

# On the development of diagnostic support algorithms based on CPET biosignals data *via* machine learning and wavelets

Rafael F. Pinheiro and Rui Fonseca-Pinto

Center for Innovative Care and Health Technology (ciTechCare), School of Health Sciences (ESSLei), Polytechnic University of Leiria, Leiria, Leiria, Portugal

## ABSTRACT

For preventing health complications and reducing the strain on healthcare systems, early identification of diseases is imperative. In this context, artificial intelligence has become increasingly prominent in the field of medicine, offering essential support for disease diagnosis. This article introduces an algorithm that builds upon an earlier methodology to assess biosignals acquired through cardiopulmonary exercise testing (CPET) for identifying metabolic syndrome (MS), heart failure (HF), and healthy individuals (H). Leveraging support vector machine (SVM) technology, a well-known machine learning classification method, in combination with wavelet transforms for feature extraction, the algorithm takes an innovative approach. The model was trained on CPET data from 45 participants, including 15 with MS, 15 with HF, and 15 healthy controls. For binary classification tasks, the SVM with a polynomial kernel and 5-level wavelet transform (SVM-POL-BW5) outperformed similar methods described in the literature. Moreover, one of the main contributions of this study is the development of a multi-class classification algorithm using the SVM employing a linear kernel and 3-level wavelet transforms (SVM-LIN-MW3), reaching an average accuracy of 95%. In conclusion, the application of SVM-based algorithms combined with wavelet transforms to analyze CPET data shows promise in diagnosing various diseases, highlighting their adaptability and broader potential applications in healthcare.

Submitted 20 June 2024  
Accepted 11 October 2024  
Published 30 January 2025

Corresponding author  
Rafael F. Pinheiro,  
rafael.f.pinheiro@ipleiria.pt

Academic editor  
Andrea Brunello

Additional Information and  
Declarations can be found on  
page 20

DOI [10.7717/peerj-cs.2474](https://doi.org/10.7717/peerj-cs.2474)

© Copyright  
2025 Pinheiro and Fonseca-Pinto

Distributed under  
Creative Commons CC-BY 4.0

**Subjects** Algorithms and Analysis of Algorithms, Artificial Intelligence, Data Mining and Machine Learning

**Keywords** CPET, Multi-class classification, Early diagnosis systems, Heart disease, Metabolic diseases

## OPEN ACCESS

## INTRODUCTION

Metabolic syndrome (MS) and heart failure (HF) both constitute significant global public health issues. MS is a set<sup>1</sup> of conditions that occur simultaneously and elevate the risk of type 2 diabetes, cardiovascular diseases, and other health problems. According to *Noubiap et al. (2022)*, depending on the criteria applied, the global prevalence of MS can vary from 12.5% to 31.4%. Moreover, prevalence varies across different regions, for example, based on the Adult Treatment Panel (ATP) III definition and the World Health Organization (WHO) regions, the prevalence of MS is 32.9% in the Eastern Mediterranean Region, 26.0% in the Region of the Americas, 25.3% in the European Region, and 18.9% in the African Region. These numbers establish it as one of the most prevalent chronic diseases worldwide, being associated with higher mortality from various causes (*Li et al., 2021*). Conversely, HF is a clinical condition that has a very significant impact on day-to-day life, gradually becoming incapacitating and occurs when the heart is unable to adequately pump blood to supply oxygen and nutrient needs. The prevalence of heart failure is on the rise, affecting approximately 26 million individuals worldwide (*Bowen et al., 2020*). The key to attenuating the societal impact of these conditions lies in prevention, early detection, and appropriate treatment.

<sup>1</sup> According to the National Cholesterol Education Program (NCEP), Adult Treatment Panel (ATP) III, individuals are classified as having MS if they meet three or more of the following criteria (*Zhou et al., 2021*): (1) high-density lipoprotein (HDL) cholesterol less than 40 mg/dl (1.02 mmol/l) for men and less than 50 mg/dl (1.29 mmol/l) for women; (2) fasting glucose greater than 100 mg/dl (5.6 mmol/l); (3) triglyceride levels greater than 150 mg/dl (1.7 mmol/l); (4) diastolic blood pressure greater than 85 mmHg or systolic blood pressure greater than 130 mmHg; and (5) abdominal obesity, indicated by a waist circumference greater than 102 cm for men and greater than 88 cm for women.

Beyond the prevalence issue mentioned above, HF and metabolic syndrome (MS) are chosen for the classification algorithm in this study because of their well-established clinical interconnection. MS is a recognized risk factor for the development of HF and commonly occurs as a comorbidity in HF patients (*Purwowiyoto & Prawara, 2021*), particularly in cases of HF with preserved ejection fraction (HFpEF), where an increased risk of hospitalization is associated with MS (*Zhou et al., 2021*). Although MS can be diagnosed through straightforward clinical measurements, such as waist circumference, fasting glucose, triglycerides, HDL cholesterol, and blood pressure, cardiopulmonary exercise testing (CPET) offers a non-invasive alternative, avoiding the need for invasive blood tests. The use of CPET data in the algorithm aims to improve diagnostic accuracy and inform early intervention strategies, justifying the selection of these two conditions.

CPET is a procedure evaluating the body's reaction to exercise through integrated analysis of cardiovascular, respiratory, and metabolic functions. It furnishes vital insights for diagnosing, prognosticating, and devising treatment strategies for various medical conditions, encompassing cardiac (*Saito et al., 2023*), metabolic (*Rodriguez et al., 2022*), and pulmonary (*Luo et al., 2021*) ailments. Health professionals undertake data interpretation, aiding in evaluating cardiorespiratory capacity, optimizing physical training regimens and diagnosing diseases.

Conversely, comprehensive analysis underlies the interpretation of CPET data, focusing on the variables documented throughout the examination. Presently, interpretation aligns with guidelines and criteria established by medical and exercise physiology associations, as well as by scientific research serving as benchmarks for result comprehension (primarily utilizing the flowchart—refer to *Kaminsky et al. (2017)* and *Hansen et al. (2019)*). The flowchart method uses binary decision trees to classify test results into categories like HF or MS, based on key metrics and normative values (see Fig. 3 in *Brown et al. (2022)*). Along

this vein, it is recognized that CPET data interpretation for diagnostic purposes remains an ongoing discussion, and with advancements in artificial intelligence techniques, novel methods and algorithms have surfaced to assist physicians in delivering more precise diagnoses and treatment plans.

Within the domain of artificial intelligence, classification algorithms have been essential in developing diagnostic tools, particularly using strategies of machine learning and artificial neural networks. In the case of binary classification, the research by [Brown et al. \(2022\)](#) in artificial neural networks is noteworthy, wherein hybrid models integrating autoencoders (AE) and convolutional neural networks (CNN) alongside logistic regression (LR) and principal component analysis (PCA) are devised for classifying HF and MS, employing a dataset of 15 CPET files for each condition. Regarding machine learning, in the context of multi-classification, [Inbar et al. \(2021\)](#) demonstrates the efficacy of the support vector machine (SVM) technique in categorizing diseases such as heart failure, chronic obstructive pulmonary disease, and healthy volunteers, achieving remarkable accuracy rates of 100% with a training dataset comprising approximately 70 CPET files per ailment. Although the methodologies of these studies differ considerably, both produce highly effective results.

In line with the development of algorithms to support diagnosis using supervised binary classification and multi-classification techniques for HF and MS diseases, this article presents an extended version of [Pinheiro & Fonseca-Pinto \(2023\)](#) with significant additional content. [Pinheiro & Fonseca-Pinto \(2023\)](#) presented a methodology using 3-level Daubechies wavelet transforms for feature selection and compared the accuracy with the [Brown et al. \(2022\)](#) methods, obtaining good results for SVM (HF/MS) binary classification. Now, in this work, the theory advances with new results, for multi-classification considering the HF, MS and health (H) patient labels. The contributions are explained in more detail below:

- Based on the literature review performed by the authors, there are no prior studies that address the creation of algorithms for disease diagnosis using CPET data that integrate SVM with wavelets.
- This work addresses the use of wavelet transforms for preprocessing the data, presenting a method for extracting features from CPET data as a highly efficient alternative for reducing computational costs. An analysis is conducted to compare SVM models with wavelets at three levels and five levels. This method facilitates a considerable reduction in the feature dimensionality employed in classification algorithms for analyzing CPET data. The authors assert that this approach constitutes the primary contribution of the work, offering an efficient algorithm with minimal computational overhead compared to [Inbar et al. \(2021\)](#) and [Brown et al. \(2022\)](#).
- This work provides a detailed explanation of the conceptualization of evaluation metrics used (accuracy, precision, recall, F1-score), both for binary and multi-classification cases. During the bibliographic research for this work, a scarcity of content addressing the clear derivation of evaluation metrics for multi-classification was noted. Therefore, the mathematical operationalization for clarification is formalized through [Eqs. \(1\)–\(4\)](#),

contributing in a didactic context on how to obtain evaluation metrics for multi-classification.

- For binary classification, the SVM model utilizing a polynomial kernel combined with 5-level wavelets (SVM-POL-BW5) presents a new construction methodology and better performance based on the evaluation metrics, in comparison to other algorithms employing LR, CNN, PCA and flowcharts, proving to be able to compete with the AE +LR technique of [Brown et al. \(2022\)](#). For multi-classification, this work with the methodology and application presented (utilization of CPET data) is world-first, with the highest ranked model being the SVM with a linear kernel with wavelets for extracting 3-level features (SVM-LIN-MW3).
- Given that reducing feature dimensions is very important for biosignals originating from the brain (*e.g.*, electroencephalography data), due to the enormous amount of information, it is understood that the methodological approach of SVM with wavelets presented here can be effectively used to support the diagnosis of neurological and psychiatric diseases.

