



**UNIVERSIDADE FEDERAL DO PARÁ**  
**INSTITUTO DE TECNOLOGIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

*Ítalo Flexa Di Paolo*

***Classificação automática de arritmias cardíacas  
através de redes neurais convolucionais multimodais  
com mecanismo de atenção***

TD 2 / 2025

UFPA / ITEC / PPGEE  
Campus Universitário do Guamá  
Belém-Pará-Brasil  
2025



UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**“CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS BASEADA EM  
REDES NEURAIS CONVOLUCIONAIS MULTIMODAIS COM  
MECANISMO DE ATENÇÃO”**

AUTOR: ÍTALO FLEXA DI PAOLO

TESE DE DOUTORADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA. APROVADA EM: 19/02/2025

**BANCA EXAMINADORA:**

---

***Prof.<sup>a</sup> Dr.<sup>a</sup> Adriana Rosa Garcez Castro***  
*(Orientadora – PPGEE/ITEC/UFPA)*

---

***Prof.<sup>a</sup> Dr.<sup>a</sup> Jasmine Priscyla Leite de Araújo***  
*(Avaliadora Interna - PPGEE/ITEC/UFPA)*

---

***Prof. Dr. Roberto Célio Limão de Oliveira***  
*(Avaliador Interno - PPGEE/ITEC/UFPA)*

---

***Prof. Dr. Márcio José Moutinho da Ponte***  
*(Avaliador Externo - UFOPA)*

---

***Prof. Dr. Orlando Shigueo Ohashi Júnior***  
*(Avaliador Externo - UFRA)*

**VISTO:**

---

***Prof. Dr. Diego Lisboa Cardoso***  
*(Coordenador do PPGEE/ITEC/UFPA)*

---

D596a Di Paolo, Ítalo Flexa

Classificação automática de arritmias cardíacas através de redes neurais convolucionais multimodais com mecanismo de atenção / Ítalo Flexa Di Paolo. -- 2024.

Orientadora: Adriana Rosa Garcez Castro  
Tese (Doutorado) – Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2024.

1. Redes neurais (computação). 2. Processamento digital de sinais. 3. Electrocardiograma. I. Título.

CDD 23. ed. 006.32

---

Elaborada por Lucicléa S. de Oliveira – CRB -2/648

*Ao meu pai, Pasquale Di Paolo (in memoriam) que, embora não esteja mais fisicamente presente, permanece vivo em cada conquista que alcanço. Sinto sua falta em cada passo desta jornada, mas também sinto sua presença em cada vitória. Esta conquista é, de certa forma, também sua, e espero que, onde quer que esteja, sinta orgulho do caminho que trilhei. Com eterna saudade e amor;*

*À minha mãe, Darcy e meus irmãos, Angela e Luciano,  
pelo apoio familiar sempre imprescindível;*

*À minha esposa Flavia, cuja presença constante e amorosa  
foi um grande pilar de sustentação ao longo desses anos;*

*À minha filha Valentina, cuja inocência e alegria foram grandes fonte de energia. Que  
você sempre mantenha o espírito de buscar conhecimentos ao longo da vida.*

*Para você, que em breve me chamará de papai e cresce dia a dia na barriga da  
mamãe e já enche meu coração de amor. Que esta conquista seja não apenas um  
pedaço do mundo que quero construir para você, mas também um legado de inspiração  
e motivação, mostrando que a educação é uma poderosa fonte de transformação.*

## AGRADECIMENTOS

*À minha mãe, Profa. Dra. Darcy Flexa Di Paolo, dedico minha mais profunda gratidão. Acompanhou de perto cada desafio, cada momento de exaustão, sempre oferecendo o conforto do seu apoio, a sabedoria dos seus conselhos e a força do seu amor. Foram suas mãos estendidas e seu coração sempre aberto que me permitiram seguir em frente com resiliência, mesmo quando parecia difícil encontrar o próximo passo. Por tudo isso, e por ser o exemplo constante de dedicação e amor, sou eternamente grato;*

*À minha esposa, Flavia Fontenele Di Paolo, que nunca deixou de me motivar nos momentos difíceis, que celebrou comigo cada conquista e me deu forças para continuar. Sua paciência e companheirismo foram fundamentais para superação de muitos desafios;*

*À minha orientadora Profa. Dra. Adriana Rosa Garcez Castro, por ter acreditado desde o início nesta jornada, pelos conselhos valiosos e apoio constante. Sua sabedoria e dedicação foram fundamentais para a conclusão deste trabalho;*

*Ao colega Me. Eng. Sérgio Teixeira Corrêa Filho, que diante do isolamento social vivido durante estes estudos, disponibilizou seu tempo para valiosas trocas de ideias que contribuíram significativamente com este trabalho;*

*Aos vários pesquisadores que passaram pelo PPGEE, deixando seus legados, e que contribuíram significativamente para o sucesso deste trabalho, entre eles, Dra. Lorena dos Reis Morais, Dr. Eng. Juan Ferreira Vidal, Dr. Eng. Ewerton Cristhian Lima de Oliveira, Me. Eng. Ulrich Kauê Mendes Alencar da Silva, entre tantos outros;*

*Aos meus orientadores anteriores, nas jornadas de graduação e pós-graduação, Profs. Dra. Andréa Cristina Marques de Araújo (in memoriam), Ma. Lêda Mara de Souza Oliveira Monteiro, Dr. José Augusto Lima Barreiros, Dr. Orlando Fonseca Silva, Dr. Roberto Célio Limão de Oliveira, Me. Marcos Venícios Conceição de Araújo, Dr. Walter Barra Junior e Dr. Ádamo Lima de Santana, por compartilharem suas experiências em momentos importantes, intensos e cheios de significados, contribuindo significativamente na minha evolução individual e profissional;*

*Aos professores e funcionários do PPGEE e de todo o Instituto de Tecnologia da UFPA, sempre presentes no incentivo aos alunos da pós-graduação;*

*Aos colegas da Universidade do Estado do Pará, em especial da Diretoria de Serviços de Processamento de Dados (DSPD), dos Departamentos de Engenharia de Produção (DENG) e de Sistemas Computacionais e Infraestrutura (DSCI), e do campus de Castanhal, bem como os monitores e alunos dos cursos de Tecnologia em Análise e Desenvolvimento de Sistemas e Bacharelado em Engenharia de Software que tive oportunidade de orientar, representados por Rafael Felipe Pantoja Sales, Rafael Farias Batista e Hian Motley Stafford Correa Barroso, que muito contribuíram com palavras de apoio, trocas de ideia e indicações qualificadas de estudos, apoiando grandemente o desenvolvimento deste trabalho;*

*Ao Governo do Estado do Pará e à Universidade do Estado do Pará, pelo suporte recebido para realização desta pesquisa através da Resolução do Conselho Universitário N. 3837/2022,*

*A todos, meu muito obrigado!*

*“Aqueles que nunca se feriram não conseguem entender o verdadeiro valor da paz”.*

## RESUMO

O eletrocardiograma (ECG) é uma tecnologia não invasiva capaz de registrar os batimentos cardíacos, sendo a técnica mais utilizada para o diagnóstico de doenças do coração. Dentre as doenças que podem ser diagnosticadas, a arritmia cardíaca é uma das cardiopatias mais comuns, sendo caracterizada pela ocorrência de batimentos cardíacos irregulares. Entretanto, a interpretação de longos registros de sinais de ECG é uma tarefa cansativa e desafiadora, quando feita de forma visual, que pode demandar tempo por parte dos médicos especialistas. A evolução da tecnologia e da inteligência artificial tem permitido avanços para o estudo e desenvolvimento de sistemas automáticos para auxílio ao diagnóstico médico. Dentro desse contexto, esta tese visa apresentar a proposta de uma estrutura para classificação de arritmias cardíacas baseada em uma Rede Neural Convolutiva (CNN) multimodal com mecanismo de atenção. A estrutura recebe como entrada segmentos de sinal ECG transformados em imagens a partir das técnicas *Hilbert Space Filling Curve* (HSFC) e *Recurrence Plot* (RP) e foi desenvolvida e avaliada a partir do banco de dados público MIT-BIH e PTB, seguindo as diretrizes da AAMI (ANSI/AAMI EC57) e considerando os paradigmas interpaciente e intrapaciente. Devido ao alto desbalanceamento de classes nos bancos de dados, técnicas complementares de aumento de dados foram avaliadas durante a fase de experimentos, destacando-se duas: SMOTE e WGAN-GP. Os resultados alcançados, considerando variações na entrada da estrutura relacionadas ao número de derivações do ECG (derivação MLII e V + MLII), podem ser considerados competitivos com trabalhos apresentados no estado da arte, com destaque para os resultados da estrutura para duas derivações do ECG, tendo obtido, para a base MIT-BIH, no paradigma intrapaciente, 99,72%, 98,19%, 97,26%, 99,34% e 97,72% de acurácia global, precisão, sensibilidade, especificidade e F1-Score, respectivamente. No paradigma interpaciente, os resultados atingidos foram de 98,48%, 94,15%, 80,23%, 96,34% e 81,91%, respectivamente.

**Palavras-chave:** Classificação de arritmias, Eletrocardiograma; Redes Neurais Convolutivas; Sinais temporais como imagens; Mecanismo de atenção.

## ABSTRACT

The electrocardiogram (ECG) is a non-invasive technology capable of recording heartbeats and is the most widely used technique for diagnosing heart diseases. Among the conditions that can be diagnosed, cardiac arrhythmia is one of the most common heart disorders, characterized by irregular heartbeats. However, interpreting long ECG signal recordings is a tiring and challenging task when performed visually, which can be time-consuming for medical specialists. Advances in technology and artificial intelligence have enabled progress in the study and development of automatic systems to support medical diagnosis. In this context, this thesis aims to propose a framework for the classification of cardiac arrhythmias based on a multimodal Convolutional Neural Network (CNN) with an attention mechanism. The framework takes as input ECG signal segments transformed into images using the Hilbert Space Filling Curve (HSFC) and Recurrence Plot (RP) techniques. It was developed and evaluated using the public MIT-BIH and PTB databases, following the AAMI (ANSI/AAMI EC57) guidelines and considering both inter-patient and intra-patient paradigms. Due to the high class imbalance in the databases, complementary data augmentation techniques were evaluated during the experimental phase, with two techniques standing out: SMOTE and WGAN-GP. The results achieved, considering variations in the input structure related to the number of ECG leads (MLII lead and V + MLII leads), can be considered competitive with state-of-the-art works. Particularly noteworthy are the results of the structure for two ECG leads, which achieved, for the MIT-BIH database in the intra-patient paradigm, 99.72%, 98.19%, 97.26%, 99.34%, and 97.72% for overall accuracy, precision, sensitivity, specificity, and F1-Score, respectively. In the inter-patient paradigm, the results obtained were 98.48%, 94.15%, 80.23%, 96.34%, and 81.91%, respectively.

**Key Words:** Classification of arrhythmias, Electrocardiogram; Convolutional Neural Networks; Time series as images; Synthetic images; Attention mechanism.

## LISTA DE FIGURAS

Figura 1: Sinal de um batimento cardíaco normal.....	18
Figura 2: Anatomia do coração humano.....	23
Figura 3: Esquema simplificado de equipamento para aquisição de sinais ECG.....	25
Figura 4: Derivações dos eletrodos para medida dos sinais ECG.....	26
Figura 5: Formas de onda em um sinal de ECG padrão e um traçado normal mostrando suas ondas e complexos.....	28
Figura 6: Fluxo para classificação automática de arritmias cardíacas.....	32
Figura 7: Dez segundos do registro 205 do banco de dados de arritmia do MIT-BIH, contendo anotações de ritmo sinusal normal “N” e taquicardia ventricular “VT” da derivação MLII.....	41
Figura 8: Rede Neural Convolutiva.....	45
Figura 9: Operação de convolução.....	46
Figura 10: CNN com abordagem multimodal.....	52
Figura 11: Estrutura de <i>Self-Attention</i> clássica com <i>Query</i> , <i>Key</i> e <i>Value</i> .....	56
Figura 12: Estrutura de <i>Self-Attention</i> clássica com funções <i>f</i> , <i>g</i> e <i>h</i> .....	56
Figura 13: Estrutura de <i>Self-Attention</i> local com convolução.....	57
Figura 14: Conversão de sinal temporal em imagem. a) Gráfico do sinal temporal; b) Imagem RP.....	59
Figura 15: Curvas de Hilbert de níveis 1, 2, 3 e 4.....	60
Figura 16: Conversão de sinal temporal em imagem. a) Gráfico do sinal temporal; b) Imagem HSFC.....	61
Figura 17: CNN multimodal com mecanismo de atenção.....	76
Figura 18: Segmentação de um sinal ECG Normal da base MIT-BIH.....	79
Figura 19: Segmentos de 2 segundos da derivação MLII por classe da base MIT-BIH.....	80
Figura 20: Exemplos de segmentos obtidos da derivação V1 por classe da base MIT-BIH.....	81
Figura 21: Segmento de sinal ECG original e filtrado.....	82
Figura 22: Etapas para geração de dados sintéticos em modelos baseados em GAN.....	90
Figura 23: Alguns mapas de cores.....	92
Figura 24: Transformação de segmento de sinal ECG da base MIT-BIH em ECG... (2018) em imagens para cada classe, utilizando os métodos GAF, RP, HSFC, MTF e HM.....	94
Figura 25: Transformação em imagens de segmento de sinal ECG da base MIT-BIH para cada classe, utilizando os métodos GAF, RP, HSFC, MTF e HM.....	95
Figura 26: Transformação em imagem de segmento de sinal ECG da base PTB para cada classe, utilizando os métodos GAF, RP, HSFC, MTF e HM.....	96

Figura 27: Normalização de imagem: (a) original; (b) normalizada por média e desvio padrão dos canais de cor.....	99
Figura 28: Componente classificador com mecanismo de atenção. ....	103
Figura 29: Matriz de confusão para base de teste da Base 1 MIT-BIH. ....	110
Figura 30: Matriz de confusão para caso de uma derivação (MLII) e base de teste da Base 2 MIT-BIH. ....	111
Figura 31: Matriz de confusão para caso de duas derivações (MLII+V) e base de teste da Base 2 MIT-BIH.....	111
Figura 32: Matriz de confusão considerando base de teste da Base 3 PTB. ....	112
Figura 33: Matriz de confusão do experimento 1, com melhor resultado de <i>F1-Score</i> interpaciente com uma derivação com a base MIT-BIH.....	114
Figura 34: Matriz de confusão do experimento 2, com melhor resultado de sensibilidade de S interpaciente com uma derivação com a base MIT-BIH. ....	115
Figura 35: Matriz de confusão do experimento 3, com melhor resultado de <i>F1-Score</i> interpaciente com duas derivações com a base MIT-BIH.....	115
Figura 36: Matriz de confusão do experimento 4, com melhor resultado de sensibilidade de S interpaciente com duas derivações com a base MIT-BIH. ....	116

