



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS QUIXADÁ**  
**CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO**

**JOILSON LEITÃO DE FREITAS JÚNIOR**

**DESENVOLVIMENTO DE UM MODELO DE APRENDIZADO DE MÁQUINA PARA  
DETECÇÃO DE ARRITMIAS CARDÍACAS EM ELETROCARDIOGRAMAS**

**QUIXADÁ**  
**2025**

JOILSON LEITÃO DE FREITAS JÚNIOR

DESENVOLVIMENTO DE UM MODELO DE APRENDIZADO DE MÁQUINA PARA  
DETECÇÃO DE ARRITMIAS CARDÍACAS EM ELETROCARDIOGRAMAS

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Sistemas de Informação  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Antônio Joel Ra-  
miro de Castro

Coorientador: Prof. Me. Bruno Torres  
Marques

QUIXADÁ

2025

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

F936d Freitas Júnior, Joilson Leita de.

Desenvolvimento de um modelo de aprendizado de máquina para detecção de arritmias cardíacas em eletrocardiogramas / Joilson Leita de Freitas Júnior. – 2025.  
50 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Sistemas de Informação, Quixadá, 2025.

Orientação: Prof. Dr. Antônio Joel Ramiro de Castro.

Coorientação: Prof. Me. Bruno Torres Marques.

1. Aprendizado de máquina. 2. Eletrocardiograma . 3. AutoML. 4. Doenças Cardiovasculares. I. Título.

CDD 005

---

JOILSON LEITÃO DE FREITAS JÚNIOR

DESENVOLVIMENTO DE UM MODELO DE APRENDIZADO DE MÁQUINA PARA  
DETECÇÃO DE ARRITMIAS CARDÍACAS EM ELETROCARDIOGRAMAS

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Sistemas de Informação  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de bacharel em Sistemas de Informação.

Aprovada em: 06 de Março de 2025

BANCA EXAMINADORA

---

Prof. Dr. Antônio Joel Ramiro de  
Castro (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Me. Bruno Torres Marques (Coorientador)  
Instituto Centro de Ensino Tecnológico (CENTEC)

---

Prof. Me. Thiago Werlley Bandeira da Silva  
Universidade Federal do Ceará (UFC)

---

Prof. Me. Roberio Rodrigues Feitosa  
Secretaria da Educação do Estado do Ceará (SEDUC)

A Deus

Aos meus pais Celma e Joilson

## **AGRADECIMENTOS**

A Deus, por guiar meus passos e me dar forças nos momentos mais desafiadores.

Aos meus pais, Celma e Joilson, pelo amor incondicional, pelo apoio incansável e por sempre acreditarem no meu potencial. Mãe, sua dedicação e carinho foram essenciais para que eu nunca desistisse. Pai, sua presença e ensinamentos me deram segurança e a certeza de que nunca estive sozinho nesta caminhada.

À minha irmã, Rayane, por seu companheirismo, incentivo e por ser uma fonte constante de apoio e inspiração. Sua presença tornou essa jornada mais leve e significativa, trazendo palavras de encorajamento nos momentos difíceis e celebrando comigo cada pequena conquista. Seu carinho e dedicação foram fundamentais para que eu seguisse em frente com determinação.

Aos meus amigos mestres, Bruno Marques e Carlos Alexandre, por embarcarem comigo nessa jornada, compartilhando desafios, aprendizados e conquistas. A parceria, as conversas e o apoio incondicional tornaram esse percurso mais leve e motivador. Sou imensamente grato pela amizade, pelo incentivo e por cada troca de conhecimento que contribuiu para meu crescimento pessoal e acadêmico.

Ao meu orientador, Prof. Dr. Antônio Joel Ramiro de Castro, pelo apoio fundamental e pela orientação sempre cuidadosa e sábia. Sua expertise, paciência e visão crítica foram determinantes para o sucesso deste trabalho. Sou extremamente grato por sua dedicação em me ajudar a desenvolver o projeto, sempre oferecendo o direcionamento necessário e inspirando-me a buscar excelência.

A todos os meus amigos, colegas e aqueles que, de alguma forma, contribuíram com o meu crescimento ao longo dessa jornada, meu muito obrigado!

"A vida é como andar de bicicleta. Para manter o equilíbrio, é preciso continuar em movimento."  
(Albert Einstein)

## RESUMO

Doenças cardiovasculares são uma das principais causas de morbidade e mortalidade em todo o mundo, sendo as arritmias cardíacas um dos distúrbios mais comuns e desafiadores de diagnosticar. A análise automatizada de sinais cardíacos tem se tornado uma ferramenta essencial para auxiliar no diagnóstico dessas condições, proporcionando maior eficiência e precisão na detecção de anomalias. Este trabalho teve como objetivo desenvolver um modelo de aprendizado de máquina automatizado para identificar e classificar doenças cardíacas a partir de sinais de eletrocardiogramas, utilizando técnicas de AutoML para otimizar a escolha de modelos e hiperparâmetros. A base de dados utilizada MIT-BIH contém múltiplas classes de batimentos, das quais foram selecionadas cinco principais para este estudo. A metodologia envolveu etapas de seleção da base de dados, pré-processamento dos dados, geração de modelos e avaliação dos modelos. O algoritmo Random Forest BAG L2 apresentou o melhor desempenho, com 98,39% de acurácia e 98,34% de F1-score, destacando-se como um modelo robusto e eficiente para a classificação. Durante o desenvolvimento, enfrentaram-se desafios relacionados ao desbalanceamento das classes e à similaridade entre certos tipos de batimentos, os quais foram mitigados por meio de técnicas de ajuste de hiperparâmetros e normalização dos dados. O estudo mostrou que a metodologia empregada foi eficaz na construção de um modelo confiável para a análise de sinais cardíacos.

**Palavras-chave:** aprendizado de máquina; eletrocardiograma; AutoML; doenças cardiovasculares.



## ABSTRACT

Cardiovascular diseases are a leading cause of morbidity and mortality worldwide, with cardiac arrhythmias being one of the most common and challenging disorders to diagnose. Automated analysis of cardiac signals has become an essential tool to aid in the diagnosis of these conditions, providing greater efficiency and accuracy in the detection of anomalies. This work aimed to develop a machine learning model for the classification of heartbeats from electrocardiogram signals, using AutoML techniques to optimize the choice of models and hyperparameters. The MIT-BIH database used contains multiple classes of heartbeats, of which five main ones were selected for this study. The methodology involved stages of database selection, data preprocessing, model generation and model evaluation. The Random Forest BAG L2 algorithm presented the best performance, with 98.39% accuracy and 98.34% F1-score, standing out as a robust and efficient model for classification. During development, challenges related to class imbalance and similarity between certain types of heartbeats were faced, which were mitigated through hyperparameter tuning and data normalization techniques. The study showed that the methodology employed was effective in building a reliable model for analyzing cardiac signals.

**Keywords:** machine learning; electrocardiogram; AutoML; cardiovascular diseases.

## LISTA DE FIGURAS

Figura 1 – Ilustração Básica do Processo de AutoML. . . . .	19
Figura 2 – Medidas fisiológicas e físicas. . . . .	22
Figura 3 – Representação de Ondas de Eletrocardiograma . . . . .	23
Figura 4 – Representação de Ondas de Eletrocardiograma . . . . .	24
Figura 5 – Visão geral dos Procedimentos Metodológicos . . . . .	29
Figura 6 – Distribuição das classes . . . . .	36
Figura 7 – Resultado da Métrica Acurácia dos modelos Gerados. . . . .	40
Figura 8 – Resultado da Métrica F1 Score dos modelos Gerados. . . . .	40
Figura 9 – Resultado da Métrica Recall dos modelos Gerados. . . . .	41
Figura 10 – Resultado da Matriz de Confusão. . . . .	42

## **LISTA DE TABELAS**

Tabela 1 – Comparativo entre os trabalhos relacionados e o proposto . . . . .	28
Tabela 2 – Tabela de anotações de batimentos cardíacos e seus significados . . . . .	34
Tabela 3 – Tabela de conversão dos tipos de batimento cardíaco para valores numéricos.	35
Tabela 4 – Comparação de Desempenho dos Modelos . . . . .	43

## LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizagem de Máquina
DCV	Doenças Cardiovasculares
SBC	Sociedade Brasileira de Cardiologia
ECG	Eletrocardiograma
AC	Arritmias Cardíacas
AutoML	Aprendizado de Máquina Automatizado
CCH	Cardiopatía Chagásica
LSTM	<i>Long Short-Term Memory</i>
CNN	Rede Neural de Convolução

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>14</b>
<b>1.1</b>	<b>Objetivos . . . . .</b>	<b>15</b>
<i>1.1.1</i>	<i>Objetivo Geral . . . . .</i>	<i>15</i>
<i>1.1.2</i>	<i>Objetivos Específicos . . . . .</i>	<i>15</i>
<b>1.2</b>	<b>Organização do Trabalho . . . . .</b>	<b>16</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO . . . . .</b>	<b>17</b>
<b>2.1</b>	<b>Aprendizado de Máquina . . . . .</b>	<b>17</b>
<i>2.1.1</i>	<i>Tipos de Aprendizado de Máquina . . . . .</i>	<i>17</i>
<i>2.1.2</i>	<i>Aprendizado de Máquina Automatizado . . . . .</i>	<i>18</i>
<b>2.2</b>	<b>Métricas para Classificação de Modelos de Aprendizado de Máquina . .</b>	<b>20</b>
<b>2.3</b>	<b>Doenças Cardiovasculares . . . . .</b>	<b>21</b>
<b>2.4</b>	<b>Biossinais . . . . .</b>	<b>21</b>
<b>2.5</b>	<b>Eletrocardiograma . . . . .</b>	<b>22</b>
<b>2.6</b>	<b>Arritmias Cardíacas . . . . .</b>	<b>24</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>26</b>
<b>3.1</b>	<b>Estudo da Arritmia via Classificação do Sinal do ECG Utilizando a Rede Neural Convolucional . . . . .</b>	<b>26</b>
<b>3.2</b>	<b>Aquisição de Eletrocardiograma e Classificação de Arritmias por Machine Learning . . . . .</b>	<b>26</b>
<b>3.3</b>	<b>Aprendizado de Máquina para Predição de Diagnósticos de Doenças Cardiovasculares . . . . .</b>	<b>27</b>
<b>3.4</b>	<b>Análise Comparativa . . . . .</b>	<b>27</b>
<b>4</b>	<b>PROCEDIMENTOS METODOLÓGICOS . . . . .</b>	<b>29</b>
<b>4.1</b>	<b>Seleção da base de dados . . . . .</b>	<b>29</b>
<b>4.2</b>	<b>Pré-processamento dos dados . . . . .</b>	<b>29</b>
<b>4.3</b>	<b>Geração do Modelo de Classificação de Arritmias . . . . .</b>	<b>30</b>
<b>4.4</b>	<b>Avaliação dos modelos . . . . .</b>	<b>31</b>
<b>5</b>	<b>RESULTADOS . . . . .</b>	<b>32</b>
<b>5.1</b>	<b>Seleção do Dataset . . . . .</b>	<b>32</b>
<i>5.1.1</i>	<i>Dataset MIT-BIH . . . . .</i>	<i>32</i>

<b>5.2</b>	<b>Pré-processamento dos dados</b>	<b>34</b>
<b>5.3</b>	<b>Geração dos Modelos</b>	<b>37</b>
<b>5.4</b>	<b>Avaliação dos Modelos</b>	<b>38</b>
<b>5.4.1</b>	<i>Análise dos Resultados</i>	<b>38</b>
<b>5.4.2</b>	<i>Comparação do Melhor Modelo com os Modelos dos Trabalhos Relacionados</i>	<b>43</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>45</b>
	<b>REFERÊNCIAS</b>	<b>47</b>

## 1 INTRODUÇÃO

Os avanços tecnológicos impulsionaram o desenvolvimento de tecnologias de ponta como a inteligência artificial e contribuíram para que as máquinas substituíssem não apenas trabalhos manuais, como também tarefas relacionadas com o uso do que se considera inteligente, através da reprodução de tarefas simples e complexas (Ludermir, 2021).

A Aprendizagem de Máquina (AM), que em inglês significa *Machine Learning*, é uma abordagem de análise de dados que automatiza a criação de modelos analíticos. Pertencente à área de inteligência artificial, essa área parte do princípio de que sistemas podem adquirir conhecimento a partir de dados, reconhecer padrões e tomar decisões com mínima intervenção humana (Alpaydin, 2021).