[Figure 1](#) illustrates the design encompassing all the stages of the process, from the collection of CPET data to the post-processing of patient data using SVM with wavelet algorithms. The right side of this figure could be expanded, in the future with new labels, as the authors intend to advance this research by developing a more comprehensive algorithm. This enhanced algorithm will use CPET data to assist in diagnosing a broader range of diseases. The current success of the algorithm in differentiating between MS, HF, and H demonstrates its potential for wider application. The goal is, in the future, to apply this methodology to identify additional conditions, including pulmonary-vascular and mechanical-ventilatory disorders. By incorporating more diverse training data, the algorithm can be further refined to distinguish between a broader spectrum of pathologies and identify patients with several overlapping conditions, thereby enhancing diagnostic accuracy for a great range of diseases.

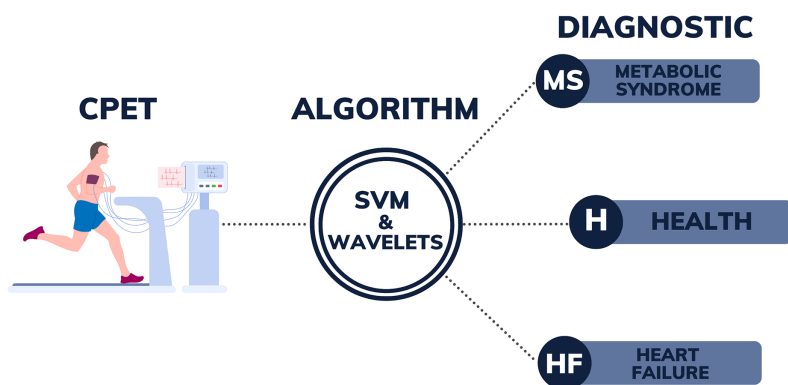
The rest of the article is structured as follows: the Methods section deals with the main theoretical basis for the development of the algorithms and methodologies are presented; in the Results section, the algorithms developed and their performance are presented with comparisons; the Discussion section presents a brief discussion of some of the issues addressed in this work; and ultimately, the article closes with the Conclusion section, which provides a summary of the findings and future work is proposed.

## METHODS

This section presents all the concepts and methods used to create and validate the algorithms presented in this work.

### Datasets

The dataset selected for this study comes from other datasets from relevant studies carried out previously. Below are details of the original datasets for MS, HF and H, as well as details of the dataset selected for training the algorithms presented in this study.



**Figure 1** Illustration of the diagnostic support system from CPET data.

Full-size DOI: 10.7717/peerj-cs.2474/fig-1

### **Original metabolic syndrome and heart failure datasets**

For MS and HF, the data used to develop the algorithms was obtained from rigorously conducted studies supported by renowned institutions, such as the National Institute of Health/National Heart Lung and Blood Institute (NIH/NHLBI) and the American Heart Association. Data for MS was gathered from a research study funded by the (NIH/NHLBI), “Exercise dose and metformin for vascular health in adults with metabolic syndrome”. Data on HF originated from patient research funded by the American Heart Association, “Personalized Approach to Cardiac Resynchronization Therapy Using High Dimensional Immunophenotyping”, as well as the NIH/NHLBI, “MRI of Mechanical Activation and Scar for Optimal Cardiac Resynchronization Therapy Implementation”. For easier access to the data, the article of [Brown et al. \(2022\)](#) made the MS and HF dataset available on *GitHub* (<https://github.com/suchethasharma/CPET>).

The study’s eligibility criteria for participants with MS and HF included specific age and health parameters. For MS, adults aged 40 to 70 years with a BMI between 27 and 47 kg/m<sup>2</sup> were included, provided they were not diagnosed with Type 2 diabetes and were not engaging in more than 60 min of exercise per week. Participants needed to meet at least three out of five Metabolic Syndrome criteria<sup>1</sup>. Exclusion criteria for MS included morbid obesity, recent significant weight changes, and a history of chronic diseases that could affect study outcomes. For HF, eligible participants were adults over 18 and 85 years old with chronic systolic heart failure and a left ventricular ejection fraction (HFrEF) of 35% or less. Some volunteers also had to meet Class I or IIa directives for the resynchronized cardiac therapy ([Epstein et al., 2008](#)). HF patients with contraindications to cardiac magnetic resonance imaging (CMR), such as those with implantable devices or certain medical conditions, were excluded. Further details can be found in other articles that were published from the projects mentioned or used their datasets, for example [Gaitán et al. \(2019\)](#), [Heiston et al. \(2019\)](#), [Malin et al. \(2019\)](#), [Bilchick et al. \(2020\)](#), [Gao et al. \(2021\)](#), [Auger et al. \(2022\)](#), [Brown et al. \(2022\)](#), and in the respective projects links provided for MS (<https://reporter.nih.gov/search/PoZvzQo230OWPUeVJHr1g/projectdetails/9934226>) and HF (<https://reporter.nih.gov/search/u0O4hSNlh0eH8Jj84y-Y0A/projectdetails/9544361>).

All patients underwent CPET on a treadmill, following a protocol divided into three phases: rest, exercise, and recovery. During the exercise phase, the treadmill's incline and speed were progressively increased. The Exercise Physiology Laboratory (EPL) at the General Clinical Research Center (GCRC) of the University of Virginia conducted the CPET, ensuring accurate data collection, which includes respiratory measurements and heart rates. All the variables provided by CPET captured from the volunteers with MS and HF, according to *GitHub* (<https://github.com/suchethasharma/CPET>) are shown in [Table 1](#). Note that this table lists all the variables used in this work, from where they were obtained and the classification application made.

### **Original healthy dataset**

CPET data from healthy volunteers was obtained from the Exercise Physiology and Human Performance Laboratory at the University of Malaga (UM). The data was found in the Physionet database (*Goldberger et al., 2000*) published by the article (*Mongin, Romero & Cruz, 2021*), as previously used in *Mongin et al. (2021)*. This dataset includes information from 992 exercise tests conducted between 2008 and 2018.

Inclusion criteria required participants to be athletes (amateur or professional), healthy, and aged between 10 and 63 years. Participants had to voluntarily agree to the maximal exercise test, with written informed consent obtained from all participants and legal guardians for those under 18 years of age. There were no specific exclusion criteria mentioned beyond standard safety guidelines for exercise testing. The research adhered to the guidelines of the Declaration of Helsinki and received approval from the Ethics Review Board of the UM, ensuring participant protection and well-being.

The procedure for gathering data required administering a maximal graded exercise test (GET) on a treadmill from the PowerJog J series. Each test began with a warm-up phase, where participants walked at 5 km/h, followed by a continuous or incremental effort, with step increments ranging from 0.5 to 1 km/h. Measurements were recorded continuously, capturing respiratory metrics on a breath-by-breath basis, while heart rate was tracked with a 12-lead ECG apparatus. The test was deemed maximal when the participants' oxygen consumption plateaued, indicating exhaustion. Upon finishing the task, the treadmill's pace was lowered to 5 km/h, and participants continued walking to prevent vasovagal syncope. Measurements were taken using specialized equipment, including the MedGraphics CPX metabolic gas analysis system and the Mortara ECG machine with 12 leads, all overseen by sports science professionals. [Table 1](#) presents all the variables collected from the CPET of healthy volunteers provided by *Mongin, Romero & Cruz (2021)*.

### **Dataset for this work**

The dataset for this work, in the treadmill test model, comprised 45 individuals, including 15 with a diagnosis of MS, 15 with HF, and 15 healthy (H) individuals. The dataset for MS and HF originates from the information provided in the subsection "Original Metabolic Syndrome and Heart Failure Datasets" and was obtained from the *GitHub* repository (<https://github.com/suchethasharma/CPET>). Meanwhile, the dataset for Healthy (H)

**Table 1** CPET variables considered for this work.

Variable description	Abbreviation	Data source	Application
Breath-by-Breath	$Time(min)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Variable not used
Time since the measurement starts	$Time(min)$	<i>Mongin, Romero &amp; Cruz (2021)</i>	Variable not used
Metabolic equivalents	<i>METS</i>	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Binary classification
Heart rate	$HR(beats/min)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> ) and <i>Mongin, Romero &amp; Cruz (2021)</i>	Binary classification and multi-classification
Peak oxygen consumption	$\dot{V}O_2(L/min)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> ) and <i>Mongin, Romero &amp; Cruz (2021)</i>	Binary classification and multi-classification
Peak oxygen consumption is measured in milliliters of oxygen used in 1 min per kilogram of body weight	$\dot{V}O_2/kg((ml/min)/kg)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Variable not used
Volume of carbon dioxide released	$\dot{V}CO_2(L/min)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> ) and <i>Mongin, Romero &amp; Cruz (2021)</i>	Binary classification and multi-classification
Respiratory exchange ratio	<i>RER</i>	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Binary classification
Ventilation	$VE(L/min)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> ) and <i>Mongin, Romero &amp; Cruz (2021)</i>	Binary classification and multi-classification
Ratio of ventilation by peak oxygen	$VE/\dot{V}O_2$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Variable not used
Ratio of ventilation by volume of carbon dioxide released	$VE/\dot{V}CO_2$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Variable not used
Respiratory rate	$RR(breaths/min)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> ) and <i>Mongin, Romero &amp; Cruz (2021)</i>	Binary classification and multi-classification
Expiratory tidal volume (expiratory time)	$V_{tex}(L)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Binary classification
Inspiratory tidal volume (inhale time)	$V_{tin}(L)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Binary classification and multi-classification
Speed of the treadmill (inhale time)	$Speed(mph)$	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> )	Variable not used
Elevation of the treadmill	<i>Elevation</i>	<i>GitHub</i> ( <a href="https://github.com/suchethassharma/CPET">https://github.com/suchethassharma/CPET</a> ) and <i>Mongin, Romero &amp; Cruz (2021)</i>	Variable not used

subjects is based on the details outlined in the subsection “Original Healthy Dataset” and was sourced from the PhysioNet database (*Mongin, Romero & Cruz, 2021*). In the case of H, the 15 oldest volunteers were taken from the original database, in order to have a sample that was closer to MS and HF in terms of age. Thus, the volunteers selected from the original database of healthy people (H) were of ID TEST: 245\_3, 296\_1, 377\_1, 389\_1,

