



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS DE SOBRAL**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO**

**LUCAS JOSÉ LEMOS BRAZ**

**CLASSIFICAÇÃO DE FIBRILAÇÃO ATRIAL EM ECG COM MÉTRICAS DE  
CONVERGÊNCIA DA DECOMPOSIÇÃO TENSORIAL BTD**

**SOBRAL**

**2025**

LUCAS JOSÉ LEMOS BRAZ

CLASSIFICAÇÃO DE FIBRILAÇÃO ATRIAL EM ECG COM MÉTRICAS DE  
CONVERGÊNCIA DA DECOMPOSIÇÃO TENSORIAL BTD

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do Campus Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Carlos Alexandre Rolim Fernandes.

SOBRAL

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)

---

B839c Braz, Lucas José Lemos.

Classificação de Fibrilação Atrial em ECG com Métricas de Convergência da  
Decomposição Tensorial BTD / Lucas José Lemos Braz. – 2025.  
70 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus  
de Sobral, Curso de Engenharia da Computação, Sobral, 2025.

Orientação: Prof. Dr. Carlos Alexandre Rolim Fernandes.

1. Decomposição Tensorial. 2. Eletrocardiograma. 3. Fibrilação Atrial. 4. Aprendizado de  
Máquina. 5. Diagnóstico Automatizado. I. Título.

CDD 621.39

---

LUCAS JOSÉ LEMOS BRAZ

CLASSIFICAÇÃO DE FIBRILAÇÃO ATRIAL EM ECG COM MÉTRICAS DE  
CONVERGÊNCIA DA DECOMPOSIÇÃO TENSORIAL BTD

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do Campus Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Aprovada em: 06 de Março de 2025

BANCA EXAMINADORA

---

Prof. Dr. Carlos Alexandre Rolim  
Fernandes (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dra. Jermana Lopes de Moraes  
Universidade Federal do Ceará (UFC)

---

Pesquisador Dr. Pedro Marinho Ramos de Oliveira  
BIOSERENITY

À minha família, por sua confiança e apoio incondicional, que me permitiram conquistar meus objetivos mesmo nos momentos de dúvida. Mãe, sua dedicação e carinho me lembraram, a cada passo, da importância de seguir em frente. Pai, sua estabilidade e sabedoria foram fundamentais para me manter focado e determinado.

## **AGRADECIMENTOS**

Este trabalho só foi possível graças ao ensino público de qualidade, representado por minha universidade e pelos professores e profissionais que, ao longo dos anos, contribuíram para minha formação acadêmica. Sou profundamente grato pelo conhecimento compartilhado e pelo ambiente de aprendizado que me permitiu crescer como estudante, pesquisador e pessoa.

Ao Prof. Dr. Carlos Alexandre, pela orientação, pelo rigor acadêmico e clareza, e por abrir caminhos fundamentais em minha trajetória científica.

Ao Dr. Pedro Marinho, pelo apoio a esta pesquisa, trazendo importantes discussões e sugestões.

À Prof. Dra. Jermana Lopes, pelos conselhos valiosos e pelo suporte em minha jornada acadêmica.

Aos meus pais, Benedito e Fatinha, minha base inabalável. Pelo amor incondicional, pela paciência nos momentos difíceis e por sempre acreditarem em mim, mesmo quando eu duvidava. Tudo o que conquistei até aqui é, na maioria, reflexo do apoio de vocês.

À minha irmã e ao meu irmão, por me escutarem e apoiarem, sou imensamente grato.

## RESUMO

Em meio ao crescente aumento da incidência de doenças cardiovasculares, as arritmias desempenham um papel central na morbimortalidade global. Dentre elas, a Fibrilação Atrial (FA) destaca-se como a arritmia sustentada mais comum, estando associada a um risco elevado de complicações como acidente vascular cerebral e insuficiência cardíaca. Diante desse cenário, torna-se essencial o desenvolvimento de ferramentas que auxiliem no diagnóstico precoce e preciso dessa condição. Neste contexto, este trabalho investiga e expande a aplicabilidade das métricas de convergência derivadas da decomposição tensorial *Block-Term Decomposition* (BTD) em sinais de Eletrocardiograma (ECG) para a classificação da FA. Em um contexto em que a acurácia e a rapidez diagnóstica são imperativas, a proposta deste estudo consiste em extrair atributos a partir dos indicadores de convergência da BTD e combiná-los com diversos modelos de Aprendizado de Máquina (AM). A metodologia adotada envolveu a análise comparativa de diversas combinações de atributos e a implementação de modelos que variaram desde abordagens tradicionais até arquiteturas avançadas baseadas em *Recurrent Neural Network* (RNN). Os resultados experimentais demonstraram que tanto os modelos convencionais quanto os modelos baseados em RNN, ao empregarem os atributos extraídos, alcançaram acurácias entorno a 95% e F1-Score de 96,50% corroborando a robustez e a eficácia dos indicadores propostos para a distinção entre FA e Ritmo Sinusal Normal (RSN). Adicionalmente, o estudo ressalta a relevância da integração entre técnicas de decomposição tensorial e algoritmos de AM para o aprimoramento dos métodos diagnósticos em cardiologia. Por fim, conclui-se que os achados apresentados incentivam a realização de investigações futuras voltadas à otimização dos vetores de atributos e ao refinamento dos parâmetros dos modelos de aprendizado profundo, bem como à avaliação da eficiência computacional das abordagens propostas, contribuindo significativamente para o avanço do diagnóstico automatizado de condições cardíacas.

**Palavras-chave:** Decomposição Tensorial; Eletrocardiograma; Fibrilação Atrial; Aprendizado de Máquina; Redes Neurais Recorrentes; Diagnóstico Automatizado.

## ABSTRACT

Amid the growing incidence of cardiovascular diseases, arrhythmias play a central role in global morbidity and mortality. Among them, atrial fibrillation stands out as the most common sustained arrhythmia and is associated with a high risk of complications such as stroke and heart failure. Given this scenario, it is essential to develop tools that help in the early and accurate diagnosis of this condition. In this context, this work investigates and expands the applicability of convergence metrics derived from the tensor decomposition BTD in ECG signals for the classification of atrial fibrillation. In a context in which diagnostic accuracy and speed are imperative, the purpose of this study is to extract attributes from the BTD convergence indicators and combine them with various machine learning models. The methodology adopted involved the comparative analysis of various combinations of attributes and the implementation of models ranging from traditional approaches to advanced architectures based on RNN. The experimental results showed that both the conventional models and the models based on RNN, when using the extracted attributes, achieved accuracies of around 95% and F1-Score of 96.50%, corroborating the robustness and effectiveness of the proposed indicators for distinguishing between atrial fibrillation and normal sinus rhythm. In addition, the study highlights the importance of integrating tensor decomposition techniques and machine learning algorithms to improve diagnostic methods in cardiology. Finally, it is concluded that the findings presented encourage future research aimed at optimizing attribute vectors and refining the parameters of deep learning models, as well as evaluating the computational efficiency of the proposed approaches, contributing significantly to the advancement of automated diagnosis of cardiac conditions.

**Keywords:** Tensor Decomposition; Electrocardiogram; Atrial Fibrillation; Machine Learning; Recurrent Neural Networks; Automated Diagnosis.

## LISTA DE TABELAS

Tabela 1 – Distribuição de segmentos (amostras) das bases de dados . . . . .	39
Tabela 2 – Especificações da função <i>lll_rnd</i> (Tensorlab). . . . .	40
Tabela 3 – Especificações da função <i>lll_nls</i> (Tensorlab). . . . .	40
Tabela 4 – Intervalo de busca dos hiperparâmetros para as RNNs . . . . .	44
Tabela 5 – Resultados do Teste U de Mann-Whitney dos Atributos Comparados . . . . .	47
Tabela 6 – Resultados usando indicadores BTD . . . . .	50
Tabela 7 – Resultados usando histograma do <i>R-R Interval</i> (RRI) . . . . .	51
Tabela 8 – Resultados usando indicadores BTD e histograma do RRI . . . . .	51
Tabela 9 – Intervalo de busca e parâmetros ajustados para os modelos . . . . .	53
Tabela 10 – Resultados usando o vetor do indicador <i>fval</i> da BTD . . . . .	54
Tabela 11 – Resultados usando os vetores dos indicadores <i>fval</i> e <i>delta</i> da BTD . . . . .	55
Tabela 12 – Resultados usando os vetores dos indicadores <i>fval</i> , <i>delta</i> e <i>relerr</i> da BTD . . . . .	55
Tabela 13 – Resultados usando o vetor do indicador <i>fval</i> da BTD e histograma dos RRIs . . . . .	56
Tabela 14 – Comparativo com resultados da literatura . . . . .	58

## LISTA DE ABREVIATURAS E SIGLAS

AA	Atividade Atrial
ACC	Acurácia
AFDB	<i>Atrial Fibrillation Database</i>
ALS	<i>Alternating Least Squares</i>
AM	Aprendizado de Máquina
AV	Atrioventricular
AVC	Acidente Vascular Cerebral
BI-LSTM	<i>Bidirectional LSTM</i>
BSS	<i>Blind Source Separation</i>
BTD	<i>Block-Term Decomposition</i>
DT	<i>Decision Tree</i>
ECG	Eletrocardiograma
FA	Fibrilação Atrial
FN	Falso Negativo
FP	Falso Positivo
GRU	<i>Gated Recurrent Unit</i>
KNN	<i>K-Nearest Neighbors</i>
LightGBM	<i>Light Gradient-Boosting Machine</i>
LSTM	<i>Long Short-Term Memory</i>
LTAfDB	<i>Long-term Atrial Fibrillation Database</i>
MLP	<i>Multilayer Perceptron</i>
NLS	<i>Nonlinear Least Squares</i>
NSRDB	<i>Normal Sinus Rhythm Database</i>
PCA	<i>Principal Component Analysis</i>
PRE	Precisão
RF	<i>Random Forest</i>
RNN	<i>Recurrent Neural Network</i>
RRI	<i>R-R Interval</i>
RSN	Ritmo Sinusal Normal
SE	Sensibilidade

SPE	Especificidade
SVM	<i>Support Vector Machine</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
XGBoost	<i>Extreme Gradient Boosting</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>2</b>	<b>OBJETIVOS</b>	<b>15</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>16</b>
<b>3.1</b>	<b>Eletrocardiograma como ferramenta de diagnóstico</b>	<b>16</b>
<b>3.2</b>	<b>Fibrilação atrial</b>	<b>17</b>
<b>3.2.1</b>	<i>Morfologia de um sinal de ECG com Fibrilação Atrial</i>	<b>18</b>
<b>3.3</b>	<i>Block-Term Decomposition</i>	<b>20</b>
<b>3.4</b>	<b>Hankel-BTD</b>	<b>22</b>
<b>3.5</b>	<b>Classificadores</b>	<b>25</b>
<b>3.5.1</b>	<i>Métodos Tradicionais de Aprendizado de Máquina</i>	<b>25</b>
<b>3.5.1.1</b>	<i>Árvore de Decisão</i>	<b>25</b>
<b>3.5.1.2</b>	<i>Random Forest e Extra Trees</i>	<b>26</b>
<b>3.5.1.3</b>	<i>XGBoost, LightGBM e Gradient Boosting</i>	<b>27</b>
<b>3.5.1.4</b>	<i>Modelos de Classificação Simples: KNN, Naive Bayes e Regressão Logística</i>	<b>28</b>
<b>3.5.2</b>	<i>Redes Neurais Recorrentes e Extensões</i>	<b>29</b>
<b>3.5.2.1</b>	<i>Arquitetura Básica das RNNs</i>	<b>29</b>
<b>3.5.2.2</b>	<i>Long Short-Term Memory</i>	<b>30</b>
<b>3.5.2.3</b>	<i>Gated Recurrent Unit</i>	<b>32</b>
<b>3.5.2.4</b>	<i>Bidirectional LSTM</i>	<b>33</b>
<b>3.5.3</b>	<i>Métricas de Desempenho</i>	<b>33</b>
<b>4</b>	<b>TRABALHOS RELACIONADOS</b>	<b>35</b>
<b>5</b>	<b>METODOLOGIA</b>	<b>37</b>
<b>5.1</b>	<b>Base de dados</b>	<b>38</b>
<b>5.2</b>	<b>Pré-processamento</b>	<b>38</b>
<b>5.3</b>	<b>Extração de atributos utilizando BTD</b>	<b>39</b>
<b>5.4</b>	<b>Randomização e divisão treino, validação e teste</b>	<b>42</b>
<b>5.5</b>	<b>Classificação e modelos</b>	<b>43</b>
<b>5.5.1</b>	<i>Modelos tradicionais de aprendizado de máquina</i>	<b>43</b>
<b>5.5.2</b>	<i>Arquitetura das RNNs</i>	<b>44</b>
<b>6</b>	<b>RESULTADOS</b>	<b>47</b>

6.1	<b>Análise dos atributos</b> . . . . .	47
6.2	<b>Resultado com valores escalares</b> . . . . .	49
6.2.1	<i>Ajuste de Hiperparâmetros do Melhor Caso</i> . . . . .	51
6.3	<b>Resultados com Valores Vetoriais</b> . . . . .	52
7	<b>DISCUSSÃO</b> . . . . .	57
7.1	<b>Comparativo com a Literatura</b> . . . . .	57
7.2	<b>O Desempenho com Indicadores da BTD</b> . . . . .	59
7.3	<b>Complexidade computacional do processo</b> . . . . .	59
8	<b>CONCLUSÕES E TRABALHOS FUTUROS</b> . . . . .	62
	<b>REFERÊNCIAS</b> . . . . .	64

## 1 INTRODUÇÃO

As doenças cardiovasculares estão entre as principais causas de mortalidade no Brasil e continuam apresentando uma tendência alarmante de crescimento (PUENTES *et al.*, 2023). Muitas dessas doenças surgem de irregularidades na condução dos impulsos elétricos do coração, conhecidas como arritmias, que podem aumentar significativamente o risco de Acidente Vascular Cerebral (AVC) e insuficiência cardíaca (MAGALHÃES *et al.*, 2016). Dentre elas, a FA é a mais comum, caracterizada por contrações descoordenadas nos átrios devido à geração anormal de impulsos elétricos. Fatores como envelhecimento populacional e maiores taxas de insuficiência cardíaca estão associados com o aumento de casos de FA (DELANEY *et al.*, 2018).

A automação e aprimoramento do diagnóstico de FA são importantes para reduzir a carga de trabalho dos especialistas, acelerar o tempo de resposta de diagnóstico e tratamento e melhorar a acessibilidade em ambientes com recursos limitados. Ademais, prover um diagnóstico precoce também é de extrema importância, uma vez que o tratamento antecipado limita e até impede o avanço da FA (NATTEL *et al.*, 2014). Essas necessidades ganham ainda mais relevância em um contexto de crescimento e envelhecimento populacional, onde há maior incidência de doenças crônicas e cardiovasculares (SCHNABEL *et al.*, 2015).

O diagnóstico para FA é geralmente baseado em sintomas como dor no peito e pulso irregular, além de uma análise visual dos sinais de ECG por especialistas (MAGALHÃES *et al.*, 2016). Nos últimos anos, métodos computacionais, especialmente técnicas de AM, têm demonstrado grande eficácia na detecção de FA, atuando como uma ferramenta de auxílio à profissionais da saúde, assim acelerando e aumentando a precisão do diagnóstico desta doença (FAUST *et al.*, 2020).

Entre as abordagens computacionais, as decomposições tensoriais vêm ganhando espaço na análise de sinais de ECG (OLIVEIRA; ZARZOSO, 2018; OLIVEIRA *et al.*, 2021). Tais técnicas extraem as fontes de um sinal de ECG, permitindo uma análise mais profunda. Em particular, a decomposição em termos de bloco BTM tem mostrado vantagens em relação às decomposições tradicionais, oferecendo unicidade e estabilidade temporal em condições menos restritivas (ZARZOSO, 2017).

Apesar de sua eficácia na separação de fontes, a aplicação da BTM na classificação de FA ainda é limitada. Poucos estudos exploraram o uso dos parâmetros de convergência gerados por essa técnica como atributos para aprendizado de máquina. Além disso, há lacunas no uso de modelos mais avançados, como RNNs, para otimizar o desempenho do diagnóstico

automatizado, este trabalho busca preencher essas lacunas.

Isto posto, propõe-se, neste trabalho, explorar mais a fundo o potencial de classificação de FA, tendo como atributos de entrada os parâmetros de convergência gerados pela BTM, ao realizar uma exploração mais extensa de modelos de aprendizado de máquina, expandindo o conjunto de técnicas a serem testadas, aplicando juntamente a algoritmos tradicionais modelos de aprendizado profundo.

Os resultados alcançados destacam a viabilidade e o impacto dessa abordagem. Utilizando apenas atributos provenientes dos parâmetros de convergência gerados pela BTM e um modelo *Bidirectional LSTM* (BI-LSTM), foi possível alcançar uma Acurácia (ACC) de 95,30%, Precisão (PRE) de 96,53% e Sensibilidade (SE) de 96,41%. Assim, este trabalho contribui para a inovação na classificação de sinais de ECG, promovendo um diagnóstico mais rápido, preciso e acessível para a FA.

## 2 OBJETIVOS

Dado o contexto supracitado, este trabalho tem como objetivo geral investigar a viabilidade do uso de métricas de convergência da BTM aplicadas a sinais de ECG, em conjunto com modelos de AM, para auxiliar profissionais de saúde no diagnóstico e classificação precoce de indivíduos com FA. Ademais, este trabalho tem os objetivos específicos listados abaixo.

- Explorar a relação entre as métricas extraídas por meio da decomposição tensorial BTM e os padrões de variabilidade observados em sinais de ECG, avaliando sua capacidade de distinguir entre FA e RSN.
- Analisar a influência do tipo de modelo de AM na eficácia da classificação de FA e RSN, investigando como diferentes arquiteturas afetam o desempenho dos classificadores ao utilizar atributos derivados da BTM.
- Realizar testes abrangentes em múltiplas bases de dados, avaliando a generalização das métricas extraídas e a robustez dos modelos propostos em diferentes cenários clínicos.
- Investigar como a combinação de métodos tradicionais e métricas baseadas em decomposições tensoriais pode contribuir para avanços na detecção de FA, possibilitando modelos mais eficientes e interpretáveis.

### 3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os conceitos que funcionam como base para o entendimento do Trabalho de Conclusão de Curso, deste modo, a Seção 3.1 apresenta as principais noções de ECG, a Seção 3.2 sobre FA, a Seção 3.3 apresenta os conceitos envolvendo decomposição tensorial, e como pode ser utilizada para extração de atributos, finalmente, a Seção 3.5 discorre sobre as técnicas de classificação a serem utilizadas.

#### 3.1 Eletrocardiograma como ferramenta de diagnóstico

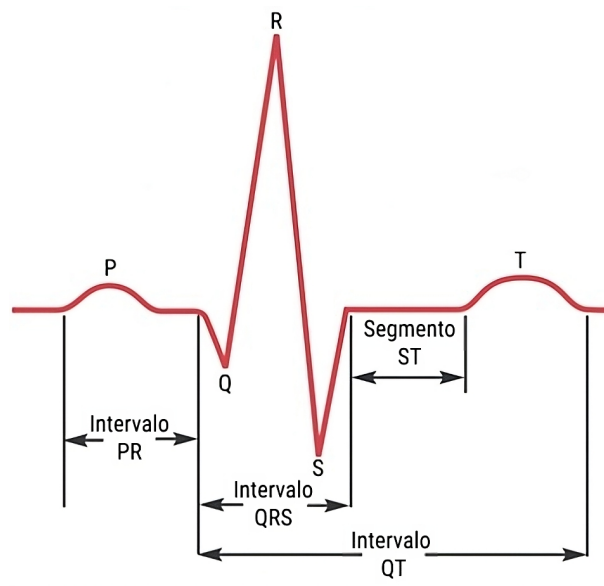
O ECG é uma ferramenta amplamente utilizada na prática médica para registrar a atividade elétrica do coração. Ele é obtido por meio de eletrodos posicionados na superfície corporal, gerando um gráfico que representa as variações elétricas do coração ao longo de cada ciclo cardíaco. Essas variações são registradas em milivolts (mV) e refletem os processos de despolarização e repolarização dos átrios e ventrículos, os quais são essenciais para o funcionamento adequado do coração (ASHLEY; NIEBAUER, 2004).

Os principais componentes do ECG incluem a onda P, que indica a despolarização dos átrios, marcando o início da contração atrial. Após isso, o impulso elétrico alcança o nó Atrioventricular (AV), gerando o intervalo PR, que reflete o tempo necessário para a condução elétrica através do nó AV, até o início do complexo QRS, responsável pela despolarização ventricular (MACFARLANE *et al.*, 2010).