Algoritmo de aprendizado de máquina relacionado à área da medicina tem sido cada vez mais útil, de maneira a auxiliar nas predições de imagens e no auxílio de todo o processo (diagnóstico, prognóstico e tratamento) de patologias transformando a forma como os profissionais de saúde diagnosticam, tratam e lidam com as doenças além de possibilitar uma melhora na precisão de diagnósticos e na sua eficiência (Braga *et al.*, 2019). Um dos principais motivos de sua importância no diagnóstico de Doenças Cardiovasculares (DCV) é devido essas doenças serem um dos principais motivos de mortalidade no Brasil, de acordo com os dados da Sociedade Brasileira de Cardiologia (SBC), cerca de quatorze (14) milhões de brasileiros têm alguma doença no coração e cerca de quatrocentos (400) mil morrem por ano em decorrência dessas enfermidades, o que corresponde a trinta por cento (30%) de todas as mortes no país.

O Eletrocardiograma (ECG) é uma ferramenta valiosa na avaliação da atividade elétrica do coração, no qual registra a variação do potencial elétrico que ocorre durante cada batimento cardíaco. No entanto, a interpretação precisa desses sinais pode ser desafiadora devido à complexidade das formas de cada pulso, e à vasta gama de Arritmias Cardíacas (AC) existentes, esta grande quantidade são provenientes de desordens no batimento do coração humano, quando ocorre essa característica pode ser a sinalização para uma possível patologia, problemas psicológicos ou provenientes do próprio órgão (Costa *et al.*, 2024). É aqui que a inteligência artificial e o AM desempenham um papel importante, oferecendo a capacidade de automatizar a análise do ECG e identificar anomalias cardíacas (Otsuzi, 2023).

Assim a utilização de Aprendizado de Máquina Automatizado (AutoML) para identificar padrões de AC, desempenha um papel fundamental na medicina moderna, uma vez que AC são desordens no batimento do coração humano, que tendem a formar padrões

nos resultados de ECG, como as arritmias na Cardiopatia Chagásica (CCH), bradiarritmias e as taquicardias, permitindo que os médicos possam usufruir dessa tecnologia para realizar diagnósticos mais precisos e assim um tratamento adequado para condições cardíacas. Um exemplo é o modelo de AM utilizando o modelo de predição de redes neurais *Long Short-Term Memory* (LSTM) que se destacou ao ser utilizado para predições de DCV resultando em uma acurácia de acerto de setenta e sete por cento (77%) em diagnósticos (Coutinho; Filho, 2022). Assim, o desenvolvimento de um modelo de AM para a detecção de arritmias cardíacas em sinais de ECG tem se destacado como uma área de pesquisa de grande relevância e impacto (Souza *et al.* 2019).

Neste contexto, este trabalho busca explorar e desenvolver um modelo de aprendizado de máquina, que seja capaz de identificar arritmias cardíacas em ECG e que sirvam de auxílio na tomada de decisões dos médicos na área da cardiologia. O presente trabalho envolve a coleta de um conjunto de dados coletados de um banco de dados MIT-BIH de resultados ECG, o pré-processamento dos sinais, a seleção de características que venham a formar padrões nos ECG, e a implementação de algoritmos de AutoML que auxiliem os profissionais da saúde no diagnóstico preciso de arritmias cardíacas, contribuindo assim para a melhoria da qualidade de vida dos pacientes e a detecção precoce de DCV.

## **1.1 Objetivos**

### ***1.1.1 Objetivo Geral***

Desenvolver um modelo de aprendizado de máquina automatizado, com intuito de identificar e classificar DCV através de eletrocardiogramas, a fim de auxiliar os cardiologistas.

### ***1.1.2 Objetivos Específicos***

Os objetivos específicos deste trabalho são:

1. Coletar uma base de dados representativa de sinais de ECG, que inclua uma variedade de tipos de arritmias cardíacas.
2. Desenvolver um modelo de AutoML para a detecção de arritmias cardíacas com base no resultados de ECG.
3. Analisar e avaliar os resultados obtidos a partir do modelo desenvolvido.



## **1.2 Organização do Trabalho**

Este trabalho está organizado da seguinte forma. O Capítulo 2 apresentará o referencial teórico do trabalho. No Capítulo 3 são apresentados trabalhos relacionados com o presente trabalho. No Capítulo 4 são apresentados os procedimentos metodológicos. No Capítulo 5 são apresentados os resultados obtidos durante o desenvolvimento deste trabalho. E por fim, o Capítulo 6 apresenta as considerações finais e os trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

Nesta seção serão apresentados os principais conceitos necessários para o entendimento deste trabalho, sendo apresentados da seguinte forma: Seção 2.1 apresenta o conceito de Aprendizado de máquina e aprendizado de máquina automatizado; Seção 2.2 define as métricas para classificação de modelos de AM; Seção 2.3 aborda o assunto doenças cardiovasculares; Seção 2.4 e Seção 2.5 apresentam o conceito de biossinais e eletrocardiograma, respectivamente. E por fim a Seção 2.6 que aborda a definição de arritmias cardíacas.

### 2.1 Aprendizado de Máquina

A aprendizagem de máquina é uma subárea da inteligência artificial que se dedica a criar algoritmos e modelos capazes de permitir que sistemas computacionais aprendam a partir de dados e aprimorem seu desempenho em tarefas específicas ao longo do tempo, sem a necessidade de programação explícita (Station, 2023). Tem como base uma sólida fundamentação teórica que abrange conceitos de estatística, álgebra linear, cálculo, probabilidade e otimização. Além disso, ela incorpora princípios da teoria da informação e da psicologia cognitiva para compreender como os seres humanos aprendem e se adaptam, com o objetivo de replicar esses processos em sistemas computacionais.

A abordagem teórica da aprendizagem de máquina possibilita a criação de modelos matemáticos e algoritmos que conseguem generalizar a partir de dados de treinamento, tornando-se ferramentas poderosas em diversas aplicações, tais como reconhecimento de padrões, processamento de linguagem natural, visão computacional, medicina, finanças e muito mais, ou seja, uma compreensão sólida dos princípios teóricos é crucial para o desenvolvimento e o sucesso das aplicações de aprendizagem de máquina (Ludermir, 2021).

#### 2.1.1 Tipos de Aprendizado de Máquina

A área de AM na inteligência artificial está focada na criação de algoritmos e modelos que capacitam um sistema a identificar padrões em uma grande quantidade de dados e tentar melhorar seus resultados obtidos ao longo do treinamento. Essas categorias representam abordagens distintas na resolução de problemas de AM, sendo cada uma adaptada a diferentes tipos de dados e contextos de aplicação (Monard; Baranauskas, 2003). Existem diversos tipos de AM, sendo os principais:

**Supervisionado:** Modelo que será utilizado no presente trabalho, este modelo funciona de maneira que a base de dados contém informações de entrada juntamente com as correspondentes saídas (variável resposta), é realizado um processo de divisão da base em conjuntos de treino e teste. Em seguida, um algoritmo de aprendizado supervisionado analisa o conjunto de treino para criar uma função inferida. Essa função é então testada no conjunto de dados de teste para avaliar sua eficácia, possibilitando seu uso na geração de previsões confiáveis para a variável de interesse. Os algoritmos de AM supervisionado são empregados de duas maneiras, abordando problemas de regressão e classificação (Monard; Baranauskas, 2003).

**Não Supervisionado:** É uma abordagem de AM na qual os usuários não necessitam supervisionar diretamente o modelo. Em vez disso, permite que o modelo opere de forma autônoma para identificar padrões e informações que não foram previamente detectados, uma vez que lida com dados que não possuem rótulos associados (variável de resposta). Quando a base de dados contém exclusivamente informações de entrada, buscando identificar padrões ou identificando grupos nos dados de entrada que sejam tendências (Mahesh, 2020).

**Por reforço:** O AM interage de forma contínua com um ambiente em constante mudança, buscando alcançar um objetivo específico. Esse tipo visa melhorar as estratégias do algoritmo através do *feedback* de recompensa, denominado sinal de reforço, tornando o sistema volátil e interativo com o seu ambiente. A solução mais otimizada para o problema é alcançada pelo algoritmo por meio de iterações que envolvem erros e tentativas repetidas (Araújo, 2024).

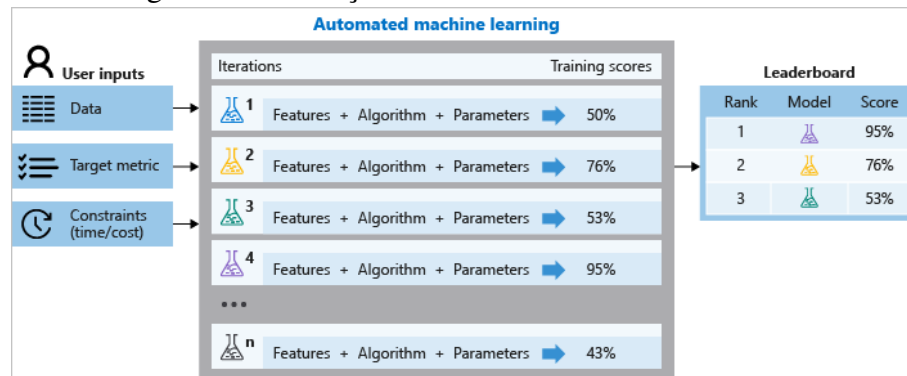
### 2.1.2 *Aprendizado de Máquina Automatizado*

O AutoML pode ser entendido como uma tarefa que objetiva automatizar várias etapas do processo de construção de modelos de AM, desde a seleção e pré-processamento de dados até a otimização dos hiperparâmetros do modelo, além disso, acelera o desenvolvimento de modelos e pode melhorar significativamente a eficiência na resolução de problemas que envolvem grandes volumes de dados (Nagarajah; Poravi, 2019).

O AM envolve o uso automatizado de algoritmos e técnicas para permitir que sistemas computacionais aprendam padrões e sejam direcionados sem intervenção humana direta. Inicialmente, os dados relevantes são coletados e pré-processados. Na Figura 1 ilustra como basicamente ocorre o processo de AutoML, inicialmente é inserido o banco de dados a ser utilizado, a métrica a ser utilizada como medida de desempenho para possíveis comparações entre modelos. Em seguida, algoritmos de aprendizado automático são selecionados e aplicados aos

dados. Durante o treinamento, o modelo é ajustado iterativamente para melhorar seu desempenho em um conjunto de dados de treinamento. Uma vez treinado, o modelo pode ser aplicado a novos dados para fazer variações, em seguida métricas de avaliação são usadas para medir a precisão do modelo. No caso da Figura 1 foi utilizada a métrica score, quanto mais alto o score melhor foi o desempenho do modelo treinado.

Figura 1 – Ilustração Básica do Processo de AutoML.



Fonte: Learn Microsoft (2023).

O AutoML tem se mostrado uma abordagem promissora para otimizar o desenvolvimento de modelos de aprendizado de máquina, reduzindo significativamente o tempo e os recursos necessários para a construção de sistemas preditivos eficientes. Com o avanço das técnicas de AutoML, tornou-se possível automatizar a seleção de modelos, a engenharia de características e a otimização de hiperparâmetros, facilitando a aplicação dessas tecnologias em domínios complexos, como a medicina (Hutter *et al.*, 2019).

Na área da saúde, o uso de AutoML tem permitido avanços significativos na análise de dados clínicos e no diagnóstico de doenças. Em particular, a detecção de arritmias cardíacas por meio de eletrocardiogramas tem-se beneficiado dessas técnicas, possibilitando a construção de modelos altamente precisos sem a necessidade de intervenção manual intensiva. Estudos recentes demonstram que abordagens baseadas em AutoML podem superar métodos tradicionais ao identificar padrões em grandes volumes de dados de ECG e sugerir diagnósticos com alta confiabilidade (Shen *et al.*, 2020; Wang *et al.*, 2021).

A eficiência do AutoML decorre da capacidade dessas ferramentas de explorar automaticamente diferentes configurações de algoritmos e parâmetros, buscando a melhor combinação para um determinado conjunto de dados. Ferramentas como Auto-sklearn, H2O.ai e Google AutoML têm sido amplamente utilizadas para esse propósito, demonstrando resultados promissores em diversas aplicações médicas, incluindo a predição de eventos cardíacos adversos

(Feurer *et al.*, 2019; He *et al.*, 2021).

Além disso, técnicas recentes de AutoML baseadas em redes neurais profundas têm permitido avanços na análise de sinais fisiológicos, como ECGs. Modelos gerados automaticamente por AutoML têm alcançado desempenho comparável ao de especialistas humanos na detecção de arritmias, tornando essa tecnologia uma aliada essencial na modernização dos sistemas de saúde (Liu *et al.*, 2022).

A integração de AutoML na detecção de arritmias cardíacas tem o potencial de democratizar o uso de aprendizado de máquina na medicina, permitindo que profissionais da saúde se beneficiem dessas tecnologias sem a necessidade de expertise em ciência de dados. À medida que a pesquisa nessa área avança, espera-se que soluções cada vez mais robustas e acessíveis sejam desenvolvidas, ampliando a adoção do AutoML em diagnósticos clínicos automatizados.

