

Diagnosis of patients with chronic heart failure implementing wavelet transform and machine learning techniques

Carlos Arizmendi¹, Jhon Reinemer¹, Hernando Gonzalez¹, and Beatriz F. Giraldo^{2,3}

¹Department of Mechatronics Engineering, Faculty of Engineering, Autonomous University of Bucaramanga, Bucaramanga, Colombia

²Department of Automatic Control (ESAIL), Barcelona East School of Engineering (EEBE), Universitat Politècnica de Catalunya, Barcelona, Spain

³Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, and CIBER de Bioingeniería, Biomateriales y Nanomedicina, Barcelona, Spain

Article Info

Article history:

Received Jul 3, 2022

Revised Apr 10, 2024

Accepted Apr 16, 2024

Keywords:

Chronic heart failure

Discrete wavelet transforms

Feature selection

K-nearest neighbors

Support vector machines

ABSTRACT

Chronic heart failure (CHF) is a significant public health concern due to its increasing prevalence, high number of hospital admissions, and associated mortality. Its prevalence is progressively increasing due to the aging of the population and the decrease in mortality from acute myocardial infarction, among other medical advancements. Consequently, the incidence of CHF predominantly affects older age groups, doubling its prevalence every decade, becoming one of the main causes of mortality in patients older than 65 years. The main objective of this study is to apply machine learning based techniques to determine the best models to classify patients with chronic heart failure through their respiratory pattern. These patterns have been characterized from time series such as inspiratory and expiratory times, breathing duration, and tidal volume obtained from the respiratory flow signal. Based on the behavior of the respiratory pattern, CHF patients were classified into patients with non-periodic breathing, with periodic breathing, and with Cheyne-Stokes respiration (CSR). Time-frequency and statistical techniques have been implemented to analyze these features, and then various classification methods have been applied to define the optimal model with the best accuracy rates. These models could help to better understand the evolution of this disease and in early diagnosis.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Carlos Arizmendi

Department of Mechatronics Engineering, Faculty of Engineering, Autonomous University of Bucaramanga
AV 42#48-11, Bucaramanga 680003, Colombia

Email: carizmendi@unab.edu.co

1. INTRODUCTION

Chronic heart failure (CHF) is a clinical syndrome that results from structural or functional damage to the heart, preventing its proper functioning as a pump that propels blood through the circulatory system to provide the body's metabolic needs. Its main manifestations are pulmonary edema (fluid accumulation in the lungs), dyspnea (breathing difficulty), decreased exercise tolerance and fatigue [1], [2]. CHF is associated with some diseases, including coronary ischemic heart disease, hypertensive heart disease, valvopathies, infective endocarditis, rheumatic valve disease, diabetes mellitus, myocardial infarction. It is a serious public health problem due to its growing prevalence, the highest number of hospital admissions and associated mortality [3], [4]. In addition, it constitutes a growing problem related to the aging of the population, and problems such as acute myocardial infarction double their prevalence every decade. Many studies are being carried out on CHF since, if the current trend continues, in a few years it could be one of the main causes of mortality in patients older than 65 years [5]–[7].

CHF patients often develop an irregular respiratory pattern, characterized by cyclic fluctuations in breathing drive, a sign of respiratory control system instability. Depending on the patient's level of severity these fluctuations are marked by alternating periods of fast and slow respirations, called periodic breathing (PB). Cheyne-Stokes respiration (CSR) is a more severe form of the PB pattern with alternating episodes of apnea and hypopnea [8], [9]. PB has a prevalence of up to 70% in patients with CHF and is associated with higher mortality, especially in patients with CSR [10], [11]. In general, three out of four patients admitted for CHF will not live for more than five years, which is comparable to the most aggressive neo-plasms between 30% and 40% of patients will die during the first year after diagnosis [12], [13]. Therefore, it is important to establish an accurate risk stratification of CHF patients to optimize the allocation of limited resources and contribute to the clinical decision.

In the initial stages of CHF, a reduction in the elasticity of the lungs is observed, which increases their work. Occasionally, dyspnea occurs at night and the patient awakens abruptly with a choking sensation or paroxysmal nocturnal dyspnea [14]. Patients are classified based on the New York Heart Association Index, which assesses the patient's physical activity [15], [16]. One way to characterize the respiratory pattern is through the analysis of its respiratory signals. Using the wavelet transform method and some classification techniques, the best parameters to characterize the respiratory pattern were selected [17]. The main objective of this study is to analyze parameters extracted from the respiratory flow signal using machine learning techniques, to classify CHF patients with periodic or non-periodic breathing. The time series of these signals, such as inspiration and expiration time, breathing duration, tidal volume, among others, can define the dynamics of the respiratory system and characterize different conditions of the CHF patients.

2. MATERIAL AND METHOD

2.1. Datasets

Respiratory flow signals were recorded from 27 elderly patients who were admitted to the short stay unit at Santa Creu i Sant Pau Hospital in Barcelona, Spain. The study was conducted in accordance with a protocol that was previously approved by the local ethics committee. A pneumotachograph, consisting of a Datex-Ohmeda monitor with a Validyne model MP45-1-871 variable reluctance transducer, was used to acquire the respiratory flow signal. The signals were recorded at a sampling rate of 250 Hz and a resolution of 12 bits for 15 minutes.

The patients were classified into three groups based on clinical criteria and respiratory flow signal behavior: 19 patients with non-periodic breathing (CHF-nPB), 5 patients with periodic breathing (CHF-PB), and 3 patients with CSR, a pattern characterized by cyclical episodes of hyperventilation and apneas or hypopneas (CHF-CSR). Additionally, a control group of 35 healthy subjects was included. Figure 1(a) to (d) shows the respiratory flow signal of CHF-CSR, CHF-PB, CHF-nPB patients, and a control subject [18].

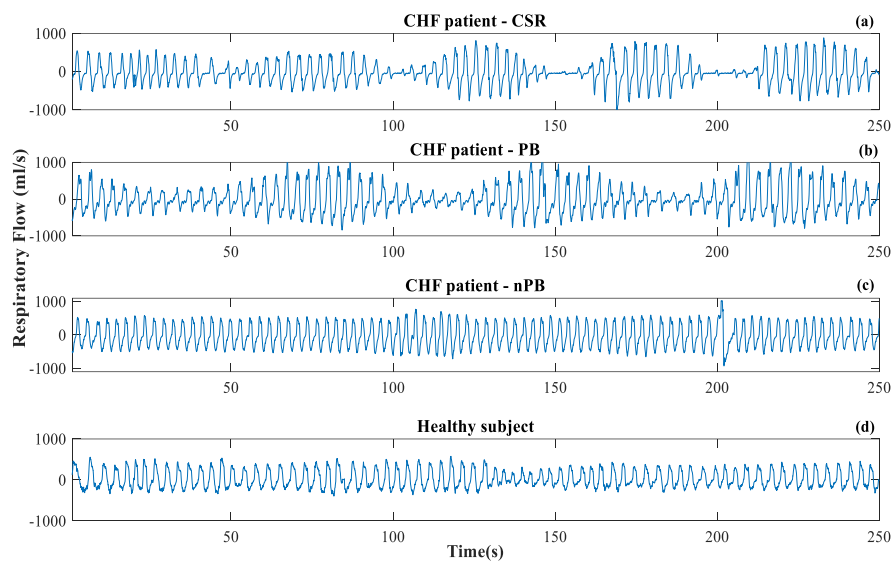


Figure 1. Excerpt of the respiratory flow signal from CHF patients with (a) CHF-CSR patient, (b) CHF-PB patient, (c) CHF-nPB patient, and (d) respiratory flow signal from a healthy subject

2.2. Signal processing

Respiratory flow signals were preprocessed using in-house tools to reduce artifacts, noise, and outliers. Afterward, the following time series were extracted: inspiratory time (TI), *Exp_iratory* time (TE), breathing duration ($TTot = TI + TE$), tidal volume (VT), respiratory rate (Rr) per minute, inspiratory fraction as the ratio of inspiratory time to breathing duration ($TT = TI/TTot$), mean inspiratory flow as the ratio of tidal volume to inspiratory time ($VI = VT/TTot$), and respiratory frequency- tidal volume ratio ($FV = Rr/VT$).

