



CATÓLICA

ESCOLA SUPERIOR DE BIOTECNOLOGIA

PORTO

Deteção de Insuficiência Cardíaca em Pacientes com Doenças de Chagas a partir de
Algoritmos de Extração Automática de Parâmetros e de Aprendizagem de Máquina em
Sinais ECG

Por

Afonso Manuel Silveira de Sousa Coutinho França

Maio de 2025



CATÓLICA

ESCOLA SUPERIOR DE BIOTECNOLOGIA

PORTO

Deteção de Insuficiência Cardíaca em Pacientes com Doenças de Chagas a partir de
Algoritmos de Extração Automática de Parâmetros e de Aprendizagem de Máquina em
Sinais ECG

Tese apresentada à Escola Superior de Biotecnologia da Universidade Católica
Portuguesa para obtenção do grau de Mestre em Engenharia Biomédica

Por

Afonso Manuel Silveira de Sousa Coutinho França

Orientação: Pedro Miguel Rodrigues (Orientador)

João Paulo Madeiro (Co-orientador)

Maio de 2025

Resumo

A Doença de Chagas é uma infecção parasitária crônica causada pelo *Trypanosoma cruzi*, sendo uma das principais causas de insuficiência cardíaca na América Latina. A detecção precoce da cardiomiopatia chagásica pode permitir uma intervenção mais eficaz, reduzindo complicações graves. Neste estudo, propõe-se uma nova abordagem na utilização de algoritmos de *Machine Learning* (ML) para a detecção automática de insuficiência cardíaca em pacientes com Doença de Chagas, utilizando características espectrais extraídas do eletrocardiograma (ECG). A nova abordagem recorre ao uso de características espectrais, como a **SEF95 (*Spectral Edge Frequency*)**, **MF (*Mean Frequency*)**, **SE (*Spectral Entropy*)**, potências e raios extraídos do Espectro de potências do sinal ECG. Posteriormente, classificadores de *Machine Learning* foram aplicados para prever o grau de insuficiência cardíaca. Os resultados indicaram que os modelos conseguem distinguir bem entre estágios extremos da doença (Exatidão: 83,3%; AUC-ROC (Área sob a Curva ROC): 0.80-0.88), mas enfrentam dificuldades na estratificação da insuficiência cardíaca moderada, devido à heterogeneidade clínica da doença.

Palavras-chave: Doença de Chagas, Insuficiência Cardíaca, *Machine Learning*, Eletrocardiograma, Análise Espectral.

Abstract

Chagas disease is a chronic parasitic infection caused by *Trypanosoma cruzi* and is one of the leading causes of heart failure in Latin America. Early detection of Chagasic cardiomyopathy can enable more effective intervention, reducing the risk of severe complications. This study proposes a novel approach using Machine Learning algorithms for the automatic detection of heart failure in patients with Chagas disease, based on spectral features extracted from the electrocardiogram (ECG). The proposed method focuses on the use of spectral features such as **SEF95 (Spectral Edge Frequency)**, **MF (Mean Frequency)**, **SE (Spectral Entropy)**, power values, and ratios. These features were then used as input to various Machine Learning classifiers to predict the degree of heart failure. The results showed that the models performed well in distinguishing between extreme stages of the disease (Accuracy: 83.3%; AUC-ROC: 0.80–0.88), but had difficulties in correctly identifying moderate heart failure, mainly due to the clinical heterogeneity of the condition.

Keywords: Chagas disease, ECG, spectral analysis, Heart failure detection, Machine learning.

ÍNDICE

| | |
|--|--------|
| 1. Introdução | - 4 - |
| 1.1 Motivação (Relevância) | - 4 - |
| 1.2 Objetivos..... | - 5 - |
| 2. Contextualização da Doença de Chagas e as Suas Complicações Cardíacas | - 6 - |
| 2.1 Epidemiologia e Modos de Transmissão | - 6 - |
| 2.2 Fases Clínicas da Doença de Chagas | - 7 - |
| 2.3 Insuficiência Cardíaca Chagásica e suas Manifestações | - 9 - |
| 2.4 Métodos Tradicionais de Diagnóstico | - 10 - |
| 3. Processamento de Sinal Eletrocardiográfico (ECG) e Extração de Características..... | - 11 - |
| Relevantes no MATLAB | - 11 - |
| 3.1 Conceitos Fundamentais do Eletrocardiograma (ECG) | - 11 - |
| 3.2 Aquisição e Pré-Processamento do Sinal ECG..... | - 12 - |
| 3.3 Extração de Características Relevantes no MATLAB | - 13 - |
| 4. <i>Machine Learning</i> na Análise de Sinais Eletrocardiográficos (ECG)..... | - 16 - |
| 4.1 Introdução ao <i>Machine Learning</i> | - 16 - |
| 4.2 Aplicações de <i>Machine Learning</i> na Análise de Sinais ECG..... | - 17 - |
| 4.3 Algoritmos de ML Utilizados no Estudo | - 18 - |
| 4.4 Utilização e Avaliação de Modelos de <i>Machine Learning</i> | - 21 - |
| 5. Materiais e Métodos | - 23 - |
| 5.2 Base de Dados | - 23 - |
| 5.2.1 Características da Base de Dados | - 24 - |
| 5.3 Pré-Processamento dos Sinais ECG no MATLAB..... | - 25 - |
| 5.4 Utilização dos Modelos de <i>Machine Learning</i> no Python..... | - 27 - |
| 5.5 Análise Estatística dos Resultados..... | - 28 - |
| 6. Resultados e Discussão | - 30 - |
| 6.1 Introdução..... | - 30 - |
| 6.2 Resultados | - 30 - |
| 6.3 Discussão..... | - 32 - |
| 7. Conclusões Gerais..... | - 34 - |
| 8. Trabalho Futuro | - 37 - |
| 9. Anexo | - 39 - |
| 10. Bibliografia | - 44 - |

1. Introdução

A Doença de Chagas, também conhecida como tripanossomíase americana, é uma infecção parasitária crônica causada pelo *Trypanosoma cruzi*, sendo um dos principais problemas de saúde pública na América Latina. Estima-se que entre 6 a 7 milhões de pessoas estejam infetadas globalmente, com uma crescente disseminação da doença para regiões não endêmicas, como a América do Norte e a Europa, devido aos fluxos migratórios (WHO, 2021). Entre as complicações da doença, a cardiomiopatia chagásica crônica destaca-se como a principal causa de insuficiência cardíaca nestes pacientes, estando associada a arritmias, tromboembolismo e morte súbita (Rassi *et al.*, 2010).

A detecção precoce da insuficiência cardíaca é essencial para melhorar o prognóstico dos pacientes, permitindo intervenções médicas antes da progressão da doença. O ECG é uma das principais ferramentas diagnósticas utilizadas para identificar alterações cardíacas associadas à Doença de Chagas. No entanto, a interpretação manual do ECG pode ser subjetiva e limitada, tornando fundamental a utilização de métodos computacionais para melhorar a precisão diagnóstica.

O avanço das técnicas de *Machine Learning* tem permitido o desenvolvimento de modelos preditivos que analisam sinais de ECG de forma automatizada, identificando padrões característicos da insuficiência cardíaca com elevada precisão. Neste contexto, este trabalho de mestrado propõe a utilização de algoritmos de *Machine Learning* para a detecção automática de insuficiência cardíaca em pacientes com Doença de Chagas, utilizando características espectrais extraídas do ECG.

1.1 Motivação (Relevância)

A insuficiência cardíaca associada à Doença de Chagas continua a ser um desafio diagnóstico, principalmente nos estágios moderados da doença. Métodos convencionais baseados em ECG apresentam limitações na identificação precoce da progressão da doença, levando a atrasos no tratamento. Além disso, como motivação, ressalta-se a dificuldade de acesso de parte da população a centros hospitalares com opção de realização de exame de imagem (como o ecocardiograma), sendo o ECG um método mais simples e econômico.

A aplicação de técnicas de processamento de sinais biomédicos com recurso a algoritmos de *Machine Learning* pode melhorar significativamente a deteção da insuficiência cardíaca, possibilitando um diagnóstico mais rápido e preciso. Além disso, o desenvolvimento de um modelo preditivo eficaz para pacientes com Doença de Chagas pode ter um impacto clínico relevante, auxiliando na estratificação de risco e permitindo intervenções personalizadas para cada paciente.

1.2 Objetivos

O objetivo principal deste trabalho de mestrado é desenvolver um sistema baseado em *Machine Learning* para a classificação automática do grau de insuficiência cardíaca em pacientes com Doença de Chagas, utilizando-se sinais ECG. Para alcançar esse objetivo, foram estabelecidos os seguintes objetivos específicos:

1. Extrair características espectrais dos sinais ECG utilizando-se métodos de análise espectral, como a Transformada de Fourier e a Densidade Espectral de Potência (PSD).
2. Utilização de algoritmos de *Machine Learning* para diferenciar os pacientes em diferentes graus de insuficiência cardíaca.
3. Avaliar a precisão dos modelos utilizando-se métricas de desempenho, como AUC-ROC (Área sob a Curva ROC), exatidão e sensibilidade.
4. Por fim, responder à seguinte pergunta: É possível distinguir diferentes níveis de insuficiência cardíaca associados à Doença de Chagas a partir do processamento automático do sinal ECG?

2. Contextualização da Doença de Chagas e as Suas Complicações Cardíacas

Este capítulo apresenta um enquadramento clínico, epidemiológico e diagnóstico da Doença de Chagas, uma infecção parasitária crónica causada pelo *Trypanosoma cruzi*. Inicialmente restrita à América Latina, a doença expandiu-se para outras regiões do mundo, impulsionada pelos fluxos migratórios. A transmissão ocorre, principalmente, por via vetorial (através do contato com as fezes do inseto triatomíneo), mas também pode dar-se por transfusões sanguíneas, transplantes de órgãos, transmissão congénita e ingestão de alimentos contaminados.

Clinicamente, a doença evolui em três fases: aguda, indeterminada e crónica. A cardiomiopatia chagásica, presente na fase crónica em cerca de 30% dos pacientes, constitui a principal complicação, estando associada a insuficiência cardíaca, arritmias, eventos tromboembólicos e morte súbita. A literatura descreve as manifestações cardíacas típicas da doença e a sua fisiopatologia, marcada pela inflamação miocárdica e fibrose progressiva.