## 2.2 Métricas para Classificação de Modelos de Aprendizado de Máquina

Os modelos de classificação de dados são técnicas de AM que visam classificar rótulos a dados com base em padrões identificados durante o treinamento do modelo. Esses modelos são utilizados em várias áreas de aplicação de AM, incluindo reconhecimento de padrões, processamento de linguagem natural, visão computacional, entre outros, a métrica utilizada para determinação do “melhor modelo” depende do problema analisado (Mariano, 2021). Dentre os modelos mais comuns de classificação estão:

**Acurácia** é considerada uma das métricas mais simples e importantes. Essa métrica vai avaliar simplesmente o percentual de acertos, ou seja, ela pode ser obtida pela razão entre a quantidade de acertos e o total de entradas.

**Precisão** é uma métrica de avaliação que mede a proporção de exemplos classificados corretamente em relação ao total de exemplos. É uma medida comum para avaliar o desempenho de modelos de classificação (Schneider, 2018). Essa métrica é fundamental em problemas onde a identificação correta de casos positivos é mais importante do que evitar falsos positivos, como no diagnóstico de doenças graves (Han *et al.*, 2022).

**F1-score** é simplesmente uma maneira de observar apenas 1 métrica ao invés de duas (precisão e recall) em alguma situação. É uma média harmônica entre as duas, que está muito mais próxima dos valores menores do que uma média aritmética simples (Mariano, 2021).

## 2.3 Doenças Cardiovasculares

Enfermidades cardíacas, também conhecidas como DCV, compreendem um conjunto de problemas de saúde que afetam tanto o coração quanto os vasos sanguíneos, e representam uma das principais causas de morbidade e mortalidade em escala global. Essas condições abrangem uma ampla variedade de questões, que vão desde a aterosclerose, hipertensão arterial, doença coronariana, insuficiência cardíaca, arritmias até doenças cardíacas congênitas (Ribeiro *et al.*, 2012).

Anualmente cardiopatias são responsáveis por 15,9 milhões de óbitos, os fatores de risco para o desenvolvimento de doenças cardíacas incluem práticas de vida inadequadas, como uma dieta desequilibrada, falta de atividade física, consumo excessivo de álcool, além de predisposição genética e outros fatores relacionados à saúde. A prevenção, a identificação precoce e o tratamento apropriado desempenham um papel essencial na gestão das doenças cardíacas, com o objetivo de melhorar a qualidade de vida e reduzir os riscos associados a possíveis DCV, adotar um estilo de vida saudável e buscar avaliações médicas regulares são medidas fundamentais para manter um coração saudável (SAÚDE, Organização Pan-Americana de , 2017).

## 2.4 Biossinais

Os biossinais são grandezas fisiológicas mensuráveis que refletem o funcionamento dos sistemas biológicos. Eles podem ser adquiridos por meio de sensores e dispositivos eletrônicos, permitindo a análise de funções corporais e a identificação de possíveis anormalidades (Sornmo; Laguna, 2005). Dentre os principais tipos de biossinais, destacam-se os sinais elétricos, como eletrocardiograma (ECG), eletroencefalograma (EEG) e eletromiograma (EMG), bem como sinais mecânicos, químicos e térmicos (Rangayyan, 2015).

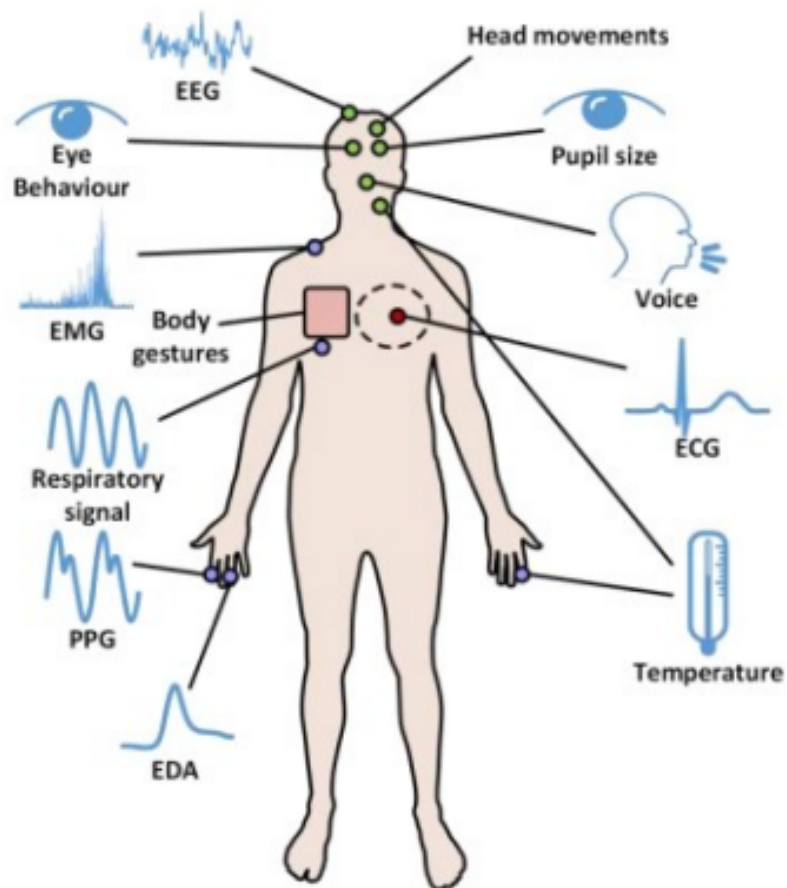
O eletrocardiograma (ECG) é um dos biossinais mais utilizados na medicina diagnóstica, sendo essencial para a análise da atividade elétrica do coração. Ele permite detectar condições cardíacas como arritmias, isquemias e bloqueios de condução (Inan *et al.*, 2020). O desenvolvimento de modelos de aprendizado de máquina para a análise automática do ECG tem sido uma área crescente de pesquisa, proporcionando maior precisão e rapidez na identificação de padrões anormais (Acharya *et al.*, 2017).

A aquisição e o processamento de biossinais envolvem diversas etapas, como a

captação dos sinais por meio de eletrodos, a filtração para remoção de ruídos e a extração de características para a análise quantitativa (Chen *et al.*, 2019). O avanço das técnicas de aprendizado de máquina tem possibilitado a implementação de algoritmos capazes de classificar padrões de biossinais e auxiliar no diagnóstico médico.

Dessa forma, o estudo e a aplicação de biossinais na medicina têm-se mostrado fundamentais para o aprimoramento da detecção precoce e do tratamento de doenças cardíacas. A integração de técnicas de aprendizado de máquina nesse campo representa um grande avanço tecnológico, promovendo diagnósticos mais precisos e acessíveis.

Figura 2 – Medidas fisiológicas e físicas.



Fonte: Giannakakis *et al.* (2019)

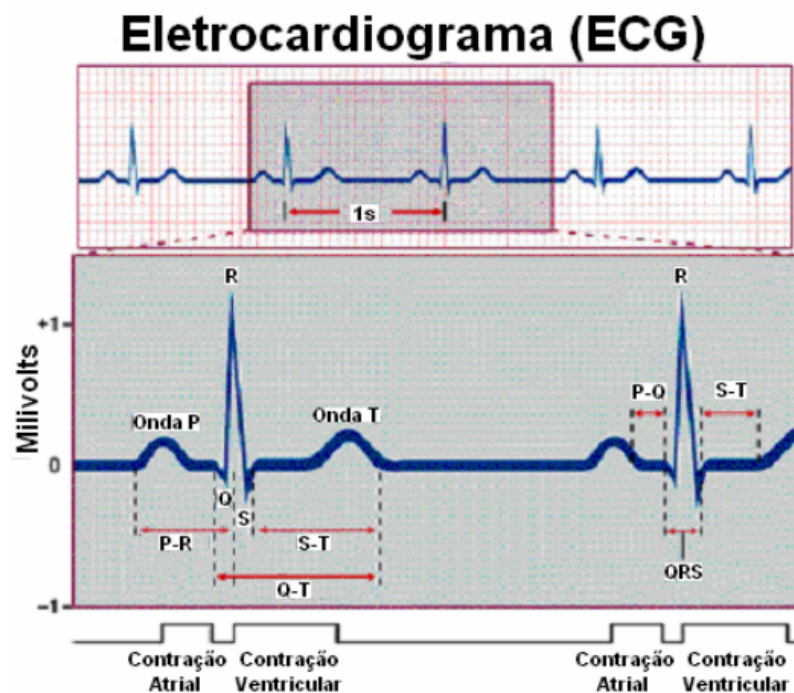
## 2.5 Eletrocardiograma

O ECG é um método que surgiu por volta de mil novecentos e dois (1902), desenvolvido pelo médico Willem Einthoven, com objetivo principal de auxiliar na identificação de mudanças de arritmia do coração. Existem critérios eletrocardiográficos para a caracterização

de arritmias como sobrecargas das câmaras cardíacas, bloqueios intraventriculares, isquemia, lesão e área eletricamente inativa do miocárdio, entre outros. É um exame fundamental para o diagnóstico cardíaco, em que o resultado é um gráfico que é representado por uma linha que define os sinais elétricos provenientes dos eletrocardiogramas (Godwasserand, 2019).

A Figura 3 ilustra um ECG que é um gráfico que representa a atividade elétrica do coração ao longo do tempo. Geralmente, consiste em linhas que mostram as variações nos potenciais elétricos do coração. A linha de base representa o estado elétrico em repouso do coração. As ondas P indicam a despolarização dos átrios, seguidas pelo complexo de ondas QRS, em que a primeira onda tem polaridade negativa (onda Q), a segunda tem polaridade positiva (onda R) e a terceira é polarizada negativamente (onda S), e juntas representam a despolarização dos ventrículos. O intervalo QT indica a duração total da despolarização e repolarização ventricular. As ondas T representam a repolarização dos ventrículos. O ECG é fundamental para diagnosticar arritmias, distúrbios cardíacos e avaliar a saúde cardíaca de um paciente, fornecendo informações cruciais sobre a função elétrica do coração (Schwarz, 2009).

Figura 3 – Representação de Ondas de Eletrocardiograma .



Fonte: Repositório UFSM (Acesso em 2023).

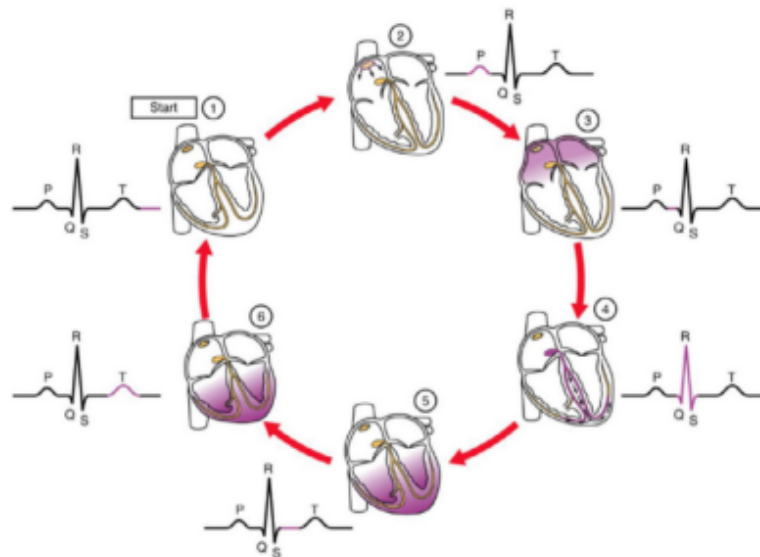
Além de sua importância diagnóstica, o ECG é amplamente utilizado em pesquisas médicas e desenvolvimento de novas tecnologias voltadas para a saúde cardiovascular. Com o avanço da inteligência artificial e aprendizado de máquina, novas abordagens vêm sendo



exploradas para a análise automática de sinais eletrocardiográficos, permitindo a detecção precoce de anomalias cardíacas e aprimorando a precisão dos diagnósticos (Berkaya *et al.*, 2018). O eletrocardiograma (ECG) é um dos principais métodos de monitoramento e análise da atividade elétrica do coração. Ele é amplamente utilizado na detecção de anomalias cardíacas, como arritmias, infarto do miocárdio e bloqueios de condução (Berkaya *et al.*, 2018). O ECG registra a variação do potencial elétrico gerado pelo coração ao longo do tempo, utilizando eletrodos posicionados na superfície do corpo. A análise do ECG envolve a identificação de padrões específicos nas ondas P, QRS e T, que representam diferentes fases do ciclo cardíaco. Estudos recentes destacam a aplicação de algoritmos de aprendizado de máquina para aprimorar a classificação automática de sinais de ECG, permitindo diagnósticos mais rápidos e precisos (Acharya *et al.*, 2017).

Para ilustrar a morfologia do ECG e suas principais ondas, a Figura 4 é possível observar sinais de ECG associados a um ciclo cardíaco comum. E como os pontos citados (P, Q, R, S e T) estão associados ao ciclo eletrocardiográfico.