**Table 2** Mean and standard deviation of the demographic variables of the volunteers in this study.

Condition	HF	MS	H	ALL
Sample size	15	15	15	45
Gender (Female)	26%	93%	26%	48%
Age	69.3 [58.6, 80.0]	56.9 [49.9, 63.8]	55.2 [49.5, 60.9]	60.5 [50.3, 70.6]
Height (cm)	169.9 [160.2, 179.6]	155.2 [112.1, 198.2]	167.3 [159.2, 175.4]	164.1 [138.0, 190.2]
Weight (kg)	101.8 [81.1, 122.5]	98.1 [83.2, 113.0]	67.0 [53.7, 80.3]	89.0 [66.3, 111.6]
BMI	35.4 [28.2, 42.6]	35.5 [30.7, 40.2]	23.7 [20.9, 26.6]	31.5 [24.0, 39.1]

390\_1, 486\_1, 596\_1, 597\_1, 598\_1, 609\_1, 651\_1, 653\_1, 755\_1, 756\_1, 856\_3. [Table 2](#) displays the demographic information of the samples utilized in this research.

Binary classification was performed using HF and MS data, while multi-classification included the dataset of healthy (H) volunteers. The CPET dataset offers a comprehensive array of information garnered during the test. For the binary classification, variables from the CPET were selected as per [Table 1](#). The same table shows the variables used for multi-classification. This selection was guided by the availability of data across both databases. Note that in [Table 1](#), the variables *METS*, *RER*, *Vtex* and *Vtin* were not included in the multi-classification, as they were not provided in the dataset of healthy volunteers.

## Wavelet transforms

Wavelet transforms, a potent mathematical tool extensively employed in signal and image analysis, differ from conventional transforms by offering a multiresolution approach. This approach efficiently captures both local and global signal information. They have emerged as a promising technique in classification algorithms ([Serhal et al., 2022](#); [Iniyani, Singh & Hazra, 2023](#)). Utilizing wavelet technique in data analysis enables the extraction of relevant features across various scales and frequencies, thereby providing a more comprehensive representation of dataset patterns. This capacity to discern discriminative information across multiple resolutions has spurred the creation of more precise and resilient classification models in various domains, including pattern recognition, image processing, and medical diagnostics.

In the application of a wavelet transform, the signal undergoes decomposition into levels ( $d_1, d_2, d_3, \dots$ ), each representing details at distinct frequencies. The coefficients within these decompositions elucidate the contributions of each level to the overall portrayal of the original signal. These wavelet coefficients facilitate a detailed examination of the signal across various resolutions. This study employed the Daubechies wavelet of second order with three ( $d_1, d_2, d_3$ ) and five levels ( $d_1, d_2, d_3, d_4, d_5$ ) along with an approximation for the CPET variables (*ap*).

## Features and labels

Features represent the data components utilized in training classification algorithms, while labels denote the classifications assigned to these features. For instance, consider the heart rate (HR) and respiratory rate (RR) data of a patient. Features can encompass the raw

variables themselves (HR and RR), resulting in a sizable dataset. However, for computational efficiency, parameters can be derived from these variables, such as the mean. Consequently, the features for this patient would comprise the mean HR and RR. Conversely, labels correspond to the classifications associated with the patient's features; for instance, a non-diabetic patient may be assigned label 0, whereas a diabetic patient would receive label 1. Hence, a collection of data from multiple patients forms the basis for algorithm training. The larger the patient cohort for training, the more adept the algorithm becomes at disease detection. Further insights into feature extraction are available in [Subasi \(2007\)](#), [Xing et al. \(2011\)](#).

For this work, the features were extracted from the CPET variables presented in [Table 1](#). In order to experiment and identify the best models, the features were organized into two main categories: one set for binary classification tasks and another set for multi-class classification tasks. [Figure 2](#) lists all the types of features used in this work.

In binary classification, three types of features are adopted. The first type, called  $X$  ([Fig. 2A](#)), is the simplest and consists of the mean and variance of each variable, with the first 15 rows corresponding to the data of HF patients and the other 15 rows for data of MS patients. The  $X$  characteristic was constructed using the means and variances of the variables. [Figure 2A](#) shows how the data is presented to the algorithm, *i.e.*, an  $X$  matrix with a dimension of 30 rows and 16 columns. The second set of features ([Fig. 2B](#)), referred to as  $BW3$ , it includes the average and variance of the wavelet transform coefficients across three levels ( $d_1$ ,  $d_2$  and  $d_3$ ), so the matrix of features for this case has a dimension of 30 rows by 64 columns. The last type of feature for binary classification is shown in [Fig. 2C](#), which has five levels and a dimension of 30 rows by 96 columns. For the labels, the first 15 rows correspond to data from HF patients and the other 15 rows to data from MS patients.

In multi-classification, two types of features are extracted. The first type ([Fig. 2D](#)), includes the mean and variance of the wavelet transform coefficients of five levels, called  $MW5$ . The  $MW5$  matrix has 45 rows by 60 columns. The second type ([Fig. 2E](#)), contains the variance and mean of the wavelet transform coefficients of three levels, called  $MW3$ . This matrix has 45 rows by 40 columns. For the labels, the first 15 rows correspond to data from patients with HF, from the 16th to the 30th rows correspond to volunteers with MS, and from the 31st to the 45th correspond to healthy volunteers. To improve understanding, the content of the features is summarized with the type of feature, the number of levels and the number of rows in [Table 3](#).

To extract all the features containing wavelets, specific code was used *via* Matlab R2020a that integrated the SVM models to create the diagnostic algorithms that will be presented in the results section.

## Support vector machine

Proposed by [Boser, Guyon & Vapnik \(1992\)](#), the support vector machine (SVM) is a supervised machine learning technique used for both classification and regression, aiming to identify the optimal hyperplane in multidimensional spaces to separate classes of data. Stemming from its effectiveness is the process of enlarging the margin between the support vectors, which represent the nearest points to the decision boundaries (see [Fig. 3](#)). This

## FEATURES FOR BINARY CLASSIFICATION

	METS		HR		...	Vtin	
	Mean	Var	Mean	Var	...	Mean	Var
1	<b>X</b>						
2							
3							
⋮							
30							

(A)

	METS				...	Vtin			
	Mean		Var		...	Mean		Var	
	$d_1$	$d_2$	$d_3$	$ap$		$d_1$	$d_2$	$d_3$	$ap$
1	<b>BW3</b>								
2									
⋮									
30									

(B)

	METS					...	Vtin						
	Mean		Var			...	Mean		Var				
	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$ap$		$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$ap$
1	<b>BW5</b>												
2													
⋮													
30													

(C)

## FEATURES FOR MULTI – CLASSIFICATION

	HR					...	VE						
	Mean		Var			...	Mean		Var				
	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$ap$		$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$ap$
1	<b>MW5</b>												
2													
⋮													
45													

(D)

	HR				...	VE			
	Mean		Var		...	Mean		Var	
	$d_1$	$d_2$	$d_3$	$ap$		$d_1$	$d_2$	$d_3$	$ap$
1	<b>MW3</b>								
2									
⋮									
45									

(E)

**Figure 2** Features used to build the algorithms: (A) features for binary classification extracted from the means and variances of the CPET variables; (B) and (C) features for binary classification extracted from the means and variances of the coefficients of the wavelet transforms of 3 and 5 levels respectively; (D) and (E) features for binary classification extracted from the means and variances of the coefficients of the wavelet transforms of 5 and 3 levels respectively.

Full-size  DOI: [10.7717/peerj-cs.2474/fig-2](https://doi.org/10.7717/peerj-cs.2474/fig-2)

**Table 3** Summary of the features with their applications and dimensions.

Feature type	Application	Wavelet levels	Rows	Columns
X	Binary	Not applicable	30	16
BW3	Binary	3	30	64
BW5	Binary	5	30	96
MW3	Multi-class	3	45	64
MW5	Multi-class	5	45	96

approach permits robust generalization even in complex, high-dimensional datasets, rendering it a prominent choice in various data analysis and pattern recognition applications.

