha^{k-1} + ... + b_k\alpha + b_{k+1}, \quad \alpha = 1, 2, ..., s. \tag{2}$$

Calculate the mean squared error and then calculate the $q-$order wave function on this basis. When $\mathcal{I} = 1, 2, \ldots, W_s$:

$$f^2 = \frac{1}{s}\sum_{\alpha=1}^{s}\{t[(\mathcal{I}-1)s + \alpha] - y_{\mathcal{I}}\}. \tag{3}$$

When $\mathcal{I} = W_s, W_{s+2}, \ldots, 2W_s$:

$$f^2 = \frac{1}{s} \sum_{\alpha=1}^{s} \{t[(W - (\mathcal{I} - W_s)s + \alpha] - y_{\mathcal{I}}\}. \tag{4}$$

$$F_q(s) = \begin{cases} \left\{ \frac{1}{W_s} \sum_{w=1}^{W_s} [f^2]^q \right\}^{\frac{1}{q}}, & \text{if } q \neq 0, \\ \exp\left( \frac{1}{W_s} \sum_{w=1}^{W_s} \ln[f^2] \right), & \text{if } q = 0. \end{cases} \tag{5}$$

With the increase of $s$, $F_q(s)$ will increase in a power-law relationship, which can be interpreted as $F_q(s) \propto s^{h(q)}$. $h(q)$ is called generalized Hurst exponent. Also, when $q = 2$, we call it standard DFA model. When $h(q)$ varies with $q$, the time series will exhibit multifractal properties. Otherwise, when the original sequence is a single fractal, $h(q)$ is a constant independent of $q$.

## 2.3 Support vector machine

Support vector machine (SVM) is a commonly used classifier [30], which is widely used in binary classification problems. SVM plays a major role not only in linear classification but also nonlinear classification. When it comes to nonlinear classification, a suitable Kernel function is selected to transform nonlinear classification problem into linear classification problem, allowing us to avoid computing dot product in high-dimensional feature space. Therefore, the problem needs to be solved is given as follows.

$$\min_{\lambda} \left[ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j (\phi(x_i) \cdot \phi(x_j)) - \sum_{i=1}^{n} \lambda_i \right], \tag{6}$$

$$s.t. \sum_{i=1}^{n} \lambda_i y_i = 0, \quad 0 \leq \lambda_i \leq C, \quad i =, 1, 2, \ldots, N, \tag{7}$$

where $x$ represents the original sample and $\phi(x)$ represents the new feature vector. $y_i$ and $y_j$ are the classification labels, where $y_i, y_j \in \{+1, -1\}$ and $C$ is a constant greater than 0, representing the penalty function. Meanwhile, from (7), the best solution can be obtained $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*)^T$.

Next, calculate the following formulas:

$$w^* = \sum_{i=1}^{N} \lambda_i^* y_i (x_i \cdot x_j),$$

$$b^* = y_j - \sum_{i=1}^{N} \lambda_i^* y_i (x_i \cdot x_j). \tag{8}$$

Here, choose a component $\lambda_j^*$ of $\lambda^*$, which should satisfy $0 < \lambda_j^* < C$.

Finally, separate hyperplane.

$$w^* \cdot x + b^* = 0. \tag{9}$$

The split hyperplane can be expressed as $f(x) = sign(w^*x + b^*)$ or $f(x) = sign(\sum_{i=1}^{N} \lambda_i^* y_i (x_i \cdot x_j) + b^*)$. The inner product can be replaced by kernel function, which can obtain the nonlinear SVM. The commonly used kernel functions are shown in Table 1.

**Table 1** Kernel functions

| Name | Expression |
|---|---|
| Linear kernal function | $k(x-i, x_j) = x_i^T x_j$ |
| Polynomial kernel function | $k(x_i, x_j) = (x_i^T x_j)^d$ |
| Gaussian kernel function | $k(x_i, x_j) = exp(-\frac{||x_i - x_j||^2}{2\sigma^2})$ |

# 3 Data collection

All ECG signals were obtained from the MIT-BIH Arrhythmia Database. All ECG signals were recorded with a sampling frequency of 360 HZ and a sampling frequency of 200 [adu/mV]. We collect ECG signals from 50 subjects with atrial premature beat (APB) and 50 people with normal sinus rhythm (NSR). Each piece of ECG signals consists of 3600 points.

We use the sliding window with a fixed size of 100 to acquire a local Hurst exponent. Then, move the sliding window to the end in a step size of 15 each time. Therefore, we can obtain 234 local Hurst exponents from each time series. Subsequently, we treated these 234 local Hurst exponents as a new time series and utilize the MF-DFA method to require a set of generalized Hurst exponents which is then used in the classification by SVM.