## LISTA DE QUADROS

Quadro 1: Organização das anotações contidas da base MIT-BIH. ....	42
Quadro 2: Distribuição dos batimentos cardíacos na base MIT-BIH. ....	42
Quadro 3: Distribuição dos dados rotulados na base PTB. ....	44
Quadro 4: Indicadores de desempenho de trabalhos correlatos para o paradigma intrapaciente com MIT-BIH. ....	71
Quadro 5: Indicadores de desempenho de trabalhos correlatos para o paradigma intrapaciente com PTB. ....	72
Quadro 6: Indicadores globais de desempenho de trabalhos correlatos para o paradigma interpaciente. ....	74
Quadro 7: Indicadores por classe de desempenho de trabalhos correlatos para o paradigma interpaciente. ....	74
Quadro 8: Organização dos dados pelo paradigma interpaciente. ....	78
Quadro 9: Registros das derivações V utilizados. ....	80
Quadro 10: Distribuição dos segmentos extraídos do MIT-BIH. ....	82
Quadro 11: Organização dos dados da Base 1 intrapaciente MIT-BIH. ....	83
Quadro 12: Distribuição de dados da Base 2 - MIT-BIH intrapaciente. ....	84
Quadro 13: Organização dos dados pelo paradigma intrapaciente da base PTB (Base 3). ....	85
Quadro 14: Conjunto de dados para o paradigma intrapaciente da Base 1 com aumento de dados. ....	86
Quadro 15: Conjunto de dados para o paradigma interpaciente da base MIT-BIH com aumento de dados. ....	87
Quadro 16: Conjunto de dados para o paradigma intrapaciente da base 2 MIT-BIH com aumento de dados. ....	87
Quadro 17: Características do sinal ECG nas imagens geradas por RP e HSFC. ....	97
Quadro 18: Características do conjunto de dados no formato PNG, paradigma intrapaciente com ECG... (2018). ....	100
Quadro 19: Características do conjunto de dados da derivação MLII no formato PNG, paradigma interpaciente filtrado. ....	100
Quadro 20: Características do conjunto de dados no formato PNG da derivação V, paradigma interpaciente filtrado. ....	101
Quadro 21: Estrutura de cada uma das sequências de camadas convolucionais. ....	102
Quadro 22: Indicadores globais para CNN-AM, com paradigma intrapaciente e base de teste da Base1 ....	109
Quadro 23: Indicadores por classe para CNN-AM, com paradigma intrapaciente e base de teste da Base1 ....	109

Quadro 24: Indicadores globais para CNN-AM com paradigma intrapaciente e base de teste da Base 2 MIT-BIH.....	110
Quadro 25: Indicadores por classe para CNN-AM com paradigma intrapaciente e base de teste da Base 2 MIT-BIH. ....	110
Quadro 26: Indicadores globais para CNN-AM para base de teste da Base 3 PTB.....	112
Quadro 27: Indicadores por classe para CNN-AM para base de teste da Base 3 PTB.	112
Quadro 28: Indicadores globais para o paradigma interpaciente considerando a base de teste.....	114
Quadro 29: Indicadores por classe para o paradigma interpaciente considerando a base de teste. ....	114
Quadro 30: Indicadores globais interpaciente MIT-BIH. ....	120
Quadro 31: Indicadores por classe interpaciente MIT-BIH. ....	121
Quadro 32: Indicadores globais intrapaciente com a Base 2 MIT-BIH.....	121
Quadro 33: Indicadores por classe intrapaciente com a Base 2 MIT-BIH.....	121
Quadro 34: Indicadores globais intrapaciente com a Base PTB. ....	121
Quadro 35: Comparação de indicadores globais para a base de teste entre diferentes estruturas para o paradigma intrapaciente com a Base 1.....	124
Quadro 36: Comparação de indicadores globais para a base de teste da Base 2.....	125
Quadro 37: Comparação de indicadores por classe para a base de teste entre diferentes estruturas para o paradigma intrapaciente com a Base 2.....	125
Quadro 38: Comparação de indicadores globais para a base de teste entre diferentes estruturas para o paradigma intrapaciente com a Base 3.....	126
Quadro 39: Comparação de indicadores globais para a base de teste entre diferentes estruturas para o paradigma interpaciente com a base MIT-BIH. ....	126
Quadro 40: Comparação com trabalhos anteriores para o paradigma interpaciente dos indicadores globais. ....	127
Quadro 41: Comparação com trabalhos anteriores para o paradigma interpaciente dos indicadores por classe.....	128
Quadro 42: Comparação com trabalhos anteriores para o paradigma intrapaciente com a base MIT-BIH. ....	129
Quadro 43: Comparação com trabalhos anteriores para o paradigma intrapaciente com a base PTB.....	130

## LISTA DE ABREVIATURAS E SIGLAS

1D	Unidimensional
2D	Bidimensional
Acc.	<i>Accuracy</i> ou Acurácia
AAMI	<i>Association for the Advancement of Medical Instrumentation</i> ou Associação para o Avanço da Instrumentação Médica
ACGAN	<i>Auxiliary Classifier Generative Adversarial Network</i> ou Classificador Auxiliar de Redes Neurais Adversariais
AM	<i>Attention Mechanism</i> ou Mecanismo de Atenção
ANSI	<i>American National Standards Institute</i> ou Instituto Americano de Padrões Nacionais
BIH	<i>Beth Israel Hospital</i>
CNN(s)	<i>Convolutional Neural Network(s)</i> ou Rede(s) Neural(is) Convolutacional(is)
DCGAN(s)	<i>Deep Convolutional GANs</i> ou GANs Convolucionais Profundas
DS1	<i>Dataset 1</i> ou Conjunto de Dados 1
DS2	<i>Dataset 2</i> ou Conjunto de Dados 2
ECG	Eletrocardiograma ou Eletrocardiográfico
F1-Score	Média harmônica entre precisão e sensibilidade
FFT	Transformada rápida de Fourier de curto prazo
GAF	<i>Gramian Angular Field</i> ou Campo Angular Gramiano
GAN(s)	<i>Generative Adversarial Network(s)</i> ou Redes Neurais Adversariais
GPU	<i>Graphics Processing Unit</i> ou Unidade de Processamento Gráfico
HP	<i>Heatmap</i> ou mapa de calor
HSFC	<i>Hilbert Space Filling Curve</i> ou Curva de Preenchimento do Espaço de Hilbert
IA	Inteligência Artificial
MGAF	<i>Multistage Gated Average Fusion</i> ou Fusão Média Fechada de Múltiplos Estágios
MIT	Instituto de Tecnologia de Massachusetts
ML	<i>Modified Lead</i> ou Derivação Modificada
MLP	<i>Multi Layer Perceptron</i> ou Redes Perceptron Multicamadas
MTF	<i>Markov Transition Field</i> ou Campo de Transição de Markov
Prec.	<i>Precision</i> ou Precisão ou Acertos Positivos
PSO	<i>Particle Swarm Optimization</i> ou Enxame de Partículas
PTB	<i>Physikalisch-Technische Bundesanstalt</i> ou Instituto Nacional de Metrologia da Alemanha
Recall	<i>Sensitivity</i> ou Sensibilidade
RGB	Sistema de cores aditivas em que combinam o Vermelho ( <i>Red</i> ), o Verde ( <i>Green</i> ) e o Azul ( <i>Blue</i> )
RNA(s)	Rede(s) Neural(is) Artificial(is)
RNN(s)	<i>Recurrent Neural Network(s)</i> ou Rede(s) Neural(is) Recorrente(s)
RP	<i>Recurrence Plot</i> ou Gráfico de Recorrência
SMOTE	<i>Synthetic Minority Over-sampling Technique</i> ou Técnica de Sobreamostragem de Minoria Sintética
Spe.	<i>Specificity</i> ou Especificidade
SVM(s)	<i>Support Vector Machine(s)</i> ou Máquina(s) de Vetores de Suporte
WGAN-GP	<i>Wasserstein Generative Adversarial Network with Gradient Penalty</i> ou Redes Neurais Adversariais Wasserstein com Gradiente de Penalidade

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>15</b>
1.1	CONTEXTUALIZAÇÃO.....	15
1.2	JUSTIFICATIVA.....	16
1.3	OBJETIVOS.....	21
1.4	ORGANIZAÇÃO DO TEXTO.....	21
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>23</b>
2.1	O CORAÇÃO E O SINAL CARDÍACO.....	23
2.2	ARRITMIAS CARDÍACAS.....	29
2.3	SISTEMAS AUTOMÁTICOS PARA CLASSIFICAÇÃO DE ARRITMIAS ....	31
2.3.1	<i>Sistemas automáticos e o padrão AAMI.....</i>	<i>37</i>
2.3.2	<i>Métricas de avaliação do classificador.....</i>	<i>39</i>
2.3.3	<i>Base de dados MIT-BIH.....</i>	<i>40</i>
2.3.4	<i>Base de dados PTB.....</i>	<i>42</i>
2.4	REDES NEURAIS CONVOLUCIONAIS.....	44
2.4.1	<i>Modelos de CNN.....</i>	<i>48</i>
2.4.2	<i>Modelos de CNN multimodais.....</i>	<i>51</i>
2.4.3	<i>Modelos de CNN com mecanismo de atenção.....</i>	<i>53</i>
2.5	CONVERSÃO DE SINAIS TEMPORAIS EM IMAGENS.....	57
2.5.1	<i>Gráfico de Recorrência.....</i>	<i>58</i>
2.5.2	<i>Curva de Preenchimento do Espaço de Hilbert.....</i>	<i>59</i>
2.6	CONSIDERAÇÕES FINAIS.....	61
<b>3</b>	<b>ESTADO DA ARTE DE CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS.....</b>	<b>62</b>
3.1	ESTADO DA ARTE.....	62
3.2	TRABALHOS CORRELATOS.....	69
3.2.1	<i>Abordagem intrapaciente.....</i>	<i>69</i>
3.2.2	<i>Abordagem interpaciente.....</i>	<i>72</i>
3.3	CONSIDERAÇÕES FINAIS.....	74
<b>4</b>	<b>CLASSIFICAÇÃO DE ARRITMIAS COM CNN MULTIMODAL E MECANISMO DE ATENÇÃO.....</b>	<b>76</b>
4.1	BANCOS DE DADOS DE SINAIS ECG.....	76
4.1.1	<i>Banco de dados para abordagem interpaciente.....</i>	<i>77</i>
4.1.2	<i>Banco de dados para abordagem intrapaciente.....</i>	<i>83</i>

4.2	AUMENTO DE DADOS.....	85
4.2.1	<i>SMOTE</i> .....	88
4.2.2	<i>WGAN-GP</i> .....	89
4.3	TRANSFORMAÇÃO DE SINAIS ECG PARA IMAGENS .....	90
4.3.1	<i>Pré-processamento das imagens</i> .....	97
4.3.2	<i>Armazenamento das imagens</i> .....	99
4.4	CAMADAS CONVOLUCIONAIS MULTIMODAIS .....	101
4.5	CLASSIFICADOR COM MECANISMO DE ATENÇÃO.....	103
4.6	CONSIDERAÇÕES FINAIS .....	105
<b>5</b>	<b>RESULTADOS EXPERIMENTAIS</b> .....	<b>107</b>
5.1	RESULTADOS COM O PARADIGMA INTRAPACIENTE .....	109
5.1.1	<i>Experimentos com CNN-AM com Base 1 – MIT-BIH</i> .....	109
5.1.2	<i>Experimentos com CNN-AM com Base 2 – MIT-BIH</i> .....	110
5.1.3	<i>Experimentos com CNN-AM com Base 3 – PTB</i> .....	111
5.1.4	<i>Considerações sobre a abordagem intrapaciente</i> .....	112
5.2	RESULTADOS COM O PARADIGMA INTERPACIENTE.....	113
5.2.1	<i>Considerações sobre a abordagem interpaciente</i> .....	116
5.3	CONSIDERAÇÕES FINAIS .....	116
<b>6</b>	<b>ANÁLISE E DISCUSSÃO DOS RESULTADOS</b> .....	<b>118</b>
6.1	ANÁLISE GERAL DOS RESULTADOS .....	118
6.1.1	<i>Abordagem Interpaciente</i> .....	118
6.1.2	<i>Abordagem Intrapaciente</i> .....	119
6.2	ESTUDO DE ABLAÇÃO .....	120
6.3	COMPARAÇÃO COM ESTRUTURAS DA FASE EXPLORATÓRIA .....	122
6.3.1	<i>Explorações na abordagem intrapaciente</i> .....	123
6.3.2	<i>Exploração na abordagem interpaciente</i> .....	126
6.4	COMPARAÇÃO COM TRABALHOS CORRELATOS .....	127
6.5	CONSIDERAÇÕES FINAIS .....	130
<b>7</b>	<b>CONCLUSÃO</b> .....	<b>132</b>
7.1	PONTOS DE INOVAÇÃO.....	133
	<b>REFERÊNCIAS</b> .....	<b>135</b>

# 1 INTRODUÇÃO

## 1.1 CONTEXTUALIZAÇÃO

O corpo humano é um sistema complexo, com muitos subsistemas interconectados que trabalham em harmonia para manter a saúde e o equilíbrio. Entre esses sistemas, o cardiovascular desempenha um papel crucial ao transportar nutrientes, oxigênio e resíduos metabólicos por todo o corpo, desempenhando um papel vital na manutenção da homeostase corporal (HALL, 2021). No entanto, as disfunções nesse sistema podem levar a uma série de anomalias e doenças cardíacas que, por sua vez, têm repercussões significativas em todo o organismo.

O subsistema cardíaco consiste no coração, vasos sanguíneos e sangue. Todos trabalham de forma coordenada para fornecer oxigênio e nutrientes para as células e remover produtos residuais. O coração, como uma bomba muscular, impulsiona o sangue rico em oxigênio para os tecidos e órgãos e retorna o sangue pobre em oxigênio aos pulmões para reoxigenação, num processo contínuo de circulação essencial para manter as funções celulares, o metabolismo e a saúde geral do corpo.

Diversos fatores, como genética, estilo de vida e envelhecimento, podem contribuir para o desenvolvimento de anomalias e doenças cardíacas que variam desde arritmias cardíacas comuns até condições mais graves, como: doença arterial coronariana, insuficiência cardíaca e cardiomiopatias. As consequências dessas doenças podem ser debilitantes, levando a redução da capacidade funcional; aumento do risco de acidente vascular cerebral; insuficiência de órgãos e, em casos extremos, morte súbita (HALL, 2021).

O diagnóstico de doenças cardíacas envolve uma combinação de avaliação clínica, exames de imagem (como ecocardiograma, eletrocardiograma, ressonância magnética cardíaca) e testes funcionais (como teste ergométrico). Com base no diagnóstico, diferentes abordagens terapêuticas podem ser adotadas, incluindo mudanças no estilo de vida, medicamentos, intervenções cirúrgicas (como angioplastia e cirurgia de revascularização) e, em casos extremos, transplante cardíaco (HONG *et al.*, 2020; PETMEZAS *et al.*, 2022).

A evolução da tecnologia e da Inteligência Artificial (IA) vem permitindo o desenvolvimento de sistemas de apoio à decisão que podem auxiliar médicos no

diagnóstico e tratamento de doenças cardíacas. Algoritmos de IA podem analisar grandes conjuntos de dados, como exames de imagem e histórico médico, para identificar padrões e auxiliar na tomada de decisões clínicas mais precisas (LUZ *et al.*, 2016; HONG *et al.*, 2020). Esses sistemas não substituem a expertise médica, mas podem complementar o julgamento clínico, melhorando a eficiência diagnóstica e a escolha de tratamentos personalizados.

