Computer Methods and Programs in Biomedicine 153 (2018) 171–184

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III

K.H. Boon*, M. Khalil-Hani, MB Malarvili

Faculty of Electrical Engineering, Universiti Tekonologi Malaysia, Skudai, Johor 81310, Malaysia

ARTICLE INFO

Article history:
Received 17 February 2017
Revised 27 August 2017
Accepted 10 October 2017

Keywords:
Heart rate variability
Paroxysmal atrial fibrillation
Arrhythmia prediction
Multi-objective optimization
Non-dominated Sorting genetic algorithm III
Feature selection

ABSTRACT

This paper presents a method that able to predict the paroxysmal atrial fibrillation (PAF). The method uses shorter heart rate variability (HRV) signals when compared to existing methods, and achieves good prediction accuracy. PAF is a common cardiac arrhythmia that increases the health risk of a patient, and the development of an accurate predictor of the onset of PAF is clinical important because it increases the possibility to electrically stabilize and prevent the onset of atrial arrhythmias with different pacing techniques. We propose a multi-objective optimization algorithm based on the non-dominated sorting genetic algorithm III for optimizing the baseline PAF prediction system, that consists of the stages of pre-processing, HRV feature extraction, and support vector machine (SVM) model. The pre-processing stage comprises of heart rate correction, interpolation, and signal detrending. After that, time-domain, frequency-domain, non-linear HRV features are extracted from the pre-processed data in feature extraction stage. Then, these features are used as input to the SVM for predicting the PAF event. The proposed optimization algorithm is used to optimize the parameters and settings of various HRV feature extraction algorithms, select the best feature subsets, and tune the SVM parameters simultaneously for maximum prediction performance. The proposed method achieves an accuracy rate of 87.7%, which significantly outperforms most of the previous works. This accuracy rate is achieved even with the HRV signal length being reduced from the typical 30 min to just 5 min (a reduction of 83%). Furthermore, another significant result is the sensitivity rate, which is considered more important that other performance metrics in this paper, can be improved with the trade-off of lower specificity.

1. Introduction

Atrial Fibrillation (AF) is the common non-life-threatening cardiac arrhythmia that can lead to stroke, heart failure, and other heart related disease [1,2]. Patients often start with episodes of paroxysmal atrial fibrillation (PAF), which last from seconds to days but it is self-terminating. It also can be treated by medication or electrical shock issued by the Implantable Defibrillator Device (ICD) [3]. However, the PAF can slowly evolve to the chronic AF that cannot return to normal sinus rhythm even with external treatment. Therefore, the development of an accurate predictor of the onset of PAF is clinically important because it increases the possibility to electrically stabilize and prevent the onset of atrial arrhythmias with different pacing techniques [4]. This can lead to decrease in symptoms, and possibly a decrease in atrial remodeling that causes increased susceptibility to future episodes of PAF [3].

Much research have been done for developing a method that can predict the onset of PAF based on electrocardiogram (ECG) signal. The works can be divided into premature atrial complexes (PAC) detection [5–7] and heart rate variability (HRV) analysis [8–13]. Tables 5 and 6 summarizes their methodology and prediction performance. Almost all existing methods, which achieved acceptable prediction accuracies (around 80% and above), employed 30 min signal for feature extraction [8]. Some of them [5,9,10] even could achieve same or above the level of 90%.

Previous works also attempted to use the HRV signal shorter than 15 min for prediction. However, their accuracy rates were lower when compared to the methods that used 30 min signal. For example, Boon et al. [8] proposed a HRV analysis prediction method based on HRV analysis, and they achieved accuracy rate of 79.3% and 68.9% for 15 min and 10 min respectively. Yang and Yin [12] achieved lowest accuracy rate with 57% when they extracted features from 10 min HRV signal based on footprint analysis. With spectral features, Hickey and Heneghan [6] achieved prediction accuracies of 68%, 70%, and 66% for 5, 10 and 30 min of HRV signal respectively.

* Corresponding author.
E-mail addresses: khboon2@live.utm.my (K.H. Boon), khalil@fke.utm.my (M. Khalil-Hani), malarvili@biomedical.utm.my (M. Malarvili).

https://doi.org/10.1016/j.cmpb.2017.10.012
0169-2607/

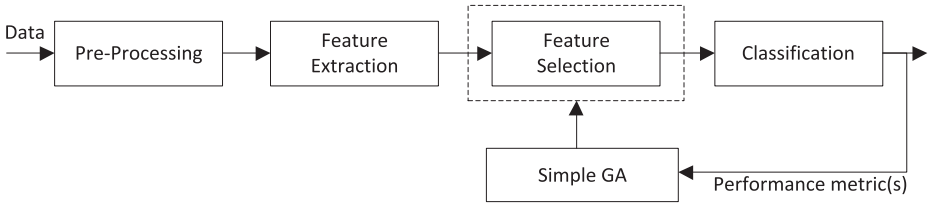

Fig. 1. Block diagram of typical feature selection based on simple GA.

Therefore, this main objective of this paper is to propose a PAF onset prediction method that requires shorter time duration of the HRV signal during the feature extraction, while improving the prediction accuracy level.

2. Optimization method

Multiple features are usually extracted during the HRV analysis. However, using all features does not always give the best classification performance due to the curse of dimensionality [14]. Therefore, feature selection algorithm is usually used to improve the performance. The feature selection is an optimization problem that involves selecting the best combination of features from the original extracted features without transforming them. It can enhance the performance, reduce the number of features, and help the researchers to understand which features are important to the classification model.

Genetic algorithm (GA) is one of the popular methods for feature selection. Fig. 1 shows the block diagram of the typical feature selection model based on simple GA for the existing research based on HRV analysis [8,15–19]. After the features are extracted from the pre-processed data, the simple GA is used to select the feature subset with high classification performance. Based on critical review, there are a few shortcomings for the feature selection model in Fig. 1.

One of the shortcomings is the parameter values and settings in both HRV pre-processing and feature extraction stages are not optimized (tuned) for maximum classification performance. As shown in Fig. 1, before the feature selector is applied, the HRV features must be extracted based on certain pre-defined values and fixed setting of HRV feature extraction algorithms. According to Rashedi et al. [20], to maximize the performance of extracted features, these parameters should be tuned simultaneously with the feature selection process for different application and database. They proposed a heuristic search method called gravitational search algorithm (GSA) to optimize the image recognition system. It simultaneously optimized both the parameters of feature extraction algorithms (wavelet transform and color histogram) and the feature subset, and this improved detection rate of their system. Inspired by their work, this paper intends to propose an optimization algorithm that can simultaneously optimizes the parameters in preprocessing and feature extraction stages, feature subset in feature selection process, and parameters of classification model. Such optimization model is shown in Fig. 2.

Another shortcoming is that the trade-off between the classification sensitivity and specificity rate is not considered. Improving certain amount of sensitivity may need to sacrifice certain degree of specificity, or vice versa [21]. Some medical applications require high sensitivity while other need high specificity [21]. For example, Xie and Minn [22] and Koley and Dey [23] developed the algorithm for detecting the sleep apnea based on ECG signal. They were interested in high sensitivity rate because it reduced the risk of overlooking the apnea events that could pose threats to the patients. In this paper, we are more interested in improving the sensitivity at the expense of acceptable reduction in specificity rate when predicting the PAF onset. Therefore, the trade-off between different prediction performance metrics is considered when developing the optimization algorithm.

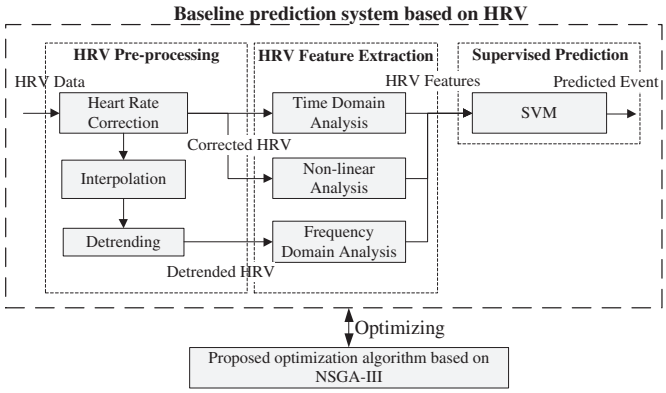

Fig. 2. Overview of the proposed method.

In this paper, optimizing the PAF onset prediction model is a multiple-objective problem. When GA is used for this class of problems, there are two common approaches: weighted sum, and Pareto dominance concept [24]. The former case is used in simple GA, in which multiple objective fitness functions are linearly combined with different weight coefficients into a single composite function. It was employed in previous works [8,18,19] to combine different performance metrics (i.e., sensitivity, specificity, accuracy rate and feature count). There are several drawbacks with this approach [24]. One of them is that the trial and error is required for tuning the weights values in order to obtain a solution with desired performance. Moreover, the simple GA only can return single solution per optimization run. As a result, the GA needs to be run multiple times for obtaining multiple solutions before trade-off among the solutions can be analyzed, which is not convenience for designer. It should be noted that, although the fitness functions in previous works [8,18,19] had multiple metrics, they did not perform the trade-off analysis because their interest was obtaining a single solution with highest accuracy. The drawbacks in the simple GA can be tackled by using the Pareto dominance concept based GA. In this paper, the state-of-the-art Pareto dominance based GA, which is called non-dominated sorting genetic algorithm III (NSGA-III) [25, 26], is adopted for optimization.