# 4 Computational results

First of all, we fix the size of the sliding window to 100 and move it at an interval of 15 each time. Next, we fix $q$ and calculate the local Hurst exponent $H$ of ECG signals in each sliding
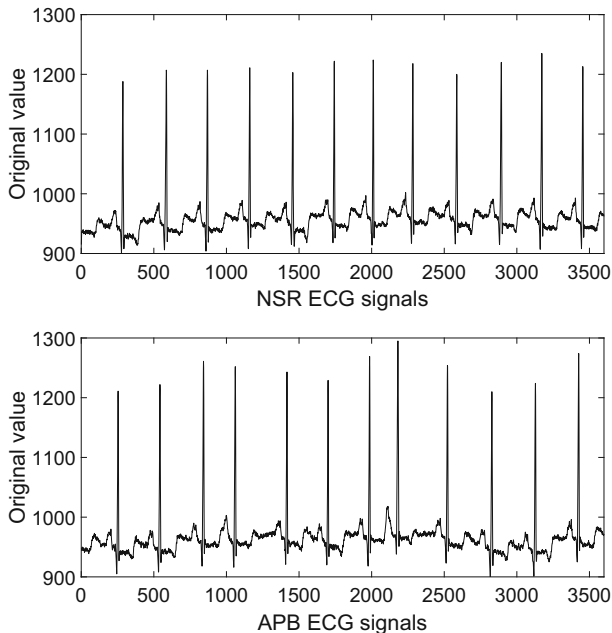


**Fig. 1** The original time series of NSR ECG signals and APB ECG signals
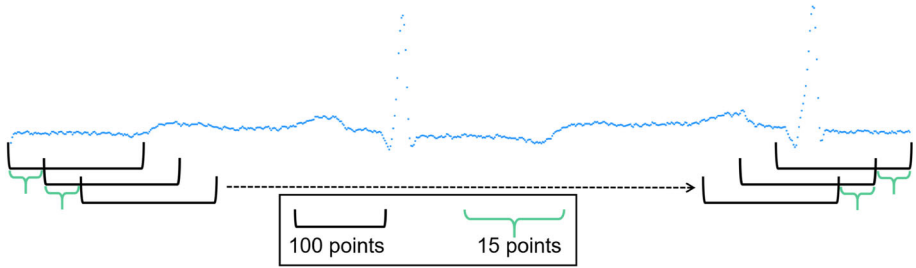
**Fig. 2** Schematic diagram of the process of constructing a new time series. 100 points represent the size of the sliding window and we calculate the Hurst exponent of the 100 points. The sliding window moves to the right with an interval of 15 points

window. All calculations were conducted using MATLAB R2020a on a 2.50GHz Intel(R) Core(TM) i5-7200U CPU.

An original ECG sequence is made up of 3600 points. Figure 1 exhibits the initial time series. First, we consider the case when $q$ equals 0. Therefore, we can obtain a new series (for convenience, we call it DFA series). The process of extracting local Hurst exponents from the original ECG signal sequence by using the sliding window can be referred to Fig. 2. Figure 3 shows the new time series consisting of local Hurst exponents.

DFA method is used in the first step. It is a common method to handle non-stationary signals like meteorological signals [31]. DFA can quantitatively analyze the correlation nature of signal sequences and determine the different properties of the signals by calculating the
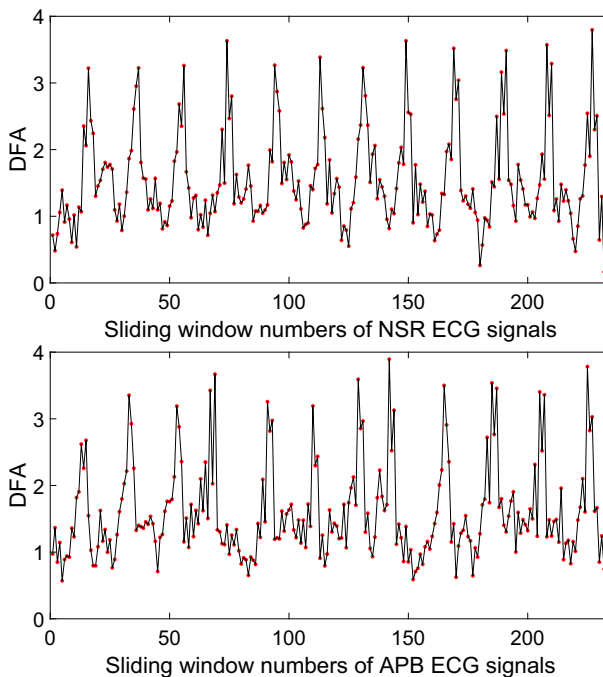


**Fig. 3** A new times series composed by local Hurst exponents when $q$ equals 0

correlation index. As well, it is capable of effectively filtering out trend components of various orders in the sequence and detecting long-range correlations containing noise, which is suitable for the analysis of non-stationary time series with long-range power-law correlation.

The following step involves using the MF-DFA method to extract 6 feature values, referred to as Sliding-Window-DFA-MF-DFA model (SWD-MF-DFA). SVM is utilized to examine the accuracy based on the extracted generalized Hurst exponents and Gaussian kernel function is adopted. When constructing the DFA sequence, we adopt different values of $q$ and calculate the corresponding accuracies. Since the common range of $q$ used in MF-DFA is (-10, 10), so the $q$ adopted to obtain local Hurst exponents are integers within this range, totaling 21 samples. Subsequently, extract 6 generalized Hurst exponents and use them as a feature vector. Then, apply the feature vector to SVM to calculate the accuracy, as well as the sensitivity and the specificity. Since SVM is supervised learning, we label the feature vectors of NSR ECG signal with the value 1 as the class 1 and label the feature vectors of NSR ECG signals with the value -1 as the class 2. Due to the small number of samples, we use leave-one-out cross-validation method. In this paper, we randomly disrupt the samples 100 times. Each time, the samples are divided into 10 disjoint subsets, with one serving as the test set and the rest as the training set. Therefore, the samples can be fully utilized so that the results are the closest to the expectations of training on the entire test set. Table 2 displays the classification accuracy, sensitivity, and specificity. We observe that when $q$ deviates from 0, the accuracy shows a downward trend generally. However, all the accuracies are higher than 60%.