Figura 4 – Representação de Ondas de Eletrocardiograma .



Fonte: (Berkaya *et al.*, 2018).

## 2.6 Arritmias Cardíacas

Arritmias cardíacas, também conhecida como disritmia cardíaca, é uma condição cardíaca identificada por modificações no ritmo ou na frequência dos batimentos cardíacos.

Esses batimentos cardíacos desordenados podem ter a sua origem relacionados a doenças cardíacas Estruturais: Problemas nas estruturas do coração, como doenças das válvulas cardíacas, cardiomiopatias ou cicatrizes de infarto do miocárdio, podem predispor a arritmias, esses fatores contribuem para alteração no ECG e são categorizadas conforme as mudanças que induzem na frequência cardíaca (Sousa, 2018).

Há várias formas de arritmias cardíacas, cada uma associada a um padrão distinto que facilita sua detecção e categorização em possíveis doenças. Essas arritmias podem ser classificadas em dois grupos principais. O primeiro grupo engloba as arritmias que se caracterizam por um único batimento cardíaco irregular, denominadas arritmias morfológicas. O segundo grupo inclui as arritmias que consistem em uma sequência de batimentos cardíacos irregulares, conhecidas como arritmias rítmicas (Luz *et al.*, 2016).

### 3 TRABALHOS RELACIONADOS

Esta seção apresenta os trabalhos relacionados a este trabalho, com uma breve descrição sobre os mesmos, sendo eles: (i) Estudo da Arritmia via Classificação do Sinal do ECG Utilizando a Rede Neural Convolucional; (ii) Aquisição de Eletrocardiograma e Classificação de Arritmias por Machine Learning; (iii) Aprendizado de Máquina para Predição de Diagnósticos de Doenças Cardiovasculares. Ao final, serão apontadas semelhanças e diferenças entre os trabalhos relacionados e este trabalho.

#### 3.1 Estudo da Arritmia via Classificação do Sinal do ECG Utilizando a Rede Neural Convolucional

O artigo desenvolvido por Wu *et al.* (2021a) aborda a crescente importância das DCV uma das principais causas de morte atualmente. Com base nisso, os autores buscam desenvolver um modelo de AM para facilitar a identificação de DCV já que o modelo convencional dificulta a identificação dessas doenças por envolver a análise manual dos ECG.

Para a realização do estudo, os autores utilizaram a base de dados MIT-BIH para segmentação de dados em que cada batimento cardíaco é marcado com uma anotação de doença. Além disso, selecionam cinco batimentos cardíacos para classificação: normal, bloqueio de ramo esquerdo, bloqueio de ramo direito, batimentos atriais prematuros, batimentos ventriculares prematuros e propõem uma estrutura de rede unidimensional de Rede Neural de Convolução (CNN) de doze (12) camadas para classificar as cinco (5) subclasses de arritmia cardíaca.

As CNN consistem na criação de camada de entrada, camada de convolução, camada de *pooling*, camada totalmente conectada e camada de saída que, em conjunto, têm por objetivo classificar cinco subclasses de arritmias cardíacas. Em contraste com as redes neurais tradicionais, a CNN possui camadas de convolução e *pooling*, que podem extrair e mapear recursos dos dados de entrada para acelerar o aprendizado e reduzir o ajuste excessivo.

#### 3.2 Aquisição de Eletrocardiograma e Classificação de Arritmias por Machine Learning

O projeto desenvolvido por Silva *et al.* (2022a) propõe a criação de um dispositivo de monitoramento cardíaco pessoal que usa as técnicas de Inteligência Artificial, de maneira mais específica a aprendizagem de máquina, para identificar arritmias cardíacas. A função desse dispositivo é detectar possíveis arritmias cardíacas e em seguida notificá-las via email para

os médicos ou profissionais da área, a fim de obter uma resposta de emergência mais rápida. Quatro classes de arritmias foram selecionadas como base: normal, supraventricular ectópico, ventricular ectópico e fusão de normais e ventriculares.

Os dados que foram utilizados para a realização de testes do trabalho foram provenientes da base de dados MIT-BIH *Arrhythmia Database* e o aplicativo *Classification Learner* do Matlab, os melhores modelos foram identificados como *Ensemble*(SubspaceKNN) e obteve a melhor acurácia (74,4%) durante os testes.

Foi implementado um aplicativo com interface para o usuário com a finalidade de melhorar a usabilidade do dispositivo, dessa forma essa inovação científico-tecnológica contribui para a acessibilidade ao monitoramento cardíaco.

### **3.3 Aprendizado de Máquina para Predição de Diagnósticos de Doenças Cardiovasculares**

O estudo desenvolvido pelos autores Coutinho e Filho (2022) apresenta modelos de AM para classificar diagnósticos de DCV visando orientar tratamentos. Foram exploradas três abordagens, incluindo algoritmos tradicionais, *ensemble learning* (aprendizado por agrupamento) e redes neurais LSTM.

Para o desenvolvimento do trabalho foi utilizada a arquitetura LSTM que se destacou atingindo setenta e sete por cento (77%) de precisão, por ser bem adequada para classificar e processar séries temporais de dados, reforçando sua superioridade. No trabalho foi utilizada a base de dados PTB-XL uma base de dados contendo resultados de exames ECG de pacientes portadores de DCV, a partir disso foi realizado o pré-processamento dos dados e dividido em cinco (5) superclasses, após essas etapas foi treinado o modelo de AM para fazer a predição de possíveis DCV.

### **3.4 Análise Comparativa**

Em Wu *et al.* (2021a) os autores desenvolvem um modelo de AM para facilitar a identificação de DCV utilizando a base de dados MIT-BIH para realizar a segmentação de dados. Ele é similar ao proposto pois utiliza o AM para detecção de AC, além de utilizar a base de dados MIT-BIH arritmia, mesma base que este trabalho pretende utilizar. Além disso, eles desenvolvem o modelo com o intuito de alcançar uma melhor acurácia em relação a outros tipos de redes neurais tradicionais, alcançando um melhor desempenho em precisão e capacidade de antirruídos

nos sinais, que

No trabalho de Silva *et al.* (2022a) é criado um dispositivo de monitoramento cardíaco pessoal que utiliza técnicas de AM para reconhecer arritmias, enquanto este busca desenvolver um modelo de AM para detecção de AC. Além disso, o trabalho relacionado se diferencia ao utilizar Arduino e desenvolver seu próprio aparelho de ECG. Como semelhanças, é possível citar que ambos os trabalhos utilizam a mesma base de dados MIT-BIH e selecionam tipos de arritmias como base, caso o paciente apresente alguma alteração durante o exame.

O trabalho de Coutinho e Filho (2022) aplica um modelo de AM baseado em redes neurais LSTM para predição de DCV. Ele apresenta semelhanças com este trabalho, uma vez que ambos possuem a mesma finalidade que é prever DCV utilizando o AM, e divergem quanto à base de dados já que os autores utilizam o PTB-XL e este trabalho pretende utilizar o MIT-BIH, o modo de seleção de superclasses e a técnica para a predição, que são baseadas em redes neurais.

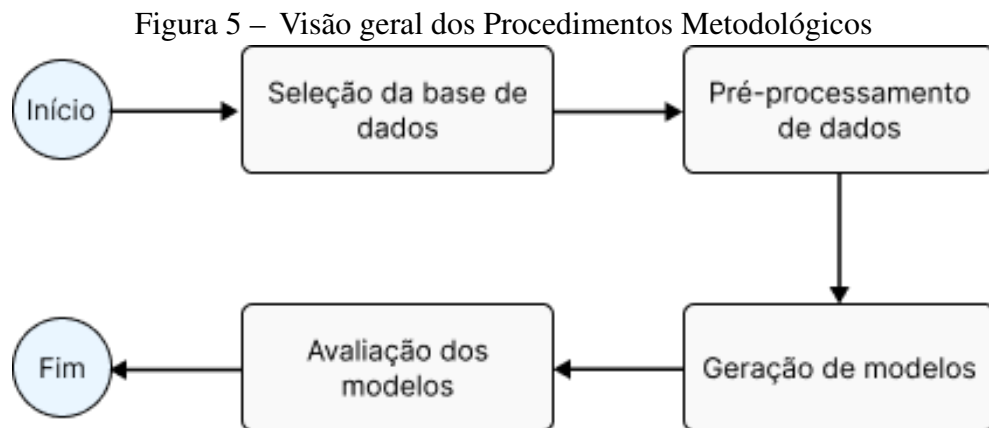
Tabela 1 – Comparativo entre os trabalhos relacionados e o proposto

Referência do trabalho	Utiliza AutoML	Utiliza modelo supervisionado	Métrica Acurácia	Mit-Bih arrhythmia
(Wu <i>et al.</i> , 2021a)	Não	Sim	Não	Sim
(Silva <i>et al.</i> , 2022a)	Não	Não	Não	Sim
(Coutinho; Filho, 2022)	Não	Sim	Sim	Não
Trabalho proposto	Sim	Sim	Sim	Sim

Fonte: Elaborado pelo autor.

## 4 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo são apresentados os passos para a execução deste trabalho. A Figura 5 apresenta um fluxograma geral das etapas dos procedimentos metodológicos que serão mais detalhadas nas seções seguintes. A etapa 1 corresponde à seção 4.1, a etapa 2 é referente à seção 4.2, e as etapas 3 e 4 representam as seções 4.3 e 4.4, respectivamente.



Fonte: Elaborado pelo autor (2025)

### 4.1 Seleção da base de dados

A primeira etapa para o desenvolvimento deste trabalho consiste em entender a base de dados escolhida. Desse modo, nessa etapa foi realizado um estudo dos dados do dataset MIT-BIH<sup>1</sup> e buscou-se entender a estrutura dos arquivos e as informações disponíveis. Essa base de dados foi criada entre os anos de mil novecentos e setenta e cinco (1975) e mil novecentos e setenta e nove (1979), ela contém cerca de quarenta e oito (48) registros de sinais de dados de ECG, cada registro com duração de trinta (30) minutos de leitura de sinais consistindo em vários registros de ECG de diferentes pacientes, cada um em um arquivo separado.

### 4.2 Pré-processamento dos dados

Nesta etapa foi realizado o pré-processamento dos dados de ECG com o objetivo de garantir a confiabilidade dos resultados posteriores. Devido à natureza das arritmias cardíacas e à sua representação em ECG, esse processo consiste em extrair características relevantes a partir desses dados, isso permitirá que, em vez de treinar o modelo diretamente com os dados originais,

<sup>1</sup> <https://physionet.org/content/mitdb/1.0.0/>

os quais são séries temporais, o treinamento seja realizado com base em dados gerados, que são tabulares.

Com essa abordagem diversos problemas que prejudicam a interpretabilidade dos modelos serão evitados. Pois, é frequente que os dados provenientes de biossinais sejam incompletos, contendo ruído e *outliers*, o que demanda a aplicação de técnicas específicas de limpeza e tratamento (Shende *et al.*, 2022). Além disso, os modelos de aprendizado de máquina aplicados a biossinais frequentemente exigem uma quantidade abundante de dados e recursos computacionais. Essa complexidade dificulta a interpretação do processo de tomada de decisão e identificação do modelo (Alsharef *et al.*, 2022).

A importância da realização dessa fase ocorre pelo fato de frequentemente os sinais de ECG serem capturados em ambientes sujeitos a interferências e isso pode resultar em leituras imprecisas ou em interpretações errôneas do sinal elétrico do coração.

### 4.3 Geração do Modelo de Classificação de Arritmias

A ferramenta de AutoML escolhida para este trabalho foi o *AutoGluon*, uma biblioteca de AutoML de código aberto desenvolvida pela AWS<sup>2</sup> (Amazon Web Services) em parceria com a Apache MXNet. De acordo com Erickson *et al.* (2020) o *AutoGluon* se destaca por ser mais rápido, robusto e significativamente mais preciso em tarefas de classificação e regressão quando comparado a outras ferramentas de AutoML, como TPOT, H2O, AutoWEKA, auto-sklearn e Google AutoML Tables.

A utilização da ferramenta AutoML automatiza várias etapas do processo de desenvolvimento de modelos, desde o pré-processamento de dados até a seleção do modelo, contribuindo no momento da escolha dos algoritmos mais adequados. Sendo assim, nesta etapa será gerados modelos utilizando o *AutoGluon* devido sua capacidade de lidar com dados tabulares e na classificação AC.

A ferramenta selecionada gerou modelos de AM, ajustando automaticamente seus hiperparâmetros para otimizar seu desempenho. Além disso, também foram analisados e comparados diferentes algoritmos de AM, incluindo abordagens clássicas, como *Random Forest*, *KNeighbors* e *Extra trees*. A escolha final foi baseada na capacidade de cada modelo em lidar com os dados do eletrocardiograma.