O complexo QRS, formado pelas ondas Q, R e S, é um dos elementos mais importantes do ECG, pois está diretamente relacionado à contração dos ventrículos e à análise do ritmo cardíaco. Em seguida, o segmento ST representa a fase inicial da repolarização ventricular, a onda T reflete a repolarização ventricular. Duas ondas R do complexo QRS é chamado de RRI, este representa um ciclo cardíaco completo, e é comumente utilizado para calcular a variabilidade da frequência cardíaca. O diagrama na Figura 1 apresenta a disposição destas ondas (GARCIA; HOLTZ, 2012).

Em termos de eletrodos, um ECG padrão utiliza dez eletrodos para gerar doze derivações, cada uma oferecendo uma perspectiva diferente da atividade elétrica do coração, permitindo uma análise mais abrangente do funcionamento cardíaco. Essas derivações são divididas em três grupos principais: bipolares dos membros, unipolares aumentadas dos membros e precordiais (ASHLEY; NIEBAUER, 2004).

Figura 1 – Padrão básico da atividade elétrica através do coração



Fonte: Adaptado de (ASHLEY; NIEBAUER, 2004)

As derivações bipolares dos membros formam o triângulo de Einthoven e são obtidas a partir de três eletrodos fixados no braço esquerdo, braço direito e perna esquerda. Essas derivações medem a diferença de potencial entre dois desses pontos. Os mesmos eletrodos também permitem calcular as derivações unipolares aumentadas dos membros, que registram a diferença de potencial entre um único eletrodo e um potencial zero estimado. Além disso, um quarto eletrodo é posicionado na perna direita, funcionando como referência ou terra (ASHLEY; NIEBAUER, 2004).

Ademais, seis eletrodos adicionais são posicionados em localizações específicas do tórax, formando as derivações unipolares precordiais. Cada uma dessas derivações mede a diferença de potencial entre seu respectivo eletrodo torácico e um potencial de referência estimado (ASHLEY; NIEBAUER, 2004).

Dessa forma, o ECG se configura como uma ferramenta diagnóstica indispensável para a avaliação da frequência, ritmo e regularidade cardíaca, sendo fundamental para a detecção de diversas condições cardíacas, como a FA, que apresenta um comportamento diferente de um RSN.

### 3.2 Fibrilação atrial

Um distúrbio do ritmo cardíaco, ou arritmia, caracteriza-se por alterações no padrão normal de batimento do coração, que pode ocorrer de forma antecipada, tardia, lenta, rápida

ou irregular (ACHARYA *et al.*, 2016; HAGIWARA *et al.*, 2018). A FA, é a arritmia cardíaca sustentada mais comum na população adulta, com 59.7 milhões de casos relatados no mundo em 2019 (ROTH *et al.*, 2020). Ademais, estima-se que 25% dos indivíduos com mais de 55 anos da Europa e Estados Unidos da America irão desenvolver FA durante suas vidas (LLOYD-JONES *et al.*, 2004).

A FA, distingue-se por uma atividade elétrica desorganizada nos átrios resultante em contrações rápidas e irregulares (BRUNDEL *et al.*, 2022). Durante a FA, a ativação anormal do coração compromete a sua capacidade de bombear sangue de maneira eficiente, o que pode levar a uma série de complicações, incluindo AVC, insuficiência cardíaca e aumento do risco de mortalidade (MAGALHÃES *et al.*, 2016). A prevalência da FA cresce com a idade, especialmente em indivíduos com mais de 70 anos. Além disso, fatores como hipertensão, diabetes e obesidade desempenham um papel significativo na predisposição à condição, ressaltando a importância de sua detecção precoce para prevenir eventos adversos (MAGALHÃES *et al.*, 2016).

Essa condição pode ser classificada em quatro categorias principais, paroxística, caracterizada por episódios de até 7 dias que podem cessar espontaneamente ou com intervenção; persistente, quando os episódios duram mais de 7 dias e requerem intervenção farmacológica ou elétrica para cessar; persistente de longa duração, definida por episódios com duração mínima de 12 meses em que há tentativa de controle; e permanente, quando paciente e médico optam por não buscar a restauração para o RSN (HINDRICKS *et al.*, 2021).

Enquanto pacientes podem ser assintomáticos para FA, existem diversos possíveis sintomas, dentre elas, palpitações, fadiga, dor no peito e desmaios (BRUNDEL *et al.*, 2022). Assim, diversas abordagens clínicas podem ser utilizadas para auxiliar na identificação da FA, como a palpação do pulso arterial, a ausculta cardíaca, questionários de triagem baseados em fatores de risco, dispositivos automáticos de medição de pressão arterial com detecção de irregularidades e a avaliação de sinais clínicos durante o exame físico geral. No entanto, apesar dessas alternativas, o ECG é reconhecido como o padrão-ouro para o diagnóstico da FA, oferecendo precisão e informações diretas sobre a atividade elétrica do coração (HINDRICKS *et al.*, 2021).

### **3.2.1 Morfologia de um sinal de ECG com Fibrilação Atrial**

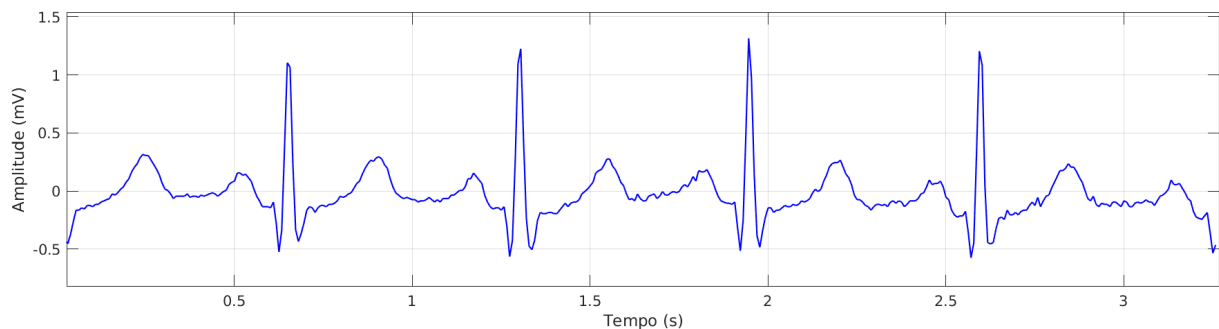
Em um sinal de ECG, a FA se manifesta de maneira bem distinta. Em vez das ondas P típicas, podem ser observadas pequenas oscilações irregulares, chamadas de ondas

F, que representam a atividade elétrica atrial caótica e altamente irregular. Essas ondas F são de amplitude variável e aparecem em diferentes formas, mas são características essenciais na detecção de FA. A principal consequência dessa irregularidade é a ausência de um intervalo PR definido, já que não há uma condução normal da atividade elétrica do átrio para o ventrículo. Em vez disso, os intervalos entre os complexos QRS tornam-se altamente irregulares, sem um padrão fixo, o que é uma das características mais proeminentes da FA no ECG (PEREZ *et al.*, 2009).

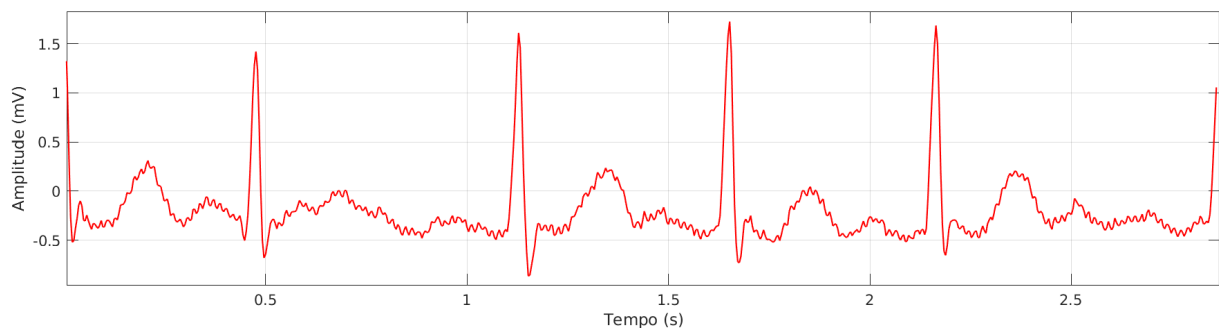
Além da irregularidade dos intervalos entre os complexos QRS, o ritmo ventricular pode ser extremamente variável. O grau de irregularidade pode ser muito variável, com os RRIs – distância entre os complexos QRS sucessivos – mostrando grandes flutuações, o que é um indicativo claro de FA (PEREZ *et al.*, 2009).

Esta característica pode ser observada na Figura 2, onde se compara um ECG com RSN e um ECG com FA. No ECG normal, as ondas P são bem definidas, e os intervalos entre os complexos QRS seguem um padrão regular. Já no ECG com FA, as ondas P são ausentes, e as ondas F geram uma flutuação no ritmo, resultando na irregularidade dos intervalos QRS.

Figura 2 – Comparativo de sinais reais de ECG



(a) Ritmo Sinusal Normal



(b) Fibrilação Atrial

Fonte: Autor

Além da análise visual do ECG convencional, métodos computacionais são incorporados para aprimorar a detecção de FA, principalmente em casos assintomáticos ou quando a

arritmia é de curta duração. Técnicas baseadas em aprendizado de máquina têm mostrado grande potencial para identificar padrões de FA que podem não ser facilmente reconhecíveis por uma análise visual convencional (FAUST *et al.*, 2020). Essas abordagens, quando aplicadas a grandes bases de dados de ECG, podem detectar a FA com alta precisão, auxiliando os profissionais de saúde no diagnóstico e acompanhamento da arritmia.

O ECG de 12 derivações, dentre outras ferramentas, como o monitoramento Holter e dispositivos portáteis de ECG, têm se mostrado eficazes na captura de episódios de FA ao longo do tempo. O monitoramento contínuo ou a longo prazo permite uma análise mais completa do comportamento da FA, que pode ser intermitente ou ocorrer apenas durante períodos específicos do dia, facilitando o diagnóstico em casos onde a FA não é constante ou está presente somente por breves intervalos (ZIEGLER *et al.*, 2006).

A análise de sinais de ECG com FA exige métodos que possam separar componentes do sinal, como a Atividade Atrial (AA), da atividade dos ventrículos e de outros ruídos. Nesse contexto, técnicas de decomposição surgem como ferramentas poderosas. Essas técnicas permitem a extração de informações específicas da AA, oferecendo representações multidimensionais que preservam a estrutura complexa dos sinais de ECG. Com a aplicação da decomposição tensorial BTM, é possível isolar padrões relacionados à AA de maneira mais robusta, facilitando a identificação de características discriminatórias para a classificação de FA.

### **3.3 Block-Term Decomposition**

As decomposições tensoriais são uma extensão das decomposições matriciais para dados multidimensionais. Enquanto a decomposição de matrizes lida com dados bidimensionais, as decomposições de tensores permitem analisar informações organizadas em múltiplas dimensões, capturando padrões mais complexos e inter-relacionados. Um tensor pode ser entendido como uma generalização de matrizes para dimensões superiores, em que uma matriz corresponde a um tensor de segunda ordem (modo dois) e um tensor de terceira ordem (modo três) pode ser visualizado como um cubo. Essa estrutura possibilita a preservação de relações intrínsecas nos dados, sendo especialmente útil em aplicações como processamento de sinais, reconhecimento de padrões e aprendizado de máquina. Assim, as decomposições tensoriais oferecem uma maneira eficiente de representar e processar grandes volumes de informações de forma compacta e interpretável.

O objetivo principal de uma decomposição tensorial é expressar um tensor de ordem

elevada em componentes de ordem inferior, revelando padrões latentes e simplificando a análise de dados. A decomposição PARAFAC (*Parallel Factor Analysis*), também chamada de CP (*Canonical Polyadic Decomposition*) (HITCHCOCK, 1927; CARROLL; CHANG, 1970) e a decomposição de Tucker (TUCKER, 1966) são os métodos mais comuns. No entanto, a BTD, apresentada por Lathauwer (2008), é uma técnica mais geral, que oferece maior flexibilidade, preservando a unicidade da decomposição sob condições mais brandas. A BTD permite modelar dados que possuem uma estrutura de blocos específica, como os sinais de ECG analisados neste trabalho.

A BTD expande a ideia da decomposição PARAFAC/CP ao permitir que um tensor seja representado como uma soma de tensores de menor ordem. Assim, a BTD possui uma característica de que seus componentes possuem um baixo *rank* multilinear, logo, pode-se definir um caso particular onde cada termo tem um *rank*  $(L_r, L_r, 1)$  (LATHAUWER, 2008).

Matematicamente, a BTD de um tensor  $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$  de terceira ordem e com *rank*  $(L_r, L_r, 1)$  pode ser descrita como:

$$\mathcal{T} = \sum_{r=1}^R \mathbf{E}_r \circ \mathbf{c}_r, \quad (3.1)$$

onde  $R$  é o *rank* do tensor, ou o número de blocos,  $\mathbf{E}_r \in \mathbb{R}^{I \times J}$  é uma matriz com *rank*  $L_r$ ,  $\mathbf{c}_r \in \mathbb{R}^K$  é um vetor e  $\circ$  denota o produto de Kronecker. É possível ainda realizar uma fatoração em  $\mathbf{E}_r$ , encontrando  $\mathbf{A}_r \mathbf{B}_r^T$ , onde  $\mathbf{A}_r \in \mathbb{R}^{I \times L}$  e  $\mathbf{B}_r \in \mathbb{R}^{J \times L}$ , obtendo:

$$\mathcal{T} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{c}_r \quad (3.2)$$

O *rank*  $L_r$  em cada componente permite controlar o grau de liberdade do sistema, possibilitando uma análise mais detalhada ou mais compacta conforme necessário. Para sinais biomédicos como o ECG, essa capacidade de decomposição em blocos permite uma discriminação mais precisa das fontes de sinal, como a atividade atrial e ventricular (RIETA *et al.*, 2004). Ademais, Lathauwer (2011) demonstra que a BTD é essencialmente única caso as seguintes condições sejam respeitadas:

1) As matrizes de fatores

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_R \end{bmatrix} \in \mathbb{R}^{I \times \sum_{r=1}^R L_r},$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \dots & \mathbf{B}_R \end{bmatrix} \in \mathbb{R}^{J \times \sum_{r=1}^R L_r},$$

tenham posto completo nas colunas, ou seja, linearmente independentes, o que necessita que  $\sum_{r=1}^R L_r \leq \min(I, J)$ .

2) A matriz  $\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \dots \ \mathbf{c}_R] \in \mathbb{R}^{K \times R}$  não tenha colunas proporcionais.

A implementação computacional da BTM é realizada por meio de algoritmos de otimização que minimizam a diferença entre o tensor original e a decomposição estimada. Lathauwer e Nion (2008) propõe o uso do método *Alternating Least Squares* (ALS), o qual visa minimizar a função de erro quadrático entre os dados reais e a representação aproximada. Sorber *et al.* (2013) propôs o uso de um método mais comum, o *Nonlinear Least Squares* (NLS), que ajusta iterativamente os fatores da decomposição para convergir em uma solução ótima.

No entanto, a escolha da estrutura do modelo BTM representa um desafio, uma vez que a formulação padrão baseada em mínimos quadrados pode ser mal posta. Para lidar com essa dificuldade, Goulart *et al.* (2020) propõem um algoritmo baseado em *group lasso* alternado, permitindo estimar simultaneamente os parâmetros do modelo e sua estrutura de forma bem posta. Essa abordagem mostrou-se mais robusta em relação à inicialização, conforme demonstrado em testes com dados sintéticos, e também eficaz na extração da AA em gravações de ECG durante episódios de fibrilação atrial, possibilitando a identificação de padrões fisiologicamente interpretáveis.

### 3.4 Hankel-BTM

A tensorização através da hankelização é um passo utilizado para transformar os sinais das derivações do ECG em um tensor multidimensional que possa ser utilizado na BTM. A hankelização é uma técnica usada em análise de séries temporais e compressão de dados, que transforma uma sequência unidimensional em uma matriz de Hankel. No caso da decomposição tensorial, esse processo é estendido para formar tensores.

Seja  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  um vetor de ECG de comprimento  $N$ . A matriz de Hankel do sinal pode ser construída mapeando as amostras na anti-diagonal, dessa forma, a  $\mathbf{H}$  correspondente é dada por:

$$\mathbf{H} = \begin{bmatrix} x_1 & x_2 & \dots & x_L \\ x_2 & x_3 & \dots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-L+1} & x_{N-L+2} & \dots & x_N \end{bmatrix}, \quad (3.3)$$

onde  $L$  é a dimensão desejada da matriz de Hankel.

A extensão do conceito de hankelização pode ser realizada não somente sobre um único vetor de ECG, como também em múltiplas derivações do sinal, formando um tensor de ordem superior a partir de várias matrizes de Hankel. Seja  $N$  o número de derivações do sinal de ECG de interesse, para cada segmento  $\mathbf{x}_i$ , pode-se construir uma matriz de Hankel  $\mathbf{H}_i$ . Assim, a partir de  $N$  derivações, obtemos  $N$  matrizes de Hankel:

$$\mathcal{H} = \left[ \mathbf{H}_1 \quad \mathbf{H}_2 \quad \dots \quad \mathbf{H}_n \right] \quad (3.4)$$

A concatenação dessas  $N$  matrizes em uma nova dimensão resulta em um tensor  $\mathcal{H} \in \mathbb{R}^{L \times (N-L+1) \times N}$ . Esse tensor então pode ser decomposto utilizando a BTM, onde as dependências temporais e espaciais no ECG são capturadas de forma simultânea. Esse método, conhecido como Hankel-BTD, permite explorar padrões latentes de forma mais profunda, modelando a estrutura temporal em um único segmento de ECG, através de técnicas de separação de fontes cega, ou *Blind Source Separation* (BSS), na qual se visa recuperar sinais de origem, ou fontes, sem conhecimento prévio da matriz de mistura, além das relações entre diferentes segmentos do sinal (LATHAUWER, 2011; OLIVEIRA; ZARZOSO, 2019).

Uma técnica inovadora para a detecção de FA foi proposta utilizando os indicadores de convergência e otimização da decomposição tensorial BTM como atributos (CARDOSO, 2024). A hipótese central dessa abordagem é que a decomposição BTM de um segmento de sinal de ECG com FA apresenta comportamento diferente em relação a um segmento com o RSN especialmente no que diz respeito à eficiência de convergência do algoritmo. Ou seja, espera-se que os sinais de ECG saudáveis convergirão de maneira lenta, enquanto sinais com FA resultarão em uma convergência mais eficiente, refletindo-se nos indicadores de otimização do processo (CARDOSO, 2024).

A justificativa matemática para essa hipótese está na unicidade da decomposição Hankel-BTD quando aplicada a sinais com FA. Como discutido por Zarzoso (2017), as fontes atriais presentes em sinais com FA podem ser modeladas por filtros do tipo só-polos, ou modelos exponenciais, que garantem a unicidade da decomposição. Essa modelagem é representada pela seguinte equação:

$$s_{r,n} = \sum_{l=1}^{L_r} \lambda_{l,r} z_{l,r}^{(n-1)}, \quad (3.5)$$

no qual  $r$  indica a  $r$ -ésima fonte atrial estimada,  $n$  é o índice de tempo discreto,  $L_r$  é o número de termos exponenciais,  $z_{l,r}$  é o  $l$ -ésimo polo da  $r$ -ésima fonte, e  $\lambda_{l,r}$  é o coeficiente de escala.

Essa modelagem permite que as condições de unicidade sejam satisfeitas ao aplicar a decomposição Hankel-BTD em sinais com FA. De acordo com Zarzoso (2017), as matrizes de Hankel geradas pela aplicação da BTD podem ser descritas por uma decomposição de Vandermonde, garantindo a unicidade da decomposição para determinados valores dos polos:

$$\mathbf{H}_S^{(r)} = \mathbf{V}_r \text{diag}(\lambda_{1,r}, \lambda_{2,r}, \dots, \lambda_{L_r,r}) \hat{\mathbf{V}}_r^T \in \mathbb{R}^{I \times J}, \quad (3.6)$$

em que  $\mathbf{V}_r$  e  $\hat{\mathbf{V}}_r$  são matrizes de Vandermonde definidas conforme a seguir:

$$\mathbf{V}_r = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_{1,r} & z_{2,r} & \dots & z_{L_r,r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1,r}^{J-1} & z_{2,r}^{J-1} & \dots & z_{L_r,r}^{J-1} \end{bmatrix} \in \mathbb{R}^{I \times L_r}, \quad (3.7)$$

$$\hat{\mathbf{V}}_r = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_{1,r} & z_{2,r} & \dots & z_{L_r,r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1,r}^{J-1} & z_{2,r}^{J-1} & \dots & z_{L_r,r}^{J-1} \end{bmatrix} \in \mathbb{R}^{J \times L_r}, \quad (3.8)$$

Conforme Zarzoso (2017), quando os polos são diferentes, a matriz de Vandermonde com  $L_r \leq I, J$  possui posto completo nas colunas  $L_r$ , garantindo a unicidade da decomposição para FA. No caso de polos idênticos, condições mais brandas ainda podem garantir a unicidade da decomposição. No entanto, em casos de RSN, a condição de unicidade não é garantida.