O diagnóstico tradicional combina testes parasitológicos, sorológicos e exames cardíacos. O ECG tem um papel fundamental na deteção precoce de alterações cardíacas, sendo complementado por ecocardiografia e ressonância magnética cardíaca para avaliação estrutural. Por fim, o capítulo introduz a motivação para o uso de abordagens computacionais, como a extração automática de parâmetros do ECG e a aplicação de algoritmos de aprendizagem de máquina, com o objetivo de melhorar a precisão e precocidade do diagnóstico.

2.1 Epidemiologia e Modos de Transmissão

A Doença de Chagas, também denominada tripanossomíase americana, é uma infecção parasitária crónica causada pelo protozoário *Trypanosoma cruzi*. Descoberta em 1909 pelo médico brasileiro Carlos Chagas, essa doença tornou-se um dos maiores desafios de saúde pública na América Latina, onde a sua incidência é mais elevada. Estima-se que entre 6 a 7 milhões de pessoas estejam infetadas em todo o mundo, sendo que, devido aos fluxos migratórios, a sua distribuição geográfica expandiu-se para regiões não endémicas, como a América do Norte e a Europa (WHO, 2021).

A Doença de Chagas está fortemente associada a condições socioeconômicas desfavoráveis, sendo mais prevalente em áreas rurais da América Latina, onde o inseto vetor prolifera em habitações precárias. No entanto, a urbanização e a migração populacional levaram à disseminação global da doença, desafiando os sistemas de saúde de regiões não endêmicas, como os Estados Unidos e diversos países europeus (WHO, 2021).

A transmissão da doença ocorre predominantemente por meio de um inseto vetor, pertencente à subfamília *Triatominae*, popularmente conhecido como “barbeiro”. Durante a picada, esse inseto hematófago deposita as fezes na pele do hospedeiro, permitindo que o *T. cruzi* entre no organismo através de pequenas feridas ou mucosas. No entanto, outras vias de transmissão também são possíveis, incluindo transfusões de sangue, transplantes de órgãos, transmissão congênita e ingestão de alimentos contaminados (Rassi *et al.*, 2010; Nunes *et al.*, 2018).

A doença apresenta uma evolução dividida em três fases clínicas. A fase aguda, que pode durar entre quatro e oito semanas, pode ser assintomática ou manifestar sintomas inespecíficos, como febre, mal-estar e linfadenopatia. Nessa fase, alguns pacientes podem apresentar sinais típicos, como o sinal de Romaña, caracterizado por edema palpebral unilateral, ou o chagoma de inoculação, uma lesão cutânea localizada na região da picada do inseto (Dias *et al.*, 2016). Em seguida, a doença pode progredir para uma fase indeterminada, na qual cerca de 70% dos indivíduos infetados permanecem assintomáticos por décadas. No entanto, aproximadamente 30% dos casos evoluem para a fase cardíaca crônica, caracterizada pelo desenvolvimento de cardiomiopatia chagásica crônica, que pode resultar em insuficiência cardíaca, arritmias e até morte súbita cardíaca (Rassi *et al.*, 2010; Nunes *et al.*, 2018).

2.2 Fases Clínicas da Doença de Chagas

A transmissão do *T. cruzi* pode ocorrer de diferentes formas (Tabela 2.1). O modo vetorial continua a ser a principal via, ocorrendo quando o *Triatoma infestans* defeca na pele do hospedeiro após a picada, permitindo que o parasita penetre no organismo. Além disso, a transmissão por transfusão de sangue representa um risco significativo em regiões onde os bancos de sangue não realizam rastreios adequados. A transmissão congênita pode ocorrer durante a gravidez, quando o parasita atravessa a placenta e infeta o feto. Outra forma relevante de transmissão é a oral, que ocorre pela ingestão de alimentos contaminados com fezes de

“barbeiros” infetados, sendo esta via responsável por surtos documentados no Brasil e na Venezuela (Rassi *et al.*, 2010).

Tabela 2.1 – Modos de Transmissão da Doença de Chagas

| Modo de Transmissão | Descrição |
|----------------------------|--|
| Vetorial | Ocorre quando o <i>Triatoma infestans</i> defeca na pele do hospedeiro após a picada, permitindo que o <i>T. cruzi</i> entre no organismo. |
| Transfusão de sangue | Risco presente em bancos de sangue sem rastreio adequado. |
| Transmissão congênita | O parasita pode atravessar a placenta e infetar o feto durante a gestação. |
| Transplante de órgãos | Transmissão ocorre quando órgãos contaminados são transplantados para um recetor saudável. |
| Via oral | Ingestão de alimentos contaminados com fezes de “barbeiros” infetados. |

Fonte: Adaptado de WHO (2021).

A evolução da Doença de Chagas é dividida em três fases clínicas distintas: aguda, indeterminada e crônica. A fase aguda ocorre logo após a infecção e pode apresentar sintomas inespecíficos, como febre, fadiga, dor muscular e hepatoesplenomegalia. Os sinais mais característicos dessa fase incluem o sinal de Romaña e lesões cutâneas inflamatórias na zona de inoculação (Chagoma) (Dias *et al.*, 2016). Se não for tratada nessa fase, a doença entra na fase indeterminada, na qual o paciente permanece assintomático durante anos ou até décadas, sem apresentar sinais clínicos evidentes da infecção.

Cerca de 30% dos pacientes infetados evoluem para a fase cardíaca crônica, desenvolvendo cardiomiopatia chagásica, que pode levar a insuficiência cardíaca, arritmias cardíacas, tromboembolismo e morte súbita (Rassi *et al.*, 2010).

2.3 Insuficiência Cardíaca Chagásica e suas Manifestações

A cardiomiopatia chagásica é a manifestação mais grave da Doença de Chagas e está associada a processos inflamatórios persistentes que resultam na destruição progressiva do miocárdio. Com o tempo, ocorre o desenvolvimento de fibrose, levando à perda da função contrátil do coração e consequente disfunção ventricular.

Entre as principais manifestações clínicas da insuficiência cardíaca chagásica (Tabela 2.2), destacam-se as arritmias ventriculares, que ocorrem devido a alterações na condução elétrica do coração; os bloqueios auriculoventriculares, que impedem a transmissão do impulso elétrico entre os aurículos e os ventrículos; a cardiomegalia, caracterizada pela dilatação do coração devido à disfunção ventricular progressiva; os eventos tromboembólicos, nos quais a formação de coágulos pode resultar em acidente vascular cerebral (AVC) ou embolia pulmonar; e a morte súbita cardíaca, frequentemente associada a arritmias ventriculares malignas (Nunes *et al.*, 2018).

Tabela 2.2 – Principais Manifestações Cardíacas na Doença de Chagas

| Manifestação | Descrição |
|---------------------------------|---|
| Arritmias | Batimentos cardíacos irregulares devido a alterações na condução elétrica do coração. |
| Bloqueios auriculoventriculares | Interrupção na passagem do impulso elétrico entre os aurículos e os ventrículos. |
| Cardiomegalia | Dilatação do coração devido à disfunção ventricular progressiva. |
| Eventos tromboembólicos | Formação de coágulos que podem levar a acidente vascular cerebral (AVC) e embolia pulmonar. |
| Morte súbita cardíaca | Principalmente associada a arritmias ventriculares malignas. |

Fonte: Adaptado de Nunes et al. (2018).

2.4 Métodos Tradicionais de Diagnóstico

O diagnóstico da Doença de Chagas é baseado em exames laboratoriais e testes cardíacos, sendo que a escolha do método diagnóstico depende da fase da infecção. Durante a fase aguda, o parasita pode ser identificado diretamente no sangue por meio de exames parasitológicos, como o esfregaço em gota espessa e o método de Strout. Já na fase crônica, o diagnóstico é realizado principalmente por testes sorológicos, que identificam anticorpos específicos contra o *T. cruzi* (Dias *et al.*, 2016).

O eletrocardiograma (ECG) desempenha um papel essencial na detecção precoce de alterações cardíacas associadas à cardiomiopatia chagásica. O exame permite identificar bloqueios de condução, extrassístoles ventriculares e alterações do segmento ST, que podem indicar o início da disfunção miocárdica. Além disso, exames de imagem, como ecocardiografia e ressonância magnética cardíaca, são fundamentais para avaliar a dilatação ventricular, a presença de trombos intracardíacos e a fibrose miocárdica (Nunes *et al.*, 2018).

Nos últimos anos, o avanço das técnicas computacionais tem permitido melhorias na análise do ECG, possibilitando a extração automática de parâmetros cardíacos e a aplicação de *Machine Learning* para a detecção precoce da insuficiência cardíaca. Estes avanços tecnológicos são explorados nos próximos capítulos desta tese.

3. Processamento de Sinal Eletrocardiográfico (ECG) e Extração de Características Relevantes no MATLAB

Este capítulo aborda o processamento de sinais eletrocardiográficos (ECG) com ênfase na extração de características espectrais no domínio da frequência, tendo como base a análise computacional de sinais biomédicos em MATLAB. Inicialmente, são apresentados os fundamentos fisiológicos do ECG, incluindo a descrição das ondas e intervalos característicos, bem como a sua relevância clínica. Em seguida, detalham-se as etapas de pré-processamento do sinal, que visam a remoção de ruídos e artefactos por meio de filtros digitais e técnicas como a Análise de Componentes Principais (PCA).

A principal contribuição metodológica do capítulo reside na utilização da Densidade Espectral de Potência (PSD), calculado através do método de Welch, como base para a extração de métricas espectrais. Entre estas, destacam-se o SEF95 (frequência abaixo da qual se concentra 95% da potência total), a frequência média (MF), a entropia espectral (SE), bem como as potências em diferentes bandas e rácios entre elas. Estas características foram utilizadas para caracterizar a atividade elétrica do coração.