Here, $q$ represents the order of the fluctuation function in DFA series and $Q$ represents the range of $q$ used in the MF-DFA model. *Acc*, *Sen*, and *Spe* represent accuracy, sensitivity

**Table 2** SWD-MF-DFA model based on different $q$

| q | Q | ACC | SEN | SPE |
|---|---|---|---|---|
| −10 | (−10, 10) | 64.30% | 74.18% | 73.88% |
| −9 | (−10, 10) | 64.50% | 73.68% | 74.07% |
| −8 | (−10, 10) | 64.34% | 73.05% | 74.15% |
| −7 | (−10, 10) | 65.43% | 72.55% | 75.92% |
| −6 | (−10, 10) | 64.99% | 71.31% | 49.05% |
| −5 | (−10, 10) | 64.84% | 70.78% | 77.00% |
| −4 | (−10, 10) | 64.84% | 70.06% | 78.91% |
| −3 | (−10, 10) | 64.20% | 71.75% | 76.04% |
| −2 | (−10, 10) | 57.16% | 68.55% | 71.55% |
| −1 | (−10, 10) | 79.04% | 77.28% | 88.97% |
| 0 | (−10, 10) | 88.00% | 79.02% | 100.00% |
| 1 | (−10, 10) | 85.86% | 78.13% | 100% |
| 2 | (−10, 10) | 83.24% | 79.40% | 93.31% |
| 3 | (−10, 10) | 77.96% | 78.54% | 86.04% |
| 4 | (−10, 10) | 77.10% | 76.10% | 86.79% |
| 5 | (−10, 10) | 65.35% | 72.83% | 75.49% |
| 6 | (−10, 10) | 65.30% | 73.17% | 75.62% |
| 7 | (−10, 10) | 66.98% | 74.89% | 75.92% |
| 8 | (−10, 10) | 67.33% | 74.47% | 76.24% |
| 9 | (−10, 10) | 68.33% | 77.72% | 77.72% |
| 10 | (−10, 10) | 67.87% | 73.83% | 77.96% |

and specificity, respectively. The specific calculations are given as follows:

$$
\begin{aligned}
Acc &= \frac{TP + TN}{TP + FN + FP + TN}, \\
Sen &= \frac{TP}{TP + FP}, \\
Spe &= \frac{TP}{TP + FN},
\end{aligned}
\tag{10}
$$

where the meanings of $TP$, $TN$, $FP$ and $FN$ are True positive, True negative, False positive, and False negative, respectively.

Figure 4 shows the distribution of the features. The blue line represents the generalized Hurst exponent when $q$ equals 0. We observe that the points above the blue line represent generalized Hurst exponent when $Q$ belongs to (-10, 0), are quite scattered. Conversely, below the blue line, when $Q$ is (0, 10), the distribution of the points is clustered. When the fluctuation of the features is larger, the self-similarity is sacrificed more. In other words, the similarity of NSR ECG signals and APB ECG signals increases, indicating that it is challenging to distinguish NSR and APB ECG signals, leading to a decrease in classification accuracy.

It is difficult to determine the distinctiveness of the features from Fig. 4. As well, we observe the fluctuation in APB ECG signals is evident. When $Q$ is in the range of (-10,0), the dissimilarity before the APB and NSR ECG signals appears to be smaller compared to when
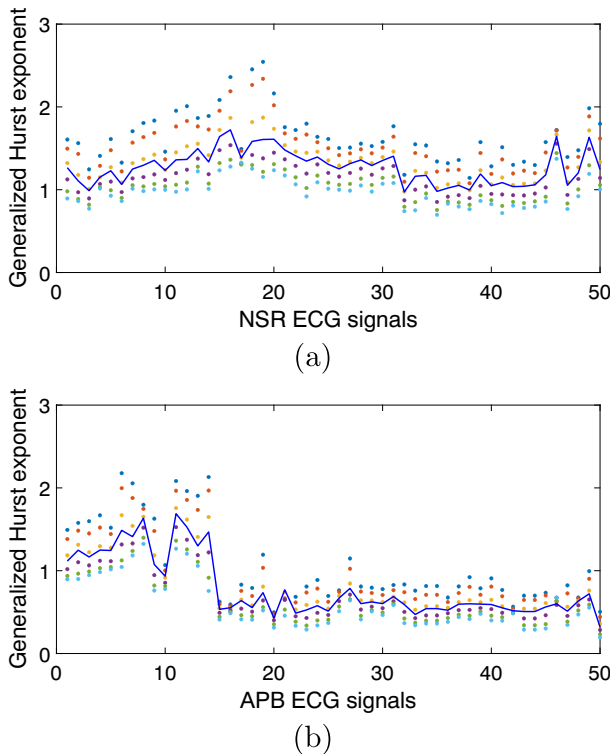


