escola superior de tecnologia e gestão
instituto politécnico de leiria

Algoritmos de Classificação Multiclasse via SVM para Apoio ao Diagnóstico a partir de Dados do Teste de Esforço Cardiopulmonar (CPET)

Flávio Bueno dos Santos

Escola Superior de Tecnologia e Gestão
Departamento de Engenharia Informática
Mestrado em Ciência de Dados

Leiria, Novembro 2025

Página intencionalmente deixada em branco.

escola superior de tecnologia e gestão
instituto politécnico de leiria

Algoritmos de Classificação Multiclasse via SVM para Apoio ao Diagnóstico a partir de Dados do Teste de Esforço Cardiopulmonar (CPET)

Flávio Bueno dos Santos

Supervisor: Professor Rafael Fernandes Pinheiro
Professor, Politécnico de Leiria

Co-supervisor: Professor Rui Fonseca Pinto
Professor, Politécnico de Leiria

Escola Superior de Tecnologia e Gestão
Departamento de Engenharia Informática
Mestrado em Ciência de Dados

Dissertação

Leiria, Novembro 2025

Página intencionalmente deixada em branco.

Algoritmos de Classificação Multiclasse via SVM para Apoio ao Diagnóstico a partir de Dados do Teste de Esforço Cardiopulmonar (CPET)

Copyright © 2025 - Flávio Bueno dos Santos, Escola Superior de Tecnologia e Gestão.

A presente dissertação é um trabalho original, elaborado exclusivamente para este fim, tendo sido devidamente citados todos os autores cujos estudos contribuíram para a sua elaboração. É permitida a sua reprodução parcial com indicação do autor e referência ao grau, ano letivo, instituição - *Politécnico de Leiria* - e data da defesa pública.

Página intencionalmente deixada em branco.

Agradecimentos

Gostaria de expressar os meus sinceros agradecimentos a todos os que contribuíram para a realização deste trabalho. Agradeço especialmente ao meu orientador, Professor Rafael Pinheiro, pela orientação técnica, paciência e pelo apoio constante ao longo da pesquisa. Ao meu coorientador, Professor Rui Fonseca Pinto, agradeço pelas suas valiosas sugestões e pela perspetiva enriquecedora que trouxe a este projeto. Agradeço à minha colega Clara Neves pela valiosa colaboração na etapa inicial da revisão bibliográfica. À memória da minha mãe, minha eterna inspiração. A toda a minha família, a minha mais profunda gratidão: ao meu pai, pelo amor e incentivo, e aos meus irmãos, pelo apoio e união. Por fim, agradeço aos meus amigos pelo companheirismo ao longo desta jornada.

Página intencionalmente deixada em branco.

Resumo

O Teste de Esforço Cardiopulmonar (*CPET*) é uma ferramenta vital para o diagnóstico funcional, mas a aplicação de modelos de aprendizagem automática é frequentemente limitada pela escassez de séries temporais completas para diversas condições clínicas. Esta dissertação aborda esta lacuna, partindo de um *framework* de classificação baseado em *Support Vector Machines* (*SVM*) e na Transformada *Wavelet* Discreta (*DWT*), originalmente desenvolvido para três classes (Insuficiência Cardíaca (IC), Síndrome Metabólica (SM) e Saudáveis (S)).

O objetivo central foi expandir esta metodologia para um cenário mais complexo de cinco classes, através da geração de Dados Semi-Sintéticos para os grupos de Limitação Pulmonar (LP) e Limitação Musculoesquelética (LM), guiada por parâmetros estatísticos de pacientes reais. Subsequentemente, para validar a eficácia da Transformada Wavelet Discreta (*Discrete Wavelet Transform*) (DWT) neste novo contexto, foi conduzida uma análise comparativa, avaliando o desempenho do modelo contra três métodos alternativos de extração de características: a Transformada de Fourier de Curto Tempo (*Short-Time Fourier Transform*) (STFT), a Transformada *Wavelet* por Pacotes (*Wavelet Packet Transform*) (WPT) e a Decomposição por Modos Empíricos (*Empirical Mode Decomposition*) (EMD). Todos os modelos foram avaliados sob um protocolo experimental consistente para garantir uma comparação justa.

Os resultados da análise comparativa foram consistentes. O modelo `SVM-Linear-MW5`, que utiliza a DWT, alcançou uma acurácia de 93.60% e um *F1-Score* de 84.14%, um desempenho que se destacou em relação ao das outras transformadas. A análise demonstrou que a STFT foi a alternativa mais competitiva (*F1-Score* de 74.25%), enquanto a WPT e a EMD não se mostraram tão eficazes para este problema.

Este trabalho conclui que a combinação de dados semi-sintéticos com a extração de características via DWT é uma abordagem viável para a expansão de modelos de diagnóstico. A metodologia de referência foi expandida para cinco classes e, na análise comparativa realizada, a sua abordagem de processamento de sinal obteve o desempenho mais elevado entre as técnicas testadas, o que estabelece um *baseline* sólido para futuras investigações na área, incluindo a otimização de Hiperparâmetros.

Palavras-Chave: Teste de Esforço Cardiopulmonar (CPET), Support Vector Machines (SVM), Transformada *Wavelet*, Geração de Dados Semi-Sintéticos, Séries Temporais Fisiológicas, Decisão Clínica.

Página intencionalmente deixada em branco.

Abstract

Cardiopulmonary Exercise Testing (CPET) is a vital tool for functional diagnosis, yet the application of machine learning models is often limited by the scarcity of complete time-series across clinical conditions. This dissertation addresses that gap by extending a classification framework based on Support Vector Machines (SVM) and the Discrete Wavelet Transform (DWT), originally developed for three classes (Heart Failure, Metabolic Syndrome, and Healthy).

The central goal was to scale this methodology to a five-class scenario by generating semi-synthetic data for Pulmonary Limitation and Musculoskeletal Limitation, guided by summary statistics from real patients. To assess the effectiveness of DWT in this expanded setting, a comparative study was conducted against three alternative feature-extraction methods: Short-Time Fourier Transform (STFT), Wavelet Packet Transform (WPT), and Empirical Mode Decomposition (EMD). All models were evaluated under a consistent experimental protocol to ensure fair comparison.

Results were consistent: the SVM-Linear-MW5 model, using DWT-based features, achieved 93.60% accuracy and an F1-score of 84.14%, outperforming the alternatives. STFT was the most competitive contender (F1-score of 74.25%), while WPT and EMD were less effective for this problem.

This work concludes that combining semi-synthetic data with DWT-based feature extraction is a viable path to expand diagnostic models. The reference methodology was successfully extended to five classes and set a strong baseline for future research, including hyperparameter optimisation.

Keywords: Cardiopulmonary Exercise Testing (CPET), Support Vector Machines (SVM), Wavelet Transform, Semi-synthetic Data Generation, Physiological Time Series, Clinical Decision-Making.

Página intencionalmente deixada em branco.

Conteúdo

Lista de Figuras xiv

Lista de Tabelas xvii

Glossário xx

Lista de siglas xxiv

1 Introdução **1**

1.1 Contexto e Motivação 1

1.2 Objetivos e Contribuições 3

1.3 Organização do Documento 4

2 Revisão da Literatura **5**

2.1 Metodologia da Revisão (PRISMA) 5

2.1.1 Fontes de Pesquisa 5

2.1.2 Estratégia de Busca 6

2.1.3 Critérios de Inclusão e Exclusão 7

2.1.4 Processo de Seleção e Diagrama PRISMA 7

2.2 Análise Crítica dos Estudos Fundamentais 10

2.3 Fundamentação Teórica 11

2.3.1 O Teste Cardiopulmonar de Esforço (CPET) em Detalhe 11

2.3.2 Técnicas de Processamento de Sinal para Análise de Dados Fisiológicos 13

2.3.3 Aprendizagem Automática para Classificação Supervisionada . 16

2.4 Síntese e Análise Crítica 17

2.4.1 Panorama da Literatura 18

2.4.2 Lacunas Identificadas e Oportunidades de Investigação 18

2.4.3 Justificação da Abordagem da Dissertação 19

2.5 Síntese e Transição para os Métodos 20

3 Materiais e Métodos **21**

3.1 Descrição do *Dataset* 21

3.2 Geração de Dados Semi-Sintéticos 23

3.3 Extração de Características 24

3.3.1 Pré-processamento e Controlo de Fuga de Informação 24

3.3.2 Justificação da Escolha dos Métodos de Extração 24
3.3.3 Transformada Wavelet Discreta (DWT) 25
3.3.4 Transformada de Fourier de Curto Tempo (STFT) 26
3.3.5 Wavelet Packet Transform (WPT) 26
3.3.6 Decomposição por Modos Empíricos (EMD) 26
3.4 Metodologia de Classificação e Avaliação 26
3.4.1 Codificação de Rótulos e Dimensões da Matriz de Características 26
3.4.2 Máquinas de Vetores de Suporte (SVM) 27
3.5 Protocolo de Avaliação 27
3.6 Validação dos Sinais Semi-Sintéticos 27

4 Resultados e Discussão **29**
4.1 Resultados Experimentais 29
4.1.1 Desempenho Detalhado do Modelo DWT 29
4.1.2 Análise Comparativa dos Métodos de Extração 30
4.2 Discussão 30
4.3 Limitações do Estudo 32
4.4 Conclusões e Trabalhos Futuros 32

5 Conclusões **34**
5.1 Conclusões 34

Referências 37

Apêndices

A Códigos-Fonte Principais **45**
A.1 Geração e Validação de Dados Semi-Sintéticos 45
A.2 Extração de Características e Avaliação Comparativa 45

B Artigo Aceite para Publicação **49**

Página intencionalmente deixada em branco.

Lista de Figuras

1.1 Visão geral da metodologia proposta. O trabalho expande um algoritmo de diagnóstico baseado em *SVM* e *Wavelets* para classificar cinco grupos distintos a partir de dados de *CPET*, integrando dados reais e semi-sintéticos para as classes de limitação pulmonar e musculoesquelética. Fonte: Adaptado de Pinheiro et al., 2025. 3

2.1 Fluxograma *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020 ilustrando o processo de identificação, triagem, elegibilidade e inclusão dos estudos. Modelo adaptado de Page et al. (2021). Template licenciado sob CC BY 4.0. 8

3.1 Fluxograma da metodologia experimental. O processo inicia-se com a construção do *dataset* híbrido, seguido por uma análise comparativa de quatro métodos de extração de características. Cada método é avaliado sob um protocolo consistente para garantir uma comparação justa, culminando na identificação do modelo com melhor desempenho. 22

4.1 Comparação da *Acurácia* (Macro) para o melhor modelo de cada método de extração de características. Os resultados destacam o desempenho da abordagem baseada em DWT em comparação com as alternativas. Fonte: Elaborado pelo autor. 31

Página intencionalmente deixada em branco.

Lista de Tabelas

2.1 Resumo dos estudos-chave da revisão 10
2.2 Resumo comparativo de algoritmos aplicáveis a dados de CPET. 19

4.1 Métricas de desempenho detalhadas para o modelo `SVM-Linear-DWT-MW5`, apresentadas por classe. Estes resultados correspondem ao melhor modelo identificado no artigo (Santos et al., 2025). 30
4.2 Resultados comparativos de desempenho para os quatro métodos de extração de características. Conforme a metodologia, os resultados reportados correspondem ao desempenho obtido no protocolo de validação. A linha a negrito indica o modelo com o melhor desempenho geral. 30

Página intencionalmente deixada em branco.

Glossário

Análise Multirresolução Técnica de processamento de sinal que decompõe um sinal em diferentes escalas de tempo e frequência. Permite analisar tanto eventos rápidos e de alta frequência como tendências lentas e de baixa frequência, sendo a principal vantagem da Transformada *Wavelet*. 15, 25

Dados Semi-Sintéticos Dados gerados artificialmente, mas guiados por parâmetros estatísticos (como média, desvio padrão e tendências) extraídos de um conjunto de dados reais. São usados para aumentar o volume de dados para treino, especialmente quando os dados reais são escassos, como foi o caso para duas das classes neste estudo. iv, 2

F1-Score Uma métrica de avaliação para modelos de classificação que calcula a média harmónica entre a Precisão (*Precision*) e a Sensibilidade (*Recall*). É particularmente útil para avaliar o desempenho em *datasets* com classes desbalanceadas. iv, 27, 30, 31, 34

Hiperparâmetros Parâmetros de um modelo de aprendizagem automática que não são aprendidos a partir dos dados, mas sim configurados antes do início do processo de treino. iv, 32

Hiperplano Em geometria, é uma generalização de um plano para mais de três dimensões. No contexto das *SVM*, é a fronteira de decisão (uma linha, um plano ou um hiperplano) que o algoritmo encontra para separar as diferentes classes de dados (por exemplo, para separar os pacientes saudáveis dos que têm insuficiência cardíaca). 27

Teste de Esforço Cardiopulmonar (*CPET*) Um teste de diagnóstico não invasivo que mede as respostas dos sistemas cardiovascular e respiratório ao exercício. É a fonte principal de dados utilizada nesta dissertação. iv

Transformada *Wavelet* Discreta (*DWT*) Uma ferramenta matemática para decompor um sinal (como os sinais do *CPET*) em diferentes níveis de frequência, preservando a informação temporal. É a base da metodologia de extração de características deste trabalho. iv, 1

Validação Cruzada (*Cross-validation*) Uma técnica robusta para avaliar a capacidade de generalização de um modelo. O processo consiste em dividir o *dataset* em vá-

rias partes (neste trabalho, 5 partes ou "*folds*"), usando iterativamente uma parte para teste e as restantes para treino, e no final calcular a média dos resultados. 27

Página intencionalmente deixada em branco.

Lista de siglas

cA coeficientes de aproximação. 15, 25
cD coeficientes de detalhe. 15, 25
CNN *Convolutional Neural Networks*. 10, 18, 32
CPET Teste de Esforço Cardiopulmonar. 1, 3, 5–7, 9–14, 17–21, 24, 25, 30, 32

DWT Transformada Wavelet Discreta (*Discrete Wavelet Transform*). iv, xiv, 2, 3, 10, 14–16, 18, 19, 24–26, 29–32, 34

ECG Eletrocardiograma. 11
EMD Decomposição por Modos Empíricos (*Empirical Mode Decomposition*). iv, 2, 3, 14, 16, 19, 21, 24–26, 29–32, 34
EN *Elastic Net*. 19

FFT Transformada Rápida de Fourier (*Fast Fourier Transform*). 14, 26

GAN *Generative Adversarial Network*. 32

HMM *Hidden Markov Model*. 19
HR Frequência Cardíaca (*Heart Rate*). 24

IC Insuficiência Cardíaca. iv, 1, 3, 10–12, 18, 19, 21, 26, 27, 29, 30, 34
IMF Função de Modo Intrínseco. 16, 26

kNN *k-Nearest Neighbors*. 19

LA Limiar Anaeróbio. 12, 19
LIME *Local Interpretable Model-agnostic Explanations*. 33
LM Limitação Musculoesquelética. iv, 2, 3, 21, 23, 26, 29, 30, 32, 34
LP Limitação Pulmonar. iv, 2, 3, 21, 23, 26, 29, 30, 32, 34
LSTM Long Short-Term Memory. 19

MW3 Wavelet Multirresolução – Nível 3. 25, 27, 29
MW5 Wavelet Multirresolução – Nível 5. 25, 27, 29, 31, 34

PCA Análise de Componentes Principais (*Principal Component Analysis*). 18, 24
PCR Ponto de Compensação Respiratória. 12, 19
PICO *Population, Intervention, Comparison, and Outcome*. 6

PRISMA *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*. xiv, 5, 7, 8, 10

RBF Função de Base Radial (*Radial Basis Function*). 17, 27, 29
RER Rácio de Trocas Gasosas (*Respiratory Exchange Ratio*, RER). 12, 21, 24, 25, 28, 32
RNN *Recurrent Neural Network*. 32
RR Frequência Respiratória (*Respiratory Rate*, RR). 24, 32

S Saudáveis. iv, 1–3, 18, 21, 26, 27, 29, 30, 34
SHAP *SHapley Additive exPlanations*. 33
SM Síndrome Metabólica. iv, 1, 3, 10, 18, 21, 26, 27, 29, 30, 34
STFT Transformada de Fourier de Curto Tempo (*Short-Time Fourier Transform*). iv, 2, 3, 14–16, 19, 21, 24–26, 29–32, 34
SVM *Support Vector Machine*. 1, 3, 10, 17, 18, 24, 27, 29, 34

VAE *Variational Autoencoder*. 32
VCO_2 Produção de Dióxido de Carbono. 11, 12, 18, 21, 24
VE Ventilação. 18, 24
VO_2 Consumo de Oxigénio. 11, 12, 18, 21, 24

WPT Transformada *Wavelet* por Pacotes (*Wavelet Packet Transform*). iv, 2, 3, 14, 15, 19, 21, 24–26, 29–32, 34

Página intencionalmente deixada em branco.

1 Introdução

Este capítulo estabelece as bases da presente dissertação. Inicia-se com a contextualização do problema, destacando a relevância clínica do Teste de Esforço Cardiopulmonar (CPET) e as limitações que motivaram esta investigação. De seguida, são formalmente apresentados os objetivos e as principais contribuições científicas do trabalho. Por fim, é descrita a organização estrutural dos restantes capítulos do documento.