3.1 Conceitos Fundamentais do Eletrocardiograma (ECG)

O sinal de ECG é composto por diversas ondas e intervalos, cada um correspondendo a um evento fisiológico específico do ciclo cardíaco (Tabela 3.1). A onda P representa a despolarização auricular, sendo fundamental para a análise da condução elétrica nos aurículos. O intervalo PR mede o tempo entre a ativação dos aurículos e dos ventrículos, e seu prolongamento pode indicar bloqueios auriculoventriculares. O complexo QRS representa a despolarização ventricular e está diretamente associado à contração dos ventrículos, sendo um dos componentes mais importantes para o diagnóstico de doenças cardíacas. A onda T, por sua vez, indica a repolarização ventricular, um evento essencial para a recuperação do coração antes do próximo ciclo. O intervalo QT mede o tempo total de despolarização e repolarização ventricular, sendo um parâmetro crucial para a detecção de arritmias. O segmento ST representa o período entre a despolarização e a repolarização ventricular, e sua elevação ou depressão pode indicar isquemia miocárdica (Guyton & Hall, 2020).

Tabela 3.1 – Componentes do Sinal ECG e Seu Significado Fisiológico

| Componente | Descrição | Relevância Clínica |
|-------------------|--|---|
| Onda P | Representa a despolarização auricular, ou seja, a ativação dos aurículos. | Alterações podem indicar bloqueios auriculares ou fibrilação auricular. |
| Intervalo PR | Mede o tempo entre a ativação dos aurículos e a ativação dos ventrículos. | Prolongamento pode indicar bloqueios auriculoventriculares. |
| Complexo QRS | Representa a despolarização ventricular, sinalizando a contração dos ventrículos. | Alargamento pode estar associado a bloqueios de ramo ou cardiomiopatia chagásica. |
| Onda T | Corresponde à repolarização ventricular, ou seja, a recuperação do coração para um novo ciclo. | Alterações podem indicar isquemia miocárdica. |
| Intervalo QT | Mede o tempo total da despolarização e repolarização ventricular. | Prolongamento pode predispor a arritmias ventriculares. |
| Segmento ST | Representa o período entre a despolarização e repolarização dos ventrículos. | Elevação ou depressão são sinais de enfarte do miocárdio. |

Fonte: Adaptado de Guyton & Hall (2020).

Pacientes com Doença de Chagas frequentemente apresentam anormalidades eletrocardiográficas, como bloqueios de condução, fibrilação auricular, extrassístoles ventriculares, bradicardias, ritmo irregular e disfunção do nó sinusal. A detecção precoce dessas alterações é essencial para prevenir a progressão da insuficiência cardíaca e melhorar o prognóstico dos pacientes.

3.2 Aquisição e Pré-Processamento do Sinal ECG

A análise computacional do ECG inicia-se com a aquisição do sinal através de elétrodos conectados ao paciente. No entanto, o sinal bruto contém ruídos e interferências que podem prejudicar sua interpretação (Tabela 3.2). As principais fontes de ruído incluem a interferência

da rede elétrica (50/60 Hz), causada pela proximidade de equipamentos eletrônicos e que pode ser removida por meio de filtros Notch; o ruído mioelétrico, produzido pela atividade elétrica dos músculos esqueléticos e que pode ser minimizado com filtros passa-baixa; os movimentos do paciente, que geram variações bruscas no sinal e que podem ser corrigidos com técnicas de análise de componentes principais (PCA); e a deriva da linha de base, resultante de oscilações lentas causadas pela respiração e variações na impedância da pele, sendo corrigida com filtros passa-alta (Sörnmo & Laguna, 2005).

Tabela 3.2 – Principais Fontes de Ruído no ECG e Métodos de Remoção

| Fonte de Ruído | Descrição | Método de Remoção |
|---|--|---|
| Interferência da rede elétrica (50/60 Hz) | Ocorre devido à proximidade de equipamentos eletrônicos. | Utilização de filtros Notch para remover essa frequência específica. |
| Ruído mioelétrico | Produzido pela atividade elétrica dos músculos esqueléticos próximos aos eletrodos. | Aplicação de filtros passa-baixa para remover frequências elevadas. |
| Movimentos do paciente | O deslocamento dos eletrodos pode gerar variações bruscas no sinal. | Uso de técnicas de correção automática, como PCA. |
| Deriva de linha de base | Pequenas oscilações lentas causadas pela respiração e variações na impedância da pele. | Aplicação de filtros passa-alta para eliminar flutuações de baixa frequência. |

Fonte: Adaptado de Sörnmo & Laguna (2005).

Após a remoção do ruído, o sinal ECG é segmentado em janelas menores para a extração de características relevantes.

3.3 Extração de Características Relevantes no MATLAB

O foco principal deste estudo foi a análise do ECG no domínio da frequência, através da Densidade Espectral de Potência (*Power Spectral Density* – PSD). Esta análise permite

compreender como a energia do sinal está distribuída nas várias frequências, sendo particularmente útil para detetar padrões associados à insuficiência cardíaca (Celler & Chazal, 1997).

A PSD representa a distribuição da potência de um sinal em função da frequência. Em termos simples, mostra quais frequências estão mais presentes e com que intensidade num determinado sinal. A sua estimativa é geralmente feita com base na Transformada de Fourier (FFT), que transforma o sinal do domínio do tempo para o da frequência.

$$PSD(f) = \lim_{T \rightarrow \infty} \frac{1}{T} |x_t(f)|^2 \quad (1)$$

Em que (f) é a transformada de Fourier de uma janela temporal do sinal e $x_t(f)$ é a Transformada de Fourier desse sinal limitado a um intervalo de tempo (t) .

No MATLAB, essa estimativa foi realizada com o método de Welch, que aplica janelas com sobreposição para melhorar a estabilidade da análise espectral. A partir da PSD, foram extraídas as seguintes características:

SEF95: frequência abaixo da qual 95% da potência total está concentrada.

MF: representa a frequência média ponderada do espectro, calculada como:

$$MF = \frac{\sum f_i \times PSD(f_i)}{\sum PSD(f_i)} \quad (2)$$

SE: mede a complexidade e imprevisibilidade do espectro. É baseada no conceito de entropia de Shannon e indica o grau de desorganização do sinal:

$$.SE = - \sum P(f_i) \log_2 P(f_i) \quad (3)$$

Em que $P(f_i)$ representa a probabilidade associada à frequência (f) .

As potências espectrais, que representam os valores absolutos de energia em faixas de frequência específicas do sinal ECG, permitem identificar alterações na distribuição da

atividade elétrica cardíaca associadas a diversas condições patológicas. Já os raios espectrais, definidos como a razão entre potências de diferentes faixas de frequência, constituem métricas complementares que podem refletir desequilíbrios e alterações na condução elétrica do coração, sendo úteis na análise não invasiva da função cardíaca (Celler & de Chazal, 1997; Sörnmo & Laguna, 2005). Estas métricas fornecem uma visão detalhada da estrutura de frequência do ECG, permitindo a detecção de alterações que não são facilmente visíveis no domínio do tempo.

4. *Machine Learning* na Análise de Sinais Eletrocardiográficos (ECG)

Este capítulo apresenta uma revisão detalhada sobre a aplicação de algoritmos de *Machine Learning* (ML) na análise de sinais ECG. Inicialmente, são discutidas as principais aplicações do ML em contexto clínico, nomeadamente a deteção de arritmias, a classificação de doenças cardíacas e a previsão da progressão da disfunção ventricular.

Adicionalmente, descrevem-se os passos fundamentais para a utilização de sistemas baseados em ML, desde a aquisição e pré-processamento dos sinais, passando pela extração de características, até à avaliação dos modelos com base em métricas robustas como exatidão, precisão, sensibilidade, especificidade, F1-score e AUC-ROC. A importância destas métricas é discutida, com ênfase na sua relevância clínica, especialmente em contextos com classes desbalanceadas. Assim, este capítulo fornece o enquadramento teórico necessário para a compreensão da abordagem computacional adotada no presente estudo, justificando a sua aplicabilidade na estratificação automática de pacientes com Doença de Chagas.

4.1 Introdução ao *Machine Learning*

Nos últimos anos, a aplicação de ML na área da saúde tem transformado significativamente a forma como doenças são diagnosticadas e tratadas. O aumento da capacidade de processamento computacional, aliado à crescente disponibilidade de grandes bases de dados biomédicos, permitiu o desenvolvimento de modelos preditivos capazes de identificar padrões complexos em sinais fisiológicos, como o ECG.

A Doença de Chagas, devido ao seu curso clínico prolongado e às suas potenciais complicações cardíacas, representa um grande desafio para a medicina. A deteção precoce de alterações no ECG associadas à cardiomiopatia chagásica pode permitir uma intervenção mais rápida e eficaz, reduzindo o risco de insuficiência cardíaca e morte súbita. O uso de ML nesse contexto desempenha um papel essencial, pois permite a análise automatizada dos sinais ECG, identificando padrões anómalos com alta precisão e rapidez. Modelos preditivos avançados podem ser treinados para reconhecer assinaturas eletrofisiológicas características da doença, facilitando a estratificação dos pacientes de acordo com o grau de insuficiência cardíaca.

Dada a importância dessa abordagem, este capítulo aborda as principais aplicações do ML na análise do ECG, os algoritmos mais utilizados e os desafios enfrentados na sua implementação clínica.

4.2 Aplicações de *Machine Learning* na Análise de Sinais ECG

Os algoritmos de ML podem ser aplicados ao processamento do ECG para diferentes finalidades clínicas. Uma das aplicações mais relevantes é a detecção automática de arritmias, onde modelos de *Deep Learning*, especialmente redes neurais convolucionais (CNNs), conseguem identificar automaticamente padrões no sinal e classificá-los em diferentes tipos de arritmia, reduzindo o tempo de diagnóstico e aumentando a precisão da análise (Rajpurkar *et al.*, 2017).

Além disso, o ML tem sido amplamente utilizado na classificação de doenças cardíacas, permitindo diferenciar ECGs normais de ECGs indicativos de patologias, incluindo a cardiomiopatia chagásica. Algoritmos como *Support Vector Machines* (SVMs) e Redes Neurais Artificiais (ANNs) são frequentemente aplicados para esse fim, possibilitando diagnósticos mais rápidos e eficientes (Hannun *et al.*, 2019).