**Fig. 4** Six generalized Hurst exponents of (a) NSR ECG signals and (b) APB ECG signals

$Q$ is in the range (0, 10). Therefore, we divide the extracted feature values in Fig. 4 into 2 parts equally, delimited by the blue line, where the blue line is the Hurst value when $Q$ equals 0. Thus, three feature values are extracted when $Q$ is (-10, 0) and the other feature values are extracted when $Q$ belongs to (0, 10). We use 3D coordinates to present their distribution. Figure 5 exhibits a better fluctuation phenomenon. The red circles represent the features extracted from APB ECG signals and the black circles are the features extracted from NSR ECG signals. The one the left means the generalized Hurst exponents when $Q < 0$, and the one on the right means $Q > 0$. Comparing with Fig. 5(a) and (b), the intersecting points in Fig. 5(a) are more than those in Fig. 5(b). When the degree of mixing is higher, the similarity of two signals is higher, indicating that it is hard to distinguish the signals, resulting in low classification accuracy. It is evident that the degree of mixing is higher when $Q$ is (-10, 0). Few points are cross-distributed. Therefore, we select the range (0, 10) as the experimental interval.

In order to better discriminate this pattern, for the previous 21 samples, we fix $Q$ to (-10, 0) and extract 6 generalized Hurst exponents. Afterwards, the extracted 6 feature values form a feature vector, which is then applied it to SVM to calculate the classification accuracy of SWD-MF-DFA model. Subsequently, change $Q$ to (0, 10) and repeat the previous operations. All the calculated results are shown in Table 3. Under the same q, the bolded data represents the higher accuracy obtained by reextracting the features from the newly obtained sequence and classifying it. It is obvious to see from Table 3 that the accuracy when $Q$ is (0, 10) is higher than that when $Q$ is (-10, 0) in most samples. It is clear from Fig. 6 that the extracted feature values of NSR ECG signals in the interval (0, 10) have fewer intersections with those of APB ECG signals. That means (0, 10) is the better interval that should be selected.

Figure 6 clearly illustrates the similarity of the extracted feature values. When $Q$ belongs to (-10, 10) and (-10, 0), various feature values of APB ECG signals are intertwined with those of NSR ECG signals. However, in Fig. 6(c), the eigenvalues are basically separate, suggesting that the accuracy of the classification will increase. In the two exceptions noted in Table 3, we find they still satisfy the pattern. Observing Fig. 7, the Hurst exponents in Fig. 7(a) and (b) have more intersections than those in Fig. 7(c). In some parts, a higher accuracy is achieved when $Q$ is in (-10, 0) because the intersectionality of feature values in (-10, 0) is lower than that of (0, 10). Therefore, features in (-10, 0) are more differentiated.
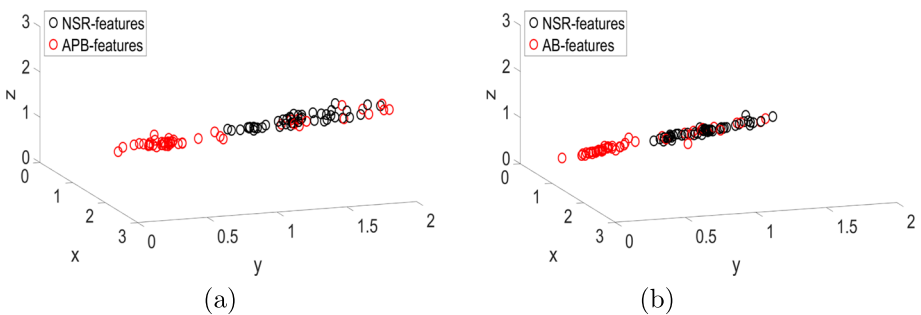


**Fig. 5** (a): $x$ represents the first generalized Hurst exponent; $y$ represents the second generalized Hurst exponent; $z$ represents the third generalized Hurst exponent; the three generalized Hurst exponents are extracted when $Q$ is (-10, 0), and (b): $x$ represents the forth generalized Hurst exponent; $y$ represents the fifth generalized Hurst exponent; $z$ represents the sixth generalized Hurst exponent; the three generalized Hurst exponents are extracted when $Q$ is (0, 10)