Finally, the GA based feature selectors in HRV based previous works [8,15–19] are belong to the type of wrapper method [27] because only the machine learning classifier is used to evaluate the fitness of the chromosome. Wrapper method has a well-known shortcoming: the risk of selecting a subset that is overfitting to the trained supervised classifier [8,27]. In the non-HRV research, the hybrid feature selection based on simple GA [28–30] has been proposed to mitigate this issue to certain degree. The hybrid GA uses the filter method (i.e., statistical test or correlation measure) to evaluate the feature, and only selects the feature that can pass certain evaluation criterion during the formation of feature subset. In the research based on HRV analysis, Boon et al. [8] cascaded

the filter method called Mann-Whitney U test with the simple GA based feature selector. Before the GA was applied, they filtered out the HRV features that could not pass the U test at 20% significance level. Therefore, in this paper, the filter method is used as part of the proposed optimization algorithm for evaluating the features before they are selected to form feature subset.

In this paper, a multi-objective optimization algorithm based on NSGA-III is proposed to optimize the PAF onset prediction system while considering the above-mentioned shortcomings. The layout of the paper is as follows. Section 3 presents the database and proposed method. Results and discussion are presented in Section 4. Finally, conclusion and future works are given in the Section 5.

3. Proposed method

Fig. 2 shows overview of the proposed prediction method. It comprises of the stages of pre-processing, HRV feature extraction, and support vector machine (SVM) model, which are simultaneously optimized by the proposed optimization algorithm based on NSGA-III. Initially, 5 min HRV data that immediately precedes the PAF event is fed to the pre-processing stage for heart rate correction, interpolation and signal detrending. They are two types of pre-processed output: corrected HRV and detrended HRV. It should be noted that the detrended HRV is also corrected and interpolated. Then, 6 time-domain and 4 non-linear features are extracted based on the corrected HRV, while 43 frequency-domain features are extracted based on the detrended HRV. Finally, these features are used as input to the SVM model for predicting the PAF event. The SVM is implemented by using the C++ library called LIBSVM [31].

3.1. Experimental data

Based on the previous works [6–13], 106 data from 53 pairs of ECG recordings (each pair is recorded from different PAF patients) are obtained from the standard database called Atrial Fibrillation Prediction Database (AFPDB) [32]. Each pair of data contains one 30 min ECG segment that ends just prior to the onset of PAF event and another 30 min ECG segment at least 45 min distant from any onset of PAF. Each ECG segment contains two-channel traces from Holter recording with sampling rate of 128 Hz and 12-bit resolution.

In this paper, the 5 min HRV segment that at least 45 min distant from the PAF event is assigned a class label of “NORMAL”, while the HRV segment that immediately precedes the PAF event is given a class label of “ABNORMAL”.

3.2. Preprocessing

In the preprocessing stage, the RR intervals (intervals between successive R peaks) are derived from the ECG signal by using the Hamilton and Tompkins algorithm [33]. Then, the HRV data is computed as the reciprocal of RR intervals. After that, the HRV data segment sequentially goes through the heart rate correction, interpolation and signal detrending. The heart rate correction and signal detrending are performed based on the McNames algorithm [34] and the high pass filter proposed by Tarvainen et al. [35] respectively. As for the interpolation process, either linear or cubic spline method is used to resample the HRV data to certain frequency (4 Hz or 7 Hz) [36]. The methodology and reason for optimizing these 3 pre-processing steps with the proposed algorithm are discussed in Section 3.5.2.

3.3. HRV feature extraction

Fifty-three well-established HRV features are extracted using time-domain, frequency domain and non-linear analysis [37]. Their abbreviations are listed in Table A.1 and explained in the following sub-sections. It should be noted that, they are one or more input parameters for the feature extraction algorithms that belong to frequency domain and non-linear analysis. In this paper, these parameters are optimized by the proposed optimization algorithm, that is discussed in Section 3.5.2. However, the recommended values for the parameters are also presented because they are used during the result analysis in Section 4.

3.3.1. Time domain features

Six time-domain HRV features are computed by using statistical analysis. They are the mean of HRV (Mean), standard deviation of HRV (SDRR), root mean square of successive difference intervals (RMSSD), number of adjacent RR intervals differing by more than 50 ms (NN50), and sum of NN50 divided by the total number of all RR intervals (pNN50). RR triangular index (RRTri) [37] is also extracted as a geometric feature. It is defined as total number NN intervals divided by number of RR intervals that fall to modal bin.

3.3.2. Spectral features

The total spectral power in low frequency (LF) band (0.04–0.15) and high frequency (HF) band (0.15–0.4 Hz) of the power spectral density (PSD) can be related to the sympathetic and parasympathetic activities of the autonomic nervous system respectively [14]. In this paper, both auto-regressive (AR) model [10] and fast Fourier transforms (FFT) [11] are used to estimate the PSD. Three features are computed from each estimation method. They are total spectral power in LF band, HF band, and ratio of LF to HF.

The coefficients of AR model are estimated with burg method, and it has one input parameter: order of the model. The order is set to 16 based on the recommendation in [38]. As for the FFT, the HRV data segment is multiplied with the temporal smoothing window function before the PSD is estimated. The default window function is rectangular window. Furthermore, each estimated PSD is normalized before the HRV features are extracted.

3.3.3. Bispectrum features

Higher order spectral (HOS) analysis has been used to estimate the bispectrum in the recent HRV analysis based research [8,10,19]. In this paper, HOS up to third-order cumulant is employed to estimate the bispectrum from HRV data. The estimation is based on the direct method described in [39]. The HRV data is divided into several segments with each segment contains 512 data points with 50% overlapping. After that, each segment is smoothed by rectangular window, and zeroes are padded at the end the segment if the segment data length is not power of 2. Then, FFT is applied to each segment for computation of the bispectrum.

Bispectrum of HRV signal can be divided into 3 subband regions inside region of interest (ROI) according to Yu and Lee [19]. They are LF–LF (LL), LF–HF (LH), and HF–HF (HH) regions which cover different ranges of frequencies. Formulas in [19,40] are employed to compute bispectrum features from each subband region and ROI. These features include mean magnitude (M_{ave}), mean of sum of squared magnitude (P_{ave}), normalized bispectral entropy (P1), normalized bispectral squared entropy (P2), sum of logarithmic amplitudes of the bispectrum (H1), sum of logarithmic amplitudes of diagonal elements in bispectrum (H2), first-order spectral moment of the amplitudes of diagonal elements in the bipsectrum (H3), Second-order spectral moment of the amplitudes of diagonal elements in the bispectrum (H4), weighted center of the bipsectrum, WCOB (f_{1m}, f_{2m}). For LH region, H2, H3 and H4 are not computed because the diagonal elements are not existed.

3.3.4. Nonlinear dynamics features

Poincare plot and sample entropy are used to extract non-linear features from HRV data. Poincare plot is drawn by plotting each

Algorithm 1 Proposed optimization algorithm.

Input: $p_c, p_m, N, M, d^1, d^2, GEN_{MAX}$
Output: F_1 *when* $t = GEN_{MAX}$
$t = 0$
Z^r = Initialize-reference − point (M, d^1,d^2)
P_t = Initialize-population()
repeat
  $S_t = \emptyset$, $i = 1$
  Q_t = Generatic-operation (P_t,p_c,p_m)
  if $t = 0$ **then**
    $V_t = P_t \cup Q_t$
  else
    $V_t = Q_t$
end if
//————————————————————————————
//Proposed Fitness Evaluation Stage
HRV-feature-extraction (V_t)
Statistical-performance-evaluation (V_t)
Local-search-operation (V_t)
Duplication-handling (V_t)
Prediction-performance-evaluation (V_t)
Feature-count-evaluation(V_t)
//————————————————————————————
if $t = 0$ **then**
  $R_t = V_t$
else
  $R_t = V_t \cup P_t$
end if
Fronts: $(F_1,F_2,\ldots)$ = NDS(R_t) /*Non-dominated sorting*/
repeat
  $S_t = S_t \cup F_i$ and $i = i = 1$
until $|S_t| \geq N$
Last front to be included: $F_l = F_i$
If $|S_t| = N$ **then**
  $P_{t+1} = S_t$, **break**
else
  $P_{t+1} = \cup_{j=1}^{l-1} F_j$
  Number of individuals to be chosen from F_l: $K = N - |P_{t+1}|$
  Normalize (S_t,$\boldsymbol{f}^1$)
  $[\pi(\boldsymbol{s}),\ d(\boldsymbol{s})]$ = Associate (S_t,Z^r) /*Associate each $\boldsymbol{s} \in S_t$ with a reference point*/
  /* $\pi(\boldsymbol{s})$: closest reference point*/
  /* $d(\boldsymbol{s})$: distance between $\boldsymbol{s}$ and $\pi(\boldsymbol{s})$*/
  $\rho_j = \sum_{\boldsymbol{s} \in S_t/F_l} (\pi(s) = j ? 1 : 0)$ /*Compute niche count of reference point $j \in Z^r$*/
  Niching (K, ρ_j,$\pi(\boldsymbol{s})$, $d(s)$, Z^r,F_l,$P_{(t+1)}$) /*Choose K members one at a time from F_l to construct P_{t+1})*/
end if
$t = t + 1$
until $t = GEN_{MAX}$
Fronts: $(F_1,F_2,\ldots)$ = NDS-without − statistic − fitness(P_t)

RR interval against next RR interval. Three features are computed from the Poincare plot. They are denoted by $SD1$, $SD2$ and the ratio $SD1/SD2$ [37]. Sample entropy (SampEn) is a statistic measure that quantifies the regularity of times series data. The method proposed in [31] is used to compute the SampEn of the HRV signal. Based on the recommendation in [41], the input parameters are set as follows: embedding dimension, $m = 2$ and tolerance distance, $r = 0.2$.