O estudo do sistema cardíaco humano é permeado pela complexidade intrínseca do corpo humano, sendo de importância vital para a saúde e o bem-estar. A compreensão das anomalias e doenças cardíacas, bem como o desenvolvimento de métodos de diagnóstico e tratamento avançados, são fundamentais para mitigar os impactos negativos dessas condições. A integração de sistemas de apoio à decisão baseados em IA oferece uma promissora abordagem para aprimorar a precisão diagnóstica e otimizar os cuidados cardíacos e, em última instância, melhorando a qualidade de vida dos pacientes.

## 1.2 JUSTIFICATIVA

A técnica não invasiva do Eletrocardiograma (ECG) destaca-se como uma ferramenta fundamental para registrar os ritmos cardíacos, desenvolvida em 1901 (GIFFONI; TORRES, 2009) e permanecendo como a abordagem mais prevalente no diagnóstico de enfermidades cardíacas até os dias atuais, como para identificação de arritmia cardíaca, notabilizada pela irregularidade dos batimentos do coração.

A Organização Mundial da Saúde (WHO..., 2021) estima que, a cada ano, cerca de 17,9 milhões de indivíduos sucumbem devido a doenças cardiovasculares, o que corresponde a 32% de todas as fatalidades globais. Entre esses óbitos, aproximadamente 85% resultam de eventos como ataques cardíacos e acidentes vasculares cerebrais. Intrigantemente, mais de três quartos dessas mortes associadas a doenças cardiovasculares acontecem em nações de renda média e baixa.

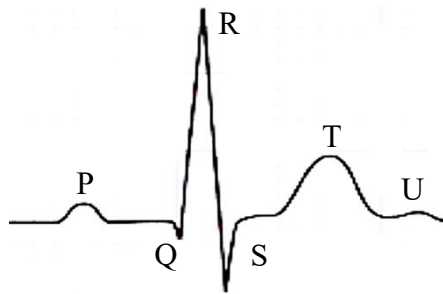
Em dados atualizados por Tsao *et al.* (2023), foram estimadas em 2020, globalmente, 19,05 milhões de mortes por doenças cardíacas, o que representou um aumento de 18,71% em relação a 2010. Ademais, tal período também foi marcado por um crescimento de 29,01% nas incidências dessas enfermidades.

Entretanto, o processo de identificação e classificação de arritmias pode ser extremamente penoso para um ser humano e ainda sujeito a erros, visto que é necessária

uma análise de batimento a batimento ao longo de horas ou até mesmo dias de registros de eletrocardiogramas (ECG), que são representações gráficas da atividade elétrica do coração ao longo do tempo, capturados, por exemplo, por um *Holter* cardíaco. Uma alternativa para auxiliar no diagnóstico médico consiste no uso de técnicas inteligentes para a classificação automática de arritmias. Dessa forma, o aprendizado de máquina surge com muitas oportunidades de investigação, destacando-se, entre elas, as técnicas de aprendizado profundo.

Os mecanismos de aprendizado profundo, em particular as Redes Neurais Convolucionais (CNNs), são amplamente utilizados na análise de sinais eletrocardiográficos (ECG) como em Gu *et al.* (2018); Zhai e Tin (2018); Izci *et al.* (2019); Benalcazar *et al.* (2020); Bragin e Spitsyn (2020); Quiao *et al.* (2020); Shaker *et al.* (2020); Xu, Jeong e Li (2020); Ahmad *et al.* (2021); Belkadi, Daamouche e Melgani, (2021); Fradi *et al.* (2021); He *et al.* (2021); Seitanidis, Gialelis e Papaconstantinou (2022); Ahmed *et al.* (2023); Islam *et al.* (2023); Mewada (2023); Rawal, Prajapati e Darji (2023); Singh e Krishnan (2023), e Zhou e Fang, (2024). As CNNs são um tipo de arquitetura de rede neural projetada especificamente para processar dados que têm uma estrutura de grade, como imagens, sinais de áudio e sinais temporais, como os sinais ECG.

As CNNs podem ser utilizadas tanto em sinais ECG 1D, que representam a amplitude do sinal ao longo do tempo em um único canal, quanto em sinais ECG 2D, que são representações bidimensionais dos sinais obtidos de múltiplos canais. Para melhor compreensão, o ECG é uma tecnologia não invasiva capaz de realizar o registro do sinal cardíaco, onde eletrodos são posicionados na pele do paciente, em diferentes áreas do corpo, para detecção de variações nos sinais elétricos durante as contrações e relaxamentos do coração. A Figura 1 ilustra o sinal de um batimento cardíaco normal que consiste da onda P, complexo QRS e ondas T e U, características estas que refletem as atividades elétricas de repolarização e despolarização do átrio e ventrículo.

**Figura 1:** Sinal de um batimento cardíaco normal.

**Fonte:** Di Paolo e Castro (2024).

No caso dos sinais ECG 1D, as CNNs podem ser usadas para tarefas como a classificação de arritmias cardíacas, detecção de eventos específicos no ECG, como o pico R ou a onda P, como apresentado por Hannun *et al.* (2019), Ahmed *et al.* (2023) e Rawal, Prajapati e Darji (2023). Já nos sinais ECG 2D, as CNNs podem ser aplicadas em tarefas mais complexas, como a localização de áreas isquêmicas no coração, a segmentação de estruturas cardíacas, como o ventrículo esquerdo como apresentado por Muthulakshmi e Kavitha (2019), e a análise espacial-temporal de eventos cardíacos como apresentados por Seitanidis, Gialelis e Papaconstantinou (2022); Islam *et al.* (2023), e Singh e Krishnan (2023).

A principal vantagem das CNNs é a sua capacidade de aprendizado de características relevantes de forma automática a partir dos dados de treinamento, sem a necessidade de extração manual de características. Isso permite que as CNNs sejam altamente eficazes na análise de sinais ECG, que podem ser complexos e difíceis de serem interpretados por humanos. Além disso, as CNNs também são capazes de lidar com variações nos sinais ECG, como variações de forma de onda e ruídos, o que é essencial na análise de dados clínicos reais.

Avançar nas investigações de aplicação de CNNs em análise de sinais ECG pode trazer uma série de benefícios significativos para a medicina e a saúde em geral, dentre eles:

- **Diagnóstico mais preciso:** As CNNs podem ajudar a melhorar a precisão e a confiabilidade dos diagnósticos de doenças cardíacas, como arritmias, infarto do miocárdio e outras condições cardíacas, ao automatizar a detecção de padrões complexos nos sinais ECG que podem ser difíceis de serem identificados por humanos. Isso pode levar a um diagnóstico mais

rápido e preciso, permitindo o tratamento adequado em estágios iniciais das doenças cardíacas, e consequentes melhores resultados para os pacientes.

- **Monitoramento contínuo:** As CNNs podem ser usadas para análise em tempo real de sinais ECG, permitindo o monitoramento contínuo de pacientes em tempo real, especialmente em ambientes de cuidados intensivos, onde a detecção precoce de eventos cardíacos é crítica, ou até em dispositivos vestíveis com o avanço dos *smartwatches* com sensores cada vez mais precisos, podendo auxiliar na identificação de alterações na atividade elétrica do coração, possibilitando a intervenção médica imediata quando necessário.
- **Personalização do tratamento:** Com a análise avançada dos sinais ECG por meio de CNNs, é possível obter informações detalhadas sobre a atividade elétrica do coração de cada paciente, permitindo um tratamento mais personalizado e adaptado às necessidades individuais. Isso pode levar a uma melhor seleção de terapias e ajuste de doses de medicamentos com base nas características específicas de cada paciente, levando a melhores resultados clínicos.
- **Acesso a cuidados de saúde em áreas remotas:** As CNNs podem ser utilizadas em dispositivos de monitoramento remoto, permitindo o acesso a cuidados de saúde em áreas remotas ou com recursos limitados. Isso pode facilitar a detecção precoce de eventos cardíacos e o acompanhamento de pacientes em regiões distantes, contribuindo para a melhoria da saúde em comunidades remotas.
- **Redução de custos de saúde:** O uso de CNNs pode ajudar a otimizar o fluxo de trabalho clínico e reduzir custos de saúde, ao automatizar a análise de sinais ECG e possibilitar a triagem mais eficiente de pacientes. Além disso, a detecção precoce de doenças cardíacas e o tratamento adequado em estágios iniciais podem reduzir a necessidade de procedimentos médicos mais invasivos e custosos em estágios avançados de doenças cardíacas.

As CNNs são classificadores robustos para imagens, que são dados em duas

dimensões, que guardam padrões espaciais (KRIZHEVSKY, SUTSKEVER, HINTON, 2012; KRIZHEVSKY, SUTSKEVER, HINTON, 2012; HE *et al.*, 2015; IZCI *et al.*, 2019, HALEEM *et al.*, 2021). Os sinais ECG são originalmente sinais temporais, ou seja, dados em uma dimensão. Trabalhos recentes já vêm apontando bons resultados na conversão de sinais temporais em imagens para classificação de arritmias, como em Izci *et al.* (2019), Loka, Kavitha e Kurita (2019), Ahmad *et al.* (2021) e Seitanidis, Gialelis e Papaconstantinou (2022), sendo um campo de estudo ainda ativo, como também apontado nas revisões apresentadas por Luz *et al.* (2016), Hong *et al.* (2020) e Petmezas *et al.* (2022).

Ressalta-se que a interpretação clínica dos resultados obtidos com o uso de CNNs em sinais ECG deve ser feita por profissionais de saúde qualificados, uma vez que o uso desses modelos é uma ferramenta de apoio à decisão clínica e não substitui a experiência clínica e o julgamento médico, sendo fundamental garantir que essas tecnologias sejam validadas de forma rigorosa e implementadas de forma ética e responsável na prática clínica.

A busca por um desempenho de classificação mais sólido na detecção de arritmias cardíacas de forma automática muitas vezes envolve a combinação de distintos métodos de classificação ou a adoção de abordagens híbridas. Além disso, identifica-se duas abordagens distintas, ou paradigmas, no âmbito da análise de dados clínicos, especialmente relacionados aos registros de pacientes, sendo conhecidas como abordagens interpaciente e intrapaciente, bem discutidos nas revisões apresentadas por Luz *et al.* (2016) e Hong *et al.* (2020).

O paradigma interpaciente foca na análise de dados entre diferentes indivíduos, como apresentados em Garcia *et al.* (2017), Wang *et al.* (2021) e Dias *et al.* (2021). Nessa abordagem, o propósito reside na extração de informações e na realização de comparações dentro de um conjunto de dados que engloba múltiplos pacientes. Esse processo envolve a alocação dos dados para treinamento de um modelo classificador a partir de registros de determinados pacientes, enquanto os dados de validação e teste pertencem a registros de outros pacientes.

Já o paradigma intrapaciente concentra-se na compreensão das mudanças e padrões individuais que emergem dentro de um conjunto de pacientes ao longo do tempo, como apresentados em He *et al.* (2021), Ahmed *et al.* (2021), Islam *et al.* (2023) e Zhou e Fang (2024). Nessa abordagem, os dados de treinamento, validação e teste de um

modelo classificador são organizados de maneira aleatória, independente da identificação dos pacientes a que pertencem. Essa estruturação implica que os dados dos mesmos pacientes estão presentes em todos os conjuntos de dados.

### 1.3 OBJETIVOS

Esta tese tem como objetivo geral desenvolver um classificador automático de arritmias cardíacas baseado em Rede Neural Convolutacional multimodal com mecanismo de atenção, para auxílio ao diagnóstico médico, considerando dados de ECGs em duas abordagens metodológicas: intrapaciente e interpaciente. Para atingir esse objetivo geral, foram definidos os seguintes **objetivos específicos**:

- Definição das melhores técnicas para transformação de sinais temporais de ECG em imagens para uso na entrada da estrutura;
- Definição das melhores técnicas para aumento de dados para banco de dados desbalanceados usados nos experimentos;
- Avaliação sobre melhoria de resultados de classificação de arritmias com abordagem multimodal e introdução de mecanismos de atenção;
- Avaliação de melhoria de resultados de classificação com uso de diferentes derivações de sinais ECG, e
- Análise de resultados através de estudo comparativo de indicadores quantitativos com o estado da arte.

### 1.4 ORGANIZAÇÃO DO TEXTO

Este trabalho está organizado em cinco Capítulos além do Capítulo introdutório.

O Capítulo 2 apresenta a fundamentação teórica para o bom entendimento deste trabalho, apresentando o funcionamento do coração e o sinal cardíaco que pode ter sua dinâmica captada através do ECG. São também apresentados mecanismos automáticos para detecção de anomalias no sinal ECG e as bases de dados utilizadas neste estudo, fundamentos sobre redes neurais convolucionais e suas abordagens multimodais, mecanismos de atenção e técnicas de conversão de sinais temporais em imagens. Também é apresentada a norma e suas recomendações usadas para padronizar e avaliar experimentos relacionados à classificação automática de sinais ECG.

O Capítulo 3 apresenta do estado da arte e de trabalhos correlatos a proposta desta

tese, apresentando o campo ativo de pesquisa no qual o trabalho foi conduzido. O Capítulo 4 apresenta a estrutura proposta nesta tese para classificação de arritmias apresentando a metodologia aplicada em seu desenvolvimento, destacando a organização dos bancos de dados de sinais ECG utilizados, seu tratamento e posterior transformação em imagens. Também são apresentadas as técnicas usadas para geração de dados sintéticos para as classes minoritárias e como foram desenvolvidos os classificadores.

O Capítulo 5 apresenta os resultados experimentais alcançados, sendo dividido na abordagem intrapaciente e interpaciente. O Capítulo 6 apresenta a análise e discussão dos resultados e comparação com diferentes modelos de classificação e com resultados apresentados no estado da arte. Por fim, o Capítulo 7 apresenta as conclusões obtidas, incluindo os pontos de inovação atingidos.

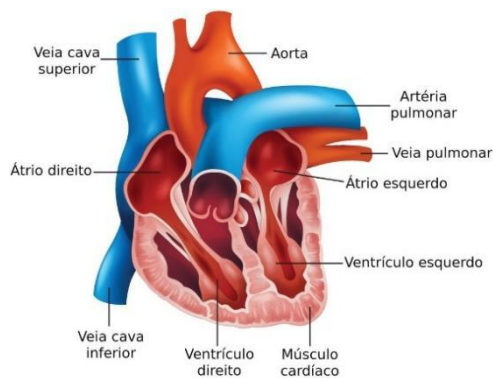
## 2 FUNDAMENTAÇÃO TEÓRICA

Este Capítulo apresenta a fundamentação teórica necessária para um bom entendimento da proposta deste trabalho. Inicia-se apresentando conceitos básicos sobre o coração, sinal cardíaco e arritmias cardíacas. Em seguida, são explicados fundamentos sobre sistemas automáticos de classificação de arritmias, incluindo padrões normativos para classificação automatizadas, que têm se apresentado como um campo de pesquisa ativo para apoio aos profissionais da saúde cardíaca e contribuição ao bem-estar humano. Para finalizar, é apresentada uma fundamentação teórica sobre Redes Neurais Convolucionais, mecanismos de atenção e técnicas de transformação de sinais temporais em imagens.