**Table 3** SWD-MF-DFA model based on different $q$

| $q$ | Q | Accuracy | Sensitivity | Specificity | Q | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|---|
| −10 | (−10, 0) | 51.46% | 68.73% | 66.19% | (0, 10) | **68.21%** | 71.59% | 80.75% |
| −9 | (−10, 0) | 50.32% | 67.64% | 66.32% | (0, 10) | **68.35%** | 72.15% | 82.02% |
| −8 | (−10, 0) | 46.36% | 66.58% | 64.08% | (0, 10) | **67.46%** | 69.68% | 81.27% |
| −7 | (−10, 0) | 45.86% | 65.07% | 65.20% | (0, 10) | **66.53%** | 69.82% | 80.80% |
| −6 | (−10, 0) | 65.92% | 65.53% | 66.40% | (0, 10) | **64.60%** | 70.85% | 77.70% |
| −5 | (−10, 0) | 54.13% | 70.00% | 67.23% | (0, 10) | **64.04%** | 71.69% | 75.04% |
| −4 | (−10, 0) | 51.67% | 69.06% | 66.09% | (0, 10) | **67.16%** | 76.12% | 75.08% |
| −3 | (−10, 0) | 55.87% | 69.80% | 68.99% | (0, 10) | **63.14%** | 70.37% | 76.02% |
| −2 | (−10, 0) | 60.69% | 71.46% | 71.70% | (0, 10) | **68.49%** | 71.76% | 80.09% |
| −1 | (−10, 0) | 77.06% | 75.92% | 87.84% | (0, 10) | **81.65%** | 77.49% | 92.87% |
| 0 | (−10, 0) | 87.00% | 79.63% | 100.00% | (0, 10) | **90.00%** | 81.93% | 98.10% |
| 1 | (−10, 0) | 83.34% | 77.01% | 97.23% | (0, 10) | **85.97%** | 78.13% | 100.00% |
| 2 | (−10, 0) | 82.85% | 82.71% | 88.26% | (0, 10) | **83.84%** | 79.01% | 95.09% |
| 3 | (−10, 0) | 74.53% | 77.91% | 81.37% | (0, 10) | **79.94%** | 80.30% | 86.55% |
| 4 | (−10, 0) | 69.43% | 73.90% | 79.06% | (0, 10) | **77.30%** | 83.12% | 79.63% |
| 5 | (−10, 0) | **68.45%** | 73.45% | 78.43% | (0, 10) | 68.20% | 75.25% | 76.77% |
| 6 | (−10, 0) | **68.66%** | 73.57% | 79.44% | (0, 10) | 67.28% | 74.48% | 76.88% |
| 7 | (−10, 0) | 69.80% | 71.82% | 82.30% | (0, 10) | **70.00%** | 75.44% | 78.31% |
| 8 | (−10, 0) | 70.25% | 72.10% | 82.76% | (0, 10) | **70.30%** | 74.84% | 78.99% |
| 9 | (−10, 0) | 67.68% | 72.30% | 79.60% | (0, 10) | **68.86%** | 74.19% | 78.70% |
| 10 | (−10, 0) | 68.32% | 72.61% | 79.91% | (0, 10) | **68.77%** | 74.42% | 78.43% |
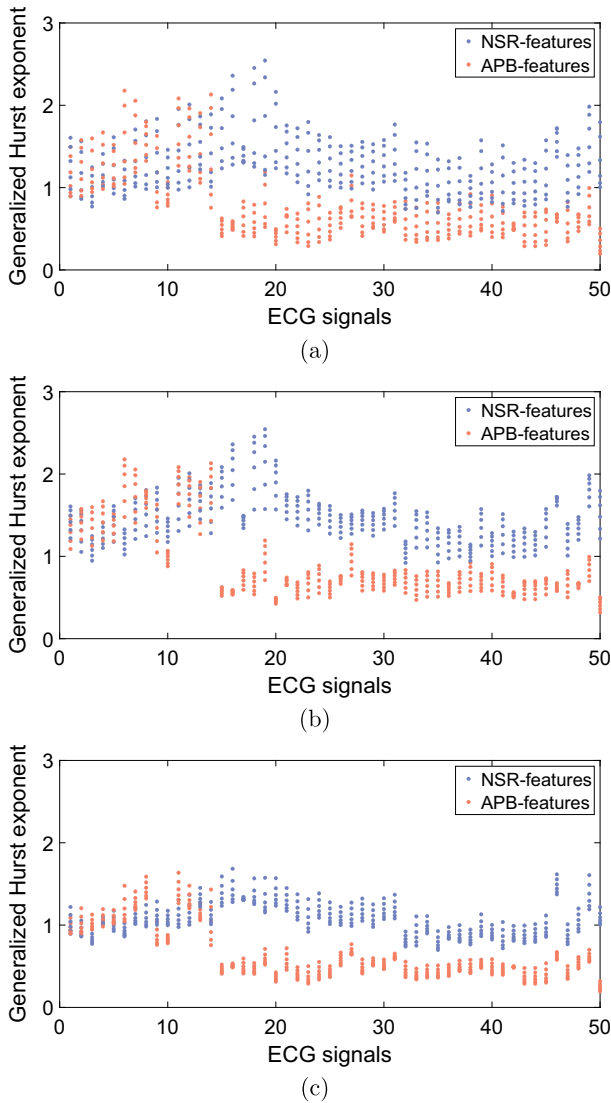
**Fig. 6** Based on MF-DFA, the distribution of feature values of the DFA series constructed when $q$ equal 0. (a): Six features are extracted when $Q$ is (-10, 10), (b): Six features are extracted when $Q$ is (-10, 0), and (c): Six features are extracted when $Q$ is (0, 10)

Then, we use MF-DFA to obtain six feature values without processing the original series. The results are listed in Table 4:

Comparing with the traditional MF-DFA method, we find that the accuracy of SWD-MF-DFA is higher. Table 5 shows when $q$ equals 0 during constructing DFA series and the range of $Q$ in MF-DFA is (0, 10), the accuracy of classification is the best. At last, we pick it as the parameter in our method. 90% of the samples are used as the training set, and the remaining
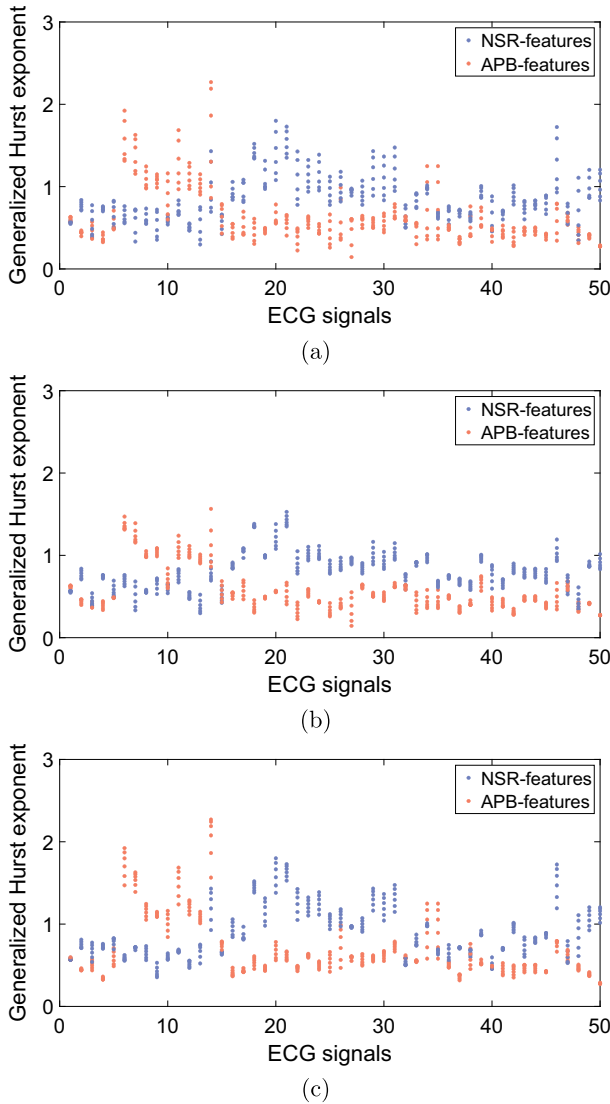
**Fig. 7** Based on MF-DFA, the distribution of feature values of the DFA series constructed when $q$ equals 6. (a): Six features are extracted when $Q$ is (-10, 10), (b): Six features are extracted when $Q$ is (0, 10), and (c): Six features are extracted when $Q$ is (-10, 0)

| Table 4 The classification accuracy, sensitivity and specificity of the MF-DFA method without prior treatment | Range of $q$ | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| | $(-10, 10)$ | 85.00% | 87.72% | 86.17% |
| | $(0, 10)$ | 77.03% | 80.98% | 84.36% |
| | $(-10, 0)$ | 84.96% | 85.97% | 85.65% |

| | $Q$ | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **Table 5** The accuracy, sensitivity and specificity based on SWD-MF-DFA model and DFA sequence constructed when $q$ is 0 | (0, 10) | 90.00% | 81.93% | 98.10% |

10% are used as the test set. We use SVM to perform 100 cross-validations and pick up the mean value. At the same time, we calculate the sensitivity and specificity in Fig. 8.

## 5 Discussion

From Tables 4 and 5, it is evident that our proposed method has a higher accuracy, indicating that our approach is more effective at capturing the key information in the input data compared to the sole use of the MF-DFA method. Meanwhile, we compare the performance of the proposed method with existing approaches. Here, the experimental ECG data comes from
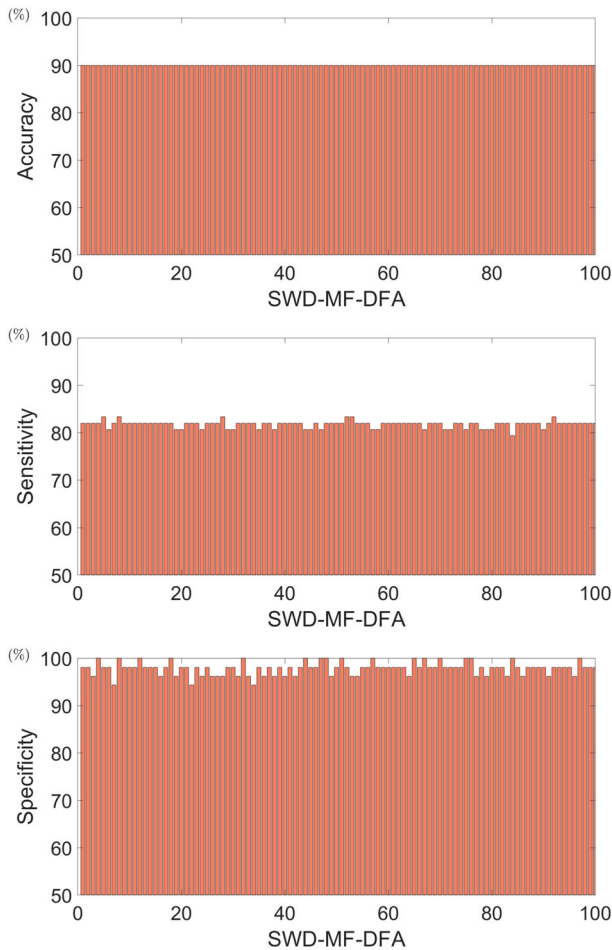