3.4. Classification: support vector machine (SVM)

The support vector machine (SVM) algorithm based on the statistical learning theory was proposed by Chang and Lin [31]. SVM maps the training samples from the input space to higher-dimensional features space via a kernel function. Product between vectors of training sample is used to generate a hyper-plane that can separate two classes. Optimization process of SVM classifier is aimed to find the optimal hyper-plane that maximizes the distance between training samples of two classes. The largest distance to the support vectors (training samples that is closest to the hyper-plane) is so-called functional margin. Larger functional margin indicates the generalization error of the classifier is lower.

In this paper, SVM is used as supervised classifier to classify the HRV segments to either "NORMAL" (segment that at least 45 min distant from the PAF event) or "ABNORMAL" (segment that immediately precede the PAF event). Input of the SVM is the HRV features that are extracted during the feature extraction stage. The radial basis function (RBF) is used as the kernel function for SVM. Parameters of kernel function–kernel width γ and penalty constant C–are optimized by NSGA-III to achieve the best result.

3.5. Proposed optimization algorithm

Algorithm 1 shows the proposed optimization method that is developed based on the non-dominated sorting genetic algorithm III (NSGA-III) [25]. The optimization procedure is similar to the original NSGA-III except the chromosome design, fitness evaluation stage, genetic operators and other some minor modifications.

There are 7 input parameters for the proposed optimization algorithm: (1) M that specifies number of fitness functions, (2) d^1 and d^2 that specify number of divisions for boundary layer and inside layer respectively during the two-layer reference point generation process, (3) population size N that specifies the size of the populations denoted by P_t and Q_t, (4) crossover rate p_c and mutation rate p_m that control the probability of genetic operation, and (5) maximum generation GEN_{MAX} that limits the maximum number of iterations for the optimization process. The output of the

Algorithm 2 Normalization (S_t,f^n).

1: **Input:** $\boldsymbol{s} \in S_t$ /*chromosome $\mathbf{s}^*$/
2: **Output:** f^n /*Normalized vector*/
3: **for** $j = 1$ *to* M **do**
4: Compute ideal point: $z_j^{min} = min_{\boldsymbol{s}\in S_t} f_j(\boldsymbol{s})$
5: Compute the maximum point: $z_j^{max} = max_{\boldsymbol{s}\in S_t} f_j(\boldsymbol{s})$
6: Translate objectives: $f_j^n(\boldsymbol{s}) = \frac{f_j(\boldsymbol{s}) - z_j^{min}}{z_j^{max} - z_j^{min}} \forall \boldsymbol{s} \in S_t$ /* f_j^n is a normalized jth objective value of a chromosome $\boldsymbol{s}^*$/
end for

Feature Set	Time Domain	RRTri	Poincare Plot	Sample Entropy	Auto-regressive Model	FFT	Higher Order Spectral	SVM Penalty Constant	SVM Gamma
$B_f^1, B_f^2 ..... B_f^{53}$	B_t^1	B_R^1	B_p^1	$B_s^1, B_s^2 ..... B_s^{10}$	$B_{ar}^1, B_{ar}^2 ..... B_{ar}^{10}$	$B_{fft}^1, B_{fft}^2 ..... B_{fft}^9$	$B_{hos}^1, B_{hos}^2, ... B_{hos}^{10}$	$B_C^1, B_C^2 ..... B_C^{20}$	$B_\gamma^1, B_\gamma^2 ..... B_\gamma^{20}$

Fig. 3. Chromosome design.

proposed optimization algorithm is the chromosomes (Pareto optimal solutions) that are belong to the front F_1 of population P_t at the end of the optimization.

Like in the original NSGA-III, two-layer reference points denoted by the Z^r are generated according to the Das and Dennis approach [42]. The initialization function takes the input parameters M, d^1 and d^2. The parent population $P_{t=0}$ with population size N are also randomly initialized. Then, two-point crossover and bit-flip mutation operators [29] are applied to P_t for producing the offspring population Q_t with same population size N. The crossover and mutation rates are determined by p_c and p_m respectively.

After that, both population P_t and Q_t are combined to formed the population set V_t if generation $t = 0$. Otherwise, V_t only consists of offspring population Q_t. The population V_t sequentially goes through the different stages of fitness evaluation: HRV feature extraction, statistical performance evaluation, local search operation (LSO), duplication handling, prediction performance evaluation and feature count evaluation. After the fitness evaluation stage, R_t is assigned V_t if generation $t = 0$. Otherwise, it consists of both V_t and P_t. Then, all chromosomes in the population R_t are sorted into different non-domination levels called fronts (F_1,F_2,...) using the dominance principle [25]. A new population S_t is constructed by selecting individuals from different fronts (F_1,F_2,...) until the size of S_t is equal to N or greater than N for the first time. The last front (F_i), which is assigned to S_t, is denoted by front F_l.

After that, P_{t+1} is assigned S_t if the number of individuals in S_t is exactly equal to N. Otherwise, P_{t+1} is formed by the chromosomes from front F_1 up to F_{l-1}. In this case, they are still K chromosomes need to be chosen from front F_l to fill the P_{t+1} until $|P_{t+1}| = N$. The K chromosomes are chosen based on a series of operations, which are adaptive normalization, association and niching. Except the normalization process in line 38, the procedures of Algorithm 1 from line 36 to 41 are same as the original NSGA-III [25]. The mix-max normalization proposed by Yuan et al. [43] is adopted to normalize the fitness vector of every chromosome that belong to set S_t. This process is summarized in Algorithm 2. The procedure above is repeated until GEN_{MAX}. After that, the function NDS-without-statistic-fitness is applied to P_t in order to obtain the solutions that belong to front F_1.

3.5.2. Chromosome design

The structure of the proposed chromosome design is shown in Fig. 3. The parameters and settings of the HRV based arrhythmia prediction system are encoded into binary digits in the chromosome for optimization. The chromosome contains 135 bits that can be divided into 10 different segments.

The feature set segment is a 53-bit binary string that represents the selected HRV feature subset for a chromosome. Each bit in the segment represents the feature selection status of one HRV feature. The bit "1" indicates selection while "0" represents the deletion of the specific feature from the feature set.

The last 2 segments are two 20-bit binary strings that represent the encoded value of the parameters C and γ respectively for the SVM model. The remaining seven segments, which are in the middle of a chromosome, represent various parameters and settings for 7 different HRV feature extraction algorithms. Each of them is explained in Table 1.

Firstly, some of the binary digits in the Table 1 are used to control the behavior of the heart rate correction, interpolation and signal detrending in the pre-processing stage of the baseline prediction system (refer to Fig. 2). Two separate bits are used to decide whether the heart rate correction and detrending are performed on the HRV data respectively. As for the interpolation process, 2 different bits are used to control its behavior. One of them is used to determine either linear or cubic spline method is chosen to resample (interpolate) the HRV data. Another one is used to select the resampling frequency (either 4 Hz or 7 Hz).

The heart rate correction enabling status is important because the abnormal heart rate may contain the required information to predict the arrhythmia event. In previous works [5–7], authors counted the number of the occurrence of premature atrial contractions (PACs) (a type of arrhythmia) in the ECG signal. The counted number was used to predict the PAF onset. The occurrences of the arrhythmia cause the abnormal heart rates in the time series HRV data. Therefore, the abnormal heart rate prior to the PAF onset may be useful for the PAF onset prediction. These abnormal heart rates are eliminated when the heart rate correction algorithm is applied. As for the interpolation and detrending, they can influence the energy value of the power spectrum estimated by frequency domain algorithms.

For sample entropy analysis, the parameter values of the embedding dimension and tolerance distance are tuned by the proposed algorithm. Their values are set to integer range of 1–4 and floating point range of 0.1–0.5 respectively according to Lake et al. [44]. Furthermore, the order of the auto-regressive (AR) model is also set to integer range of 6–32 based on the investigation in Anita et al. [38]. In both fast Fourier transform (FFT) and higher order spectral (HOS) analysis, one type of temporal smoothing window function is selected among 8 different types of window function when the spectral analysis is performed. They are Rectangular, Parzen, Hanning, Hamming, Blackman, Blackman Harris, Welch, and Barlette window. Furthermore, one bit denoted by "FFT segmentation" in the Table 1 is used to decide whether regular FFT or Welch based FFT is chosen for performing the Fourier transform. Besides that, there is one extra bit that controls whether the PSD estimated by the AR and FFT are normalized to between 0 and 1 before the related HRV features are computed. Finally, the segmen-

Table 1
Definition of the chromosome representation.