2.3. Wavelet transform and feature selection techniques

The continuous wavelet transforms (CWT) of a signal $x(t)$ with mother wavelet $\psi(\cdot)$ is defined as (1):

$$W(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-\tau}{s}\right) dt, \quad (1)$$

where the transformed signal $W(\tau, s)$ is a function of the translation parameter τ and the scale s . The signal energy is normalized to each scale by dividing the wavelet coefficients by $1/\sqrt{|s|}$. The original signal can be reconstructed with the inverse CWT, defined by (2) [19].

$$x(t) = \frac{1}{C_{\psi}^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(\tau, s) \frac{1}{s^2} \psi\left(\frac{t-\tau}{s}\right) d\tau ds, \quad \text{with } C_{\psi} = \int_0^{\infty} \frac{|\Psi(f)|^2}{f} df < \infty. \quad (2)$$

The discrete wavelet transform (DWT) allows multiresolution analysis by applying a bank of high-pass filters, where each filter represents a level of decomposition. A cascade filter bank of eight-order is applied, a low-pass $L(z)$ and high-pass $H(z)$ filters, followed by a resampling step. Next, the original signal is reconstructed from a new bank of synthesis filters, introducing zero values between the sampled signal calculated with a new high-pass filter $H'(z)$ and a new low-pass filter $L'(z)$, respectively [19], [20]. Next, the original signal is reconstructed from a bank of synthesis filters, introducing zero values between the sampled signal calculated with a new high-pass filter $H'(z)$ and a new low-pass filter $L'(z)$, respectively [19], [20]. Afterward, the forward selection method [21], genetic algorithms [22], [23], and moving window [24], [25] with variance analysis are implemented to obtain the parameters that characterize the respiratory pattern.

2.4. Classification techniques

Classification methods make it possible to determine the characteristics that differentiate the study data into subsets called classes. Based on the properties of these subsets, they are compared within each classification model. Based on the properties of these subsets, they are compared within each classification model. In this paper, the support vector machine (SVM) method [26], [27] and K-nearest neighbor [28], [29] are proposed and then a feature selection technique is applied [30].

3. DATA ANALYSYS

A statistical analysis of the patient's data is performed, after preprocessing the time series that describes their respiratory pattern, to rule out outlier values, and information unrelated to the physiological process, and normalize them by standard scaling. Then, the principal component analysis (PCA) is applied to analyze the variability of these times series (i.e., TI , TE , $TTot$, VT , TT , VI , FV). Figure 2(a) illustrates the multidimensional projection of the means of these seven-time series for each patient, considering the first ten principal components, with a cumulative variance of 93.1%. Figure 2(b) represents a three-dimensional projection of the first three PCs, considering all study groups (CHF patients with non-periodic breathing, periodic breathing and Cheyne-Stokes respiration patterns, and healthy subjects) representing a cumulative variance of 58.7% [31].

The CHF patients studied were classified into three groups: non-periodic breathing (CHF-nPB), periodic breathing (CHF-PB), and Cheyne-Stokes respiration (CHF-CSR). Due to the small number of patients in PB and CSR groups, they were regrouped into one, periodic breathing and Cheyne-Stokes respiration (CHF-PBCSR). For the diagnosis-oriented classification, four clinical interest groups were defined: i) G1 group, healthy subjects versus CHF patients; ii) G2 group, healthy subjects versus CHF-PBCSR patients; iii) G3 group, healthy subjects versus CHF-nPB patients; and iv) G4 group, CHF-PBCSR versus CHF-nPB patients. To evaluate the behavior of these groups, combined strategies of signal processing techniques, dimensionality reduction and classification methods were analyzed considering a set of 10 experiments (Exp_). Figure 3 is a schematic representation of the applied process.

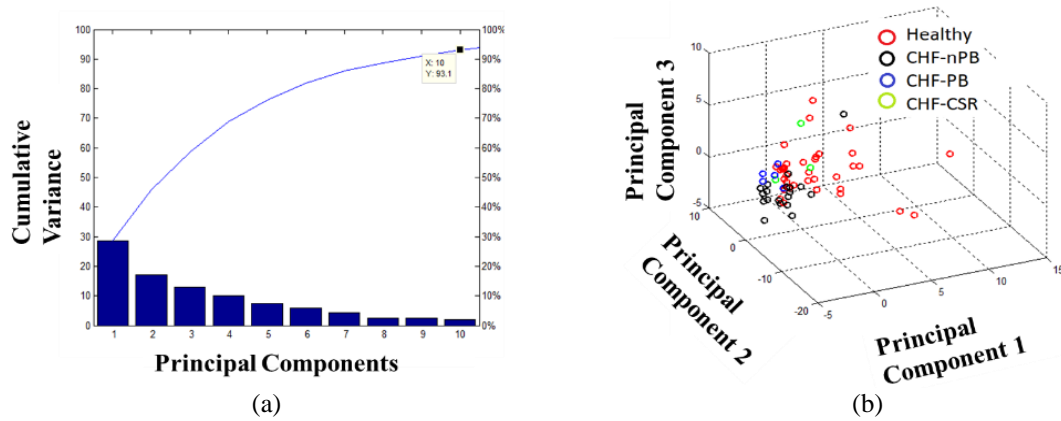


Figure 2. Representation of the principal component’s behavior for the mean values of the time series: (a) the cumulative variance collected through the first ten principal components and (b) the contribution of the first three principal components associated with the classification of patients with CHF (Cheyene-Stokes respiration, periodic breathing and non-periodic breathing) and healthy subjects