**Fig. 8** The accuracy, the sensitivity, and the specificity of SWD-MF-DFA model. The accuracy is the mean value after 100 calculations, so as the sensitivity and the specificity

**Table 6** Comparison with other methods about ECG signals in heart disease

| Model | Method | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| Li et al. [32] | Integrate convolutional neural network and entropy | 81.85% | 84.87% | 65.60% |
| Liang et al. [33] | Integrate convolutional neural network and long short-term memory methods | 85.00% | 99.00% | 84.00% |
| Kuila et al. [34] | Combine MF-DFA with Kolmogorov complexity, Shannon entropy and Higuchi'?s Hurst exponent | 95.29% | 96.28% | 94.55% |
| Romdhane and Pr [35] | Combine convolutional nerual work with segmentation algorithm | 98.41% | 98.37% | 98.41% |
| Our method | Combine sliding window and MF-DFA | 90.00% | 81.93% | 98.10% |

MIT-BIH Arrhythmia Database. Observing Table 6, although the classification accuracy is not the highest, the accuracy of our proposed method is acceptable. As well, the sensitivity needs to be further improved. However, it is gratifying that the specificity is satisfactory.

In fact, sliding window has been widely in various models [36–38]. Meanwhile, the proposed method can effectively extract local features, so that the the local structures and patterns in the data can be focused on. This aids in capturing the local information, thereby increasing sensitivity to local variations. As well, the extraction of local features can decrease the dimension of the data, indicating that the original data can be mapped to a more compact representation, helping to reduce computational burden and prevent overfitting.

## 6 Conclusions

In this paper, considering MF-DFA performs well in ECG classification, and in order to obtain more detailed local information of ECG signals, we introduced a sliding window algorithm to the MF-DFA model. We took 100 points as a group, then used the MF-DFA model to extract the Hurst exponent of this set of points. Subsequently, we took another 100 points to the right at an interval of 15, and calculated the Hurst exponent until the last point was reached. Therefore, a total of 234 Hurst exponents can be obtained, which were used to form a new series. Next, we extracted 6 generalized Hurst exponents using MF-DFA of the new series, which were regarded as the input vector of SVM. Finally, we used SVM to examine the performance of the proposed method.

Here, we employed the influence of $q$ when forming the new sequence. Generally, when forming the new sequence, we conducted 21 experiments, iterating over q with an interval of 1, ranging from -10 to 10. The results showed that when $q = 0$, the classification accuracy of the new formed series was the highest. As well, we observed the difference of the fluctuation when extracting 6 generalized Hurst exponents. In most conditions, when $Q$ was in (0, 10), the accuracy was higher than that when $Q$ was (-10, 0). As well, from the spatial distribution of feature values, when $Q$ was in the range of (0, 10), the intersection between NSR ECG signals and APB ECG signals was smaller, indicating that the similarity was low so that the accuracy was higher. Finally, the SWD-MF-DFA model obtained a classification accuracy, sensitivity, and specificity of 90.00%, 81.93%, and 98.10%, respectively. Meanwhile, we compared our method to only using MF-DFA, our method performed better. As well, we compared the proposed method with other methods, although there was still room for improvement, the results were satisfactory. In the future, we will modify this model and render it suitable for 2D image classification.

**Data Availibility Statement**  The datasets of ECG signals analysed during the current study are available in the MIT-BIH Arrhythmia Database and the data can be acquired from the corresponding author upon reasonable request.

## Declarations

**Conflicts of interest**  The authors declare that there is no conflict of interests regarding the publication of this article.