1.1 Contexto e Motivação

O Teste de Esforço Cardiopulmonar (CPET) transcende o simples teste de esforço, atuando como uma ferramenta diagnóstica não invasiva fundamental na cardiologia, pneumologia e reabilitação. A sua capacidade de fornecer uma avaliação integrada e funcional dos sistemas cardiovascular, respiratório e musculoesquelético permite a identificação precoce de patologias e a estratificação de risco de pacientes com condições complexas, como a Insuficiência Cardíaca (IC) e a Síndrome Metabólica (SM) (Wasserman et al., 2011). No entanto, a interpretação clínica tradicional, que se baseia em parâmetros de pico ou em limiares fisiológicos, frequentemente negligencia a riqueza da informação contida na dinâmica temporal completa dos sinais, o que leva à perda de padrões subtis e não lineares que são essenciais para um diagnóstico mais preciso.

Nos últimos anos, a aplicação de técnicas de aprendizagem automática tem emergido como uma promissora solução para extrair essa informação latente e desenvolver sistemas de apoio ao diagnóstico mais objetivos (Brown et al., 2022; Al Rahhal et al., 2016). Um trabalho de referência, que serve de alicerce para esta dissertação, já havia demonstrado o potencial de um *framework* inovador que combina a Transformada *Wavelet* Discreta (*DWT*) para extração de características e as *Support Vector Machine* (SVM) para classificar três grupos diagnósticos: IC, SM e Saudáveis (S) (Pinheiro et al., 2023; Pinheiro et al., 2025).

Contudo, este avanço inicial levantou questões importantes que motivam a presente investigação:

1. A escassez de dados como barreira ao diagnóstico multiclasse: A principal barreira para a expansão do modelo original era a dificuldade em obter *datasets* públicos com séries temporais completas para outras classes clinicamente relevantes, como limitações pulmonares (LP) e musculoesqueléticas (LM). A geração de Dados Semi-Sintéticos, guiada por parâmetros estatísticos de pacientes reais (como média, desvio padrão e inclinações), surge, portanto, como uma estratégia viável para superar esta barreira, permitindo a expansão do modelo para novas categorias de diagnóstico (Pezoulas et al., 2024).
2. Inclusão de variáveis fisiológicas adicionais: O estudo de base, devido à indisponibilidade de dados no *dataset* do grupo S, não pôde incluir a Razão de Troca Respiratória (RER) na sua análise. Este trabalho complementa esta investigação ao incorporar a variável RER na *pipeline* de análise, utilizando tanto valores reais como derivados (calculados como a razão entre VCO_2 e VO_2). Esta inclusão permite uma avaliação mais completa e abrangente da contribuição do RER para o desempenho do modelo.
3. A necessidade de validar a escolha da transformada para um problema mais complexo: O trabalho de referência demonstrou a eficácia da DWT para um problema de três classes. No entanto, ao expandir o modelo para um cenário de cinco classes, mais desafiador e com dados de natureza híbrida, torna-se fundamental verificar se a DWT continua a ser a abordagem mais adequada. Uma validação rigorosa requer uma análise comparativa com outras técnicas consagradas de extração de características, como a Transformada de Fourier de Curto Tempo (STFT), a Transformada *Wavelet* por Pacotes (WPT) e a Decomposição por Modos Empíricos (EMD), para garantir que a escolha metodológica se mantém como a mais robusta.

> ***Nota***
>
> Realizar uma análise comparativa com outras técnicas consagradas é uma boa prática na investigação em *machine learning*. Isso não só valida a escolha metodológica, mas também fortalece a robustez dos resultados, demonstrando que a abordagem selecionada é genuinamente a mais adequada para o problema em questão (Demšar, 2006).

Esta dissertação aborda diretamente estas questões, propondo a expansão do modelo para cinco classes através de uma metodologia de geração de dados semi-sintéticos e, posteriormente, conduz uma análise comparativa rigorosa para determinar se a DWT continua a ser a abordagem mais adequada de extração de características neste novo contexto.

1.2 Objetivos e Contribuições

O objetivo principal deste trabalho é expandir e validar um sistema de apoio ao diagnóstico multiclasse, ampliando um *framework* SVM-*Wavelet* existente para novas categorias clínicas e confirmando a sua eficácia em relação a outras abordagens de processamento de sinal.

As principais contribuições científicas e tecnológicas desta dissertação são:

- A expansão de um modelo de classificação para cinco grupos diagnósticos (IC, SM, S, LP e LM) através da construção de um *dataset* híbrido, combinando dados reais e semi-sintéticos.
- A geração fisiologicamente guiada de séries temporais de CPET semi-sintéticas, baseada em estatísticas derivadas de pacientes reais, como método para enriquecer *datasets* clínicos.
- A inclusão e avaliação da variável RER, tanto medida como derivada, para uma avaliação comparativa completa do seu impacto no desempenho do modelo de diagnóstico.
- A validação da DWT para esta tarefa, cuja análise comparativa demonstrou o desempenho mais elevado em relação à STFT, WPT e EMD.
- Uma análise comparativa sistemática entre quatro diferentes transformadas de sinal (DWT, STFT, WPT, EMD), conduzida sob um protocolo experimental unificado para esta aplicação.

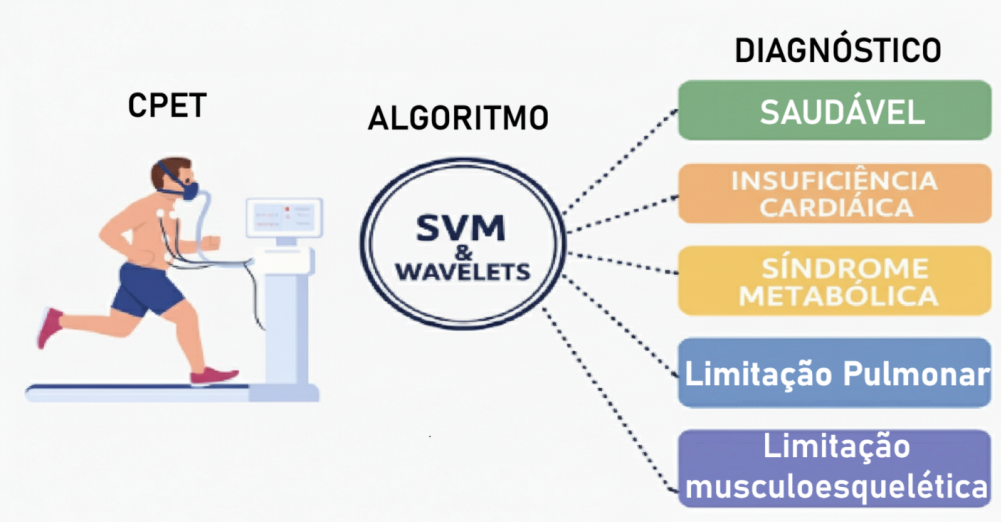

Figura 1.1: *Visão geral da metodologia proposta. O trabalho expande um algoritmo de diagnóstico baseado em SVM e Wavelets para classificar cinco grupos distintos a partir de dados de CPET, integrando dados reais e semi-sintéticos para as classes de limitação pulmonar e musculoesquelética. Fonte: Adaptado de Pinheiro et al., 2025.*

1.3 Organização do Documento

O restante do documento está estruturado da seguinte forma: O Capítulo 2 apresenta uma revisão sistemática da literatura sobre *datasets* de CPET. O Capítulo 3 detalha os materiais e métodos, incluindo a descrição do *dataset*, a metodologia de geração de dados semi-sintéticos e a abordagem de classificação. O Capítulo 4 apresenta e discute os resultados obtidos. Finalmente, o Capítulo 5 apresenta as conclusões do trabalho e aponta direções para investigação futura.

Por fim, os códigos-fonte principais desenvolvidos e o artigo resultante deste trabalho encontram-se, respetivamente, nos Apêndices A e B.

2
Revisão da Literatura

Este capítulo apresenta uma revisão sistemática da literatura, conduzida com o objetivo de identificar e analisar os conjuntos de dados (*datasets*) de CPET disponíveis, bem como algoritmos de diagnóstico que utilizam estes dados. A metodologia adotada segue as diretrizes PRISMA (Page et al., 2021), para garantir um processo transparente e rigoroso. A análise subsequente dos resultados foca-se em detalhar os trabalhos que forneceram os dados fundamentais para esta investigação, permitindo identificar as lacunas existentes na literatura que justificam a abordagem desta dissertação.

2.1 Metodologia da Revisão (PRISMA)

O fluxograma PRISMA 2020 foi utilizado como template para documentar o processo de seleção dos estudos e o respetivo diagrama foi obtido a partir do *website* oficial do PRISMA (Page et al., 2021; PRISMA Statement, 2024).[1]

Para a realização desta revisão sistemática da literatura, foi desenvolvida uma metodologia estruturada, focada em identificar trabalhos relevantes que disponibilizam *datasets* de CPET ou que aplicam algoritmos de inteligência artificial a estes dados. O processo foi realizado em colaboração com a colega investigadora Clara Neves, sob a orientação científica do Professor Rafael, que definiu os critérios metodológicos a serem seguidos, utilizando como base a estrutura de revisão proposta em Pinheiro et al. (2024).

2.1.1 Fontes de Pesquisa

A pesquisa foi realizada em quatro bases de dados eletrónicas de elevada relevância científica e académica: *Scopus*, *Web of Science*, *PubMed* e *PEDro*. A seleção destas fontes foi estratégica para abranger de forma complementar as áreas de engenharia e ciências da computação (*Scopus*, *Web of Science*) e, de forma importante, as áreas da saúde, biomédica e reabilitação (*PubMed*, *PEDro*), garantindo um vasto espectro de publicações

[1] PRISMA 2020 flow diagram template. Licensed under CC BY 4.0.

relevantes. Embora a base de dados *IEEE Xplore* seja uma fonte importante de artigos de engenharia, ela foi considerada indiretamente através da sua indexação na *Scopus*.

2.1.2 Estratégia de Busca

A estratégia de busca foi desenhada com base no modelo *Population, Intervention, Comparison, and Outcome* (PICO), que estrutura a questão de investigação em quatro pilares: População/Problema (P), Intervenção (I), Comparação (C) e *Outcome*/Resultados (O). No presente trabalho, estes componentes correspondem a: (P) pacientes com condições cardiovasculares e metabólicas; (I) o uso do CPET e de algoritmos de *machine learning*; (C) a presença de comparação entre técnicas (quando existente) ou a sua ausência; e (O) a disponibilidade de *datasets* e/ou a performance de modelos de diagnóstico.

Este modelo foi então **adaptado** aos objectivos desta revisão para efeitos de busca bibliográfica, mapeando-se os componentes em três *blocos de pesquisa*: População (P) → **DOENÇAS**, Intervenção (I) → **CPET**, e Outcome (O) → **COMPUTACIONAL** (com foco na identificação de *datasets* públicos e de algoritmos/classificadores e respectivas métricas de desempenho). O componente Comparação (C) **não foi incluído como bloco de busca separado**, uma vez que o foco principal é mapear dados e abordagens, e não realizar uma comparação directa entre intervenções no momento da busca.

Para cada componente, foi construída uma *search string* optimizada, combinando sinónimos e termos relacionados com o operador booleano "OR" para maximizar a sensibilidade dentro de cada bloco, e ligando os blocos entre si com "AND" para manter a especificidade da pergunta.

As *search strings* finais utilizadas foram as seguintes:

```
 % <-- Substitui o \texttt e resolve TUDO
String 1 (DOENÇAS): "heart failure" OR "cardiac insufficiency" OR
"cardiac failure" OR "cardiac arrhythmia" OR "cardiac rehabilitation" OR
"cardiovascular disease" OR "disease" OR "metabolic syndrome" OR
"cardiomyopathies" OR "pulmonary hypertension" OR "hypertension" OR
"cardiac assessment" OR "health"

String 2 (CPET): "CPET" OR "cardiopulmonary exercise test" OR
"cardiopulmonary exercise" OR "cardiopulmonary test" OR
"exercise tests" OR "treadmill exercise" OR "cardiorespiratory fitness"

String 3 (COMPUTACIONAL): "artificial intelligence" OR "machine
learning" OR "classification algorithm" OR "diagnostics system" OR
"clustering algorithm" OR "diagnosis system" OR "database" OR
"datasets" OR "deep learning" OR "predictive analytics"
```

A pesquisa principal combinou as três *strings* com o operador "AND" para identificar os artigos mais relevantes que cruzam as três áreas de interesse. Foi também

realizada uma busca mais ampla, combinando apenas as *strings* 1 e 2, para mapear a literatura geral sobre CPET em patologias cardíacas. As buscas foram limitadas aos campos de título, resumo e palavras-chave.

As buscas foram realizadas entre 10 e 15 de agosto de 2024, abrangendo as bases *Scopus*, *PubMed*, *Web of Science* e *PEDro*.

2.1.3 Critérios de Inclusão e Exclusão

Após a recolha dos artigos, foi realizado um processo de triagem em múltiplas fases, aplicando um conjunto de critérios de elegibilidade para refinar a seleção.

Os critérios de inclusão foram:

- Artigos que descrevem ou disponibilizam conjuntos de dados (*datasets*) originais de CPET.
- Artigos que desenvolvem ou aplicam algoritmos de diagnóstico, classificação ou prognóstico utilizando dados de CPET.
- Artigos de revisão sobre o desenvolvimento de algoritmos ou a criação de bases de dados de CPET.

Os critérios de exclusão foram:

- Artigos cujo texto completo não estava acessível (PDF indisponível).
- Artigos que, após análise do título e resumo, se revelaram fora do tema principal (ex.: estudos exclusivamente fisiológicos sem componente de dados, ou estudos de validação de equipamentos de medição).
- Artigos duplicados entre as diferentes bases de dados.

2.1.4 Processo de Seleção e Diagrama PRISMA

O processo de seleção seguiu as quatro fases do fluxograma PRISMA: identificação, triagem, elegibilidade e inclusão. A Figura 2.1 ilustra o fluxo detalhado desta revisão.

- Identificação: A busca inicial nas bases de dados, utilizando as combinações de *strings* definidas, resultou num total de 8.162 registos exportados para o gestor de referências *Mendeley*. Note-se que, embora a contagem inicial reportada pelas bases de dados pudesse ser ligeiramente superior (n=8.335), este número final (8.162) representa os registos únicos e processáveis que foram efetivamente importados para análise, após a resolução de quaisquer falhas de exportação ou incompatibilidades iniciais. Estes registos são provenientes das seguintes fontes:
  - Scopus: N=6.447
  - Web of Science: N=1.144
  - PubMed: N=249
  - PEDro: N=322

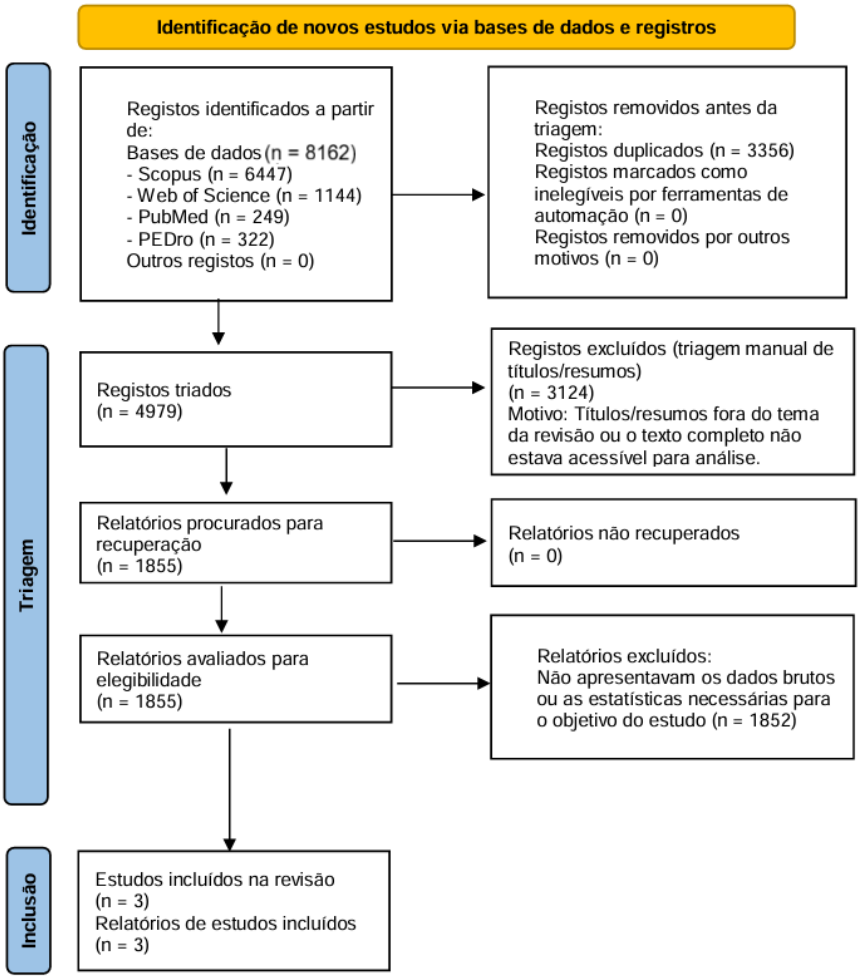

Figura 2.1: *Fluxograma PRISMA 2020 ilustrando o processo de identificação, triagem, elegibilidade e inclusão dos estudos. Modelo adaptado de Page et al. (2021). Template licenciado sob CC BY 4.0.*

- Triagem (*Screening*): Na primeira fase de triagem, de um total de 8.162 registos exportados, foram identificados e removidos 3.356 registos duplicados, inicialmente pela ferramenta automática do gestor de referências *Mendeley* e, posteriormente, por uma verificação manual focada em títulos e resumos. Esta ação resultou numa base de 4.979 registos únicos para a fase seguinte de análise. Na etapa de triagem manual de títulos e resumos, 3.124 registos foram excluídos. Esta exclusão ocorreu porque os artigos não se adequavam aos critérios de inclusão definidos, principalmente por dois motivos: relevância temática (estudos que se revelaram fora do foco da revisão, como pesquisas exclusivamente fisiológicas ou de validação de equipamentos) e inacessibilidade do texto completo (PDF indisponível para a equipa de revisão).
  Distribuição da triagem: Esta etapa foi realizada de forma colaborativa entre dois revisores independentes (o autor desta dissertação e a investigadora Clara Neves). Os registos foram divididos entre ambos, garantindo que cada subconjunto fosse avaliado em conformidade com os critérios definidos. Conflitos ou dúvidas foram discutidos em conjunto e resolvidos com base nas orientações do Professor Rafael.
- Elegibilidade: Os 4.979 registos foram submetidos a uma análise mais aprofundada do título e do resumo para avaliar a sua relevância para os objetivos da revisão. Artigos que não abordavam o uso de dados de CPET nem o desenvolvimento de algoritmos foram excluídos. Esta fase reduziu o número de registos elegíveis para 1.855.
  Nova divisão de tarefas: Também nesta etapa os registos foram divididos entre os dois revisores (Flávio e Clara), repetindo o processo de triagem independente, com discussão posterior em casos de discordância.
- Inclusão (texto integral): Após a leitura integral dos 1.855 registos, procedeu-se a uma análise de texto integral minuciosa e criteriosa, com o objetivo específico de identificar e selecionar *datasets* publicados. Para garantir robustez, esta triagem final foi dividida de forma equilibrada entre os dois revisores, assegurando uma cobertura total.
  Deste processo, foram selecionados 3 estudos para inclusão na síntese qualitativa. Note-se que o estudo-base de Pinheiro e Fonseca-Pinto, referido na Tabela 2.1, é contado como uma única entrada, apesar de detalhado em duas publicações (a original de 2023 e a sua expansão de 2025). Estes constituem a base fundamental deste trabalho, pois disponibilizam os *datasets* de dados reais e as estatísticas-chave necessárias para o trabalho experimental desenvolvido nos capítulos subsequentes desta dissertação. Os restantes 1.852 artigos foram excluídos por não cumprirem o critério essencial de disponibilização dos dados brutos necessários para os objetivos do estudo.