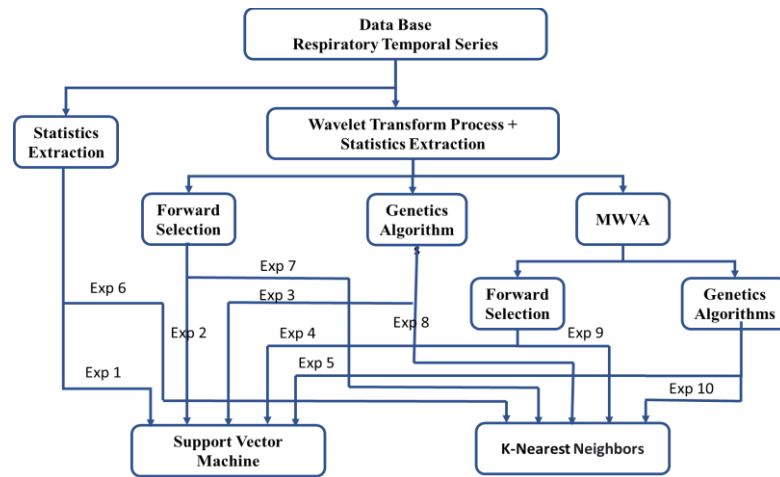


Figure 3. Explanatory schematic of the set of 10 experiments

3.1. Statistical parameter extraction

From the time series that characterize the respiratory pattern, TI, TE, TTot, VT, TT, VI and FV, the following statistics were calculated: mean (\bar{X}), standard deviation (SD), kurtosis (K), interquartile range (IQR) and variance (V). A total of 35 parameters were obtained from each patient (e.g., $\bar{X}(T_I)$, SD (TI), K(TI), IQR(TI), V(TI)). Then, these parameters were included in the Exp_1 and Exp_2 experiments to classify the patients and healthy subjects, according to the schema presented in Figure 3.

3.2. Wavelet transform process and statistics extraction

The DWT is implemented to obtain information about the approximation coefficients and signal detail for each time series. The mother wavelets to be computed are: 'haar', 'db2', 'db3', 'db4', 'db5', 'db6', 'db7', 'db8', 'db9', 'db10', 'bior1.3', 'bior1.5', 'bior2.2', 'bior2.4', 'bior2.6', 'bior2.8', 'bior3.1', 'bior3.3', 'bior3.5', 'bior3.7', 'bior3.9', 'bior4.4', 'bior5.5', 'bior6.8', 'coif1', 'coif2', 'coif3', 'coif4', 'coif5', 'sym2', 'sym3', 'sym4', 'sym5', 'sym6', 'sym7', 'sym8' and 'sym9'. To determine the optimal mother wavelet (A_i : approximates coefficient and D_i : details coefficients), the Q index is calculated according to (3). Each signal is decomposed and reconstructed using five decomposition levels, calculating the average mean square error (\overline{MSE}), the average signal-to-noise ratio (\overline{SNR}), and the numbers of average decay coefficients (\overline{NDC}), being $x(i)$ the original signal and $x'(i)$ the reconstructed signal [17]. Table 1 presents the mother wavelets that obtained the highest values of the Q index for each time series, with the *haar* being the optimal mother wavelet for this decomposition analysis.

$$\overline{SNR} = 10 \log \frac{\sum_{i=1}^N [x(i)]^2}{\sum_{i=1}^N [x(i) - x'(i)]^2} \quad Q = \frac{\overline{SNR}}{\overline{MSE} + \overline{NDC}} \quad (3)$$

Table 1. Best values of the Q index for each time series

Series	Mother Wavelet	\overline{MSE}	\overline{SNR}	Q
FV	<i>haar</i>	2,373E-32	278.462	0.775
TE	<i>haar</i>	2,543e-30	278.073	0.773
TI	<i>haar</i>	1,764e-30	278.139	0.772
Ttot	<i>haar</i>	9,793e-32	277.638	0.772
TT	<i>haar</i>	7,889e-30	278.051	0.773
VI	<i>haar</i>	5,959e-27	278.406	0.774
VT	<i>haar</i>	9,589e-27	278.412	0.774

After selecting the optimal wavelet, a signal was obtained for each of the five decomposition levels and for each of the approximation (A_i) and detail coefficients (D_i) (e.g., T_{1-A_i} , T_{1-D_i} , approximation and detail coefficients of the inspiratory time serie (TI), for the i -decomposition level), resulting in 70 signals for each patient [11]. Next, a preselection of variables is performed for each of the four groups of clinical interest, calculating the mean p-value of the Mann-Whitney test for all-time series, for coefficients of approximation and detail, at each level of decomposition. Afterwards, the coefficients with p-value ≤ 0.05 were selected for each group of the study and for each time series as shown in Table 2.

Next, the statistics \bar{X} , SD , K , IQR and V are calculated for the preselected signals in Table 2 (e.g., G4 serie VT: $\bar{X}(V_T-A_1)$, $SD(V_T-A_1)$, $IQR(V_T-A_1)$, $K(V_T-A_1)$, $V(V_T-A_1)$), obtaining a total of 145 variables for G1, 95 variables for G2, 140 variables for G3, and 135 variables for G4. Figure 4 illustrates the multidimensional projections of the preselected variables considering the first three principal components for each group when comparing in Figure 4(a) G1 group, healthy subjects versus CHF patients, Figure 4(b) G2 group, healthy subjects versus CHF-PBCSR patients, Figure 4(c) G3 group, healthy subjects versus CHF-nPB patients, and Figure 4(d) G4 group, CHF-PBCSR versus CHF-nPB patients. According to the results, the PCA contributions in each case are: G1 = 54.28%, G2 = 67.76%, G3 = 56.17%, and G4 = 60.74% of the information, respectively.

Table 2. Coefficients of approximation and detail of the decomposition levels whose p-value ≤ 0.05 in each time series for each classification group

Series	G1	G2	G3	G4
FV	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁
TE	A ₁ , A ₂ , A ₃	A ₁	A ₁ , A ₂	A ₁ , A ₂ , A ₃ , A ₄ , A ₅
TI	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅
TTot	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	-----	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅
TT	A ₁ , A ₂	A ₁ , A ₂ , A ₃	A ₁	A ₁ , A ₂ , A ₃ , A ₄ , A ₅
VI	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	-----	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅
VT	A ₁ , A ₂ , A ₃ , A ₄	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁ , A ₂ , A ₃ , A ₄ , A ₅	A ₁

3.3. Dimensionality reduction and classification

Once the variables have been preselected and their statistics calculated, the dimensionality reduction and classification process are implemented with the experiments *Exp_2*, *Exp_3*, *Exp_4*, *Exp_5*, *Exp_7*, *Exp_8*, *Exp_9* and *Exp_10* in Figure 3. To evaluate these results, maintaining the independence between the training-test data, a 4-fold cross-validation with 150 iterations is implemented. The value of average accuracy of the test results is calculated and used in the validation group to tune the parameters in the dimensionality reduction process with their respective metric. Then, the best fit result was used to calculate the balanced accuracy.

Applying the forward selection method, the stopping criterion used was when the variables reached the maximum accuracy of the validation group in its respective classification method, after evaluating all the variables of the Experiment. Using the moving window and variance analysis (MWVA), the most relevant set of variables in the optimal window was selected according to the elbow method [18]. For the genetic algorithm method (GA) the following characteristics were defined: generations=100, population size=10, selection function=spinner, crossover function=scattered, crossover fraction=0.8, and mutation function=mutation adapts feasible. With the SVM classification methods, the kernel functions used were linear, polynomial order 3, polynomial order 2 and radial basic function (RBF). The adjustable parameters of the SVM were calculated by implementing the Bayesian optimization [32], [33]. For the k-NN method, the Euclidean distance was implemented, with the maximum K being twice the number of patients in the minority class.