# References

1. Ahmad S, Mehfuz S, Mebarek-Oudina F, Beg J (2022) RSM analysis based cloud access security broker: a systematic literature review. Cluster Comput 25(5):3733–3763
2. Nyo MT, Mebarek-Oudina F, Hlaing SS, Khan NA (2022) Otsu?s thresholding technique for MRI image brain tumor segmentation. Multimedia Tools Appl 81(30):43837–43849
3. Dong X, Wang C, Si W (2017) ECG beat classification via deterministic learning. Neurocomputing 240:1–12
4. Dalal S, Vishwakarma VP (2021) Classification of ECG signals using multi-cumulants based evolutionary hybrid classifier. Sci Rep-UK 11(1):15092
5. Sangaiah AK, Arumugam M, Bian GB (2020) An intelligent learning approach for improving ECG signal classification and arrhythmia analysis. Artif Intell Med 103:101788
6. Wang S, Li R, Wang X, Shen S, Zhou B, Wang Z (2021) Multiscale residual network based on channel spatial attention mechanism for multilabel ECG classification. J Healthc Eng 2021:6630643
7. Einthoven W (1902) Galvanometrische registratie van het menschelijk electrocardiogram. Herinnerings-bundel Professor S S Rosenstein 101–107
8. Vafaie MH, Ataei M, Koofigar HR (2014) Heart diseases prediction based on ECG signals? classification using a genetic-fuzzy system and dynamical model of ECG signals. Biomed Signal Proces 14:291–296
9. Ji Y, Zhang S, Xiao W (2019) Electrocardiogram classification based on faster regions with convolutional neural network. Sensors 19(11):2558
10. He R, Liu Y, Wang K, Zhao N, Yuan Y, Li Q, Zhang H (2019) Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM. IEEE Access 7:102119–102135
11. Diker A, Sönmez Y, Özyurt F, Avcı E, Avcı D (2021) Examination of the ECG signal classification technique DEA-ELM using deep convolutional neural network features. Multimed Tools Appl 80:24777–24800
12. Venkatesan C, Karthigaikumar P, Varatharajan R (2018) A novel LMS algorithm for ECG signal prepro-cessing and KNN classifier based abnormality detection. Multimed Tools Appl 77(8):10365–10374
13. da Silva Filho AM, Zebende GF, Guedes EF (2021) Analysis of intentional lethal violent crimes: a sliding windows approach. Physica A 567:125653
14. Jeon H, Kim SL, Kim S, Lee D (2020) Fast wearable sensor?based foot?ground contact phase classification using a convolutional neural network with sliding-window label overlapping. Sensors 20(17):4996
15. Papaloukas C, Fotiadis DI, Liavas AP, Likas A, Michalis LK (2001) A knowledge-based technique for automated detection of ischaemic episodes in long duration electrocardiograms. Med Biol Eng Comput 39(1):105–112
16. Tsipouras MG, Fotiadis DI, Sideris D (2005) An arrhythmia classification system based on the RR-interval signal. Artif Intell Med 33(3):237–250
17. Yang R, Zha X, Liu K, Xu S (2021) A CNN model embedded with local feature knowledge and its application to time-varying signal classification. Neural Networks 142:564–572
18. Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49(2):1685–1689
19. Peng CK, Havlin S, Stanley HE, Goldberger AL (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5(1):82–87
20. Pranata AA, Adhane GW, Kim DS (2017, January) Detrended fluctuation analysis on ECG device for home environment. In: 2017 14th IEEE annual consumer communications & networking conference (CCNC) IEEE, pp 126–130
21. Kantelhardt JW, Zschiegner SA, Koscielny-Bunde E, Havlin S, Bunde A, Stanley HE (2002) Multifractal detrended fluctuation analysis of nonstationary time series. Physica A 316(1–4):87–114
22. Méndez-Gordillo AR, Cadenas E (2021) Wind speed forecasting by the extraction of the multifractal patterns of time series through the multiplicative cascade technique. Chaos Soliton Fract 143:110592
23. Li S, Xu N, Hui X (2020) International investors and the multifractality property: evidence from accessible and inaccessible market. Physica A 559:125029
24. Zhang Z, Wen T, Huang W, Wang M, Li C (2017) Automatic epileptic seizure detection in EEGs using MF-DFA, SVM based on cloud computing. J X-ray Sci Technol 25(2):261–272
25. Jiao D, Wang Z, Li J, Feng F, Hou F (2020) The chaotic characteristics detection based on multifractal detrended fluctuation analysis of the elderly 12-lead ECG signals. Physica A 540:123234
26. Wang J, Shao W, Kim J (2020) ECG classification comparison between Mf-DFA and Mf-DXA. Fractals 29(02):2150029
27. Wang J, Jiang W, Yan Y, Chen W, Kim J (2021) ECG classification using multifractal detrended moving average cross-correlation analysis. Int J Mod Phys B 35(32):2150327

28. Mishra AK, Raghav S (2010) Local fractal dimension based ECG arrhythmia classification. Biomed Signal Proces 5(2):114–123
29. Li HF, Lee SY (2009) Mining frequent itemsets over data streams using efficient window sliding techniques. Expert Syst Appl 36(2):1466–1477
30. Vapnik V (1999) The nature of statistical learning theory. Springer science & business media
31. Liu D et al (2016) Precipitation complexity measurement using multifractal spectra empirical mode decomposition detrended fluctuation analysis. Water Resour Manag 30(2):505–522
32. Li Y et al (2018) Combining convolutional neural network and distance distribution matrix for identification of congestive heart failure. IEEE Access 6:39734–39744
33. Liang Y, Yin S, Tang Q, Zheng Z, Elgendi M, Chen Z (2020) Deep learning algorithm classifies heartbeat events based on electrocardiogram signals. Front Physiol 11:569050
34. Kuila S, Dhanda N, Joardar S (2022) ECG signal classification and arrhythmia detection using ELM-RNN. Multimedia Tools Appl 81(18):25233–25249
35. Romdhane TF, Pr MA (2020) Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. Comput Biol Med 123:103866
36. Tseng KK, Wang C, Xiao T, Chen CM, Hassan MM, de Albuquerque VHC (2021) Sliding large kernel of deep learning algorithm for mobile electrocardiogram diagnosis. Comput Electr Eng 96:107521
37. Naqvi SF et al (2020) Real-time stress assessment using sliding window based convolutional neural network. Sensors 20(16):4400
38. Ma S, Cui J, Chen CL, Chen X, Ma Y (2022) An effective data enhancement method for classification of ECG arrhythmia. Measurement 203:111978

**Publisher's Note**  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.