### 2.1 O CORAÇÃO E O SINAL CARDÍACO

O coração é um órgão muscular vital que desempenha um papel central no sistema circulatório humano, sendo responsável por bombear o sangue, fornecendo oxigênio e nutrientes para todas as partes do corpo. Anatomicamente, como observado na Figura 2, o coração é dividido em quatro câmaras: dois átrios (aurículas) e dois ventrículos. Ao átrio esquerdo está ligada a veia pulmonar, ao direito a veia cava superior. Ao ventrículo esquerdo está ligada a aorta, e ao direito a artéria pulmonar. Os átrios são as câmaras superiores e os ventrículos são as câmaras inferiores que recebem o sangue que retorna do corpo (no caso do átrio direito) e dos pulmões (no caso do átrio esquerdo), enquanto os ventrículos bombeiam o sangue para fora do coração, sendo o ventrículo direito o que bombeia o sangue para os pulmões e o ventrículo esquerdo o que bombeia o sangue para o resto do corpo (HALL, 2021).

**Figura 2:** Anatomia do coração humano.



**Fonte:** Banco de Imagens Shutterstock.

O sistema elétrico do coração é responsável por coordenar as contrações rítmicas e regulares do músculo cardíaco, que são conhecidas como batimentos cardíacos. Esse sistema elétrico envolve uma série de células especializadas que geram e conduzem os sinais elétricos pelo coração. O nó sinusal, localizado na aurícula direita, é conhecido como o marcapasso natural do coração, pois gera os sinais elétricos iniciais dos batimentos cardíacos. Esses sinais elétricos se espalham pelas células musculares dos átrios, fazendo com que elas se contraíam e empurrem o sangue para os ventrículos (CLIFFORD, 2006; HALL, 2021).

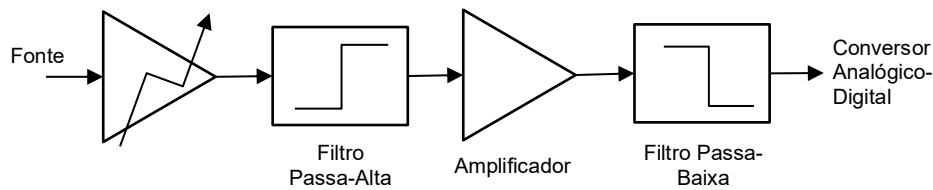
Os sinais elétricos são transmitidos para o nó atrioventricular, localizado entre os átrios e os ventrículos. Esse nó retarda a propagação dos sinais elétricos, permitindo que os ventrículos se encham completamente de sangue antes de serem estimulados a se contrair. Após atravessar o nó atrioventricular, os sinais elétricos são conduzidos por uma rede de fibras especializadas e seus ramos, que se ramificam pelos ventrículos. Essas fibras permitem uma rápida propagação dos sinais elétricos, fazendo com que os ventrículos se contraíam de forma coordenada e eficiente, impulsionando o sangue para fora do coração (CLIFFORD, 2006).

A atividade elétrica do coração pode ser observada e registrada através de eletrodos ligados ao corpo e um equipamento apropriado que gera um ECG, que é um exame não invasivo que mede os sinais elétricos produzidos pelo coração. Os eletrodos são aplicados na pele do paciente em diferentes áreas do corpo, como: como no peito, braços e pernas. Essa ligação dos eletrodos formam as chamadas derivações, e cada derivação oferece uma perspectiva única do sinal cardíaco, permitindo uma análise detalhada da atividade elétrica do coração. Esses eletrodos detectam as variações dos sinais elétricos à medida que o coração se contrai e relaxa, e essas informações são registradas pelo equipamento apropriado em um gráfico, permitindo que os médicos avaliem a atividade elétrica do coração de forma a identificar normalidade ou possíveis problemas cardíacos.

Um esquema simplificado de equipamento para aquisição de sinais ECG é apresentado na Figura 3, sendo a diferença de potencial entre os pontos marcados pelos eletrodos na pele realçada com o auxílio de um amplificador instrumental com isolação ótica. Em seguida, o sinal é submetido a um filtro passa-alta, depois um segundo estágio de amplificação, depois um filtro de passa-baixa, e, por fim, é amostrado por um conversor analógico-digital que faz o registro gráfico desta aquisição, o qual se chama

ECG pré-processado.

**Figura 3:** Esquema simplificado de equipamento para aquisição de sinais ECG.



**Fonte:** Adaptado de Clifford (2006).

O primeiro aparelho de ECG foi desenvolvido pelo Dr. Willem Einthoven, um fisiologista holandês, em 1901, cunhando os termos “derivação” e “tripé” para descrever o uso de eletrodos posicionados em diferentes partes do corpo para registrar a atividade elétrica do coração de diferentes ângulos. O trabalho pioneiro do Dr. Einthoven levou à evolução dos conceitos modernos de derivações eletrocardiográficas, que são amplamente utilizados até os dias atuais (GIFFONI; TORRES, 2009).

Os pontos onde os eletrodos são posicionados na pele são chamadas de derivações, conforme ilustra a Figura 4.a, as quais registram a atividade elétrica do coração em relação a uma referência neutra. Esses sinais são então usados para gerar o traçado gráfico do ECG, que é uma representação visual da atividade elétrica do coração ao longo do tempo.

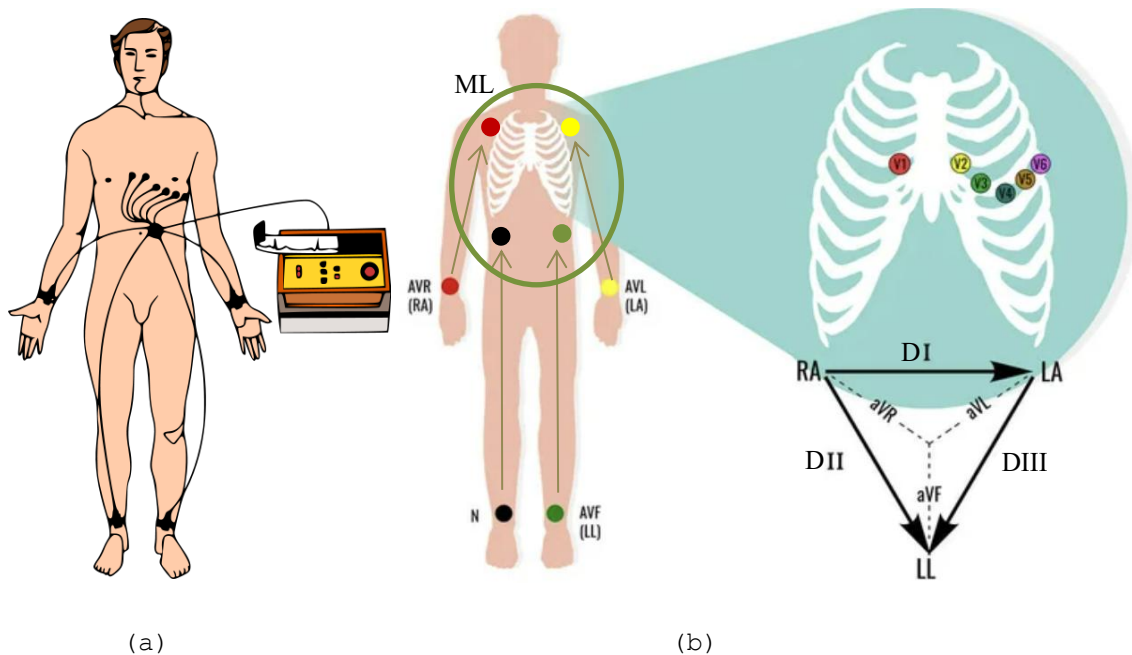
As derivações do ECG podem ser divididas em duas categorias principais (FRANCIS, 2016; SOUTO, 2016):

- **Derivações bipolares ou periféricas:** Essas derivações são obtidas usando eletrodos positivos e negativos colocados nos membros superiores e inferiores do corpo. Na ver Figura 4.b pode-se observar as derivações DI com o percurso aVR, aVL, DII com o percurso aVR, aVF, e DIII com o percurso aVL, aVF, que são obtidas colocando-se os eletrodos nos membros superiores e inferiores.
  - Também pode-se observar na Figura 4.b as derivações modificadas (*Modified Lead, ML*) nos 4 pontos posicionados no abdome, sendo variações das derivações bipolares padrão, geralmente adaptadas para uso em monitoramento contínuo de mais longa duração, permitindo maior conforto do paciente que pode se mover com

maior facilidade. Estas derivações modificadas seguem a mesma lógica das derivações bipolares padrão, modificando a posição dos eletrodos dos membros superiores e inferiores para as posições equivalentes no tórax, sendo então denominadas de derivações MLI, MLII e MLIII, respectivamente, com os mesmos percursos já citados.

- **Derivações precordiais:** Essas derivações são obtidas usando eletrodos posicionados em pontos específicos do tórax (V1, V2, V3, V4, V5, V6). Essas derivações fornecem uma visão frontal do coração e são úteis na detecção de problemas cardíacos anteriores e laterais.

**Figura 4:** Derivações dos eletrodos para medida dos sinais ECG.



**Fonte:** a, imagem de domínio público. b, adaptado do Banco de Imagens DepositPhotos.

As derivações são medidas em relação a uma referência neutra, sendo esta a derivação “zero” ou “terra”, localizada normalmente no tornozelo direito (ver Figura 4.b). A atividade elétrica do coração é registrada como diferenças de potencial entre os eletrodos positivos e negativos em cada derivação, e essas diferenças de potencial são representadas graficamente no ECG como ondas e complexos.

As ondas em um sinal ECG são deflexões positivas ou negativas na linha de base do traçado que representam eventos elétricos específicos durante o ciclo cardíaco (BETTS

*et al.*, 2022). Cada onda corresponde a uma fase particular da atividade elétrica do coração. A Figura 5 apresenta parte de um ECG destacando um intervalo de um segundo onde se pode observar as principais ondas de um sinal cardíaco padrão (HALL, 2021; BETTS *et al.*, 2022):

- **Onda P:** Representa a despolarização atrial, que é a contração dos átrios.
- **Onda Q:** É a primeira deflexão negativa após a onda P. Pode indicar um evento de condução anômala ou infarto.
- **Onda R:** É a primeira deflexão positiva após a onda Q que é geralmente alta e proeminente. Representa a despolarização ventricular.
- **Onda S:** É a deflexão negativa após a onda R e antes da onda T. Também está relacionada à despolarização ventricular.
- **Onda T:** Indica a repolarização ventricular, quando as câmaras do coração se preparam para a próxima contração.
- **Onda U:** Uma pequena onda que pode aparecer após a onda T, e antes da P.

Os complexos no ECG são conjuntos de ondas que ocorrem juntas e representam fases específicas do ciclo cardíaco, destacando-se (HALL, 2021; BETTS *et al.*, 2022):

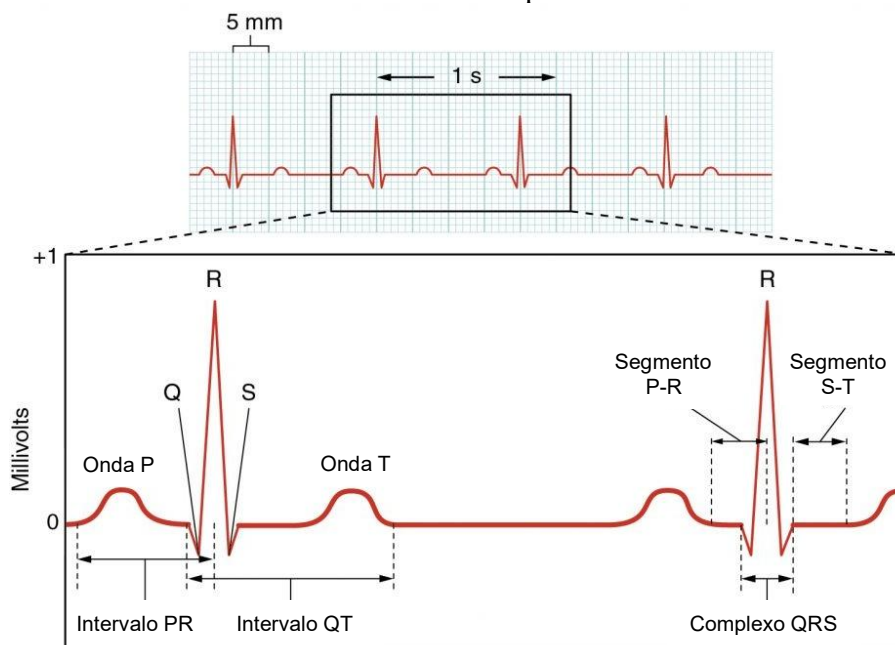
- **Complexo QRS:** Composto pelas ondas Q, R e S, reflete a despolarização ventricular, que é a contração dos ventrículos.

Além dos complexos QRS, o ECG também apresenta outros componentes importantes tais como os intervalos RR, PR e QT e segmentos PR e ST que são definidos como:

- **Intervalo RR:** É a medida da distância entre dois complexos QRS consecutivos e representa o intervalo de tempo entre dois batimentos cardíacos consecutivos. É um indicador importante para a análise da frequência cardíaca e pode ser usado para identificar arritmias e outros distúrbios do ritmo cardíaco.
- **Segmento PR:** É o segmento do ECG após a onda P até a onda R e representa o tempo de condução elétrica do nódulo atrioventricular. É um indicador da duração da condução elétrica do impulso cardíaco entre as câmaras atriais e ventriculares e pode ser utilizado para identificar anormalidades na condução cardíaca.

- **Segmento ST** É o segmento do ECG após a onda S e antecede a onda T. Representa o período de repolarização ventricular, ou seja, a recuperação dos ventrículos após a contração. Alterações no segmento ST podem indicar isquemia ou lesão cardíaca e são usadas na avaliação de infarto do miocárdio e outras condições cardíacas.
- **Intervalo PR:** é o período que se estende desde o início da onda P até o pico do complexo QRS. Ele representa o tempo necessário para a ativação elétrica do átrio (onda P) e a propagação dessa ativação através do nó atrioventricular até os ventrículos (complexo QRS). O Intervalo PR é uma medida da condução elétrica do coração e é usado para avaliar atrasos ou bloqueios na transmissão elétrica entre os átrios e os ventrículos.
- **Intervalo QT:** é o período que se estende desde o início do complexo QRS até o final da onda T no ECG. Ele representa a duração total da despolarização e repolarização dos ventrículos cardíacos, ou seja, o tempo que leva para os ventrículos se contraírem e depois relaxarem elétrica e mecanicamente. O Intervalo QT é crítico na avaliação da função cardíaca, pois sua duração pode estar relacionada a problemas de ritmo cardíaco, incluindo o risco de arritmias graves.

**Figura 5:** Formas de onda em um sinal de ECG padrão e um traçado normal mostrando suas ondas e complexos.



**Fonte:** Adaptado de Betts *et al.* (2022, p. 780).

Em resumo, ondas são as deflexões individuais que representam eventos elétricos específicos, enquanto complexos são combinações de ondas que indicam fases completas do ciclo cardíaco. Essas definições são essenciais para a interpretação de um ECG e a análise da função cardíaca.