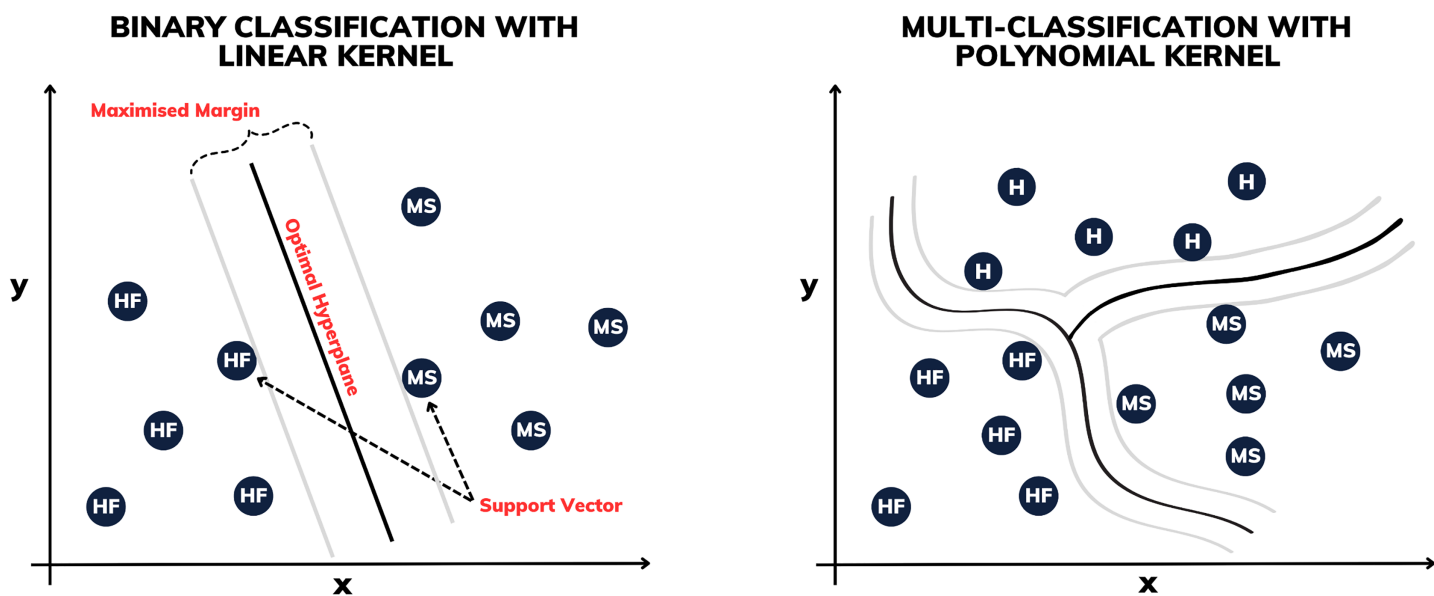


Figure 3 Illustration of binary classification and multi-classification with different kernels.

Full-size DOI: 10.7717/peerj-cs.2474/fig-3

Some of the SVM variants are determined by their kernels. In this work, constructions using linear, polynomial and radial basis function (RBF) kernels are made. The linear SVM is used to separate linearly separable classes, while polynomial SVMs are used for non-linearly separable datasets through transformations. The RBF SVM, on the other hand, is highly powerful, mapping the data into a high-dimensional space to separate complex classes.

For multi-classification, several methods have been proposed to combine multiple SVMs from two classes in order to build a multi-class classifier. In this work, using Matlab's *fitcecoc* function, the “one-vs-one” with error-correcting output codes (ECOC) model approach is used, which was developed by [Dietterich & Bakiri \(1994\)](#) and applied to support vector machines by [Allwein, Schapire & Singer \(2000\)](#).

### Validation process

To validate the algorithm, the cross-validation k-fold was the validation procedure used (see more in [Wong & Yeh \(2019\)](#)). The k-fold cross-validation method is a valuable technique in machine learning, especially when working with limited datasets, as is the case in this study. As made by [Pinheiro & Fonseca-Pinto \(2023\)](#), the k-fold cross-validation method splits the dataset into  $k$  equal parts. The model is trained  $k$  times, using one subset for testing and the remaining  $k - 1$  for training. After  $k$  iterations, the results are averaged to provide a single performance metric. This ensures efficient data usage and unbiased model evaluation. In this work, a 5-fold cross-validation is applied, where the dataset is divided into five parts. Each model is trained on 4/5 of the data and tested on the remaining 1/5, repeating this process five times. The results are the mean performance of the five rounds of 5-fold cross-validation. For cross-validation, the *crossval* function in

Matlab is used, which does not implement stratification and shuffling automatically (more on this issue is covered in the Discussion section).

The evaluation metrics were gathered following the 5-fold cross-validation process. The metrics used in this work include precision, accuracy, F1-score and recall. These metrics have also been defined and employed in various studies, such as *Brown et al. (2022)* and *Chen et al. (2023)*. The metrics are defined as follows.

- **Accuracy** indicates the model's performance as a whole by calculating the ratio of correct classifications (positive and negative) to the total number of predictions. Although it provides a good general overview, it may be misleading, for example, in cases of class imbalance, descriptors overlapped or outliers (*Morales et al., 2020; Michelucci et al., 2021*).
- **Precision** calculates the ratio of true positives to all positive predictions made by the model, indicating the number of predicted positives that are truly positives.
- **Recall**, also known as revocation or sensitivity, measures the percentage of true positive cases correctly detected by the model. This is crucial in scenarios where failing to identify positive cases is critical (False Negatives).
- **F1-score** is the harmonic mean of precision and recall, and offers a single metric that balances the trade-off that exists between them. A low F1-score indicates that either precision or recall is low, making it a useful measure when both metrics are important.

Each formula for these metrics is derived from the confusion matrix (see Fig. 4).

For a binary classification, the confusion matrix generated has a dimension of  $2 \times 2$ , where the values of the evaluation metrics ( $A$ ,  $R$ ,  $P$  and  $F1$ ) are obtained directly according to Fig. 4. In the case of multi-classification, the confusion matrix will be dimensioned according to the number of classes (labels). In order to extract the evaluation metrics in multi-classification, it is necessary to reduce the confusion matrix for each classification label, obtaining the shape of Fig. 4. In this work, to calculate the evaluation metrics in multi-classification ( $A^m$ ,  $P^m$ ,  $R^m$  and  $F1^m$ ), one considers obtaining a real overall confusion matrix  $C_{3 \times 3}$ , where the values  $TP^m$ ,  $FP^m$ ,  $FN^m$  and  $TN^m$  are obtained according to Eqs. (1)–(4), respectively, where, if  $m = 1$ , one has label HF; if  $m = 2$ , the label is MS; and if  $m = 3$ , the label is H. For a number of  $n$  labels, take  $m = 1, 2, 3, \dots, n$ .

$$TP^m = C_{m,m}, \quad (1)$$

$$FP^m = \sum_{j=1}^n C_{m,j} - TP^m, \quad (2)$$

$$FN^m = \sum_{i=1}^n C_{i,m} - TP^m, \quad (3)$$

$$TN^m = \sum_{i=1}^n \sum_{j=1}^n C_{i,j} - TP^m - FP^m - FN^m. \quad (4)$$

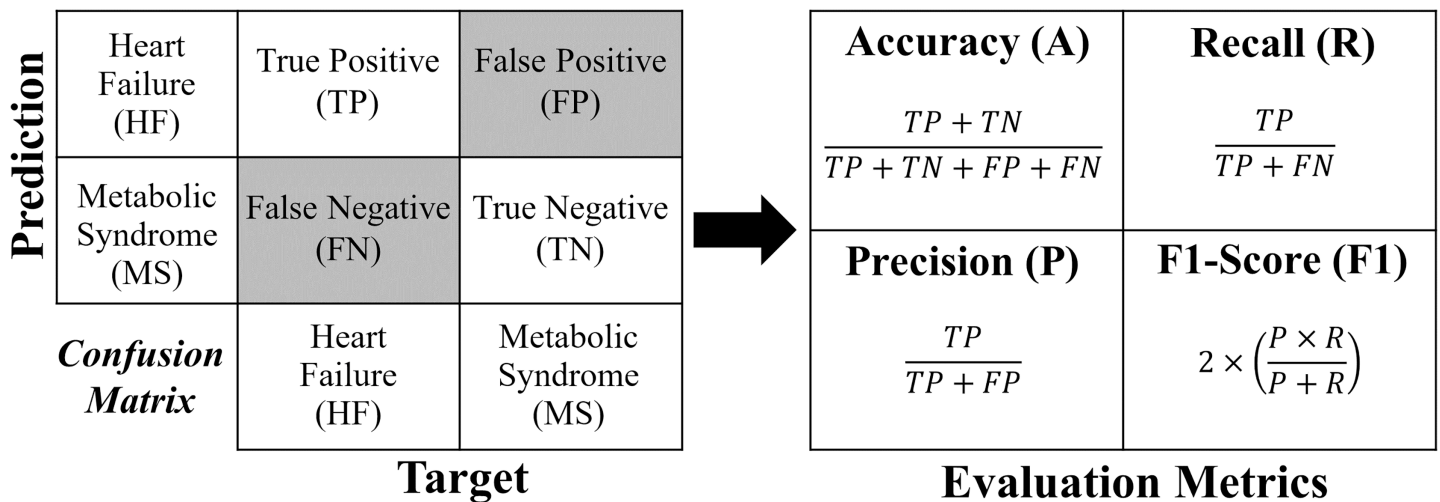


Figure 4 Confusion matrix and formulas of the evaluation metrics.