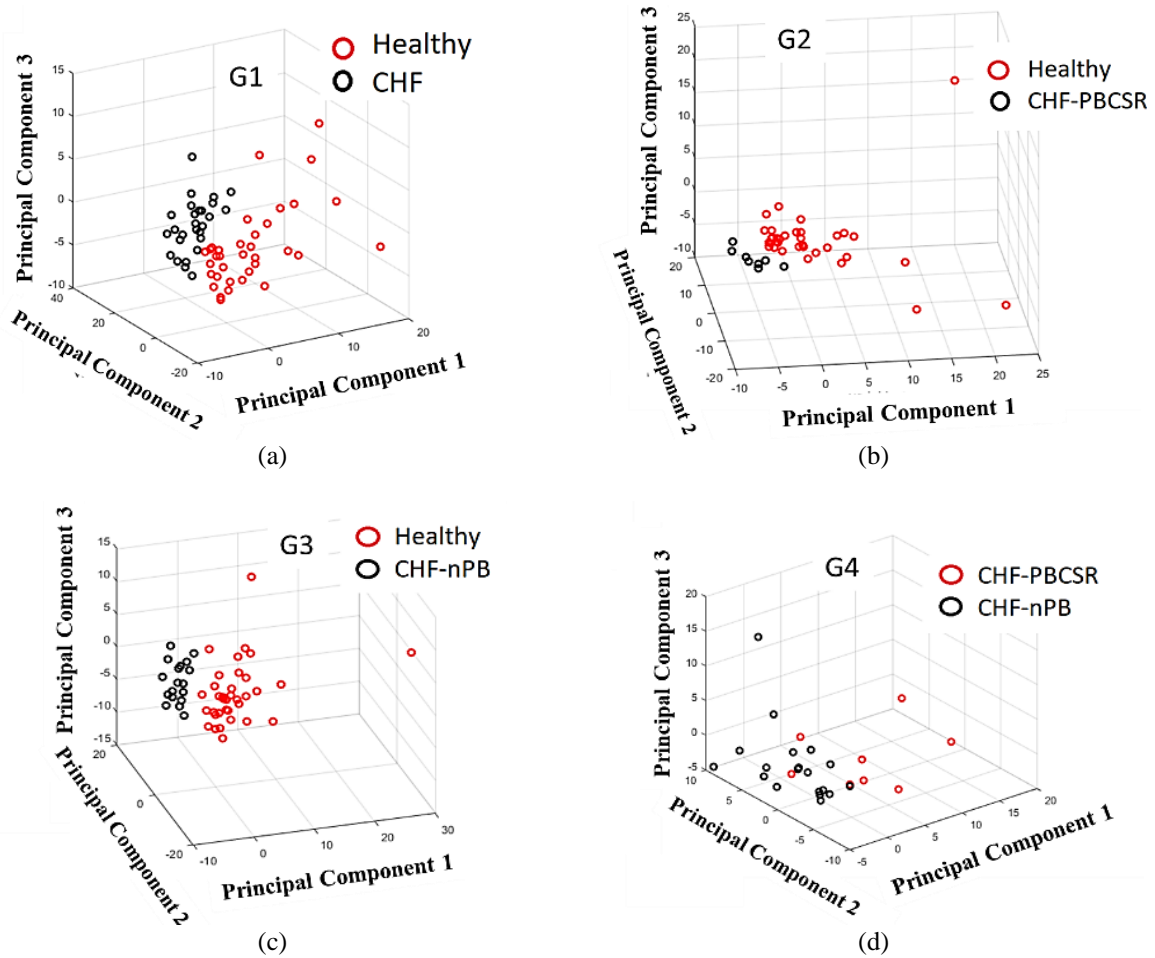


Figure 4. Contributions of the first three principal components to classify (a) G1 group, healthy subjects versus CHF patients, (b) G2 group, healthy subjects versus CHF-PBCSR patients, (c) G3 group, healthy subjects versus CHF-nPB patients, and (d) G4 group, CHF-PBCSR versus CHF-nPB patients

4. RESULTS AND DISCUSSION

Tables 3 and 4 present the best results of all the experiments, in terms of accuracy and their balanced accuracy of the comparison groups G1, G2, G3 and G4, respectively. When comparing healthy subjects versus chronic heart failure patients (G1) the balanced accuracies are greater than 91.9% in all cases. Analyzing the performance of healthy subjects versus patients with periodic breathing and Cheyne-Stokes respiration (G2), the best results were obtained with *Exp_3* and *Exp_7*, while the lowest value was 64.5%, obtained using the forward selection method and classification with SVM. Table 5 presents the best accuracy results of each group with the most relevant parameter of each system. Of all cases, the best result was obtained with *Exp_9*, classified with SVM, and the optimal window of 1 in the MWVA calculation.

Table 3. Accuracy and balanced accuracy of G1 and G2 groups

Experiment	G1		G2	
	Accuracy (%)	Balanced accuracy (%)	Accuracy (%)	Balanced accuracy (%)
Exp1	96.25 ± 2.18	96.42	94.86 ± 3.66	92.62
Exp2	92.25 ± 5.29	91.90	77.95 ± 16.48	64.49
Exp3	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp4	100.00 ± 0.00	100.00	98.63 ± 1.36	97.54
Exp5	99.89 ± 0.11	99.89	99.90 ± 0.10	99.85
Exp6	98.97 ± 1.03	98.90	84.97 ± 14.42	73.53
Exp7	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp8	100.00 ± 0.00	100.00	85.43 ± 12.26	74.73
Exp9	100.00 ± 0.00	100.00	82.06 ± 5.48	79.45
Exp10	100.00 ± 0.00	100.00	88.57 ± 10.73	81.84

Table 4. Accuracy and balanced accuracy of G3 and G4 groups

Experiment	G3		G4	
	Accuracy (%)	Balanced accuracy (%)	Accuracy (%)	Balanced accuracy (%)
Exp1	95.41 ± 2.25	95.72	96.81 ± 1.48	96.38
Exp2	91.45 ± 7.30	89.71	95.29 ± 4.70	90.66
Exp3	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp4	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp5	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp6	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp7	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp8	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp9	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00
Exp10	100.00 ± 0.00	100.00	100.00 ± 0.00	100.00

Table 5. Accuracy, balanced accuracy, and the most relevant variables of each group

Groups	Accuracy (%)	Balanced accuracy (%)	Variables
G1	100.00 ± 0.00	100.00	$\bar{X}(VI_{A_5})$ and $\bar{X}(TTot_{A_3})$
G2	100.00 ± 0.00	100.00	IQR(FV_{A_1}), IQR(TI_{A_1}) IQR(VT_{A_1}).
G3	100.00 ± 0.00	100.00	IQR(TI_{A_2}), $\bar{X}(VI_{A_3})$
G4	100.00 ± 0.00	100.00	IQR(VI_{A_1})

Figure 5 presents the best result of the G1 group preselecting an initial set of 145 parameters and implementing the MWVA method for different window widths and amplitudes in Figure 5(a). Based on the relationship between the number of variables and the window widths, and the elbow criterion to select the best parameters, with 57 preselected, 99.98% of the energy was obtained in Figure 5(b). Then, the forward selection method is applied in Figure 5(c) and finally, the two best parameters are selected in Figure 5(d).