## 2.2 ARRITMIAS CARDÍACAS

De acordo com Hall (2021), arritmias cardíacas são alterações na frequência, ritmo ou condução elétrica do coração, podendo ser classificadas como:

### 1) Arritmias por frequência:

- a) Taquicardia: frequência cardíaca rápida, geralmente definida em uma pessoa adulta como mais rápida que 100 batimentos/min. Algumas de suas causas incluem aumento da temperatura corporal, estimulação do coração pelos nervos simpáticos ou condições tóxicas do coração.
- b) Bradicardia: de forma oposta, ocorre quando a frequência cardíaca se torna lenta, geralmente definida como menos de 60 batimentos/min.
- c) Arritmia sinusal: sendo um distúrbio do ritmo cardíaco, causando batidas irregulares, mais rápidas ou mais lentas que o normal, com variações vinculadas ao ritmo respiratório.

### 2) Arritmias por origem anatômica:

- a) Supraventriculares: Geradas acima dos ventrículos, ou seja, nos átrios ou no nó atrioventricular. É caracterizada por um ritmo cardíaco irregular e rápido causado pela ativação desorganizada dos átrios, reduzindo a eficácia do bombeamento atrial causando palpitações, cansaço e falta de ar.
- b) Ventriculares: Origem nos ventrículos, que são as câmaras inferiores do coração, sendo mais graves porque podem comprometer a ejeção de sangue. Podem ser mais perigosas devido ao impacto direto na função de bombeamento do coração e ao maior risco de evoluírem para fibrilação ventricular, levando à perda de contração e risco iminente de parada cardíaca, causando tontura, desmaios e colapso circulatório.

### 3) Arritmias por condução:

- a) Bloqueios atrioventriculares: quando há atraso ou interrupção na condução entre

os átrios e os ventrículos, com condução mais lenta do impulso elétrico, com intervalo PR prolongado no ECG. Muitas vezes é assintomático, mas pode causar tonturas em graus avançados.

- b) Bloqueios de ramo: quando há falha na condução nos ramos direito ou esquerdo de setores dos ventrículos, podendo causar desconforto no peito, fadiga e falta de ar.
- 4) Outras arritmias:
- a) Extrasístoles: são batimentos cardíacos “extras” ou prematuros, que podem surgir nos átrios (extrasístoles atriais) ou nos ventrículos (extrasístoles ventriculares), causando batimentos fortes no peito.
  - b) Síndrome do QT longo: sendo um distúrbio no tempo de repolarização ventricular, prolongando o intervalo QT no ECG, causando palpitações e desmaios.

A observação de arritmias ocorre a partir do ciclo elétrico cardíaco, que é o processo responsável por gerar e conduzir os impulsos elétricos que controlam a contração coordenada do coração, o qual envolve três estruturas principais (HALL, 2021):

- 1) **Nó Sinotrial - O Marcapasso Natural:** gerador primário dos impulsos elétricos do coração. O impulso elétrico se espalha rapidamente pelos átrios direito e esquerdo, causando sua contração e empurrando o sangue para os ventrículos. Representado no ECG pela onda P, que reflete a despolarização atrial.
- 2) **Nó Atrioventricular:** atua atrasando levemente o impulso elétrico antes de transmiti-lo aos ventrículos. Esse atraso permite que os átrios esvaziem completamente o sangue nos ventrículos antes de sua contração. Esse intervalo é visível no ECG como o segmento PR, que representa o tempo de condução entre átrios e ventrículos.
- 3) **Fibras de Purkinje e Condução pelos Ventrículos:** promovem a rápida despolarização e contração simultânea dos ventrículos (sístole ventricular), ejetando o sangue para a circulação pulmonar (ventrículo direito) e sistêmica (ventrículo esquerdo). No ECG, essa fase é representada pelo complexo QRS, que reflete a despolarização ventricular, e a Onda T, que representa a repolarização ventricular, preparando o coração para o próximo ciclo.

### 2.3 SISTEMAS AUTOMÁTICOS PARA CLASSIFICAÇÃO DE ARRITMIAS

A análise direta dos registros de sinais de ECG, para identificação de arritmias cardíacas, deve ser realizada por especialistas, e considerando o volume de dados gerados pelos gráficos dos sinais ECG, é um trabalho extremamente desafiador e que exige muita atenção do profissional. A implementação de sistemas automáticos para classificação de arritmias e apoio ao diagnóstico médico pode proporcionar benefícios tanto para os pacientes quanto para os profissionais de saúde (O'BRIEN, 2011), podendo-se citar as seguintes vantagens para seu uso:

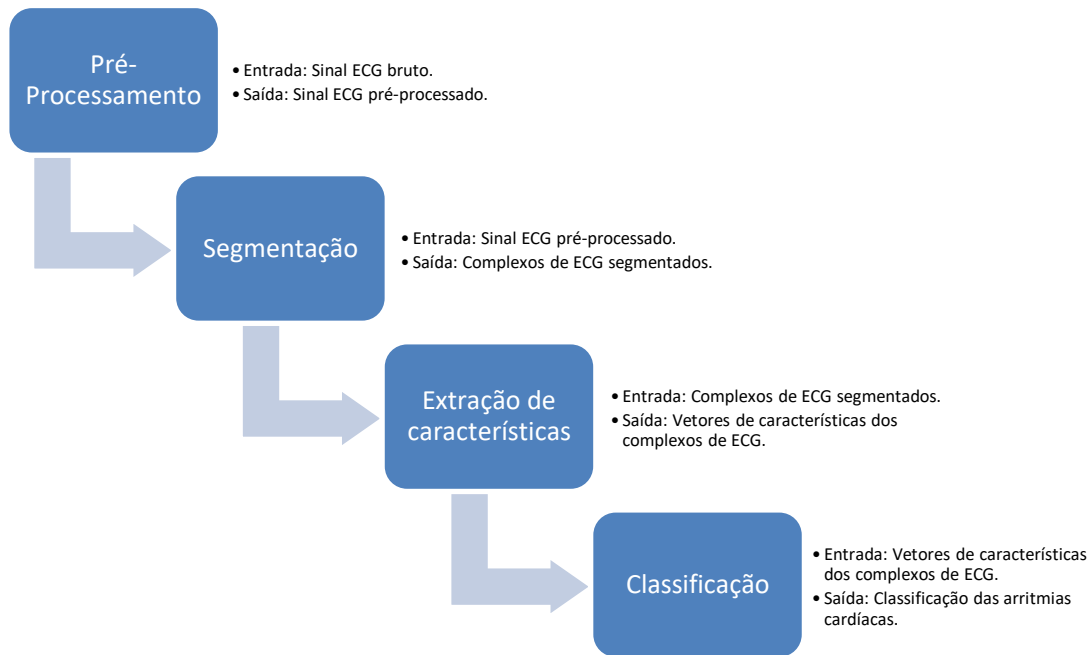
- **Eficiência:** Um sistema automático pode processar grandes volumes de dados, como de sinais de ECG, de forma rápida e eficiente, superando a capacidade humana de análise manual. Isso é especialmente útil em cenários onde o tempo é crucial, como em unidades de emergência ou ambientes hospitalares ocupados.
- **Padronização:** A interpretação de sinais de ECG é um processo altamente especializado e pode variar entre profissionais de saúde. Um sistema automático oferece uma abordagem padronizada e consistente para a classificação de arritmias, reduzindo as variações e melhorando a precisão dos diagnósticos.
- **Deteção precoce:** Algoritmos de classificação automatizada podem identificar padrões sutis ou características ocultas nos sinais de ECG que podem passar despercebidos durante a análise visual do especialista. Isso pode levar à deteção precoce de arritmias, permitindo um tratamento mais rápido e melhor prognóstico para os pacientes.
- **Redução de erros:** A análise manual de sinais de ECG está sujeita a erros humanos, como fadiga, distração ou subjetividade na interpretação dos padrões. Um sistema automático ajuda a minimizar esses erros, melhorando a precisão e a confiabilidade dos resultados.
- **Suporte à tomada de decisão:** Um sistema automatizado pode fornecer informações valiosas aos médicos, auxiliando na tomada de decisões clínicas. Ele pode apresentar resultados de classificação, indicar a gravidade da arritmia, informar onde foram detectadas para a conferência

do especialista, e até mesmo sugerir opções de tratamento com base em protocolos ou diretrizes médicas.

- **Monitoramento contínuo:** Em situações em que o monitoramento contínuo é necessário, um sistema automático pode analisar e alertar imediatamente sobre qualquer mudança nas condições cardíacas do paciente, permitindo uma avaliação e intervenção rápida quando necessário.

Um sistema automático para classificação de arritmias cardíacas tem seu fluxo amplamente discutido na literatura (LUZ *et al.*, 2016; HALL, 2021; DIAS *et al.*, 2021) e geralmente envolve várias etapas, incluindo pré-processamento dos sinais cardíacos, segmentação dos complexos de ECG, extração de características relevantes e classificação das arritmias, como ilustrado na Figura 6.

**Figura 6:** Fluxo para classificação automática de arritmias cardíacas.



**Fonte:** Próprio Autor.

Na etapa de Pré-processamento, o sinal de ECG é submetido a técnicas para remover artefatos, filtrar ruídos indesejados e melhorar a qualidade do sinal. Isso pode envolver filtragem passa-baixa e/ou passa-alta, remoção de linha de base, cancelamento de artefatos musculares, normalização e digitalização do sinal. Com o sinal pré-processado e digitalizado, inicia-se a fase de segmentação do sinal sendo esta etapa

utilizada para identificação e isolamento do complexo QRS, segmentos e intervalos, que são as principais características do ECG que representam a atividade elétrica do ventrículo cardíaco, entre outras ondas. A segmentação é uma etapa crucial em muitas aplicações de processamento de ECG, como detecção de arritmias, análise de ritmo cardíaco, análise de variabilidade da frequência cardíaca e outras análises quantitativas. É importante ressaltar que a segmentação é uma etapa desafiadora devido à variabilidade nas características, sendo o complexo QRS um dos mais importantes, pois representa a despolarização ventricular, ou seja, a ativação elétrica dos ventrículos do coração (HALL, 2021). A análise do complexo QRS é essencial para avaliar a função cardíaca, diagnosticar arritmias e identificar possíveis distúrbios na condução elétrica do coração. Além da segmentação dos complexos QRS, o ECG também pode ser analisado em relação a outros componentes importantes, como os intervalos RR, segmentos PR, ST e as ondas U, P e T.

A segmentação é realizada para obter as informações relevantes do sinal as quais serão usadas na etapa seguinte, de Extração de Características, que está focada na identificação e quantificação de características relevantes do sinal ECG, que fornecem informações distintivas para diferenciar as arritmias, cujos métodos são executados no domínio do tempo ou da frequência (SINGH; KRISHNAN, 2023):

- **No domínio do tempo:** As características temporais estão relacionadas às medidas de tempo do sinal de ECG, como duração do complexo QRS, intervalo PR, intervalo QT e intervalos RR. Essas medidas podem fornecer informações sobre a frequência cardíaca, a condução atrioventricular e a repolarização ventricular. Existem várias técnicas e métodos de extração de recursos disponíveis, tais como recursos estatísticos (ALOTAIBY *et al.*, 2019), modelagem autorregressiva multivariada e autorregressiva escalar (VANEGHI *et al.*, 2012), Algoritmo Pan-Tompkins (PAN; TOMPKINS, 1985; ELGENDI, 2013), Codificação preditiva linear (LIN; CHANG, 1989), entre outras.
- **No domínio frequência:** As características de domínio da frequência exploram as propriedades espectrais do sinal de ECG. Elas podem incluir frequências dominantes, banda de frequência de energia, coeficientes de transformada de Fourier ou características de análise espectral de curto prazo, como: poder espectral ou razão de potência. Esta etapa foca em

características morfológicas relacionadas à forma e à amplitude dos complexos de ECG. Elas incluem amplitude máxima e mínima do complexo QRS, morfologia das ondas P e T, presença de anormalidades como inversões ou alargamento do complexo QRS. Podem ser extraídas utilizando técnicas de análise de forma de onda, como a transformada de Fourier (RODRÍGUEZ *et al.*, 2015) ou análise *Wavelet* (SEENA; YOMAS, 2014; MADEIRO *et al.*, 2020). Já no domínio do tempo, existem várias técnicas e métodos de extração de recursos disponíveis, tais como recursos estatísticos (ALOTAIBY *et al.*, 2019), modelagem autorregressiva multivariada e autorregressiva escalar (VANEGHI *et al.*, 2012), Algoritmo Pan-Tompkins (PAN; TOMPKINS, 1985; ELGENDI, 2013), Codificação preditiva linear (LIN; CHANG, 1989), entre outras.

Por fim, a etapa de classificação de arritmias é a fase final do processo de processamento automatizado de sinais de ECG, onde um modelo ou algoritmo é aplicado para identificar e classificar os diferentes tipos de arritmias cardíacas com base nas características extraídas. Existem diversos métodos de classificação que podem ser utilizados, sendo que entre eles os mais comumente utilizados são (ADEWOLE *et al.*, 2022):

- **Algoritmos de Floresta Aleatória:** são um tipo de aprendizado de máquina que combina várias árvores de decisão para tomar decisões. Cada árvore é treinada em uma amostra diferente dos dados e, durante a classificação, as árvores votam em uma decisão final como aplicado por Li e Zhou (2016).
- **Máquinas de Vetores de Suporte (SVM):** são algoritmos de aprendizado de máquina que buscam encontrar um hiperplano que melhor separe as diferentes classes de dados em um espaço de características. O objetivo é encontrar o hiperplano que maximize a margem entre as classes, ou seja, a distância entre o hiperplano e os pontos mais próximos de cada classe (os vetores de suporte), como aplicado por Zhang *et al.* (2014), Garcia *et al.* (2017) e Xu, Jeong e Li (2020).
- **Redes Neurais Recorrentes (RNNs):** são um tipo de arquitetura de redes neurais projetada para lidar com sequências de dados, como sinais

temporais ou linguagem natural. Elas possuem conexões recorrentes que permitem que informações sejam transmitidas entre os passos de tempo. Isso é especialmente útil para capturar dependências temporais em dados como aplicado por Islam *et al.* (2023).

- **Redes Neurais Convolucionais (CNNs):** Esses classificadores são capazes de aprender padrões complexos nos dados de ECG e são particularmente eficazes na extração de características relevantes de arritmias cardíacas, sendo amplamente utilizadas em pesquisas recentes para a classificação de arritmias como aplicados por Dias *et al.* (2021), Ahmad *et al.* (2021), Mewada (2023) e Zhou e Fang (2024).
- **Lógica *fuzzy*:** A lógica *fuzzy* é uma abordagem que lida com a incerteza e a imprecisão nos dados como aplicado por Frahan, Li e Ahmed (2018).
- **Redes bayesianas:** Esses classificadores utilizam redes bayesianas para modelar a relação entre as características extraídas dos sinais de ECG e as classes de arritmias. As redes bayesianas são capazes de lidar com a incerteza dos dados e podem fornecer probabilidades de classificação para cada arritmia, como aplicado por Casas *et al.* (2018).