---

<sup>2</sup> <https://aws.amazon.com/pt/>

#### 4.4 Avaliação dos modelos

Nesta etapa foi realizado uma comparação entre os modelos com o intuito de analisar os resultados obtidos entre ambos e verificar qual consegue uma melhor precisão no diagnósticos das DCV em relação aos outros. Em um estudo realizado por Saba *et al.* (2018), a eficácia de diferentes modelos de aprendizado de máquina na classificação de arritmias foi avaliada, mostrando que, embora os modelos tradicionais apresentem bons resultados, os modelos baseados em AutoML têm o potencial de melhorar a precisão do diagnóstico, uma vez que automatizam a seleção de algoritmos e otimização de parâmetros, reduzindo o viés humano e possibilitando uma exploração mais eficiente do espaço de modelos.

Em seguida os modelos foram avaliados por meio de métricas de avaliação, apresentadas na seção 2.2 deste trabalho, ou seja, precisão, revocação e f1-score. O desempenho de cada modelo por meio das métricas será apresentado através de gráficos.



## 5 RESULTADOS

Neste capítulo, são apresentados os resultados obtidos a partir das etapas metodológicas realizadas para atingir os objetivos deste trabalho. As seções a seguir detalham os procedimentos aplicados para a seleção do dataset e preparação dos dados, o pré-processamento, bem como os impactos dessas transformações na análise final, geração e treinamento dos modelos e avaliação dos modelos.

### 5.1 Seleção do Dataset

A base de dados MIT-BIH Arrhythmia foi escolhida para este estudo devido à sua relevância na pesquisa de problemas cardíacos e à ampla adoção na comunidade científica. O MIT-BIH contém registros de sinais ECG adquiridos de pacientes com diferentes condições cardíacas, sendo amplamente utilizados como padrão para treinamento e validação de modelos de aprendizado de máquina.

Além disso, contém uma diversidade de padrões de arritmia, permitindo um treinamento mais robusto do modelo proposto em comparação com outros conjuntos de dados disponíveis. Para isso, foi realizada inicialmente uma análise exploratória para compreender a estrutura do dataset detalhados na 5.1.1.

#### 5.1.1 Dataset MIT-BIH

O banco de dados MIT-BIH, utilizado no desenvolvimento deste trabalho, é um banco de dados fornecido pelo Instituto de Tecnologia de Massachusetts (MIT) e amplamente utilizado em pesquisas de detecção de arritmias cardíacas (Wu *et al.*, 2021a).

O banco contém 48 registros de meia hora de sinais de ECG ambulatorial de dois canais, obtidos de 47 pacientes. As gravações foram digitalizadas a 360 amostras por segundo por canal com resolução de 11 bits em faixa de 10mv. Além disso, cada registro foi anotado de forma independente por dois ou mais cardiologistas, totalizando aproximadamente 110.000 anotações de batimentos cardíacos (Goldberger *et al.*, 2000).

O MIT-BIH contém anotações detalhadas que classificam os batimentos cardíacos em diferentes categorias, totalizando 20 tipos de classificações (Tabela 2). No entanto, os trabalhos que utilizam esse conjunto de dados podem apresentar diferentes quantidades de classes, dependendo dos seus objetivos. Por exemplo, alguns estudos optam por agrupar os

batimentos em cinco categorias principais (Wu *et al.*, 2021b), enquanto outros transformam o problema de classificação multiclasse em binário, diferenciando apenas entre batimentos normais e anormais (Singh *et al.*, 2019).

Neste trabalho, foram consideradas apenas cinco classes principais de batimentos cardíacos: N (Batida normal), S (Batimento prematuro supraventricular), V (Contração ventricular prematura), F (Batida de fusão de ritmo normal e ventricular) e Q (Batida desconhecida). A escolha dessas classes foi motivada por diversos fatores, incluindo sua importância clínica, frequência no dataset e impacto na detecção de arritmias.

A classe N (batida normal) foi incluída para representar o comportamento padrão do coração e servir de referência para a identificação de arritmias. A classe S (batimento prematuro supraventricular) foi escolhida porque representa batimentos que se originam em locais anômalos acima dos ventrículos, sendo frequentemente associados a palpitações e outras condições cardíacas benignas ou patológicas (García, 2013).

Já a classe V (contração ventricular prematura - PVC) foi considerada devido à sua relevância clínica, uma vez que esses batimentos podem ser sinais precoces de arritmias ventriculares mais graves, como taquicardia ventricular e fibrilação ventricular (Hamilton, 2002). A inclusão da classe F (batida de fusão de ritmo normal e ventricular) justifica-se pelo fato de que tais batimentos são indicativos de interações entre ritmos normais e ventriculares, podendo ocorrer em pacientes com marcapasso ou em algumas condições cardíacas particulares.

Por fim, a classe Q (batida desconhecida) foi mantida para representar batimentos que não se encaixam claramente nas demais categorias ou que podem ser resultado de ruído no sinal de ECG. A presença dessa classe permite uma melhor adaptação dos modelos de aprendizado de máquina, garantindo que batimentos não categorizados adequadamente durante a anotação original não sejam simplesmente descartados ou atribuídos erroneamente a outra classe (Chawla *et al.*, 2002).

A escolha dessas cinco classes permitiu um bom equilíbrio entre a complexidade do problema e a identificação dos principais padrões de batimentos cardíacos para o diagnóstico de arritmias. Além disso, ajudou a reduzir o impacto do desequilíbrio entre as classes, já que algumas categorias no MIT-BIH têm muito menos exemplos do que outras, o que poderia prejudicar o desempenho do modelo de classificação.

Tabela 2 – Tabela de anotações de batimentos cardíacos e seus significados

<b>Significado</b>	<b>Anotação de batimento</b>
Batida normal	N
Batimento de bloqueio de ramo esquerdo	L
Batimento de bloqueio de ramo direito	R
Batimento atrial prematuro	A
Batimento atrial prematuro aberrado	a
Batimento prematuro nodal (juncional)	J
Batimento prematuro supraventricular	S
Contração ventricular prematura	V
Fusão de batimento ventricular e normal	F
Início do flutter/fibrilação ventricular	[
Onda de vibração ventricular	!
Fim do flutter/fibrilação ventricular	]
Batimento de escape atrial	e
Batida de escape nodal (juncional)	j
Batida de escape ventricular	E
Batida ritmada	/
Fusão de ritmo normal e ritmo	f
Onda P não conduzida (APB bloqueada)	x
Batida inclassificável	Q
Artefato isolado semelhante a QRS	

Fonte: Elaborado pelo autor (2025).

## 5.2 Pré-processamento dos dados

O pré-processamento dos dados é uma etapa crucial para garantir a qualidade dos dados utilizados no treinamento do modelo. Por esse motivo, uma vez analisados os dados originais e extraídas as características relevantes, como mencionado anteriormente neste trabalho, foi observado que os dados originais consistiam em séries temporais que foram convertidas para um formato tabular, em que cada linha corresponde a uma amostra individual e as colunas representam características obtidas dessas séries.

Após o carregamento dos dados, foi realizada uma inspeção por meio do uso dos comandos *df.info()*, *df.describe()*, *df.head()* que permitiu a identificação das variáveis numéricas e categóricas e uma visão geral da distribuição dos valores das variáveis. Além disso, também foi verificada a existência de valores nulos utilizando *df.isnull().sum()* e foi constatado que o dataset MIT-BIH não possui valores ausentes e, por esse motivo, não foi necessário imputar dados. Para padronizar os dados, utilizou-se a normalização min-max aos sinais ECG, garantindo que todas as amplitudes estivessem no mesmo intervalo. Além disso, foi realizada a conversão

das características categóricas em variáveis numéricas utilizando *one-hot encoding*, tornando os dados adequados para o treinamento do modelo de aprendizado de máquina.

#### Algoritmo 1- Normalização com MinMaxScaler

```
1 from sklearn.preprocessing import MinMaxScaler
2 scaler = MinMaxScaler()
3 df_normalizado = scaler.fit_transform(df)
```

Fonte: Elaborado pelo autor

Além disso, a coluna que continha letras representando diferentes tipos de batimentos cardíacos foi transformada em valores numéricos, visando facilitar o treinamento do modelo (Tabela 3).

Tabela 3 – Tabela de conversão dos tipos de batimento cardíaco para valores numéricos.

Tipo de Batimento	Descrição	Label
N	Batimento normal (Non-ectopic beats)	0
S	Batimento com arritmia tipo 1 (Supraventricular)	1
V	Batimento com arritmia tipo 2 (Ventricular)	2
F	Batimento de fusão (Fusion beats)	3
Q	Batimento desconhecido (Unknown beats)	4

Fonte: Elaborado pelo autor (2025).

O Dataset foi dividido em dois conjuntos: o de treino e o de teste. A divisão adotou uma prática comum para garantir uma validação adequada em que considerou 80% (oitenta por cento) dos dados para treino, que ficou com 87.554 (oitenta e sete mil quinhentos e cinquenta e quatro) amostras e 188 (cento e oitenta e oito) colunas, e 20% (vinte por cento) dos dados para teste, totalizando 21.892 (vinte e um mil oitocentos e noventa e dois) amostras e 188 (cento e oitenta e oito) colunas. Após o carregamento dos dados, a coluna de índice 187, referente aos tipos de batimentos cardíacos, foi transformada para o tipo inteiro e, em seguida, realizada uma amostragem da distribuição das classes no conjunto de treino (Figura 6).

## Algoritmo 2- Visualização de Distribuição com Gráfico de Pizza

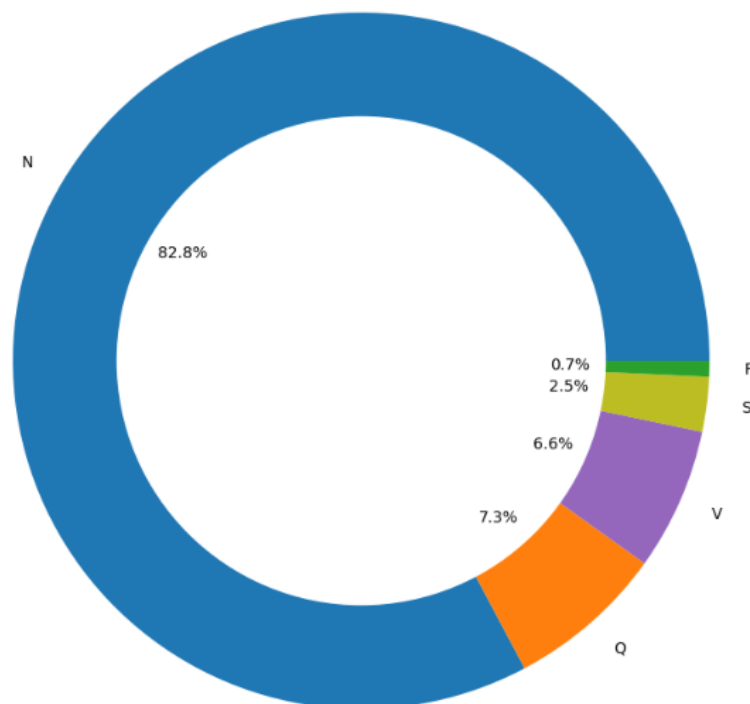
```

1 train_df[187]=train_df[187].astype(int)
2 equilibre=train_df[187].value_counts()
3 print(equilibre)
4
5 plt.figure(figsize=(20,10))
6 my_circle=plt.Circle( (0,0), 0.7, color='white')
7 plt.pie(equilibre, labels=['N','Q','V','S','F'], colors=['
    tab:blue', 'tab:orange', 'tab:purple', 'tab:olive', 'tab
    :green'], autopct='%1.1f%%')
8 p=plt.gcf()
9 p.gca().add_artist(my_circle)
10 plt.show()

```

Fonte: Elaborado pelo autor (2024)

Figura 6 – Distribuição das classes



Fonte: Elaborado pelo autor (2024)

O gráfico da Figura 6 apresenta o balanceamento das classes no conjunto de treino para verificar possíveis desequilíbrios, em que na classe 0 (N) contém 72.471 amostras, a classe 1 contém (S) 2.223 amostras, a 2 (V) 5.788 amostras, a 3 (F) 641 amostras e a 4 (Q) 6.431 amostras. Após a análise, a coluna 187 foi renomeada para 'label' nos conjuntos de treino e teste

para facilitar sua identificação.

#### Algoritmo 3- Renomeação de Coluna

```
1 train_df.rename(columns={187: 'label'}, inplace=True)
2 test_df.rename(columns={187: 'label'}, inplace=True)
```

Fonte: Elaborado pelo autor (2024)

### 5.3 Geração dos Modelos

A construção do modelo preditivo foi realizada utilizando AutoML, permitindo a seleção automática de algoritmos e o ajuste de hiperparâmetros de forma otimizada, sem necessidade de intervenção manual. Essa abordagem garantiu eficiência na busca por modelos de alto desempenho, reduzindo o tempo necessário para experimentação e tuning (Feurer *et al.*, 2015). Para isso foi escolhida a ferramenta de AutoML, o AutoGluon, que testa diferentes algoritmos para resolução de problemas de AM.