Full-size DOI: 10.7717/peerj-cs.2474/fig-4

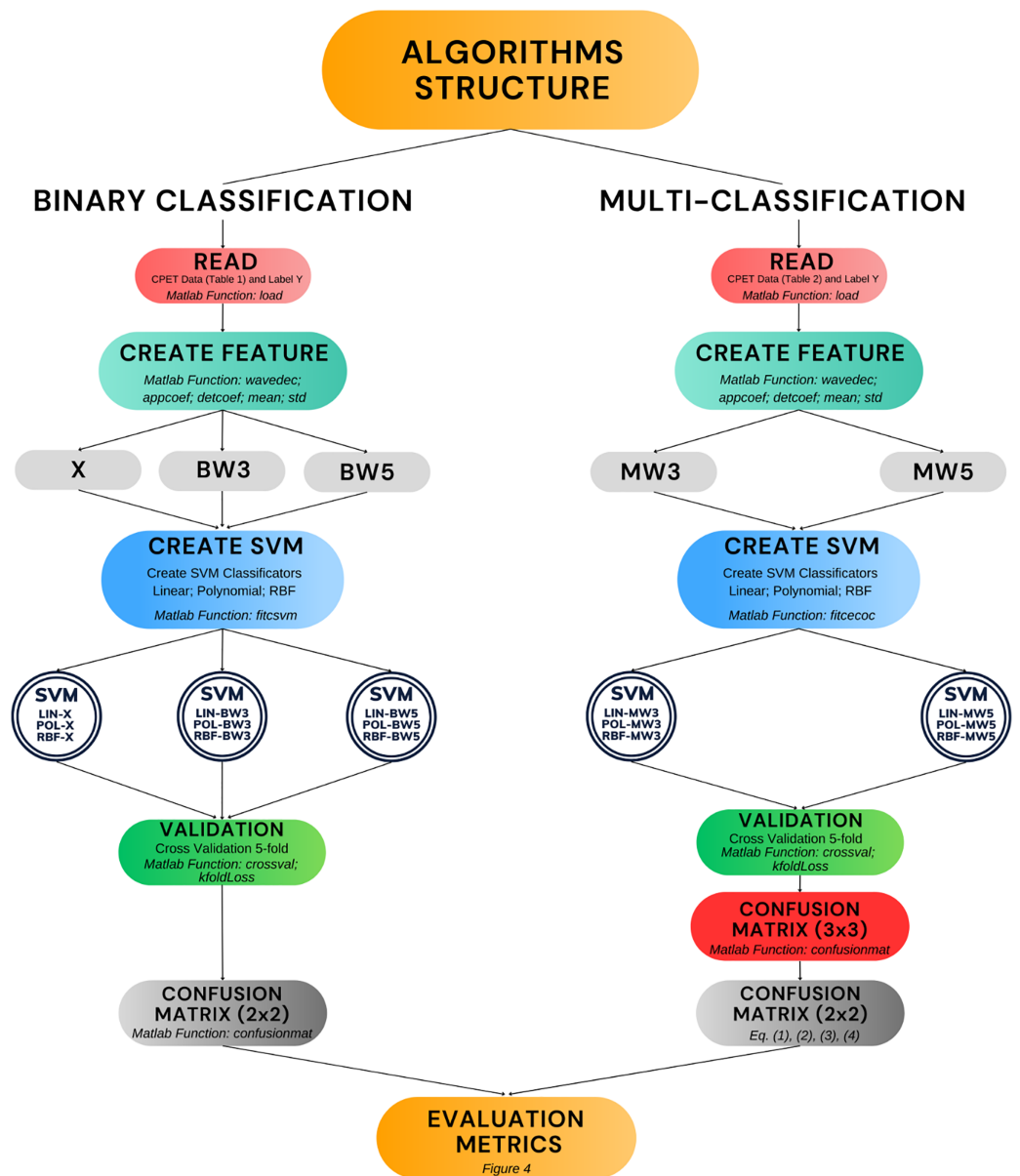
## RESULTS

Figure 5 presents the algorithms developed in this work (double circles in the figure). Moreover, it shows the process of construction and validation, including the main Matlab functions used. The diagram covers the entire structure that encompasses the phases of reading the raw data, creating the features, creating the SVM model and validation. The remainder of this results section is divided into two parts. The first part covers the performance of the algorithms in binary classification, including comparisons with other results in the literature. The second part presents the performance of the algorithms in multi-classification.

### Results for binary classification

Table 4 shows the evaluation metrics of the binary classification algorithms (HF or MS) created in this study. For the model with feature X, the SVM algorithm with a linear kernel (SVM-LIN-X) gave the best result compared to the polynomial kernel and RBF algorithms. Looking at this table, it can be seen that the SVM model with a polynomial kernel for binary classification with 5-level wavelets (SVM-POL-BW5) is the best performer (highlighted in bold). Also note that the confusion matrix generated in this case has dimension  $2 \times 2$ , so the metrics were calculated by directly applying the formulae in Fig. 4.

Table 5 shows comparisons of the metrics of the SVM-POL-BW5 algorithm (the best performer according to Table 4) with other algorithms developed in other articles. The algorithms being compared are the Flowchart method, convolutional neural networks (CNN), and hybrids: principal component analysis with logistic regression (PCA+LR); and autoencoder with logistic regression (AE+LR). It should be noted that this comparison was conducted honestly, *i.e.*, the same databases were used for all the algorithms, which also carried out the binary classification of HF and MS diseases. It can be seen from this table



**Figure 5** Algorithm development and validation process.

Full-size DOI: 10.7717/peerj-cs.2474/fig-5

that the algorithm developed in this study performed better than the other three types of algorithms compared, coming second only to the AE+LR algorithm.

### Results for multi-classification

Table 6 shows the evaluation metrics obtained for the multi-classification algorithms. These metrics were obtained using the Eqs. (1)–(4), and then the equations in Fig. 4 for each label. Next, Table 7 shows the ranking of the algorithms according to their performance. The values for Table 7 were obtained by calculating the simple mean of the results in Table 6, for example, the accuracy of the SVM-LIN-MW3 algorithm was

**Table 4** Comparisons between the algorithms in this study and the number of wavelet levels for binary classification. Bold indicates the method with the best performance.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM-POL-BW(3/5)	89/ <b>94</b>	87/ <b>90</b>	93/ <b>100</b>	90/ <b>94</b>
SVM-LIN-BW(3/5)	92/93	88/88	98/100	93/93
SVM-RBF-BW(3/5)	87/90	85/87	89/94	87/90
SVM-LIN-X	83	91	73	81
SVM-POL-X	74	74	73	73
SVM-RBF-X	71	74	68	70

**Table 5** Comparisons with other methods in the literature for binary classification. Bold indicates the method with the best performance.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1. AE+LR ( <i>Brown et al., 2022</i> )	<b>97</b>	<b>94</b>	<b>100</b>	<b>97</b>
2. SVM-POL-BW5	94	90	100	94
3. CNN ( <i>Brown et al., 2022</i> )	90	100	80	86
4. PCA+LR ( <i>Brown et al., 2022</i> )	90	93	87	90
5. Flowchart ( <i>Kaminsky et al., 2017</i> )	77	78	93	85
6. Flowchart ( <i>Hansen et al., 2019</i> )	70	100	53	70

**Table 6** Comparisons between the SVM (multi-classification) algorithms in this study and the number of wavelet levels. Bold indicates the method with the best performance.

Method	Label	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM-LIN-MW(3/5)	HF	<b>95/91</b>	<b>95/93</b>	<b>91/83</b>	<b>93/87</b>
	MS	<b>93/88</b>	<b>90/81</b>	<b>90/84</b>	<b>90/83</b>
	H	<b>97/97</b>	<b>93/92</b>	<b>100/100</b>	<b>96/95</b>
SVM-POL-MW(3/5)	HF	87/87	73/67	86/93	79/78
	MS	80/78	86/94	65/61	74/74
	H	93/90	81/72	98/100	88/83
SVM-RBF-MW(3/5)	HF	46/44	81/69	36/33	49/45
	MS	64/56	24/33	34/35	28/33
	H	68/61	13/17	68/7	20/10

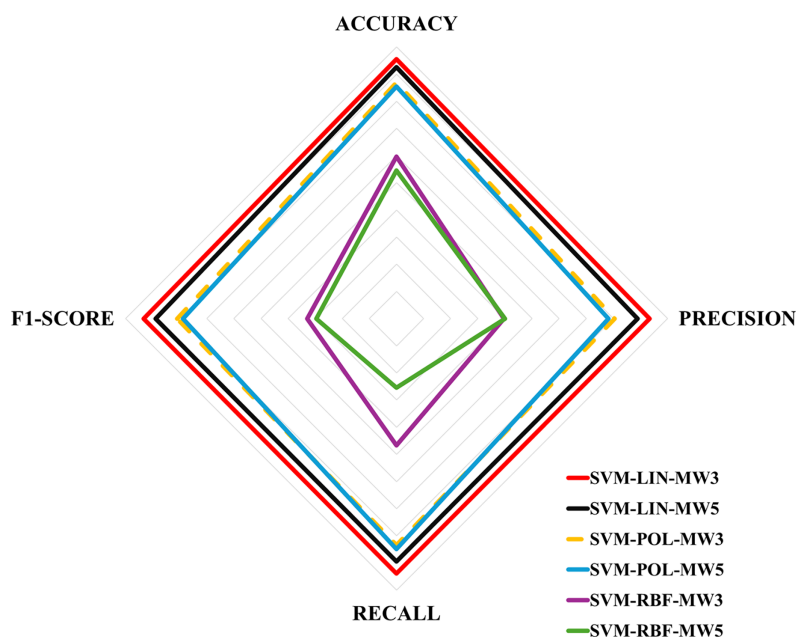
obtained by calculating the mean of the accuracy of HF, MS and H, *i.e.*,  $(95 + 93 + 97)/3$ . Finally, Fig. 6 shows a multicriteria analysis chart built from the results of Table 7. In this way, it is possible to see graphically the performance of each algorithm based on its domain region, *i.e.*, the larger the region of coverage, the better the algorithm's performance.