Para desenvolvimento e testes dos sistemas automáticos classificadores de arritmia, dois paradigmas vêm sendo utilizados por diversos autores na literatura, sendo eles o paradigma interpaciente e o intrapaciente (LUZ *et al.*, 2016; HONG *et al.* 2020). Na abordagem interpaciente as bases de dados para treinamento e teste do classificador não possuem sinais segmentados de ECG de um mesmo paciente, enquanto na abordagem intrapaciente, uma parte dos sinais segmentados de ECG de um paciente pode ser utilizada na base de dados treino enquanto a outra parte, pode ser utilizada na base de dados de teste.

Long *et al.* (2015) e Zhao *et al.* (2021) realizaram estudos aplicados com estas duas abordagens, apresentando suas características, vantagens e desvantagens:

- **Paradigma intrapaciente:** Nesse paradigma, os modelos de classificação são treinados e testados ignorando a quais pacientes cada registro pertence. Ou seja, todos os dados são embaralhados e separados em treino e teste, de forma que diferentes dados do mesmo paciente estão presentes nos dados usados para treinamento e teste do modelo. Isso permite que o

modelo seja avaliado em sua capacidade de generalização para diferentes condições dos mesmos pacientes. Tem como vantagem levar em consideração a variabilidade intra-individual dos sinais ECG, como variações de ritmo cardíaco, alterações na morfologia das ondas, entre outros. E, geralmente, resulta em melhor desempenho do modelo, já que é treinado e testado nos dados específicos dos pacientes, considerando suas características únicas. No entanto, como desvantagens pode resultar em superestimação do desempenho do modelo, já que, embora não sejam os mesmos registros, o fato de serem registros de um mesmo paciente presente nos dois conjuntos treino e teste, pode levar a uma avaliação otimista da capacidade de generalização do modelo, não sendo adequado para modelos que visam generalização para diferentes pacientes ou populações.

- **Paradigma interpaciente:** Nesse paradigma, os modelos de classificação são treinados em um conjunto de dados de um grupo de pacientes e testados em outro conjunto de dados de pacientes diferentes. Ou seja, os dados de treinamento e teste são provenientes de pacientes distintos. Isso permite que o modelo seja avaliado em sua capacidade de generalização para novos pacientes, já que o modelo não teve acesso ao comportamento dinâmico dos registros dos pacientes cujos registros estão no conjunto de teste. Tem como vantagens a avaliação mais realista da capacidade de generalização do modelo, já que é testado em pacientes diferentes daqueles usados no treinamento, sendo útil quando o objetivo é criar um modelo de classificação que possam ser aplicados em diferentes populações de pacientes. No entanto, requer um conjunto de dados grande o suficiente para permitir a divisão entre conjuntos de treinamento e teste, o que pode ser difícil de obter em algumas situações práticas. Além disso, pode haver variações entre os pacientes, como diferenças na idade, sexo, etnia, presença de comorbidades, que podem afetar a performance do modelo. Geralmente, resulta em menor desempenho do modelo, já que os dados de treinamento podem ser limitados em sua representatividade da população geral de pacientes.

Em resumo, o paradigma interpaciente é mais adequado quando se busca a

generalização para diferentes populações de pacientes, enquanto o paradigma intrapaciente é mais adequado quando se busca considerar as variações intra-individuais dos sinais ECG. Ambos têm suas vantagens e desvantagens e a escolha do paradigma adequado depende do objetivo específico da pesquisa ou aplicação clínica.

### 2.3.1 *Sistemas automáticos e o padrão AAMI*

A Associação para o Avanço da Instrumentação Médica (AAMI, *Association for the Advancement of Medical Instrumentation*) é uma organização que estabelece diretrizes e padrões para a classificação automática de sinais ECG (ANSI/AAMI, 2020). A AAMI desenvolveu um conjunto de recomendações conhecido como Diretrizes da AAMI para a Classificação de Eletrocardiogramas (ANSI/AAMI EC57), que fornece orientações sobre a terminologia, os critérios de desempenho e as métricas de avaliação para sistemas de classificação automática de arritmias cardíacas.

As diretrizes da AAMI definem cinco classes principais de arritmias, que são amplamente utilizadas na classificação automática de sinais ECG, sendo elas:

- **Classe N (Normal):** Essa classe abrange batimentos cardíacos normais, sem nenhuma anormalidade significativa. Esses batimentos podem incluir o complexo QRS normal, ondas P e T normais e intervalos PR e QT dentro dos limites normais.
- **Classe S (Supraventricular):** Essa classe inclui arritmias que se originam acima dos ventrículos, como taquicardia sinusal, taquicardia atrial, fibrilação atrial, vibração atrial e bloqueio atrioventricular.
- **Classe V (Ventricular):** Essa classe abrange arritmias que se originam nos ventrículos, como taquicardia ventricular, fibrilação ventricular e complexos de prematuros ventriculares.
- **Classe F (Fusão de Normal e Ventricular):** Essa classe inclui arritmias que têm componentes tanto do complexo QRS normal e do complexo QRS ventricular ectópico (anormal), geralmente com uma aparência intermediária entre as duas formas.
- **Classe Q (Desconhecido):** Essa classe é reservada para casos em que a classificação não pode ser definitivamente determinada ou quando há

dúvidas sobre a interpretação do sinal de ECG.

A adoção das diretrizes da AAMI na classificação automática de sinais ECG ajuda a padronizar a terminologia e a abordagem de classificação, permitindo a comparação de resultados entre diferentes sistemas e facilitando a interpretação clínica dos resultados. Isso contribui para o desenvolvimento de sistemas mais confiáveis e precisos na detecção e classificação de arritmias cardíacas a partir de sinais de ECG.

A ANSI/AAMI (2020) destaca a importância de dados de alta qualidade e representativos para o desenvolvimento e avaliação de algoritmos de classificação de arritmias. Portanto, é prática comum na comunidade científica utilizar bases de dados conhecidas e amplamente aceitas para desenvolvimento e avaliação de algoritmos de classificação, e para isso, ela recomenda o uso de cinco bases, sendo elas:

1. **AHA:** *The American Heart Association Database for Evaluation of Ventricular Arrhythmia Detectors* (80 registros de pacientes de 35 minutos cada);
2. **CU:** *The Creighton University Sustained Ventricular Arrhythmia Database* (35 registros de pacientes de 8 minutos cada);
3. **ESC:** *The European Society of Cardiology ST-T Database* (90 registros de pacientes de 2 horas cada);
4. **MIT-BIH:** *The Massachusetts Institute of Technology - Beth Israel Hospital Arrhythmia Database* (48 registros de pacientes de 30 minutos cada);
5. **PTB:** *Physikalisch-Technische Bundesanstalt* (Instituto Nacional de Metrologia da Alemanha, em Berlin, com 549 registros de pacientes de 30 segundos a 5 minutos cada);
6. **NST:** *The Noise Stress Test Database* (12 registros de pacientes de 30 minutos cada, mais 3 registros com excesso de ruídos).

A escolha da base de dados depende dos objetivos da pesquisa, dos tipos de arritmias que se deseja estudar e das características específicas do algoritmo de classificação a ser desenvolvido. Considerando que a base MIT-BIH e a base PTB são bases usadas em nosso trabalho, elas serão apresentadas a seguir com mais detalhes.

### 2.3.2 Métricas de avaliação do classificador

A norma ANSI/AAMI EC57:2012 R2020 (ANSI/AAMI, 2020) especifica os indicadores que devem ser usados para avaliação de desempenho de classificadores de arritmias, sendo eles acurácia, precisão ou acertos positivos, sensibilidade (ou *recall*) e especificidade e adicionalmente, como bastante observado na literatura e recomendável para avaliação em bases muito desbalanceadas, a média harmônica entre precisão e sensibilidade (*F1-Score*).

A Acurácia indica uma performance geral do modelo, ou seja, dentre todas as classificações, quantas o modelo classificou corretamente, podendo ser calculada conforme:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

A Precisão ou Acertos Positivos é compreendida como a quantidade correta de todas as classificações de classe Positivas que o modelo fez. Em outras palavras, quantas arritmias ou normalidades detectadas de fato o modelo acertou, sendo calculada conforme:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

A Sensibilidade, também chamada de *Recall* é compreendida como os acertos de todas as situações de classe Positivo como valor esperado. Ou seja, indica o quanto o modelo está identificando os casos positivos corretamente, sendo útil para minimizar os falsos negativos, sendo calculado conforme:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

A Especificidade é a contrapartida da sensibilidade. No caso de diagnóstico de arritmias, usa-se sensibilidade para descrever quão bem o teste pode identificar a sua presença, e a especificidade para dizer quão bem o teste pode identificar a sua ausência, sendo calculada conforme:

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

O *F1-Score* é média harmônica entre precisão e *recall*, sendo de especial importância ao avaliar bases de dados muito desbalanceadas entre suas classes, indicando o quão alta ou baixa as duas são, sendo calculada conforme:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

### 2.3.3 Base de dados MIT-BIH

A MIT-BIH vem sendo utilizada para desenvolvimento dos classificadores, sendo esta base uma das mais antigas e amplamente utilizada na área de processamento de sinais ECG (LUZ *et al.*, 2016; HONG *et al.*, 2020), sendo disponibilizada publicamente na PhysioNet (GOLDBERGER *et al.*, 2000). A MIT-BIH contém uma variedade de arritmias cardíacas, incluindo taquicardia sinusal, taquicardia atrial, fibrilação atrial, vibração atrial, bloqueio atrioventricular e diferentes tipos de taquicardia ventricular. Essa diversidade de arritmias permite o desenvolvimento e a avaliação de algoritmos de classificação capazes de identificar uma ampla gama de anormalidades cardíacas, sendo por isso, uma das bases mais utilizadas na literatura para desenvolvimento de sistemas classificadores.

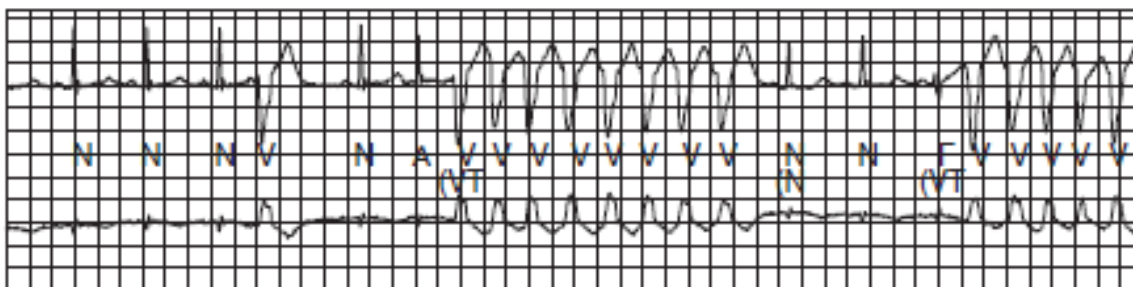
A MIT-BIH foi criada no início dos anos 1980 por pesquisadores do MIT e do *Beth Israel Hospital* em Boston, EUA, com registros coletados entre 1975 e 1979, e tem sido uma referência importante para a avaliação e desenvolvimento de algoritmos de processamento de ECG, detecção de arritmias, análise de complexos QRS e outras aplicações em ECG.

É composta por uma coleção de 48 registros de ECG de 47 pacientes reais, que foram digitalizados e anotados para fins de pesquisa. Os registros de ECG foram coletados de pacientes hospitalizados e ambulatoriais do *Beth Israel Hospital*, com consentimento informado dos pacientes e em conformidade com os regulamentos éticos e de privacidade aplicáveis.

Os registros de ECG na base MIT-BIH são organizados em pares, com cada par contendo um registro de ECG de prolongado com a derivação MLII e mais uma derivação precordial. Além disso, a base MIT-BIH também inclui anotações detalhadas dos complexos QRS em todos os registros de ECG (ver Figura 7) com indicação do tipo de arritmia. Essas anotações foram feitas manualmente por especialistas em ECG, que marcaram os complexos QRS, segmentos ST, ondas P e outros eventos relevantes em

cada registro. As anotações foram revisadas e validadas por outros especialistas em ECG para garantir sua precisão.

**Figura 7:** Dez segundos do registro 205 do banco de dados de arritmia do MIT-BIH, contendo anotações de ritmo sinusal normal “N” e taquicardia ventricular “VT” da derivação MLII.



**Fonte:** Moody e Mark (2001).

Os registros e anotações na base MIT-BIH têm sido amplamente utilizados para o desenvolvimento e validação de algoritmos de processamento de ECG, detecção de arritmias, classificação de complexos QRS, análise de variabilidade da frequência cardíaca e outras aplicações em pesquisa cardiovascular. Eles têm sido uma referência importante para a comunidade científica na avaliação de desempenho e comparação de diferentes métodos e algoritmos em ECG.

Os 48 registros de ECG na base de dados MIT-BIH são categorizados em várias classes, que representam diferentes tipos de arritmias e condições cardíacas, com pacientes variando de 32 a 89 anos de idade, com frequência de amostragem de 360 Hz, totalizando aproximadamente 30 minutos de registro de sinal ECG por paciente. Algumas das classes incluídas na base de dados MIT-BIH são: taquicardia supraventricular, bloqueio de ramo, arritmias ventriculares, fibrilação atrial, entre outros. As classes são identificadas nas anotações dos registros de ECG e são usadas para avaliar o desempenho de algoritmos de diagnóstico automático em diferentes condições clínicas. O Quadro 1 apresenta a organização das anotações contidas da base MIT-BIH considerando as diretrizes da AAMI que definem as cinco classes principais de arritmias.

**Quadro 1:** Organização das anotações contidas da base MIT-BIH.

Classes	Símbolo	Classe
N	N	Batimento normal.
<b>Qualquer batimento não categorizado como S, V, F ou Q.</b>	L	Batimento de bloqueio de ramo esquerdo.
	R	Batimento de bloqueio de ramo direito.
	e	Batimento de escape atrial.
	j	Batida de escape nodal (juncional).
SVEB (S)	A	Batimento atrial prematuro.
<b>Batimento ectópico supraventricular.</b>	a	Batimento atrial prematuro aberrado.
	J	Batimento prematuro nodal (juncional).
	S	Batimento prematuro supraventricular.
VEB (V)	V	Contração ventricular prematura.
<b>Batimento ectópico Ventricular.</b>	E	Batimento de escape ventricular.
	F	Fusão dos batimentos ventriculares e normais.
<b>Batimento de Fusão.</b>	F	
Q	P ou /	Batida ritmada.
<b>Batimento desconhecido.</b>	f	Fusão de batida ritmada e normal.
	U	Batida inclassificável.

**Fonte:** Moody e Mark (2005).

O Quadro 2 apresenta a distribuição de batimentos por classes de cada um dos 48 registros contidos na base, sendo 4 registros de pacientes com marca-passo. Já é possível observar um importante desafio para a classificação automatizada que é o forte desbalanceamento entre os registros de batimentos disponíveis para cada classe, contendo mais de 90 mil registros para a classe Normal e as demais classes com menos de 10 mil.