	Feature extraction algorithm	Algorithm parameters	Bit length	Representation
1	Time domain	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
2	RR Triangular Index	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
3	Poincare plot	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
4	Sample entropy	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
		• Embedding dimension	2 bit	• Represent an integer value within the range of 1–4.
		• Tolerance Distance	7 bit	• Represent a floating value within the range of 0.1–0.5.
5	Auto Regressive (AR) analysis	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
		• Heart rate detrending enabling	1 bit	• '0' represents disabling while '1' represents enabling of the detrending.
		• Interpolation method	1 bit	• '0' represents linear interpolation while '1' represents cubic spline interpolation.
		• Resampling frequency	1 bit	• '0' represents 4 Hz while '1' represents 7 Hz.
		• Order of model	5 bits	• Represent an integer value within the range of 6–32.
		• Normalization of AR spectrum	1 bit	• '0' represents no normalization while '1' represents the spectrum is normalized by the maximum value of AR spectrum.
6	Fast Fourier Transform (FFT)	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
		• Heart rate detrending enabling	1 bit	• '0' represents disabling while '1' represents enabling of the detrending.
		• Interpolation method	1 bit	• '0' represents linear interpolation while '1' represents cubic spline interpolation.
		• Resampling frequency	1 bit	• '0' represents 4 Hz while '1' represents 7 Hz.
		• Temporal smoothing window function	3 Bit	• Represents 8 types of symmetric smoothing window function. • They are Rectangular, Parzen, Hanning, Hamming, Blackman, Blackman Harris, Welch and Barlette.
		• FFT segmentation	1 Bit	• '0' represents normal FFT while '1' represents welch based FFT with 50% data segment overlap.
		• Normalization of FFT spectrum	1 bit	• '0' represents no normalization while '1' represents the spectrum is normalized by the maximum value of FFT spectrum.
7	Higher Order Spectral (HOS) analysis	• Heart rate correction enabling	1 bit	• '0' represents disabling while '1' represents enabling of the ectopic beat correction.
		• Heart rate detrending enabling	1 bit	• '0' represents disabling while '1' represents enabling of the detrending.
		• Interpolation method	1 bit	• '0' represents linear interpolation while '1' represents cubic spline interpolation.
		• Resampling frequency	1 bit	• '0' represents 4 Hz while '1' represents 7 Hz
		• Temporal smoothing window function	3 bit	• Represents 8 types of symmetric smoothing window function. • They are Rectangular, Parzen, Hanning, Hamming, Blackman, Blackman Harris, Welch and Barlette.
		• Segmentation size	1 bit	• '0' represents 256 samples per segment while '1' represents 512 samples per segment.
		• Zero-padding size for the segmented data.	1 bit	• '0' represents the padding size is equal to segmentation size while '1' represents twice of the segmentation size.
		• Overlapping of the segmented data	1 bit	• '0' represents no overlapping in the segmented data while '1' represents 50% overlap of the segmented data.

tation size, zero-padding size and overlapping of the segmented data are also determined for the HOS analysis.

In the proposed chromosome design, the real number R of a parameter that is encoded into binary string p (that has more than 1 bit) can be computed as:

$$R = \min_p + \frac{\max_p - \min_p}{2^l - 1} \times d \quad (1)$$

where d is decimal value of binary string p, $\max_p$ is maximum value of parameter, $\min_p$ is minimum value of parameter, l is length of binary string p. The $\max_p$ and $\min_p$ for each real number parameter shown in Table 1 are different. They are set according to their minimum and maximum value of the specified integer or floating point range. As for the SVM parameters C and γ, their minimum and maximum value are set to 0.1 and 1000 respectively.

3.5.3. Fitness evaluation stage

The proposed fitness evaluation stage consists of HRV feature extraction, statistical performance evaluation, local search operation, duplication handling, prediction performance evaluation and feature count evaluation. Before entering the fitness evaluation stage, the duplicate chromosomes from population Q_t are identified by comparing them to all chromosomes from P_t and Q_t. The fitness evaluation of these duplicate chromosomes are skipped and reserved until they are modified by duplicate handling process (line 19 of Algorithm 1). The detail of the duplicate chromosome is explained in Section 3.5.3.4 of this section.

3.5.3.1. HRV feature extraction. During the HRV feature extraction stage, for every chromosome that belong to the set V_t (referring to Algorithm 1), 53 HRV features in Table A.1 are extracted based on the decoded parameter values and settings. These parameter values are presented in Table 1, and decoded based on the Eq. (1). After

the feature extraction process, all HRV features are fed to statistical performance evaluation stage.

3.5.3.2. Statistical performance evaluation. The Mann–Whitney U test is a non-parametric method used to test whether two independent samples of observations are drawn from the same or identical distributions. Perez et al. [45] proposed a feature selection method called uFilter method, which is based on the Mann-Whitney U test, for single-objective optimization: feature selection. It is a standalone algorithm that evaluates the quality of a feature. In their work, the statistic value denoted by w_i was computed for each feature. Then, top 20 features (with highest statistic values w_i in descending order) were selected to form a feature set. This feature set was fed to the classification models for training and testing.

In this paper, the uFilter method is used to evaluate the HRV features of each chromosome for two optimization objectives: (1) to optimize the parameters and settings that belong to HRV feature extraction algorithms, and (2) to be used as an evaluation criterion in the local search operation (LSO). In the former objective, the optimization is achieved by minimizing the sum of the statistic values of all HRV features. It is computed as:

$$\text{Sum of statistic value} = -\sum_{i=1}^{53} w_i \tag{2}$$

where w_i is the absolute value of numerical difference between Z_{NORMAL} and $Z_{ABNORMAL}$ for ith feature (one of the 53 HRV features). The algorithm that computes the w_i, Z_{NORMAL} and $Z_{ABNORMAL}$ can be found in [45]. The Z_{NORMAL} is the Z-indicator of feature values that are extracted from HRV segments with the class label of "NORMAL", while the $Z_{ABNORMAL}$ is the Z-indicator of feature values extracted from HRV segments with the class label of "ABNORMAL". These two class labels are explained in Section 3.1.

3.5.3.3. Local search operation (LSO). In this paper, the local search operation (LSO) is developed based on the hybrid feature selection process of the simple GA proposed by Huang and Rong [30]. The objectives of proposed LSO are: (1) to remove the non-significant features from the feature subset in the chromosome so that only the good features are retained. (2) to maintain the minimum number of selected significant features in the chromosome after the removal process. The "significant feature" is the feature that can pass the evaluation criterion: two tailed Mann-Whitney U test at 20% significance level. Otherwise, it is referred as "non-significant feature".

Fig. 4 shows the flow chart of the proposed LSO. Initially, the HRV features, which are selected within the feature subset segment of the chromosome, are examined. The non-significant features are removed from the feature subset. This approach was used in [8] for manually removing the features that could not pass the U test before the simple GA was applied for feature selection.

After the above removal process, the chromosome is updated to reflect the changes. Then, the minimum number of selected significant features is evaluated. If the number equal or more than the integer number of 3 (denoted by "min_num" in the Fig. 4), the LSO is ended. Otherwise, the handling process is started to increase the feature count to 3. In the handling process, the first step is to examine whether there is enough number of significant features for the selection. If it is enough, then one additional unselected significant feature is randomly selected and added to the feature subset. Otherwise, the chromosome goes through the mutation operation (genetic operation), HRV feature extraction stage and statistical evaluation stage again. Finally, the LSO is applied to this new chromosome. The handling process is repeated until the chromosome has same or more than the minimum number of significant features.

3.5.3.4. Duplication handling. Duplicate chromosome is the creation of the same chromosome that has been evaluated before the current generation during the GA optimization process. It causes the computation wastage due to the redundant evaluation of duplicate chromosomes [46].

In this paper, the duplicate handling process is proposed to handle (remove and replace) the duplicate chromosome. It is developed based on the method proposed by Saroj and Devraj [46]. They modified the duplicate chromosomes when there was a same chromosome exists in the one previous and the most current generations of the simple GA. Their results showed that the optimization performance of the simple GA was improved. Therefore, the same concept is applied to the proposed optimization algorithm in this paper because by default, the NSGA-III stores the previous population (population P) and current population (population Q) in one generation.

Fig. A.1 shows the flow chart of the proposed duplicate handling process. The algorithm can be divided into 3 main parts: identifying stage, modification stage, and genetic operation stage. In the identifying stage, the binary digit pattern of a chromosome is compared against all the chromosomes in population P_t and Q_t. If its binary digit pattern is unique, the duplication handling process is ended. Otherwise, the chromosome is fed to the modification stage.

In the modification stage, the content of the duplicate chromosome is modified so that it become unique among the population. When a chromosome enters this stage for the first time, its status represented by Duplication_Code is set to "0". Then, if the number of selected significant features in the feature subset is equal to total number of significant features (sum of both selected and unselected features in a chromosome), the duplicate chromosome is fed to reset route (by setting Duplication_Code to (1). If the number of selected significant features is equal to minimum number of features denoted by "min_num", it will go to set route (by setting Duplication_Code to (2). If neither of the conditions is met, then it is randomly fed to either set route or reset route. In the reset route, if the number of selected significant features greater than min_num, then one of the significant features in the feature subset of the chromosome is removed. Otherwise, the chromosome is fed to genetic operation stage. In the set route, if the number of selected significant features is less than total number of significant features, then one of the unselected significant features is randomly selected and added to the feature subset.

In the genetic operation stage, the first step is to find whether there is another new but different duplicate chromosome exists. If it exists, both crossover and mutation operators are applied to the modify both duplicate chromosomes. Otherwise, only mutation operator is applied to the current duplicate chromosome. The process in the genetic operation stage is repeated until the chromosome(s) becomes unique. Then, they are fed to HRV feature extraction stage, statistical performance evaluation, and local search operation. Finally, the duplicate handling process is repeated if necessary.

Several considerations are taken into account when developing the duplicate handling process. Firstly, the duplicate handling process is placed after the LSO because the duplicate chromosomes can be formed not only due to the genetic operation, but also due to the LSO. Secondly, HRV feature extraction process is avoided as much as possible during the handling process because it is the most computation intensive part of the algorithm. Therefore, the modification stage only alters the feature subset segment of the duplicate chromosome. The parameters and settings of HRV feature extraction algorithms are only modified in genetic operation stage.