2.2 Análise Crítica dos Estudos Fundamentais

Uma verificação mais cuidadosa dos artigos selecionados revelou uma escassez de *datasets* de CPET publicamente disponíveis que contivessem as séries temporais completas (*breath-by-breath*). No entanto, o processo de triagem PRISMA foi fundamental para identificar três fontes de dados essenciais que, em conjunto, formaram a base para a construção do *dataset* híbrido utilizado neste trabalho, conforme sintetizado na Tabela 2.1.

Tabela 2.1: *Visão geral dos estudos utilizados como fonte de dados e metodologia nesta dissertação.*

Autor/Estudo	População/Ano	Descrição e Uso na Dissertação
Pinheiro et al., 2023; Pinheiro et al., 2025	N=45 (2023, 2025)	Base metodológica (SVM + DWT) e fonte dos dados de controlo (S), apresentada originalmente em 2023 e expandida em 2025. As séries temporais de CPET eram parciais.
Brown et al., 2022	N=30 (2022)	Principal fonte de dados para pacientes (IC/SM). A metodologia original do estudo utilizava modelos de CNN e *autoencoders* com estas séries temporais completas (*breath-by-breath*).
Portella et al., 2022	N=219 (2022)	Fonte de parâmetros estatísticos (média, DP, *slopes*) usados para gerar os dados semi-sintéticos (LP e LM). O estudo original aplicou um modelo Random Forest para a sua própria tarefa de classificação.

As secções seguintes detalham cada uma destas fontes, analisando as suas contribuições e limitações no contexto desta investigação.

1. Os trabalhos de Pinheiro e Fonseca-Pinto (Pinheiro et al., 2023; Pinheiro et al., 2025) estabelecem a referência fundamental que é expandida nesta investigação. O *framework* de classificação original, que utilizou a DWT e o SVM para três classes, serve como ponto de partida. A metodologia é então aplicada a um problema mais complexo de cinco classes, e a sua eficácia é validada através da comparação com outras técnicas, fortalecendo as conclusões para a comunidade científica.

2. O trabalho de Brown et al. (Brown et al., 2022) é fundamental para esta investigação, pois constituiu a fonte principal dos dados reais para as classes IC e SM, disponibilizados num repositório público no *GitHub*[2]. A sua abordagem, com *autoencoders* e *Convolutional Neural Networks* (CNN), é distinta da metodologia baseada em SVM e DWT aqui explorada. Neste estudo, os seus dados são utilizados para expandir e validar uma abordagem diferente, em vez de replicar os resultados originais.

3. O trabalho de Portella et al. (Portella et al., 2022) evidencia a escassez de dados brutos na área. Embora o trabalho original forneça um *dataset* de parâmetros estatísticos através do *IEEE DataPort*, ele não disponibiliza as séries temporais completas. Em

[2] O repositório está disponível em: `https://github.com/suchethassharma/CPET`

função disso, essas estatísticas foram utilizadas para gerar dados semi-sintéticos para as classes de Limitação Pulmonar e Limitação Musculoesquelética, demonstrando a viabilidade de uma nova abordagem para superar uma limitação fundamental da área.

> ***Nota***
>
> A escassez de séries *breath-by-breath* públicas é uma limitação recorrente em CPET; muitos estudos disponibilizam apenas dados estatísticos agregados/medidas derivadas, o que dificulta métodos que exploram a dinâmica temporal completa (Franssen et al., 2023).

2.3 Fundamentação Teórica

Nesta secção, são apresentados os conceitos teóricos fundamentais que servem de alicerce para a metodologia desenvolvida nesta dissertação. O objetivo é detalhar os princípios do CPET, as suas variáveis e limiares, e introduzir as bases do processamento de sinal e da aprendizagem automática aplicadas a este contexto.

2.3.1 O Teste Cardiopulmonar de Esforço (CPET) em Detalhe

O Teste Cardiopulmonar de Esforço (CPET) é um procedimento de diagnóstico não invasivo que avalia de forma integrada as respostas dos sistemas cardiovascular, respiratório, metabólico e musculoesquelético a um stresse fisiológico induzido pelo exercício (Wasserman et al., 2011). Ao contrário de um teste de esforço convencional, que se foca primariamente na resposta do Eletrocardiograma (ECG), o CPET analisa a troca gasosa, respiração a respiração (*breath-by-breath*), permitindo uma visão holística e quantitativa da capacidade funcional do indivíduo.

Variáveis Fundamentais do CPET

A riqueza do CPET reside na medição simultânea de múltiplas variáveis, cujas dinâmicas temporais contêm informação diagnóstica importante. As principais variáveis incluem:

- Consumo de Oxigénio (Consumo de Oxigénio (VO_2)): Representa a quantidade de oxigénio que o corpo capta e utiliza para o metabolismo energético. O seu valor de pico (VO_2 pico) é considerado o "padrão-ouro" para medir a capacidade cardiorrespiratória e é um dos mais importantes preditores de prognóstico em pacientes com IC (Wasserman et al., 2011).
- Produção de Dióxido de Carbono (VCO_2): Mede a quantidade de dióxido de carbono que o corpo elimina. A sua relação com o VO_2 reflete a intensidade do metabolismo e o tipo de substrato energético utilizado (gorduras vs. hidratos de carbono).

- Rácio de Trocas Gasosas (*Respiratory Exchange Ratio*, RER): É a razão entre a produção de dióxido de carbono e o consumo de oxigénio (RER = VCO_2 / VO_2). Em repouso e exercício de baixa intensidade, assume valores inferiores a 1.0. À medida que a intensidade aumenta, o Rácio de Trocas Gasosas (*Respiratory Exchange Ratio*, RER) (RER) sobe e pode ultrapassar 1.0, indicando a transição para o metabolismo anaeróbio e a tamponização do lactato. Valores de pico superiores a 1.10 são frequentemente usados como critério de esforço máximo do paciente (Wasserman et al., 2011).
- Ventilação Minuto (VE): Corresponde ao volume total de ar expirado por minuto. Reflete a resposta do sistema respiratório para satisfazer a crescente demanda metabólica durante o exercício.

Limiares Fisiológicos e a sua Importância Clínica

A análise da relação entre as variáveis do CPET ao longo do tempo permite identificar pontos de inflexão, conhecidos como limiares fisiológicos, que são de extrema importância para o diagnóstico:

- Limiar Anaeróbio (LA): É o ponto em que o metabolismo anaeróbio começa a contribuir de forma significativa para a produção de energia, resultando num aumento sustentado do lactato no sangue. É identificado de forma não invasiva (através do método *V-slope*) como o ponto em que o VCO_2 começa a aumentar de forma desproporcional em relação ao VO_2. Um LA que ocorre numa percentagem baixa do VO_2 pico previsto é um indicador de disfunção cardiovascular ou de descondicionamento físico (Wasserman et al., 2011).
- Ponto de Compensação Respiratória Ponto de Compensação Respiratória (PCR): Ocorre numa intensidade de exercício superior ao LA. Neste ponto, a acidose metabólica é tão acentuada que o sistema respiratório responde com uma hiperventilação para compensar (tamponizar) o excesso de H^+ no sangue. É identificado pelo aumento acentuado da VE em relação ao VCO_2.
- Eficiência Ventilatória (declive VE/VCO_2): Representa a relação entre a ventilação e a eliminação de dióxido de carbono. Um valor elevado indica que o paciente necessita de ventilar um volume de ar excessivo para eliminar o VCO_2, o que sugere uma troca gasosa ineficiente. Este é um poderoso marcador de prognóstico, especialmente em pacientes com IC e hipertensão pulmonar (Wasserman et al., 2011).

A interpretação tradicional do CPET foca-se na identificação destes limiares e na análise de valores de pico. Contudo, a análise da dinâmica temporal completa de todas as variáveis — frequentemente negligenciada pela análise convencional — apresenta-se como uma oportunidade para a descoberta de padrões diagnósticos adicionais, justificando a aplicação de técnicas de processamento de sinal e aprendizagem automática.

2.3.2 Técnicas de Processamento de Sinal para Análise de Dados Fisiológicos

A análise de séries temporais fisiológicas, como as obtidas no CPET, apresenta desafios únicos que exigem ferramentas de processamento de sinal sofisticadas. As propriedades estatísticas destes sinais mudam constantemente ao longo do tempo, em resposta à carga de exercício crescente e às respostas fisiológicas do indivíduo.

Sinais Estacionários vs. Não-Estacionários

Um conceito central no processamento de sinal é a distinção entre estacionariedade e não estacionariedade. Um sinal é considerado estacionário se as suas propriedades estatísticas (como a média e a variância) permanecem constantes ao longo do tempo. Sinais como um ruído branco ou uma onda sinusoidal pura são exemplos de sinais estacionários. Para estes, ferramentas clássicas de análise no domínio da frequência são altamente eficazes.

No entanto, os sinais biomédicos são quase invariavelmente não estacionários. Durante um CPET, a frequência cardíaca aumenta, a frequência respiratória acelera e a amplitude da ventilação expande-se. Isto significa que as frequências e amplitudes que compõem estes sinais estão em constante mudança. Aplicar métodos tradicionais a estes sinais pode levar a interpretações erróneas, pois a informação temporal — *quando* uma determinada mudança ocorre — é perdida.

> ***Nota***
>
> A aplicação da Transformada de Fourier de Tempo Curto (STFT) apresenta limitações significativas na análise de sinais não estacionários, como os biossinais obtidos em CPET. Em contraste, a Transformada *Wavelet* Discreta (DWT) permite uma análise multirresolução, oferecendo maior capacidade de capturar variações temporais e frequenciais desses sinais (Addison, 2005).

Análise no Domínio da Frequência: A Transformada de Fourier e as suas Limitações

A ferramenta mais clássica para a análise de sinais no domínio da frequência é a Transformada de Fourier. Decompõe um sinal numa soma de ondas sinusoidais de diferentes frequências, revelando eficazmente "quais" frequências estão presentes no sinal como um todo. No entanto, o seu principal inconveniente para a análise de sinais não estacionários é que perde toda a informação sobre a localização temporal. A Transformada de Fourier pode indicar que uma frequência de alta intensidade ocorreu, mas não consegue especificar se foi no início, no meio ou no fim do teste de esforço. Para uma análise clínica, onde a cronologia dos eventos fisiológicos é fundamental, esta limitação torna-a inadequada.

Esta limitação fundamental da Transformada de Fourier motivou o desenvolvimento de técnicas de análise tempo-frequência, que procuram representar o sinal em

ambos os domínios simultaneamente. As quatro transformadas utilizadas nesta dissertação — STFT, DWT, WPT e EMD — inserem-se nesta categoria de ferramentas avançadas, cada uma com uma abordagem distinta para resolver o desafio da análise de sinais não estacionários.

Análise Tempo-Frequência: A Transformada de Fourier de Curto Tempo (STFT)

A Transformada de Fourier de Curto Tempo (STFT) foi uma das primeiras e mais intuitivas soluções para contornar a limitação da Transformada de Fourier na análise de sinais não estacionários (Oppenheim et al., 1999). A sua abordagem baseia-se no princípio de "dividir para conquistar": em vez de analisar o sinal inteiro de uma só vez, a STFT divide o sinal em pequenos segmentos temporais, ou "janelas", que são suficientemente curtos para que se possa assumir que o sinal é localmente estacionário dentro de cada um deles.

O processo funciona da seguinte forma:

1. Janelamento: Uma função de janela (como a janela de Hamming ou Hanning) de um determinado comprimento é selecionada. Esta janela é deslizada ao longo do sinal.
2. Cálculo da FFT: Para cada posição da janela, o segmento do sinal dentro dela é isolado e uma Transformada Rápida de Fourier (*Fast Fourier Transform*) (FFT) é calculada. Isto gera o espectro de frequências para esse pequeno instante de tempo.
3. Construção do Espectrograma: O processo é repetido, deslizando a janela ao longo de todo o sinal (geralmente com alguma sobreposição entre janelas adjacentes). Os espectros de frequência resultantes de cada janela são então combinados para formar um espetrograma, que é uma representação tridimensional que mostra a energia do sinal em função do tempo (eixo x) e da frequência (eixo y).

A principal vantagem da STFT é que preserva a informação temporal e de frequência simultaneamente. No entanto, a sua maior limitação reside na resolução fixa da sua janela de análise. Esta característica impõe um compromisso fundamental, muitas vezes referido como o Princípio da Incerteza de Heisenberg-Gabor:

- Uma janela larga no tempo captura um segmento longo do sinal, o que permite uma excelente resolução de frequência (consegue-se distinguir frequências muito próximas), mas oferece uma péssima resolução temporal (não se sabe exatamente *quando* um evento ocorreu dentro daquela janela larga).
- Uma janela estreita no tempo oferece uma excelente resolução temporal (localiza eventos com precisão), mas resulta numa péssima resolução de frequência (as componentes de frequência ficam "esborratadas"e difíceis de distinguir).

A escolha do tamanho da janela é, portanto, crítica e fixa para toda a análise. Esta rigidez é uma desvantagem significativa para sinais como os do CPET, que contêm si-

multaneamente eventos rápidos (detalhes de alta frequência) e tendências lentas (componentes de baixa frequência). É precisamente esta limitação que a Transformada Wavelet, com a sua capacidade de análise multirresolução, procura resolver.

Análise Multirresolução: A Família das Transformadas Wavelet (DWT e WPT)

Para superar a limitação da resolução fixa da STFT, foi desenvolvida a Análise *Wavelet*. A sua principal vantagem, e o conceito que a torna tão poderosa para a análise de sinais não estacionários, é a Análise Multirresolução (Polikar, 1996). Em vez de usar uma janela de análise de tamanho fixo, a Transformada Wavelet utiliza uma onda protótipo de curta duração, denominada wavelet-mãe, que é flexível em escala.

O processo de análise envolve duas operações fundamentais sobre a wavelet-mãe:

- Translação (ou Deslocamento): A wavelet-mãe é deslizada ao longo do eixo do tempo do sinal. Esta operação permite localizar *quando* um determinado evento de frequência ocorre, preservando a informação temporal.
- Escala (ou Dilatação/Compressão): A wavelet-mãe é esticada ou comprimida. Esta operação de escala está inversamente relacionada com a frequência. Uma wavelet "comprimida"(escala pequena) é curta e de alta frequência, enquanto uma wavelet "esticada"(escala grande) é longa e de baixa frequência.

A combinação destas duas operações resulta na capacidade de análise multirresolução. Ao analisar o sinal, a transformada utiliza versões comprimidas (janelas curtas) da wavelet-mãe para detetar eventos de alta frequência, obtendo uma excelente resolução temporal. Simultaneamente, utiliza versões esticadas (janelas longas) para analisar as componentes de baixa frequência do sinal, obtendo uma excelente resolução de frequência. Desta forma, a Transformada Wavelet adapta-se dinamicamente à natureza do sinal, ao contrário da STFT e do seu compromisso fixo.

Nesta dissertação, foram exploradas duas variantes desta família de transformadas:

Transformada Wavelet Discreta (DWT) A DWT é uma implementação computacionalmente eficiente da transformada que decompõe o sinal em níveis discretos. Em cada nível, a DWT divide o sinal em duas componentes através de filtros digitais: os coeficientes de aproximação (cA), que representam a componente de baixa frequência (a "tendência"do sinal), e os coeficientes de detalhe (cD), que representam a componente de alta frequência (as "variações rápidas"). O processo é então repetido de forma recursiva apenas sobre os cA, aprofundando a análise das baixas frequências a cada nível subsequente (Polikar, 1996).