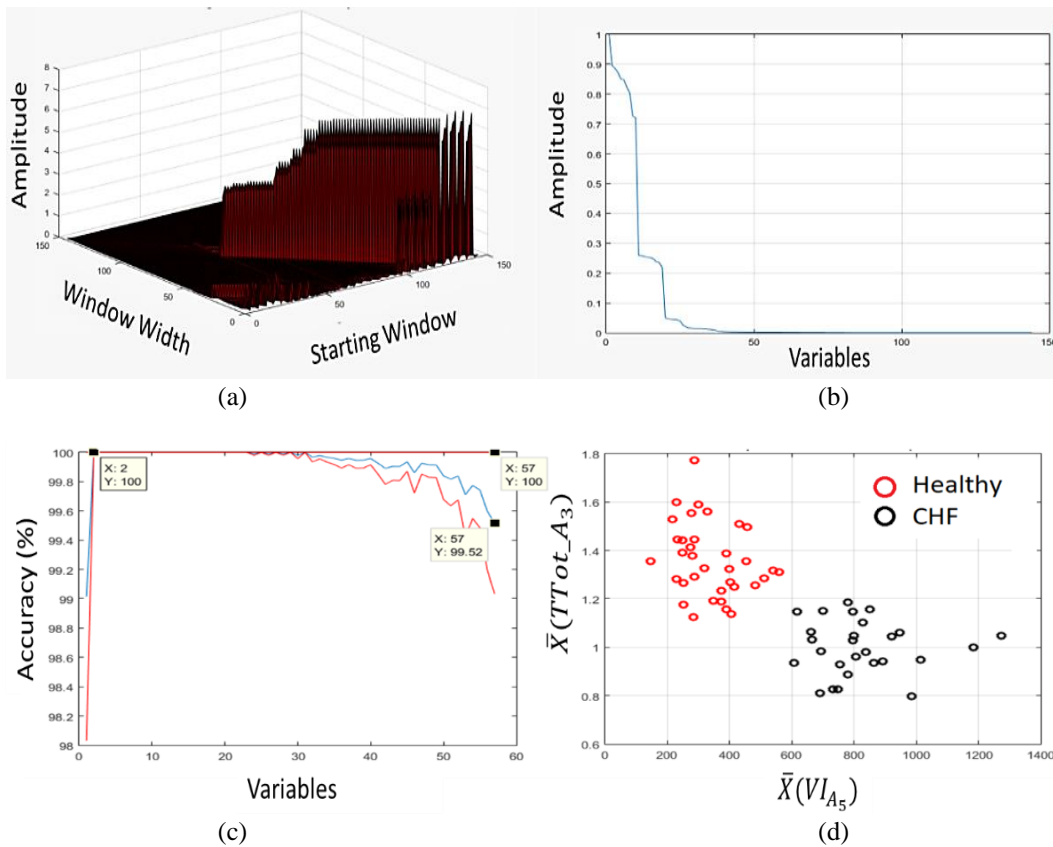
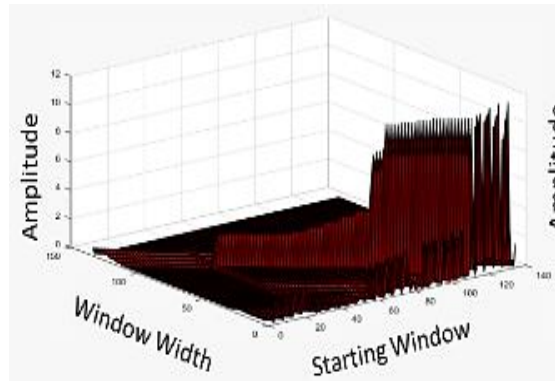


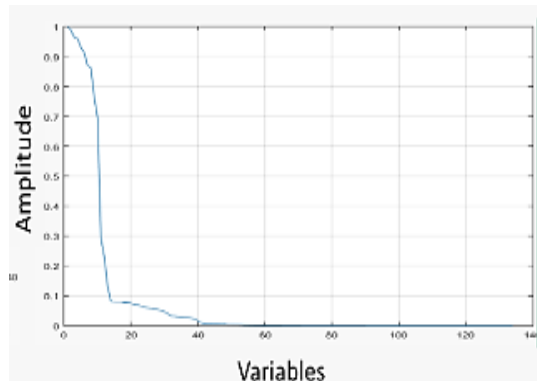
Figure 5. Behavior of the best result of G1 group when comparing healthy subjects versus chronic heart failure patients (CHF): (a) for different width and amplitude of window selection according to MWVA dissimilarity, (b) relationship between number of variables and window amplitude, (c) results of the forward selection process, and (d) a scatterplot of the two most relevant parameters

Figure 6 presents the best result of the G2 group preselecting an initial set of 135 parameters and implementing the MWVA method for different window widths and amplitudes in Figure 6(a). Based on the relationship between the number of variables and the window widths, and the elbow criterion to select the best parameters, with 59 preselected, 99.87% of the energy was obtained in Figure 6(b). Then, the forward selection method is applied, obtaining an accuracy of 100% with the first selected parameter in Figure 6(c).

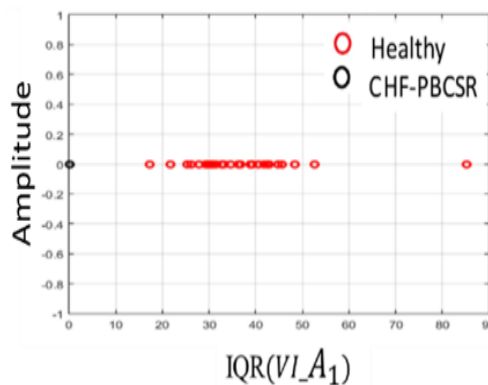
Figure 7 presents the best result of the G3 group preselecting an initial set of 140 parameters and implementing the MWVA method for different window widths and amplitudes in Figure 7(a). Based on the relationship between the number of variables and the window widths, and the elbow criterion to select the best parameters, with 70 preselected, 99.86% of the energy was obtained in Figure 7(b). Then, the forward selection method is applied Figure 7(c), and finally the two best parameters are selected in Figure 7(d).