Essa diferença nos comportamentos de convergência entre sinais de FA e sinais saudáveis sugere que os parâmetros de otimização, como o número de iterações até a convergência, o erro residual e a norma dos passos iterativos, podem ser usados como atributos discriminativos. Esses resultados foram corroborados por estudos experimentais anteriores, que demonstraram a superioridade desta abordagem na detecção de FA em sinais de ECG (OLIVEIRA; ZARZOSO, 2018; OLIVEIRA; ZARZOSO, 2019).

### 3.5 Classificadores

A classificação de FA em ECG é um problema que requer métodos robustos para lidar com dados médicos frequentemente ruidosos e de natureza temporal. As abordagens mais comuns envolvem algoritmos supervisionados de aprendizado de máquina, que aprendem a partir de um conjunto de dados rotulados a identificar padrões nos atributos extraídos dos sinais de ECG. Esses algoritmos podem ser divididos em duas categorias principais: métodos tradicionais de aprendizado de máquina e redes neurais, com foco em arquiteturas recorrentes. A seguir, discutimos cada abordagem em detalhe.

#### 3.5.1 Métodos Tradicionais de Aprendizado de Máquina

Os métodos tradicionais de aprendizado de máquina baseiam-se em teorias estatísticas e matemáticas para modelar os dados e realizar a classificação. Entre os modelos mais utilizados para análise de ECG, destacam-se os baseados em árvores de decisão, modelos de proximidade e redes neurais artificiais simples.

##### 3.5.1.1 Árvore de Decisão

Modelos baseados em *Decision Tree* (DT) organizam os dados em uma hierarquia de decisões binárias, onde cada nó da árvore representa uma regra de divisão baseada em um atributo. Esse tipo de modelo é altamente interpretável e fácil de implementar, sendo útil em tarefas onde a transparência do processo de decisão é crítica (BREIMAN *et al.*, 2017).

As DTs são construídas dividindo-se o conjunto de treinamento com base nos atributos, visando gerar dois nós terminais (ou folhas) a cada divisão. Para determinar a melhor característica ou preditor em cada nó, uma das métricas utilizadas é a impureza de Gini. Esta métrica avalia a probabilidade de uma amostra ser classificada incorretamente, se escolhida aleatoriamente, sendo calculada da seguinte maneira:

$$G(i) = \sum_{x \neq y} f(x|i)f(y|i), \quad (3.9)$$

onde  $G(i)$  representa o índice de impureza de Gini no nó  $i$ , e  $f(x|i)$  e  $f(y|i)$  indicam as probabilidades, ou densidade de probabilidades em casos contínuos, das duas classes presentes no nó  $i$ . Além disso, a profundidade máxima da árvore é um parâmetro crucial que controla o nível de

complexidade do modelo, limitando a quantidade de divisões que a árvore pode realizar. Esse controle é fundamental para evitar o sobreajuste, assegurando que o modelo generalize bem para dados não vistos (JAHAN *et al.*, 2022).

### 3.5.1.2 *Random Forest e Extra Trees*

Árvores de decisão individuais são suscetíveis ao sobreajuste, ou *overfitting*, especialmente em conjuntos de dados complexos ou ruidosos, assim, métodos como o *bagging* (agregação de amostras) ajudam a mitigar esse problema. O *bagging* combina múltiplas árvores de decisão independentes, construídas a partir de subconjuntos aleatórios dos dados e atributos. Ao combinar essas árvores, o modelo pode reduzir a variância e melhorar a capacidade de generalização, uma vez que cada árvore tende a aprender diferentes aspectos dos dados, evitando que o modelo se ajuste excessivamente a padrões específicos e não generalizáveis do conjunto de treinamento (BREIMAN, 2001).

*Random Forest* é um dos algoritmos mais populares baseados em *bagging*. Sua construção envolve a criação de inúmeras árvores de decisão, cada uma treinada em um subconjunto aleatório dos dados, extraído aleatoriamente um subconjunto das características em cada nó da árvore. Isso garante que as árvores sejam descorrelacionadas, ou seja, aprendam padrões diferentes, o que, por sua vez, contribui para melhorar a robustez e a precisão do modelo, especialmente quando aplicado a dados heterogêneos, como sinais de ECG (BREIMAN, 2001). O modelo final é construído a partir da média ou votação das previsões feitas por cada árvore individual, tornando o modelo mais estável e com uma maior capacidade de generalização.

Embora o *Random Forest* seja eficaz em diversas tarefas, o método *Extra Trees* (ou Árvores Extremamente Aleatórias) propõe um aprimoramento no processo de construção das árvores. Enquanto o *Random Forest* seleciona aleatoriamente um subconjunto de características em cada nó, o *Extra Trees* vai um passo além, realizando uma randomização ainda mais acentuada. Durante a construção de cada árvore, o algoritmo seleciona um ponto de divisão aleatório para as características escolhidas, ao invés de calcular a melhor divisão com base em algum critério como a impureza de Gini ou a entropia. Esse processo torna o treinamento mais rápido, uma vez que as divisões são escolhidas aleatoriamente em vez de serem calculadas de forma determinística. Além disso, o *Extra Trees* tende a ser mais robusto a ruídos devido à maior aleatoriedade durante o treinamento.

### 3.5.1.3 XGBoost, LightGBM e Gradient Boosting

Os algoritmos baseados em *boosting*, diferentemente de técnicas de *bagging*, constroem modelos sequencialmente, onde cada árvore subsequente é treinada para corrigir os erros cometidos pela árvore anterior. Esse processo de correção contínua permite que o modelo se concentre nos exemplos mais difíceis de classificar, melhorando seu desempenho iterativamente. O *Gradient Boosting*, como método genérico, ajusta a previsão do modelo por meio do cálculo do gradiente da função de perda, focando nos resíduos do modelo anterior. No entanto, a técnica tradicional de *Gradient Boosting* pode ser computacionalmente cara e lenta em conjuntos de dados grandes, o que motivou o desenvolvimento de versões otimizadas como o *Extreme Gradient Boosting* (XGBoost) e o *Light Gradient-Boosting Machine* (LightGBM) (KE *et al.*, 2017).

XGBoost é uma implementação amplamente utilizada e altamente otimizada do *Gradient Boosting*. Ele incorpora técnicas avançadas de regularização, como *L1* e *L2*, para evitar o sobre ajuste, além de usar uma abordagem de *shrinkage* (redução dos ajustes feitos por cada árvore), que contribui para melhorar a generalização do modelo. O XGBoost também se destaca por seu uso de técnicas eficientes de *parallel processing* e *tree pruning* (poda de árvores), permitindo um treinamento mais rápido sem sacrificar a acurácia. Além disso, a capacidade do XGBoost de lidar com dados desbalanceados, comum em aplicações como a classificação de sinais de ECG, é uma de suas maiores vantagens, por permitir ajustar pesos para diferentes classes ou amostras no treinamento (CHEN; GUESTRIN, 2016).

O LightGBM é uma variante recente do *boosting* que se destaca pela utilização de uma técnica chamada *leaf-wise splitting*, que realiza a divisão das folhas das árvores de maneira diferente do XGBoost. Em vez de buscar o melhor ponto de divisão em todos os níveis da árvore, o LightGBM foca em dividir a folha que apresenta a maior melhoria no erro, permitindo uma construção de árvores mais profundas e com maior capacidade de representar padrões complexos nos dados. Esse método acelera significativamente o treinamento, especialmente em conjuntos de dados grandes, ao reduzir o número de divisões necessárias para construir uma árvore. O *leaf-wise splitting* permite que o LightGBM se beneficie de um treinamento mais rápido, mantendo a precisão do modelo, o que o torna particularmente útil em tarefas que envolvem grandes volumes de dados, como a classificação de sinais de ECG (KE *et al.*, 2017).

### 3.5.1.4 Modelos de Classificação Simples: KNN, Naive Bayes e Regressão Logística

Além dos modelos mais complexos, existem métodos de classificação simples que também podem ser eficazes em tarefas de diagnóstico médico, como a análise de sinais de ECG. Embora esses modelos geralmente não tenham a mesma flexibilidade dos algoritmos mais sofisticados, sua simplicidade e rapidez os tornam opções valiosas, especialmente em cenários onde a separação entre as classes é clara ou quando a interpretabilidade é uma prioridade. Três desses modelos amplamente utilizados são o *K-Nearest Neighbors* (KNN), o *Naive Bayes* e a Regressão Logística.

O KNN é um algoritmo baseado em proximidade, que classifica uma amostra desconhecida segundo os seus  $k$  vizinhos mais próximos no espaço de atributos. A principal suposição do KNN é que amostras de classes semelhantes estão localizadas próximas umas das outras. Embora simples, o KNN pode se tornar computacionalmente ineficiente em bases de dados grandes, devido à necessidade de calcular a distância entre a amostra de teste e todas as amostras do conjunto de treinamento (KRAMER, 2013).

Por outro lado, o *Naive Bayes* é um classificador probabilístico que utiliza o teorema de Bayes, assumindo a independência entre as características dos dados e atributos respeitando a distribuição gaussiana. Embora essa suposição de independência seja frequentemente uma simplificação excessiva, ela torna o *Naive Bayes* extremamente eficiente, com baixo custo computacional. Em contextos como o diagnóstico de ECG, esse modelo pode ser particularmente útil quando as características extraídas apresentam uma relação estatística simples, mas independente, entre si. A sua simplicidade e rapidez tornam-no uma escolha interessante em sistemas de diagnóstico que exigem respostas rápidas (WEBB *et al.*, 2010).

A Regressão Logística, embora comumente associada a problemas de regressão, é amplamente utilizada para classificação binária e multiclasse. Baseada em uma função sigmoide, a Regressão Logística mapeia as entradas para uma probabilidade que pode ser interpretada como a chance de uma amostra pertencer a uma classe específica. Sua principal vantagem está na facilidade de interpretação, além de ser uma técnica robusta para cenários onde as classes são separadas linearmente. No contexto de ECG, a Regressão Logística é eficaz quando as variáveis apresentam relações lineares claras, sendo frequentemente utilizada quando a explicabilidade do modelo é uma prioridade (LAVALLEY, 2008).

### 3.5.2 Redes Neurais Recorrentes e Extensões

As redes neurais artificiais são modelos computacionais inspirados no funcionamento do cérebro humano e são amplamente utilizadas em aprendizado de máquina. Entre suas arquiteturas mais comuns está o Perceptron Multicamadas, ou *Multilayer Perceptron* (MLP), um tipo de rede *feedforward* composto por três tipos de camadas: a camada de entrada, que recebe os dados; uma ou mais camadas ocultas, responsáveis pelo processamento da informação; e a camada de saída, que gera a predição final. Cada neurônio em uma camada está totalmente conectado a todos os neurônios da camada seguinte, permitindo que a rede aprenda padrões complexos a partir dos dados (HAYKIN, 2009).

Matematicamente, o funcionamento de um neurônio em um MLP pode ser descrito pela seguinte equação:

$$y = \varphi(Wx + b), \quad (3.10)$$

onde  $x$  representa a entrada,  $W$  é a matriz de pesos associada às conexões,  $b$  é o viés e  $\varphi$  é a função de ativação não linear, como a sigmoide. Essa estrutura permite que o modelo encontre padrões e realize transformações não lineares nos atributos, permitindo uma separação entre classes de interesse (HAYKIN, 2009).

A RNN foi projetada para lidar com dados sequenciais, como texto, vídeo e sinais de ECG, nos quais o comportamento de uma entrada atual pode depender fortemente de entradas anteriores. Ao contrário dos métodos tradicionais de aprendizado de máquina, como as redes *feedforward*, as RNNs possuem conexões recorrentes que permitem a retenção de informações ao longo do tempo, facilitando a modelagem de dependências temporais em sequências de dados. Essa característica as torna adequadas para tarefas no qual a sequência de eventos é crucial, como a análise de séries temporais e sinais biológicos (YU *et al.*, 2019).

#### 3.5.2.1 Arquitetura Básica das RNNs

A principal inovação das RNNs é o uso de um ciclo na rede, o que significa que a saída de um neurônio em um determinado instante pode ser alimentada de volta como entrada para o próximo ciclo, permitindo que a rede lembre informações passadas. A equação fundamental

para uma RNN simples pode ser expressa da seguinte forma:

$$h_t = f(W_h h_{t-1} + W_x x_t + b), \quad (3.11)$$

$$y_t = h_t, \quad (3.12)$$

onde  $h_t$  é o estado oculto no instante  $t$ ,  $x_t$  é a entrada no instante  $t$ ,  $W_x$  e  $W_h$  são os pesos das conexões de entrada e ocultas, respectivamente,  $b$  é o viés, e  $f$  é uma função de ativação, como a tangente hiperbólica ou ReLU. O estado oculto  $h_t$  é passado para a próxima etapa do processo de recorrência, permitindo que a rede processe dependências temporais.

Embora as RNNs simples sejam eficazes para dados sequenciais, elas enfrentam um problema significativo: o desaparecimento ou explosão do gradiente. Esses problemas ocorrem durante o treinamento, especialmente quando a sequência de dados é muito longa. O gradiente, ao ser propagado para trás na rede durante a retro propagação, pode diminuir exponencialmente (desaparecimento) ou crescer descontroladamente (explosão), dificultando aprender dependências de longo prazo.

### 3.5.2.2 Long Short-Term Memory

Para superar as limitações das RNNs tradicionais, a arquitetura de *Long Short-Term Memory* (LSTM) foi proposta por Hochreiter e Schmidhuber (1997). O LSTM introduz uma estrutura de células de memória, projetada para controlar melhor o fluxo de informações através das etapas de tempo. Uma célula LSTM é composta por originalmente por dois principais componentes chamados *gates* (portões), que regulam a informação que entra ou sai da célula. A LSTM foi aprimorada com portão de esquecimento, introduzido por Gers *et al.* (2000), e segue a estrutura:

1. Portão de Entrada ( $i_t$ ): Este *gate* controla a quantidade de nova informação que será armazenada na célula de memória. Ele decide, com base na entrada  $x_t$  e no estado oculto anterior  $h_{t-1}$ , quais informações precisam ser atualizadas.
2. Portão de Esquecimento ( $f_t$ ): Esse *gate* controla quais informações armazenadas na célula de memória devem ser descartadas. Ele analisa tanto a entrada quanto o estado anterior para determinar a relevância de manter ou esquecer certas informações.
3. Portão de Saída ( $o_t$ ): Este *gate* decide qual parte da célula de memória será usada para gerar a saída da célula. A saída do LSTM é determinada pela multiplicação do *gate* de saída com a célula de memória filtrada.

A equação para a célula LSTM pode ser expressa da seguinte maneira:

$$i_t = \sigma(W_x x_t + W_h h_{t-1} + b_i), \quad (3.13)$$

$$f_t = \sigma(W_x x_t + W_h h_{t-1} + b_f), \quad (3.14)$$

$$o_t = \sigma(W_x x_t + W_h h_{t-1} + b_o), \quad (3.15)$$

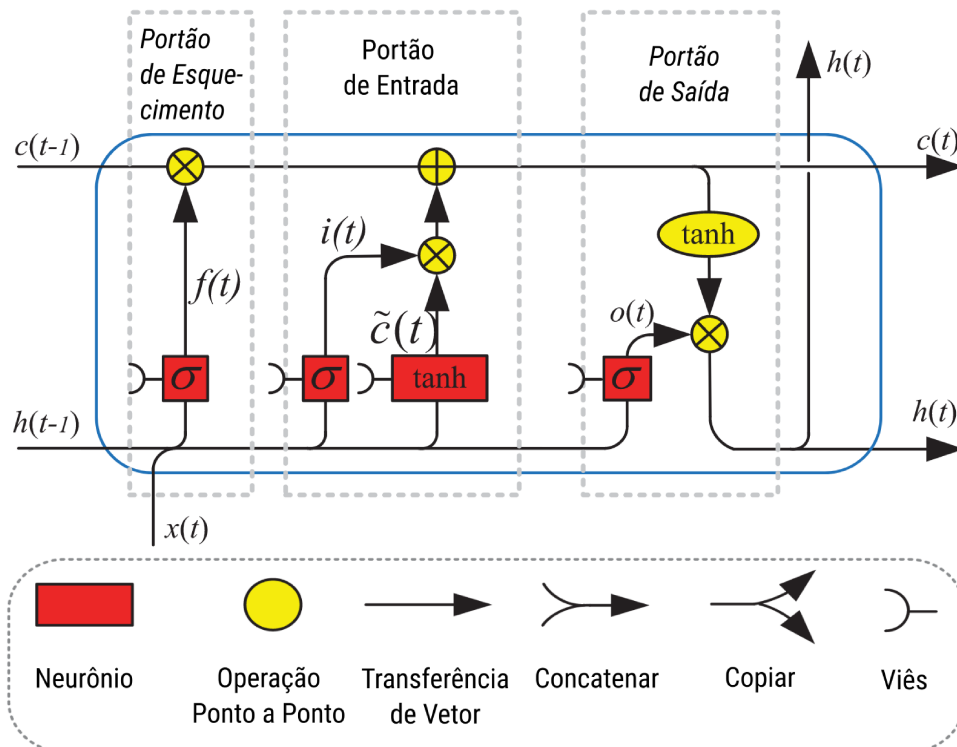
$$\tilde{C}_t = \tanh(W_x x_t + W_h h_{t-1} + b_C), \quad (3.16)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \quad (3.17)$$

$$h_t = o_t \cdot \tanh(C_t), \quad (3.18)$$

onde  $\sigma$  é a função sigmoide e  $\tanh$  é a tangente hiperbólica.  $W_x$  e  $W_h$  os pesos dos neurônios. As variáveis  $C_t$  e  $\tilde{C}_t$  representam a célula de memória e a memória candidata, respectivamente. A Figura 3 apresenta a estrutura descrita.

Figura 3 – Arquitetura de uma LSTM com portão de esquecimento



Fonte: Adaptado de (YU *et al.*, 2019)

Essa estrutura de portões permite que as LSTMs retenham informações de longo prazo, enquanto controlam com precisão o que deve ser mantido e o que deve ser esquecido, tornando-as extremamente eficazes para capturar dependências temporais longas em séries temporais, como os sinais de ECG, ou a convergência do BTB.

### 3.5.2.3 Gated Recurrent Unit

A *Gated Recurrent Unit* (GRU), proposta por Cho *et al.* (2014), é uma variação das LSTMs que simplifica a arquitetura ao combinar os *gates* de entrada e esquecimento em um único *gate* de atualização. Esse design mais simples resulta em menos parâmetros a serem treinados, tornando a GRU uma opção atraente em cenários onde a eficiência computacional é crucial. Além disso, a GRU muitas vezes apresenta um desempenho semelhante ao das LSTMs, apesar de sua simplicidade.