Um avanço significativo proporcionado pelo ML é a capacidade de prever a insuficiência cardíaca com base na análise de sinais de ECG. A avaliação de parâmetros eletrofisiológicos ao longo do tempo permite identificar pacientes com maior risco de evolução para disfunção ventricular, especialmente no contexto da Doença de Chagas. Algoritmos como *Gradient Boosting* têm demonstrado elevada eficácia na detecção automática de padrões ocultos em sinais ECG, contribuindo para modelos preditivos robustos. Por exemplo, Huang *et al.* (2023) relataram que modelos baseados em *XGBoost* alcançaram precisão superior a 99% na detecção de insuficiência cardíaca, utilizando abordagens híbridas que combinam transformadas *wavelet* e redes neurais convolucionais. Além disso, González *et al.* (2022) demonstraram que análises de ECG de curta duração (30 segundos) podem ser eficazes na estimativa do risco de hospitalização por insuficiência cardíaca, com modelos de *Gradient Boosting* atingindo AUCs superiores a 0,85, evidenciando o potencial do ML na estratificação de risco e no suporte à decisão clínica.

Adicionalmente, técnicas de ML também têm sido aplicadas para a redução de ruído e aprimoramento do sinal ECG. Modelos baseados em *autoencoders* e outras arquiteturas neurais são usados para remover artefactos e melhorar a qualidade do sinal antes da análise. Essa abordagem permite a filtragem inteligente do ECG sem perda de informações fisiologicamente relevantes, garantindo uma melhor interpretação clínica (Moody & Mark, 2001).

4.3 Algoritmos de ML Utilizados no Estudo

A seleção dos algoritmos de ML neste estudo foi guiada pela sua ampla adoção e eficácia comprovada em tarefas de classificação binária (Tabela 4.1). Essa escolha é particularmente relevante, dado que há uma escassez de estudos que exploram características espectrais como SEF95, MF e SE na análise de ECG. Portanto, optou-se por algoritmos com desempenho robusto e bem documentado na literatura científica.

Os classificadores utilizados:

- *Random Forest* (RF): É um algoritmo de *ensemble* composto por múltiplas árvores de decisão, onde cada árvore contribui com um voto para a classificação final. Este método é resistente ao *overfitting* e adequado para conjuntos de dados com grande número de atributos (Breiman, 2001). Tem sido amplamente utilizado em problemas biomédicos devido à sua robustez e capacidade de interpretar variáveis de alta dimensionalidade (Jackins *et al.*, 2021).
- *Logistic Regression* (LR): Modelo estatístico clássico para problemas de classificação binária. Apesar da sua simplicidade, é eficaz e amplamente utilizado em contextos médicos para distinguir entre padrões normais e anormais em sinais ECG (Escalona-Morán *et al.*, 2015).
- *Support Vector Machine* (SVM): Utiliza um hiperplano ótimo para separar as classes de forma a maximizar a margem entre elas. Este algoritmo é eficaz para a detecção de padrões subtis em dados biomédicos, sendo particularmente útil quando há sobreposição entre as classes (Osowski *et al.*, 2004).
- *K-Nearest Neighbors* (KNN): Classifica as amostras com base na proximidade dos seus vizinhos mais próximos no espaço das características. A sua simplicidade torna-o útil em bases de dados pequenas e com distribuição espacial clara das classes (de Chazal & Reilly, 2006).
- *Naive Bayes* (NB): Baseado no Teorema de Bayes, este classificador probabilístico assume independência entre as características. É utilizado em contextos clínicos pela sua eficiência computacional e bom desempenho mesmo com conjuntos de dados maiores (Jackins *et al.*, 2021).
- *Decision Tree* (DT): Utiliza estruturas hierárquicas para tomar decisões com base em regras simples. É facilmente interpretável por clínicos, tornando-se

ideal para tarefas em que a transparência do modelo é essencial (Podgorelec *et al.*, 2002).

- *Gradient Boosting* (GB): Constrói modelos sequenciais, onde cada novo classificador corrige os erros do anterior. É amplamente utilizado em previsão de risco e estratificação de pacientes (Chen & Guestrin, 2016).
- *Extra Trees* (ET): Variante mais aleatória do *Random Forest*. Em vez de procurar os melhores pontos de divisão, utiliza seleções aleatórias, o que pode melhorar a generalização e acelerar o tempo de treino (Geurts *et al.*, 2006).

A escolha desses algoritmos deve-se à sua comprovada eficácia em tarefas de classificação binária de ECG, conforme documentado em diversos estudos. Apesar da escassez de pesquisas que utilizem características espectrais como SEF95, MF e SE, a robustez e popularidade desses algoritmos os tornam candidatos ideais para explorar essas características menos estudadas.

Tabela 4.1 – Comparação de Algoritmos de ML utilizados e os seus respetivos Hiperparâmetros.

| Algoritmo | Descrição | Hiperparâmetros |
|-------------------------------|--|---|
| <i>Random Forest</i> | Conjunto de árvores de decisão que combinam previsões para melhorar a precisão. | <i>n_estimators</i> : número de árvores na floresta. <i>max_depth</i> : profundidade máxima das árvores. <i>min_samples_split</i> : mínimo de amostras para dividir um nó. <i>min_samples_leaf</i> : número mínimo de amostras por folha. <i>class_weight</i> : ajusta o peso das classes para lidar com desequilíbrios. |
| <i>Logistic Regression</i> | Modelo estatístico para classificação binária. | <i>max_iter</i> : número máximo de iterações para convergência do otimizador. |
| <i>Support Vector Machine</i> | Classificador que separa dados com um hiperplano ótimo. | <i>kernel</i> : função usada para projetar os dados em outra dimensão (ex.: 'rbf'). <i>C</i> : controla a penalização dos erros de classificação (<i>trade-off</i> entre margem e erro). <i>gamma</i> : define a influência de pontos individuais sobre a forma da fronteira de decisão. <i>probability</i> : ativa a estimativa probabilística. <i>class_weight</i> : ajusta pesos entre classes. |
| <i>K-Nearest Neighbors</i> | Classificação baseada nos vizinhos mais próximos. | <i>n_neighbors</i> : número de vizinhos considerados para votação. <i>weights</i> : define como os vizinhos contribuem (ex.: 'distance' dá mais peso aos mais próximos). <i>metric</i> : métrica de distância usada (ex.: 'manhattan'). |
| <i>Naive Bayes</i> | Modelo probabilístico baseado no Teorema de Bayes. | Não requer ajuste de hiperparâmetros específicos — usa estatísticas dos dados (média e variância) para cada classe. |
| <i>Decision Tree</i> | Modelo de decisão hierárquico baseado em regras simples. | <i>random_state</i> : garante reprodutibilidade ao controlar a aleatoriedade na criação da árvore. |
| <i>Gradient Boosting</i> | Algoritmo baseado em aprendizado sequencial, corrigindo erros anteriores. | <i>n_estimators</i> : número de árvores no modelo. <i>learning_rate</i> : taxa de aprendizagem que controla o impacto de cada árvore. <i>max_depth</i> : profundidade máxima de cada árvore. <i>subsample</i> : fração dos dados usados para treinar cada árvore. |
| <i>Extra Trees</i> | Variante do <i>Random Forest</i> com maior aleatoriedade nas divisões das árvores. | <i>random_state</i> : controla a aleatoriedade para reprodutibilidade. |

Fonte: Adaptado de *Scikit-learn: Machine Learning in Python*

4.4 Utilização e Avaliação de Modelos de *Machine Learning*

A utilização de um modelo de ML para análise de ECG segue um processo estruturado. Primeiramente, os sinais ECG são adquiridos e submetidos a técnicas de pré-processamento, onde são aplicados filtros para remoção de ruídos e artefactos. Posteriormente, são extraídas características relevantes, como a variabilidade da frequência cardíaca e métricas espectrais.

Após a extração, os dados são divididos em conjuntos de treino e teste, garantindo que o modelo possa generalizar adequadamente. O modelo é então treinado, onde aprende os padrões nos dados e ajusta os seus parâmetros de acordo com a estrutura do sinal. Finalmente, o desempenho do modelo é avaliado com base em métricas como exatidão, precisão, sensibilidade, especificidade utilizando a matriz de confusão e AUC-ROC (Área sob a Curva da Característica de Operação do Recetor).

Exatidão representa a proporção de classificações corretas (tanto verdadeiros positivos como verdadeiros negativos) em relação ao total de amostras avaliadas. É uma métrica global que oferece uma visão geral do desempenho do modelo (Sokolova & Lapalme, 2009):

$$\text{Exatidão} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Onde:

- *TP (True Positive)* refere-se aos casos positivos corretamente classificados.
- *TN (True Negative)* representa os casos negativos corretamente identificados.
- *FP (False Positive)* corresponde aos negativos incorretamente classificados como positivos.
- *FN (False Negative)* diz respeito aos positivos não detetados pelo modelo.

Precisão mede a proporção de previsões positivas corretas, ou seja, quantos dos casos classificados como positivos pelo modelo eram, de facto, positivos. Esta métrica é especialmente relevante em contextos onde os falsos positivos têm um custo elevado (Powers, 2011):

$$\text{Precisão} = \frac{TP}{TP+FP} \quad (5)$$

Especificidade indica a capacidade do modelo em identificar corretamente os casos negativos, ou seja, quantas vezes o modelo foi capaz de evitar falsos positivos. Um modelo com alta especificidade é eficaz na identificação dos casos que não pertencem à classe-alvo:

$$\text{Especificidade} = \frac{TN}{TN+FP} \quad (6)$$

Sensibilidade, também conhecida como *recall*, avalia a capacidade do modelo em detetar corretamente os casos positivos reais, isto é, a proporção de verdadeiros positivos em relação ao total de positivos existentes:

$$\text{Sensibilidade} = \frac{TP}{TP+FN} \quad (7)$$

F1-score é a média harmónica entre a precisão e a sensibilidade, oferecendo um equilíbrio entre estas duas métricas. É especialmente útil quando se pretende comparar o desempenho entre modelos em cenários com classes desequilibradas (Sokolova & Lapalme, 2009):

$$\text{F1- score} = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (8)$$

A métrica AUC-ROC (Área sob a Curva ROC) é amplamente utilizada para avaliar o desempenho de modelos de classificação binária, especialmente em contextos médicos, onde é crucial medir a capacidade discriminativa dos algoritmos. A Curva ROC (*Receiver Operating Characteristic*) representa graficamente a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 – especificidade), para diferentes limiares de decisão. A AUC (*Area Under the Curve*) corresponde à área total sob essa curva, variando entre 0 e 1. Um valor de AUC próximo de 1 indica que o modelo tem uma excelente capacidade para distinguir entre classes positivas e negativas, enquanto um valor próximo de 0,5 sugere desempenho semelhante ao de uma classificação aleatória (Fawcett, 2006).