(a)



(b)



(c)

Figure 6. Behavior of the best result of G2 group when comparing healthy subjects versus CHF-PBCSR respiration: (a) for different width and amplitude of window selection according to MWVA dissimilarity, (b) relationship between number of variables and window amplitude, and (c) a scatterplot with the best most relevant parameter that achieved 100% accuracy

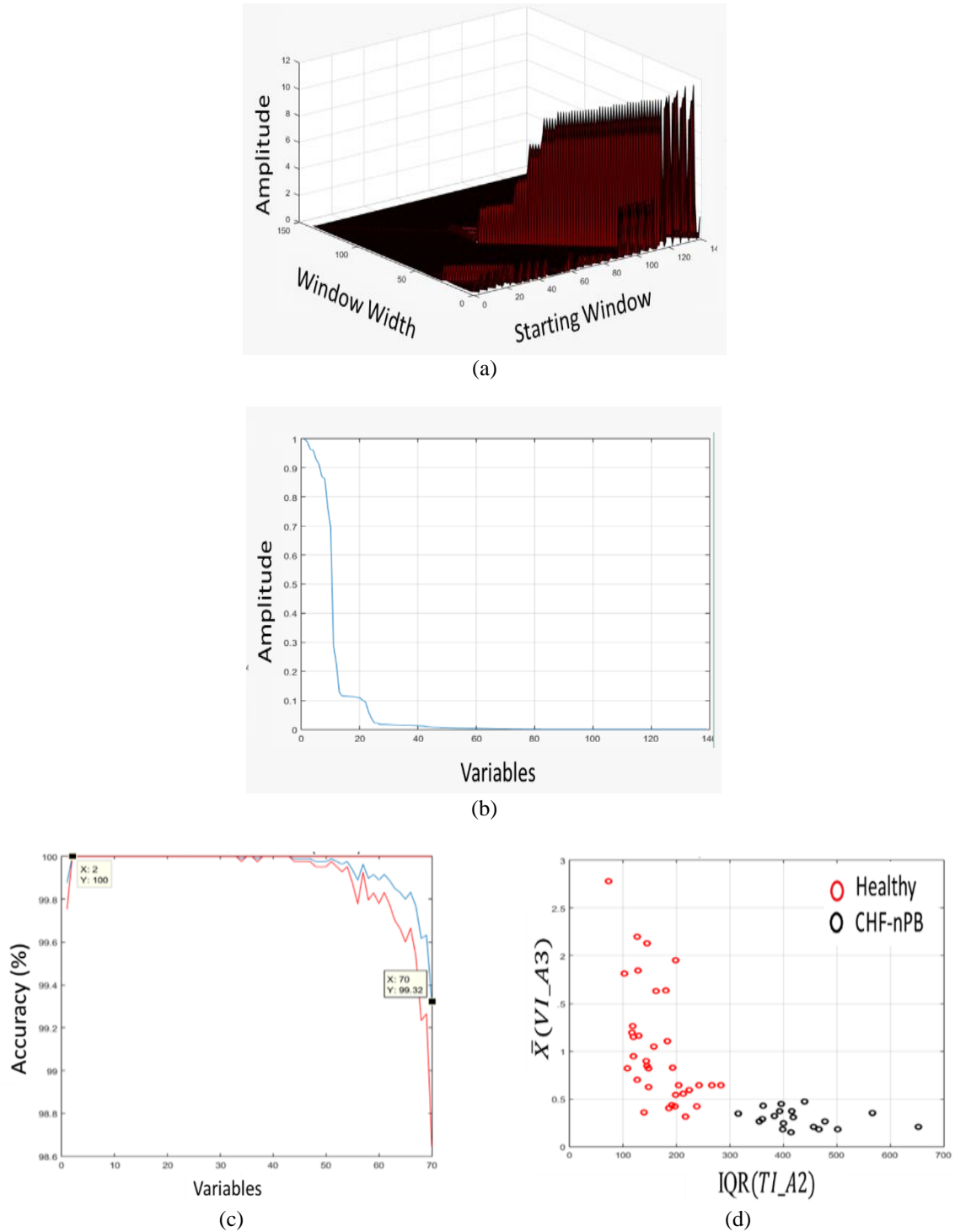


Figure 7. Behavior of the best result of G3 group when comparing healthy subjects versus CHF-nPB patients: (a) for different width and amplitude of window selection according to MWVA dissimilarity, (b) relationship between number of variables and window amplitude, (c) results of the forward selection process, and (d) a scatterplot of the two most relevant parameters

Figure 8 presents the best result of the G4 group preselecting an initial set of 95 parameters and implementing the MWVA method for different window widths and amplitudes in Figure 8(a). Based on the relationship between the number of variables and the window widths, and the elbow criterion to select the best parameters, with 80 preselected, 99.99% of the energy was obtained in Figure 8(b). Then, the forward selection method is applied in Figure 8(c), and finally the three best parameters are selected in Figure 8(d).

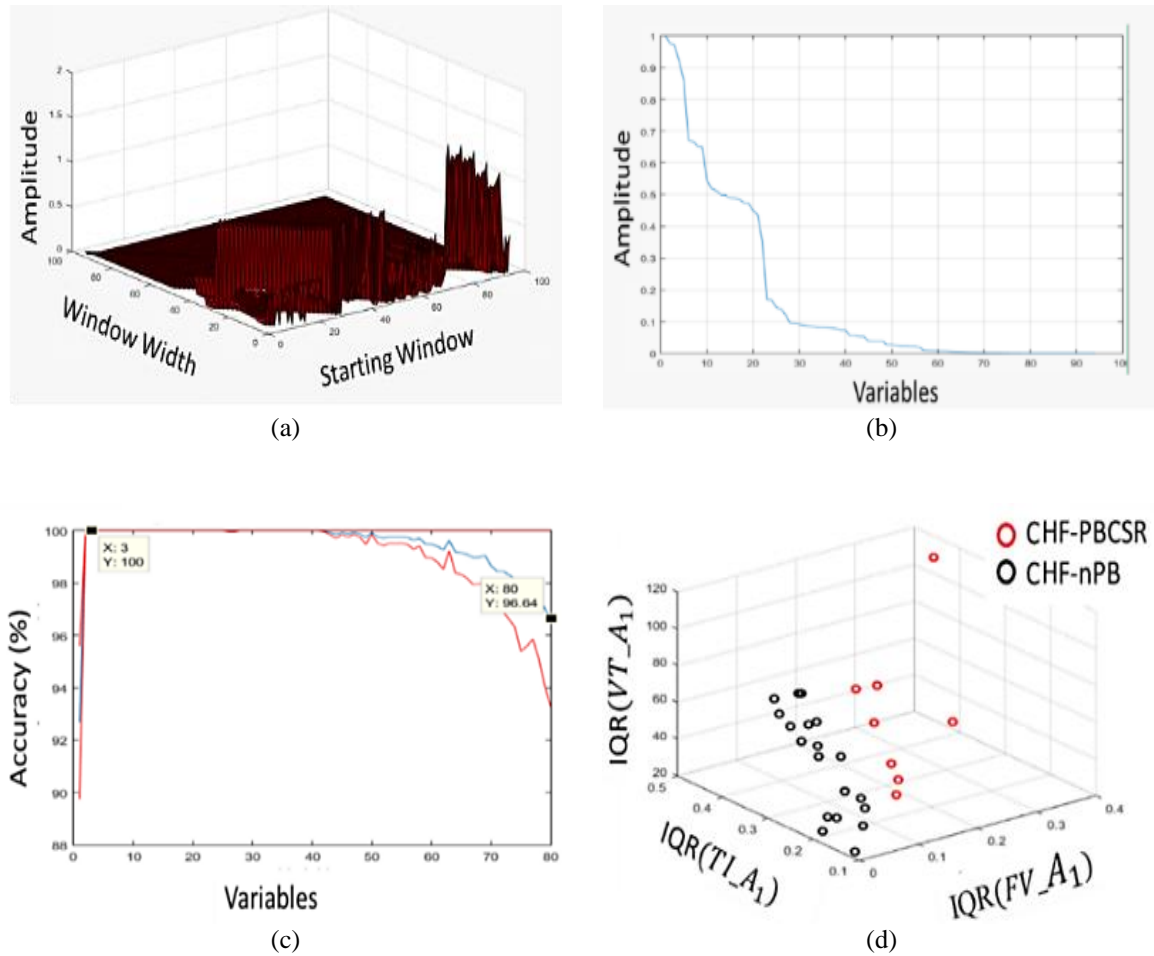


Figure 8. Behavior of the best result of G4 group when comparing CHF-PBCSR versus CHF-nPB patients: (a) for different width and amplitude of window selection according to MWVA dissimilarity, (b) relationship between number of variables and window amplitude, (c) results of the forward selection process, and (d) a scatterplot of the three most relevant parameters