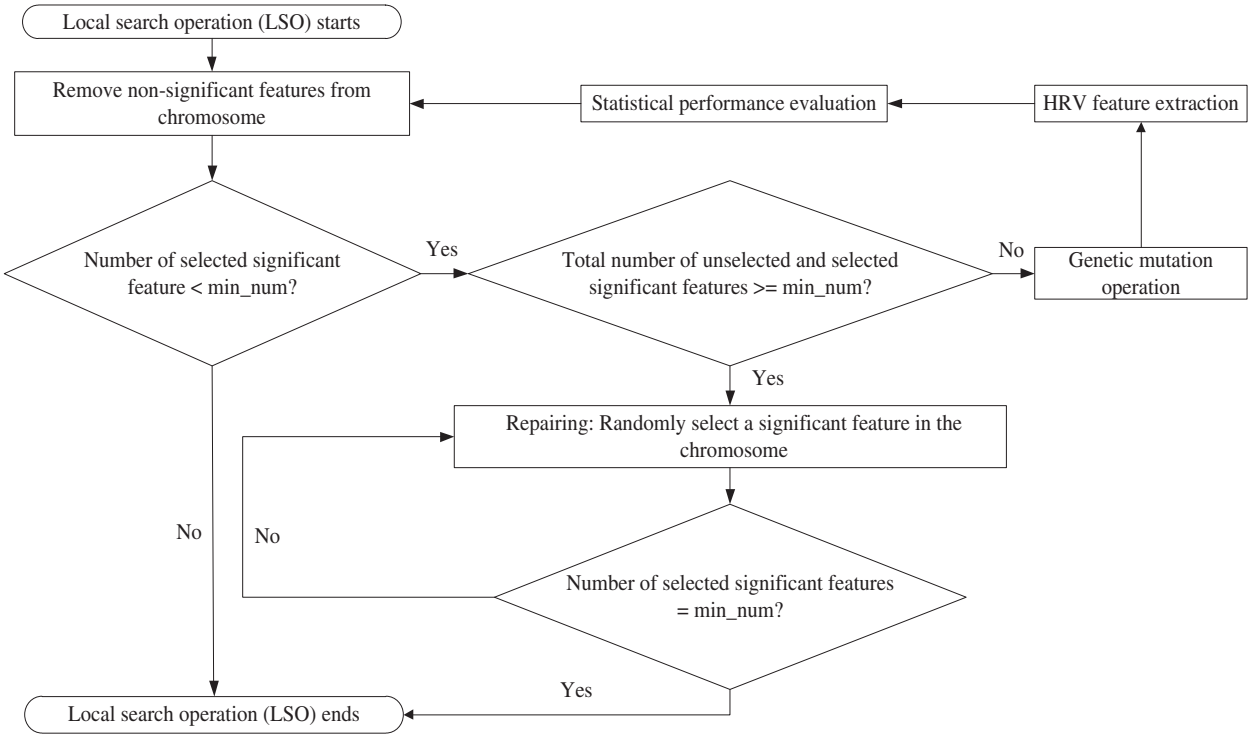

Fig. 4. Flow chart of the local search operation (LSO).

3.5.3.5. Prediction performance evaluation. The prediction performance of each chromosome is evaluated. Initially, the chromosome is decoded in order to determine the selected feature subset, and the values of SVM parameters (penalty constant C and gamma γ). After that, the selected HRV features are taken from the training dataset to train SVM model by using the decoded SVM parameter values. The trained classifier is then evaluated with the testing dataset. Sensitivity (SEN), specificity (SPE), and accuracy (ACC) are used to measure the prediction performance. They are defined as follows:

$$SEN = \frac{TP}{TP + FN} \tag{3}$$

$$SPE = \frac{TN}{TN + FP} \tag{4}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP is number of abnormal event (arrhythmia is occurred) that is correctly predicted, TN is number of normal event (No arrhythmia is occurred) that is correctly predicted, FN is number of arrhythmia event that is incorrectly predicted as normal event, FP is number of normal event that is incorrectly predicted as arrhythmia event.

10-fold cross validation is applied to evaluate the proposed method. 10 ECG recordings (5 distant and 5 prior to onset of PAF event) are selected as testing set to measure the performance of classifier while remaining recordings (96 recordings) are used to train the classifier. This procedure is repeated over 10 times in order to cover entire database. In the 10th evaluation of 10-fold cross validation, training set contains 90 ECG recordings and testing set contains 16 ECG recordings. Then, average SEN, SPE and ACC are calculated as the measures for classification performance. Training set and testing set are subject independent (belong to different patient). Since it is well known that GA suffers from local optima issue, the optimization is repeated 5 times and the best results are selected for performance analysis.

3.5.3.6. Feature count. In the proposed optimization algorithm, the feature count is defined as the number of HRV features that are selected in a feature subset. The feature count is optimized by minimizing the fitness value given by:

$$\text{Feature count} = \sum_{i=1}^{N} F_i \tag{6}$$

where F_i is '1' if ith feature is selected and '0' if not selected, and N is the total number of features (regardless the feature is selected or not) in the feature set segment of the chromosome. The N is set to 53 for this paper.

3.5.3.7. Non-dominated sorting (NDS). During the non-dominated sorting (NDS) (line 28 of the Algorithm 1), the chromosomes are divided and assigned to different sets (represented $F_1, F_2, \ldots$) [25] starting from F_1. Each set F_i contains a group of solutions that are not dominated to each other. The dominance operator denoted by "$\prec$" is the standard symbol for representing the dominance relationship between two chromosomes. Assume that there is a minimization problem, the $p \prec q$ represents that the chromosome p is said to strictly dominate another chromosome q, if and only if the $z_i(p) \leq z_i(q)$ for $i = 1, \ldots, M$ and $z_i(p) < z_i(q)$ for at least one fitness function indexed by i, where M is total number of fitness functions, and $z_i(p)$ represents ith fitness value of the chromosome p.

In this paper, five fitness functions are proposed for the optimization process. They are sum of statistic value (Eq. (2)), prediction sensitivity (Eq. (3)), prediction specificity (Eq. (4)), prediction accuracy (Eq. (5)), and feature count (Eq. (6)). To fit into the above-described dominance operator $\prec$, some maximization fitness functions are turned into minimization fitness function by multiplying each of them with a -1.0. These functions are sensitivity, specificity, and accuracy rate. The remaining fitness functions are minimization functions by default.

The procedure of the NDS-without-statistic-fitness (line 45 of the Algorithm 1) is same as the process in NDS. However, there is one minor difference between them: the operation in dominance

Table 2
Performance comparison between typical feature selection method and the proposed optimization method for PAF onset prediction.

Method	Performance			
	SEN	SPE	ACC	NF
Typical Method	88.7	66.0	77.4	4
Proposed Method	86.8	88.7	87.7	7

operator "$\prec$". The solutions in fronts F_i, which are produced by the NDS-without-statistic-fitness(), are non-dominated to each other with respect to all proposed fitness functions except the Eq. (2). In contrast, the solutions produced by the NDS() are non-dominated to each other with respect to all fitness functions. It is because this paper is interested in analyzing the prediction performance of the solutions.

4. Results and discussion

4.1. Performance comparison between the typical method and the proposed optimization algorithm

In this section, the performance difference between the PAF onset prediction systems (solutions), which are optimized by the proposed algorithm and the typical GA based feature selection method respectively, is investigated.

The detail of the proposed optimization algorithm is described by the Algorithm 1. The input parameters of the algorithm are set as follows: $N=212$, $p_c=0.7$, $p_m=0.1$, $M=5$, $d^1=6$, $d^2=0$, $GEN_{MAX}=1000$. As for the typical GA based feature selection (shown in Fig. 1), the optimization methodology from [8] is employed for selecting the optimal feature subset and tuning the SVM parameters. The procedure of this method can be summarized as follows: Firstly, the same 53 HRV features are extracted based on the pre-defined HRV parameter values and settings (which are explained in the Section 3.2 and 3.3). After that, each HRV feature is evaluated with two tailed Mann-Whitney U test. Only the features that can pass the U test at 20% significance level are selected and used to form a significant feature set. Finally, the simple GA feature selector is applied to this significant feature set for selecting best feature subset. It should be noted that, unlike the proposed algorithm in this paper, the HRV parameters and settings in typical method are not tuned. Therefore, the HRV feature extraction process is performed one time only since the parameter values and settings are same for every chromosome.

Each method is repeated 5 times with different initial populations. After that, the best solution represented by a chromosome is selected from each method for performance analysis. Table 2 compares the prediction performance between the prediction systems that are optimized by typical feature selection method and the proposed optimization algorithm respectively. With typical method, the PAF onset prediction system achieves accuracy rate of 77.4%. The accuracy rate of the prediction system is improved by 10.3% when it is optimized by proposed algorithm. The prediction performance improvement is achieved because the proposed algorithm simultaneously optimizes all stages of the arrhythmia prediction system, while the typical method only optimizes the feature subset and SVM classifier parameters.

The higher prediction performance can also be attributed to more number of HRV features can pass the Mann–Whitney U test. Table 3 shows the HRV features that able to pass the two tailed U test at 5%, 10% and 20% significance levels for typical method and proposed method. In typical method, only 10 HRV features can pass the U test up to 20% significance level. In contrast, the number is improved to 32 when the proposed optimization algorithm is used, and majority of them pass the 5% significance level. The improvement is achieved because the proposed optimization algorithm explicitly tunes the HRV parameter values and settings for improving the statistic value of each HRV feature by minimizing the fitness value of Eq. (2). Furthermore, in Table 3, it is observed that all significant features from the typical method also re-appear as the significant features from the proposed method, except the HH-WCOB(f_{2m}). Finally, Table A.2 shows the difference between the pre-defined and optimized values for the parameters and settings of HRV feature extraction algorithms.

When the number of features that pass the U test increases, the NSGA-III has opportunity to explore and evaluate more combination of HRV features [8]. This leads to higher possibility in obtaining the optimal feature set with higher accuracy. However, only 10 HRV features are available for typical method while 32 HRV features are available for the proposed optimization algorithm.

In addition to the prediction performance, Table 2 shows that although the optimal subset selected by typical method has lower (better) feature count than proposed method, but at the expense of significant lower accuracy rate. The typical method selects 4 features (out of 10 HRV features) to form the feature subset, which reduces the feature count by 60%. The selected features are AR-LF, LL-H2, LL-H3, ROI-H2. In contrast, the proposed method selects 7 features (out of 32 HRV features), which reduces the feature count by 78%. The selected features are NN50, pNN50, SampEn, SD2, AR-LF, LL-H1, ROI-WCOB (f_{2m}), and all of them can pass the U test at 5% significance level.