Transformada Wavelet por Pacotes (WPT) A WPT é uma generalização e extensão da DWT. A principal diferença é que, enquanto a DWT decompõe apenas os coeficientes de aproximação em cada nível, a WPT decompõe *ambos*, os cA e os cD. Isto resulta numa árvore de decomposição muito mais rica e completa, permitindo uma análise

mais fina e detalhada de todas as bandas de frequência do sinal, e não apenas das baixas frequências.

Abordagens Adaptativas: A Decomposição por Modos Empíricos (EMD)

A Decomposição por Modos Empíricos (EMD) representa uma abordagem fundamentalmente diferente das transformadas de Fourier e Wavelet. Enquanto a STFT e a DWT são métodos *a priori*, que decompõem um sinal com base em funções pré-definidas (sinusoides ou uma wavelet-mãe), a EMD é uma técnica *a posteriori*, ou seja, totalmente adaptativa e guiada pelos próprios dados (Schmitt et al., 2014).

O objetivo da EMD é decompor qualquer sinal complexo e não estacionário num conjunto de componentes mais simples, chamadas Função de Modo Intrínsecos (IMFs). Cada IMF representa um modo oscilatório simples e monocomponente presente no sinal. Para que uma função seja considerada uma IMF, deve satisfazer duas condições:

1. O número de extremos (máximos e mínimos) e o número de cruzamentos com o zero devem ser iguais ou diferir no máximo em um.
2. Em qualquer ponto do sinal, a média dos seus invólucros superior (definido pelos máximos locais) e inferior (definido pelos mínimos locais) deve ser zero.

A extração das IMFs é realizada através de um processo iterativo chamado *sifting* (peneiramento). De forma conceptual, o algoritmo funciona da seguinte maneira:

- Identifica todos os máximos e mínimos locais do sinal.
- Gera um invólucro superior e um inferior, interpolando entre os pontos extremos.
- Calcula a média local destes dois invólucros.
- Subtrai esta média do sinal original. O resultado é uma "proto-IMF".
- O processo é repetido sobre a proto-IMF até que as duas condições de uma IMF sejam satisfeitas.

Uma vez que a primeira IMF (que representa a componente de frequência mais alta do sinal) é extraída, ela é subtraída do sinal original, resultando num resíduo. O processo de *sifting* é então aplicado a este resíduo para encontrar a segunda IMF, e assim sucessivamente, até que o resíduo final seja uma função monotónica, a partir da qual não se podem extrair mais IMFs. A sua natureza adaptativa torna-a uma ferramenta poderosa para analisar sinais não lineares, constituindo uma alternativa interessante às abordagens baseadas em transformadas.

2.3.3 Aprendizagem Automática para Classificação Supervisionada

O terceiro pilar tecnológico desta dissertação é a aplicação de técnicas de Aprendizagem Automática (*Machine Learning*) para realizar o diagnóstico multiclasse. Especificamente, o problema enquadra-se na categoria de aprendizagem supervisionada. Neste paradigma, o objetivo é aprender uma função que mapeia um vetor de entradas

(as características extraídas dos sinais de CPET) a um vetor de saídas (os rótulos de diagnóstico), a partir de um conjunto de dados de treino onde ambos são conhecidos.

A tarefa específica é a de classificação, onde a saída desejada é uma categoria discreta de um conjunto finito de possibilidades — no caso deste trabalho, uma das cinco condições fisiológicas. O algoritmo é "treinado"com os dados de múltiplos pacientes e aprende a reconhecer os padrões que distinguem cada classe. Uma vez treinado, o modelo deve ser capaz de generalizar o seu conhecimento para classificar corretamente novos pacientes, nunca antes vistos.

Máquinas de Vetores de Suporte (SVM)

Dentro do vasto leque de algoritmos de classificação, as Máquinas de Vetores de Suporte (SVM), introduzidas por Cortes e Vapnik, são um dos métodos mais poderosos e robustos, especialmente eficazes em espaços de características de alta dimensão e com amostras limitadas (Cortes et al., 1995).

A intuição geométrica por trás do SVM é encontrar um hiperplano de separação ótimo que divida os pontos de dados de diferentes classes no espaço de características. Para um problema com duas classes linearmente separáveis, existiriam infinitos hiperplanos (linhas, em 2D) que poderiam separar os dados. O SVM procura o "melhor" hiperplano, que é aquele que maximiza a margem — a distância entre o hiperplano e os pontos de dados mais próximos de cada classe. A lógica é que um separador com uma margem maior terá uma menor probabilidade de classificar incorretamente novos dados, conferindo ao modelo uma melhor capacidade de generalização.

Os pontos de dados que se encontram exatamente na margem são chamados de vetores de suporte, e são eles que definem a posição e orientação do hiperplano. Uma propriedade notável do SVM é que apenas estes vetores de suporte são relevantes para a definição do modelo; os restantes pontos de dados poderiam ser removidos sem alterar a fronteira de decisão.

Para lidar com problemas em que os dados não são linearmente separáveis no seu espaço original, o SVM emprega a célebre "manobra do kernel"(*kernel trick*). Em vez de tentar separar os dados com uma linha ou plano, uma função de *kernel* mapeia os dados, de forma implícita e computacionalmente eficiente, para um espaço de características de dimensão muito superior, onde se torna possível encontrar um hiperplano que os separe linearmente. A escolha da função de *kernel* (Linear, Polinomial, Função de Base Radial (*Radial Basis Function*) (RBF)) define a complexidade e a forma da fronteira de decisão no espaço original.

2.4 Síntese e Análise Crítica

Esta revisão sistemática da literatura permitiu mapear o panorama atual da utilização de dados de CPET, com um foco especial na disponibilidade de *datasets* para o desenvolvimento de modelos computacionais.

2.4.1 Panorama da Literatura

A literatura oferece uma quantidade considerável de estudos que demonstram a utilidade clínica e prognóstica dos parâmetros derivados do CPET. Grandes registos como o *FIT-CPX* e o *MECKI score* solidificaram o valor de variáveis como o consumo de oxigénio de pico (VO_2) e o declive Ventilação (VE)/VCO_2. Além disso, existe um interesse crescente na aplicação de técnicas de *machine learning* para extrair informação mais detalhada e preditiva dos dados do CPET. Exemplos incluem Brown et al., 2022, que aplicaram CNN e *autoencoders* para distinguir pacientes com IC e SM, e Pinheiro et al. (2025), que utilizaram SVM combinado com DWT para classificação multiclasse entre IC, SM e S. Estes exemplos reforçam que o CPET é não apenas uma ferramenta diagnóstica tradicional, mas também um recurso valioso para aplicações avançadas de IA em saúde. Outros estudos também exploraram o SVM em conjunto com outras técnicas, como a Análise de Componentes Principais (*Principal Component Analysis*) (PCA) (Subasi et al., 2010; Choudhary et al., 2017; Azizi, 2024), para tarefas de classificação de sinais biomédicos.

Outras Abordagens de Classificação na Literatura de CPET

Para além dos algoritmos de SVM e redes neuronais explorados nos estudos fundamentais, a literatura de aplicação de *machine learning* ao CPET abrange uma variedade de outras técnicas. Algoritmos de *ensemble*, como *Random Forests* e *Gradient Boosting* (ex: XGBoost), são frequentemente utilizados devido à sua robustez e elevado desempenho na classificação baseada em parâmetros clínicos extraídos dos sinais (Deng et al., 2022). Métodos baseados em árvores de decisão também são frequentes para a criação de *scores* de risco e auxílio à decisão clínica, demonstrando uma acurácia comparável à de especialistas na identificação de limitações de exercício (Schwendinger et al., 2024). A regressão logística, embora um modelo linear, permanece como um *baseline* importante e interpretável para tarefas de prognóstico e estratificação de risco em cirurgia oncológica com base em dados de CPET (West et al., 2024). A Tabela 2.2 resume as características, pontos fortes e limitações das principais famílias de algoritmos neste contexto.

Neste trabalho privilegia-se o SVM com DWT pela boa relação amostra-dimensão encontrada no problema e pela interpretabilidade tempo-escala com dados escassos. As abordagens de *Deep Learning* e *ensembles* são discutidas como alternativas potentes quando existem grandes *datasets* brutos e a necessidade de aprender representações diretamente das séries.

2.4.2 Lacunas Identificadas e Oportunidades de Investigação

Apesar da riqueza da literatura, a principal lacuna identificada é a escassez de *datasets* publicamente disponíveis que contenham os dados brutos, sinal a sinal (*breath-by-breath*), do CPET. A maioria dos estudos publica apenas os parâmetros derivados, o que impede o desenvolvimento e a validação de novos algoritmos. Adicionalmente,

observou-se que muitos *datasets* não são públicos e de pequena dimensão. Existe uma concentração de dados em patologias específicas (como IC), e a falta de padronização nos protocolos de aquisição dificulta a combinação de dados.

Tabela 2.2: *Resumo comparativo de algoritmos aplicáveis a dados de CPET.*

Algoritmo	**Entrada típica**	**Pontos fortes em CPET**	**Limitações**
Regressão Logística (	Picos, limiares, slopes, índices tabulares	Interpretável e calibrável; bom baseline	Fronteira linear; perde padrões complexos
kNN / DTW	Séries VO_2, VCO_2, VE, RR alinhadas	Sensível à forma temporal; pouco *tuning*	Escalabilidade; sensível a ruído e escala
Random Forest / XGBoost	*Features* tabulares derivadas	Não linear; lida com interações e *missing* moderado	Hiperparâmetros; explicabilidade global
SVM (lin/RBF)	Coeficientes DWT/WPT ou *features*	Bom com poucos sujeitos e alta dimensão	Escolha de kernel e escalonamento
HMM	Séries segmentadas por fases (repouso–rampa–recuperação)	Modela estados latentes (LA/PCR)	Requer segmentação consistente e suposições Markov
1D-CNN / LSTM	Séries *breath-by-breath* multivariadas	Capta padrões temporais complexos	Demanda dados/-regularização; custo computacional

2.4.3 Justificação da Abordagem da Dissertação

As lacunas identificadas fundamentam diretamente a necessidade da presente dissertação. A dificuldade em obter acesso a *datasets* de CPET com dados brutos é uma barreira significativa. Neste contexto, a geração de dados semi-sintéticos surge como uma solução inovadora e pragmática para contornar esta limitação, permitindo desenvolver e testar novos algoritmos num ambiente controlado e explorar cenários com dados escassos.

A abordagem aqui proposta distingue-se ao realizar uma análise comparativa sistemática entre a DWT e outras transformadas de sinal (STFT, WPT, EMD) sob um protocolo consistente. Esta análise é fundamental para validar se a metodologia base, comprovadamente eficaz para 3 classes, mantém a sua robustez e elevado desempenho num cenário mais desafiador de 5 classes, algo que não havia sido explorado na literatura.

2.5 Síntese e Transição para os Métodos

Esta revisão permitiu compreender a relevância clínica do CPET e o seu potencial para aplicações computacionais. A carência significativa de *datasets* públicos contendo séries temporais completas reforça a importância de estratégias inovadoras como a geração de dados semi-sintéticos. Adicionalmente, a necessidade de validar as metodologias existentes em cenários mais complexos justifica a análise comparativa realizada neste trabalho. Diante disto, o capítulo seguinte detalha a metodologia completa, incluindo a forma como os dados foram gerados e os modelos foram avaliados.

3
Materiais e Métodos

Este capítulo descreve em detalhe a metodologia empregue no presente trabalho, desde a caracterização do conjunto de dados até aos processos de extração de características, classificação e validação. A metodologia aqui apresentada constitui uma expansão do trabalho aceite para publicação e apresentado no Apêndice B (Santos et al., 2025), aplicando o *framework* original a um problema mais complexo de cinco classes. Adicionalmente, detalha-se a implementação de três métodos alternativos de extração de características – baseados na STFT, na WPT e na EMD – para realizar uma análise comparativa abrangente e determinar a abordagem mais eficaz para a classificação de sinais de CPET. A Figura 3.1 ilustra o fluxo completo desta metodologia experimental.

3.1 Descrição do *Dataset*

O conjunto de dados utilizado nesta investigação é composto por séries temporais de parâmetros fisiológicos, recolhidos durante a realização de CPET em cinco grupos de diagnóstico distintos: IC, SM, S, LP e LM.

As fontes dos dados são heterogéneas. Os dados para os grupos IC e SM, contendo as séries temporais completas para variáveis como a frequência cardíaca (HR), consumo de oxigénio (VO_2), produção de dióxido de carbono (VCO_2), ventilação (VE), frequência respiratória (RR) e o RER, foram obtidos a partir de um repositório público no *GitHub* associado ao trabalho de Brown et al. (2022). Os dados para o grupo de controlo (S) provêm de um outro conjunto de dados, conforme descrito em Pinheiro et al. (2025). Neste último, os valores de RER não estavam explicitamente disponíveis, tendo sido calculados manualmente através da seguinte equação, uma prática fisiologicamente válida e aceite na literatura (Wasserman et al., 2011):

$$\text{RER} = \frac{\text{VCO}_2}{\text{VO}_2}. \tag{3.1}$$

Os dados para os grupos LP e LM foram gerados de forma semi-sintética, utilizando estatísticas fisiológicas extraídas de pacientes reais, disponíveis no repositório *IEEE*

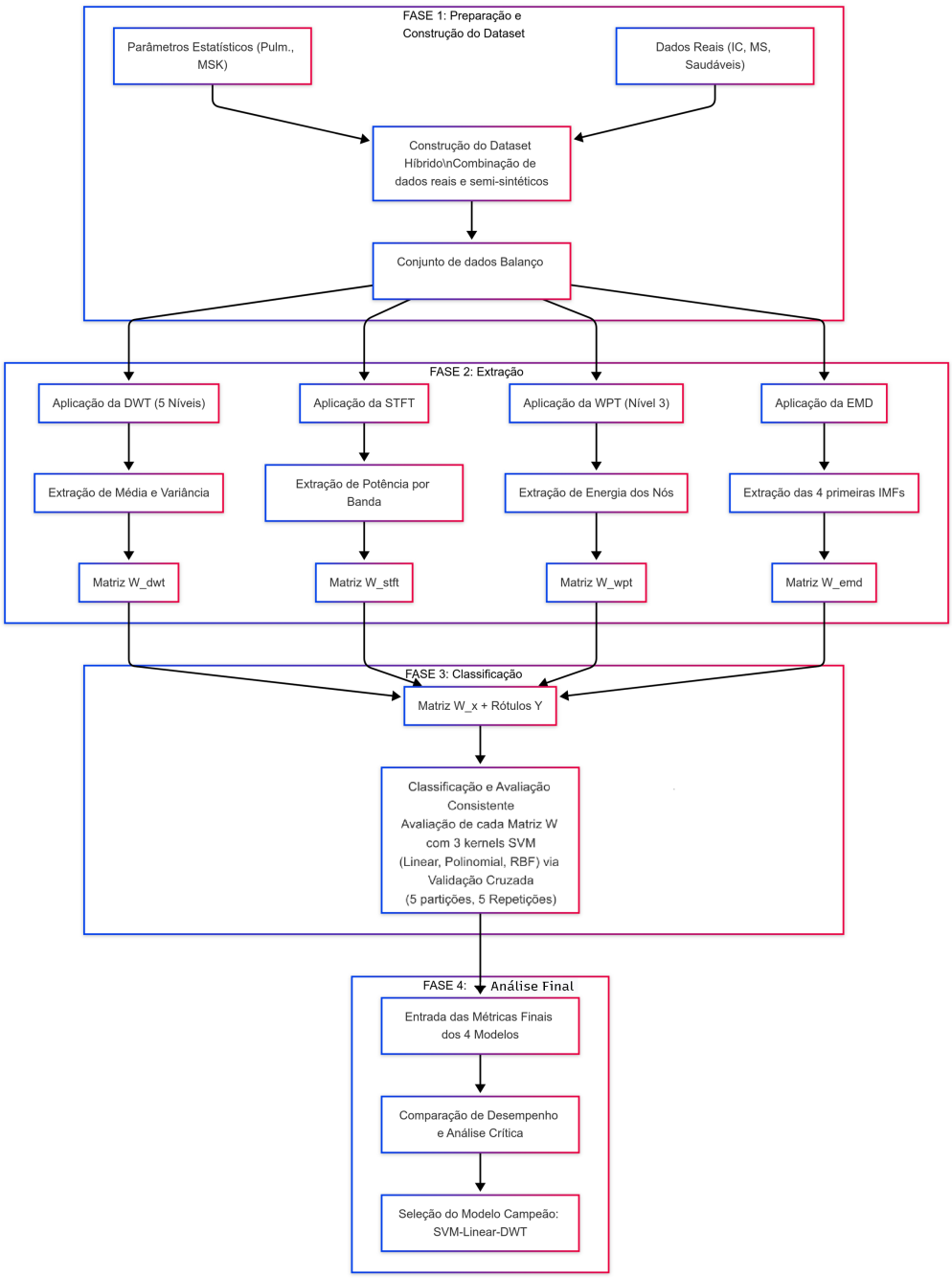

Figura 3.1: *Fluxograma da metodologia experimental. O processo inicia-se com a construção do dataset híbrido, seguido por uma análise comparativa de quatro métodos de extração de características. Cada método é avaliado sob um protocolo consistente para garantir uma comparação justa, culminando na identificação do modelo com melhor desempenho.*

DataPort (Portella et al., 2022).