5. CONCLUSION

In this study, the classification of patients with CHF has been proposed through the parameters extracted from the respiratory flow signal. According to the results, techniques related to biomedical signal processing and machine learning can contribute to determining characteristics to classify these patients. Seven time series extracted from the respiratory flow signal allow describing the different respiratory patterns of these patients. The analysis based on the Q index allows selecting the best mother wavelet applied to these time series that characterize these respiratory patterns. Additionally, from the statistical analysis in Table 2, the low frequencies of the respiratory series predominate in the representation of the separability of the data classification as initial preselection of variables. Tables 3 and 4 show the strengths of the implementation in experiment 9, where the extraction of the most frequent information through the DWT and the statistical parameters generates a high-dimensional system. Subsequently, the selection of new variables with MWVA and then with the forward selection method allow avoiding the course of dimensionality, generating models for each of the four groups of clinical interest, with a maximum of 3 relevant variables, to make an outstanding diagnosis.

The results of the classification showed that the SVMs present a reliable performance in the study of these characteristics with high accuracy. The procedure applied for the preselection of parameters, that can represent the most information about the process, has contributed to improving the selection of the minimum number of these best parameters that contain the maximum information. In all comparisons, models with a maximum of three parameters that presented the highest precision were obtained. Comparing the methodologies of this study, increases in the performance of the classifiers are observed when the proposed processing techniques and dimensionality reduction methods are implemented. A special contribution of this

study is the adaptability of parameter preselection to balance the performance of different types of groups regardless of the number of samples. However, these results should be validated with a larger number of patients, especially those with periodic breathing and Cheyne-Stokes respiration. All of this could contribute to improving the early diagnosis of these patients.

6. STATEMENTS AND DECLARATIONS

6.1. Funding

This work was supported in part by the CERCA program/*Generalitat de Catalunya*, the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya under grant GRC 2021 SGR 01390; in part by the Spanish Ministry of Science and Innovation (PID2021-126455OB-I00 MCIN/AEI/FEDER); in part by the UNAB CONVOCA program of the Universidad Autónoma de Bucaramanga, project code 2021/00003/002/001/006; and in part by the Ministry of Education, Colombian Republic, under the bicentennial doctoral excellence grant to doctoral student Hernando González Acevedo.

6.2. Ethics approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The database worked on this publication has been used in different scientific publications. This database was recorded between January to March 2006, with patients from the Short-State Unit at the Santa Creu i Sant Pau Hospital, Barcelona, Spain. All patients were studied in accordance with the corresponding protocol, approved by the clinical research ethics committee (CEIC—HSant-Pau). All data, clinical information and records, were de-identified and coded.

Data supporting the findings of this study are available from Dr. Beatriz F. Giraldo, but restrictions apply to their availability, used under license for this study, and are therefore not publicly available. The data set is not publicly available due to the conditions that were established when the protocol was defined and patients were registered.

ACKNOWLEDGMENTS

This work was supported in part by Universidad Autonoma de Bucaramanga (UNAB) by the grant GIN02-03-FO-03 associated to the project number 00142 of biennial scientific call UNAB, the CERCA Program/*Generalitat de Catalunya*, in part by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya under grant GRC 2021 SGR 01390, in part by the Spanish grant and PID2021-126455OB-I00 (MCIN/AEI/FEDER). Also was supported in part by the Colombian Ministry under grant Becas de Excelencia Doctoral del Bicentenario to PhD student Hernando González Acevedo. This paper is a contribution under the program “Colombia Científica-Pasaporte a la ciencia”, solution of the focus Health, Challenge 3: Scientific and technological development for timely treatment.




REFERENCES

- [1] P. Ponikowski *et al.*, “2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure,” *European Journal of Heart Failure*, vol. 18, no. 8, pp. 891–975, Aug. 2016, doi: 10.1002/ejhf.592.
- [2] G. Savarese and L. H. Lund, “Global public health burden of heart failure,” *Cardiac Failure Review*, vol. 03, no. 01, 2017, doi: 10.15420/cfr.2016:25:2.
- [3] E. E. S. van Riet, A. W. Hoes, K. P. Wagenaar, A. Limburg, M. A. J. Landman, and F. H. Rutten, “Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review,” *European Journal of Heart Failure*, vol. 18, no. 3, pp. 242–252, Mar. 2016, doi: 10.1002/ejhf.483.
- [4] K. Kanaoka *et al.*, “Quality indicators for acute cardiovascular diseases: a scoping review,” *BMC Health Services Research*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12913-022-08239-0.
- [5] O. Bebb *et al.*, “Performance of hospitals according to the ESC ACCA quality indicators and 30-day mortality for acute myocardial infarction: national cohort study using the United Kingdom Myocardial Ischaemia National Audit Project (MINAP) register,” *European Heart Journal*, vol. 38, no. 13, pp. 974–982, Apr. 2017, doi: 10.1093/eurheartj/ehx008.
- [6] J. Wu *et al.*, “Patient response, treatments, and mortality for acute myocardial infarction during the COVID-19 pandemic,” *European Heart Journal - Quality of Care and Clinical Outcomes*, vol. 7, no. 3, pp. 238–246, May 2021, doi: 10.1093/ehjqcco/qcaa062.
- [7] M. Arrigo *et al.*, “Acute heart failure,” *Nature Reviews Disease Primers*, vol. 6, no. 1, Mar. 2020, doi: 10.1038/s41572-020-0151-7.
- [8] P. A. Gould and D. M. Kaye, “Clinical treatment regimens for chronic heart failure: a review,” *Expert Opinion on Pharmacotherapy*, vol. 3, no. 11, pp. 1569–1576, Nov. 2002, doi: 10.1517/14656566.3.11.1569.
- [9] D. P. Francis, K. Willson, L. C. Davies, A. J. S. Coats, and M. Piepoli, “Quantitative general theory for periodic breathing in chronic heart failure and its clinical implications,” *Circulation*, vol. 102, no. 18, pp. 2214–2221, Oct. 2000, doi: 10.1161/01.CIR.102.18.2214.