A arquitetura da GRU pode ser expressa através das seguintes equações:

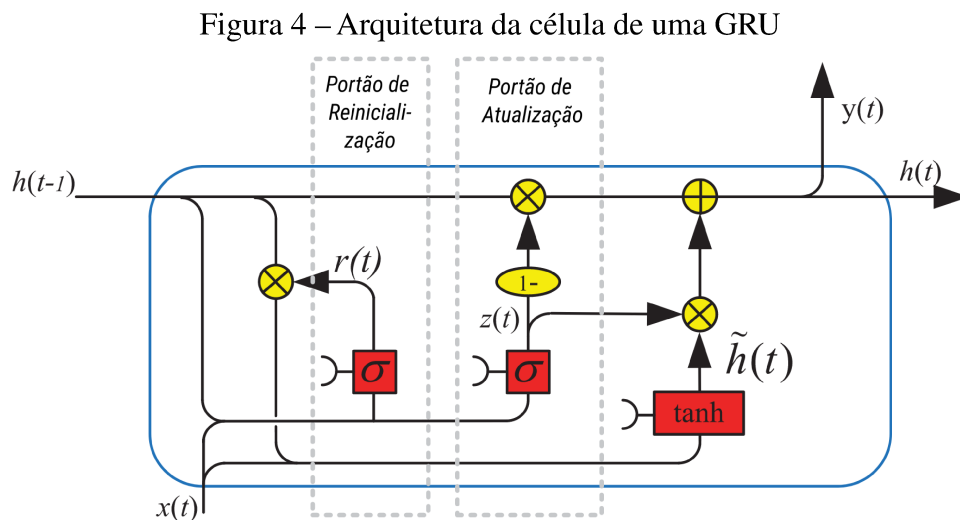
$$r_t = \sigma(W_x x_t + W_h h_{t-1} + b_r), \quad (3.19)$$

$$z_t = \sigma(W_x x_t + W_h h_{t-1} + b_z), \quad (3.20)$$

$$\tilde{h}_t = \tanh(W_x x_t + W_h (r_t \cdot h_{t-1}) + b_h), \quad (3.21)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t, \quad (3.22)$$

onde  $r_t$  é o portão de reinicialização,  $z_t$  é o portão de atualização e  $\tilde{h}_t$  é o candidato a estado oculto. A função de ativação  $\sigma$  é a sigmoide e  $\tanh$  é a tangente hiperbólica. A sua estrutura básica é apresentada na Figura 4.



Fonte: Adaptado de (YU *et al.*, 2019)

A principal vantagem das GRUs em relação às LSTMs é a menor complexidade computacional, o que as torna adequadas para dispositivos com recursos limitados, mantendo uma capacidade de modelagem de dependências temporais eficazes.

### 3.5.2.4 Bidirectional LSTM

Embora as LSTMs sejam eficazes para capturar dependências de longo prazo em uma direção temporal, elas ainda podem ter limitações quando o contexto de uma sequência depende de informações que ocorrem após um determinado ponto, no entanto, muitas das aplicações o sinal completo já está disponível, logo, não há razão para não aproveitar do contexto futuro (GRAVES; JAITLEY, 2014). Logo, para superar essa limitação, a BI-LSTM foi introduzida.

A ideia central da BI-LSTM é processar a sequência de entrada em duas direções: uma do passado para o futuro (como nas LSTMs tradicionais) e outra do futuro para o passado. Esse processamento bidirecional permite que o modelo capture informações contextuais em ambas as direções da sequência, o que é particularmente útil em sinais de ECG, onde as relações temporais anteriores e posteriores a um ponto específico podem ser igualmente importantes.

A arquitetura do BI-LSTM pode ser formulada como:

$$h_t^{\text{forward}} = f(W_x x_t + W_h h_{t-1}^{\text{forward}} + b), \quad (3.23)$$

$$h_t^{\text{backward}} = f(W_x x_{T-t} + W_h h_{t+1}^{\text{backward}} + b), \quad (3.24)$$

onde  $h_t^{\text{forward}}$  e  $h_t^{\text{backward}}$  representam os estados ocultos da LSTM para as direções para frente e para trás, respectivamente. A saída final para cada etapa de tempo  $t$  é obtida pela concatenação ou soma dos estados ocultos nas duas direções (ANDERSEN *et al.*, 2019).

### 3.5.3 Métricas de Desempenho

A avaliação do desempenho dos modelos de classificação é baseada na matriz de confusão, que organiza os resultados das previsões em quatro categorias fundamentais. O Verdadeiro Positivo (VP) representa os casos em que a classe positiva foi corretamente identificada, enquanto o Verdadeiro Negativo (VN) indica as instâncias da classe negativa classificadas corretamente. Por outro lado, um Falso Positivo (FP) ocorre quando um exemplo negativo é erroneamente classificado como positivo, e um Falso Negativo (FN) acontece quando um exemplo positivo é incorretamente classificado como negativo.

Com base nesses valores, diversas métricas podem ser calculadas para medir o desempenho do modelo. A ACC, por exemplo, expressa a proporção de previsões corretas em relação ao total de amostras:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}. \quad (3.25)$$

A PRE quantifica a proporção de previsões positivas que são realmente corretas, sendo especialmente relevante em cenários onde falsos positivos devem ser minimizados:

$$\text{Precisão} = \frac{VP}{VP + FP}. \quad (3.26)$$

Já a SE avalia a capacidade do modelo de identificar corretamente as amostras da classe positiva, sendo crucial em aplicações onde a detecção de falsos negativos deve ser minimizada:

$$\text{Sensibilidade} = \frac{VP}{VP + FN}. \quad (3.27)$$

Por fim, o *F1-score* é a média harmônica entre PRE e SE, fornecendo uma métrica balanceada para cenários em que há um desequilíbrio entre classes:

$$\text{F1-Score} = \frac{2 \times \text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}. \quad (3.28)$$

## 4 TRABALHOS RELACIONADOS

Diversos estudos utilizam técnicas de aprendizado de máquina para a detecção de FA de curto prazo, explorando abordagens variadas, desde modelos clássicos a métodos de aprendizado profundo. Cada estudo tem em vista aprimorar a precisão e a robustez dos modelos, seja através da combinação de atributos ou do uso de diferentes bancos de dados.

Por exemplo, Jahan *et al.* (2022) compara quatro algoritmos populares —DT, XGBoost, AdaBoost e *Support Vector Machine* (SVM) — usando as bases MIT-BHI *Atrial Fibrillation Database* (AFDB), MIT-BIH *Long-term Atrial Fibrillation Database* (LTAfDB), MIT-BIH *Arrhythmia Database* (MITDB) e MIT-BIH *Normal Sinus Rhythm Database* (NSRDB). Com base em sete atributos temporais dos segmentos de RRI, o estudo conclui que modelos *ensemble*, como o AdaBoost, apresentam melhor desempenho, alcançando SE de 87,58% e Especificidade (SPE) de 89,27%.

Em outra abordagem, Duan *et al.* (2022) explora o uso da função densidade de probabilidade do RRI de um sinal de ECG para a classificação de FA, ao invés de apenas atributos numéricos. Utilizando as mesmas bases de dados (AFDB, LTAfDB e NSRDB), foi construído histogramas dos RRI e aplica um modelo baseado em SVM, alcançando resultados impressionantes com ACC de 96,97%, SPE de 95,24% e SE de 99,94%.

Além disso, Hirsch *et al.* (2021) investigou a combinação de atributos derivados do RRI e da extração de características do AA por meio da Decomposição Empírica de Modos. Com a base AFDB, o estudo consegue alcançar uma ACC de 95,9%, SE de 96,1% e SPE de 97,4%, utilizando o modelo *Random Forest* (RF), relata-se que foi possível melhorar o desempenho ao introduzir os atributos provenientes da AA.

Em relação à variabilidade cardíaca, amplamente empregada para classificar FA, atributos que indicam a irregularidade dos intervalos entre batimentos foram explorados. Com SVM, Czabanski *et al.* (2020) realiza a detecção de episódios de FA em sinais de longa duração, alcançando ACC de 98,6%, SE de 98,9% e SPE de 98,4%.

O uso de modelos de aprendizado profundo também tem ganhado destaque. Andersen *et al.* (2019) apresenta um modelo baseado em RNNs com camadas convolucionais para extração de atributos dos RRI, que, avaliado em uma base não vista previamente, alcança ACC de 87,40% SPE de 98,96% e SE de 86,04%. Paralelamente, Aldughayfiq *et al.* (2023) propõe utilizar uma CNN unidimensional para extração de características locais do sinal de ECG e PPG, seguida por uma RNN (Bi-LSTM) para captura de informações contextuais, alcançando uma ACC de 95,0%

e PRE de 88%.

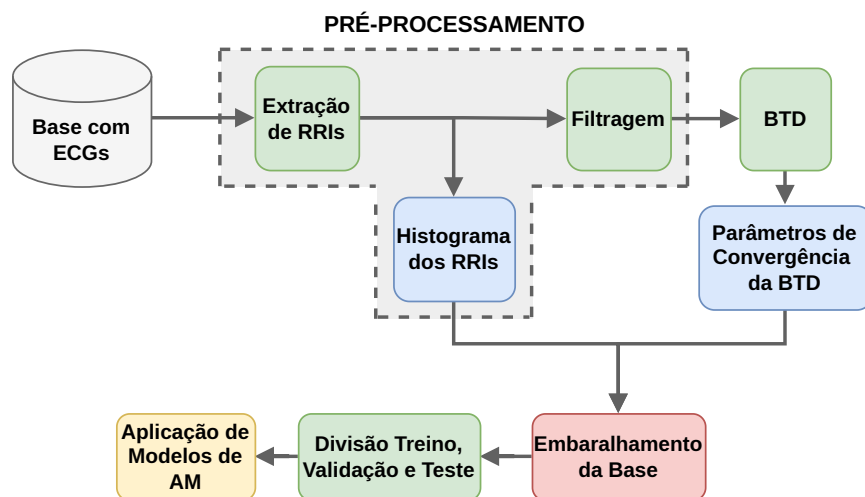
Nota-se como métodos tradicionais de aprendizado de máquina, como aqueles baseados em DT, e técnicas de aprendizado profundo, como as RNNs, são amplamente utilizados para a classificação de FA, utilizando diversos atributos, desde estatísticas extraídas do RRI até a extração de AA. Isto posto, este trabalho tem em vista aplicar os modelos previamente apresentados como eficazes para classificação de FA, juntamente com os parâmetros de convergência da BTD.

Mais recentemente, Cardoso (2024) explorou a combinação de parâmetros de convergência da BTD com histogramas de RRI como atributos para a classificação de FA em sinais de ECG. Testando diversos algoritmos *ensemble*, o estudo identifica o LightGBM como o mais eficaz, com uma ACC de 97,04%, SE de 97,44% e SPE de 97,02%. No entanto, o estudo não utiliza técnicas de aprendizado profundo, que poderiam permitir uma extração mais rica de informações. Além disso, embora vários parâmetros de convergência do BTD sejam vetores, o trabalho considera apenas o último valor, resultando na perda de informações potencialmente discriminatórias.

## 5 METODOLOGIA

O método proposto para a classificação de FA combina diversas etapas para explorar o potencial de dados derivados de sinais de ECG. Três bases de dados conhecidas contendo sinais de FA e RSN são utilizadas. Inicialmente, os RRIs são extraídos e, simultaneamente, é gerado o histograma desses intervalos. Os RRIs passam por um processo de filtragem e, em seguida, são submetidos à decomposição tensorial BTD, cuja análise gera indicadores de convergência. Esses atributos são então combinados, embaralhados e divididos em conjuntos de treino, validação e teste. Modelos compostos por técnicas clássicas e RNNs são treinados e validados utilizando esses conjuntos. Isto posto, a Figura 5 apresenta a metodologia proposta.

Figura 5 – Metodologia proposta



Fonte: Autor

Assim, nesta seção apresenta-se o processo metodológico proposto para a classificação da FA. A seção está estruturada em três partes principais: a Seção 5.1 aborda as bases de dados utilizadas, detalhando suas distribuições de classes e características. A Seção 5.2 descreve o pré-processamento dos dados, a Seção 5.3 detalha extração de atributos dos sinais de ECG com BTB, a Seção 5.4 descreve a divisão e embaralhamento da base resultante, já a Seção 5.5 explora o processo de classificação, que integra abordagens de aprendizado de máquina.

## 5.1 Base de dados

As bases de dados utilizadas na pesquisa foram: MIT-BHI AFDB, MIT-BIH LTAfDB e MIT-BIH NSRDB (GOLDBERGER *et al.*, 2000). Estas bases foram escolhidas por serem amplamente utilizadas em pesquisas envolvendo arritmias (RIZWAN *et al.*, 2021). Cada gravação possui anotações sobre seus meta-dados, contendo informações sobre a taxa de amostragem, derivações e anotações clínicas, como a presença de batimentos cardíacos e a marcação de todo o complexo QRS. As condições cardíacas associadas a cada instante na gravação também são documentadas. É possível acessar essas informações utilizando bibliotecas e ferramentas disponíveis, como a *WFDB* (SILVA; MOODY, 2014; SILVA *et al.*, 2021; XIE *et al.*, 2021).

A base MIT-BHI AFDB é composta por 25 gravações de ECG com duas derivações de humanos com FA paroxística. Entre elas, duas gravações não possuem anotações dos batimentos e, por isso, devem ser excluídas da base. Cada gravação individual tem 10 horas de duração, com uma taxa de amostragem de 250 Hz, resolução de 12 bits e amplitude de  $\pm 10$  mV. A base MIT-BIH LTAfDB contém 84 gravações, com duas derivações, de indivíduos com FA paroxística e persistente. Cada gravação é composta por duas derivações, com duração variando entre 24 e 25 horas e uma taxa de amostragem de 128 Hz, resolução de 12 bits e amplitude de  $\pm 20$  mV.

Finalmente, a base MIT-BIH NSRDB é composta por 18 gravações de pacientes com RSN, isto é, pessoas sem doenças cardíacas. As gravações têm duração variando entre 20 e 24 horas e possuem taxa de amostragem de 128 Hz.

## 5.2 Pré-processamento

Inicialmente, utilizando as anotações dos sinais de ECG disponibilizados pelos autores das bases, foram extraídos os RRIs para serem utilizados como amostra. Nominalmente, utilizou-se um segmento de sinal de 5 RRIs, valor encontrado por Cardoso (2024) como suficiente, demonstrando um bom compromisso entre desempenho e custo computacional.

Na base MIT-BHI AFDB, foram extraídos apenas os segmentos positivos (FA), descartando-se os trechos marcados como RSN. Dentre as 25 gravações, as de código 00735 e 03665, que não foram anotadas pelos autores, foram excluídas. Ademais, gravações 04043, 08405, e 08434, contem blocos de 10 segundos de sinal ilegível, marcadas com amplitude zero pelos autores. Os trechos onde mais de 20% do sinal apresentava amplitude zero foram

descartados, evitando problema de convergência no BTD. Para MIT-BHI LTAfDB e NSRDB, foram extraídos os trechos contendo somente FA e RSN, respectivamente.

Para cada um dos trechos, foi salvo o histograma dos RRIs utilizando 30 *bins*. Em seguida, os sinais brutos de ECG, foram filtrados utilizando um filtro Chebyshev passa-banda tipo II, com corte inferior de 0,35 Hz e superior de 60 Hz para suprimir ruídos da flutuação da linha de base e de alta frequência (SöRNMO; LAGUNA, 2006). Os segmentos foram salvos para na próxima etapa ser possível extrair os parâmetros de convergência da BTD. Para cada par, histograma e sinal, foi associado um código de identificação único, a fim de relacionar cada conjunto de atributos e a devida classe. Realizando esta divisão, a base tem a distribuição apresentada na Tabela 1.

Tabela 1 – Distribuição de segmentos (amostras) das bases de dados

Base	Qtd. Segmentos (FA)	Qtd. Segmentos (RSN)	
AFDB	101491	-	
NSRDB	-	361348	
LTAfDB	618183	-	
<b>Total:</b>	719674 (66%)	361348 (33%)	1081022

### 5.3 Extração de atributos utilizando BTD

Para realizar o cálculo da BTD, foi utilizado o programa MATLAB juntamente do *framework* de análise e operações com dados multidimensionais *TensorLab* (VERVLIET *et al.*, 2016). Nominalmente, para realizar o procedimento Hankel-BTD de ordem  $(L_r, L_r, 1)$ , utilizou-se a função *lll\_rand* para gerar uma estrutura pré-definida e pseudoaleatória utilizada como uma estimativa inicial para a decomposição, juntamente com um tensor  $\mathcal{T}$ , construído a partir do segmento de ECG através do processo de Hankelização. A Tabela 2 descreve o uso desta função, contando com as entradas e saídas. A decomposição foi realizada usando a função *lll\_nls*, que computando a BTD que gerou os parâmetros de convergência que utilizados como atributos de classificação. A Tabela 3 descreve as especificações desta função. Em seu trabalho, Zarzoso (2017) propõe utilizar um vetor R de tamanho 5, o rank multilinear  $L_r$  como 17 e o número de iterações limitado a 1000.

Ao realizar a decomposição, a função *lll\_nls* retorna o atributo **iter**, sendo o número necessário de iterações até a convergência, além dos vetores, contendo os seguintes parâmetros:

- **relfval**: A diferença no valor da função objetivo entre duas iterações sucessivas, em relação

Tabela 2 – Especificações da função *lll\_rnd* (Tensorlab).

<b>Nome da Função</b>	<code>lll_rnd</code>
<b>Propósito</b>	Inicialização pseudoaleatória para a decomposição $(L_r, L_r, 1)$ de um tensor.
<b>Entrada</b>	<ul style="list-style-type: none"> <li>– <code>size_tens</code>: Dimensões do tensor a ser estimado.</li> <li>– <code>L</code>: Vetor de tamanho <math>R</math>, onde cada posição é <math>L_r</math> de cada bloco.</li> <li>– <code>options</code> (opcional): Opções para a geração dos termos.</li> </ul>
<b>Saída</b>	<ul style="list-style-type: none"> <li>– <code>U_init</code>: Matrizes fatores iniciais pseudoaleatórias correspondentes à decomposição <math>(L_r, L_r, 1)</math>.</li> </ul>
<b>Uso</b>	<code>[U_init] = lll_rnd(size_tens, L).</code>

Fonte: (CARDOSO, 2024)

Tabela 3 – Especificações da função *lll\_nls* (Tensorlab).

<b>Nome da Função</b>	<code>@lll_nls</code>
<b>Propósito</b>	Calcular a decomposição em termos de blocos com rank multilinear $(L_r, L_r, 1)$ a partir do tensor $\mathcal{T}$ usando o método dos mínimos quadrados não lineares.
<b>Entrada</b>	<ul style="list-style-type: none"> <li>– <math>\mathcal{T}</math>: Tensor a ser estimado.</li> <li>– <code>Uinit</code>: Estimativa inicial para as matrizes de fatores.</li> <li>– <code>Lr</code> (opcional): Número de colunas em cada matriz fator.</li> <li>– <code>algorithm</code>: Escolha do método de otimização (<code>@nls_gndl</code>, <code>@nls_lm</code>).</li> <li>– <code>output format</code>: Formato de saída (btd ou cpd). Se não for fornecido, o mesmo formato da entrada será utilizado.</li> <li>– Outras opções podem ser repassadas através do método de otimização selecionado em <code>algorithm</code>.</li> </ul>
<b>Saída</b>	<ul style="list-style-type: none"> <li>– <code>U</code>: Matrizes fatores correspondentes à decomposição <math>(L_r, L_r, 1)</math>.</li> <li>– <code>output</code>: Parâmetros de otimização e convergência obtidos na última iteração do cálculo.</li> </ul>
<b>Uso</b>	<code>[U, output] = lll_nls(<math>\mathcal{T}</math>, Uinit).</code>

Fonte: (CARDOSO, 2024)

- ao seu valor inicial;
- *fval*: O valor da função objetivo em cada iteração;
- *relstep*: O tamanho do passo em relação à norma da iteração atual em cada iteração;
- *delta*: O raio da região de confiança em cada iteração;
- *rho*: A confiabilidade em cada tentativa de iteração;

Além disso, pode-se criar um atributo adicional utilizando as matrizes fator correspondentes à decomposição para construir um tensor estimado e calcular o erro relativo entre o tensor original e o modelo BTD estimado, chamado de *relerr*. O processo de decomposição e construção dos atributos para cada um dos trechos de RRI foi cronometrado; para isso, foi utilizado um computador com Ryzen 5700x, 32GB de RAM no sistema operacional Ubuntu 22.04.

Para confirmar a capacidade dos atributos em separar as classes, foi realizado o Teste U de Mann-Whitney, ou teste da soma dos postos de Wilcoxon, que testa se ao combinarmos os dados e colocarmos ordem, a posição dos elementos de cada grupo varia significativamente (MANN; WHITNEY, 1947). Assim, para o presente teste a hipótese ( $H_0$ ) nula foi definida como "as duas amostras são de distribuições contínuas com medianas iguais", onde o valor de significância foi definido em  $\alpha = 1 \times 10^{-2}$ , e o teste foi realizado utilizando o último valor de cada um dos vetores dos atributos supracitados.

Além disso, para avaliar quantitativamente o quanto de informação cada um dos atributos carrega, utilizou-se o algoritmo XGBoost, para determinar a importância de cada atributo. O processo foi realizado unindo todas as bases, juntamente de uma validação cruzada com 10 dobras. Após cada dobra, a média e desvio padrão dos da importância foi gerada. Ainda nesse sentido também foi gerada uma matriz de correlação de Pearson entre os atributos, com o objetivo de entender a covariância entre cada um dos atributos.