Inicialmente foi instalado o AutoGluon e suas dependências e depois os dados foram preparados e iniciado o treinamento do modelo. Nessa etapa, utilizou-se o TabularPredictor que serve para encontrar o melhor modelo para o problema de forma automática. Ele usa diferentes algoritmos e otimiza os hiperparâmetros buscando a obtenção do melhor desempenho possível. O treinamento foi realizado considerando o tempo de 4 horas que corresponde a 14.400 segundos e foi utilizado o parâmetro *best\_quality* para obter o melhor desempenho.

#### Algoritmo 4- Treinamento de Modelo com AutoGluon

```
1 from autogluon.tabular import TabularPredictor as task
2
3 predictor = task(label="label", eval_metric="f1_weighted",
4     problem_type="multiclass").fit(
5         train_data=train_df, time_limit=14400, presets='
6         best_quality'
7     )
8 predictor.fit_summary()
```

Fonte: Elaborado pelo autor (2024)

A Limitação do tempo buscou atingir a eficiência computacional e assim obter um

modelo de alta qualidade. Durante o treinamento, a ferramenta AutoGluon reserva uma parte do conjunto de treinamento para realizar uma validação interna por meio do algoritmo de validação cruzada *cross-validation*, sendo assim uma parte dos dados é reservada a cada interação do treinamento para avaliar o desempenho do modelo.

Durante a execução, múltiplos modelos foram gerados e avaliados. Foram gerados 17 modelos distintos, explorando diferentes abordagens de AM para maximizar a precisão e a robustez das previsões. Dentre os principais, destacam-se o *Random Forest* (RandomForestGini e RandomForest), amplamente utilizado devido à sua robustez e resistência ao overfitting (Breiman, 2001), Extra Trees BAG conhecido por sua capacidade de capturar padrões locais nos dados e melhorar a variabilidade do modelo ao introduzir aleatoriedade adicional na divisão dos nós (Geurts *et al.*, 2006) e o KNeighborsUnif BAG, que se mostrou eficaz na classificação ao considerar a média dos vizinhos mais próximos para suavizar variações locais e reduzir o impacto de outliers (Shao *et al.*, 2019).

## 5.4 Avaliação dos Modelos

A avaliação dos modelos é uma etapa fundamental para garantir sua eficácia na classificação dos sinais de ECG. Para isso, foram utilizadas diferentes métricas de avaliação, incluindo acurácia, precisão, *recall* e *F1-score*, que permitem mensurar a performance dos modelos em diferentes aspectos.

### 5.4.1 Análise dos Resultados

Inicialmente, os resultados de múltiplos modelos de AM foram comparados utilizando diversas métricas de desempenho. Entre elas, destaca-se o *F1\_weighted*, que representa a média harmônica entre precisão (*precision*) e revocação (*recall*), sendo especialmente útil em cenários com classes desbalanceadas. Além disso, foram analisadas a métrica de *recall*, que avalia a capacidade do modelo de identificar corretamente as instâncias positivas, e a acurácia, que mede a proporção total de classificações corretas.

A avaliação considerou um conjunto diversificado de modelos, abrangendo variações de algoritmos baseados em *K-Nearest Neighbors* (KNN), *Random Forest* e *Extra Trees*, além de combinações via *Weighted Ensemble*.

Os resultados baseados na métrica *F1\_weighted* indicam que o *Random Forest* é

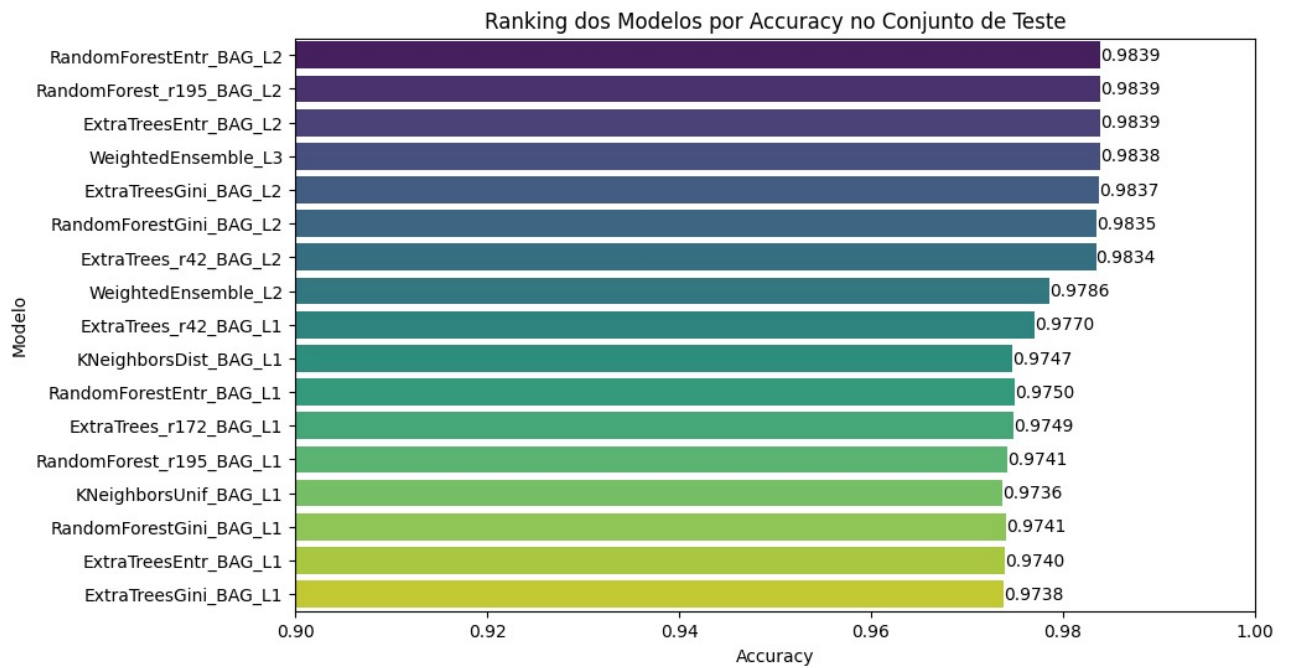
o modelo mais adequado para a tarefa de classificação dos sinais de ECG apresentando um resultado de 0,98% de precisão, uma vez que combina precisão e recall no aprendizado dos padrões dos sinais cardíacos. Esse bom desempenho pode ser explicado por sua capacidade de criar múltiplas árvores de decisão e combinar os resultados, reduzindo o impacto de ruídos nos dados e aumentando a generalização do modelo. Estudos recentes confirmam que o *Random Forest* tem sido amplamente utilizado na classificação de sinais biomédicos, especialmente no reconhecimento de padrões cardíacos (Zhao *et al.*, 2022).

Além disso, os modelos baseados no *Random Forest* mostraram-se mais eficazes do que os baseados em *Extra Trees*, sugerindo que o processo de amostragem controlada das árvores contribui para um desempenho mais consistente (Geurts *et al.*, 2006). Por outro lado, os modelos baseados em KNN apresentaram dificuldades em identificar padrões complexos nos dados, reforçando a necessidade de algoritmos mais sofisticados para esse tipo de problema.

As Figuras 7, 8 e 9 ilustram os resultados obtidos em termos de acurácia, *F1-score* e *recall* para os modelos gerados, respectivamente. Esses indicadores são fundamentais para avaliar a performance preditiva, fornecendo uma visão abrangente sobre a capacidade dos modelos de classificar as amostras. A acurácia reflete a proporção de classificações corretas em relação ao total de previsões, enquanto o *F1-score* equilibra precisão e *recall* e revocação (*recall*) destaca a capacidade do modelo de identificar as instâncias positivas. Dessa forma, a análise desses resultados permite validar a eficiência dos modelos gerados.

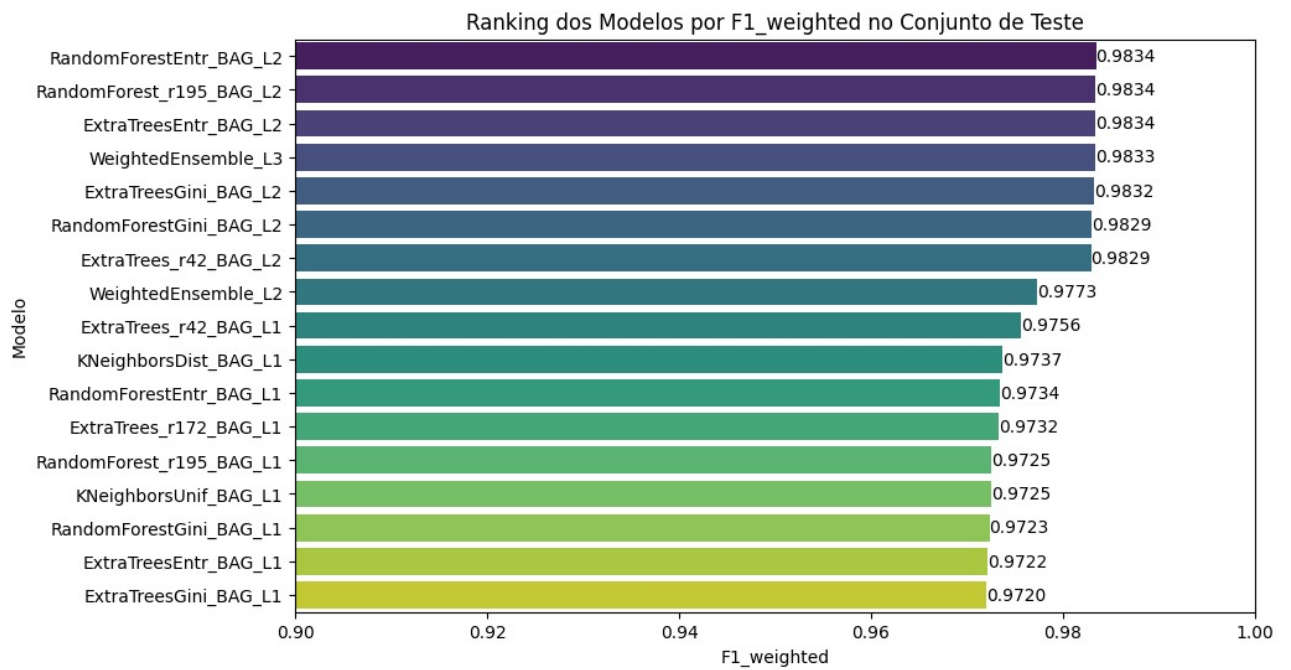


Figura 7 – Resultado da Métrica Acurácia dos modelos Gerados.



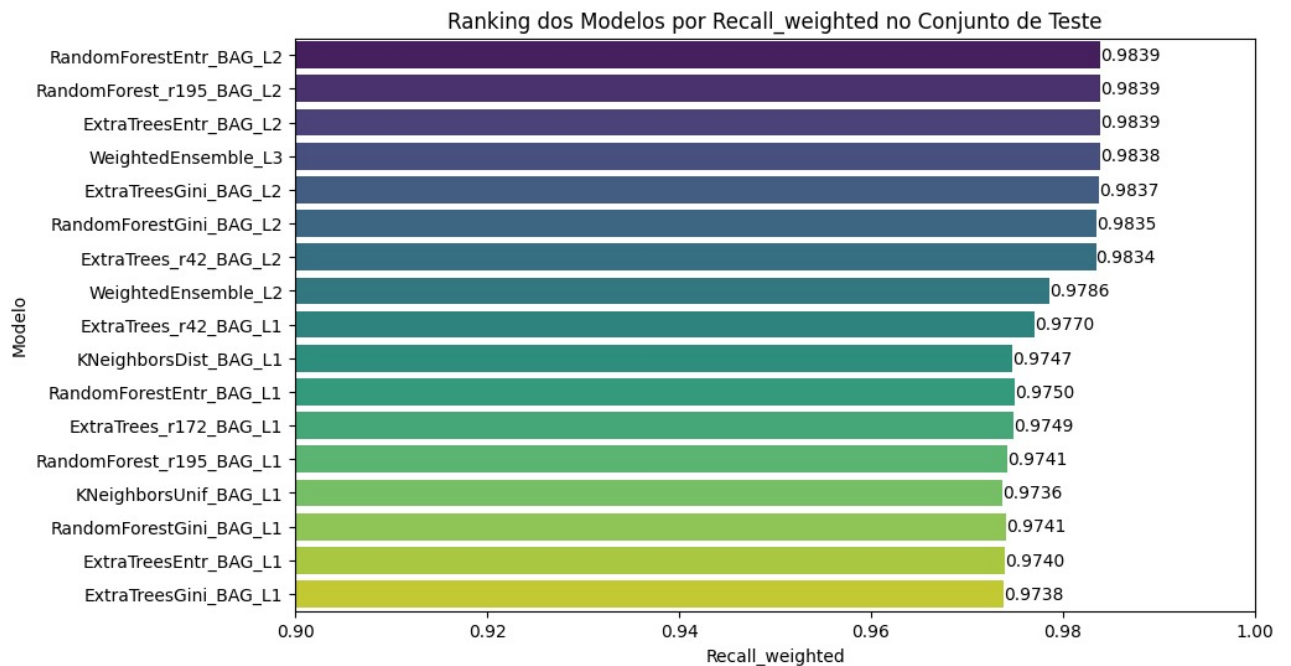
Fonte: Elaborado pelo autor (2025)

Figura 8 – Resultado da Métrica F1 Score dos modelos Gerados.