O *dataset* final foi balanceado, sendo composto por um total de 75 pacientes, distribuídos uniformemente com 15 sujeitos por cada uma das cinco classes para minimizar o viés de treino. A escolha deste número de amostras foi uma decisão metodológica deliberada para garantir uma comparação direta e justa com os estudos de base (Pinheiro et al., 2023; Pinheiro et al., 2025; Brown et al., 2022), que adotaram precisamente o mesmo tamanho de amostra.

3.2 Geração de Dados Semi-Sintéticos

Devido à ausência de séries temporais completas para os grupos de LP e LM, foi implementado um processo de geração de dados "semi-sintéticos". Esta designação é utilizada porque os sinais, apesar de artificiais, são gerados com base em parâmetros estatísticos extraídos de dados de pacientes reais (média, mínimo, máximo, desvio padrão e inclinações), uma abordagem inspirada em modelos de reconstrução de séries temporais fisiológicas (Pezoulas et al., 2024; McSharry et al., 2003).

Nos casos em que o desvio padrão ($\hat{\sigma}$) não estava explicitamente disponível, este foi aproximado pela *range rule of thumb*, uma regra empírica clássica para dados aproximadamente normais (Ramirez et al., 2012), definida por:

$$\hat{\sigma} \approx \frac{\text{Amplitude}}{4}. \tag{3.2}$$

onde a amplitude corresponde à diferença entre o valor máximo e o mínimo do conjunto de observações.

> ***Nota***
>
> A utilização do divisor 4 constitui uma simplificação heurística. Embora em distribuições normais ideais o fator de escala ótimo dependa do tamanho da amostra (n), tendendo a ser ligeiramente superior para n maiores, optou-se por manter esta regra fixa como critério de padronização, assumindo-se esta aproximação como uma limitação metodológica do processo de geração.

O processo de criação de cada sinal sintético seguiu um protocolo rigoroso para garantir o realismo fisiológico:

- Nos grupos semi-sintéticos (LP e LM) adotou-se um comprimento fixo de 200 pontos para padronizar a decomposição, nos grupos reais, utilizou-se o comprimento efetivamente observado em cada registo.
- Foram aplicadas as inclinações (*slopes*) a cada segmento para refletir as transições fisiológicas do exercício.

- Foram adicionadas oscilações controladas e ruído gaussiano para simular a variabilidade biológica natural.
- Os valores foram recortados (*clipped*) para se manterem dentro de limites fisiológicos de mínimo e máximo.

> ***Nota***
>
> A designação "semi-sintético" é utilizada para enfatizar que, apesar de gerados artificialmente, os sinais são guiados por parâmetros estatísticos de pacientes reais. Esta abordagem garante que as séries temporais mantêm tendências e variabilidade fisiologicamente plausíveis, sendo mais realistas do que dados puramente aleatórios (McSharry et al., 2003).

3.3 Extração de Características

3.3.1 Pré-processamento e Controlo de Fuga de Informação

Os sinais brutos de CPET (HR, VO_2, VCO_2, RR, VE e RER) foram usados sem normalização prévia, filtragem ou suavização antes da extração de características. Para cada sujeito e variável, utilizou-se o comprimento real do registo (sem truncamento ou *padding*). As características foram extraídas diretamente dos sinais através das quatro transformadas (DWT, STFT, WPT e EMD). A única etapa de padronização foi a normalização z-score, aplicada apenas às *features* já extraídas. Esta etapa foi realizada internamente pelo classificador SVM, através do parâmetro `Standardize=true`, e ajustada dentro de cada *fold* de treino durante a validação cruzada (5-*fold*), assegurando assim a ausência de *data leakage*.

> ***Nota***
>
> A fuga de informação (*data leakage*) é um erro metodológico crítico em *machine learning*. Ocorre quando dados do conjunto de teste (ou informação externa ao treino) influenciam qualquer etapa de modelação — por exemplo, estimar parâmetros de pré-processamento (normalização, seleção de *features*, PCA, imputação) com todo o *dataset* ou fora do ciclo de validação. Tal prática tende a inflacionar as estimativas de desempenho e compromete a validade externa. Para o evitar, os parâmetros de pré-processamento devem ser aprendidos exclusivamente no conjunto de treino em cada *fold* e só depois aplicados ao teste (Sasse et al., 2023).

3.3.2 Justificação da Escolha dos Métodos de Extração

O ponto de partida do presente trabalho foi a DWT, uma técnica já validada para a análise de sinais biomédicos, conforme o artigo de base (Pinheiro et al., 2025). No

entanto, a literatura oferece diversas alternativas para a extração de características de sinais não estacionários. Para garantir que a abordagem com a DWT é, de facto, a mais adequada para a tarefa de classificação de sinais de CPET, foram comparadas três outras metodologias: a STFT (Oppenheim et al., 1999), a WPT (Coifman et al., 1992) e a EMD (Schmitt et al., 2014). Esta comparação sistemática e empírica foi realizada para validar a escolha da transformada de processamento de sinal num cenário de classificação mais complexo.

3.3.3 Transformada Wavelet Discreta (DWT)

Fundamentação Teórica

A DWT é uma técnica de processamento de sinal que decompõe um sinal nos domínios do tempo e da frequência (Polikar, 1996). A sua principal vantagem reside na sua capacidade de Análise Multirresolução. Sinais fisiológicos como os do CPET são inerentemente não estacionários. A DWT utiliza uma *wavelet-mãe* que é escalada e transladada, permitindo analisar o sinal com diferentes resoluções. A transformada é definida pela equação geral da Transformada Wavelet Contínua:

$$T(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt \tag{3.3}$$

onde $x(t)$ é o sinal, $\psi(t)$ é a *wavelet-mãe*, a é o parâmetro de escala, b é o parâmetro de translação, e * denota o conjugado complexo. A decomposição iterativa da DWT gera dois tipos de componentes:

- Coeficientes de Aproximação cA: representam o conteúdo de baixa frequência do sinal.
- Coeficientes de Detalhe cD: representam o conteúdo de alta frequência do sinal.

Aplicação e Configuração

Para a extração de características com a DWT, todas as séries temporais foram processadas utilizando a *wavelet-mãe* Daubechies 2 (`db2`). Foram consideradas duas configurações de decomposição:

- Wavelet Multirresolução – Nível 3 (MW3): decomposição em 3 níveis;
- Wavelet Multirresolução – Nível 5 (MW5): decomposição em 5 níveis.

Relativamente ao trabalho de referência, que utilizava 5 sinais fisiológicos, acrescentou-se o RER como sexto sinal. As características foram obtidas a partir da média e da variância dos coeficientes de detalhe de cada nível e do coeficiente de aproximação final.

Assim, na configuração MW3 extraíram-se 48 características por paciente (6 sinais $\times$ (3 detalhes + 1 aproximação) $\times$ 2 estatísticas), enquanto na configuração principal MW5 se obtiveram 72 características (6 sinais $\times$ (5 detalhes + 1 aproximação) $\times$ 2 estatísticas), em contraste com as 60 características do trabalho original.

3.3.4 Transformada de Fourier de Curto Tempo (STFT)

Fundamentação Teórica

A STFT adapta a Transformada de Fourier para sinais não estacionários, dividindo o sinal em pequenos segmentos. O resultado é um espetrograma. A sua principal limitação é a janela de análise de tamanho fixo, que implica um compromisso entre a resolução temporal e a de frequência. A sua definição matemática é dada por:

$$\text{STFT}(t, f) = \int_{-\infty}^{\infty} x(\tau) \cdot w(\tau - t) \cdot e^{-j2\pi f \tau} d\tau \tag{3.4}$$

onde $x(\tau)$ é o sinal, $w(\tau - t)$ é a função janela deslocada no tempo, e f é a frequência.

Aplicação e Configuração

A implementação da STFT foi configurada com uma janela de Hamming de 30 amostras, aplicada com uma sobreposição de 50%. A FFT para cada janela foi calculada com 64 pontos (`nfft = 64`). As características extraídas consistiram na potência média do sinal em três bandas de frequência pré-definidas: 0–0.1 Hz, 0.1–0.2 Hz e 0.2–0.5 Hz.

3.3.5 Wavelet Packet Transform (WPT)

A WPT foi avaliada como uma generalização da DWT. Foi utilizada a mesma *wavelet-mãe* Daubechies 2 (`db2`), com uma decomposição até ao nível 3, resultando numa árvore com $2^3 = 8$ nós terminais. As características extraídas para cada sinal consistiram na energia normalizada de cada um destes 8 nós, resultando num vetor de 48 características por paciente.

3.3.6 Decomposição por Modos Empíricos (EMD)

A EMD foi a abordagem adaptativa avaliada. O sinal foi decomposto em Funções de Modo Intrínseco (IMFs), e as características extraídas consistiram na energia das quatro primeiras IMFs, resultando num vetor de 24 características por paciente.

3.4 Metodologia de Classificação e Avaliação

3.4.1 Codificação de Rótulos e Dimensões da Matriz de Características

Para a tarefa de classificação, cada amostra foi rotulada numericamente de 0 a 4, correspondendo às seguintes categorias: S (0), IC (1), SM (2), LP (3) e LM (4). O vetor de rótulos foi definido como $Y \in \{0, 1, 2, 3, 4\}^{75}$. A ordem (S, IC, SM, LP, LM) foi mantida de forma consistente em todas as matrizes e tabelas de resultados.

As matrizes de características finais para os modelos baseados em DWT, após a extração e concatenação das estatísticas de todos os sinais, tiveram as seguintes dimensões:

- Para a configuração MW3: $X_{MW3} \in \mathbb{R}^{75\times48}$.
- Para a configuração MW5: $X_{MW5} \in \mathbb{R}^{75\times72}$.

3.4.2 Máquinas de Vetores de Suporte (SVM)

Fundamentação Teórica

Foi utilizado o algoritmo de SVM (Boser et al., 1992), um método de aprendizagem supervisionada que encontra um Hiperplano de separação ótimo que maximiza a margem entre as classes. Para dados não linearmente separáveis, a "manobra do *kernel*" mapeia os dados para um espaço de dimensão superior.

- *Kernel* Linear: constrói um separador linear.
- *Kernel* Polinomial: cria fronteiras de decisão curvas.
- *Kernel* Gaussiano (RBF): cria fronteiras de decisão complexas e não lineares.

Aplicação e Configuração

Foram testadas as três funções de *kernel* (Linear, Polinomial e RBF). Para lidar com o problema de cinco classes, foi adotada a estratégia multiclasse *one-vs-one*, implementada através da função `fitcecoc` do MATLAB. Em todas as configurações, foi aplicada a padronização das características (`Standardize=true`) como etapa de pré-processamento. Para o *kernel* RBF, o parâmetro de escala foi definido como automático (`KernelScale='auto'`).

3.5 Protocolo de Avaliação

Os modelos foram avaliados por Validação Cruzada de 5 partições (*5-fold cross-validation*) com padronização das características ajustada apenas no conjunto de treino em cada *fold*. Como prática de verificação interna, executámos corridas adicionais em ambiente separado e observámos variação reduzida nas métricas agregadas, o que sustenta a representatividade dos resultados reportados. A métrica principal foi a (acurácia global) e, como métrica complementar, o *F1-Score* (macro).

3.6 Validação dos Sinais Semi-Sintéticos

Para garantir que os sinais gerados replicavam um comportamento fisiológico realista, foi conduzido um processo de validação. Foram selecionados nove pacientes com dados reais (três de cada grupo: IC, SM e S). As inclinações (s_i) para cada um dos quatro segmentos temporais (i) foram extraídas dos sinais reais usando a seguinte equação:

$$s_i = \frac{x_{\text{end},i} - x_{\text{start},i}}{N_i} \tag{3.5}$$

onde $x_{\text{end},i}$ e $x_{\text{start},i}$ representam os valores final e inicial do segmento i, e N_i é o número de pontos nesse segmento.

Com base nestes parâmetros, foram gerados novos sinais sintéticos, que foram depois comparados com os originais usando o coeficiente de correlação de Pearson (r) e a Raiz do Erro Quadrático Médio (*Root Mean Square Error*, RMSE) (Hodson, 2022), cuja fórmula é:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

onde y_i é o valor do sinal real, $\hat{y}_i$ o do sinal sintético e n o número de amostras. Para avaliar a fidelidade dos sinais gerados, estabeleceu-se como critério de aceitação um RMSE inferior a 15% da amplitude pico-a-pico dos sinais reais. Este valor foi considerado adequado, garantindo que o erro de reconstrução se manteve substancialmente abaixo da variabilidade fisiológica natural. Embora as variáveis Frequência Respiratória (*Respiratory Rate*, RR) e RER tenham apresentado uma correlação mais fraca, a validação geral confirmou que o processo preservava as características temporais relevantes.

4

Resultados e Discussão

Este capítulo apresenta e analisa os resultados experimentais obtidos a partir da metodologia de comparação descrita no capítulo anterior. A análise é dividida em duas partes: primeiro, uma apresentação detalhada do desempenho do *framework* de referência baseado na DWT; segundo, a avaliação comparativa deste *framework* em relação aos três métodos alternativos (STFT, WPT e EMD). A análise culmina numa discussão aprofundada que sintetiza as conclusões, aborda as limitações do estudo e explora as suas implicações e direções para trabalho futuro.

4.1 Resultados Experimentais

A fase experimental consistiu em avaliar quatro abordagens distintas de extração de características, mantendo um protocolo de classificação e validação consistente, conforme detalhado na Secção 3.4. O desempenho de cada transformada foi avaliado utilizando classificadores SVM com três *kernels* distintos (Linear, Polinomial e RBF).

4.1.1 Desempenho Detalhado do Modelo DWT

O modelo de referência, baseado na DWT, foi avaliado em duas configurações: com três (MW3) e cinco (MW5) níveis de decomposição. Embora o modelo MW3 já demonstrasse um desempenho robusto, a configuração MW5, proposta no artigo, alcançou melhores resultados em todas as métricas com o *kernel* Linear, o que resultou no melhor desempenho global entre os modelos testados. A Tabela 4.1 apresenta o desempenho detalhado do modelo `SVM-Linear-DWT-MW5`, discriminado por cada uma das cinco classes de diagnóstico.

Os resultados por classe demonstram que o modelo alcançou um desempenho quase perfeito na distinção dos grupos com dados reais (IC, SM e S), mas exibiu maior dificuldade com os grupos cujos dados foram gerados semi-sinteticamente (LP e LM).

Tabela 4.1: *Métricas de desempenho detalhadas para o modelo `SVM-Linear-DWT-MW5`, apresentadas por classe. Estes resultados correspondem ao melhor modelo identificado no artigo (Santos et al., 2025).*

Classe	Precisão	Recall	*F1-Score*	Acurácia
IC	93.33%	100.00%	96.55%	98.67%
SM	100.00%	93.75%	96.77%	98.67%
S	93.33%	100.00%	96.55%	98.67%
LP	73.33%	64.71%	68.75%	86.67%
LM	60.00%	64.29%	62.07%	85.33%
Média Macro	**84.00%**	**84.55%**	**84.14%**	**93.60%**

4.1.2 Análise Comparativa dos Métodos de Extração

Para validar a DWT como uma abordagem robusta, o seu desempenho foi comparado com os métodos baseados em STFT, WPT e EMD. A Tabela 4.2 apresenta os resultados completos desta análise, utilizando o modelo `SVM-Linear-DWT-MW5` como *baseline*.

Tabela 4.2: *Resultados comparativos de desempenho para os quatro métodos de extração de características. Conforme a metodologia, os resultados reportados correspondem ao desempenho obtido no protocolo de validação. A linha a negrito indica o modelo com o melhor desempenho geral.*

Transformada	Kernel	Acurácia	Precisão	Recall	*F1-Score*
DWT (Artigo)	**Linear**	**93.60%**	**84.00%**	**84.55%**	**84.14%**
STFT	Linear	69.33%	69.86%	69.33%	69.07%
	Polinomial	74.67%	75.33%	74.67%	**74.25%**
	RBF	65.33%	68.27%	65.33%	65.29%
WPT	Linear	53.33%	53.02%	53.33%	52.87%
	Polinomial	57.33%	57.84%	57.33%	**57.01%**
	RBF	49.33%	51.77%	49.33%	50.12%
EMD	Linear	61.33%	62.32%	61.33%	**60.59%**
	Polinomial	56.00%	55.34%	56.00%	55.09%
	RBF	50.67%	55.74%	50.67%	51.48%

A análise da Tabela 4.2 indica uma vantagem de desempenho consistente para a abordagem da DWT em comparação com as outras transformadas. O modelo `SVM-Linear-DWT` alcançou a melhor performance global, com uma acurácia de 93.60% e um *F1-Score* de 84.14%. Este resultado é aproximadamente 19 pontos percentuais superior em acurácia e 10 pontos em *F1-Score* à melhor alternativa, a STFT (com *kernel* Polinomial), que obteve uma acurácia de 74.67%. As abordagens baseadas em WPT e EMD revelaram-se consideravelmente menos eficazes.

4.2 Discussão

O desempenho mais elevado do método DWT reforça a hipótese de que a sua capacidade de análise multirresolução é a chave para modelar as complexas dinâmicas fisiológicas do CPET. Ao contrário da STFT com a sua janela fixa, a DWT consegue captu-

rar tanto as tendências lentas como os eventos rápidos com uma resolução adequada para cada escala, extraindo características mais ricas. A STFT, sendo a segunda melhor abordagem, mostra que a informação no domínio da frequência é relevante, mas a sua resolução fixa limita o seu potencial. O desempenho inferior da WPT e da EMD sugere que estas técnicas podem gerar um espaço de características demasiado complexo ou ruidoso para um *dataset* de dimensão limitada.