Outra análise relevante inclui a média, moda, mediana, desvio padrão e outras estatísticas descritivas do número de iterações para cada classe. Isto porque, como muitos dos atributos são vetores, é necessário entender o tamanho destes para construir, em seguida, um vetor de características com estes que tenha um tamanho representativo. Os resultados encontrados serão discutidos no capítulo seguinte, e os valores encontrados serão usados mais a frente para definir o tamanho do vetor de atributos.

Assim, usando os indicadores de convergência, foram construídos conjuntos de atributos, entre esses, têm-se:

- Vetor composto pelos valores referentes às métricas da última iteração da BTM;
- Histograma dos RRIs;
- Vetor de atributos da última iteração da BTM, juntamente com o histograma dos RRIs.

Ademais, outras combinações de atributos foram elaborados. O processo BTM, como mencionado, gera diferentes vetores de atributos que representam métricas em cada iteração até atingir a convergência. Assim, considerando os resultados a partir da análise dos atributos, foram criados vetores limitados a um tamanho de 150 iterações ou pontos. Esse valor foi definido com base no número médio de iterações por classe (129,56 para RSN e 59,58 para FA), buscando preservar o máximo de informações possíveis.

Durante a construção dos vetores de atributos, sinais com menos de 150 elementos tiveram zeros adicionados até atingir o tamanho necessário, enquanto sinais com mais de 150 elementos foram truncados. Seguindo esse procedimento, foram selecionados os atributos mais informativos, especificamente *fval* e *delta*. Dessa forma, os seguintes conjuntos de atributos foram definidos:

- *fval*, um vetor com 150 elementos;
- *fval* e *delta*, dois vetores de 150 elementos cada, totalizando 300 atributos;
- *fval*, *delta* e *relerr*, dois vetores de 150 elementos cada e um valor escalar, totalizando 301 atributos;
- *fval* e histograma dos RRIs, um vetor de 150 elementos e um vetor de 30 elementos, respectivamente, totalizando 180 atributos.

#### 5.4 Randomização e divisão treino, validação e teste

Finalmente, com os indicadores coletados para cada segmento de ECG e suas respectivas classes, pode-se embaralhar as amostras para evitar qualquer viés do modelo. Em seguida, realizou-se uma divisão estratificada, preservando o balanço de classes original, da base de dados entre treino e teste, nas proporções 80% e 20%, respectivamente. Inicialmente, para os modelos tradicionais, um teste geral foi realizado com a divisão 80%-20%. Em seguida, para o conjunto de atributos com melhor desempenho, os melhores modelos foram selecionados e submetidos a um ajuste dos parâmetros utilizando uma validação cruzada com 10 dobras na base de treino. Por fim, os hiperparâmetros ajustados foram avaliados na base de teste.

Para os modelos baseados em RNNs, partiu-se da divisão treino-teste inicial, onde a base de treino foi subdividida em 80% para treino e 20% para validação, resultando em

uma divisão final de 64% treino, 16% validação e 20% teste. Os parâmetros das RNNs foram ajustados utilizando os 64% de treino e acompanhando o desempenho nos 16% de validação. Após encontrar o melhor modelo, ele foi aplicado aos diferentes experimentos, sendo sempre treinado nos 64% de treino, acompanhado pelos 16% de validação e avaliado nos 20% de teste, que permaneceram intactos. Essa abordagem garante uma avaliação robusta e representativa do desempenho dos modelos em um cenário real.

## 5.5 Classificação e modelos

Após o embaralhamento e a divisão das bases, deu-se início à aplicação dos modelos para classificação. Diversos conjuntos de atributos foram criados para medir a influência de cada parâmetro. Além disso, como mencionado anteriormente, foram aplicados diferentes modelos.

Para realizar a classificação da FA utilizando esses conjuntos de atributos, foram empregadas tanto técnicas tradicionais de aprendizado de máquina quanto abordagens de aprendizado profundo. A configuração e disposição dessas técnicas são descritas nas subseções a seguir.

### 5.5.1 Modelos tradicionais de aprendizado de máquina

Os modelos tradicionais de aprendizado de máquina utilizados foram os seguintes:

- *Random Forest*;
- Extra Trees;
- XGBoost;
- LightGBM;
- *Gradient Boosting*;
- KNN;
- Regressão Logística;
- *Naive Bayes*.

Todos esses modelos foram treinados nos diferentes atributos mencionados. Para identificar o caso de melhor desempenho, foi realizada uma análise comparativa, selecionando os modelos mais promissores para ajuste de hiperparâmetros. Esse ajuste foi realizado por meio de validação cruzada com 10 dobras na base de treino. O processo foi conduzido com o auxílio do *framework* de otimização *Optuna*, que permite ajustes mais rápidos e eficientes ao empregar

algoritmos de estado da arte para amostragem de configurações e exclusão de combinações pouco promissoras (AKIBA *et al.*, 2019).

O objetivo da otimização foi maximizar a acurácia dos modelos. O número máximo de combinações testadas por modelo foi limitado a 100. Os modelos selecionados e os respectivos parâmetros otimizados são apresentados na seção de resultados. Paralelamente, além dos modelos tradicionais, foram também explorados modelos baseados em RNNs.

### 5.5.2 Arquitetura das RNNs

Dado que os atributos possuem dependência temporal — isto é, os valores de cada iteração variam em relação às iterações anteriores — optou-se pelo uso de RNNs. As arquiteturas específicas utilizadas, baseadas em LSTM e GRU, são apresentadas na Figura 6. Esta arquitetura foi encontrada mediante experimentos e ajuste dos parâmetros, visando encontrar um equilíbrio entre arquiteturas propostas e o custo computacional associado ao treinamento e teste (ALDUGHAYFIQ *et al.*, 2023; ANDERSEN *et al.*, 2019; ALDHYANI; ALKAHTANI, 2021). Para esse ajuste, utilizou-se a biblioteca *Keras Tuner* (O’MALLEY *et al.*, 2019).

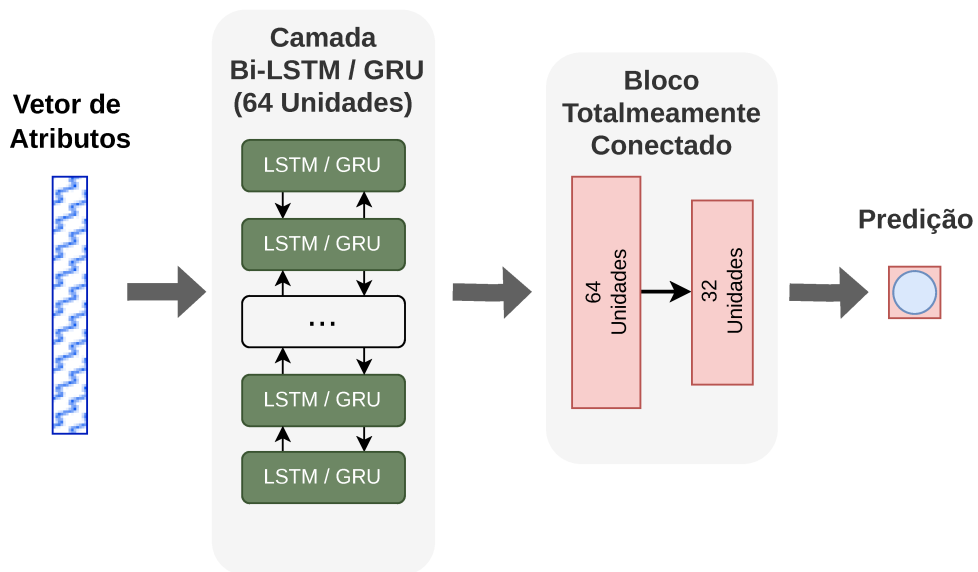
A busca pelos hiperparâmetros foi realizada sobre a base de treinamento subdividida, composta por 64% dos dados totais, enquanto a base de validação, com 16% dos dados, foi utilizada para avaliar os modelos durante o ajuste, especificamente, utilizou-se o conjunto de atributos do vetor *fval*. A função *BayesianOptimization* foi utilizada para otimizar o desempenho dos modelos. A busca foi limitada a um período de 36 horas, com um máximo de 100 épocas por modelo. A Tabela 4 apresenta o intervalo dos hiperparâmetros explorados, ademais, para cada camada, houve otimização do número de unidades considerando o intervalo apresentado. Devido a limitações computacionais, foram testadas somente 20 combinações de modelos dentro desse intervalo.

Tabela 4 – Intervalo de busca dos hiperparâmetros para as RNNs

Hiperparâmetro	Mínimo	Máximo	Passo
Unidades LSTM	64	256	32
Unidades Densas 2	16	256	16
Quantidade de Camadas BI-LSTM	1	3	1
Quantidade de Camadas Totalmente Conectado	1	3	1
Taxa de <i>Dropout</i>	0,1	0,5	0,05
Regularização L2	$1 \times 10^{-5}$	$1 \times 10^{-3}$	Log
Taxa de Aprendizado	{ 1e-3, 1e-4, 5e-5 }		
Função de Ativação	ReLU, Swish		

A arquitetura que apresentou o melhor desempenho consiste em: um vetor de entrada, uma camada bidirecional de LSTM ou GRU com 64 neurônios, uma camada de *Dropout* com magnitude de 0,3, que desativa aleatoriamente 30% dos neurônios para reduzir o *overfitting*; duas camadas densamente conectadas, com 64 e 32 unidades, respectivamente. Por fim, uma camada de saída com um único neurônio, utilizando a função de ativação *sigmoid* para realizar a predição.

Figura 6 – Arquitetura utilizada para o modelo LSTM e GRU

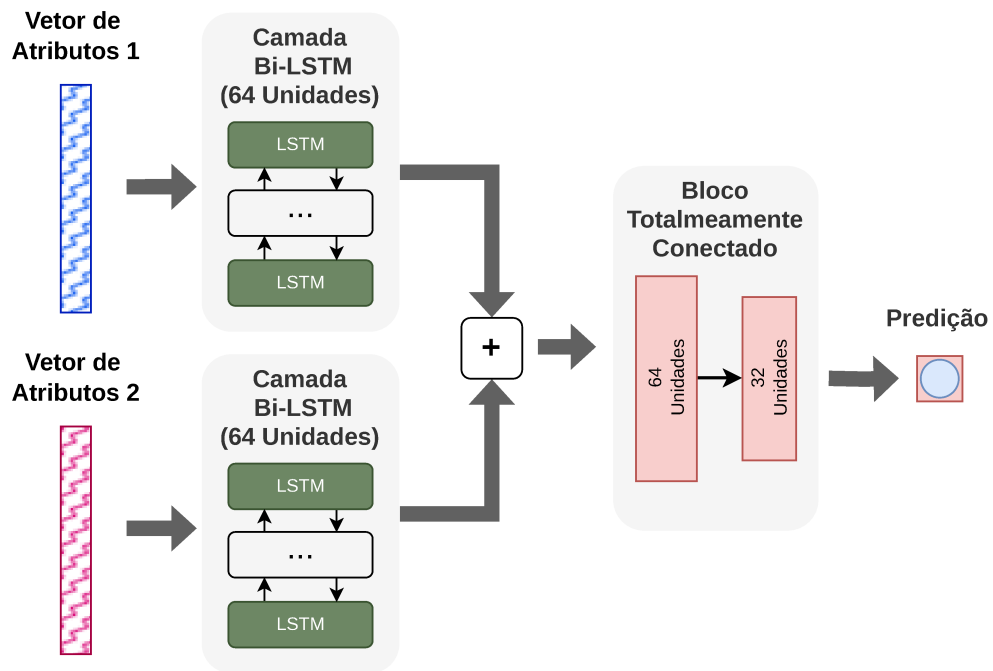


Fonte: Autor

Para alguns dos vetores de atributos, pensou-se necessário utilizar um bloco BI-LSTM para cada vetor de entrada, uma vez que os vetores possuem estruturas diferentes e utilizar um bloco para cada pode extrair mais informação. A estrutura seguiu o mesmo princípio da arquitetura anterior, mas com as saídas de cada camada BI-LSTM concatenadas e conectadas ao bloco densamente conectado, como ilustrado na Figura 7.

Em todos os modelos de RNN desenvolvidos neste trabalho, foi utilizado o otimizador *Adam* com taxa de aprendizado inicial de 0,001 e a função de perda de entropia binária cruzada. Além disso, definiu-se um máximo de 100 épocas, com lotes de 128 amostras por iteração. Para monitorar o treinamento, foi implementada uma função de ajuste programado da taxa de aprendizado. Caso a perda de validação permanecesse estável por 3 épocas consecutivas, a taxa de aprendizado era reduzida para 20% do valor atual. Também foi utilizada uma função de parada antecipada, ativada quando a perda não apresentava melhora por 5 épocas consecutivas.

Figura 7 – Arquitetura utilizada para o modelo BI-LSTM



Fonte: Autor

O modelo, então, era restaurado para a época com menor perda registrada. Estes valores foram encontrados baseando-se na literatura, ajustados mediante experimentações (ALDUGHAYFIQ *et al.*, 2023; ALDHYANI; ALKAHTANI, 2021).

## 6 RESULTADOS

Este capítulo apresenta os resultados obtidos durante os testes. A Seção 6.1 apresenta uma análise dos atributos, a Seção 6.2 descreve os resultados obtidos utilizando valores escalares do BTD e o histograma do RRI, enquanto a seção 6.3 investiga a utilização de atributos vetoriais derivados da convergência do BTD, explorando uma abordagem que, até o momento, não havia sido testada na literatura.

### 6.1 Análise dos atributos

Nesta seção, analisam-se os atributos de convergência gerados pela decomposição BTD com o objetivo de compreender sua relevância para a classificação de FA. Foram realizados testes de hipótese, cálculo da importância dos atributos e análise de correlação, cujos resultados são apresentados e discutidos a seguir.

Inicialmente, aplicou-se o Teste U de Mann-Whitney para identificar diferenças significativas entre as classes de FA e RSN. Os resultados, apresentados pela Tabela 5, mostraram que quatro dos sete atributos analisados rejeitaram a hipótese nula, indicando variância significativa entre as classes, nominalmente, *relfval*, *relerr*, *fval* e *iterations*, estes atributos estão diretamente relacionados à qualidade e à eficiência do processo de convergência.

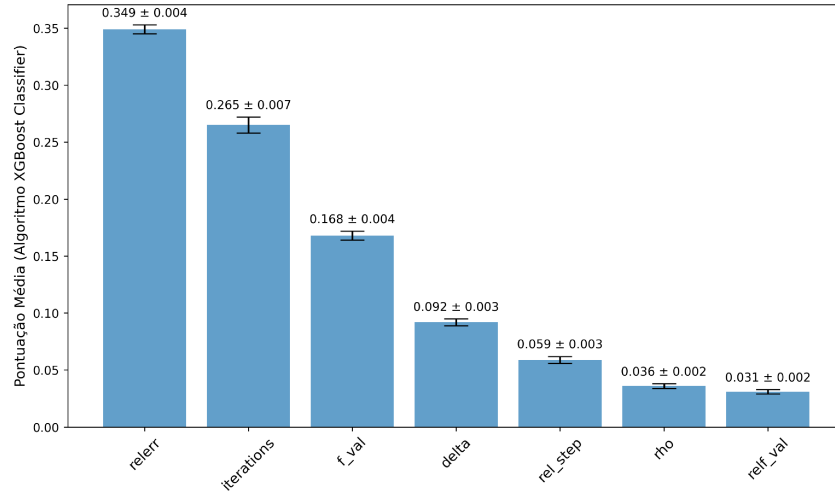
Tabela 5 – Resultados do Teste U de Mann-Whitney dos Atributos Comparados

Atributo	H <sub>0</sub>	p-valor
<i>relfval</i>	Rejeitada	$< 1 \times 10^{-6}$
<i>relerr</i>	Rejeitada	$< 1 \times 10^{-6}$
<i>fval</i>	Rejeitada	$< 1 \times 10^{-6}$
<i>relstep</i>	Não Rejeitada	0.5916
<i>delta</i>	Não Rejeitada	1.0
<i>rho</i>	Não Rejeitada	1.0
<i>iterations</i>	Rejeitada	$< 1 \times 10^{-6}$

Complementando essa análise, utilizou-se o algoritmo XGBoost para calcular a importância média relativa de cada atributo. A Figura 8 ilustra a importância e desvio padrão de cada modelo em ordem decrescente. Observou-se que *relerr*, que quantifica o erro relativo entre o tensor original e o modelo estimado, apresentou a maior relevância, seguido por *iterations* e *fval*, estes resultados são consistentes com o teste de hipótese. Ademais, o atributo *relfval*, apesar de apresentar variação significativa no teste de hipótese, exibiu baixo ganho de informação, enquanto apresenta alta covariância com *relstep*, conforme apresentado pela matriz de correlação

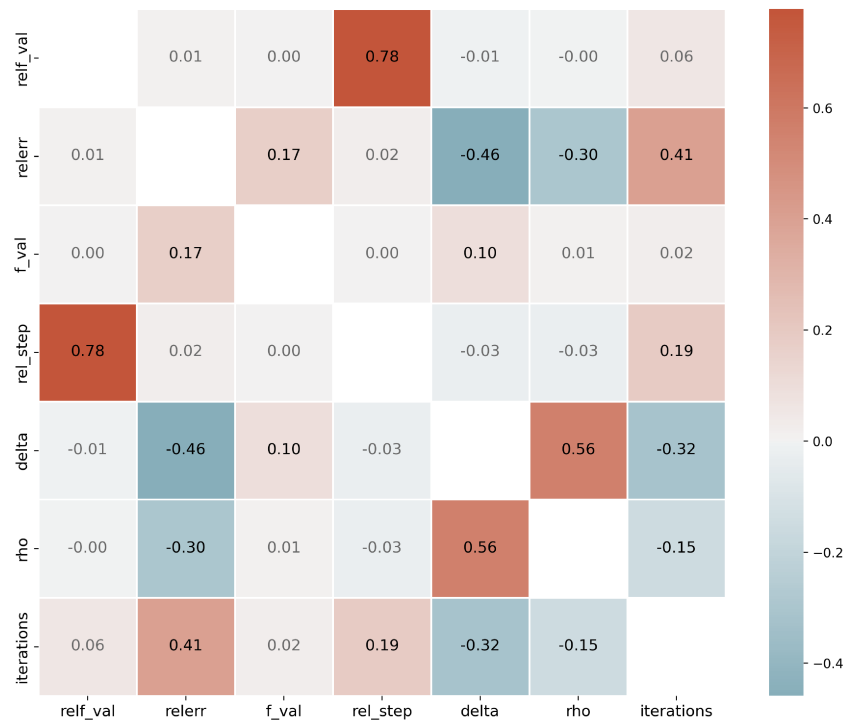
de Pearson na Figura 9, onde cores avermelhadas apresentam uma covariância positiva maior, enquanto próximas do azul uma covariância negativa.

Figura 8 – Média da importância dos atributos de convergência da BTM



Fonte: Autor

Figura 9 – Matriz de Correlação de Pearson dos atributos de convergência da BTM



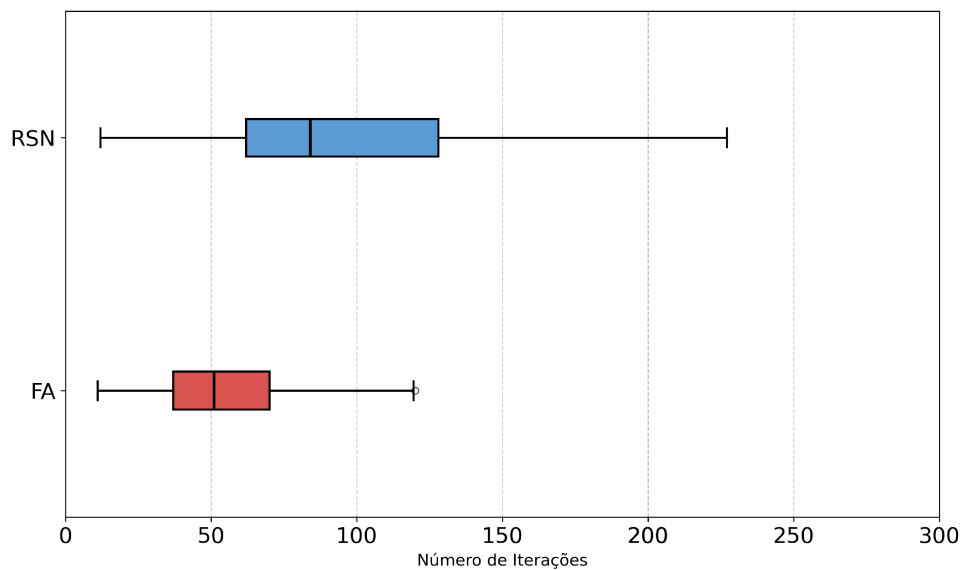
Fonte: Autor

A matriz revela correlações importantes entre os atributos. Por exemplo, *relerr*, *fval* e *iterations* exibem baixa ou moderada correlação entre si, indicando que fornecem informações complementares para a classificação. Por outro lado, a alta correlação entre *relfval* e *relstep*

justifica o menor impacto individual de *relfval* na classificação.