Fonte: Elaborado pelo autor (2025)

Figura 9 – Resultado da Métrica Recall dos modelos Gerados.



Fonte: Elaborado pelo autor (2025)

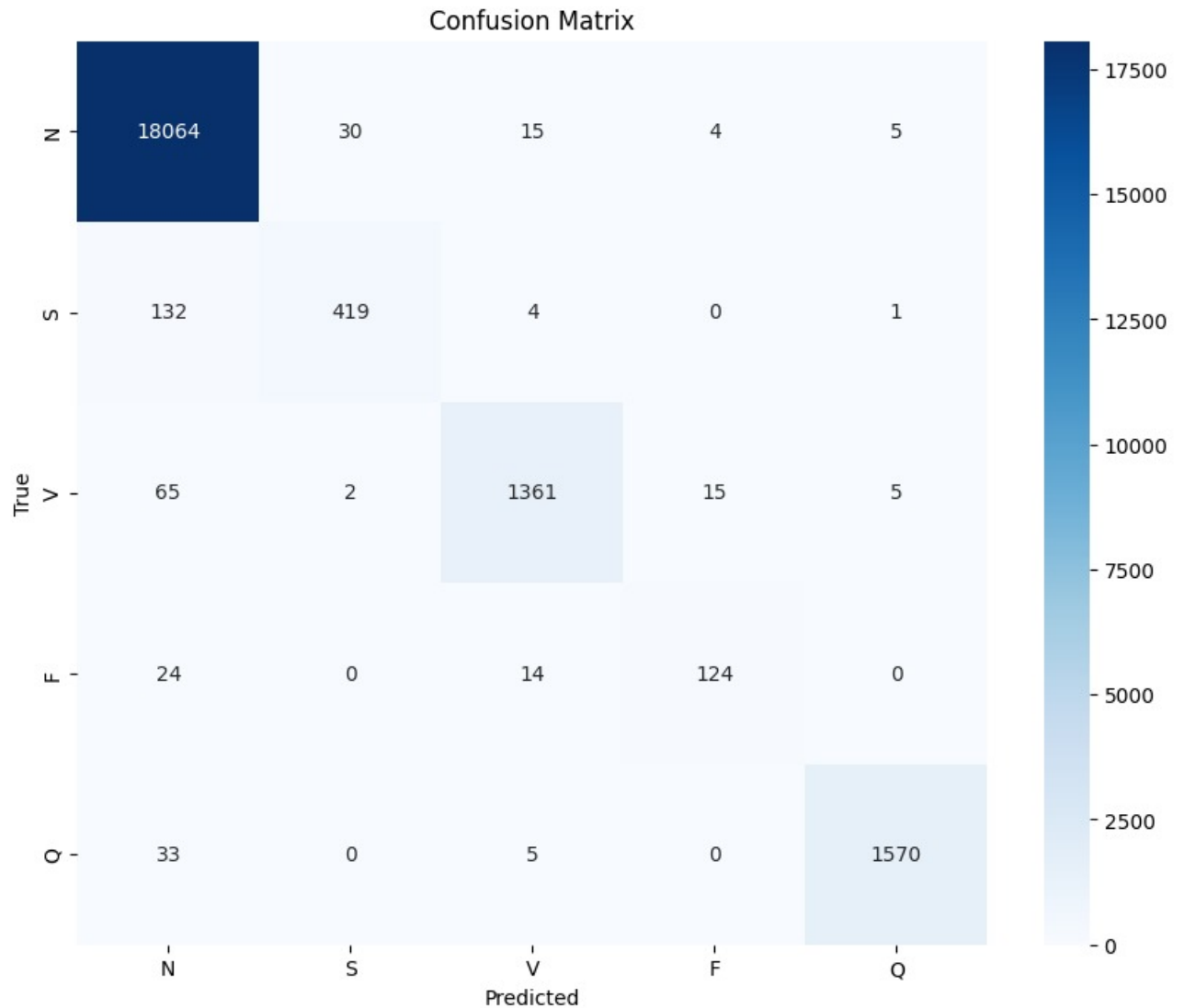
Com base nos resultados, foi possível observar a eficiência do modelo *Random Forest BAG L2* que alcançou valores de 98.39% em acurácia e *Recall* e 98,34% de *F1\_score* se mostrando o modelo mais promissor em relação aos demais. Enquanto que o *KNeighborsUnif BAG* apresentou resultados inferiores em comparação aos outros modelos obtendo uma acurácia de 97.36% e um *F1-score* de 97.25%. Sendo assim, a metodologia empregada foi eficaz e mostrou um modelo robusto e aplicável ao problema estudado.

A Figura 10 apresenta a matriz de confusão do modelo Random Forest BAG L2, que obteve o melhor desempenho geral. A matriz demonstra a distribuição das classificações corretas e incorretas para cada classe do problema. Observa-se que a maioria das previsões foi corretamente atribuída às classes verdadeiras, destacando a alta acurácia do modelo.

Observa-se que a maior parte das previsões foi correta, como no caso da classe N (batimentos normais), onde o modelo acertou 18.064 classificações, errando apenas algumas amostras. Da mesma forma, a classe Q (batimentos desconhecidos) teve um desempenho bastante preciso, com 1.570 acertos e pouquíssimos erros.

No entanto, algumas classes apresentaram mais confusão. Por exemplo, na classe S (batimentos supraventriculares), o modelo fez 419 classificações corretas, mas confundiu 132 amostras como N e 4 como V. Essa confusão pode ser explicada pelo fato de que os batimentos supraventriculares possuem algumas semelhanças morfológicas com os batimentos normais,

Figura 10 – Resultado da Matriz de Confusão.



Fonte: Elaborado pelo autor (2025)

tornando a distinção mais difícil para o modelo.

Outro ponto de atenção é a classe F (batimentos fusionados), que teve 124 acertos, mas algumas amostras foram classificadas incorretamente como V (batimentos ventriculares). Isso sugere que há características compartilhadas entre essas duas categorias, o que pode ter levado o modelo a cometer erros.

Apesar dessas dificuldades, os valores baixos fora da diagonal principal indicam que o modelo teve um bom desempenho geral e foi eficiente na identificação correta dos diferentes tipos de batimentos cardíacos.

### 5.4.2 Comparação do Melhor Modelo com os Modelos dos Trabalhos Relacionados

Para avaliar o modelo Random Forest BAG L2, melhor modelo obtido após a análise dos resultados dos modelos gerados, realizou-se uma comparação entre o desempenho deste modelo e os resultados dos trabalhos relacionados sobre a classificação de arritmias cardíacas por meio do uso de técnicas de AM.

A Tabela 4 apresenta o modelo obtido em cada trabalho, base de dados, quantidade de classes utilizadas na sua construção e um comparativo em relação as métricas de acurácia, *recall* e F1 Score. Ressalta-se que alguns dos trabalhos utilizam outras métricas de classificação, no entanto, para essa análise comparativa foram consideradas apenas as que foram utilizadas no decorrer deste trabalho.

Tabela 4 – Comparação de Desempenho dos Modelos

Trabalho	Base de dados	Modelo	Classes	Acurácia (%)	F1-Score (%)	Recall (%)
(Wu <i>et al.</i> , 2021a)	MIT-BIH	CNN (Rede Neural Convolucional)	5 classes	97.41%	Não possui	Não possui
(Silva <i>et al.</i> , 2022b)	MIT-BIH	Ensemble (Subspace KNN), SVM (Cubic SVM)	5 classes	74.4% (Ensemble)	Não possui	Não possui
(Coutinho; Filho, 2022)	PTB-XL	LSTM (Rede Neural Recorrente)	5 superclasses	77%	52% (F1-Score)	68% (Recall)
<b>Trabalho Proposto</b>	MIT-BIH	Random Forest Bag L2 (AutoML)	5 classes	98.39%	98.34%	98.39%

Fonte: Elaborado pelo autor (2025).

Com base na análise, o trabalho de Wu *et al.* (2021a) alcançou um resultado de 97,41% de acurácia no seu modelo baseado em CNN ao classificar cinco micro-classes de arritmias cardíacas. Enquanto que o trabalho de Silva *et al.* (2022b) gerou dois modelos finais, o *Ensemble (Subspace KNN)* e *SVM (Cubic SVM)* utilizando cinco classes de arritmias cardíacas. O Ensemble obteve um resultado melhor que o outro modelo gerado alcançando uma acurácia de 74,4%, no entanto, seu desempenho foi inferior quando comparado ao modelo obtido neste trabalho que demonstrou maior robustez com uma acurácia de 98,34%, significativamente

superior.

O Trabalho de Coutinho e Filho (2022) exploram o uso de redes neurais recorrentes (LSTM) na classificação de arritmias, alcançando uma acurácia de 77%, F1-Score de 52% e recall de 68%, sendo o único que apresenta todas as métricas de avaliação utilizadas neste trabalho. No entanto, todos os valores obtidos foram inferiores ao do presente trabalho. Acredita-se que a utilização do AutoML tenha contribuído com uma melhor adaptação e um desempenho superior no resultado das métricas.

Dessa forma, o modelo Random Forest Bag L2, obtido no desenvolvimento do presente trabalho utilizando AutoML, obteve uma acurácia de 98,34%, um F1-Score de 98,3% e um recall de 98%, destacando-se entre os modelos analisados, especialmente em termos de precisão e capacidade na identificação das arritmias cardíacas.

## 6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo desenvolver um modelo de aprendizado de máquina capaz de classificar diferentes tipos de batimentos cardíacos a partir de sinais eletrocardiográficos. Para isso, foi utilizada uma abordagem baseada em AutoML, permitindo a seleção automática de hiperparâmetros e modelos mais adequados ao problema.

A base de dados utilizada contém diversas classes de batimentos cardíacos, porém, neste estudo, foram consideradas cinco classes principais, representando diferentes tipos de arritmias e batimentos normais. Esse recorte foi feito para garantir uma análise mais precisa e alinhada com os objetivos do trabalho.

Durante a fase experimental, foram gerados diversos modelos de classificação, entre eles Random Forest, K-Nearest Neighbors (KNN) e Extra Trees, integrados em um conjunto de técnicas de ensemble learning. Após a avaliação dos resultados, o Random Forest BAG L2 se destacou como o modelo de melhor desempenho, atingindo uma acurácia de 98,39% e um F1-score de 98,34%, demonstrando sua robustez na classificação dos batimentos cardíacos.

Ademais, a utilização de AutoML se mostrou promissora neste trabalho, por automatizar o processo de seleção de algoritmos até a otimização dos parâmetros, facilitando a obtenção de resultados mais rápidos. Uma das vantagens identificadas é sua capacidade de ajustar automaticamente e gerar melhores modelos o que pode acelerar a detecção de DCV e contribuir em diagnósticos mais precisos.

Ao longo do desenvolvimento, algumas dificuldades foram enfrentadas, principalmente relacionadas ao desbalanceamento das classes e à diferenciação de batimentos com características semelhantes, o que gerou certa confusão em algumas previsões. A utilização de técnicas de normalização e ajuste de hiperparâmetros ajudou a mitigar esses desafios, permitindo um desempenho mais equilibrado do modelo.

A aplicação de técnicas de AutoML, com o uso do AutoGluon, foi o diferencial para alcançar um bom desempenho na geração do modelo. Esse resultado mostra o potencial de ferramentas automatizadas na classificação de sinais de eletrocardiogramas bem como o uso desse tipo de técnica pode auxiliar médicos cardiologistas no diagnóstico mais preciso de doenças cardiovasculares. Além disso, embora a técnica utilizada seja automatizada ainda requer uma análise crítica sobre a escolha dos parâmetros certos. Apesar do bom desempenho nos resultados, o modelo pode ser aprimorado ao considerar diferentes abordagens de pré-processamento, como o uso de outros tipos de normalização e, posteriormente, a introdução de novas classes de

arritmias.