> ***Nota***
>
> O principal trunfo da DWT em sinais não estacionários é a análise multirresolução: "janelas" estreitas para eventos rápidos (altas frequências) e largas para tendências lentas (baixas frequências), ao contrário da janela fixa da STFT (Addison, 2005).

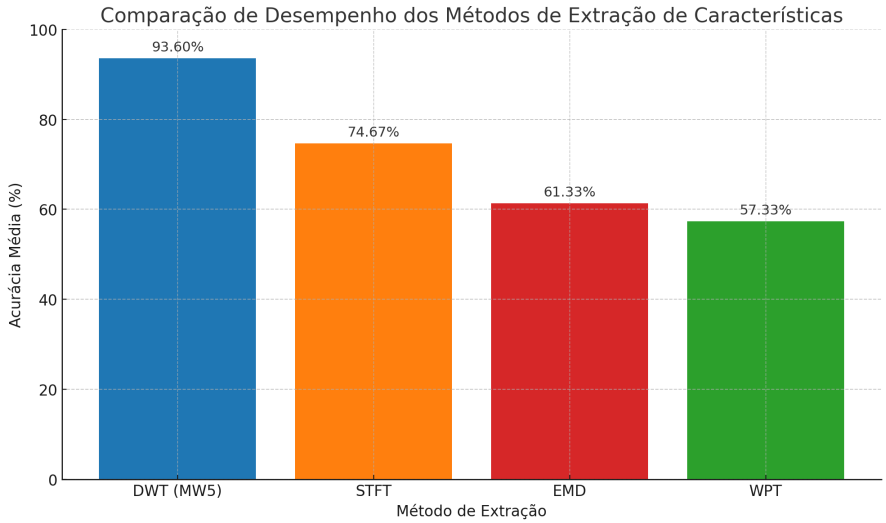

Figura 4.1: *Comparação da Acurácia (Macro) para o melhor modelo de cada método de extração de características. Os resultados destacam o desempenho da abordagem baseada em DWT em comparação com as alternativas. Fonte: Elaborado pelo autor.*

É relevante contextualizar estes resultados com os do trabalho de referência de 3 classes (Pinheiro et al., 2025). Nesse estudo, o modelo MW5 alcançou 92.00% de acurácia e 88.00% de *F1-Score*. O modelo expandido para 5 classes, mesmo com uma complexidade de classificação superior, apresentou um valor de acurácia superior (93.60%) e manteve um *F1-Score* robusto (84.14%), indicando que a expansão do modelo para mais classes não comprometeu severamente o seu desempenho.

Numa perspetiva clínica, o modelo proposto poderia servir como uma ferramenta de apoio à decisão, ajudando a orientar um diagnóstico mais direcionado para pacientes com intolerância ao exercício não específica. Contudo, a transição para um ambiente real enfrenta desafios como a heterogeneidade de dados provenientes de diferentes equipamentos, protocolos de teste e populações de pacientes com múltiplas comorbidades, para as quais a robustez do modelo ainda precisa ser determinada.

4.3 Limitações do Estudo

> ***Nota***
>
> A discussão transparente das limitações não é sinal de fraqueza, mas de rigor científico. Explicitar como erros, viéses e decisões metodológicas podem condicionar os resultados ajuda a contextualizar a evidência e a credibilidade das conclusões (Ioannidis, 2007).

É imperativo discutir as limitações inerentes a este estudo. A principal limitação continua a ser o tamanho restrito do *dataset* (15 amostras por classe). Embora justificada para permitir a comparação direta, esta dimensão impõe um limite estatístico à generalização dos modelos.

Outra limitação crítica reside na fidelidade dos sinais semi-sintéticos. Como visto na Tabela 4.1 e consistente em todos os modelos, a dificuldade em distinguir as classes LP e LM sugere que as descrições estatísticas usadas para a sua geração podem ser demasiado semelhantes. O modelo gerador utilizado, baseado em segmentos lineares, é uma simplificação e não captura dinâmicas não lineares complexas, o que é particularmente evidente na baixa correlação obtida durante a validação para os sinais de Frequência Respiratória (*Respiratory Rate*, RR) (RR) e RER.

Adicionalmente, a estrutura de comprimento fixo dos sinais sintéticos (200 pontos), embora necessária para padronizar a extração de características, não capta a variação natural na duração das sessões de CPET entre pacientes, o que representa um compromisso a ser considerado.

4.4 Conclusões e Trabalhos Futuros

Os resultados desta dissertação indicam que o *framework* de diagnóstico foi expandido de 3 para 5 classes e que a técnica DWT obteve o desempenho mais elevado entre as alternativas testadas (STFT, WPT, EMD) nas métricas de avaliação utilizadas.

As direções para investigação futura são claras:

- Otimização de Hiperparâmetros: Aplicar técnicas de otimização de Hiperparâmetros (ex.: otimização *bayesiana*) para extrair o desempenho máximo de cada combinação de transformada e *kernel*.
- Modelos Gerativos Avançados: Explorar modelos como *Generative Adversarial Network*s (GANs) ou *Variational Autoencoder*s (VAEs) para produzir sinais sintéticos com uma dinâmica mais realista.
- Validação em *datasets* de Maior Dimensão: Validar o *framework* em *datasets* multicêntricos e mais vastos para avaliar a sua capacidade de generalização.
- Exploração de Arquiteturas de *Deep Learning*: Comparar o *baseline* estabelecido com arquiteturas de *deep learning* (CNNs e *Recurrent Neural Network*s (RNNs)).

- Interpretabilidade dos Modelos: Aplicar técnicas de Inteligência Artificial Explicável (*Explainable Artificial Intelligence*, XAI), como *Local Interpretable Model-agnostic Explanations* (LIME) e *SHapley Additive exPlanations* (SHAP), para identificar as características mais importantes para cada diagnóstico, aumentando a confiança e a aceitação clínica da ferramenta.

5 Conclusões

5.1 Conclusões

Esta dissertação teve como objetivo central a expansão e validação de um *framework* de diagnóstico baseado em SVM e DWT, aplicando-o a um problema de classificação multiclasse mais complexo. O trabalho foi estruturado em duas frentes principais: primeiro, a expansão do modelo de três para cinco classes clínicas (IC, SM, S, LP e LM) através da geração de dados semi-sintéticos; segundo, a realização de uma análise comparativa sistemática para validar a eficácia da DWT em relação a outras técnicas de extração de características.

A primeira grande conclusão deste trabalho é que a combinação de dados reais com dados semi-sintéticos, guiados por parâmetros estatísticos de pacientes reais, constitui uma estratégia viável para superar a escassez de dados, o que permitiu expandir o modelo de diagnóstico para novas classes clínicas e aumentar a sua relevância e aplicabilidade.

A segunda conclusão principal é que os resultados da análise comparativa indicam que a DWT foi o método de extração de características com o desempenho mais elevado para esta tarefa. O modelo de referência, `SVM-Linear-DWT-MW5`, obteve um resultado global robusto com uma acurácia macro-média de 93.60%, precisão de 84.00%, *recall* de 84.55% e um *F1-Score* de 84.14%. Este desempenho foi numericamente superior ao das alternativas baseadas em STFT, WPT e EMD. A análise também demonstrou que níveis de decomposição *wavelet* mais profundos (MW5) resultaram num melhor desempenho da classificação. Estes achados reforçam a validade da escolha metodológica do trabalho de referência e a eficácia da abordagem para um problema mais complexo.

Em suma, este trabalho alcançou os seus objetivos de expandir o *framework* de diagnóstico e, através de uma comparação metodologicamente consistente, validou a abordagem baseada em DWT como a de melhor desempenho neste estudo, estabelecendo assim um *baseline* para os trabalhos futuros detalhados no capítulo anterior.

Página intencionalmente deixada em branco.

Bibliografia

Addison, P. S. (2005). «Wavelet transforms and the ECG: a review». Em: *Physiological Measurement* 26, R155–R199. DOI: 10.1088/0967-3334/26/5/R01.

Al Rahhal, M. M. et al. (2016). «Deep learning approach for active classification of electrocardiogram signals». Em: *Information Sciences* 345, pp. 340–354. DOI: 10.1016/j.ins.2016.01.082.

Azizi, T. (2024). «Comparative analysis of statistical, time-frequency, and SVM techniques for change detection in nonlinear biomedical signals». Em: *Signals* 5.4, pp. 736–755. DOI: 10.3390/signals5040041.

Boser, B. E., I. M. Guyon e V. N. Vapnik (1992). «A Training Algorithm for Optimal Margin Classifiers». Em: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. COLT '92. New York, NY, USA: ACM, pp. 144–152. ISBN: 0-89791-498-8. DOI: 10.1145/130385.130401.

Brown, D. E. et al. (2022). «Neural network methods for diagnosing patient conditions from cardiopulmonary exercise testing data». Em: *BioData Mining* 15.16. DOI: 10.1186/s13040-022-00299-6.

Choudhary, S. e S. S. Mehta (nov. de 2017). «Analysis of Automatic Classification of Electrocardiogram (ECG) Beats Using Wavelet Transform and SVM and PCA-SVM». Em: *International Journal of Advance Engineering and Research Development* 4.11. URL: https://www.ijaerd.org/index.php/IJAERD/article/view/4120.

Coifman, R. R. e M. V. Wickerhauser (1992). «Entropy-based algorithms for best basis selection». Em: *IEEE Transactions on Information Theory* 38.2, pp. 713–718. DOI: 10.1109/18.119732.

Cortes, C. e V. Vapnik (1995). «Support-vector networks». Em: *Machine Learning* 20.3, pp. 273–297. DOI: 10.1007/BF00994018.

Demšar, J. (2006). «Statistical comparisons of classifiers over multiple data sets». Em: *Journal of Machine Learning Research* 7, pp. 1–30. URL: http://www.jmlr.org/papers/v7/demsar06a.html.

Deng, J. et al. (2022). «Automatic Cardiopulmonary Endurance Assessment: A Machine Learning Approach Based on GA-XGBOOST». Em: *Diagnostics* 12.10, p. 2538. DOI: 10.3390/diagnostics12102538.

Franssen, R. F. W. et al. (2023). «Influence of different data-averaging methods on mean values of selected variables derived from preoperative cardiopulmonary exercise testing in patients scheduled for colorectal surgery». Em: *PLOS ONE* 18.3, e0283129. DOI: 10.1371/journal.pone.0283129.

Hodson, T. O. (2022). «Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not». Em: *Geoscientific Model Development* 15.15, pp. 5481–5487. DOI: 10.5194/gmd-15-5481-2022.

Ioannidis, J. P. A. (2007). «Limitations are not properly acknowledged in the scientific literature». Em: *Journal of Clinical Epidemiology* 60.4, pp. 324–329. DOI: 10.1016/j.jclinepi.2006.09.011.

McSharry, P. E. et al. (2003). «A dynamical model for generating synthetic electrocardiogram signals». Em: *IEEE Transactions on Biomedical Engineering* 50.3, pp. 289–294. DOI: 10.1109/TBME.2003.808805.

Oppenheim, A. V., R. W. Schafer e J. R. Buck (1999). *Discrete-Time Signal Processing*. 2nd. Upper Saddle River, NJ, USA: Prentice Hall. ISBN: 0-13-754920-2.

Page, M. J. et al. (2021). «The PRISMA 2020 statement: an updated guideline for reporting systematic reviews». Em: *BMJ* 372, n71. DOI: 10.1136/bmj.n71.

Pezoulas, V. C. et al. (2024). «Synthetic data generation methods in healthcare: a review on open-source tools and methods». Em: *Computational and Structural Biotechnology Journal* 23, pp. 2892–2910. DOI: 10.1016/j.csbj.2024.07.005.

Pinheiro, R. F. e R. Fonseca-Pinto (2023). «Algorithm for diagnosis of metabolic syndrome and heart failure using CPET biosignals via SVM and wavelet transforms». Em: *International Conference on Smart Objects and Technologies for Social Good*. Springer, pp. 159–171. DOI: 10.1007/978-3-031-52524-7_12.

Pinheiro, R. F. e R. Fonseca-Pinto (2025). «On the development of diagnostic support algorithms based on CPET biosignals data via machine learning and wavelets». Em: *PeerJ Computer Science* 11, e2474. DOI: 10.7717/peerj-cs.2474.

Pinheiro, R. F., R. Fonseca-Pinto e D. Colón (2024). «A review of the Lurie problem and its applications in the medical and biological fields». Em: *AIMS Mathematics* 9.11, pp. 32962–32999. DOI: 10.3934/math.20241577.

Polikar, R. (1996). *The Wavelet Tutorial*. Rowan University; PDF arquivado. URL: https://cseweb.ucsd.edu/~baden/Doc/wavelets/polikar_wavelets.pdf.

Portella, J. J. et al. (2022). «Using Machine Learning to Identify Organ System Specific Limitations to Exercise via Cardiopulmonary Exercise Testing». Em: *IEEE Journal of Biomedical and Health Informatics* 26.8, pp. 4228–4237. DOI: 10.1109/JBHI.2022.3163402.

PRISMA Statement (2024). *PRISMA 2020 flow diagram templates*. https://www.prisma-statement.org/prisma-2020-flow-diagram. Acesso em: 16 de setembro de 2025. Template licenciado sob CC BY 4.0.

Ramirez, Alfredo e Charles Cox (2012). «Improving on the Range Rule of Thumb». Em: *Rose-Hulman Undergraduate Mathematics Journal* 13.2, Article 1. URL: https://scholar.rose-hulman.edu/rhumj/vol13/iss2/1/.

Santos, F. B., R. F. Pinheiro e R. Fonseca-Pinto (2025). «Semi-Synthetic Expansion of CPET Time-Series for Multiclass Diagnosis Using SVM and Wavelet Transform». Em: *16th IEEE/IAS International Conference on Industry Applications* (*INDUSCON*). Aceite para publicação. URL: https://induscon.org/.

Sasse, L. et al. (2023). *Data Leakage in Machine Learning: A Tutorial and Taxonomy*. arXiv: 2311.04179 [cs.LG]. URL: https://arxiv.org/abs/2311.04179.

Schmitt, J. et al. (2014). «2D Hilbert-Huang transform». Em: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*). IEEE, pp. 3720–3724. DOI: 10.1109/ICASSP.2014.6854630.

Schwendinger, F. et al. (2024). «Using Machine Learning-Based Algorithms to Identify and Quantify Exercise Limitations in Clinical Practice: Are We There Yet?» Em: *Medicine & Science in Sports & Exercise* 56.2, pp. 159–169. DOI: 10.1249/MSS.0000000000003293.

Subasi, A. e M. I. Gursoy (2010). «EEG signal classification using PCA, ICA, LDA and support vector machines». Em: *Expert Systems with Applications* 37.12, pp. 8659–8666. DOI: 10.1016/j.eswa.2010.06.065.

Wasserman, K., J. E. Hansen e D. Y. Sue (2011). *Principles of exercise testing and interpretation: including pathophysiology and clinical applications*. 5th. Wolters Kluwer Health.

West, M. A. et al. (2024). «Cardiopulmonary exercise variables and their association with postoperative morbidity and mortality after major oesophagogastric cancer surgery-a multicentre observational study». Em: *BJA Open* 10, p. 100289. DOI: 10.1016/j.bjao.2024.100289.

Página intencionalmente deixada em branco.

Apêndices

Página intencionalmente deixada em branco.

A

Códigos-Fonte Principais

Neste apêndice, são apresentados os códigos-fonte mais relevantes desenvolvidos para este trabalho, demonstrando a metodologia de geração de dados, a extração de características para cada transformada e a avaliação consistente dos modelos de classificação.

A.1 Geração e Validação de Dados Semi-Sintéticos

A Listagem 1 apresenta o *script* utilizado para gerar e validar os sinais semi-sintéticos. O *script* lê parâmetros estatísticos de pacientes reais, gera um sinal artificial através da função `gerarSerie` e, em seguida, compara o sinal gerado com o sinal real correspondente para calcular a correlação e o RMSE, validando assim o processo.

A.2 Extração de Características e Avaliação Comparativa

As listagens seguintes demonstram os scripts utilizados para a extração de características de cada uma das quatro transformadas. A avaliação de cada modelo seguiu um protocolo robusto baseado em validação cruzada (5-*fold*).