**Quadro 2:** Distribuição dos batimentos cardíacos na base MIT-BIH.

Classe AAMI	Tipos de batimentos	% do total	Quantidade de batimentos
Normal	N, e, j, L, R	82,78	90.631
SVEB (S)	A, a, J, S	2,54	2.781
VEB (V)	V, E	6,60	7.236
F	F	0,73	803
P (marca-passo)	/, f	7,32	8.010
Q	U	0,03	33
Total		<b>100,00</b>	<b>109.494</b>

**Fonte:** Moody e Mark (2005).

### 2.3.4 Base de dados PTB

A base de dados PTB (BOUSSELJOT, 2004) foi desenvolvida no Instituto Nacional de Metrologia da Alemanha, em Berlin, através dos trabalhos de Boussejot (1994), Boussejot e Kreiseler (1995) e Kreiseler e Boussejot (1995), sendo também disponibilizada publicamente na PhysioNet (GOLDBERGER *et al.*, 2000). É uma coleção de 549 sinais ECGs com 15 derivações diferentes e de 294 indivíduos, entre saudáveis e pacientes com uma variedade de doenças cardíacas.

É importante destacar que esta base de dados não possui os dados rotulados de acordo com a recomendação da AAMI, sendo altamente desbalanceada em relação à quantidade de registros das classes disponíveis. Seguem as características observadas:

1. 290 indivíduos são documentados com idade entre 17 a 87 anos, com média de 57,2 anos. 209 são homens com média de idade de 55,5, e 81 mulheres, com média de idade de 61,6. No entanto, não há registro de idade para 1 mulher e 14 homens. Estes indivíduos totalizam 549 registros de sinais ECG.
2. Para 22 indivíduos, não há nenhuma informação documentada sobre seu estado clínico, sem nenhuma rotulação nos dados.
3. Para 4 registros, identificados com os códigos 124, 132, 134 e 161, não há nenhuma identificação.
4. Todos os registros incluem 15 sinais medidos simultaneamente, sendo as 12 derivações convencionais (DI, DII, DIII, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6) junto com mais 3 outras derivações denominadas de Frank (Vx, Vy e Vz). Cada sinal é digitalizado a 1.000 amostras por segundo.
5. Cada registro possui um tempo de gravação e quantidade de amostras diferente, variando de 30 s a pouco mais de 5 min.

Portanto, do total de 294 indivíduos, apenas 268 apresentam dados clínicos documentados e rotulados, úteis para experiências em classificação automatizada, de acordo com o Quadro 3. Em estudos de classificação automatizada que utilizam esta base, predomina a escolha por dados de registros “Saudáveis” e “Infarto do Miocárdio” por questões práticas, como o desbalanceamento de classes e a relevância clínica. São classes altamente majoritárias, e Infarto do Miocárdio é uma condição clínica de alta prevalência e gravidade, sendo uma das principais causas de mortalidade global, tendo alta relevância prática. Além disso, usar indivíduos saudáveis como referência facilita a validação de modelos, pois permite estabelecer um “padrão normal” em oposição às condições patológicas, sendo o infarto no miocárdio, nesses trabalhos, chamado de “padrão anormal”.

**Quadro 3:** Distribuição dos dados rotulados na base PTB.

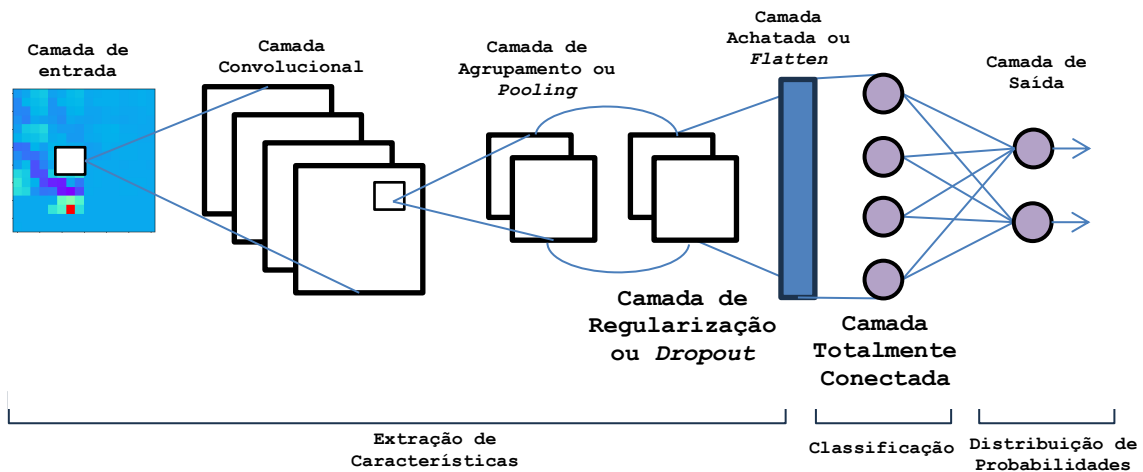
Classes rotuladas	Quantidade de indivíduos
Infarto do miocárdio	148
Cardiomiopatia/Insuficiência cardíaca	18
Bloqueio de ramo	15
Disritmia	14
Hipertrofia miocárdica	7
Doença cardíaca valvular	6
Miocardite	4
Saudáveis	52

**Fonte:** Bousseljot (2004).

## 2.4 REDES NEURAIAS CONVOLUCIONAIS

Uma Rede Neural Convolutiva (CNN – *Convolutional Neural Network*) é uma arquitetura de rede neural artificial, variante do Perceptron de múltiplas camadas (MLP - *Multi Layer Perceptron*) proposto por Rumelhart, Hinton e Williams (1986), sendo comumente usada em aplicações de detecção, classificação e reconhecimento em imagens e vídeos. As CNNs, como ilustrado na Figura 8, tem sua estrutura elaborada a partir de três etapas principais: extração de características, classificação e distribuição de probabilidade, com cada uma das etapas com camadas específicas que desempenham papéis distintos. Na etapa de extração de características, a camada de entrada recebe imagens, que são representados por tensores multidimensionais que representam os padrões de cores, por exemplo. A camada convolutiva extrai padrões locais, como bordas, texturas e formas, usando filtros (*kernels*) aplicados de forma espacial. Em seguida, a camada de agrupamento ou *pooling* reduz a dimensionalidade espacial, agrupando os valores em regiões menores, o que aumenta a robustez do modelo a variações e reduz o custo computacional. A camada de regularização, ou *dropout*, pode ser inserida, zerando aleatoriamente valores para prevenir *overfitting*. A etapa de classificação começa com a camada achatada ou *flatten*, que transforma os mapas de características 2D em um vetor 1D. Esse vetor é passado para a camada totalmente conectada, onde todos os neurônios se conectam para aprender combinações globais das características extraídas. Por fim, na etapa de distribuição de probabilidade, a camada de saída gera a predição final, geralmente utilizando uma função de ativação como *Softmax* (para classificação) ou linear (para regressão), fornecendo probabilidades associadas às classes ou valores contínuos. Essa organização permite que a CNN aprenda hierarquias complexas de características em busca de resultados precisos.

Figura 8: Rede Neural Convolucional



Fonte: Próprio Autor.

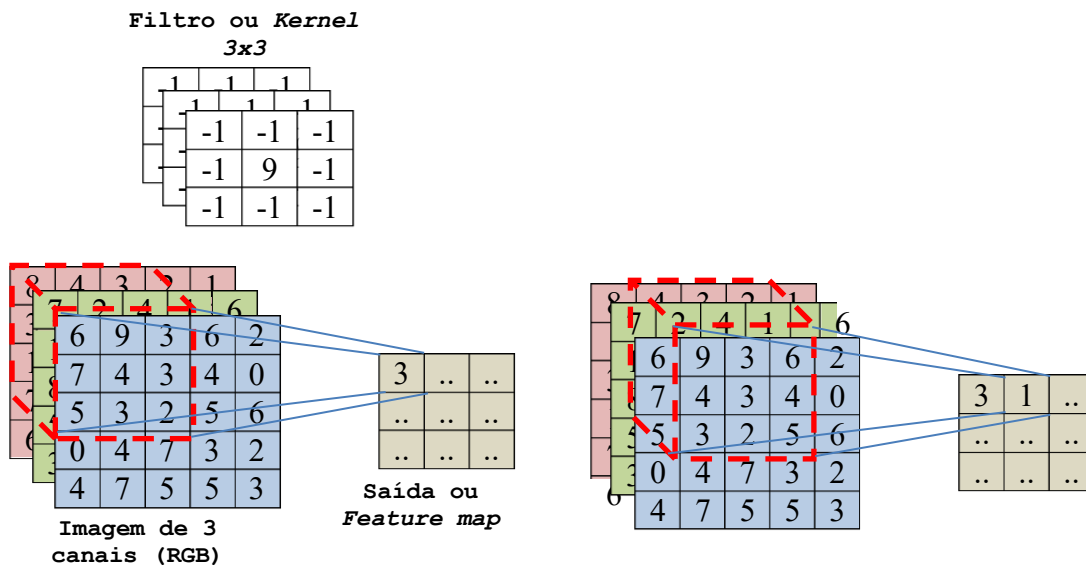
Uma RNA é considerada “profunda” quando possui múltiplas camadas intermediárias entre a camada de entrada e a camada de saída. O número mínimo de camadas intermediárias necessárias para uma rede neural ser considerada “profunda” pode variar dependendo do contexto, mas se aplica ao contexto das Redes Neurais Convolucionais (CNNs) devido à quantidade de camadas intermediárias para sua construção. Redes neurais mais profundas têm a capacidade de aprender características mais complexas e abstratas dos dados de entrada, permitindo um melhor desempenho em tarefas mais desafiadoras.

A Figura 9 ilustra como ocorre a operação de convolução em uma rede CNN iniciando com uma imagem RGB (3 canais) e gerando como saída o mapa de recursos (*feature map*). A convolução é um processo matemático que consiste em somar os produtos dos elementos de duas funções, sendo uma delas a função de entrada e a outra, uma função denominada filtro ou *kernel*. Na CNN, a convolução é usada para extrair recursos (*features*) da imagem de entrada que são importantes para a tarefa de classificação. O passo (*stride*) refere-se ao número de *pixels* que o filtro da convolução (cubo tracejado na Figura 9) se desloca a cada iteração.

O preenchimento (*padding*) é uma técnica utilizada para manter o tamanho da imagem de entrada após a convolução, uma vez que a aplicação de filtros de convolução pode diminuir a dimensão da entrada. O *padding* adiciona uma camada de *pixels* com valores predefinidos ao redor da entrada, de modo que a convolução possa ser aplicada sem alterar suas dimensões. Existem dois tipos principais de *padding*. Esse

preenchimento pode ser de um valor constante, como uma moldura ao redor da imagem, ou de forma reflexiva, completando com *pixels* já existentes na imagem, entre outras formas. O uso de *padding* também ajuda a destacar dados nas bordas das imagens que sejam relevantes para a extração de recursos, pois sem ele, o filtro passará apenas uma vez nos *pixels* de borda da imagem. Tal característica de *padding* pode não ser relevante para características localizadas no centro de uma imagem, como um rosto humano, mas pode ser relevante para outros casos cujos dados de borda contenham mais informação.

**Figura 9:** Operação de convolução.



**Fonte:** Próprio Autor.

A CNN então aplica ao mapa de características resultante uma função de ativação para introduzir não-linearidade na rede. Isso é importante para que a CNN possa aprender a representar e discriminar padrões complexos na imagem de entrada. As funções de ativação mais comumente utilizada é a função ReLU (*Rectified Linear Unit*).

*Pooling* é uma operação utilizada para reduzir o tamanho espacial do mapa de características, diminuindo o número de parâmetros e cálculos na rede neural, o que pode ajudar a evitar o *overfitting* (sobreajuste). Essa operação é geralmente realizada com um filtro de tamanho fixo (como 2x2), que é deslizado sobre o mapa de características. Existem diferentes tipos de *pooling*, sendo os mais comuns: *Max Pooling*, que seleciona o maior valor presente na região coberta pelo filtro, e *Average Pooling*, que calcula a média dos valores dentro da região do filtro. Além disso, variantes menos comuns, como o *Global Pooling*, aplicam *pooling* em toda a dimensão espacial de um mapa de

características (produzindo um único valor por canal), e o *L2 Pooling*, que calcula a raiz quadrada da soma dos quadrados dos valores na região. Esses métodos de *pooling* objetivam a compactação das informações mais relevantes do mapa de características, contribuindo para uma maior eficiência e generalização da rede neural.

As etapas de convolução, ativação e *pooling* são repetidas várias vezes em camadas sucessivas da CNN, formando uma hierarquia de recursos cada vez mais abstratos. Isso permite que a rede neural aprenda a representar padrões cada vez mais complexos na imagem de entrada.

A operação de convolução é o bloco de construção básico das CNNs e é responsável por sua grande eficácia em tarefas de visão computacional, como reconhecimento de imagens e detecção de objetos.

Assim, as CNNs são especialmente úteis na análise de dados com estrutura espacial, como imagens e sinais de áudio. Por exemplo, uma CNN pode ser treinada para reconhecer objetos em imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), detectar anomalias em sinais de vibração de máquinas (YUN *et al.*, 2021) ou classificar sinais ECG em diferentes arritmias cardíacas (HONG *et al.*, 2020).

O aprendizado da CNN é realizado por meio do treinamento supervisionado, onde a rede é apresentada com dados de entrada e sua saída correspondente. O objetivo é ajustar os pesos das camadas de forma a minimizar o erro entre a saída da rede e a classe verdadeira do dado de entrada. Para calcular o erro, são usadas diferentes funções de custo, como a entropia cruzada (*Cross Entropy*) e perda de dobradiça (*Hinge Loss*).

A entropia cruzada é uma medida comum de erro usada em classificação multiclasse. Ela calcula a diferença entre a distribuição de probabilidade prevista pela rede e a distribuição verdadeira das classes. A perda de dobradiça, por outro lado, é particularmente aplicada em algoritmos de máquinas de vetores de suporte (SVM, *Support Vector Machine*), medindo a distância entre a pontuação prevista pelo modelo e o limite de decisão. Se a pontuação já está além do limite de decisão, a perda é zero. Caso contrário, a perda é proporcional à distância até o limite de decisão. O objetivo é maximizar a margem de separação entre as classes, garantindo que os exemplos de treinamento estejam corretamente classificados.

A quantidade de parâmetros de uma CNN é determinada pela sua arquitetura, ou seja, pelo número de camadas e pelo tamanho dos filtros e mapas de recursos em cada

camada. A complexidade computacional da rede é diretamente proporcional ao número de parâmetros, o que afeta o tempo de treinamento e a memória necessária para armazenar o modelo, além da quantidade de imagens de entrada e de seus formatos, como resolução e quantidade de canais. Além disso, uma rede com muitos parâmetros pode levar a *overfitting*, ou seja, a uma rede que se adapta demais aos dados de treinamento e não generaliza bem para novos dados. Portanto, é importante encontrar um equilíbrio entre a complexidade da rede e seu desempenho na tarefa de classificação, sendo este também um desafio de estudo.