Como trabalhos futuros, sugere-se a ampliação da base de dados com novas classes de batimentos e a aplicação de técnicas de data augmentation podem contribuir para melhorar ainda mais a precisão dos modelos. Além disso pode-se explorar de redes neurais profundas, como redes convolucionais (CNNs) e redes recorrentes (RNNs), que podem capturar padrões temporais mais complexos nos sinais cardíacos.

## REFERÊNCIAS

- ACHARYA, U. R. *et al.* Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ecg signals: A comparative study. **Information Sciences**, [S.l.], v. 377, p. 17–29, 2017.
- ALPAYDIN, E. **Machine learning**. [S. l.]: Mit Press, 2021.
- ALSHAREF, A.; AGGARWAL, K.; SONIA; KUMAR, M.; MISHRA, A. Review of ml and autotml solutions to forecast time-series data. **Archives of Computational Methods in Engineering**, Springer, [S.l.], v. 29, n. 7, p. 5297–5311, 2022.
- ARAÚJO, M. V. A. de. Métodos de clustering em aprendizado de máquinas não supervisionado. **Revista de Inteligência Artificial e Machine Learning**, Editora Acadêmica Brasileira, [S.l.], v. 10, n. 2, p. 123–134, 2024. Disponível em: <https://revml.org/artigo/2024-clustering>. Acesso em: 8 jun 2024.
- BERKAYA, S. K. *et al.* A survey on ecg analysis: techniques, applications, and challenges. **Biomedical Signal Processing and Control**, [S.l.], v. 43, p. 216–235, 2018.
- BRAGA, A. V.; LINS, A. F.; SOARES, L. S.; FLEURY, L. G.; CARVALHO, J. C.; PRADO, R. S. do. Machine learning: o uso da inteligência artificial na medicina. **Brazilian Journal of Development**, [S.l.], v. 5, n. 9, p. 16407–16413, 2019.
- BREIMAN, L. Random forests. **Machine Learning**, [S.l.], v. 45, p. 5–32, 2001.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: Synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, [S.l.], v. 16, p. 321–357, 2002.
- CHEN, W. *et al.* A convolutional attention-based neural network for ecg arrhythmia classification. **Physiological Measurement**, [S.l.], v. 40, n. 6, p. 065002, 2019.
- COSTA, A. d. C.; FRANCO, A. C. L.; PEREIRA RAFAEL FERREIRA E GONÇALVES, M. V. V. e. N. E. S. A. Arritmias cardíacas: Diagnóstico, tratamento e prevenção. **Brazilian Journal of Implantology and Health Sciences**, [S.l.], v. 6, n. 2, p. 348–360, 2024. Disponível em: <https://bjih.s.emnuvens.com.br/bjih/article/view/1374>. Acesso em: 15 nov 2023.
- COUTINHO, E. F.; FILHO, F. R. da S. Aprendizado de máquina para predição de diagnósticos de doenças cardiovasculares. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA à SAÚDE, 22., Porto Alegre. **Anais [...]**. Porto Alegre, RS: SBC, 2022. p. 358–369. ISSN 2763-8952. Disponível em: <https://sol.sbc.org.br/index.php/sbcas/article/view/21646>. Acesso em: 25 nov. 2023.
- ERICKSON, N.; MUELLER, J. W.; SHIRKOV, A.; ZHANG, H.; LARROY, P.; LI, M.; SMOLA, A. Autogluon-tabular: Robust and accurate autotml for structured data. **ArXiv**, abs/2003.06505, 2020. Disponível em: <https://api.semanticscholar.org/CorpusID:212725762>. Acesso em: 25 out 2024.
- FEURER, M.; KLEIN, A.; EGGENSPERGER, K.; SPRINGENBERG, J.; BLUM, M.; HUTTER, F. Efficient and robust automated machine learning. **Advances in neural information processing systems**, [S.l.], v. 28, 2015.



FEURER, M.; KLEIN, A.; EGGENSPERGER, K.; SPRINGENBERG, J. T.; BLUM, M.; HUTTER, F. Auto-sklearn efficient and robust automated machine learning: Methods, systems, challenges. In: HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. (Ed.). **Automated Machine Learning**. Cham: Springer International Publishing, 2019. p. 113–134. ISBN 978-3-030-05318-5. Disponível em: [https://doi.org/10.1007/978-3-030-05318-5\\_6](https://doi.org/10.1007/978-3-030-05318-5_6). Acesso em: 17 jun 2024.

GARCÍA, T. B. **12-Lead ECG: The art of interpretation**. 2. ed. Burlington, MA, USA: Jones & Bartlett Learning, 2013.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine Learning**, Springer, [S.l.], v. 63, n. 1, p. 3–42, 2006.

GIANNAKAKIS, G.; GRIGORIADIS, D.; GIANNAKAKI, K.; SIMANTIRAKI, O.; RONIOTIS, A.; TSIKNAKIS, M. Review on psychological stress detection using biosignals. **IEEE Transactions on Affective Computing**, IEEE, v. 13, n. 1, p. 440–460, 2019.

GODWASSERAND, P. 119 eletrocardiograma normal e patológico. **Periodico da EMSM**, [S.l.], v1, n. 2, p. 1–25, 2019.

GOLDBERGER, A.; AMARAL, L.; GLASS, L.; HAUSDORFF, J.; IVANOV, P. C.; MARK, R.; STANLEY, H. E. Physiobank, physiotoolkit e physionet: Componentes de um novo recurso de pesquisa para sinais fisiológicos complexos. **Circulation**, [S.l.], v. 101, n. 23, p. e215–e220, 2000. Disponível em: <https://physionet.org/content/mitdb/1.0.0/>. Acesso em: 25 nov 2023.

HAMILTON, P. S. Open source ecg analysis. In: **Computers in Cardiology**. [S. l.: s. n.], 2002. p. 101–104.

HAN, J.; PEI, J.; TONG, H. **Data mining: concepts and techniques**. [S. l.]: Morgan kaufmann, 2022.

HE, X.; ZHAO, K.; CHU, X. Automl: A survey of the state-of-the-art. **Knowledge-Based Systems**, v. 212, p. 106622, 2021. ISSN 0950-7051. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950705120307516>. Acesso em: 10 maio 2024.

HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. (Ed.). **Automatic machine learning: methods, systems, challenges**. Germany: Springer, 2019. (Challenges in Machine Learning). ISBN 978-3-030-05317-8.

INAN, O. T. *et al.* Wearable ambulatory cardiovascular monitoring: present and future. **IEEE Reviews in Biomedical Engineering**, [S.l.], v. 13, p. 187–198, 2020.

LIU, T.; SIEGEL, E.; SHEN, D. Deep learning and medical image analysis for covid-19 diagnosis and prediction. **Annual review of biomedical engineering**, Annual Reviews, [S.l.], v. 24, n. 1, p. 179–201, 2022.

LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. **Estudos Avançados**, SciELO Brasil, [S.l.], v. 35, p. 85–94, 2021.

LUZ, E. J. da S.; SCHWARTZ, W. R.; CÁMARA-CHÁVEZ, G.; MENOTTI, D. Ecg-based heartbeat classification for arrhythmia detection: A survey. **Computer Methods and Programs in Biomedicine**, [S.l.], v. 127, p. 144–164, 2016. ISSN 0169-2607. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169260715003314>. Acesso em: 27 nov 2023.

MAHESH, B. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR)**, [S.l.], v. 9, n. 1, p. 381–386, 2020.

MARIANO, D. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e f-score. **BIOINFO – Revista Brasileira de Bioinformática**, [S. l.], 2021.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, [S.l.], v. 1, n. 1, p. 32, 2003.

NAGARAJAH, T.; PORAVI, G. A review on automated machine learning (automl) systems. In: **2019 IEEE 5th International Conference for Convergence in Technology (I2CT)**. [S. l.: s. n.], 2019. p. 1–6.

RANGAYYAN, R. M. **Biomedical Signal Analysis::** A case-study approach. [S. l.]: John Wiley Sons, 2015.

RIBEIRO, A. G.; COTTA, R. M. M.; RIBEIRO, S. M. R. A promoção da saúde e a prevenção integrada dos fatores de risco para doenças cardiovasculares. **Ciência & Saúde Coletiva**, SciELO Public Health, [S.l.], v. 17, p. 7–17, 2012.

SABA, T.; AHMAD, I.; NAEEM, K.; USMAN, M.; KORA, R.; MOHAMMAD, A. L. Automated diagnosis of arrhythmias in electrocardiogram signals using machine learning algorithms. **Healthcare**, [S.l.], v. 6, n. 4, p. 103, 2018.

SAÚDE, Organização Pan-Americana de . **Causas de Doenças Cardíacas**. 2017. Disponível em: <https://www.paho.org/pt/topicos/doencas-cardiovasculares#:~:text=A%20causa%20de%20ataques%20card%C3%ADacos,%2C%20hipertens%C3%A3o%2C%20diabetes%20e%20hiperlipidemia>. Acesso em: 07 out. 2023.

SCHNEIDER, C. F. Machine learning aplicado na previsão de resultados de partidas de futebol: um estudo de caso para comparação de diferentes classificadores. **Revista Brasileira de Análise de Dados e Inteligência Artificial**, Editora Acadêmica Brasileira, [S.l.], v. 4, n. 1, p. 45–57, 2018. Disponível em: <https://revbadai.org/artigo/2018-machine-learning>. Acesso em: 20 fev 2025.

SCHWARZ, L. Artigo de revisão: Eletrocardiograma. **Revista Ilha Digital**, [S.l.], v. 1, p. 3–19, dez. 2009. Disponível em: <https://ilhadigital.florianopolis.ifsc.edu.br/index.php/ilhadigital/article/view/4>. Acesso em: 3 dez 2023.

SHAO, Z.; XIA, T.; ZHANG, W.; Lü, J. An efficient weighted k-nearest neighbor algorithm based on incomplete distance matrix. **Neurocomputing**, Elsevier, [S.l.], v. 365, p. 51–60, 2019.

SHEN, X. *et al.* Automated machine learning for arrhythmia detection. **IEEE Transactions on Biomedical Engineering**, [S.l.], v. 67, n. 12, p. 3451–3462, 2020.

SHENDE, M. K.; FEIJÓO-LORENZO, A. E.; BOKDE, N. D. cleants: Automated (automl) tool to clean univariate time series at microscale. **Neurocomputing**, [S.l.], v. 500, p. 155–176, 2022. ISSN 0925-2312. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231222006117>. Acesso em: 25 nov. 2024.

SILVA, C. N.; MATOS, J. A.; LOPES, F. F. **Aquisição de eletrocardiograma e classificação de arritmias por machine learning**. Trabalho de Conclusão de Curso Engenharia Elétrica – Centro Universitário FEI, São Bernardo do Campo, 2022. Disponível em:

<https://repositorio.fei.edu.br/items/6f52c290-c638-486f-964d-c4a5fc1a0f1b>. Acesso em: 02 abr. 2024.

SILVA, J. *et al.* Análise de sinais biomédicos: Técnicas e aplicações. **Revista de Engenharia Biomédica**, v. 15, n. 2, p. 45–60, 2022.

SINGH, V.; TEWARY, S.; SARDANA, V.; SARDANA, H. K. Arrhythmia detection—a machine learning based comparative analysis with mit-bih ecg data. In: **2019 IEEE 5th International Conference for Convergence in Technology (I2CT)**. Bombay, India: IEEE, 2019. p. 1–5.

SORNMO, L.; LAGUNA, P. **Bioelectrical Signal Processing in Cardiac and Neurological Applications**. [S. l.]: Academic Press, 2005.

SOUSA, L. A. N. de. **Aquisição e processamento do eletrocardiograma para auxílio ao diagnóstico de arritmias cardíacas**. 64 f. Monografia (Graduação em Engenharia Elétrica) – Universidade Federal Rural do Semi-Árido, Mossoró, 2018.

STATION, R. **Conceitos Machine Learning**. 2023. Disponível em: <https://resultadosdigitais.com.br/marketing/machine-learning/>. Acesso em: 09 out 2023.

WANG, Y. *et al.* Autotml-based deep learning for ecg classification. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 25, n. 7, p. 2348–2356, 2021.

WU, M.; LU, Y.; YANG, W.; WONG, S. Y. A study on arrhythmia via ecg signal classification using the convolutional neural network. **Frontiers in Computational Neuroscience**, [S.l.], v. 14, p. 564015, 2021. ISSN 1662-5188. Disponível em: <https://www.frontiersin.org/articles/10.3389/fncom.2020.564015>. Acesso em: 02 mar. 2024.

WU, Y.; YANG, F.; WANG, Q.; XU, Z.; ZHOU, B.; DENG, X. A comparison of different deep learning models for arrhythmia classification using ecg signals. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 25, n. 7, p. 1876–1887, 2021. Acesso em: 02 abr. 2024.

ZHAO, Z.; DENG, Y.; ZHANG, M.; WANG, X. Ecg classification using improved random forest algorithm. **Biomedical Signal Processing and Control**, [S.l.], v. 71, p. 103076, 2022.