Listagem 1: *Script principal para a geração e validação dos dados semi-sintéticos.*

```
% Script para Geração e Validação de Sinais Semi-Sintéticos
statsFile = 'Caminho/Para/FicheiroDeEstatisticas.xlsx';
realFile  = 'Caminho/Para/FicheiroDeDadosReais.xlsx';
numPontos = 200;
tableStats = readtable(statsFile, 'VariableNamingRule', 'preserve');
realData = readtable(realFile);
variaveis = {'HR','VO2','VCO2','RR','VE','RER'};
for paciente = 1:9
    for v = 1:length(variaveis)
        % (Lógica para extrair stats e nomes de colunas omitida)
        realSig = realData.(colName);
        realSig = realSig(~isnan(realSig));
        if length(realSig) > numPontos
            realSig = realSig(1:numPontos);
        elseif length(realSig) < numPontos
            realSig(end+1:numPontos) = realSig(end);
        end
        artSig = gerarSerie(minVal, maxVal, stdVal, slopes, numPontos);
        rmse = sqrt(mean((realSig - artSig).^2));
        [r, p] = corrcoef(realSig, artSig);
        fprintf('Correlação: r = %.2f (p = %.4f)\n', r(1,2), p(1,2));
    end
end
function s = gerarSerie(minVal, maxVal, stdVal, slopes, N)
    q = floor(N / 4);
    y = zeros(1, 5);
    y(1) = minVal;
    y(2) = y(1) + slopes(1) * q;
    y(3) = y(2) + slopes(2) * q;
    y(4) = y(3) + slopes(3) * q;
    y(5) = y(4) + slopes(4) * (N - 3*q);
    x = linspace(1, N, 5);
    s = spline(x, y, 1:N)';
    t = (1:N)';
    osc = stdVal * 0.1 * sin(2 * pi * 0.01 * t);
    noise = stdVal * 0.2 * randn(N, 1);
    s = s + osc + noise;
    s = min(max(s, minVal), maxVal);
end
```

Listagem 2: *Script de avaliação para o modelo de referência `SVM-DWT-MW5`.*

```
% Script de Avaliação para DWT (Modelo de Referência)
% (Carregamento de dados omitido por ser idêntico aos outros scripts)
W_dwt = zeros(75, 72);
for i = 1:75
    signals = {HR(1:n(i,1), i), VO(1:n(i,2), i), VCO(1:n(i,3), i), RR(1:n(i,4), i), VE(1:n(i,5), i), RER(1:n(i,6), i)};
    feature_vector = [];
    for s = 1:length(signals)
        [C, L] = wavedec(signals{s}, 5, 'db2');
        approx_coeffs = appcoef(C, L, 'db2');
        means_details = zeros(1, 5);
        vars_details = zeros(1, 5);
        for level = 1:5
            detail_coeffs = detcoef(C, L, level);
            means_details(level) = mean(detail_coeffs);
            vars_details(level) = var(detail_coeffs);
        end
        signal_features = [means_details, mean(approx_coeffs), vars_details, var(approx_coeffs)];
        feature_vector = [feature_vector, signal_features];
    end
    W_dwt(i,:) = feature_vector;
end
% (Secção de avaliação SVM idêntica aos scripts seguintes, usando W_dwt)
```

Listagem 3: *Script de avaliação para os modelos `SVM-STFT`.*

```
% Script de Avaliação para STFT
% (Carregamento de dados omitido)
fs = 1; window_length = 30; overlap_percent = 50; nfft = 64;
freq_bands = [0, 0.1; 0.1, 0.2; 0.2, 0.5];
num_bands = size(freq_bands, 1);
W_stft = zeros(75, 6 * num_bands);
for i = 1:75
    signals = {HR(1:n(i,1), i), VO(1:n(i,2), i), VCO(1:n(i,3), i), RR(1:n(i,4), i), VE(1:n(i,5), i), RER(1:n(i,6), i)};
    feature_vector = [];
    for s = 1:length(signals)
        current_signal = signals{s};
        if length(current_signal) < window_length
            signal_features = zeros(1, num_bands);
        else
            [S, F, ~] = stft(current_signal, fs, 'Window', hamming(window_length), 'OverlapLength', floor(window_length*overlap_percent/100), 'FFTLength', nfft);
            psd = abs(S).^2 / nfft;
            band_powers = zeros(1, num_bands);
            for b = 1:num_bands
                freq_indices = F >= freq_bands(b, 1) & F < freq_bands(b, 2);
                band_powers(b) = mean(mean(psd(freq_indices, :)));
            end
            signal_features = band_powers;
        end
        feature_vector = [feature_vector, signal_features];
    end
    W_stft(i, :) = feature_vector;
end
% (Secção de avaliação SVM idêntica, usando W_stft)
```

Listagem 4: *Script de avaliação para os modelos SVM-WPT.*

```
% Script de Avaliação para WPT
% (Carregamento de dados omitido)
wavelet_name = 'db2'; level = 3;
num_features_per_signal = 2^level;
W_wpt = zeros(75, 6 * num_features_per_signal);
for i = 1:75
    signals = {HR(1:n(i,1), i), VO(1:n(i,2), i), VCO(1:n(i,3), i), RR(1:n(i,4), i), VE(1:n(i,5),
    ↪  i), RER(1:n(i,6), i)};
    feature_vector = [];
    for s = 1:length(signals)
        current_signal = signals{s};
        if length(current_signal) < 2^level
            signal_features = zeros(1, num_features_per_signal);
        else
            wpt = wpdec(current_signal, level, wavelet_name);
            energies = wenergy(wpt);
            total_energy = sum(energies);
            if total_energy > 0
                signal_features = energies / total_energy;
            else
                signal_features = zeros(1, num_features_per_signal);
            end
        end
        feature_vector = [feature_vector, signal_features];
    end
    W_wpt(i, :) = feature_vector;
end
% (Secção de avaliação SVM idêntica, usando W_wpt)
```

Listagem 5: *Script de avaliação para os modelos SVM-EMD.*

```
% Script de Avaliação para EMD
% (Carregamento de dados omitido)
num_imfs_to_use = 4;
W_emd = zeros(75, 6 * num_imfs_to_use);
for i = 1:75
    signals = {HR(1:n(i,1), i), VO(1:n(i,2), i), VCO(1:n(i,3), i), RR(1:n(i,4), i), VE(1:n(i,5),
    ↪  i), RER(1:n(i,6), i)};
    feature_vector = [];
    for s = 1:length(signals)
        signal_features = zeros(1, num_imfs_to_use);
        try
            imfs = emd(signals{s});
            num_imfs_generated = size(imfs, 2);
            for imf_idx = 1:min(num_imfs_generated, num_imfs_to_use)
                signal_features(imf_idx) = sum(imfs(:, imf_idx).^2);
            end
        catch
            fprintf('Aviso: EMD falhou para o sinal %d do paciente %d.\n', s, i);
        end
        feature_vector = [feature_vector, signal_features];
    end
    W_emd(i, :) = feature_vector;
end
% (Secção de avaliação SVM idêntica, usando W_emd)
```

B

Artigo Aceite para Publicação

O trabalho, intitulado *Semi-Synthetic Expansion of CPET Time-Series for Multiclass Diagnosis Using SVM and Wavelet Transform*, foi aceite para apresentação e publicação na conferência **16th IEEE/IAS International Conference on Industry Applications** (**INDUSCON**). O artigo será disponibilizado na biblioteca digital **IEEE Xplore**, uma base de dados indexada à **Scopus**.

Semi-Synthetic Expansion of CPET Time-Series for Multiclass Diagnosis Using SVM and Wavelet Transform

1st Flávio Bueno dos Santos
School of Technology and Management (ESTG)
Polytechnic University of Leiria
Leiria, Portugal
2230395@my.ipleiria.pt

2nd Rafael Fernandes Pinheiro
Center for Innovative Care and Health Technology (ciTechCare), School of Health Sciences (ESSLei)
Polytechnic University of Leiria
Leiria, Portugal
rafael.f.pinheiro@ipleiria.pt

3rd Rui Fonseca-Pinto
Center for Innovative Care and Health Technology (ciTechCare), School of Health Sciences (ESSLei)
Polytechnic University of Leiria
Instituto de Telecomunicacoes (IT),
Leiria, Portugal
rui.pinto@ipleiria.pt

***Abstract*—Cardiopulmonary Exercise Testing (CPET) provides critical physiological data for diagnosing limitations across cardiac, pulmonary, and musculoskeletal systems, and detecting physiological alterations related to conditions such as metabolic syndrome. However, the lack of full time-series data in some clinical groups limits the scalability of machine learning-based diagnostic tools. This study enhances a classification framework based on Support Vector Machines (SVM) and wavelet-based feature extraction by generating synthetic CPET time-series to complement existing real datasets. Artificial signals for Pulmonary and Musculoskeletal limitations were generated using real patient-derived statistics (mean, standard deviation, slopes, and peak timing), ensuring that the synthetic data reflect true physiological patterns. These were combined with full time-series CPET data from real Healthy, Heart Failure, and Metabolic Syndrome subjects, forming a five-class classification model. Classification was conducted using SVM with three wavelet decomposition levels (MW3) and five levels (MW5). The best model achieved an accuracy of 93.60%, precision of 84.00%, recall of 84.55%, and F1-score of 84.14% using MW5.**

***Index Terms*—cardiopulmonary exercise testing (CPET), support vector machines (SVM), wavelet transform, synthetic data generation, physiological time-series, clinical decision support**

I. Introduction

The Cardiopulmonary Exercise Test (CPET) is an essential tool for evaluating the functional limitations of patients, providing detailed physiological insights into cardiac, pulmonary, and musculoskeletal systems [1]. Traditional CPET analyses are often based on specific threshold variables or peak values, potentially overlooking subtle patterns embedded in the full exercise time-series data [2].

Recent advances have introduced machine learning techniques, such as Support Vector Machines (SVM) combined with wavelet-based feature extraction, to classify different physiological limitations more accurately [3]. Other approaches, such as neural networks, have also been explored for CPET data analysis [4]. A previous study successfully applied this approach to distinguish between Heart Failure (HF), Metabolic Syndrome (MS), and Healthy subjects [5]. However, that study did not include the Respiratory Exchange Ratio (RER), due to its unavailability in the healthy dataset.

Furthermore, full time-series data were not available for certain clinical groups, such as Pulmonary and Musculoskeletal limitations, which restricted the scalability of the model.

To overcome current limitations, we propose a method that supplements full time-series data from real patients with semi-synthetic signals generated from patient-derived statistics (mean, standard deviation, slopes, and peak timing). These signals reflect physiologically plausible patterns and enable the expansion of the model to new diagnostic categories. In this work, we refer to these signals as "semi-synthetic" because they are not entirely artificial, but rather generated using real patient-derived physiological statistics (e.g., mean, standard deviation, and signal slope), ensuring that they preserve plausible temporal and physiological patterns.

In this work, we include the RER variable in the analysis pipeline, using both real RER values (available for Pulmonary and Musculoskeletal groups) and derived values for the Healthy group—calculated directly as the ratio of VCO_2 to VO_2. This inclusion enables a comprehensive evaluation of RER's contribution to model performance.

This study expands the original three-class SVM-Wavelet model [5] to classify five distinct groups: HF, MS, Healthy, Pulmonary Limitation (PulmonaryLim), and Musculoskeletal Limitation (MSK), using both real and semi-synthetic data. Classification is performed using three- and five-level wavelet decomposition (MW3 and MW5), and metrics are compared with and without RER to assess its impact. In summary, this work presents the following contributions:

- Integration of two new diagnostic groups (PulmonaryLim and MusculoskeletalLim) into the existing CPET classification model;
- Physiologically guided generation of semi-synthetic CPET time-series based on patient-derived statistics;
- Inclusion of the RER variable, both measured and derived, for full comparative evaluation;
- Validation of five-class SVM-Wavelet models using MW3 and MW5 decomposition levels.

The remainder of this paper is organized as follows: Section II describes the dataset, semi-synthetic data generation strategy, and wavelet-based feature extraction. Section III presents the classification results. Section IV discusses the findings and implications. Section V concludes the study and outlines directions for future research.

II. Materials and Methods

A. Dataset Description

The study utilized Cardiopulmonary Exercise Testing (CPET) data from five diagnostic groups: Heart Failure (HF), Metabolic Syndrome (MS), Healthy subjects, Pulmonary Limitation (PulmonaryLim), and Musculoskeletal Limitation (MSK).

Data for the HF and MS groups were obtained from a publicly available GitHub repository associated with the work of Brown et al. [4]. This repository includes full-length CPET time-series containing key physiological variables such as heart rate (HR), oxygen consumption (VO_2), carbon dioxide production (VCO_2), respiratory rate (RR), ventilation (VE), and the respiratory exchange ratio (RER). Data for the Healthy group were obtained from a different dataset, as described in [5]. However, RER values were not explicitly available for the Healthy group.

To address this, RER was calculated manually for the Healthy group as:

$$\text{RER} = \frac{\text{VCO}_2}{\text{VO}_2} \tag{1}$$

This method is physiologically valid and widely accepted in cardiopulmonary diagnostic literature [1].

Data for the PulmonaryLim and MSK groups were generated semi-synthetically using physiological statistics extracted from real patients, available through the IEEE DataPort repository [2].

The final dataset was balanced across all five groups, totaling 75 patients, with 15 subjects in each class to ensure fair representation and to minimize training bias.

It is important to note that the choice of using 15 samples per class was a deliberate methodological decision. This approach ensures a direct and fair comparison with the studies [3]–[5], which also adopted exactly the same sample size. By keeping the dataset size consistent, the present study can reliably assess whether the semi-synthetic expansion and wavelet analysis provide advantages over established methodologies, without the results being influenced by a larger or different dataset. To guarantee a robust evaluation of the model's performance, k-fold cross-validation was employed, as described in this study.

B. Semi-Synthetic Data Generation

Due to the lack of raw CPET time-series data for the PulmonaryLim and MSK groups, semi-synthetic signals were generated using key statistical parameters derived from real clinical data [2], inspired by prior statistical modeling approaches for physiological time-series reconstruction [6]. These parameters included the mean, minimum, maximum, slope (i.e., rate of change or signal inclination), and, when necessary, standard deviation.

In variables where standard deviation (σ) was not explicitly provided, it was estimated using the following approximation:

$$\sigma \approx \frac{\text{Amplitude}}{4} \tag{2}$$

This heuristic is based on the empirical rule for normal distributions, where most values fall within $\pm 2\sigma$ of the mean. For variables such as HR, where σ was available directly, no approximation was needed.

Each synthetic signal was constructed by:

- Dividing the 200-point time-series into four segments;
- Applying slope values to each segment to reflect physiological transitions;
- Adding controlled oscillations and Gaussian noise to introduce variability;
- Clipping values between physiological minimum and maximum to ensure realism.

All semi-synthetic signals were generated with exactly 200 data points to maintain consistency in wavelet decomposition. While this standardization removes the natural variation in CPET session duration, it enables uniform extraction of wavelet features and facilitates statistical comparability.

C. Feature Extraction Using Wavelets

Each time-series signal (HR, VO_2, VCO_2, RR, VE) was processed using the Discrete Wavelet Transform (DWT) [7], a widely adopted method for analyzing non-stationary biomedical signals. The Daubechies 2 (db2) mother wavelet was applied, and two configurations were explored:

- MW3 – three decomposition levels (configuration);
- MW5 – five decomposition levels (extended configuration).

Statistical features, including the mean and variance of both approximation and detail coefficients at each level, were extracted and concatenated into the final feature matrix. This allowed dimensionality reduction while preserving important time-frequency information. The effectiveness of wavelet-based analysis for physiological signals has been demonstrated both in the extraction of statistical features for classification tasks [8], and in real-time detection of biomedical signal patterns such as ECG QRS complexes [9].

D. Label Encoding and Feature Matrix Dimensions

The dataset was structured for a five-class classification task. Each sample was labeled numerically from 0 to 4, corresponding to the diagnostic categories: Healthy (0), Heart Failure (1), Metabolic Syndrome (2), Pulmonary Limitation (3), and Musculoskeletal Limitation (4).

The final feature matrices used for training the SVM classifiers varied according to the wavelet decomposition level:

- *MW3 configuration*: Includes six physiological signals (HR, VO_2, VCO_2, RR, VE, RER). After three-level wavelet decomposition and statistical feature extraction, the resulting matrix had dimensions $\mathbf{X}_{\text{MW3}} \in \mathbb{R}^{75\times 48}$.
- *MW5 configuration*: Includes five physiological signals (HR, VO_2, VCO_2, RR, VE). After five-level wavelet decomposition and feature extraction, the resulting matrix had dimensions $\mathbf{X}_{\text{MW5}} \in \mathbb{R}^{75\times 72}$.

The label vector was defined as $\mathbf{Y} \in \{0, 1, 2, 3, 4\}^{75}$.

E. Classification with Support Vector Machines

The extracted features were used to train multiclass Support Vector Machine (SVM) classifiers. Three different kernel functions were tested [10]:

- Linear Kernel
- Polynomial Kernel

- Radial Basis Function (RBF) Kernel

A one-vs-one multiclass strategy was used, and all models were evaluated using 5-fold cross-validation. The performance metrics computed included accuracy, precision, recall, and F1-score, both per class and as macro-averages.

Metric definitions. Accuracy refers to the overall proportion of correct predictions. Precision measures the proportion of true positives among predicted positives. Recall indicates the proportion of actual positives correctly identified. F1-score is the harmonic mean of precision and recall. .

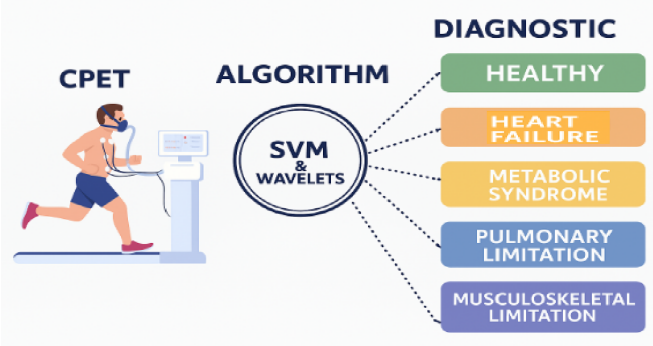

Fig. 1. Base methodology using SVM + Wavelets. The current work extends this to include semi-synthetic CPET data for Pulmonary and Musculoskeletal limitations.

F. Software Tools

All preprocessing, synthetic signal generation, feature extraction, and classification were conducted using MATLAB R2020a [11].

G. Validation of Semi-Synthetic Signals

To ensure that the semi-synthetic time-series accurately replicated realistic physiological behavior, a validation process was carried out using patients with real CPET time-series data available in the HF, MS, and Healthy groups. From these real signals, statistical summaries (mean, minimum, maximum, and standard deviation) and slope information were extracted. The slopes were calculated by dividing each signal into four temporal segments (quartiles) and computing the rate of change in each segment, which was then used to reconstruct synthetic time-series mimicking the dynamics of the original signals.