Por fim, a Figura 10 mostra algumas estatísticas descritivas do número de iterações para cada classe. Evidenciado pela Figura 10, os sinais de RSN apresentaram valores significativamente maiores em comparação aos sinais de FA, o que reflete diferenças na complexidade de convergência entre as classes. Essas informações são cruciais para a definição de configurações apropriadas na representação vetorial dos dados.

Figura 10 – Diagrama de caixa do número de iterações da BTM por classe



Fonte: Autor

Com base nessas análises, foi possível identificar os atributos mais informativos e definir parâmetros relevantes para a construção de modelos de classificação otimizados, utilizando exclusivamente os atributos de convergência da BTM. Conclui-se que os atributos gerados são propícios para classificação da FA, por apresentarem baixa covariância e um ganho de informação satisfatório.

## 6.2 Resultado com valores escalares

Feita a análise dos atributos, apresentamos os resultados dos modelos treinados utilizando os atributos contendo os valores escalares dos parâmetros de convergência da BTM. Ressalta-se que estes testes foram realizados de forma semelhante por Cardoso (2024), que encontrou resultados satisfatórios com ACC, SE e SPE de aproximadamente 97%. Assim, inicialmente, os modelos foram treinados com os parâmetros padrões e, posteriormente, ajustados para o melhor modelo, que será discutido. Os modelos foram avaliados na base de teste.

A fim de avaliar a influência de cada conjunto de atributos, apresentaremos os resultados utilizando apenas os atributos da BTD, o histograma dos RRI e, finalmente, o resultado dos modelos treinados com ambos os conjuntos de dados. Os modelos foram treinados na sub-base de treino, contendo 60% dos dados, e avaliada na sub-base de teste que contém 20%, a sub-base de validação foi utilizada apenas para acompanhar os modelos de aprendizado profundo. A Tabela 6 mostra a ACC, PRE, SE e F1-Score, obtidos pelos modelos treinados exclusivamente com os parâmetros de convergência da BTD.

Observa-se que o modelo XGBoost apresentou o melhor desempenho em todas as métricas avaliadas, seguido pelos modelos baseados em RNNs. Entre os modelos de aprendizado profundo, o GRU obteve uma diferença de apenas 0,22% no F1-Score em relação ao XGBoost, enquanto modelos baseados em técnicas mais simples, como regressão logística ou Naive Bayes, apresentaram os piores resultados.

Tabela 6 – Resultados usando indicadores BTD

Modelo	ACC(%)	PRE(%)	SE(%)	F1-Score(%)
Random Forest	93,15	94,82	94,90	94,86
Extra Trees	92,68	93,96	95,12	94,54
<b>XGBoost</b>	<b>93,60</b>	<b>95,27</b>	<b>95,10</b>	<b>95,18</b>
LightGBM	92,73	94,27	94,84	94,55
Gradient Boosting	89,33	91,35	92,76	92,05
KNN	85,74	89,23	89,36	89,30
Logistic Regression	78,29	80,72	88,53	84,44
Naive Bayes	76,94	81,14	85,16	83,10
Bi-LSTM	93,28	95,21	94,65	94,93
GRU	93,33	95,35	94,60	94,97

Em seguida, a Tabela 7 apresenta os resultados dos modelos utilizando apenas o histograma dos RRI. Observa-se que todos os modelos apresentaram um desempenho inferior ao dos modelos com os parâmetros da BTD, com todos os modelos apresentando resultados praticamente idênticos em termos de F1-Score. O modelo Extra Trees obteve o melhor desempenho, seguido pelos modelos baseados em árvore de decisão. Os modelos baseados em RNN mantiveram sua competitividade, apresentando bons resultados comparados aos outros modelos.

Por fim, a Tabela 8 apresenta os resultados dos modelos treinados com a base composta pelos indicadores da BTD e o histograma dos RRIs. Notamos uma ligeira melhoria em todos os modelos, sendo o XGBoost o modelo com o melhor desempenho, apresentando uma melhoria de 2,02% na ACC e 1,51% no F1-Score ao adicionar o histograma dos RRI. Modelos baseados em RNN ocuparam a segunda, terceira e quarta posição, apresentando resultados

Tabela 7 – Resultados usando histograma do RRI

<b>Modelo</b>	<b>ACC(%)</b>	<b>PRE(%)</b>	<b>SE(%)</b>	<b>F1-Score(%)</b>
Random Forest	81,34	83,43	89,81	86,50
<b>Extra Trees</b>	<b>81,31</b>	<b>83,07</b>	<b>90,33</b>	<b>86,55</b>
XGBoost	81,33	83,58	89,55	86,46
LightGBM	80,22	81,71	90,57	85,91
Gradient Boosting	76,97	77,24	92,74	84,28
KNN	79,19	82,44	87,35	84,82
Logistic Regression	75,40	77,32	89,20	82,84
Nayve Bayes	69,53	81,97	69,53	75,24
Bi-LSTM	81,34	83,72	89,34	86,44
GRU	81,30	83,28	89,97	86,50

bastante similares entre si. Para sintetizar, a Figura 11 mostra um comparativo entre os melhores modelos de cada conjunto de dados.

Tabela 8 – Resultados usando indicadores BTM e histograma do RRI

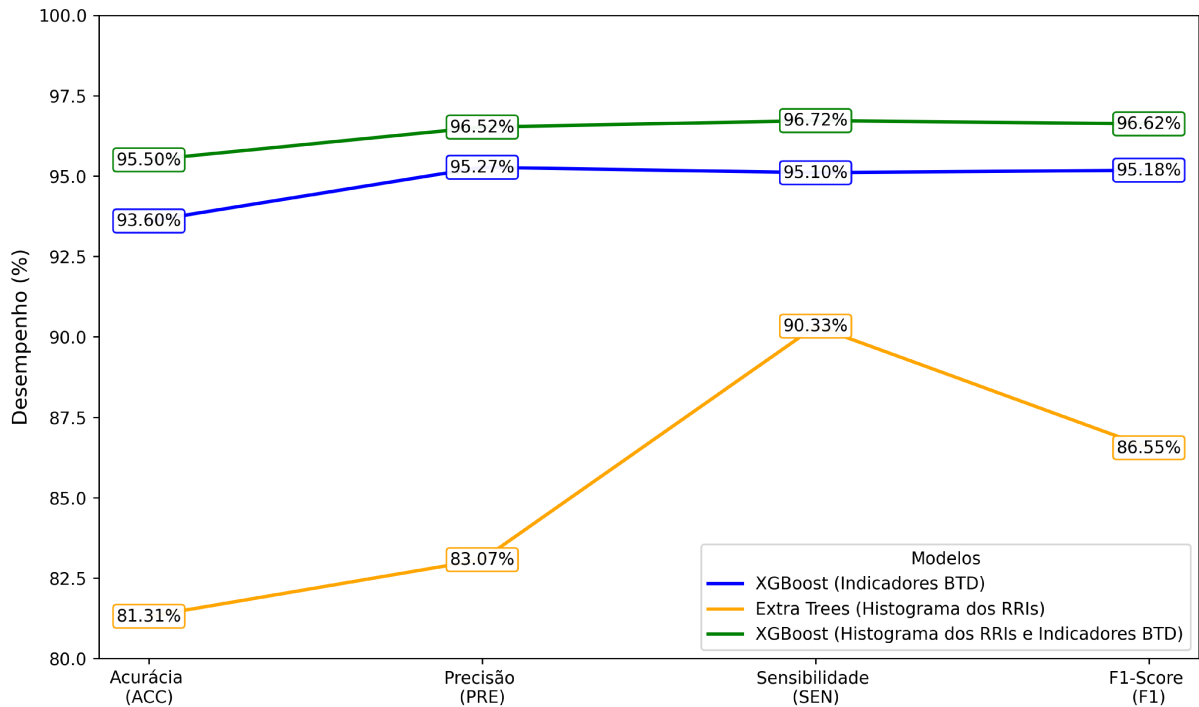
<b>Modelo</b>	<b>ACC(%)</b>	<b>PRE(%)</b>	<b>SE(%)</b>	<b>F1-Score(%)</b>
Random Forest	94,45	95,50	96,19	95,85
Extra Trees	91,98	92,45	95,77	94,08
<b>XGBoost</b>	<b>95,50</b>	<b>96,52</b>	<b>96,72</b>	<b>96,62</b>
LightGBM	94,21	95,42	95,90	95,66
Gradient Boosting	90,25	92,67	92,68	92,68
KNN	87,69	90,64	90,90	90,77
Logistic Regression	83,03	84,18	91,75	87,80
Nayve Bayes	77,24	85,66	79,04	82,22
Bi-LSTM	95,35	96,53	96,48	96,51
Bi-LSTM 2 entradas	95,42	96,79	96,52	96,53
GRU	95,14	96,34	96,35	96,35

Entre os modelos avaliados, o XGBoost obteve o melhor desempenho geral, apresentando os melhores resultados em Acurácia, Precisão, Recall e F1-Score para as bases com BTM e BTM com histograma do RRI. Quando analisada a base com apenas os histogramas dos RRIs, o modelo GRU se destacou, com o modelo Extra Trees ocupando o primeiro lugar e o GRU em segundo.

### 6.2.1 Ajuste de Hiperparâmetros do Melhor Caso

Considerando que o melhor desempenho foi obtido com os parâmetros da BTM e histograma dos RRIs, esse cenário foi escolhido para o ajuste dos hiperparâmetros. Foram selecionados os três melhores modelos para o ajuste fino: XGBoost, RF e LightGBM. O ajuste foi realizado utilizando a ferramenta Optuna, com validação cruzada de 10 dobras nas bases de

Figura 11 – Comparativo melhores modelos em cada uma das bases



Fonte: Autor

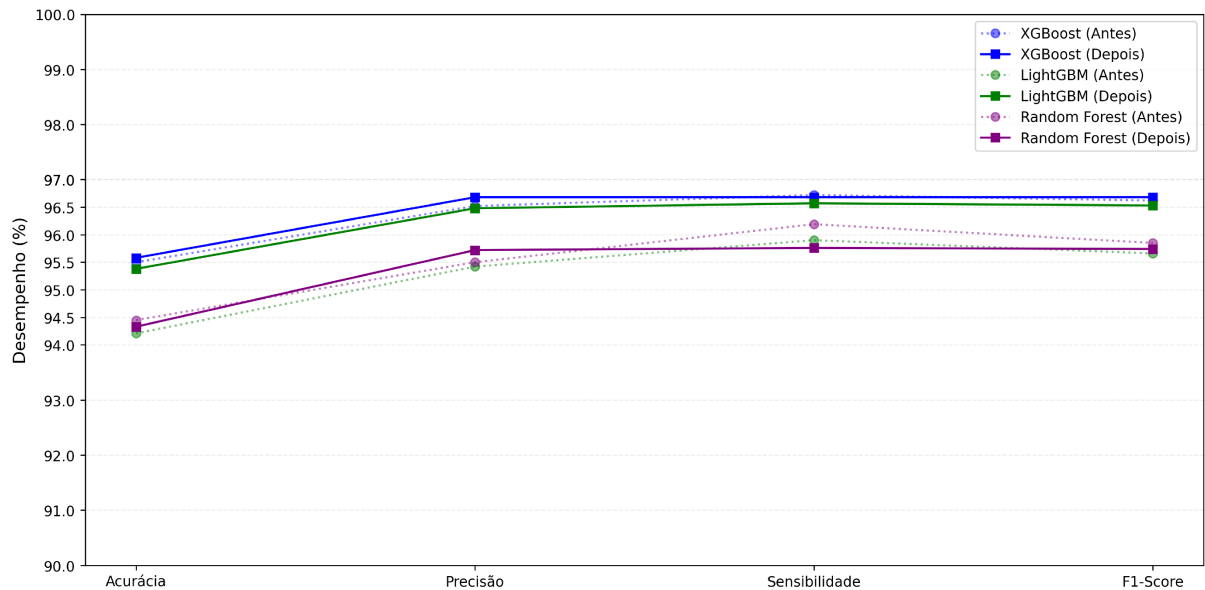
treino e validação. A Tabela 9 apresenta o intervalo de busca dos hiperparâmetros e os valores ajustados.

A Figura 12 apresenta os resultados pós-ajuste, avaliados na base de testes. Observa-se que, de forma geral, os modelos não apresentaram melhorias significativas no desempenho, sendo o LightGBM o que obteve a maior melhoria, embora não tenha superado o desempenho do XGBoost. O XGBoost manteve-se como o melhor modelo, com ACC de 95,58%, PRE de 96,68%, SE de 96,68% e F1-Score de 96,68%, representando uma melhora de aproximadamente 0,06%.

### 6.3 Resultados com Valores Vetoriais

Conforme mencionado anteriormente, alguns indicadores de convergência da BTD são representados como vetores. Para explorar de forma mais aprofundada o potencial de classificação desses indicadores, realizamos os testes a seguir utilizando alguns dos vetores de atributos, mais especificamente uma combinação dos atributos *fval* e *delta*. Esses atributos foram escolhidos por carregarem mais informações relevantes, uma vez que os atributos *relerr* e *iterations* são escalares. A hipótese é que uma rede neural consiga extrair mais informações ao utilizar um vetor de atributos em vez de valores escalares.

Figura 12 – Comparativo de resultados após o ajuste dos hiperparâmetros na base com indicadores do BTD e histograma do RRI



Fonte: Autor

Dessa forma, esta seção apresenta os experimentos realizados com atributos vetoriais, explorando uma abordagem ainda não investigada na literatura. Os resultados reforçam a

Tabela 9 – Intervalo de busca e parâmetros ajustados para os modelos

Modelo	Hiperparâmetro	Valores Candidatos	Ajustado
XGBoost	booster	{gbtree, gblinear, dart}	gbtree
	gamma	{0.0, 0.1, 0.5, 1.0}	0.0
	learning_rate	[0.01, 0.5] (log)	0.2303
	n_estimators	[5, 500]	486
	max_depth	[3, 15]	14
	reg_alpha	{0.0, 0.1, 1.0, 10.0}	0.0
	reg_lambda	{0.0, 0.1, 1.0, 10.0}	10.0
	min_child_weight	{1, 2, 3, 5}	5
	subsample	[0.6, 1.0]	0.7331
Random Forest	n_estimators	[50, 500]	174
	max_depth	{5, 10, 15, 20, 25, 30, 35, 40, None}	30
	min_samples_split	[2, 10]	9
	min_samples_leaf	[1, 4]	1
	max_features	{sqrt, log2, None}	None
	criterion	{gini, entropy}	entropy
	max_leaf_nodes	{None, 5, 10, 15, 20, 25, 30}	None
LightGBM	n_estimators	[50, 500]	321
	max_depth	{5, 10, 15, 20, 25, 30, 35, 40, None}	15
	learning_rate	[0.0001, 0.3] (log)	0.0640
	subsample	[0.6, 1.0]	0.8166
	colsample_bytree	[0.6, 1.0]	0.9722
	num_leaves	[10, 100]	69

viabilidade dessa abordagem, evidenciando o impacto da estrutura vetorial dos atributos no desempenho dos modelos e demonstrando que os indicadores de convergência da BTM, quando utilizados como vetores de atributos, podem fornecer informações suficientes para a classificação precisa de FA.

A Tabela 10 apresenta os resultados dos modelos treinados utilizando o indicador *fval* da BTM. Observa-se que o modelo BI-LSTM obteve o melhor desempenho, seguido pelo modelo GRU. A média das acurácias utilizando o parâmetro *fval* foi 0,59% inferior àquela obtida com os valores escalares. Quando comparado o melhor modelo utilizando todos os indicadores, o XGBoost, com o BI-LSTM, observamos uma redução de 0,64%.

Notamos também que todos os modelos parecem ter alcançado um limite de extração de informação para o atributo *fval*, dado que os resultados entre os modelos, como acurácia, mostraram pouca diferença. Isso sugere que a complexidade das arquiteturas não resultou em ganhos significativos ao processar somente este atributo.

Tabela 10 – Resultados usando o vetor do indicador *fval* da BTM

Modelo	ACC(%)	PRE(%)	SE(%)	F1-Score(%)
Random Forest	92,05	94,26	93,77	94,01
Extra Trees	90,49	92,67	93,07	92,87
XGBoost	92,23	94,45	93,84	94,14
LightGBM	91,15	93,67	92,99	93,33
Gradient Boosting	89,71	91,30	93,45	92,36
KNN	90,69	93,69	92,22	92,95
Logistic Regression	80,58	80,88	92,74	86,41
Nayve Bayes	79,41	79,98	92,13	85,63
<b>Bi-LSTM</b>	<b>92,95</b>	<b>95,31</b>	<b>94,03</b>	<b>94,67</b>
GRU	92,74	94,94	94,11	94,52

Para melhorar os resultados, realizamos um novo teste combinando o vetor de *fval* com o vetor *delta*, escolhido por representar o segundo maior ganho de informação. Ambos os vetores foram combinados, com cada um tendo 150 elementos, totalizando 300 pontos ou atributos. Os modelos testados foram os mesmos dos testes anteriores, mas dessa vez incluímos uma variação da rede BI-LSTM, com uma arquitetura de entrada dupla, conforme descrito na Figura 7. Nessa configuração, cada vetor (*fval* e *delta*) é processado separadamente por uma "sub-rede"BI-LSTM, e suas saídas são concatenadas em uma camada densamente conectada para a classificação.

A Tabela 11 mostra os resultados obtidos com os vetores *fval* e *delta*. O melhor desempenho foi alcançado pelo modelo BI-LSTM com uma entrada dupla, seguido pelo modelo

BI-LSTM com uma entrada única. O melhor modelo nesta combinação apresentou uma diferença de apenas 0,36% em relação ao melhor desempenho obtido com o XGBoost usando os indicadores da BTD e RRI.

Tabela 11 – Resultados usando os vetores dos indicadores *fval* e *delta* da BTD

<b>Modelo</b>	<b>ACC(%)</b>	<b>PRE(%)</b>	<b>SE(%)</b>	<b>F1-Score(%)</b>
Random Forest	92,82	94,57	94,65	94,61
Extra Trees	91,10	92,72	94,01	93,36
XGBoost	93,66	95,43	95,03	95,23
LightGBM	91,91	93,92	93,93	93,93
KNN	90,85	93,83	92,32	93,07
Logistic Regression	81,37	81,97	92,32	86,83
Nayve Bayes	81,23	85,23	86,84	86,03
<b>Bi-LSTM - 2 Entradas</b>	<b>95,13</b>	<b>96,62</b>	<b>96,04</b>	<b>96,33</b>
Bi-LSTM - 1 Entrada	94,00	96,07	94,87	95,47
GRU	94,59	96,37	95,46	95,91

Em busca de melhorar ainda mais os resultados, combinamos o atributo mais informativo, *relerr*, aos vetores *fval* e *delta*. A Tabela 12 apresenta os resultados dessa combinação. Observamos que, de maneira geral, os modelos baseados em redes neurais não apresentaram melhorias em relação ao experimento anterior. O modelo com melhor desempenho foi o BI-LSTM com uma entrada, seguido por uma leve melhoria nos modelos baseados em técnicas tradicionais, embora essa melhoria tenha sido quase insignificante.