## DISCUSSION

Firstly, this work clarifies some issues that arise in *Pinheiro & Fonseca-Pinto (2023)*. The first concerns the inclusion of data from healthy volunteers; the second is the verification of

**Table 7** The ranking of the multi-classification algorithms for diagnostic support developed in this study. Bold indicates the method with the best performance.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1. SVM-LIN-MW3	<b>95</b>	<b>93</b>	<b>93</b>	<b>93</b>
2. SVM-LIN-MW5	92	88	89	88
3. SVM-POL-MW3	86	80	83	80
4. SVM-POL-MW5	85	78	84	78
5. SVM-RBF-MW3	59	39	46	32
6. SVM-RBF-MW5	54	40	25	29



**Figure 6** Multicriteria analysis chart.

Full-size DOI: 10.7717/peerj-cs.2474/fig-6

the SVM model with different numbers of wavelet levels; and, finally, the use of more evaluation metrics. In relation to the inclusion of data from healthy volunteers, this work responds by creating a multi-classification algorithm, which is innovative in relation to *Pinheiro & Fonseca-Pinto (2023)* and *Brown et al. (2022)*. With regard to the number of wavelet levels, this work shows that a greater number of wavelet levels does not necessarily improve algorithm performance. For example, in the case of multi-classification, algorithms with three levels of wavelets performed better than algorithms with five levels. Finally, this work provides results with the calculation of more evaluation metrics, in order to confirm the partial results already obtained in *Pinheiro & Fonseca-Pinto (2023)*, and also adds the main metrics for multi-classification.

The results obtained in this research indicate that the proposed methodology, using SVM with wavelet transformations, performed competitively compared to existing solutions in the literature, even on a smaller scale dataset. However, it is important to note

that generalizing this performance for larger datasets requires additional validation. Previous studies, which have applied SVM and wavelets to larger datasets from different domains (Asgari, Mehrnia & Moussavi, 2015; Xin & Zhao, 2017; Dhivya & Vasuki, 2018; Kehri & Awale, 2018), suggest that the approach is promising and scalable.

In this study, older volunteers were intentionally selected for the Healthy (H) dataset to obtain a more comparable age range compared to the MS and HF dataset provided by Brown *et al.* (2022). This strategy was employed to minimize possible age-related bias in the classification results. However, the decision to exclude demographic variables from the feature set is in line with the methodology of Brown *et al.* (2022) and is supported by existing literature, such as the findings of Inbar *et al.* (2021), which demonstrate that machine learning models can effectively differentiate health states without relying on demographic factors.

In the preparation of data for use in classification algorithms, normalization plays an essential role in ensuring that all variables contribute equally to the decision function of the SVM. This procedure is particularly important when the data exhibit different units of measurement or scales. In this work, normalization was applied using the ‘Standardize’ option in the *fitsvm* function in Matlab. However, medical standard normalization (such as the 80–100% range of the predicted normal value), as performed by Inbar *et al.* (2021), was not implemented. This specific type of normalization may further enhance the algorithm’s effectiveness and is recommended for future work.

In Table 1, it can be seen that the variables *METS*, *RER*, *Vtex*, and *Vtin* were not included for multi-classification, as they were not provided in the dataset of healthy volunteers (Mongin, Romero & Cruz, 2021). However, specifically regarding the *RER* variable, it can be derived from the ratio of  $VCO_2$  to  $VO_2$ . To better understand how the inclusion of *RER* might affect the algorithm’s performance, it was incorporated into the model, and simulations for multi-classification with three levels of wavelets were conducted. The results showed that the algorithm’s performance was slightly worse. Indeed, the literature supports the notion that variable redundancy, especially through combinations, can degrade algorithm performance and increase the computational cost (Haq *et al.*, 2019).

Although this study does not present a detailed analysis of the physiological variables of CPET that are most relevant for a good differentiation between cohorts, the literature reveals some important indications in this regard. Analyzing some studies (Inbar *et al.*, 2021; Portella *et al.*, 2022; Zignoli, 2023), there is an intersection in variables such as  $VO_2$ ,  $VCO_2$ ,  $VE$  and  $HR$ , which are, therefore, of greater importance for classification. According to Table 1, these main variables are considered in the algorithms to extract features for binary classification and multi-class classification.

The algorithm proves useful in real-world scenarios, particularly for patients with HF with preserved ejection fraction who often experience exercise intolerance, a common reason for undergoing CPET. It aids in distinguishing between cardiac and metabolic causes of these symptoms. For MS, the algorithm is a non-invasive approach to early warning of a MS condition, since the standard determination of a MS condition requires clinical analyses with blood sampling focusing on some variables. As various conditions

can influence CPET variables, the algorithm will be further trained to classify additional and overlapping conditions. Future research will focus on expanding the dataset and incorporating data from the CPET of ciTechCare (the authors' institution) installed at Centro Hospitalar de Leiria (Portugal).

When it comes to choosing the best metric for medical diagnosis, sensitivity or recall (R) is highly relevant. False negatives can have serious consequences, for example in the detection of tumors, delaying necessary treatments (*Thölke et al., 2023; Spolaôr et al., 2024*). High sensitivity minimizes these errors, allowing for early detection and timely medical intervention, vital for favorable disease outcomes. Although in this study, the task is to identify syndromes early on, whose progression depends on several variables and their stage at the time, it is also essential to analyze them from the point of view of minimizing the number of false negatives.

In this work, the models performed poorly with the RBF kernel, which can be attributed to the linear separability characteristic of the data. The RBF kernel is effective for complex, non-linear patterns, but can be excessively complex and less efficient for data that is (quasi) linearly separable or can be separated with a low-degree polynomial. Furthermore, if the data is dispersed or lacks a clear structure, RBF can generate complicated and inefficient decision boundaries. Under these conditions, simpler kernels, such as linear or polynomial, can offer superior performance in identifying and separating the data (*Gopinath, Kumar & Ramachandran, 2018; Kumar, Shukla & Wadhvani, 2024*).

Another interesting issue to discuss is that multi-classification algorithms were developed based on the SVM methodology, despite the fact that, in binary classification, the best SVM model ranked second to the model proposed by *Brown et al. (2022)*, which combines autoencoders with logistic regression. In response to this, it can be argued that the SVM approach was preferred due to its widespread use (already with many paths in Matlab or Python) being an important factor in achieving the main objective of this work, which focuses on multi-classification. AE+LR, on the other hand, leverages artificial neural networks, which come with greater complexity and higher computational costs compared to SVM. Nevertheless, although the SVM approach with wavelets was adopted in this work, future studies could explore the performance of AE+LR as multi-classifiers on CPET data and compare those results with the findings of this study.

Cross-validation in this study was conducted without data shuffling or stratification, which introduces certain limitations. Without shuffling, inherent patterns or ordering within the dataset can skew the results, leading to subsets that may not accurately represent the overall problem, potentially resulting in misleading model evaluations, for example, different accuracies for shuffled and unshuffled (*Chakraborty & Sorwar, 2022*). Additionally, the absence of stratification can lead to imbalanced class distributions across the splits, especially in datasets with class imbalance, making it more challenging for the model to accurately detect minority classes (*Sadaiyandi et al., 2023*). In this work, the data are balanced, which can reduce the inefficiency that may be caused by the lack of stratification. However, for future work, a more rigorous stratification process could be used, for example, using Matlab's *cvpartition* function, which can be used in conjunction with the *crossval* function.

Although there are limitations, the research tackles these issues by replicating the validation process and averaging the results over five rounds of 5-fold cross-validation, so that some random shuffling can take place. In addition, the application of measurements like recall, precision and F1-score provides a more nuanced and extensive evaluation of the model's performance (Niaz, Shahariar & Patwary, 2022). This approach offers clearer insights into how effectively the model handles positive cases, false positives, false negatives, and overall class balance, even in the face of data distribution challenges.

A final point to be addressed in this article concerns the limitations of the dataset. The databases used in this study differ, suggesting that a uniform data collection protocol was likely not followed, and the CPET equipment varied across sources. This discrepancy could raise concerns if the method were to be applied directly to real-world scenarios without further validation on larger, more diverse datasets. However, at this stage, it is important to note that this work is best viewed as a pilot study. The investigation is still ongoing, and future efforts should focus on gathering more comprehensive datasets to refine the algorithms and enhance their reliability.

Models trained on small datasets often suffer performance drops due to overfitting and lack of sufficient samples for generalization (Sordo & Zeng, 2005; Prusa, Khoshgoftaar & Seliya, 2015; Rahman & Sultana, 2017). Also, as noted by Althnian et al. (2021), while SVM is relatively robust in such scenarios, relying on support vectors for hyperplane definition, its effectiveness also decreases with reduced data. Fewer samples may result in missing decisive support vectors, impacting the model's performance. Thus, despite being less sensitive than other models, SVM still experiences a decline in smaller datasets. To address this, validating on larger datasets is essential, as limited data may constrain the model's effectiveness. Enhancing generalization and reliability by utilizing larger datasets should be a focus of future research.

## CONCLUSION

This work presented algorithms to support the diagnosis of HF and MS with the option of classifying healthy people. The technique used, SVM with wavelets, proved to be effective for various models, with the SVM-POL-BW5 model excelling for the binary case and the SVM-LIN-MW3 model for the multi-class case. The results presented in this study are promising and motivating to continue this research with the aim of building a more comprehensive model to support the diagnosis of various diseases using CPET data according to Fig. 1.

In this sense, as suggestion for future research, it is proposed to search for more CPET databases with patients diagnosed with HF, MS and other diseases, in order to improve and expand the model proposed in this work. The aim in the future is to create a system capable of integrating existing CPET equipment in health centres, helping doctors to make faster diagnoses and thus improving people's quality of life.