4.2. Trade-off between performance metrics

In the typical feature selection method based on the simple GA (with weighted sum approach), each GA run returns a single solution. Therefore, the typical method in Section 4.1 only gives a solution that has highest accuracy rate. In contrast, with the proposed NSGA-III based optimization algorithm, multiple solutions that have different degree of trade-off between the performance metrics can be obtained in a single optimization run.

Table 4 shows the prediction sensitivity, specificity, accuracy rate and feature count of the Pareto optimal solutions. These solutions are obtained from the same optimization run that gives the best solution in Table 2. It should be noted that these solutions are only a portion of the Pareto optimal solutions given by the output of Algorithm 1. These solutions are selected for analysis because they have highest prediction sensitivity at certain specificity rate, or vice versa.

Table 4 shows that the sensitivity of the solutions can be improved, but at the expense of lower specificity rate. For example, from solution 1 to 10, the sensitivity rate increases from 43.4 to 100%, while the specificity rate decreases from 96.2% to 26.4%. On the other hand, the accuracy rate increases from solution 1, peaks at 5, and then decreases until solution 10. Unlike sensitivity and specificity, the changes in accuracy rate does not show either increasing or decreasing trend. It is because the increment (or decrement) in sensitivity rate may outpace the decrement (or increment) in specificity rate.

The solution 5, which has the highest accuracy rate among the Pareto optimal solutions, is used for comparison in Table 2 and benchmarking in Table 5. Furthermore, the accuracy rate of solution 1, 2 and 10, which have the best sensitivity or specificity rate, are significant lower than 80% (that is achieved by other solutions in Table 4). It shows that even the sensitivity can be improved over certain threshold value, a significant trade-off is needed in specificity, or vice versa.

Table 3
HRV features that able to pass the two tailed U test up to 20% significance level for PAF onset.

Significance Level	HRV feature	
	typical method	Proposed method
5%	AR-LF, FFT-LF, LL-H2, LL-H3, LL-H4, HH-WCOB(f_{1m}), HH-WCOB(f_{2m})	SDRR, RMSSD, NN50, pNN50, SampEn, SD1, SD2, AR-LF, AR-HF, FFT-LF, FFT-HF, LL-M_{ave}, LL-P_{ave}, LL-H1, LL-H2, LL-H3, LL-H4, LH-P1, LH-M_{ave}, LH-P_{ave}, LL-H1, ROI-P1, ROI-P2, ROI-M_{ave}, ROI-P_{ave}, ROI-WCOB(f_{2m}), ROI-H1
10%	–	LL-P1, LH-P2
20%	FFT-HF, LL-M_{ave}, ROI-H2	HH-H1, ROI-WCOB(f_{1m}), ROI-H2

Table 4
Trade-off between sensitivity, specificity, accuracy rate and feature count.

Solution	SEN (%)	SPE (%)	ACC (%)	Feature count
1	43.4	96.2	69.8	4
2	45.3	94.3	69.8	4
3	75.5	92.5	84.0	7
4	83.0	90.6	86.8	6
5	86.8	88.7	87.7	7
6	90.6	81.1	85.9	5
7	92.5	79.3	85.9	5
8	94.3	71.7	83.0	3
9	98.1	60.4	79.3	5
10	100.0	26.4	63.2	5

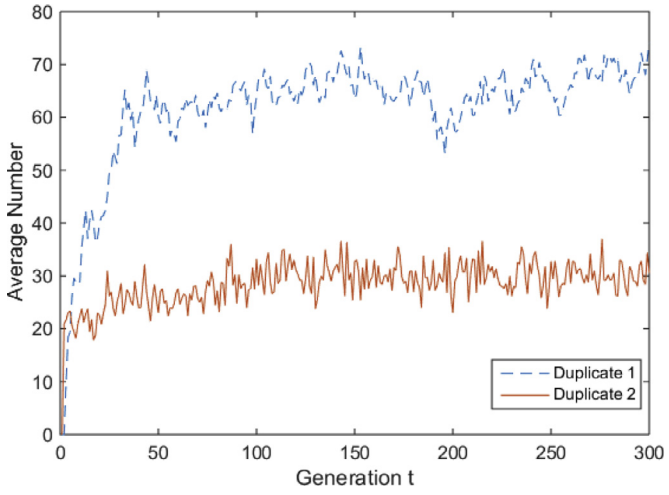

Fig. 5. Average number of the duplicate chromosomes for 5 optimization runs without handling at generation t.

4.3. Duplicate chromosomes

The results in this section show that high number of duplicates chromosomes are formed when NSGA-III is adapted for optimization in this paper. It wastes the computation resource and increases the optimization time when these duplicate chromosomes are re-evaluated. Fig. 5 shows the average number of duplicates for successive NSGA-III generations when they are not handled by the duplicate chromosome handling process. The average number is computed by summing and averaging the number of duplicates in each generation of 5 different optimization runs.

First of all, the line denoted by “Duplicate 1" is the number of duplicate chromosomes found by comparing each chromosome to other chromosomes within the population P_t only. For example, if they are 4 chromosomes that have same binary pattern within the population P_t, the count is increased by 3 $(4-1)$. Fig. 5 shows that the average number increases from 0 to around 60 within 50 generations, and then remains around that number (within range of ± 10) in the following generations. It represents approximately 28% of the chromosomes in population P_t that has total size of 212. Another line denoted by “Duplicate 2" is the number of duplicate chromosomes found by comparing each chromosome in population Q_t to all the chromosomes in population P_t after the LSO in fitness evaluation stage. The count is increased by 1 if there are two same chromosomes between two populations in a generation. Fig. 5 also shows that this number also increases rapidly within 100 generations.

With the proposed duplicate handling process, the number denoted by “Duplicate 1" becomes zero for all generations (zero duplicate), which means all the chromosomes in population P_t and P_{t+1} are unique. Although the number denoted by “Duplicate 2" remains same, the related duplicates are not re-evaluated. It is because they are formed during the normal genetic operation (line 8 of Algorithm 1), and they are identified before entering the fitness evaluation stage. Their fitness re-evaluations are skipped until modified by the duplicate handling process (line 19 of Algorithm 1).

4.4. Benchmarking

Table 5 shows the benchmarking results of our proposed method against the previous works that use less than 15 min of HRV signal for prediction. Our proposed method achieves prediction accuracy rate of 87.7% by using the shortest HRV signal length among all previous works: 5 min only. Furthermore, this accuracy rate outperforms all previous works that in Table 5. Boon et al. [8] achieved accuracy rate of 79.3% but used longest signal length with 15 min. Their accuracy rate was reduced to 68.9% when 10 min signal is used. Yang and Yin [12] employed 10 min signal and it achieved the lowest accuracy rate among previous works with 57%. Finally, Hickey and Heneghan [6] only achieved accuracy rate of 68% and 70% for 5 min and 10 min respectively.

Table 5
Benchmarking of proposed method against previous works using less than 15 min of signal length.

Previous Work	Signal length (Min)	Feature extraction method	Performance evaluation method	SEN (%)	SPE (%)	ACC (%)
Boon et al. 2016 [8]	15	HRV Features	10-fold CV	77.4	81.1	79.3
	10	HRV Features	10-fold CV	58.5	81.1	68.9
Yang and Yin, 2001 [12]	10	HRV based Footprint Analysis	Single Hold	–	–	57.0
Hickey and Heneghan, 2002 [6]	10	Spectral based HRV features	5-fold CV	53.0	80.0	70.0
	5	Spectral based HRV features	5-fold CV	51.0	79.0	68.0
Proposed Method	5	HRV Features	10-fold CV	86.8	88.7	87.7

Note: CV: Cross validation

Table 6
Prediction performance of previous works that used 30 min signal for prediction.

Previous Work	Feature extraction method	Performance evaluation method	SEN (%)	SPE (%)	ACC (%)
Zong et al. [7]	Number and timing of PACs	Single Hold	–	–	80.0
Hickey and Heneghan [6]	PACs detection and Spectral based HRV features	5-fold CV	79.0	72.0	75.0
Thong et al., [5]	PACs Analysis	Single Hold	89.00	91.00	90.0
Costin et al. [9]	HRV features and Morphological Variability of QRS complexes of ECG	Single Hold	89.3	89.4	89.4
Mohebbi and Ghassemian [10]	HRV features	Single Hold	96.2	93.1	94.5
Cheskonov [11]	HRV based spectral features	Single Hold	72.7	88.2	80.0
Lynn and Chiang [13]	HRV based Return Map and Poincare Plot features	Single Hold	–	–	64.0

In Table 6 , the prediction performances of the previous works using 30 min data (either ECG or HRV signal depending on feature extraction method) are summarized. When compared to these previous works, our proposed method outperforms all of them except [5,9,10]. Our method is 1.7% lower than Costin et al. [9], 2.3% lower than Trans et al. [5], 6.8% lower than the highest accuracy achieved by Mohebbi and Ghassemian [10]. Finally, the feature count of the proposed method is improved. The proposed method only uses 7 HRV features, which is significantly lower than 12 HRV features in best previous work [10].