Estas métricas permitem uma análise detalhada da performance dos classificadores, sendo fundamentais para a validação e comparação dos modelos aplicados neste estudo.

5. Materiais e Métodos

O sucesso da detecção de insuficiência cardíaca em pacientes com Doença de Chagas depende, além do modelo de ML em si, da qualidade dos dados utilizados e da metodologia empregada para a extração de características do sinal ECG. Neste estudo, foi utilizada uma base de dados composta por 380 pacientes provenientes do Brasil, todos submetidos a um exame Holter de 24 horas. Para facilitar a análise e processamento dos sinais, optou-se por recolher segmentos representativos de 5 minutos do sinal completo, garantindo um equilíbrio entre a preservação de informações relevantes e a otimização da complexidade computacional.

Este capítulo apresenta uma descrição detalhada dos materiais utilizados, incluindo a base de dados, os critérios de seleção dos sinais e a pré-processamento do ECG no MATLAB, que serviu como base para a extração de características espectrais posteriormente utilizadas nos modelos de ML.

A metodologia adotada neste estudo foi dividida em três grandes etapas:

- Pré-processamento e extração de características do ECG no MATLAB.
- Construção e treino de modelos de ML no *Python*.
- Análise estatística dos resultados para avaliação da eficácia dos classificadores.

Cada uma dessas etapas foi cuidadosamente estruturada para garantir a fiabilidade dos resultados e a explicabilidade das características mais relevantes para a detecção da insuficiência cardíaca na Doença de Chagas.

5.2 Base de Dados

A base de dados utilizada no estudo é composta por 380 pacientes diagnosticados com Doença de Chagas, cujos sinais ECG foram adquiridos através de monitorização contínua de 24 horas (Holter). Para efeitos de análise, foi extraído um segmento de apenas 5 minutos, de forma a facilitar a extração de características e a reduzir a carga computacional do processamento.

5.2.1 Características da Base de Dados

A base de dados foi classificada conforme o grau de insuficiência cardíaca apresentado por cada paciente, seguindo a seguinte estrutura (Tabela 5.1).

No presente estudo, a divisão dos pacientes quanto ao grau de insuficiência cardíaca foi realizada com base nos valores da fração de ejeção pelo método de Teicholz (*FE Teicholz*). A fração de ejeção é um dos parâmetros clínicos mais utilizados para avaliar a função sistólica do coração e quantificar o grau de disfunção do ventrículo esquerdo. Ela representa a porcentagem de sangue que é ejetada do ventrículo esquerdo a cada batimento cardíaco.

A classificação foi efetuada da seguinte forma:

- $FE < 35\%$ – Classe 0: Insuficiência cardíaca grave
Pacientes com severa disfunção ventricular esquerda, frequentemente associados a sintomas marcantes, elevado risco de eventos cardíacos adversos e prognóstico reservado.
- $35\% \leq FE < 45\%$ – Classe 1: Insuficiência cardíaca moderada
Indivíduos com disfunção ventricular significativa, porém com sintomas menos intensos. Este grupo representa um estado clínico intermediário, que exige monitorização contínua.
- $45\% \leq FE < 55\%$ – Classe 2: Insuficiência cardíaca leve
Pacientes com uma ligeira redução da função ventricular, podendo apresentar poucos sintomas ou estar assintomáticos. Apesar da condição ser considerada leve, há risco de progressão da doença.
- $FE \geq 55\%$ – Classe 3: Normal (sem insuficiência cardíaca)
Considerados clinicamente normais em termos de função sistólica, com fração de ejeção preservada.

Este critério baseia-se em valores de referência amplamente aceites na prática clínica cardiológica e reflete diretrizes como as da American Heart Association (AHA) (Heidenreich *et al.*, 2022), que utilizam a fração de ejeção para definir os subtipos de insuficiência cardíaca:

- $<40\%$: Insuficiência cardíaca com fração de ejeção reduzida (HFrEF).
- $40-49\%$: Fração de ejeção ligeiramente reduzida (HFmrEF).
- $\geq 50\%$: Fração de ejeção preservada (HFpEF).

A adaptação dos limites para este estudo (com divisões em 35%, 45% e 55%) permite uma classificação mais adequada à modelação computacional e à aplicação de algoritmos de ML, que exigem classes bem definidas para o treino supervisionado.

Tabela 5.1 - Classificação dos Pacientes Segundo o Grau de Insuficiência Cardíaca

| Classe | Descrição | Código Atribuído | Dimensão |
|---------------|---|-------------------------|-----------------|
| Grave | Pacientes com insuficiência cardíaca severa. | 0 | 43 |
| Moderada | Pacientes com insuficiência cardíaca moderada. | 1 | 119 |
| Leve | Pacientes com insuficiência cardíaca leve. | 2 | 44 |
| Normal | Pacientes sem sinais de insuficiência cardíaca. | 3 | 173 |

A classificação foi utilizada posteriormente para treinar e avaliar modelos de ML, permitindo a predição da gravidade da insuficiência cardíaca com base nas características extraídas do sinal ECG.

5.3 Pré-Processamento dos Sinais ECG no MATLAB

Após a seleção dos sinais ECG de 5 minutos, foi necessário aplicar um conjunto de técnicas de filtragem e segmentação para preparar os dados para análise espectral e extração de características (Figura 5.1). O pré-processamento foi realizado no MATLAB, seguindo os seguintes passos:

1. Aplicação de um filtro passa-banda (Butterworth de sexta ordem) de 1 a 40 Hz para remover ruídos indesejados e preservar componentes fisiologicamente relevantes do ECG.

2. Segmentação do sinal de 5 minutos em janelas de 10 segundos, resultando em 30 janelas de análise por paciente.
3. Cálculo da PSD, utilizando a Transformada de Fourier para analisar a distribuição de potência ao longo das frequências, por janela.
4. Normalização da PSD, garantindo que todas as características fossem comparáveis independentemente da potência do sinal original.
5. Extração de características espectrais, incluindo:
 - SEF95: frequência abaixo da qual está contida 95% da potência espectral.
 - MF: frequência média da distribuição espectral.
 - SE: entropia do espectro de potência.
 - Potências nas bandas de 1 a 64 Hz, de 1 em 1 Hz (Ver anexo 1).
 - Rácios de potência entre bandas específicas (Formula geral: $r = \frac{P_k}{P_j}$ (9))
(Ver anexo 2).
6. Cálculo de estatísticas de cada distribuição de características espectrais, incluindo média, mediana, curtose, desvio padrão, variância e percentil 95.
7. Armazenamento dos resultados em uma matriz 380×666 (Tabela 5.1), onde para cada uma das 111 características (SEF 95, MF, SE, 63 potências e 45 rácios) extraídas por janela de 5 segundos, foram calculadas 6 estatísticas descritivas: média, mediana, curtose, desvio padrão, variância e percentil 95. Assim, cada paciente foi representado por um vetor de 666 características (111×6).
8. A matriz final também incluiu uma coluna adicional com os rótulos de insuficiência cardíaca (0, 1, 2 ou 3), permitindo a posterior aplicação dos modelos de ML no Python.

Neste estudo, foram construídos empiricamente rácios espectrais a partir das potências extraídas de diferentes bandas de frequência da densidade espectral de potência do sinal ECG. A motivação para o uso destes rácios assenta na hipótese de que relações entre potências em diferentes regiões do espectro podem adquirir alterações subtis na dinâmica eletrofisiológica do coração, ainda que não possuam, individualmente, uma interpretação fisiológica universal direta.

5.4 Utilização dos Modelos de *Machine Learning* no Python

A matriz de características espectrais gerada no MATLAB foi importada para o Python para a aplicação dos modelos de ML. O fluxo de trabalho seguiu os seguintes passos (Figura 5.1):

1. Carregamento da matriz de características extraída no MATLAB.
2. Separação das características (x) e dos rótulos (y).
3. Normalização dos dados divididos pelo *target* (grau de insuficiência cardíaca) através da função *Z_score* para garantir que todas as variáveis tivessem a mesma escala.
4. Realização de comparações entre diferentes classes, testando as seguintes combinações:
 - 3 vs 1 (Normal vs Moderada);
 - 3 vs 2 (Normal vs Leve);
 - 3 vs 0 (Normal vs Grave);
 - 2 vs 1 (Leve vs Moderada);
 - 2 vs 0 (Leve vs Grave);
 - 1 vs 0 (Moderada vs Grave).
5. Filtragem e balanceamento dos dados, garantindo um número equilibrado de amostras por classe para evitar viés nos modelos.
6. Divisão dos dados em treino (80%) e teste (20%) para validação dos modelos.
7. Seleção dos classificadores utilizando a biblioteca *Scikit-learn* (Hiperparâmetros *default*), incluindo:
 - *Random Forest* (*n_estimators=200*, *max_depth=None*, *min_samples_split=5*, *min_samples_leaf=2*, *class_weight='balanced'*, *random_state=42*);
 - *Logistic Regression* (*max_iter=1000*);
 - *Support Vector Classifier* (*kernel='rbf'*, *C=1.5*, *gamma='scale'*, *probability=True*, *class_weight='balanced'*);
 - *K-Nearest Neighbors* (*n_neighbors=7*, *weights='distance'*, *metric='manhattan'*);
 - *Naive Bayes*;
 - *Decision Tree* (*random_state=42*);

- *Gradient Boosting* ($n_estimators=150$, $learning_rate=0,05$, $max_depth=3$, $subsample=0.8$);
 - *Extra Trees* ($random_state=42$);
8. Seleção das melhores características com *SelectKBest* (*ANOVA F-test*), testando todos os valores possíveis de k entre **1 e 666**.
 9. Treino e teste de cada classificador para cada comparação, avaliando o desempenho com as métricas:
 - Exatidão;
 - Precisão;
 - Sensibilidade;
 - Especificidade;
 - F1-score;
 - Área sob a curva ROC (AUC-ROC).
 10. Armazenamento dos resultados num ficheiro Excel para análise posterior.