Tabela 12 – Resultados usando os vetores dos indicadores *fval*, *delta* e *relerr* da BTD

<b>Modelo</b>	<b>ACC(%)</b>	<b>PRE(%)</b>	<b>SE(%)</b>	<b>F1-Score(%)</b>
Random Forest	93,06	94,88	94,68	94,78
Extra Trees	91,82	93,37	94,41	93,89
XGBoost	93,78	95,55	95,09	95,32
LightGBM	92,30	94,38	94,03	94,21
KNN	90,85	93,83	92,32	93,07
Logistic Regression	81,42	82,02	92,34	86,87
Nayve Bayes	81,40	85,41	86,91	86,15
Bi-LSTM - 2 Entradas	94,62	96,15	95,74	95,95
<b>Bi-LSTM - 1 Entrada</b>	<b>94,81</b>	<b>96,38</b>	<b>95,80</b>	<b>96,09</b>
GRU	93,93	95,80	95,05	95,42

Finalmente, combinamos o vetor *fval* com o histograma dos RRI. A Tabela 13 apresenta os resultados dessa combinação. Observa-se uma melhoria considerável, com o melhor desempenho obtido pelo modelo BI-LSTM com entradas duplas. O resultado alcançado se aproxima dos melhores resultados obtidos ao utilizar a combinação de valores escalares da BTD

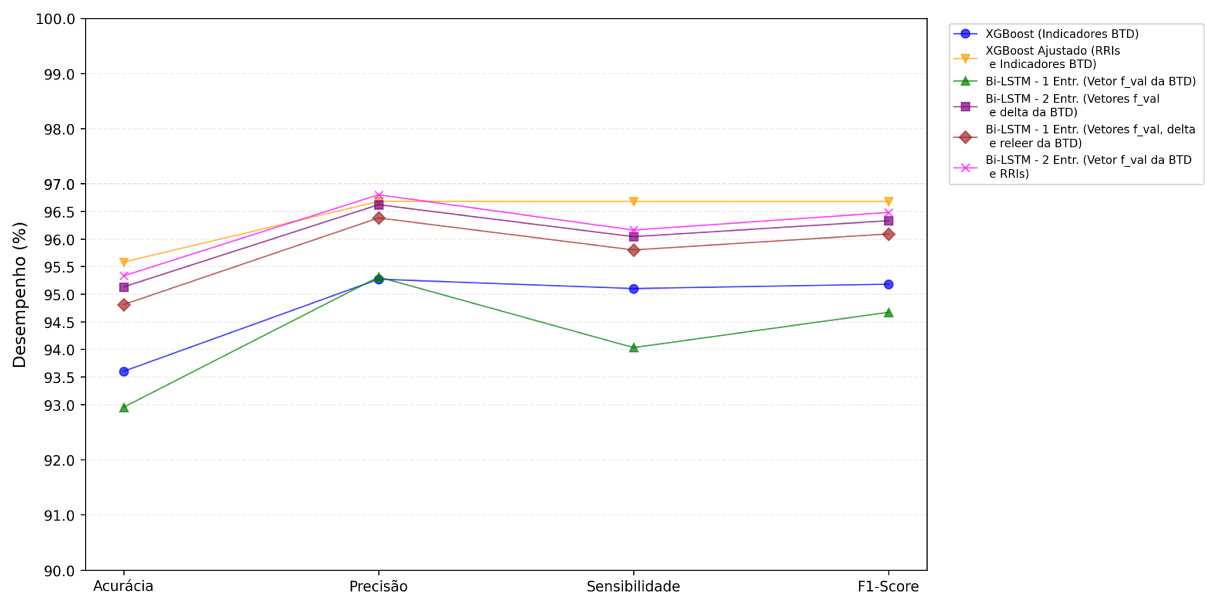
com o histograma dos RRIs.

Tabela 13 – Resultados usando o vetor do indicador *fval* da BTD e histograma dos RRIs

Modelo	ACC(%)	PRE(%)	SE(%)	F1-Score(%)
Random Forest	93,30	95,02	94,91	94,97
Extra Trees	92,67	93,94	95,11	94,52
XGBoost	94,57	95,95	95,89	95,92
LightGBM	93,09	94,82	94,80	94,81
KNN	90,69	93,69	92,23	92,95
Logistic Regression	81,21	81,58	92,70	86,79
Nayve Bayes	81,87	82,44	92,46	87,16
<b>Bi-LSTM - 2 Entradas</b>	<b>95,33</b>	<b>96,80</b>	<b>96,16</b>	<b>96,48</b>
Bi-LSTM - 1 Entrada	95,22	96,70	96,09	96,39
GRU	94,59	96,30	95,54	95,92

Concluimos que o modelo BI-LSTM com duas entradas obteve os melhores resultados em termos de acurácia e F1-Score ao utilizar essa base, aproximando-se dos melhores desempenhos obtidos com a combinação dos indicadores da BTD e RRIs, como pode ser observado pela Figura 13, que compila todos os melhor reportados neste trabalho. Nota-se ainda, que todos os resultados apresentam desempenho bem similar.

Figura 13 – Comparativo dos melhores casos em cada uma das possíveis bases



Fonte: Autor

## 7 DISCUSSÃO

Os resultados apresentados nas seções anteriores demonstram que, mesmo utilizando somente os indicadores de convergência da BTM, foi possível alcançar desempenho significativo na tarefa de classificação, aproximando-se de resultados obtidos com combinação de outros atributos, como o histograma dos RRIs. Este achado sugere que os indicadores de convergência da BTM, em sua forma escalar, já contêm informações robustas o suficiente para permitir classificações de alta qualidade, mas ao recorrer a fontes adicionais de dados, como trabalhar com os vetores ao invés do valor escalar é possível conseguir uma pequena melhora nos resultados.

A abordagem com apenas os parâmetros da BTM resulta em modelos competitivos, como exemplificado pelo desempenho do modelo XGBoost na base que utiliza os valores escalares dos indicadores, ou até mesmo o modelo BI-LSTM de duas entradas com os vetores *fval* e *delta*, que alcançou resultados similares à combinação com histograma dos RRIs, que se destacou como o melhor nas avaliações. Ademais, não somente estes resultados são similares, uma vez que quase todos os melhores casos apresentaram resultados muito parecidos, como pode ser visto na Figura 13, uma possível explicação é que os atributos gerados apresentam uma boa consistência e separabilidade, no entanto, esta separação parece atingir seu limite com aproximadamente 95% de acurácia. Esta ideia é forçada pelo comparativo com o ajuste dos hiperparâmetros (Figura 12), no qual os resultados virtualmente não apresentam melhoria.

Dessa forma, a próxima etapa consiste em posicionar os resultados obtidos neste estudo frente às abordagens presentes na literatura, a fim de avaliar a relevância e as contribuições dos métodos propostos em comparação com outros trabalhos.

### 7.1 Comparativo com a Literatura

A Tabela 14 apresenta uma comparação detalhada entre os resultados obtidos neste trabalho e outros estudos relevantes da literatura que investigaram a classificação de FA. Entre os trabalhos analisados, destaca-se o de Duan *et al.* (2022), que utiliza histogramas de RRIs como principal atributo em combinação com um modelo baseado em SVM. Apesar do desempenho elevado, é importante ressaltar que os testes realizados por Duan *et al.* (2022) utilizaram uma base externa composta exclusivamente por sinais positivos de FA, o que pode ter enviesado os resultados ao reduzir a complexidade do problema de classificação. Em contraste, o presente trabalho empregou bases que incluem tanto sinais normais quanto episódios de FA, aumentando

a robustez e generalização dos modelos propostos.

Outro trabalho relevante é o de Cardoso (2024), que combina histogramas de RRI com indicadores de BTM e utiliza o modelo LightGBM, alcançando métricas ligeiramente superiores às apresentadas neste estudo. Embora os resultados do presente trabalho não superem diretamente esse estudo, é notável que, ao utilizar exclusivamente os vetores de BTM, os modelos propostos foram capazes de atingir um desempenho próximo, reforçando a relevância desses atributos na classificação precisa de FA.

Tabela 14 – Comparativo com resultados da literatura

Autor	Atributo	Modelo	Bases de Dados	Janela	ACC(%)	PRE(%)	SEN(%)
Andersen <i>et al.</i> (2019)	ECG (Onda original)	CNN+LSTM	AFDB, MITDB, NSRDB	30 RRI	87,40	45,45	86,04
Czabanski <i>et al.</i> (2020)	Variabilidade Cardíaca	SVM	AFDB	20 RRI	98,60	-	98,90
Hirsch <i>et al.</i> (2021)	RRI e AA	Random Forest	AFDB	29 RRI	95,90	97,36*	96,10
Jahan <i>et al.</i> (2022)	RRI (Domínio do Tempo)	AdaBoost	AFDB, LTAfDB, NSRDB, MITDB	19 RRI	83,46*	90,85	87,58
Duan <i>et al.</i> (2022)	Histograma do RRI	SVM	AFDB, LTAfDB, NSRDB	29 RRI	96,97	95,24	99,94
Aldughayfiq <i>et al.</i> (2023)	ECG (Onda) e PPG	CNN+BI-LSTM	MIMIC PERform	-	95,00	88,00	85,00
Cardoso (2024)	BTM e Histograma do RRI	LightGBM	AFDB, LTAfDB, NSRDB	5 RRI	97,04	97,02	97,44
<b>Presente Trabalho</b>	BTM e Histograma do RRI	XGBoost	AFDB, LTAfDB, NSRDB	5 RRI	95,58	96,68	96,68
	Vetor da BTM	BI-LSTM	AFDB, LTAfDB, NSRDB	5 RRI	95,13	96,62	96,04

Fonte: Autor

Os resultados obtidos também mostram uma vantagem em relação a abordagens baseadas em redes neurais profundas, como a de Andersen *et al.* (2019), que utiliza modelos CNN+LSTM aplicados a sinais de ECG completos. Apesar de redes neurais geralmente apresentarem alta capacidade de generalização, o método de Andersen *et al.* (2019) exige sinais mais longos (30 RRI) e apresenta uma acurácia de 87,40%, e uma precisão de apenas 45,45%, consideravelmente inferior ao alcançado neste trabalho. Além disso, os modelos baseados em vetores de convergência da BTM demonstram que é possível obter alta precisão na classificação utilizando uma quantidade reduzida de RRI (apenas 5), representando um diferencial significativo em relação a métodos que requerem maior volume de dados.

O uso exclusivo de vetores de BTM neste trabalho destaca uma abordagem inovadora na literatura. Enquanto a maioria dos estudos utiliza combinações de atributos no domínio do tempo ou frequência, este trabalho explora as decomposições tensoriais como uma ferramenta eficaz para extrair informações essenciais dos sinais. Isso é evidenciado pela capacidade do modelo BI-LSTM de atingir métricas competitivas mesmo em cenários de curtíssima duração dos sinais. A classificação baseada em vetores de BTM não apenas confirma a relevância desses atributos para discriminar episódios de FA, mas também abre caminhos para a aplicação dessa metodologia em cenários onde dados de curta duração são predominantes, como em triagens rápidas ou monitoramento remoto.

Assim, o presente trabalho contribui para a literatura ao demonstrar que os vetores da BTM, isoladamente, são suficientes para realizar uma classificação precisa de FA. Essa abordagem contrasta com estudos que dependem de atributos adicionais ou sinais mais longos, evidenciando a robustez e a aplicabilidade dos modelos propostos em situações práticas e desafiadoras. Dessa forma, os resultados apresentados não apenas corroboram a eficácia dos métodos propostos, mas também posicionam a decomposição tensorial como uma técnica promissora para análise de sinais biomédicos. Com base no desempenho observado nos modelos propostos e sua comparação com trabalhos da literatura, torna-se pertinente uma análise mais detalhada do impacto isolado dos indicadores da BTM no desempenho de classificação.

## 7.2 O Desempenho com Indicadores da BTM

Ao considerarmos exclusivamente os indicadores da BTM, sem a combinação com outros parâmetros como o histograma dos RRIs, os resultados indicam que foi possível alcançar um desempenho bastante competitivo. A integração de dados adicionais, como os histogramas, embora tenha mostrado algumas melhorias em certos casos, não resultou em ganhos substanciais quando comparado aos resultados obtidos utilizando somente os indicadores da BTM. Este achado sugere que, em alguns cenários, a adição de mais variáveis não contribui, necessariamente, significativamente para a melhoria da classificação, e, por vezes, pode até resultar em *overfitting*, especialmente com modelos mais complexos.

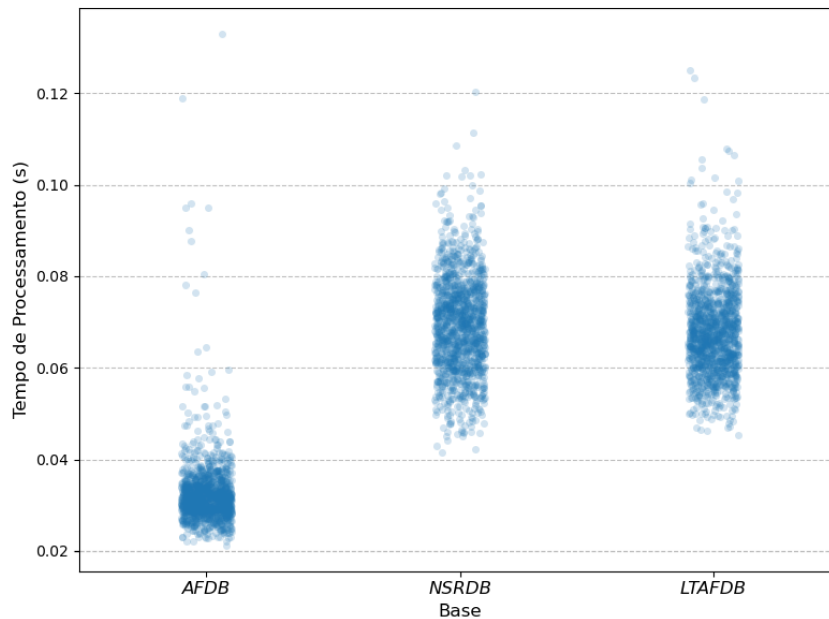
Ademais, o aumento na dimensionalidade com o uso de vetores e a introdução de parâmetros adicionais pode ser uma limitação, principalmente quando os modelos treinados com essas variáveis não apresentam melhorias significativas em relação aos modelos mais simples. O uso de técnicas como *Principal Component Analysis* (PCA), que permite a redução da dimensionalidade, aparece como uma solução interessante para contornar esse problema. Além disso, a otimização no número de parâmetros pode reduzir o risco de *overfitting* e melhorar a generalização do modelo.

## 7.3 Complexidade computacional do processo

A eficiência computacional é um importante fator ao se utilizar métodos baseados em decomposições tensoriais, especialmente em aplicações que demandam análise em tempo real ou processamento de grandes volumes de dados. Durante os experimentos, foi medido o

tempo necessário para processar cada segmento de RRIs utilizando a BTM. A Figura 14 ilustra o tempo médio de processamento por base de dados. Observa-se que a base AFDB apresentou o menor tempo médio (0,032s), enquanto as bases NSRDB e LTAfDB tiveram tempos semelhantes (0,068s). Esses resultados indicam que, mesmo processando cerca de 20 mil segmentos de 5 RRIs, o processamento completo de um sinal de 24 horas, levaria aproximadamente 22 minutos, demonstrando a viabilidade prática do método.

Figura 14 – Tempo levado para processar o segmento de RRI com BTM



Fonte: Autor

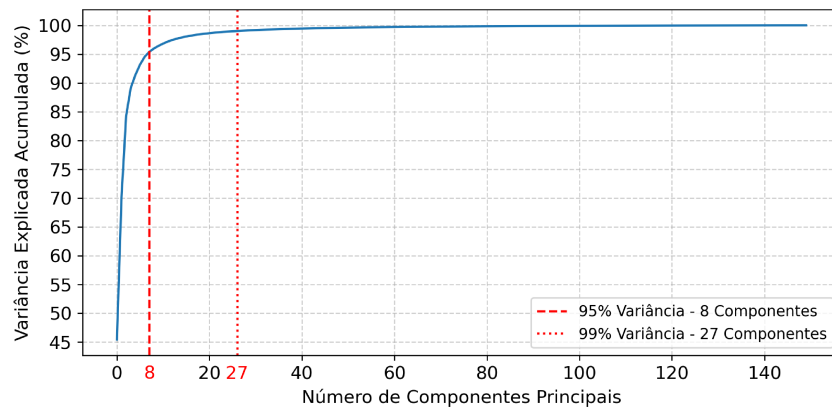
Uma vantagem importante deste modelo é sua eficiência em sinais de curtíssima duração. Por exemplo, em um exame típico de ECG de 10 segundos (com cerca de 12 RRIs), o tempo estimado para processar os segmentos e construir os atributos seria de apenas 0.16s. Esse aspecto torna o método atrativo para dispositivos portáteis e cenários de monitoramento remoto, onde rapidez e precisão são essenciais.

Adicionalmente, ao incorporar os vetores de atributos da BTM, como *fval* e *delta*, verificamos um aumento na complexidade computacional dos modelos, acompanhado de melhorias marginais no desempenho. Esse aumento é devido ao crescimento significativo no número de parâmetros, exigindo maior capacidade computacional. No entanto, a análise da variância explicada pelos vetores de atributos revela uma oportunidade de otimização. A Figura 15 mostra que 95% da variância do vetor *fval* pode ser explicada com apenas 8 componentes principais, enquanto 99% é alcançada com 27 componentes. Isso sugere que a redução dimensional via

PCA pode ser implementada para simplificar os modelos, reduzindo o custo computacional sem perdas significativas de desempenho.

Em particular, ao analisar a variância explicada pelos vetores de atributos  $f\_val$ , ficou claro que, utilizando somente um pequeno subconjunto desses atributos, já é possível explicar uma porção significativa da variabilidade nos dados. A Figura 15 mostra que 95% da variância do vetor  $f\_val$  pode ser explicada com apenas 8 elementos, e 99% com 27 elementos, utilizando a PCA. Isso sugere que uma otimização no número de parâmetros poderia reduzir significativamente a complexidade dos modelos, mantendo a eficácia, e abre a possibilidade de modelos mais leves e rápidos, sem perdas substanciais de desempenho.

Figura 15 – Variância explicada acumulada do vetor  $fval$  através do PCA



Fonte: Autor

Essa otimização é especialmente relevante ao considerar o impacto de modelos menores em sistemas de baixa capacidade, como dispositivos embarcados. Além disso, a baixa correlação entre os principais atributos  $fval$ ,  $relerr$ , e  $delta$  reforça a complementaridade das informações fornecidas, justificando a escolha desses atributos para representar o processo de convergência da BTD.

## 8 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho teve como objetivo investigar a viabilidade do uso de métricas de convergência da BTD aplicadas a sinais de ECG, explorando esses novos atributos em combinação com diversos modelos de AM para a classificação de FA. Ao longo do estudo, foram analisadas múltiplas combinações de atributos derivados dos indicadores de convergência da BTD e testados diferentes modelos de AM – desde abordagens mais simples até arquiteturas robustas baseadas em árvores e aprendizado profundo – para avaliar tanto o papel dos atributos extraídos via decomposição tensorial na distinção entre FA e RSN, quanto a influência do tipo de modelo empregado no desempenho da classificação.

Os resultados demonstraram que, ao utilizar modelos tradicionais com uma combinação de atributos provenientes dos intervalos RRI, é possível atingir uma acurácia próxima a 95%. Além disso, a aplicação exclusiva dos indicadores de convergência, quando combinada com um modelo baseado em RNN, apresentou desempenho similar – com precisão e sensibilidade compatíveis com o estado da arte. Esses achados indicam que os atributos propostos neste trabalho são promissores para a classificação robusta entre FA e RSN, mesmo quando aplicados a intervalos de ECG de curtíssima duração, contribuindo significativamente para o auxílio no diagnóstico da fibrilação atrial.

Apesar das contribuições, o estudo possui limitações que devem ser consideradas. Entre elas, destaca-se a ausência de ajuste mais refinado dos parâmetros dos modelos de aprendizado profundo. Adicionalmente, os testes foram realizados com uma proporção de amostras que não reflete exatamente as condições reais, apresentando um leve desbalanceamento em favor das amostras com FA. Tais restrições podem ser exploradas em pesquisas futuras, proporcionando aprimoramentos metodológicos e uma melhor adequação aos cenários reais de aplicação.

Nesse sentido, recomenda-se que trabalhos futuros investiguem a relação entre a complexidade dos modelos e a quantidade de informações contidas nos dados, a fim de identificar o ponto ótimo de complexidade. Essa análise incluiria a exploração do impacto do número de atributos no desempenho dos modelos e a verificação se abordagens mais simples podem oferecer resultados comparáveis aos de modelos mais complexos, desde que haja uma seleção criteriosa dos atributos, para isso pode-se utilizar técnicas como o PCA, removendo atributos que não acrescentem significativamente na variância explicada.

Além disso, torna-se relevante a avaliação da eficiência computacional dos métodos baseados em BTD, especialmente para aplicações que demandem respostas rápidas e baixo custo

de processamento. Por fim, o desenvolvimento de modelos híbridos, combinando RNNs com camadas convolucionais, pode representar uma estratégia promissora para aprimorar a precisão e a robustez da classificação de FA.