Although our proposed method does not outperform them [5,9,10], they are several factors that need to be taken into consideration during the comparison. Firstly, their methods required longer duration of HRV signal for prediction. In contrast, the signal length is reduced by 83.33% (from 30 minto 5 min) in our method. Furthermore, they only employed single hold out validation method – it is well known that separating samples to single training and testing set can bias the performance of classifier. In this paper, 10-fold cross validation method, which is better method than single hold-out, is employed to estimate prediction accuracy of proposed work. The K-fold method is considered better than hold-out method because K-fold method can reduce the overfitting problem of trained classifier model [31], and more suitable to evaluate algorithm with small sample size of dataset [47]. Furthermore, Tran et al. [5] did not use HRV analysis for prediction, and they also needed to specify both recordings belong to same subjects before classifying the data. The proposed method does not require this extra step of specifying both recordings. Costin et al. [9] also used non-HRV analysis based features called morphological variability features that are extracted from QRS complexes in ECG signal. In contrast, the proposed method employs HRV analysis based features only.

5. Conclusion

In this paper, a paroxysmal atrial fibrillation (PAF) prediction method based on HRV analysis and NSGA-III is proposed. The proposed PAF onset prediction method achieves accuracy with 87.7% that outperforms all previous works that use less than 30 min signal for prediction. It is achieved even with HRV signal length being reduced from typical 30 min to 5 min (a reduction of 83%).

The improvements are achieved by proposing the NSGA-III based multi-objective optimization algorithm that can simultaneously optimizes all stages of the PAF onset prediction system. Furthermore, the trade-off between prediction sensitivity and specificity is analyzed, in which the results show that the sensitivity can be improved at the expense of lower specificity. Mann–Whitney U test is also used as the filter method to examine the statistical significance of the features before they can be selected to form the feature subset. Finally, a duplicate handling process is proposed to reduce the computation wastage due to the duplicate chromosomes.

As for limitation of this work, the proposed method still needs improvement in order to achieve same or higher accuracy rate than best work [10]. The proposed prediction method is also limited by a small sample size of real data from patients. Therefore, our result may not represent the true characteristic of the general PAF population. Furthermore, due to small sample size, the proposed algorithm is not evaluated with the testing data not included in the cross-validation process. As a result, evaluation with such testing set will be performed after we acquire larger sample size in future. Finally, the optimization process takes hours on single personal computer because multiple fitness evaluation processes are performed on each NSGA-III solution. However, it is not an issue in practice since the proposed optimization process is run single time only to find the best prediction model before real-world deployment.

Hence, the following future works are planned to improve our method. Firstly, more types of HRV feature extraction algorithm that are reviewed in [37] can be used for prediction. Furthermore, different filter methods such as analysis of variance, *t*-test and mutual information can be used to evaluate the features. The impact of different methods on the optimization performance can be analyzed and compared to U test in this paper. Finally, the duplication handling process can be improved so that it records all the chromosomes that have been evaluated in all previous generations. Then, the history record is used to determine the duplication status of a chromosome in the most recent generation. Such approach is employed in the simple GA [48] to completely eliminate the duplicate chromosome, but at the expense of higher memory and computation resource. Another future work is extending and applying the proposed optimization algorithm to other HRV research problems such as prediction of ventricular tachyarrhythmia onset and detection sleep apnea. It is applicable as long as the classification model is similar to baseline model in Fig. 2, with some changes in pre-processing stage.

Appendix A

Table A.1
List of abbreviation for HRV features.

HRV Feature	Abbreviation
Mean	Mean of HRV
SDRR	Standard deviation of HRV (SDRR)
RMSSD	Root mean square of successive difference intervals
NN50	Number of adjacent RR intervals differing by more than 50 ms
pNN50	Sum of NN50 divided by the total number of all RR intervals (pNN50).
RRTri	RR Triangular Index
SampEn	Sample Entropy Feature
SD1	Standard deviation of the distances of RR intervals from line-of-identity($y = x$)
SD2	Standard deviation of the distances of RR intervals from $y = -x + 2RR_m$
SD1/SD2	Ratio of SD1 to SD2
AR-LF	Low frequency component of AR spectrum
AR-HF	High frequency component of AR spectrum
AR-LF/HF	Ratio of LF to HF for AR
FFT-LF	Low frequency component of FFT spectrum
FFT-HF	High frequency component of FFT spectrum
FFT-LF/HF	Ratio of LF to HF for FFT
M_{ave}	Mean of sum of magnitude of LL, LH and HH region
P_{ave}	Mean of sum of squared magnitude of LL, LH and HH region
P1	Normalized bispectral entropy of LL, LH and HH region
P2	Normalized bispectral squared entropy of LL, LH, HH and ROI region
H1	Sum of logarithmic amplitudes in bispectrum of LL, LH, HH and ROI region.
H2	Sum of logarithmic amplitudes of diagonal elements in bispectrum of LL, HH and ROI region.
H3	First-order spectral moment of the amplitudes of diagonal elements in the bipsectrum of LL, HH and ROI region.
LLH4	Second-order spectral moment of the amplitudes of diagonal elements in the bispectrum of LL, HH and ROI region.
WCOB (f_{1m}, f_{2m})	Weighted center in bispectrum of LL, LH, HH and ROI region.

Table A.2
Pre-defined and optimized parameters and settings.

Feature extraction algorithm	Algorithm parameters	Pre-defined value	Optimized value
1 Time domain	• Heart rate correction enabling	• Enabled	• Disabled
2 RR Triangular Index	• Heart rate correction enabling	• Enabled	• Disabled
3 Poincare plot	• Heart rate correction enabling	• Enabled	• Disabled
4 Sample entropy	• Heart rate correction enabling	• Enabled	• Disabled
	• Embedding dimension	• 2	• 2
	• Tolerance Distance	• 0.2 of standard deviation of RR interval	• 0.38 of standard deviation of RR interval
5 Auto Regressive (AR) analysis	• Heart rate correction enabling	• Enabled	• Disabled
	• Heart rate detrending enabling	• Enabled	• Enabled
	• Interpolation method	• Cubic	• Linear
	• Resampling frequency	• 4 Hz	• 7 Hz
	• Order of model	• 16	• 16
	• Normalization of AR spectrum	• Normalized	• Not normalized.
6 Fast fourier transform (FFT)	• Heart rate correction enabling	• Enabled	• Disabled
	• Heart rate detrending enabling	• Enabled	• Enabled
	• Interpolation method	• Cubic	• Linear
	• Resampling frequency	• 4 Hz	• 4 Hz
	• Temporal smoothing window function	• Rectangular Window	• Rectangular window
	• FFT segmentation	• Normal FFT	• Normal FFT
	• Normalization of FFT spectrum	• Normalized	• Not normalized
7 Higher order spectral (HOS) analysis	• Heart rate correction enabling	• Enabled	• Enabled
	• Heart rate detrending enabling	• Enabled	• Enabled
	• Interpolation method	• Cubic	• Cubic spline
	• Resampling frequency	• 4 Hz	• 7 Hz
	• Temporal smoothing window function	• Rectangular	• Blackman
	• Segmentation size	• 512	• 256
	• Zero-padding size for the segmented data.	• 512	• 512
	• Overlapping of the segmented data	• 50%	• 0%

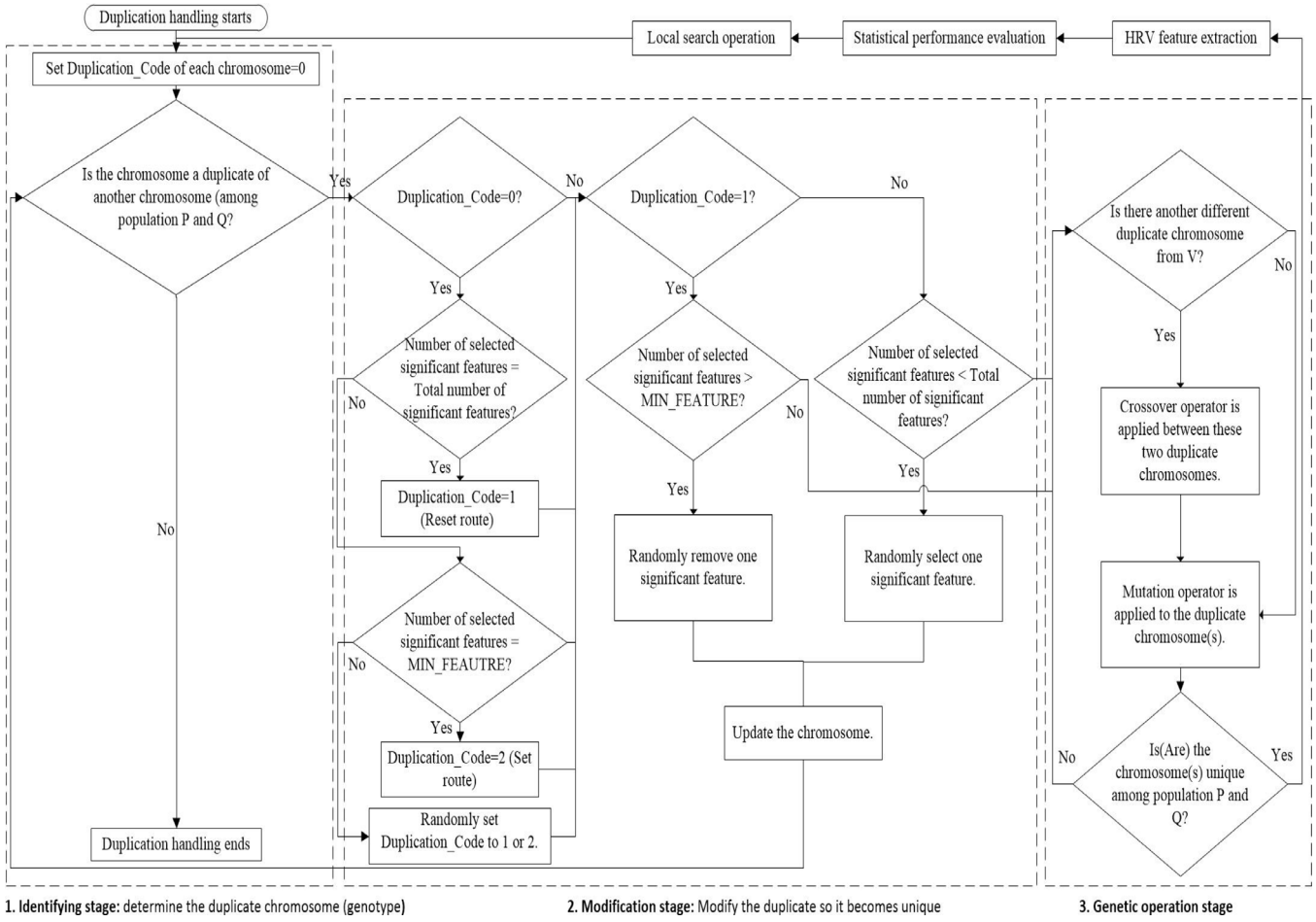

Fig. A.1. Flow chart of the proposed duplicate handling process.