5.5 Análise Estatística dos Resultados

Além da avaliação dos classificadores, foi realizado um teste *ANOVA* para verificar a relevância estatística das diferenças entre os grupos analisados (Figura 5.1). O teste *ANOVA* foi selecionado por se tratar de uma ferramenta estatística adequada à comparação das médias entre três ou mais grupos independentes.

Teste *ANOVA* para cada comparação – Permiteu identificar diferenças estatisticamente significativas entre os grupos analisados (ex.: Normal vs Grave, Leve vs Moderada, etc.).

Teste *ANOVA* global (todos contra todos) – Analisou as diferenças entre todas as classes de insuficiência cardíaca simultaneamente.

Verificação do valor prova – Foram consideradas significativas as diferenças com $p\text{-value} < 0,05$.

Com base nos resultados da ANOVA, foi possível identificar quais características espectrais foram mais predominantes na distinção entre os diferentes graus de insuficiência cardíaca.

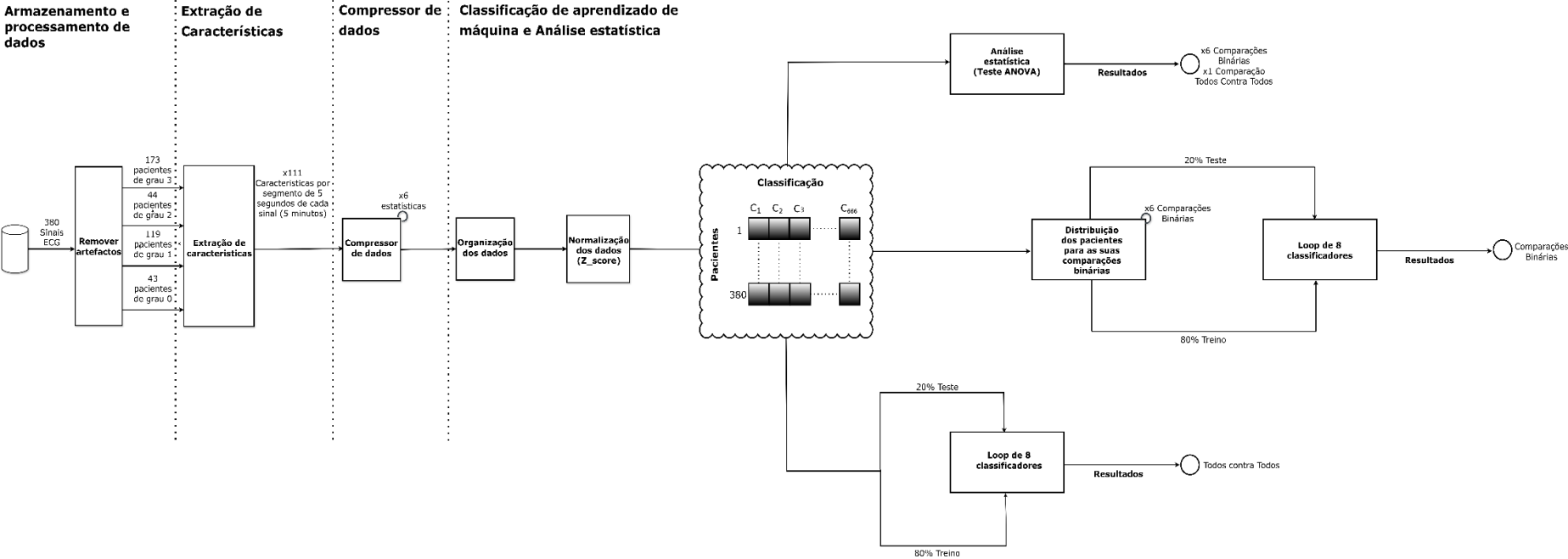


Figura 5.1 – Fluxograma do processo metodológico do estudo.

6. Resultados e Discussão

6.1 Introdução

Neste capítulo, são apresentados os resultados obtidos na classificação da insuficiência cardíaca em pacientes com Doença de Chagas, com base na análise espectral do ECG e na aplicação de algoritmos de ML. Inicialmente, são discutidos os resultados do teste ANOVA, que permitiu identificar as características espectrais mais relevantes para distinguir entre os diferentes estágios da insuficiência cardíaca. Em seguida, são analisados os desempenhos dos modelos de ML utilizados para classificar os pacientes de acordo com a gravidade da doença. Na seção de discussão, são explorados os desafios encontrados na classificação, com ênfase nas dificuldades na diferenciação da insuficiência cardíaca moderada, associadas à heterogeneidade clínica da Doença de Chagas e à influência de comorbidades. Por fim, são propostas estratégias para aprimorar a precisão dos modelos e superar as limitações observadas.

6.2 Resultados

A análise estatística das características espectrais extraídas do ECG foi conduzida através do teste ANOVA, com o objetivo de identificar parâmetros com diferenças estatisticamente significativas entre os diferentes estágios de insuficiência cardíaca. Os resultados (Tabela 6.1), indicam que as comparações envolvendo a classe moderada (1), em particular, moderada vs leve (1 vs 2) e moderada vs grave (1 vs 0), apresentaram o maior número de características significativamente distintas, com 30 (4,50%) e 37 (5,56%) variáveis, respectivamente. Este padrão sugere que, apesar de representar uma condição clínica intermediária, a classe moderada revela diferenças bem definidas em relação aos restantes grupos, refletindo possivelmente a sua complexidade fisiopatológica e heterogeneidade interna.

Adicionalmente, a comparação entre os extremos do espectro clínico, grave vs normal (0 vs 3), também evidenciou um número expressivo de características significativas (19 variáveis; 2,85%), reforçando a distinção entre os estágios mais opostos da insuficiência cardíaca. Na análise ANOVA global, considerando simultaneamente todas as classes, foram identificadas 18 características significativas, com destaque para a variância de P8,

que apresentou o menor valor de p (0,0018). Na comparação moderada vs grave (1 vs 0), a característica r28 Curtose destacou-se como a mais discriminativa, com o menor p-valor observado (0,0003).

Tabela 6.1 – Características Significativas no Teste ANOVA

| Comparação | Nº Características Significativas | Percentual (%) | Característica com Menor p-value | Menor p-value |
|--------------------------------|--|-----------------------|---|----------------------|
| 0 vs 1 (Normal vs Moderada) | 16 | 2,40% | P56 Curtose | 0,0100 |
| 0 vs 2 (Normal vs Leve) | 14 | 2,10% | P19 Curtose | 0,0086 |
| 0 vs 3 (Normal vs Grave) | 19 | 2,85% | r6 Desvio Padrão | 0,0164 |
| 1 vs 2 (Moderada vs Leve) | 30 | 4,50% | P8 Variância | 0,0026 |
| 1 vs 3 (Moderada vs Grave) | 37 | 5,56% | r28 Curtose | 0,0003 |
| 2 vs 3 (Leve vs Grave) | 18 | 2,70% | P35 Curtose | 0,0038 |
| Todos os Grupos | 18 | 2,70% | P8 Variância | 0,0018 |

Fonte: Resultados da pesquisa.

Os modelos de ML foram avaliados com base na capacidade de distinguir entre os diferentes estágios da insuficiência cardíaca (Tabela 6.2). O melhor desempenho foi observado na comparação entre insuficiência cardíaca leve e grave (2 vs 0), onde o classificador Extra Trees alcançou uma AUC-ROC de 0,88, indicando que essas classes apresentam padrões bem diferenciados no ECG. Da mesma forma, a distinção entre normal e grave (3 vs 0) apresentou bons resultados, com o modelo Naive Bayes atingindo

uma AUC-ROC de 0,80. No entanto, as comparações envolvendo a insuficiência cardíaca moderada apresentaram pior desempenho. Em particular, a classificação entre moderada e grave (1 vs 0) apresentou uma AUC-ROC de apenas 0,63, evidenciando a dificuldade dos modelos em distinguir entre esses estágios intermediários da doença. Da mesma forma, a distinção entre moderada e leve (1 vs 2) obteve uma AUC-ROC de 0,72, indicando uma sobreposição considerável entre essas duas classes.

Tabela 6.2 – Desempenho dos Classificadores para Cada Comparação

| Comparação | Classificador | Acurácia (%) | Precisão (%) | Recall (%) | F1-Score (%) | AUC-ROC |
|--------------------------------|----------------------|---------------------|---------------------|-------------------|---------------------|----------------|
| 2 vs 0 (Leve vs Grave) | Decision Tree | 83,3 | 87,5 | 77,8 | 84,2 | 0,84 |
| 1 vs 0 (Moderada vs Grave) | Decision Tree | 77,8 | 77,8 | 77,8 | 77,8 | 0,78 |
| 3 vs 1 (Normal vs Moderada) | Decision Tree | 62,5 | 62,5 | 62,5 | 62,5 | 0,63 |
| 2 vs 1 (Leve vs Moderada) | Random Forest | 77,8 | 85,7 | 66,7 | 80,0 | 0,72 |
| 3 vs 2 (Normal vs Leve) | Extra Trees | 83,3 | 87,5 | 77,8 | 88,9 | 0,88 |
| 3 vs 0 (Normal vs Grave) | Naive Bayes | 83,3 | 75,0 | 66,7 | 90,1 | 0,80 |

Fonte: Resultados da pesquisa.

6.3 Discussão

Apesar de os algoritmos de ML terem demonstrado dificuldade na correta discriminação dos pacientes com insuficiência cardíaca moderada, os resultados do teste ANOVA apontaram o contrário, as comparações envolvendo esta classe (nomeadamente

moderada vs leve e moderada vs grave) foram justamente as que apresentaram o maior número de características espectrais com diferenças estatisticamente significativas. Esta aparente contradição pode ser explicada pelas diferenças fundamentais entre as abordagens estatísticas e os métodos de aprendizagem automática.

O teste ANOVA avalia cada característica isoladamente, ou seja, de forma univariada, para determinar se existem diferenças estatisticamente significativas nas médias entre grupos. Assim, mesmo que uma variável apresente um valor de $p < 0,05$, isso não garante que ela seja eficaz para separar os grupos quando considerada em conjunto com outras variáveis. Esta limitação é relevante no contexto de modelos preditivos multivariados, como os utilizados em aprendizagem automática, que analisam simultaneamente a interação entre várias variáveis. A natureza univariada do ANOVA é amplamente reconhecida na literatura estatística (Scikit-learn, 2025; Kiptoon, 2020), sendo uma ferramenta útil para triagem inicial, mas insuficiente, por si só, para modelação complexa de padrões fisiológicos.