Conclui-se que este trabalho reforça a importância de explorar novos atributos e combinações de modelos para o avanço na classificação de fibrilação atrial em ECG. Espera-se que as contribuições aqui apresentadas sirvam de base para futuras investigações e que os caminhos sugeridos possam fomentar o desenvolvimento de soluções cada vez mais eficientes e precisas no diagnóstico dessa condição.

## REFERÊNCIAS

ACHARYA, U. R.; FUJITA, H.; ADAM, M.; Oh Shu Lih; Tan Jen Hong; SUDARSHAN, V. K.; KOH, J. E. Automated characterization of arrhythmias using nonlinear features from tachycardia ECG beats. In: **2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**. Budapest, Hungary: IEEE, 2016. p. 000533–000538. ISBN 978-1-5090-1897-0. Disponível em: <http://ieeexplore.ieee.org/document/7844294/>.

AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. **Optuna: A Next-generation Hyperparameter Optimization Framework**. arXiv, 2019. ArXiv:1907.10902 [cs]. Disponível em: <http://arxiv.org/abs/1907.10902>.

ALDHYANI, T. H. H.; ALKAHTANI, H. A Bidirectional Long Short-Term Memory Model Algorithm for Predicting COVID-19 in Gulf Countries. **Life**, v. 11, n. 11, p. 1118, out. 2021. ISSN 2075-1729. Disponível em: <https://www.mdpi.com/2075-1729/11/11/1118>.

ALDUGHAYFIQ, B.; ASHFAQ, F.; JHANJHI, N. Z.; HUMAYUN, M. A Deep Learning Approach for Atrial Fibrillation Classification Using Multi-Feature Time Series Data from ECG and PPG. **Diagnostics**, v. 13, n. 14, p. 2442, jan. 2023. ISSN 2075-4418. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute. Disponível em: <https://www.mdpi.com/2075-4418/13/14/2442>.

ANDERSEN, R. S.; PEIMANKAR, A.; PUTHUSSERYPADY, S. A deep learning approach for real-time detection of atrial fibrillation. **Expert Systems with Applications**, v. 115, p. 465–473, jan. 2019. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417418305190>.

ASHLEY, E. A.; NIEBAUER, J. Conquering the ECG. In: **Cardiology Explained**. Remedica, 2004. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK2214/>.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. ISSN 08856125. Disponível em: <http://link.springer.com/10.1023/A:1010933404324>.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification And Regression Trees**. 1. ed. Routledge, 2017. ISBN 978-1-315-13947-0. Disponível em: <https://www.taylorfrancis.com/books/9781351460491>.

BRUNDEL, B. J. J. M.; AI, X.; HILLS, M. T.; KUIPERS, M. F.; LIP, G. Y. H.; GROOT, N. M. S. de. Atrial fibrillation. **Nature Reviews Disease Primers**, v. 8, n. 1, p. 1–23, abr. 2022. ISSN 2056-676X. Publisher: Nature Publishing Group. Disponível em: <https://www.nature.com/articles/s41572-022-00347-9>.

CARDOSO, R. H. **Detecção de fibrilação atrial em ECG com o uso da decomposição tensorial BTD**. Dissertação (Dissertação de Mestrado) – Universidade Federal do Ceará, Sobral, CE, Brasil, 2024. Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Campus de Sobral. Disponível em: <http://repositorio.ufc.br/handle/riufc/78739>.

CARROLL, J. D.; CHANG, J.-J. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. **Psychometrika**, v. 35, n. 3, p. 283–319, set. 1970. ISSN 1860-0980. Disponível em: <https://doi.org/10.1007/BF02310791>.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 785–794. ISBN 9781450342322. Disponível em: <https://doi.org/10.1145/2939672.2939785>.

CHO, K.; MERRIENBOER, B. van; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. **Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation**. arXiv, 2014. Version Number: 3. Disponível em: <https://arxiv.org/abs/1406.1078>.

CZABANSKI, R.; HOROBA, K.; WROBEL, J.; MATONIA, A.; MARTINEK, R.; KUPKA, T.; JEZEWSKI, M.; KAHANKOVA, R.; JEZEWSKI, J.; LESKI, J. Detection of Atrial Fibrillation Episodes in Long-Term Heart Rhythm Signals Using a Support Vector Machine. **Sensors**, v. 20, n. 3, p. 765, jan. 2020. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/20/3/765>.

DELANEY, J. A.; YIN, X.; FONTES, J. D.; WALLACE, E. R.; SKINNER, A.; WANG, N.; HAMMILL, B. G.; BENJAMIN, E. J.; CURTIS, L. H.; HECKBERT, S. R. Hospital and clinical care costs associated with atrial fibrillation for Medicare beneficiaries in the Cardiovascular Health Study and the Framingham Heart Study. **SAGE Open Medicine**, v. 6, p. 205031211875944, jan. 2018. ISSN 2050-3121, 2050-3121. Disponível em: <http://journals.sagepub.com/doi/10.1177/2050312118759444>.

DUAN, J.; WANG, Q.; ZHANG, B.; LIU, C.; LI, C.; WANG, L. Accurate detection of atrial fibrillation events with R-R intervals from ECG signals. **PLOS ONE**, v. 17, n. 8, p. e0271596, ago. 2022. ISSN 1932-6203. Publisher: Public Library of Science. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0271596>.

FAUST, O.; CIACCIO, E. J.; ACHARYA, U. R. A Review of Atrial Fibrillation Detection Methods as a Service. **International Journal of Environmental Research and Public Health**, v. 17, n. 9, p. 3093, abr. 2020. ISSN 1660-4601. Disponível em: <https://www.mdpi.com/1660-4601/17/9/3093>.

GARCIA, T.; HOLTZ, N. **12 Lead ECG: The Art of Interpretation**. Langara College, 2012. Disponível em: <https://books.google.com.br/books?id=-u-ozwEACAAJ>.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to Forget: Continual Prediction with LSTM. **Neural Computation**, v. 12, n. 10, p. 2451–2471, out. 2000. ISSN 0899-7667, 1530-888X. Disponível em: <https://direct.mit.edu/neco/article/12/10/2451-2471/6415>.

GOLDBERGER, A. L.; AMARAL, L. A.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. **circulation**, Am Heart Assoc, v. 101, n. 23, p. e215–e220, 2000.

GOULART, J. H. D. M.; OLIVEIRA, P. M. R. D.; FARIAS, R. C.; ZARZOSO, V.; COMON, P. Alternating Group Lasso for Block-Term Tensor Decomposition and Application to ECG Source Separation. **IEEE Transactions on Signal Processing**, v. 68, p. 2682–2696, 2020. ISSN 1053-587X, 1941-0476. Disponível em: <https://ieeexplore.ieee.org/document/9057734/>.

GRAVES, A.; JAITLY, N. Towards end-to-end speech recognition with recurrent neural networks. In: XING, E. P.; JEBARA, T. (Ed.). **Proceedings of the 31st International Conference on**

**Machine Learning**. Beijing, China: PMLR, 2014. (Proceedings of Machine Learning Research, 2), p. 1764–1772. Disponível em: <https://proceedings.mlr.press/v32/graves14.html>.

HAGIWARA, Y.; FUJITA, H.; OH, S. L.; TAN, J. H.; TAN, R. S.; CIACCIO, E. J.; ACHARYA, U. R. Computer-aided diagnosis of atrial fibrillation based on ECG Signals: A review. **Information Sciences**, v. 467, p. 99–114, out. 2018. ISSN 00200255. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0020025518305917>.

HAYKIN, S. S. **Neural networks and learning machines**. 3. ed. ed. New York Munich: Prentice-Hall, 2009. ISBN 978-0-13-147139-9.

HINDRICKS, G.; POTPARA, T.; DAGRES, N.; ARBELO, E.; BAX, J. J.; BLOMSTRÖM-LUNDQVIST, C. *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). **European Heart Journal**, v. 42, n. 5, p. 373–498, February 2021. ISSN 0195-668X, 1522-9645. Disponível em: <https://academic.oup.com/eurheartj/article/42/5/373/5899003>.

HIRSCH, G.; JENSEN, S. H.; POULSEN, E. S.; PUTHUSSERYPADY, S. Atrial fibrillation detection using heart rate variability and atrial activity: A hybrid approach. **Expert Systems with Applications**, v. 169, p. 114452, maio 2021. ISSN 09574174. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0957417420311064>.

HITCHCOCK, F. L. The Expression of a Tensor or a Polyadic as a Sum of Products. **Journal of Mathematics and Physics**, v. 6, n. 1-4, p. 164–189, abr. 1927. ISSN 0097-1421. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/sapm192761164>.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, nov. 1997. ISSN 0899-7667, 1530-888X. Disponível em: <https://direct.mit.edu/neco/article/9/8/1735-1780/6109>.

JAHAN, M. S.; MANSOURVAR, M.; PUTHUSSERYPADY, S.; WIIL, U. K.; PEIMANKAR, A. Short-term atrial fibrillation detection using electrocardiograms: A comparison of machine learning approaches. **International Journal of Medical Informatics**, v. 163, p. 104790, jul. 2022. ISSN 1386-5056. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1386505622001046>.

KE, G.; MENG, Q.; FINLEY, T.; WANG, T.; CHEN, W.; MA, W.; YE, Q.; LIU, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>.

KRAMER, O. K-Nearest Neighbors. In: **Dimensionality Reduction with Unsupervised Nearest Neighbors**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. v. 51, p. 13–23. ISBN 978-3-642-38651-0 978-3-642-38652-7. Series Title: Intelligent Systems Reference Library. Disponível em: [https://link.springer.com/10.1007/978-3-642-38652-7\\_2](https://link.springer.com/10.1007/978-3-642-38652-7_2).

LATHAUWER, L. D. Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and Uniqueness. **SIAM Journal on Matrix Analysis and Applications**, v. 30, n. 3, p. 1033–1066, jan. 2008. ISSN 0895-4798, 1095-7162. Disponível em: <http://epubs.siam.org/doi/10.1137/070690729>.

LATHAUWER, L. D. Blind Separation of Exponential Polynomials and the Decomposition of a Tensor in Rank- $(L_r, L_r, 1)$  Terms. **SIAM Journal on Matrix Analysis and Applications**, v. 32, n. 4, p. 1451–1474, out. 2011. ISSN 0895-4798, 1095-7162. Disponível em: <http://epubs.siam.org/doi/10.1137/100805510>.

LATHAUWER, L. D.; NION, D. Decompositions of a Higher-Order Tensor in Block Terms—Part III: Alternating Least Squares Algorithms. **SIAM Journal on Matrix Analysis and Applications**, v. 30, n. 3, p. 1067–1083, jan. 2008. ISSN 0895-4798, 1095-7162. Disponível em: <http://epubs.siam.org/doi/10.1137/070690730>.

LAVALLEY, M. P. Logistic Regression. **Circulation**, v. 117, n. 18, p. 2395–2399, maio 2008. ISSN 0009-7322, 1524-4539. Disponível em: <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.106.682658>.

LLOYD-JONES, D. M.; WANG, T. J.; LEIP, E. P.; LARSON, M. G.; LEVY, D.; VASAN, R. S.; D'AGOSTINO, R. B.; MASSARO, J. M.; BEISER, A.; WOLF, P. A.; BENJAMIN, E. J. Lifetime Risk for Development of Atrial Fibrillation: The Framingham Heart Study. **Circulation**, v. 110, n. 9, p. 1042–1046, ago. 2004. ISSN 0009-7322, 1524-4539. Disponível em: <https://www.ahajournals.org/doi/10.1161/01.CIR.0000140263.20897.42>.

MACFARLANE, P. W.; OOSTEROM, A. V.; PAHLM, O.; KLIGFIELD, P.; JANSE, M.; CAMM, J. (Ed.). **Comprehensive Electrocardiology**. London: Springer London, 2010. ISBN 978-1-84882-045-6 978-1-84882-046-3. Disponível em: <http://link.springer.com/10.1007/978-1-84882-046-3>.

MAGALHÃES, L. P.; FIGUEIREDO, M. J. O.; CINTRA, F. D.; SAAD, E. B.; KUNIYOSHI, R. R. *et al.* II Diretrizes Brasileiras de Fibrilação Atrial. **Arquivos Brasileiros de Cardiologia**, v. 106, n. 4, p. 1–22, April 2016. ISSN 0066-782X, 1678-4170. Sociedade Brasileira de Cardiologia (SBC). Disponível em: <https://www.scielo.br/j/abc/a/QhSz5Wks4Yq7vJDknvCXwCR/?lang=pt>.

MANN, H. B.; WHITNEY, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. **The Annals of Mathematical Statistics**, v. 18, n. 1, p. 50–60, mar. 1947. ISSN 0003-4851. Disponível em: <http://projecteuclid.org/euclid.aoms/1177730491>.

NATTEL, S.; GUASCH, E.; SAVELIEVA, I.; COSIO, F. G.; VALVERDE, I.; HALPERIN, J. L.; CONROY, J. M.; AL-KHATIB, S. M.; HESS, P. L.; KIRCHHOF, P.; BONO, J. D.; LIP, G. Y. H.; BANERJEE, A.; RUSKIN, J.; BLENDEA, D.; CAMM, A. J. Early management of atrial fibrillation to prevent cardiovascular complications. **European Heart Journal**, v. 35, n. 22, p. 1448–1456, jun. 2014. ISSN 0195-668X, 1522-9645. Disponível em: <https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehu028>.

OLIVEIRA, P. M. R. D.; ZARZOSO, V.; FERNANDES, C. A. R. Coupled Tensor Model of Atrial Fibrillation ECG. In: **2020 28th European Signal Processing Conference (EUSIPCO)**. Amsterdam, Netherlands: IEEE, 2021. p. 915–919. ISBN 978-90-827970-5-3. Disponível em: <https://ieeexplore.ieee.org/document/9287494/>.

OLIVEIRA, P. M. R. de; ZARZOSO, V. Temporal Stability of Block Term Decomposition in Noninvasive Atrial Fibrillation Analysis. In: **2018 52nd Asilomar Conference on Signals, Systems, and Computers**. IEEE, 2018. p. 816–820. ISSN: 2576-2303. Disponível em: <https://ieeexplore.ieee.org/document/8645164>.

OLIVEIRA, P. M. R. de; ZARZOSO, V. Löwner-Based Tensor Decomposition for Blind Source Separation in Atrial Fibrillation ECGs. In: **2019 27th European Signal Processing Conference (EUSIPCO)**. IEEE, 2019. p. 1–5. ISSN: 2076-1465. Disponível em: <https://ieeexplore.ieee.org/document/8902493>.

O'MALLEY, T.; BURSZTEIN, E.; LONG, J.; CHOLLET, F.; JIN, H.; INVERNIZZI, L. *et al.* **KerasTuner**. 2019. <https://github.com/keras-team/keras-tuner>.

PEREZ, M. V.; DEWEY, F. E.; MARCUS, R.; ASHLEY, E. A.; AL-AHMAD, A. A.; WANG, P. J.; FROELICHER, V. F. Electrocardiographic predictors of atrial fibrillation. **American Heart Journal**, v. 158, n. 4, p. 622–628, out. 2009. ISSN 00028703. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0002870309006127>.

PUNTES, O. M. O.; ROLIM, A. C. D. S.; MACÊDO, A. M. V.; MACEDO, L. M. D.; MEDEIROS, J. H. B. D.; NETO, L. M. C.; FEITOSA, P. W. G. Análise Epidemiológica de Mortalidade por Doença Cardiovascular no Brasil. **ID on line. Revista de psicologia**, v. 17, n. 65, p. 469–479, fev. 2023. ISSN 1981-1179. Disponível em: <https://idonline.emnuvens.com.br/id/article/view/3639>.

RIETA, J.; CASTELLS, F.; SANCHEZ, C.; ZARZOSO, V.; MILLET, J. Atrial Activity Extraction for Atrial Fibrillation Analysis Using Blind Source Separation. **IEEE Transactions on Biomedical Engineering**, v. 51, n. 7, p. 1176–1186, jul. 2004. ISSN 0018-9294. Disponível em: <http://ieeexplore.ieee.org/document/1306570/>.

RIZWAN, A.; ZOHA, A.; MABROUK, I. B.; SABBOUR, H. M.; AL-SUMAITI, A. S.; ALOMAINY, A.; IMRAN, M. A.; ABBASI, Q. H. A Review on the State of the Art in Atrial Fibrillation Detection Enabled by Machine Learning. **IEEE Reviews in Biomedical Engineering**, v. 14, p. 219–239, 2021. ISSN 1937-3333, 1941-1189. Disponível em: <https://ieeexplore.ieee.org/document/9016113/>.

ROTH, G. A.; MENSAH, G. A.; JOHNSON, C. O.; ADDOLORATO, G.; AMMIRATI, E.; BADDOUR, L. M.; BARENGO, N. C.; BEATON, A. Z.; BENJAMIN *et al.* Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019. **Journal of the American College of Cardiology**, v. 76, n. 25, p. 2982–3021, dez. 2020. ISSN 07351097. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0735109720377755>.

SCHNABEL, R. B.; YIN, X.; GONA, P.; LARSON, M. G.; BEISER, A. S.; MCMANUS, D. D.; NEWTON-CHEH, C.; LUBITZ, S. A.; MAGNANI, J. W.; ELLINOR, P. T.; SESHADRI, S.; WOLF, P. A.; VASAN, R. S.; BENJAMIN, E. J.; LEVY, D. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. **The Lancet**, v. 386, n. 9989, p. 154–162, jul. 2015. ISSN 01406736. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0140673614617748>.

SILVA, I.; MOODY, B.; MOODY, G. **Waveform Database Software Package (WFDB) for MATLAB and Octave**. PhysioNet, 2021. Disponível em: <https://physionet.org/content/wfdb-matlab/0.10.0/>.

SILVA, I.; MOODY, G. B. An Open-source Toolbox for Analysing and Processing PhysioNet Databases in MATLAB and Octave. **Journal of Open Research Software**, v. 2, set. 2014. ISSN 2049-9647. Disponível em: <http://openresearchsoftware.metajnl.com/articles/10.5334/jors.bi/>.

SORBER, L.; BAREL, M. V.; LATHAUWER, L. D. Optimization-Based Algorithms for Tensor Decompositions: Canonical Polyadic Decomposition, Decomposition in Rank- $(L_r, L_r, 1)$  Terms, and a New Generalization. **SIAM Journal on Optimization**, v. 23, n. 2, p. 695–720, jan. 2013. ISSN 1052-6234, 1095-7189. Disponível em: <http://epubs.siam.org/doi/10.1137/120868323>.

SÖRNMO, L.; LAGUNA, P. Electrocardiogram (ECG) Signal Processing. In: **Wiley Encyclopedia of Biomedical Engineering**. John Wiley & Sons, Ltd, 2006. ISBN 978-0-471-74036-0. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780471740360.ebs1482>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780471740360.ebs1482>.

TUCKER, L. R. Some mathematical notes on three-mode factor analysis. **Psychometrika**, v. 31, n. 3, p. 279–311, set. 1966. ISSN 1860-0980. Disponível em: <https://doi.org/10.1007/BF02289464>.

VERVLIET, N.; DEBALS, O.; LATHAUWER, L. D. Tensorlab 3.0 — Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. In: **2016 50th Asilomar Conference on Signals, Systems and Computers**. IEEE, 2016. p. 1733–1738. Disponível em: <https://ieeexplore.ieee.org/document/7869679>.

WEBB, G. I.; KEOGH, E.; MIIKKULAINEN, R. Naïve bayes. **Encyclopedia of machine learning**, v. 15, n. 1, p. 713–714, 2010.

XIE, C.; MCCULLUM, L.; JOHNSON, A.; POLLARD, T.; GOW, B.; MOODY, B. **Waveform Database Software Package (WFDB) for Python**. PhysioNet, 2021. Disponível em: <https://physionet.org/content/wfdb-python/3.3.0/>.

YU, Y.; SI, X.; HU, C.; ZHANG, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. **Neural Computation**, v. 31, n. 7, p. 1235–1270, jul. 2019. ISSN 0899-7667, 1530-888X. Disponível em: <https://direct.mit.edu/neco/article/31/7/1235-1270/8500>.

ZARZOSO, V. Parameter estimation in block term decomposition for noninvasive atrial fibrillation analysis. In: **2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)**. IEEE, 2017. p. 1–5. Disponível em: <https://ieeexplore.ieee.org/document/8313173>.

ZIEGLER, P. D.; KOEHLER, J. L.; MEHRA, R. Comparison of continuous versus intermittent monitoring of atrial arrhythmias. **Heart Rhythm**, v. 3, n. 12, p. 1445–1452, dez. 2006. ISSN 1547-5271. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1547527106018418>.