Additionally, a promising direction for future research would be to investigate which CPET parameters aid in creating ideal feature sets to improve model performance. For example, Schwendinger et al. (2024) noted the importance of derived variables like oxygen pulse and ventilatory efficiency slopes in machine learning. However, these variables are

derived from other measures, such as oxygen pulse being the ratio of  $VO_2$  and  $HR$ . Therefore, an additional study focused on identifying an optimal set of CPET variables (similarly to what is done directly for features by *Haq et al. (2019)* and *Bezerra et al. (2024)*), combined with the wavelet-based feature extraction technique presented in this work, could lead to an innovative and efficient approach.

Going even further beyond the data provided by CPET, today there is data generated by patients themselves (smartphones, wearables, wristbands, sensor-equipped clothing, among others) that can be shared with healthcare professionals to feed these types of models. In this direction, a new line of research and development can be pursued for the establishment of increasingly personalized and rapid diagnostic support systems.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was funded by Portuguese national funds provided by the Portuguese Foundation for Science and Technology (FCT) (FCT-UIDB/05704/2020) and in the scope of the research project 2 ARTs (PTDC/EMD-EMD/6588/2020). Rafael F. Pinheiro was supported by FCT through the Institutional Scientific Employment Stimulus CEECINST/00060/2021. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Portuguese Foundation for Science and Technology (FCT): FCT-UIDB/05704/2020.

Assessing Autonomic Control in Cardiac Rehabilitation (2 ARTs): PTDC/EMD-EMD/6588/2020.

FCT through the Institutional Scientific Employment Stimulus: CEECINST/00060/2021.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Rafael F. Pinheiro conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Rui Fonseca-Pinto conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The CPET raw data are available at GitHub: <https://github.com/suchethassharma/CPET>.

The CPET data for Healthy (H) volunteers is available at Physionet: <https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1>.

The algorithm codes are available in the [Supplemental Files](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.2474#supplemental-information>.

## REFERENCES

- Allwein EL, Schapire RE, Singer Y. 2000. Reducing multiclass to binary: a unifying approach for margin classifiers. *Journal of Machine Learning Research* **1**(Dec):113–141 DOI [10.1162/15324430152733133](https://doi.org/10.1162/15324430152733133).
- Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris AB, Alzakari N, Abou Elwafa A, Kurdi H. 2021. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. *Applied Sciences* **11**(2):796 DOI [10.3390/app11020796](https://doi.org/10.3390/app11020796).
- Asgari S, Mehrnia A, Moussavi M. 2015. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Computers in Biology and Medicine* **60**:132–142 DOI [10.1016/j.compbiomed.2015.03.005](https://doi.org/10.1016/j.compbiomed.2015.03.005).
- Auger DA, Ghadimi S, Cai X, Reagan CE, Sun C, Abdi M, Cao JJ, Cheng JY, Ngai N, Scott AD, Ferreira PF, Oshinski JN, Emamifar N, Ennis DB, Loecher M, Liu Z-Q, Croisille P, Viallon M, Bilchick KC, Epstein FH. 2022. Reproducibility of global and segmental myocardial strain using cine DENSE at 3 t: a multicenter cardiovascular magnetic resonance study in healthy subjects and patients with heart disease. *Journal of Cardiovascular Magnetic Resonance* **24**(1):1–12 DOI [10.1186/s12968-022-00851-7](https://doi.org/10.1186/s12968-022-00851-7).
- Bezerra FE, Oliveira Neto GCD, Cervi GM, Francesconi Mazetto R, Faria AMD, Vido M, Lima GA, Araújo SAD, Sampaio M, Amorim M. 2024. Impacts of feature selection on predicting machine failures by machine learning algorithms. *Applied Sciences* **14**(8):3337 DOI [10.3390/app14083337](https://doi.org/10.3390/app14083337).
- Bilchick KC, Auger DA, Abdishektai M, Mathew R, Sohn M-W, Cai X, Sun C, Narayan A, Malhotra R, Darby A, Mangrum JM, Mehta N, Ferguson J, Mazimba S, Mason PK, Kramer CM, Levy WC, Epstein FH. 2020. CMR DENSE and the seattle heart failure model inform survival and arrhythmia risk after crt. *Cardiovascular Imaging* **13**(4):924–936 DOI [10.1016/j.jcmg.2019.10.017](https://doi.org/10.1016/j.jcmg.2019.10.017).
- Boser BE, Guyon IM, Vapnik VN. 1992. A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152.
- Bowen RE, Graetz TJ, Emmert DA, Avidan MS. 2020. Statistics of heart failure and mechanical circulatory support in 2020. *Annals of Translational Medicine* **8**(13):827 DOI [10.21037/atm-20-1127](https://doi.org/10.21037/atm-20-1127).
- Brown DE, Sharma S, Jablonski JA, Weltman A. 2022. Neural network methods for diagnosing patient conditions from cardiopulmonary exercise testing data. *BioData Mining* **15**(1):16 DOI [10.1186/s13040-022-00299-6](https://doi.org/10.1186/s13040-022-00299-6).
- Chakraborty PR, Sorwar G. 2022. A machine learning approach to identify fall risk for older adults. *Smart Health* **26**(10):100303 DOI [10.1016/j.smhl.2022.100303](https://doi.org/10.1016/j.smhl.2022.100303).
- Chen H, Wang N, Du X, Mei K, Zhou Y, Cai G. 2023. Classification prediction of breast cancer based on machine learning. *Computational Intelligence and Neuroscience* **2023**(1):6530719 DOI [10.1155/2023/6530719](https://doi.org/10.1155/2023/6530719).
- Dhivya P, Vasuki S. 2018. Wavelet based MRI brain image classification using radial basis function in svm. In: *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. Piscataway: IEEE, 1–9.

- Dietterich TG, Bakiri G. 1994.** Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research* 2:263–286 DOI 10.1613/jair.105.
- Epstein A, DiMarco J, Ellenbogen K, Estes N, Freedman R, Gettes L, Gillinov A, Gregoratos G, Hammill S, Hayes D, Hlatky M, Newby L, Page R, Schoenfeld M, Silka M, Stevenson L, Sweeney M, Tracy C, Darbar D, Dunbar S, Ferguson T, Karasik P, Link M, Marine J, Shanker A, Stevenson W, Varosy P. 2008.** ACC/AHA/HRS, 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (writing committee to revise the ACC/AHA/NASPE, 2002 guideline update for implantation of cardiac pacemakers and antiarrhythmia devices) developed in collaboration with the american association for thoracic surgery and society of thoracic surgeons. *Journal of the American College of Cardiology* 51(21):e1–e62 DOI 10.1016/j.jacc.2008.02.032.
- Gaitán JM, Eichner NZM, Gilbertson NM, Heiston EM, Weltman A, Malin SK. 2019.** Two weeks of interval training enhances fat oxidation during exercise in obese adults with prediabetes. *Journal of Sports Science & Medicine* 18(4):636–644.
- Gao X, Abdi M, Auger DA, Sun C, Hanson CA, Robinson AA, Schumann C, Oomen PJ, Ratcliffe S, Malhotra R, Darby A, Monfredi OJ, Mangrum JM, Mason P, Mazimba S, Holmes JW, Kramer CM, Epstein FH, Salerno M, Bilchick KC. 2021.** Cardiac magnetic resonance assessment of response to cardiac resynchronization therapy and programming strategies. *Cardiovascular Imaging* 14(12):2369–2383 DOI 10.1016/j.jcmg.2021.06.015.
- Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. 2000.** Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220 DOI 10.1161/01.CIR.101.23.e215.
- Gopinath R, Kumar CS, Ramachandran R. 2018.** Scalable fault models for diagnosis in a synchronous generator using feature mapping and transformation techniques. *International Journal of Prognostics and Health Management* 9(2) DOI 10.36001/ijphm.2018.v9i2.2737.
- Hansen D, Bonne K, Alders T, Hermans A, Copermans K, Swinnen H, Maris V, Jansegers T, Mathijs W, Haenen L, Vaes J, Govaerts E, Reenaers V, Frederix I, Dendale P. 2019.** Exercise training intensity determination in cardiovascular rehabilitation: should the guidelines be reconsidered? *European Journal of Preventive Cardiology* 26(18):1921–1928 DOI 10.1177/2047487319859450.
- Haq AU, Zhang D, Peng H, Rahman SU. 2019.** Combining multiple feature-ranking techniques and clustering of variables for feature selection. *IEEE Access* 7:151482–151492 DOI 10.1109/ACCESS.2019.2947701.
- Heiston EM, Eichner NZ, Gilbertson NM, Gaitán JM, Kranz S, Weltman A, Malin SK. 2019.** Two weeks of exercise training intensity on appetite regulation in obese adults with prediabetes. *Journal of Applied Physiology* 126(3):746–754 DOI 10.1152/jappphysiol.00655.2018.
- Inbar O, Inbar O, Reuveny R, Segel MJ, Greenspan H, Scheinowitz M. 2021.** A machine learning approach to the interpretation of cardiopulmonary exercise tests: development and validation. *Pulmonary Medicine* 2021(2):1–9 DOI 10.1155/2021/5516248.
- Iniyani S, Singh A, Hazra B. 2023.** Wavelet transformation and vertical stacking based image classification applying machine learning. *Biomedical Signal Processing and Control* 79:104103 DOI 10.1016/j.bspc.2022.104103.
- Kaminsky LA, Imboden MT, Arena R, Myers J. 2017.** Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing using cycle ergometry: data from the