- [10] A. Garde, L. Sömmo, R. Jané, and B. F. Giraldo, "Correntropy-based spectral characterization of respiratory patterns in patients with chronic heart failure," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 1964–1972, Aug. 2010, doi: 10.1109/TBME.2010.2044176.
- [11] P. Ponikowski *et al.*, "Heart failure: preventing disease and death worldwide," *ESC Heart Failure*, vol. 1, no. 1, pp. 4–25, Sep. 2014, doi: 10.1002/ehf2.12005.
- [12] G. V. Ramani, P. A. Uber, and M. R. Mehra, "Chronic heart failure: contemporary diagnosis and management," *Mayo Clinic Proceedings*, vol. 85, no. 2, pp. 180–195, Feb. 2010, doi: 10.4065/mcp.2009.0494.
- [13] M. R. Cowie *et al.*, "Survival of patients with a new diagnosis of heart failure: a population based study," *Heart*, vol. 83, no. 5, pp. 505–510, May 2000, doi: 10.1136/heart.83.5.505.
- [14] M. Guazzi *et al.*, "Exercise oscillatory ventilation may predict sudden cardiac death in heart failure patients," *Journal of the American College of Cardiology*, vol. 50, no. 4, pp. 299–308, Jul. 2007, doi: 10.1016/j.jacc.2007.03.042.
- [15] New York Heart Association Criteria Committee, *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*, 9th ed. Little, Brown, Boston, 1979.
- [16] T. A. McDonagh *et al.*, "2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure," *European Heart Journal*, vol. 42, no. 36, pp. 3599–3726, Sep. 2021, doi: 10.1093/eurheartj/ehab368.
- [17] J. Pinto, H. González, C. Arizmendi, H. González, Y. Muñoz, and B. F. Giraldo, "Analysis of the cardiorespiratory pattern of patients undergoing weaning using artificial intelligence," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, Mar. 2023, doi: 10.3390/ijerph20054430.
- [18] A. Garde *et al.*, "Assessment of respiratory flow cycle morphology in patients with chronic heart failure," *Medical & Biological Engineering & Computing*, vol. 55, no. 2, pp. 245–255, Feb. 2017, doi: 10.1007/s11517-016-1498-5.
- [19] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.
- [20] C. Arizmendi, A. Vellido, and E. Romero, "Classification of human brain tumours from MRS data using discrete wavelet transform and Bayesian neural networks," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5223–5232, Apr. 2012, doi: 10.1016/j.eswa.2011.11.017.
- [21] S. Xia and Y. Yang, "An iterative model-free feature screening procedure: forward recursive selection," *Knowledge-Based Systems*, vol. 246, Jun. 2022, doi: 10.1016/j.knsys.2022.108745.
- [22] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/s11042-020-10139-6.
- [23] W. Deng *et al.*, "An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems," *Information Sciences*, vol. 585, pp. 441–453, Mar. 2022, doi: 10.1016/j.ins.2021.11.052.
- [24] C. Arizmendi, D. A. Sierra, A. Vellido, and E. Romero, "Automated classification of brain tumours from short echo time in vivo MRS data using Gaussian decomposition and Bayesian neural networks," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5296–5307, Sep. 2014, doi: 10.1016/j.eswa.2014.02.031.
- [25] F. Liu and Y. Deng, "Determine the number of unknown targets in Open World based on Elbow method," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 986–995, May 2021, doi: 10.1109/TFUZZ.2020.2966182.
- [26] M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie, "Comprehensive review on twin support vector machines," *Annals of Operations Research*, Mar. 2022, doi: 10.1007/s10479-022-04575-w.
- [27] S. Suthaharan, "Support vector machine," Boston, MA: Springer, 2016, pp. 207–235.
- [28] J. Lu, W. Qian, S. Li, and R. Cui, "Enhanced K-nearest neighbor for intelligent fault diagnosis of rotating machinery," *Applied Sciences*, vol. 11, no. 3, Jan. 2021, doi: 10.3390/app11030919.
- [29] E. Dann, N. C. Henderson, S. A. Teichmann, M. D. Morgan, and J. C. Marioni, "Differential abundance testing on single-cell data using k-nearest neighbor graphs," *Nature Biotechnology*, vol. 40, no. 2, pp. 245–253, Feb. 2022, doi: 10.1038/s41587-021-01033-z.
- [30] S. Visalakshi and V. Radha, "A literature review of feature selection techniques and applications: review of feature selection in data mining," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, Dec. 2014, pp. 1–6, doi: 10.1109/ICCIC.2014.7238499.
- [31] B. M. S. Hasan and A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, 2021.
- [32] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, 2012.
- [33] N. Sultana, S. M. Z. Hossain, M. Abusaad, N. Alanbar, Y. Senan, and S. A. Razzak, "Prediction of biodiesel production from microalgal oil using Bayesian optimization algorithm-based machine learning approaches," *Fuel*, vol. 309, Feb. 2022, doi: 10.1016/j.fuel.2021.122184.

BIOGRAPHIES OF AUTHORS






Carlos Arizmendi    received the B.Eng. degree in electronic engineering from Universidad Industrial de Santander (UIS), Bucaramanga, Colombia in 1997, in 2008 received the diploma of advanced studies in the doctorate of biomedical engineering from UPC, in 2012 the PhD in artificial intelligent from UPC. He is currently titular professor of the mechatronics engineering program and biomedical engineering program at Universidad Autonoma de Bucaramanga (UNAB), also is director of the Research Group in Control and Mechatronics (GICYM) at UNAB, creator of the biomedical engineering program at UNAB. His research interests include machine learning, deep learning, signal treatments, electronics, mechatronics and biomedical devices. He can be contacted at email: carizmendi@unab.edu.co.






Jhon Reinemer    received received the B.Eng., degree in mechatronics engineering from UNAB. Jhon is an engineer with ability to design and implement automatic control systems, automate process, develop applications on the signal processing, machine learning and data mining field. With knowledge in pneumatics, hydraulics, robotics, power electronics and artificial intelligence. Engineer with experience in variable speed driver maintenance used in oil and gas industry for controlling the speed of installed motors in artificial lift units. He can be contacted at email: jrodriguez86@unab.edu.co. His profile can be found on: <https://co.linkedin.com/in/jhon-cesar-oswaldo-rodr%C3%ADguez-reinemer-8bb25ab6>



Hernando Gonzalez    received the B.Eng. degree in electronic engineering from UIS, master in electronic engineering from UIS, student of doctorate in engineering from UNAB, currently is associate professor of the mechatronics engineering program at UNAB and member of GICYM, His research interests include machine learning, control systems, signal treatments, robotics, electronics and mechatronics devices. He can be contacted at email: hgonzalez7@unab.edu.co.



Beatriz F. Giraldo    received the M.Sc. degree in electrical engineering from the Technical University of Pereira, Pereira, Colombia, in 1983, and the M.Sc. degree in bioengineering, and the Ph.D. degree in industrial engineering in the program of biomedical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1990 and 1996, respectively. She is associate professor at the Automatic Control Department, UPC, and senior researcher of the Institute for Bioengineering of Catalonia and the CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), in Spain. Her main research interests include biomedical signal processing and statistical analysis of cardiac, respiratory, and cardiorespiratory signals. The last studies are orientated to enhance the knowledge of the respiratory pattern and their interaction with cardiac system, working with different signals as ECG, flow and volume signal, blood pressure. Current research projects include study of cardiac and respiratory system in elderly patients, chronic heart failure patients, and patients on weaning trial process. She can be contacted at email: Beatriz.Giraldo@upc.edu, and bgiraldo@ibecbarcelona.eu.