Por outro lado, os modelos de ML operam em contexto multivariado e tentam capturar padrões complexos de interação entre variáveis para gerar decisões robustas. No caso da insuficiência cardíaca moderada, a elevada heterogeneidade clínica e espectral observada, resultado da variabilidade fisiopatológica da Doença de Chagas (Issa, 2018), pode dificultar a discriminação da mesma. Pacientes nessa categoria tendem a partilhar características tanto com estágios mais leves como mais graves da doença, o que resulta numa sobreposição entre classes que compromete a eficácia dos algoritmos preditivos.

Adicionalmente, a presença de comorbidades como hipertensão arterial, diabetes e doença pulmonar obstrutiva (DPOC), frequentemente observadas em pacientes com Doença de Chagas, altera significativamente os padrões eletrocardiográficos (Santos & Falcão, 2020), contribuindo para essa complexidade. Assim, embora a análise estatística identifique diferenças pontuais, essas não se traduzem necessariamente numa separação efetiva quando múltiplas variáveis são analisadas em conjunto por algoritmos de classificação.

Estes resultados reforçam a necessidade de explorar estratégias mais avançadas de seleção de características, incluindo abordagens não lineares e métodos baseados em redes neurais profundas, que possam lidar com padrões complexos e heterogêneos no domínio do ECG.

7. Conclusões Gerais

O presente estudo teve como objetivo desenvolver e avaliar um sistema baseado em ML para a detecção de insuficiência cardíaca em pacientes com Doença de Chagas, utilizando características espectrais extraídas do ECG. Para isso, foram analisados sinais ECG de 380 pacientes, que foram processados no MATLAB, onde foram aplicadas técnicas de filtragem e extração de características espectrais, incluindo SEF95, MF, SE, potências e raios através da PSD. Os dados extraídos foram posteriormente classificados utilizando diversos algoritmos de ML no *Python*, permitindo a diferenciação entre os diferentes graus de insuficiência cardíaca.

Os resultados indicaram que os modelos apresentaram um bom desempenho na distinção entre os estágios extremos da doença, especialmente na comparação entre pacientes normais (3) e graves (0), onde classificadores como *Extra Trees* e *Naive Bayes* obtiveram um AUC-ROC superior a 0,80. No entanto, a classificação da insuficiência cardíaca moderada (1) revelou-se um desafio significativo, com desempenho inferior nas comparações moderada vs leve (1 vs 2) e moderada vs grave (1 vs 0). Essa limitação pode ser explicada pela heterogeneidade clínica da Doença de Chagas, pela sobreposição de características espectrais entre os grupos e pela presença de comorbidades que podem mascarar ou alterar os padrões eletrocardiográficos, tornando mais difícil a distinção entre os diferentes níveis da doença (Santos & Falcão, 2020; Tassi *et al.*, 2014).

A análise estatística dos dados revelou que diversas características espectrais apresentaram diferenças significativas entre os grupos. Dentre elas, a curtose (P56, P19, P35, r28) e a variância (P8) destacaram-se como determinantes para a separação dos estágios da insuficiência cardíaca. O teste ANOVA confirmou a relevância dessas características, mostrando que a comparação moderada vs grave (1 vs 3) apresentou o maior número de características estatisticamente significativas (37 características, 5,56%), indicando que há uma maior distinção entre esses grupos. Por outro lado, a comparação normal vs leve (3 vs 2) teve o menor número de diferenças estatísticas, sugerindo uma maior similaridade entre esses estágios da doença.

Quanto ao desempenho dos classificadores, os melhores resultados foram observados em comparações que envolvem pacientes normais (3) e graves (0), com acurácia acima de 80%. No entanto, os modelos apresentaram baixa precisão nas comparações moderada vs leve (1 vs 2) e moderada vs grave (1 vs 0), com AUC-ROC inferior a 0,72. Esse desempenho inferior pode ser justificado pela variabilidade clínica e

fisiopatológica dos pacientes com insuficiência cardíaca moderada, que frequentemente apresentam características eletrofisiológicas semelhantes às dos outros grupos (Tassi *et al.*, 2014; Pino-Marín *et al.*, 2021).

Apesar dos resultados promissores, este estudo apresenta algumas limitações que podem ter impactado a precisão dos modelos. A heterogeneidade clínica dos pacientes com insuficiência cardíaca moderada dificultou a separação clara entre os grupos, uma vez que a progressão da Doença de Chagas não é linear e pode apresentar diferentes manifestações em cada indivíduo. Estudos indicam que a fibrose miocárdica pode estar presente mesmo em pacientes assintomáticos, afetando a análise do ECG e comprometendo a distinção entre os diferentes estágios da doença (Tassi *et al.*, 2014). Além disso, o modelo baseou-se exclusivamente em dados do ECG, sem considerar outros biomarcadores sanguíneos ou exames de imagem, como ecocardiografia e ressonância magnética cardíaca, que poderiam fornecer informações adicionais sobre a progressão da insuficiência cardíaca e melhorar a precisão da classificação. Outro fator limitante foi o tamanho da base de dados, que, apesar de conter 380 pacientes, apresentou uma distribuição desigual entre os grupos, o que pode ter influenciado o desempenho dos classificadores. Uma base de dados maior e mais balanceada poderia contribuir para a melhoria da robustez dos modelos e para uma generalização mais confiável dos resultados. Adicionalmente, uma limitação importante do presente estudo foi a duração dos sinais ECG utilizados. Embora todos os pacientes tivessem disponível um exame Holter de 24 horas, optou-se por utilizar apenas cinco minutos de registro. Esta decisão foi tomada com o objetivo de reduzir o tempo de processamento computacional. No entanto, tal redução pode ter comprometido a riqueza informativa do sinal, uma vez que eventos cardíacos relevantes podem ocorrer fora do intervalo analisado. Finalmente, a utilização de redes neurais profundas e técnicas de aprendizado não supervisionado poderia permitir a identificação de padrões mais complexos nos sinais ECG e melhorar a performance dos modelos de classificação.

Diante dessas limitações, estratégias futuras devem ser exploradas para aprimorar o desempenho dos modelos preditivos na análise do ECG de pacientes com Doença de Chagas. Modelos híbridos, que combinem aprendizado profundo com técnicas estatísticas tradicionais, também podem ser uma solução eficaz para aumentar a explicabilidade dos resultados e melhorar a precisão da classificação.

A inclusão de dados clínicos complementares poderia ser outra estratégia para melhorar a classificação da insuficiência cardíaca na Doença de Chagas. Da mesma

forma, a incorporação de exames de imagem, como ressonância magnética cardíaca e ecocardiografia, poderia ajudar a diferenciar os estágios da doença de forma mais precisa. A utilização de aprendizado multimodal, que combina diferentes fontes de dados, pode ser uma alternativa promissora para a construção de modelos mais robustos e eficientes.

Em suma, esta nova abordagem demonstrou ser promissora para a detecção da insuficiência cardíaca na Doença de Chagas. Os modelos conseguiram distinguir eficazmente os pacientes normais e graves, mas apresentaram dificuldades na classificação da insuficiência cardíaca moderada, devido a vários fatores. A inclusão de dados clínicos complementares, o uso de redes neurais profundas e a aplicação de métodos mais avançados de extração de características são estratégias promissoras para melhorar a precisão dos modelos. A continuidade deste trabalho poderá contribuir para o desenvolvimento de sistemas de apoio à decisão clínica, auxiliando na detecção precoce da insuficiência cardíaca e melhorando o tratamento dos pacientes com Doença de Chagas.

8. Trabalho Futuro

Os desafios identificados neste estudo abrem espaço para novas investigações e aprimoramentos nos métodos de classificação da insuficiência cardíaca na Doença de Chagas. A principal limitação observada foi a dificuldade na diferenciação dos estágios moderados da insuficiência cardíaca, o que sugere a necessidade de técnicas mais avançadas de extração de características, aplicação de modelos de aprendizado profundo e inclusão de dados complementares para melhorar a precisão dos modelos.

Uma área de aprimoramento crucial reside na compreensão da representação fisiológica das características extraídas. Os resultados obtidos podem sugerir uma nova abordagem para a extração de características em ECG, mas sempre tendo atenção às limitações referidas anteriormente.

Além da extração de características, outra estratégia relevante é a aplicação de modelos avançados de aprendizagem profunda. Os modelos tradicionais de ML utilizados neste estudo demonstraram limitações na diferenciação dos estágios moderados da insuficiência cardíaca, possivelmente devido à sobreposição de características espectrais entre os grupos. Assim, propõe-se a implementação de CNNs para permitir a análise automática do sinal ECG, eliminando a necessidade de extração manual de características e possibilitando a detecção direta de padrões complexos no sinal. Outra abordagem promissora envolve a aplicação de modelos híbridos, que combinam aprendizado profundo com técnicas estatísticas tradicionais, possibilitando maior explicabilidade dos resultados e melhor desempenho na classificação.

Outro fator importante a ser considerado para a melhoria do desempenho dos modelos preditivos é a inclusão de dados clínicos complementares, que pode aumentar significativamente a precisão da classificação da insuficiência cardíaca nos pacientes com Doença de Chagas. O uso de exames de imagem como a ecocardiografia e a ressonância magnética cardíaca (RMC) pode fornecer informações estruturais e funcionais essenciais na avaliação da cardiomiopatia chagásica. A ecocardiografia, considerada o padrão-ouro na estimativa da fração de ejeção ventricular esquerda, permite classificar os pacientes com insuficiência cardíaca segundo o grau de disfunção ventricular, com base em parâmetros funcionais como a fração de ejeção. A aplicação de técnicas de aprendizado multimodal, que combinam informações do ECG com outros exames clínicos, pode permitir a construção de modelos mais robustos e precisos, reduzindo os erros associados

à sobreposição de padrões eletrofisiológicos entre os diferentes estágios da insuficiência cardíaca.