References

[1] A.J. Camm, P. Kirchhof, G.Y. Lip, U. Schotten, I. Savelieva, S. Ernst, et al., Guidelines for the management of atrial fibrillation the task force for the management of atrial fibrillation of the european society of cardiology (ESC), Euro. Heart J. (2010) 278.

[2] A.J. Manolis, E.A. Rosei, A. Coca, R. Cifkova, S.E. Erdine, S. Kjeldsen, et al., Hypertension and atrial fibrillation: diagnostic approach, prevention and treatment. Position paper of the working group 'Hypertension Arrhythmias and Thrombosis' of the European society of hypertension, J. Hypertension. 30 (2) (2012) 239–252.

[3] E.N. Prystowsky, Management of atrial fibrillation: therapeutic options and clinical decisions, Am. J. Cardiol. 85 (10) (2000) 3–11 Supplement 1.

[4] A. Prakash, S. Saksena, P.M. Hill, R.B. Krol, A.N. Munsif, I. Giorgberidze, et al., Acute effects of dual-site right atrial pacing in patients with spontaneous and inducible atrial flutter and fibrillation, J. Am. College Cardiol. 29 (5) (1997) 1007–1014.

[5] T. Tran, J. McNames, M. Aboy, B. Goldstein, Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes, IEEE Trans. Biomed. Eng. 51 (4) (2004) 561–569.

[6] B. Hickey, C. Heneghan, Screening for paroxysmal atrial fibrillation using atrial premature contractions and spectral measures, Comput. Cardiol. (September) (2002) 22–25.

[7] W. Zong, R. Mukkamala, R.G. Mark, A methodology for predicting paroxysmal atrial fibrillation based on ECG arrhythmia feature analysis, Comput. Cardiol. (2001).

[8] K.H. Boon, M. Khalil-Hani, M.B. Malarvili, C.W. Sia, Paroxysmal atrial fibrillation prediction method with shorter HRV sequences, Comput. Methods Programs Biomed. 134 (2016) 187–196.

[9] H. Costin, C. Rotariu, A. Păsărică, Atrial fibrillation onset prediction using variability of ECG signals, in: Advanced Topics in Electrical Engineering (ATEE), 8th International Symposium, 2013, pp. 23–25.

[10] M. Mohebbi, H. Ghassemian, Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal, Comput.Methods Programs Biomed. 105 (1) (2012) 40–49.

[11] Y.V. Chesnokov, Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods, Artif. Intell. Med. 43 (2) (2008) 151–165.

[12] A.C.C. Yang, H.W. Yin, Prediction of paroxysmal atrial fibrillation by footprint analysis, Comput. Cardiol. (2001).

[13] K.S. Lynn, H.D. Chiang, A two-stage solution algorithm for paroxysmal atrial fibrillation prediction, Comput. Cardiol. (2001).

[14] A. Narin, Y. Isler, M. Ozer, Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance, Comput. Biol. Med. 45 (2014) 72–79.

[15] F. Lucena, A.K. Barros, N. Ohnishi, The performance of short-term heart rate variability in the detection of congestive heart failure, BioMed. Res. Int. (2016) 2016.

[16] H.F. Jelinek, J.H. Abawajy, D.J. Cornforth, A. Kowalczyk, M. Negnevitsky, M.U. Chowdhury, et al., Multi-layer attribute selection and classification algorithm for the diagnosis of cardiac autonomic neuropathy based on HRV attributes, AIMS Med. Sci. 2 (4) (2015) 396–409.

[17] Y. Isler, M. Kuntalp, Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure, Comput. Biol. Med. 37 (10) (2007) 1502–1510.

[18] H. Ocak, A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being, J. Med. Syst. 37 (2) (2013) 9913.

[19] S.N. Yu, M.Y. Lee, Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability, Comput. Biol. Med. 42 (8) (2012) 816–825.

[20] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, A simultaneous feature adaptation and feature selection method for content-based image retrieval systems, Know-Based Syst. 39 (2013) 85–94.

[21] F.H. Wians, Clinical laboratory tests: which, why, and what do the results mean? Lab. Med. 40 (2) (2009) 105–113.

[22] B. Xie, H. Minn, Real-time sleep apnea detection by classifier combination, IEEE Trans. Inf. Technol. Biomed. 16 (3) (2012) 469–477.

[23] B.L. Koley, D. Dey, Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices, IEEE Trans. Biomed. Eng. 60 (12) (2013) 3354–3363.

[24] A. Konak, D.W. Coit, A.E. Smith, Multi-objective optimization using genetic algorithms: A tutorial, Realib. Eng. Syst. Safety 91 (9) (2006) 992–1007.

[25] K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints, IEEE Trans. Evolution. Comput. 18 (4) (2014) 577–601.
[26] X. Bi, C. Wang, An improved NSGA-III algorithm based on objective space decomposition for many-objective optimization, Soft Comput. (2016) 1–28.
[27] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[28] T.A. El-Mihoub, A.A. Hopgood, L. Nolle, A. Battersby, Hybrid genetic algorithms: a review, Eng. Lett. 13 (2) (2006) 124–137.
[29] J. Huang, Y. Cai, X. Xu, A hybrid genetic algorithm for feature selection wrapper based on mutual information, Pattern Recognit.Lett. 28 (13) (2007) 1825–1844.
[30] J. Huang, P. Rong, A hybrid genetic algorithm for feature selection based on mutual information, in: F. Emmert-Streib, M. Dehmer (Eds.), Information Theory and Statistical Learning, Springer US, Boston, MA, 2009, pp. 125–152.
[31] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27.
[32] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215-e20.
[33] P.S. Hamilton, W.J. Tompkins, Quantitative investigation of QRS detection rules using the MIT/BIH Arrhythmia database, IEEE Trans. Biomed. Eng. 33 (12) (1986) 1157–1165 BME.
[34] J. McNames, T. Thong, M. Aboy, Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. Engineering in Medicine and Biology Society, in: Annual International Conference of the IEEE, 2004, pp. 1–5.
[35] M.P. Tarvainen, P.O. Ranta-aho, P.A. Karjalainen, An advanced detrending method with application to HRV analysis, IEEE Trans. Biomed.Eng. 49 (2) (2002) 172–175.
[36] G.D. Clifford, L. Tarassenko, Quantifying errors in spectral estimates of HRV due to beat replacement and resampling, IEEE Trans. Biomed. Eng. 52 (4) (2005) 630–638.
[37] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C.M. Lim, J.S. Suri, Heart rate variability: a review, Medical Biol.Eng. Comput. 44 (12) (2006) 1031–1051.
[38] B. Anita, S. Fernando Soares, R. Ana Paula, L. Argentina, A study on the optimum order of autoregressive models for heart rate variability, Physiol. Measur. 23 (2) (2002) 325.
[39] C.L. Nikias, M.R. Raghuveer, Bispectrum estimation: A digital signal processing framework, in: Proceedings of the IEEE, 75, 1987, pp. 869–891.
[40] S.-M. Zhou, J.Q. Gan, F. Sepulveda, Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface, Inf. Sci. 178 (6) (2008) 1629–1640.
[41] S.M. Pincus, A.L. Goldberger, Physiological time-series analysis: what does regularity quantify? Am. J. Physiol. 266 (4) (1994) H1643-H56.
[42] I. Das, J. Dennis, Normal-boundary intersection: an alternate method for generating pareto optimal points in multicriteria optimization problems, DTIC Document (1996).
[43] Y. Yuan, H. Xu, B. Wang, An improved NSGA-III procedure for evolutionary many-objective optimization, in: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation, Vancouver, BC, Canada, ACM, 2014, pp. 661–668. 2598342.
[44] D.E. Lake, J.S. Richman, M.P. Griffin, J.R. Moorman, Sample entropy analysis of neonatal heart rate variability, Am. J. Physiol. 283 (3) (2002) R789-R97.
[45] N.P. Pérez, M.A. Guevara López, A. Silva, I. Ramos, Improving the Mann–Whitney statistical test for feature selection: an approach in breast cancer diagnosis on mammography, Artif. Intell. Med. 63 (1) (2015) 19–31.
[46] D. Saroj, A non-revisiting genetic algorithm with adaptive mutation for function optimization, in: N. Meghanathan, N. Chaki, D. Nagamalai (Eds.), Advances in Computer Science and Information Technology Computer Science and Engineering: Second International Conference, CCSIT 2012, Springer Berlin Heidelberg, 2012, pp. 288–297.
[47] Y. Bengio, Y. Grandvalet, No unbiased estimator of the variance of k-fold cross–validation, J. Mach. Learn Res. 5 (2004) 1089–1105.
[48] Y. Lou, S.Y. Yuen, Non-revisiting genetic algorithm with adaptive mutation using constant memory, Memetic Comput. (2016) 1–22.