For this validation, three patients from each of the three groups (HF, MS, Healthy) were selected, totaling nine patients. This sample was considered sufficient to evaluate whether the signal generation method preserved key physiological dynamics, as the goal was not to train predictive models but to verify the fidelity of the generation process. Despite the simple selection strategy, the subset captured sufficient inter-subject variability to validate the method effectively.

The slope s_i for each segment i was computed as:

$$s_i = \frac{x_{\text{end},i} - x_{\text{start},i}}{N_i} \tag{3}$$

where $x_{\text{end},i}$ and $x_{\text{start},i}$ represent the final and initial values of segment i, and N_i is the number of time points in that segment.

The semi-synthetic series were then compared to their original real counterparts using Pearson's r and RMSE. This process allowed evaluation of how closely the simulated signals followed the temporal structure of actual physiological data.

Although the PulmonaryLim and MSK groups did not have available raw time-series data, the same statistical methodology was applied to generate their synthetic signals. The validation using real signals confirmed that the simulation process was capable of preserving relevant temporal characteristics, thereby supporting the integration of semi-synthetic data into the classification model.

III. Results

This section presents the classification results obtained from Support Vector Machine (SVM) models trained on the combined dataset, which includes both real and semi-synthetic CPET data. Three kernel functions were evaluated: Linear, Polynomial, and Radial Basis Function (RBF).

The classification task involved five classes: Healthy, Heart Failure (HF), Metabolic Syndrome (MS), Pulmonary Limitation (PulmonaryLim), and Musculoskeletal Limitation (MSK). Performance was evaluated through 5-fold cross-validation using accuracy, precision, recall, and F1-score metrics.

All performance metrics were calculated following the methodology described in [5].

A. Results for SVM-MW3 model

Table II shows the performance of the model using 3-level wavelet decomposition (MW3) and a linear SVM kernel, which achieved strong overall classification performance, with particularly high scores for the Healthy class.

Note: Metrics in tables are presented in decimal format (e.g., 0.9867), while values mentioned in the text use the equivalent percentage format (e.g., 98.67%) for better readability. This convention follows IEEE style recommendations.

TABLE I
Classification performance of the MW3 SVM model using different kernels (5 classes)

SVM Kernel	Precision	Recall	F1-Score	Accuracy
Linear	0.7733	0.7754	0.7727	0.9093
Polynomial	0.6800	0.6926	0.6774	0.8720
RBF	0.3600	0.5026	0.3726	0.7440

The "MEAN" row presents the macro-average across all five diagnostic classes, where each class contributes equally to the average regardless of its sample size.

TABLE II
Performance metrics for linear SVM with MW3 (5 classes)

Class	Precision	Recall	F1-Score	Accuracy	Samples
Heart Failure (HF)	0.8000	0.7059	0.7500	0.8933	15
Metabolic Syndrome (MS)	0.6667	0.7692	0.7143	0.8933	15
Healthy	1.0000	1.0000	1.0000	1.0000	15
Pulmonary Limitation	0.7333	0.6875	0.7097	0.8800	15
Musculoskeletal Limitation	0.6667	0.7143	0.6897	0.8800	15
MEAN	**0.7733**	**0.7754**	**0.7727**	**0.9093**	–

B. Results for SVM-MW5 model

Table III presents the classification performance using the proposed MW5 model with six physiological variables. All three kernels (Linear, Polynomial, and RBF) were tested. The Linear kernel achieved the best overall results. These results were obtained from the implementation developed in this work, which builds upon the methodology introduced in previous research. Although the MW3 model already demonstrated strong performance, the MW5 configuration proposed here provided slight improvements in all metrics, particularly with the Linear kernel.

TABLE III
CLASSIFICATION PERFORMANCE OF THE MW5 SVM MODEL USING DIFFERENT KERNELS (5 CLASSES)

SVM Kernel	Precision	Recall	F1-Score	Accuracy
Linear	0.8400	0.8455	0.8414	0.9360
Polynomial	0.6667	0.7069	0.6721	0.8666
RBF	0.2400	0.2446	0.1977	0.6967

TABLE IV
PERFORMANCE METRICS FOR LINEAR SVM WITH MW5 (5 CLASSES)

Class	Precision	Recall	F1-Score	Accuracy	Samples
Heart Failure (HF)	0.9333	1.0000	0.9655	0.9867	15
Metabolic Syndrome (MS)	1.0000	0.9375	0.9677	0.9867	15
Healthy	0.9333	1.0000	0.9655	0.9867	15
Pulmonary Limitation	0.7333	0.6471	0.6875	0.8667	15
Musculoskeletal Limitation	0.6000	0.6429	0.6207	0.8533	15
MEAN	**0.8400**	**0.8455**	**0.8414**	**0.9360**	–

C. Comparison with Previous Work

Compared to the MW3 results previously reported in [5], which achieved an accuracy of 95.00% and an F1-score of 93.00% in a three-class scenario (Table 7 that paper), our MW3 model—expanded to five diagnostic classes—obtained moderately lower results due to the increased classification complexity (macro-average F1-score of 77.27% and accuracy of 90.93%).

In contrast, the proposed MW5 model developed in this work significantly improved performance, reaching a macro-average accuracy of 93.60% and an F1-score of 84.14% using the Linear SVM kernel. These results suggest that deeper wavelet decomposition, combined with semi-synthetic data grounded in real physiological statistics, enhances both feature extraction and classification effectiveness.

When compared to the MW5 model from [5], which reported an accuracy of 92.00% and an F1-score of 88.00% in a three-class setting (Table 7 that paper). Our enhanced MW5 model (Table III) demonstrates an improvement in accuracy, reinforcing the robustness and scalability of the updated multiclass diagnostic framework.

D. Validation of Semi-Synthetic Signals

To assess the similarity between real and semi-synthetic signals, two well-established metrics were employed: the Pearson correlation coefficient (r) and the Root Mean Square Error (RMSE). This validation was performed using semi-synthetic signals generated from CPET data of real patients in the HF, Healthy, and Metabolic Syndrome groups, using three randomly selected subjects from each group as reference.

The r coefficient quantifies the strength of the linear association between two signals, with values closer to 1 indicating greater similarity. RMSE measures the average deviation between real and generated values, where lower values reflect better alignment. Thus, high r values combined with low RMSE values indicate that the semi-synthetic signals effectively replicate the dynamics of the original physiological data. For reference, RMSE values in this study were generally below 15% of the peak-to-peak amplitude of the corresponding real signals, which is considered an acceptable threshold for physiological signal reconstruction.

Metric definitions. Pearson's r assesses the linear correlation between two signals, with values closer to 1 indicating greater similarity. RMSE (Root Mean Square Error) quantifies the average difference between the real and synthetic signals, with lower values indicating better alignment.

Table V reports the r and RMSE values across six physiological variables and nine patients, confirming the viability of augmenting real datasets with synthetic signals derived from real physiological statistics.

This outcome reinforces the potential of using synthetic signals to expand real physiological datasets in diagnostic models. Despite some inter-patient variability, most synthetic signals demonstrated strong alignment with the original data, preserving physiologically relevant dynamics that support model generalization. In some cases, negative correlation coefficients (e.g., $r < 0$) were observed, particularly in low-amplitude signals such as RER or RR. These values typically reflect the dominance of noise or slight misalignment in phase rather than true inverse relationships. Such occurrences are expected in synthetic modeling when the signal amplitude is small relative to the added variability.This effect was especially noted in the RER and RR signals, which often exhibit low absolute amplitude and high noise sensitivity in both real and synthetic representations.

The evaluation metrics (accuracy, precision, recall, and F1-score) were computed following the methodology presented in [5].

IV. DISCUSSION

The results obtained in this study confirm the feasibility of expanding CPET-based diagnostic models through the integration of semi-synthetic time-series data. The use of wavelet decomposition preserved key physiological signal characteristics, enabling Support Vector Machine classifiers to achieve reliable performance across multiple diagnostic categories. The integration of machine learning methods into the CPET analysis has been increasingly explored in the literature [12], supporting the relevance of the proposed multiclass diagnostic framework.

A critical aspect of this work was the simulation of Pulmonary Limitation (PulmonaryLim) and Musculoskeletal Limitation (MSK) data. By extracting statistical features from real patient datasets and using controlled procedures—such as slope-based segmentation, Gaussian noise addition, and peak enforcement—the resulting semi-synthetic series retained physiologically coherent patterns. This methodology is consistent with previous strategies that apply statistical characteristic modeling to biomedical signal generation [13], and has also

TABLE V
VALIDATION METRICS (r AND RMSE) FOR REAL VS. SEMI-SYNTHETIC SIGNALS ACROSS SIX CPET VARIABLES (NINE PATIENTS)

Patient	HR		VO_2		VCO_2		RR		VE		RER	
	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE
Patient 1	0.83	10.02	0.85	0.20	0.85	0.18	0.33	6.38	0.79	5.06	0.88	0.13
Patient 2	0.72	9.78	0.58	0.51	0.54	0.38	0.19	22.42	0.55	12.85	-0.16	0.08
Patient 3	0.70	9.51	0.36	0.52	0.67	0.39	0.26	10.85	0.60	12.14	0.68	0.10
Patient 4	0.76	18.85	0.86	0.39	0.89	0.30	0.65	7.21	0.88	8.55	0.66	0.23
Patient 5	0.67	12.07	0.70	0.78	0.67	0.54	0.22	20.59	0.61	21.30	0.63	0.09
Patient 6	0.84	8.81	0.72	0.48	0.78	0.34	0.45	10.71	0.77	10.23	0.67	0.17
Patient 7	-0.04	44.82	0.09	0.79	0.06	0.70	-0.37	8.78	-0.07	19.50	0.21	0.20
Patient 8	0.43	20.15	0.29	0.38	0.34	0.37	0.27	10.96	0.37	8.57	-0.32	0.10
Patient 9	0.34	24.34	0.43	0.42	0.44	0.38	0.44	10.00	0.39	10.87	0.18	0.14

proven effective in classification tasks involving artificial physiological signals [14], supporting the use of statistical features in SVM classification.

Compared to previous models restricted to three classes (HF, MS, and Healthy), the addition of PulmonaryLim and MSK expanded the diagnostic scope while maintaining classification performance.

The fixed-length structure of the synthetic signals (200 data points) may partially explain the performance gap observed between real and synthetic classes, especially for PulmonaryLim and MSK. While this design enables standardized wavelet feature extraction and model training, it does not capture natural variations in CPET session duration across patients. This trade-off should be considered when interpreting results and in future improvements of temporal modeling. Several studies highlight that modeling temporal variability plays a key role in improving the realism and diagnostic utility of synthetic physiological signals [15].

An important limitation of this work lies in the fidelity of the semi-synthetic signals for the RR and RER variables, which showed weaker correlation with real data. This discrepancy is attributable not only to the higher noise sensitivity of these variables, but also to the fact that our generative model, based on linear segments, is too simplistic to capture their complex non-linear dynamics. This lower fidelity likely restricted the usefulness of these features in distinguishing the classes with synthetic data (PulmonaryLim and MSK), which may partially justify the slightly lower classification performance for these groups, as observed in Table IV. The development of more advanced generative models, capable of better simulating these specific dynamics, therefore constitutes a crucial direction for future work.

In a clinical workflow, the proposed model could serve as a valuable decision support tool. For example, for patients presenting with non-specific exercise intolerance, the algorithm's output could help guide physicians toward a more targeted diagnostic path by highlighting similarities with established physiological limitation profiles. However, the transition to a real-world environment presents significant challenges. The performance of our model was validated on a curated and balanced dataset; its robustness against data heterogeneity—arising from different CPET equipment manufacturers, testing protocols, and complex patient populations with multiple comorbidities—still needs to be determined. Additionally, it is important to note that performance in such specific datasets may exhibit sensitivity to the randomness inherent in the data partitioning during cross-validation, which reinforces the importance of validation on larger samples to confirm model stability. Therefore, extensive validation on large multicenter datasets and, ultimately, in prospective clinical trials, is an essential step before clinical adoption can be considered.

Future research should explore additional wavelet families (e.g., Symlets, Coiflets), automated hyperparameter tuning, and deep learning architectures to further improve performance and capture more complex temporal dependencies in CPET data. Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have already shown promise in physiological time series classification tasks [12], and could be valuable alternatives for future implementations.

An important methodological detail was the reconstruction of temporal dynamics from limited statistical information. The approach of segmenting time-series into quartiles and estimating signal slopes enabled the generation of structured synthetic signals even in the absence of original raw data. This process simulates how physiological responses evolve throughout the CPET protocol, ensuring that synthetic samples are not only statistically plausible but also temporally coherent.

The validation using real patient signals—where both synthetic and real series were available—demonstrated high correlation and low error, supporting the generalization of this methodology. This strategy can be especially valuable in clinical datasets where full temporal data is not consistently recorded or shared due to privacy constraints.

V. CONCLUSION

This study demonstrated the feasibility of developing robust diagnostic support algorithms based on CPET data using machine learning and wavelet decomposition. By combining real and semi-synthetic time-series signals, the proposed approach successfully classified five physiologically distinct diagnostic groups—Healthy, HF, MS, PulmonaryLim, and MSK—with high accuracy and generalization.

Among the evaluated configurations, the model based on Support Vector Machine with five-level wavelet decomposition (SVM-LIN-MW5) achieved the best overall performance, with a macro-average accuracy of 93.60% and an F1-score of 84.14%, outperforming the MW3 model in all metrics. These results represent an improvement over MW3, demonstrating that deeper decomposition levels enhance feature extraction and classification performance.

Future work may focus on expanding this framework with multicenter datasets, integrating temporal alignment strategies, and validating the approach in clinical settings. Additional

enhancements may include testing alternative wavelet families, optimizing hyperparameters, and applying deep learning architectures to explore more complex temporal patterns.

Additionally, a crucial next step will be to systematically optimize the different SVM kernels to identify their best configuration and, subsequently, to perform a benchmarking of this optimized SVM model against a broader range of other classification algorithms.

These findings are consistent with prior studies supporting wavelet-based machine learning approaches in biomedical time-series analysis [11], [14].

ACKNOWLEDGMENT

This work was funded by Portuguese national funds provided by the Portuguese Foundation for Science and Technology (FCT) (FCT-UIDB/05704/2020 and UIDP/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.

REFERENCES

[1] K. Wasserman, J. E. Hansen, and D. Y. Sue, "Principles of exercise testing and interpretation: including pathophysiology and clinical applications," 5th ed., Philadelphia, PA, USA: Wolters Kluwer Health, 2011.

[2] J. Portella et al., "Processed CPET dataset," *IEEE Dataport*, Jan. 26, 2022. DOI: 10.21227/m4h8-z187.

[3] R. F. Pinheiro and R. Fonseca-Pinto, "Algorithm for diagnosis of metabolic syndrome and heart failure using CPET biosignals via SVM and wavelet transforms," in *Int. Conf. on Smart Objects and Technologies for Social Good*, Springer, 2023, pp. 159–171.

[4] D. E. Brown et al., "Neural network methods for diagnosing patient conditions from cardiopulmonary exercise testing data," *BioData Mining*, vol. 15, no. 16, 2022.

[5] R. F. Pinheiro and R. Fonseca-Pinto, "On the development of diagnostic support algorithms based on CPET biosignals data via machine learning and wavelets," *PeerJ Comput. Sci.*, vol. 11, e2474, 2025.

[6] V. C. Pezoulas, D. I. Zaridis, E. Mylona, C. Androutsos, K. Apostolidis, N. S. Tachos, and D. I. Fotiadis, "Synthetic data generation methods in healthcare: a review on open-source tools and methods," *Comput. Struct. Biotechnol. J.*, vol. 22, pp. 2024–2042, 2024. DOI: 10.1016/j.csbj.2024.02.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2001037024002393.

[7] R. Polikar, "The wavelet tutorial," Rowan University, 1996. [Online]. Available: http://users.rowan.edu/ polikar/WAVELETS/WTtutorial.html.

[8] L. Sharma and R. K. Sunkaria, "A novel feature extraction approach for ECG classification using wavelet transform and statistical features," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 579–589, 2013. DOI: 10.1016/j.bspc.2013.05.004.

[9] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, "Wavelet transform-based analysis of QRS complex in ECG signals," arXiv preprint, arXiv:1311.6460, 2013. [Online]. Available: https://arxiv.org/abs/1311.6460.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[11] MathWorks, "Wavelet toolbox user's guide," *MATLAB R2020a Documentation*, The MathWorks, Inc., Natick, MA, USA, 2020.

[12] N. Al Rahhal et al., "Deep learning approach for time-series classification of physiological signals," *Sensors*, vol. 16, no. 1, pp. 1–15, 2016. DOI: 10.3390/s16010001.

[13] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, 2003. DOI: 10.1109/TBME.2003.808805.

[14] T. Azizi, "Comparative analysis of statistical, time–frequency, and SVM techniques for change detection in nonlinear biomedical signals," *Signals*, vol. 5, no. 4, pp. 736–755, 2024. DOI: 10.3390/signals5040041. [Online]. Available: https://www.mdpi.com/2673-5419/5/4/41.

[15] G. D. Clifford, C. Liu, B. Moody, et al., "Synthetic ECG generation and temporal modeling of arrhythmias for improved training of deep neural networks," *Physiol. Meas.*, vol. 41, no. 11, pp. 115002, 2020. DOI: 10.1088/1361-6579/abb9f8.

Página intencionalmente deixada em branco.