Dessa forma, as melhorias sugeridas para este estudo envolvem o refinamento das técnicas de extração de características, a implementação de modelos de aprendizado profundo mais avançados e a inclusão de informações clínicas complementares. A combinação dessas estratégias pode contribuir para o desenvolvimento de sistemas preditivos mais eficazes, permitindo uma detecção mais precoce e precisa da insuficiência cardíaca na Doença de Chagas, facilitando a tomada de decisão clínica e melhorando o prognóstico dos pacientes.

9. Anexo

Anexo 1 (Tabela dos intervalos de frequência definidos para cada potência):

| Potência (P _i) | Intervalo de Frequência (Hz) |
|----------------------------|------------------------------|
| P1 | [1, 2] |
| P2 | [2, 3] |
| P3 | [3, 4] |
| P4 | [4, 5] |
| P5 | [5, 6] |
| P6 | [6, 7] |
| P7 | [7, 8] |
| P8 | [8, 9] |
| P9 | [9, 10] |
| P10 | [10, 11] |
| P11 | [11, 12] |
| P12 | [12, 13] |
| P13 | [13, 14] |
| P14 | [14, 15] |
| P15 | [15, 16] |
| P16 | [16, 17] |
| P17 | [17, 18] |
| P18 | [18, 19] |
| P19 | [19, 20] |
| P20 | [20, 21] |
| P21 | [21, 22] |
| P22 | [22, 23] |
| P23 | [23, 24] |

| | |
|-----|----------|
| P24 | [24, 25] |
| P25 | [25, 26] |
| P26 | [26, 27] |
| P27 | [27, 28] |
| P28 | [28, 29] |
| P29 | [29, 30] |
| P30 | [30, 31] |
| P31 | [31, 32] |
| P32 | [32, 33] |
| P33 | [33, 34] |
| P34 | [34, 35] |
| P35 | [35, 36] |
| P36 | [36, 37] |
| P37 | [37, 38] |
| P38 | [38, 39] |
| P39 | [39, 40] |
| P40 | [40, 41] |
| P41 | [41, 42] |
| P42 | [42, 43] |
| P43 | [43, 44] |
| P44 | [44, 45] |
| P45 | [45, 46] |
| P46 | [46, 47] |
| P47 | [47, 48] |
| P48 | [48, 49] |
| P49 | [49, 50] |
| P50 | [50, 51] |

| | |
|-----|----------|
| P51 | [51, 52] |
| P52 | [52, 53] |
| P53 | [53, 54] |
| P54 | [54, 55] |
| P55 | [55, 56] |
| P56 | [56, 57] |
| P57 | [57, 58] |
| P58 | [58, 59] |
| P59 | [59, 60] |
| P60 | [60, 61] |
| P61 | [61, 62] |
| P62 | [62, 63] |
| P63 | [63, 64] |

Anexo 2 (Tabela das fórmulas dos rácios):

| Rácio | Fórmula |
|--------------|---------------------|
| R1 | $P3 / P2$ |
| R2 | $P1 / P2$ |
| R3 | $(P4 + P5) / P1$ |
| R4 | $P5 / P1$ |
| R5 | $P6 / P2$ |
| R6 | $(P7 + P8) / P1$ |
| R7 | $P9 / P3$ |
| R8 | $P10 / P4$ |
| R9 | $(P11 + P12) / P2$ |
| R10 | $P13 / P5$ |
| R11 | $(P14 + P15) / P6$ |
| R12 | $P16 / P7$ |
| R13 | $(P17 + P18) / P8$ |
| R14 | $P19 / P9$ |
| R15 | $P20 / P10$ |
| R16 | $(P21 + P22) / P11$ |
| R17 | $P23 / P12$ |
| R18 | $(P24 + P25) / P13$ |
| R19 | $P26 / P14$ |
| R20 | $(P27 + P28) / P15$ |
| R21 | $P29 / P16$ |
| R22 | $(P30 + P31) / P17$ |
| R23 | $P31 / P18$ |
| R24 | $P32 / P19$ |
| R25 | $(P33 + P34) / P20$ |

| | |
|-----|-------------------|
| R26 | P35 / P21 |
| R27 | (P36 + P37) / P22 |
| R28 | P38 / P23 |
| R29 | P39 / P24 |
| R30 | (P40 + P41) / P25 |
| R31 | P42 / P26 |
| R32 | (P43 + P44) / P27 |
| R33 | P45 / P28 |
| R34 | (P46 + P47) / P29 |
| R35 | P48 / P30 |
| R36 | (P49 + P50) / P31 |
| R37 | P51 / P32 |
| R38 | (P52 + P53) / P33 |
| R39 | P54 / P34 |
| R40 | (P55 + P56) / P35 |
| R41 | P57 / P36 |
| R42 | (P58 + P59) / P37 |
| R43 | P60 / P38 |
| R44 | (P61 + P62) / P39 |
| R45 | P63 / P40 |

10. Bibliografía

1. Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences*, 405, 81–90. <https://doi.org/10.1016/j.ins.2017.04.012>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Celler, B., & de Chazal, P. (1997). Selection of parameters from power spectral density, wavelet transforms and other methods for the automated interpretation of the ECG. In *Proceedings of the 13th International Conference on Digital Signal Processing (Vol. 1, pp. 71–74)*. IEEE. <https://doi.org/10.1109/ICDSP.1997.627970>
4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794)*. ACM. <https://doi.org/10.1145/2939672.2939785>
5. de Chazal, P., & Reilly, R. B. (2006). A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 53(12), 2535–2543. <https://doi.org/10.1109/TBME.2006.879163>
6. Dias, J. C. P. et al. (2016). II Consenso Brasileiro em Doença de Chagas, 2015. *Epidemiologia e Serviços de Saúde*, 25(21), 1–10. <https://doi.org/10.5123/s1679-49742016000500002>
7. Diwakar, N., & Raj, D. (2024). ECG-Based Cardiac Disease Detection in Time-Frequency Domain using Grid Search Optimized Wavelet Transforms. *Research Square*. <https://doi.org/10.21203/rs.3.rs-4468988/v2>
8. Escalona-Morán, M. A., Soriano, M. C., Fischer, I., & Mirasso, C. R. (2015). Electrocardiogram classification using reservoir computing with logistic regression. *IEEE Journal of Biomedical and Health Informatics*, 19(3), 892–898. <https://doi.org/10.1109/JBHI.2014.2332001>

9. Fawcett, T. (2006). *An introduction to ROC analysis*. Pattern Recognition Letters, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
10. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
11. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215–e220. <https://doi.org/10.1161/01.CIR.101.23.e215>
12. González, S., Zhao, J., Liu, C., & Nemati, S. (2022). *Interpretable estimation of the risk of heart failure hospitalization from a 30-second electrocardiogram*. *arXiv Preprint*, arXiv:2211.00819. <https://arxiv.org/abs/2211.00819>
13. Guyton, A. C., & Hall, J. E. (2020). *Tratado de Fisiología Médica* (14^a ed.). Elsevier.
14. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65–69. <https://doi.org/10.1038/s41591-018-0268-3>
15. Heidenreich, P. A., et al. (2022). 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure. *Circulation*, 145(18). <https://doi.org/10.1161/cir.0000000000001063>
16. Huang, Y., Zhang, J., Li, X., & Wang, H. (2023). *Heart failure detection using deep learning and Gradient Boosting classifier*. DergiPark. <https://dergipark.org.tr/>
17. Issa, V. S. (2018). The indeterminate form of Chagas disease. *Arquivos Brasileiros de Cardiologia*. <https://doi.org/10.5935/abc.20180027>
18. Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77(5), 5198–5219. <https://doi.org/10.1007/s11227-020-03481-x>
19. Kiptoon, J. D. (2020, October 18). *Feature selection in machine learning*. Medium. <https://medium.com/@jdkiptoon/feature-selection-in-machine-learning-20417d052b80>

20. Lomoio, U., et al. (2023). AUTAN-ECG: An autoencoder-based system for anomaly detection in ECG signals. *TechRxiv*. <https://doi.org/10.36227/techrxiv.24638856.v1>
21. Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45–50.
22. Nunes, M. C. P., et al. (2018). Chagas Cardiomyopathy: An update of current clinical knowledge and management. *Circulation*, 138(12). <https://doi.org/10.1161/cir.0000000000000599>
23. Osowski, S., Hoai, L. T., & Markiewicz, T. (2004). Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering*, 51(4), 582–589. <https://doi.org/10.1109/TBME.2004.824137>
24. Pino-Marín, A., et al. (2021). Chagas Cardiomyopathy: From Romaña sign to heart failure and sudden cardiac death. *Pathogens*, 10(5), 505. <https://doi.org/10.3390/pathogens10050505>
25. Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: An overview and their use in medicine. *Journal of Medical Systems*, 26(5), 445–463. <https://doi.org/10.1023/A:1016409317640>
26. Powers, D. M. W. (2011). *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation*. *Journal of Machine Learning Technologies*, 2(1), 37–63.
27. Rajpurkar, P., Hannun, A. Y., Haghpahani, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *arXiv preprint*. <https://arxiv.org/abs/1707.01836>
28. Rassi, A., Rassi, A., & Marin-Neto, J. A. (2010). Chagas disease. *The Lancet*, 375(9723), 1388–1402. [https://doi.org/10.1016/S0140-6736\(10\)60061-X](https://doi.org/10.1016/S0140-6736(10)60061-X)
29. Santos, É., & Falcão, L. M. (2020). Chagas cardiomyopathy and heart failure: From epidemiology to treatment. *Revista Portuguesa de Cardiologia*, 39(5), 279–289. <https://doi.org/10.1016/j.repc.2019.12.006>
30. Scikit-learn. (2025). *scikit-learn: Machine learning in Python* (v1.6.1). <https://scikit-learn.org/stable/>
31. Sokolova, M., & Lapalme, G. (2009). *A systematic analysis of performance measures for classification tasks*. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>

32. Sörnmo, L., & Laguna, P. (2005). *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Elsevier Academic Press.
33. Tassi, E. M., et al. (2014). Relationship between fibrosis and ventricular arrhythmias in Chagas heart disease without ventricular dysfunction. *Arquivos Brasileiros de Cardiologia*. <https://doi.org/10.5935/abc.20140052>
34. World Health Organization. (2025). *Chagas disease (also known as American trypanosomiasis)*. [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))