- fitness registry and the importance of exercise national database (friend) registry. *Mayo Clinic Proceedings* **92**:228–233 Elsevier DOI [10.1016/j.mayocp.2016.10.003](https://doi.org/10.1016/j.mayocp.2016.10.003).
- Kehri V, Awale R. 2018.** Emg signal analysis for diagnosis of muscular dystrophy using wavelet transform, SVM and ANN. *Biomedical and Pharmacology Journal* **11**(3):1583–1591 DOI [10.13005/bpj/1525](https://doi.org/10.13005/bpj/1525).
- Kumar L, Shukla S, Wadhvani R. 2024.** Analysis of feature noise on standard svm with linear kernel. In: *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. Piscataway: IEEE, 1–6.
- Li X, Zhai Y, Zhao J, He H, Li Y, Liu Y, Feng A, Li L, Huang T, Xu A, Lyu J. 2021.** Impact of metabolic syndrome and its components on prognosis in patients with cardiovascular diseases: a meta-analysis. *Frontiers in Cardiovascular Medicine* **8**:704145 DOI [10.3389/fcvm.2021.704145](https://doi.org/10.3389/fcvm.2021.704145).
- Luo Q, Yu X, Zhao Z, Zhao Q, Ma X, Jin Q, Yan L, Zhang Y, Liu Z. 2021.** The value of cardiopulmonary exercise testing in the diagnosis of pulmonary hypertension. *Journal of Thoracic Disease* **13**(1):178–188 DOI [10.21037/jtd-20-1061b](https://doi.org/10.21037/jtd-20-1061b).
- Malin SK, Gilbertson NM, Eichner NZ, Heiston E, Miller S, Weltman A. 2019.** Impact of short-term continuous and interval exercise training on endothelial function and glucose metabolism in prediabetes. *Journal of Diabetes Research* **2019**(9179):1–8 DOI [10.1155/2019/4912174](https://doi.org/10.1155/2019/4912174).
- Michelucci U, Sperti M, Piga D, Venturini F, Deriu MA. 2021.** A model-agnostic algorithm for bayes error determination in binary classification. *Algorithms* **14**(11):301 DOI [10.3390/a14110301](https://doi.org/10.3390/a14110301).
- Mongin D, Chabert C, Courvoisier DS, García-Romero J, Alvero-Cruz JR. 2021.** Heart rate recovery to assess fitness: comparison of different calculation methods in a large cross-sectional study. *Research in Sports Medicine* **31**(2):157–170 DOI [10.1080/15438627.2021.1954513](https://doi.org/10.1080/15438627.2021.1954513).
- Mongin D, Romero JG, Cruz JRA. 2021.** Treadmill maximal exercise tests from the exercise physiology and human performance lab of the university of malaga. *Physionet*, version 1.0.1. DOI [10.13026/7ezk-j442](https://doi.org/10.13026/7ezk-j442).
- Morales L, Aguilar J, Chávez D, Isaza C. 2020.** Lamda-had, an extension to the lamda classifier in the context of supervised learning. *International Journal of Information Technology & Decision Making* **19**(01):283–316 DOI [10.1142/S0219622019500457](https://doi.org/10.1142/S0219622019500457).
- Niaz NU, Shahariar KN, Patwary MJ. 2022.** Class imbalance problems in machine learning: a review of methods and future challenges. In: *Proceedings of the 2nd International Conference on Computing Advancements*, 485–490.
- Noubiap JJ, Nansseu JR, Lontchi-Yimagou E, Nkeck JR, Nyaga UF, Ngouo AT, Tounouga DN, Tianyi F-L, Foka AJ, Ndoadoumgue AL, Bigna JJ. 2022.** Geographic distribution of metabolic syndrome and its components in the general adult population: a meta-analysis of global data from 28 million individuals. *Diabetes Research and Clinical Practice* **188**(4):109924 DOI [10.1016/j.diabres.2022.109924](https://doi.org/10.1016/j.diabres.2022.109924).
- Pinheiro RF, Fonseca-Pinto R. 2023.** Algorithm for diagnosis of metabolic syndrome and heart failure using cpet biosignals via svm and wavelet transforms. In: *International Conference on Smart Objects and Technologies for Social Good*. Cham: Springer, 159–171.
- Portella JJ, Andonian BJ, Brown DE, Mansur J, Wales D, West VL, Kraus WE, Hammond WE. 2022.** Using machine learning to identify organ system specific limitations to exercise via cardiopulmonary exercise testing. *IEEE Journal of Biomedical and Health Informatics* **26**(8):4228–4237 DOI [10.1109/JBHI.2022.3163402](https://doi.org/10.1109/JBHI.2022.3163402).
- Prusa J, Khoshgoftaar TM, Seliya N. 2015.** The effect of dataset size on training tweet sentiment classifiers. In: *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. Piscataway: IEEE, 96–102.

- Purwowiyoto SL, Prawara AS. 2021.** Metabolic syndrome and heart failure: mechanism and management. *Medicine and Pharmacy Reports* **94(1)**:15 DOI [10.15386/mpr-1884](https://doi.org/10.15386/mpr-1884).
- Rahman MS, Sultana M. 2017.** Performance of fifth- and log<sub>f</sub>-type penalized methods in risk prediction for small or sparse binary data. *BMC Medical Research Methodology* **17(1)**:1–15 DOI [10.1186/s12874-017-0313-9](https://doi.org/10.1186/s12874-017-0313-9).
- Rodriguez JC, Peterman JE, Fleenor BS, Whaley MH, Kaminsky LA, Harber MP. 2022.** Cardiopulmonary exercise responses in individuals with metabolic syndrome: the ball state adult fitness longitudinal lifestyle study. *Metabolic Syndrome and Related Disorders* **20(7)**:414–420 DOI [10.1089/met.2021.0130](https://doi.org/10.1089/met.2021.0130).
- Sadaiyandi J, Arumugam P, Sangaiah AK, Zhang C. 2023.** Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset. *Electronics* **12(21)**:4423 DOI [10.3390/electronics12214423](https://doi.org/10.3390/electronics12214423).
- Saito Y, Obokata M, Harada T, Kagami K, Murata M, Sorimachi H, Kato T, Wada N, Okumura Y, Ishii H. 2023.** Diagnostic value of expired gas analysis in heart failure with preserved ejection fraction. *Scientific Reports* **13(1)**:4355 DOI [10.1038/s41598-023-31381-6](https://doi.org/10.1038/s41598-023-31381-6).
- Schwendinger F, Biehler A-K, Nagy-Huber M, Knaier R, Roth V, Dumitrescu D, Meyer FJ, Hager A, Schmidt-Trucksäss A. 2024.** Using machine learning-based algorithms to identify and quantify exercise limitations in clinical practice: are we there yet? *Medicine & Science in Sports & Exercise* **56(2)**:159–169 DOI [10.1249/MSS.0000000000003293](https://doi.org/10.1249/MSS.0000000000003293).
- Serhal H, Abdallah N, Marion J-M, Chauvet P, Oueidat M, Humeau-Heurtier A. 2022.** Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG. *Computers in Biology and Medicine* **142(3)**:105168 DOI [10.1016/j.combiomed.2021.105168](https://doi.org/10.1016/j.combiomed.2021.105168).
- Sordo M, Zeng Q. 2005.** On sample size and classification accuracy: a performance comparison. In: *International Symposium on Biological and Medical Data Analysis*. Cham: Springer, 193–201.
- Spolaôr N, Lee HD, Mendes AI, Nogueira CV, Parmezan ARS, Takaki WSR, Coy CSR, Wu FC, Fonseca-Pinto R. 2024.** Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets. *Multimedia Tools and Applications* **83(9)**:27305–27329 DOI [10.1007/s11042-023-16529-w](https://doi.org/10.1007/s11042-023-16529-w).
- Subasi A. 2007.** Eeg signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications* **32(4)**:1084–1093 DOI [10.1016/j.eswa.2006.02.005](https://doi.org/10.1016/j.eswa.2006.02.005).
- Thölke P, Mantilla-Ramos Y-J, Abdelhedi H, Maschke C, Dehgan A, Harel Y, Kemtur A, Berrada LM, Sahraoui M, Young T, Pépin AB, Khantour CE, Landry M, Pascarella A, Hadid V, Combrisson E, O’Byrne J, Jerbi K. 2023.** Class imbalance should not throw you off balance: choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage* **277**:120253 DOI [10.1016/j.neuroimage.2023.120253](https://doi.org/10.1016/j.neuroimage.2023.120253).
- Wong T-T, Yeh P-Y. 2019.** Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering* **32(8)**:1586–1594 DOI [10.1109/TKDE.2019.2912815](https://doi.org/10.1109/TKDE.2019.2912815).
- Xin Y, Zhao Y. 2017.** Paroxysmal atrial fibrillation recognition based on multi-scale wavelet  $\alpha$ -entropy. *Biomedical Engineering Online* **16**:1–12 DOI [10.1186/s12938-017-0406-z](https://doi.org/10.1186/s12938-017-0406-z).
- Xing Z, Pei J, Yu PS, Wang K. 2011.** Extracting interpretable features for early classification on time series. In: *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 247–258.

- Zhou Y, Fu L, Sun J, Zhu Z, Xing Z, Zhou S, Tai S, Wang Y. 2021.** Association between metabolic syndrome and an increased risk of hospitalization for heart failure in population of HFpEF. *Frontiers in Cardiovascular Medicine* **8**:698117 DOI [10.3389/fcvm.2021.698117](https://doi.org/10.3389/fcvm.2021.698117).
- Zignoli A. 2023.** Machine learning models for the automatic detection of exercise thresholds in cardiopulmonary exercising tests: from regression to generation to explanation. *Sensors* **23**(2):826 DOI [10.3390/s23020826](https://doi.org/10.3390/s23020826).