### 2.4.1 Modelos de CNN

Os modelos de CNN para classificação evoluíram significativamente nas últimas décadas, tornando-se cada vez mais profundos e precisos. A seguir tem-se alguns dos principais modelos de classificação de CNN já propostos:

- **LeNet-5:** proposto em 1998 por LeCun *et al.* (1998), sendo o primeiro modelo de CNN a ter sucesso em reconhecer caracteres manuscritos. Consiste em sete camadas, incluindo duas camadas convolucionais e três camadas totalmente conectadas.
- **AlexNet:** proposto em 2012 por Krizhevsky, Sutskever e Hinton (2012), foi o primeiro modelo de CNN a usar uma profundidade significativa, com 8 camadas (5 convolucionais e 3 totalmente conectadas), considerando ainda as limitações de GPU para a época. Foi treinado no conjunto de dados ImageNet e alcançou um erro de classificação de 15,3%, superando significativamente os modelos anteriores.
- **VGGNet:** proposto em por Simonyan Zisserman (2014), utilizou uma arquitetura uniforme com 19 camadas (16 convolucionais e 3 totalmente conectadas), tornando-o um dos modelos mais profundos até aquele momento. Foi treinado no conjunto de dados ImageNet e alcançou um erro de classificação de 7,3%.
- **GoogLeNet / Inception:** proposto em 2014 por Christian Szegedy *et al.* (2014), sendo uma arquitetura “Inception” com módulos de convolução em paralelo, sendo que reduziu a quantidade de parâmetros necessários em comparação com os modelos anteriores. Foi treinado no conjunto de dados

ImageNet e alcançou um erro de classificação de 6,7%.

- **ResNet:** proposto em 2015 por He *et al.* (2015), sendo uma arquitetura residual que permitiu o treinamento de modelos ainda mais profundos (até 152 camadas). Foi treinado no conjunto de dados ImageNet e alcançou um erro de classificação de 3,57%.
- **MobileNet:** proposto em 2017 por Howard *et al.* (2017), sendo uma arquitetura com camadas convolucionais profundas com baixa complexidade computacional, que permitiu que fosse executada em dispositivos móveis com recursos limitados. Foi treinado no conjunto de dados ImageNet e alcançou um erro de classificação de 3,3%.

Esses modelos são geralmente comparados em termos de sua precisão de classificação no conjunto de dados ImageNet (YANG *et al.*, 2021), que é uma tarefa desafiadora de reconhecimento de objetos que consiste em 1,2 milhão de imagens em 1000 classes diferentes. O desempenho é medido em termos da taxa de erro de classificação top-1 (o modelo atribui a classe correta ao objeto em sua primeira escolha) e top-5 (o modelo atribui a classe correta ao objeto em sua primeira ou segunda escolha), e assim por diante.

Investigações sobre a topologia mais eficaz para a aplicação de classificação de sinais ECG foram realizadas em nosso estudo de forma a buscar a topologia com as melhores métricas de avaliação. Nesse contexto, as arquiteturas mais exploradas foram a ResNetV2, ResNet18 e a AlexNet (DIKER *et al.*, 2019; ULLAH *et al.*, 2020; AHMAD *et al.*, 2021; MATHUNJWA *et al.*, 2021; KIRANYAZ *et al.*, 2021; XIE *et al.*, 2021; CHEN *et al.*, 2023; MEWADA, 2023; NAROTAMO *et al.*, 2024).

A arquitetura ResNetV2 é uma variante da ResNet (Rede Neural Residual), apresentada por He *et al.* (2016) com uma ênfase maior na regularização. Algumas de suas características construtivas são:

- Arquitetura muito profunda;
- Uso de camadas residuais para permitir o fluxo de informações em várias profundidades da rede. Cada bloco residual consiste em várias camadas convolucionais, seguidas por uma conexão de atalho que permite o fluxo direto de informações;

- O seu bloco residual utiliza uma técnica chamada “ativação de atalho”, que adiciona uma camada de *Batch Normalization* e uma função de ativação ReLU diretamente na conexão de atalho. Essa camada é usada para acelerar a convergência do treinamento e reduzir a sensibilidade à inicialização de pesos. Atua como uma forma de regularização, reduzindo o risco de *overfitting*. É, portanto, responsável por normalizar a distribuição das ativações durante o treinamento, garantindo que a média dos valores das ativações seja zero e a variância unitária, garantindo estabilidade numérica;
- Usa uma técnica chamada “*drop-path*”, que é semelhante ao *dropout*, mas é aplicada diretamente às conexões de atalho, em vez das camadas convolucionais, com o objetivo de prevenir *overfitting*;
- Usa uma camada de *Batch Normalization* e uma função de ativação ReLU após cada camada convolucional, o que ajuda a estabilizar o treinamento.

A ResNet18 também é apresentada em He *et al.* (2016), como uma variante menor da ResNet, projetada para ser mais rápida e mais fácil de treinar do que versões mais profundas. Algumas de suas características construtivas são:

- Total de 18 camadas, incluindo 16 camadas convolucionais e 2 camadas totalmente conectadas;
- Usa a técnica de bloco residual para permitir o fluxo de informações em várias profundidades da rede;
- Cada bloco residual consiste em duas camadas convolucionais de 3x3, seguidas por uma conexão de atalho;
- Usa uma camada de *Batch Normalization* e uma função de ativação ReLU após cada camada convolucional, e
- Usa uma camada de *Pooling Global Média* no final da rede, seguida por uma camada de saída totalmente conectada para classificar as imagens.

A AlexNet é apresentada em Krizhevsky, Sutskever e Hinton (2012) num modelo para uso em duas GPU devido às limitações de *hardware* existentes naquele momento. Com a evolução do *hardware* GPU, o modelo AlexNet foi atualizado em Krizhevsky (2014). Algumas de suas características construtivas são:

- Possui 8 camadas no total, incluindo 5 camadas convolucionais e 3 camadas totalmente conectadas;
- Usa a função de ativação ReLU após cada camada convolucional;
- Usa a técnica de sobreposição para reduzir o tamanho das camadas intermediárias;
- Usa uma camada de *Pooling Max* depois das camadas convolucionais, e
- Usa a técnica de *Dropout* para prevenir *overfitting*.

### 2.4.2 Modelos de CNN multimodais

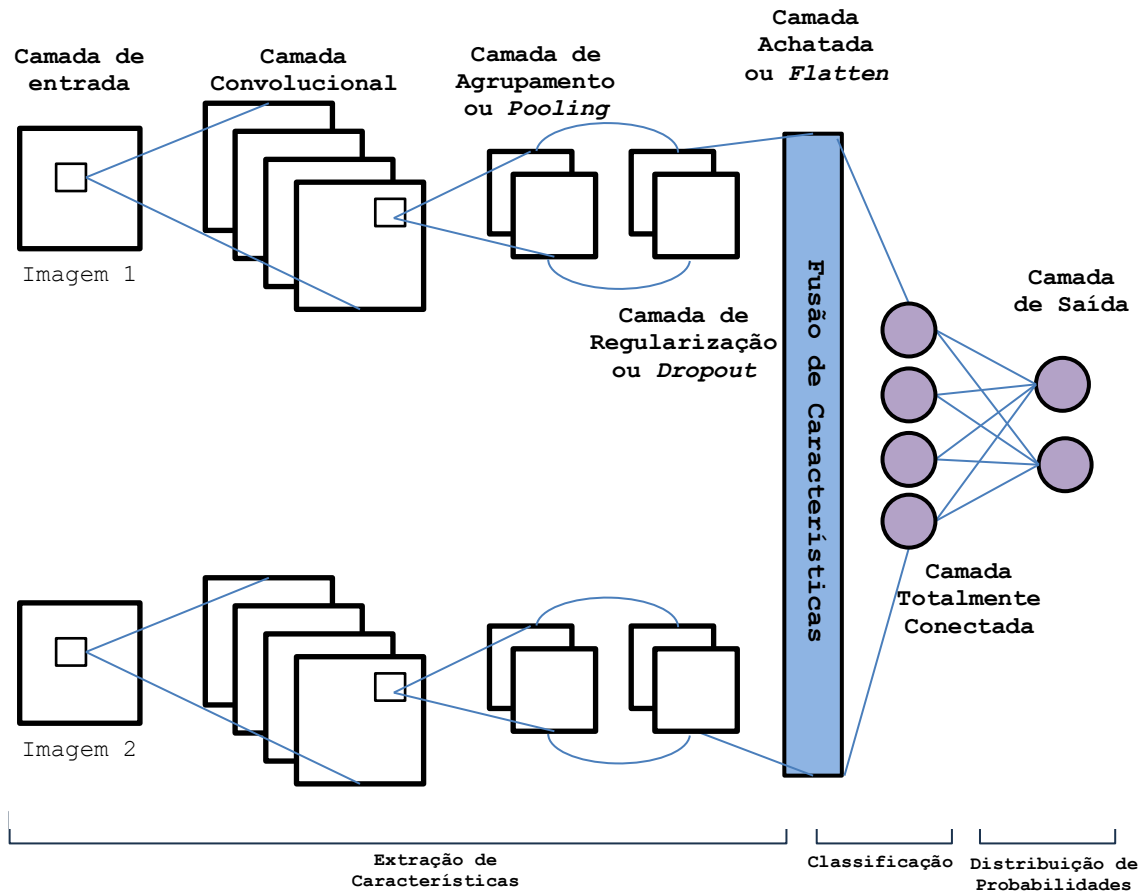
Considerando as capacidades da CNN, alguns trabalhos vêm sendo apresentados na literatura usando CNN multimodais em algumas áreas, como por exemplo para o diagnóstico médico, com resultados considerados promissores (ASFAND-E-YAR *et al.*, 2024; JIANG *et al.* 2024; TANIOKA *et al.*, 2023; WAJID *et al.*, 2023; WANG *et al.*, 2023a). Existem duas formas não exclusivas de caracterizar uma CNN como multimodal:

1. Quando utiliza dados diferentes como entrada para estrutura multimodal, tais como dados de diferentes derivações de sinal ECG ou dados combinados de sinais ECG com medicação ou dados de outros sensores, o que conduz a mais de uma camada convolucional trabalhando em paralelo para extração de características das entradas da estrutura multimodal.
2. Quando o mesmo dado é representado de forma diferente, mas complementares, como usando diferentes técnicas de representação de sinais ECG como imagens, o que conduz a mais de uma camada convolucional para extração de características de cada imagem.

A Figura 10 ilustra uma CNN multimodal em que duas imagens diferentes são usadas como entrada, apresentando camadas de extração de características independentes com posterior fusão de características e etapas de classificação e distribuição de probabilidades para saída unificadas. Portanto, cada imagem pode representar por exemplo dados diferentes como diferentes derivações de um sinal ECG, ou outros que possam estar disponíveis, como medição de temperatura, pressão arterial, etc. Mas também podem ser representações diferentes de um mesmo dado, como um mesmo sinal

ECG representado por diferentes técnicas de imagens.

**Figura 10:** CNN com abordagem multimodal.



**Fonte:** Próprio Autor.

Variações da topologia apresentada na figura 10 podem ainda ser aplicadas numa abordagem multimodal, pois o processo de fusão de características também pode ocorrer no meio das etapas de extração de características, entre camadas convolucionais, ou no meio da etapa de classificação entre camadas de neurônios. A literatura apresenta diversas abordagens multimodais diferentes utilizando sinais ECG:

- Trabalhos que adotam mais de uma derivação de sinais ECG: Zhang *et al.* (2014) usando classificador SVM; Zhang *et al.* (2023) e Garcia *et al.* (2017) fundindo diferentes derivações numa imagem única; e Hammad, Liu e Wang (2018) que projetaram um sistema de autenticação biométrica multimodal usando uma fusão de dados de ECG e impressão digital.
- Trabalhos que utilizam diferentes representações de um mesmo sinal

ECG: Wang *et al.* (2019b) que não usa CNN, mas apresenta diferentes representações do mesmo sinal como entrada de uma MLP; Wang *et al.* (2021) em que usam CNN com diferentes camadas convolucionais para diferentes representações de um mesmo sinal ECG, fundindo as características para uma camada fortemente conectada; Ahmad *et al.* (2021), em que os autores comparam duas abordagens multimodais de CNN, gerando uma única imagem com diferentes técnicas como entrada de uma única camada convolucional, e separando 3 diferentes entradas com representações diferentes do mesmo sinal ECG conectadas numa camada de classificação usando SVM; e Zhou e Fang (2024) que utilizam diferentes representações de imagens de um mesmo sinal ECG numa CNN.

- Trabalhos que fundem características na camada totalmente conectada: Hammad, Liu e Wang (2018) e Ahmad *et al.* (2021).
- Outras abordagens: Naz *et al.* (2021) apresenta uma abordagem inspirada na multimodal, usando uma mesma imagem com camadas convolucionais diferentes, fundindo as características no meio da camada totalmente conectada.

Como corroborado por Hammad, Liu e Wang (2018) e Hong *et al.* (2020), existem muitas oportunidades para estudos em projetos e modelos capazes de utilizar dados multimodais. Em especial para a abordagem com diferentes tipos de dados, com base no desenvolvimento de dispositivos médicos e de saúde, muitos sinais vitais, como temperatura, frequência respiratória e pressão arterial, podem ser coletados simultaneamente com dados de ECG e utilizados em sistemas automatizados. No entanto, deve-se garantir que todos estejam sincronizados na linha do tempo em suas frequências de amostragem, sendo um potencial desafio para esta abordagem, diferentemente da abordagem com diferentes representações de um mesmo sinal, cuja sincronização é natural.

### **2.4.3 Modelos de CNN com mecanismo de atenção**

Recentemente trabalhos vêm sendo apresentados procurando aliar CNN com mecanismos de atenção (TOĞAÇAR, ERGEN, CÖMERT, 2020; LIU, YANG, 2021;

JUN, LIYUAN, 2022, TANG *et al.*, 2023; ZHOU, FANG, 2024). Os mecanismos de atenção, inicialmente aplicados com sucesso na área de Processamento de Linguagem Natural, se baseiam na ideia de que humanos tendem a concentrar sua atenção em certas partes do espaço visual para adquirir informações, em vez de ver a imagem como um todo sendo que, quando inseridos nas CNNs, podem permitir que a rede tenha atenção seletiva para características específicas extraídas das imagens podendo desta forma aumentar o desempenho no processo de classificação. A seguir são apresentados alguns dos principais mecanismos de atenção observados na literatura:

1. *Self-Attention* ou Auto Atenção: captura relações de longo alcance entre diferentes posições de um mapa de características, considerando a interação entre todas as localizações. Nessa abordagem busca-se a dependências e relacionamentos dentro da própria sequência de entrada, de forma que o modelo identifique e pondere a importância de diferentes partes ao atender a si mesmo. Exemplos dessa abordagem são: Ahmad *et al.* (2021), Wang *et al.* 2023b, Tang *et al.* (2023) e Di Paolo e Castro (2024).
2. *Spatial Attention* ou Atenção Espacial: foca em identificar regiões importantes no espaço, destacando onde a atenção deve ser aplicada. Esse mecanismo utiliza informações espaciais (como padrões em *pixels* adjacentes) para gerar um mapa de atenção. Exemplos dessa abordagem são: Fan *et al.* (2023) e Huang *et al.* (2023).
3. *Channel Attention* ou Atenção de Canal: foca em ajustar a importância de cada canal no mapa de características, permitindo que a rede destaque canais mais relevantes. Essa abordagem utiliza uma operação global (como *global average pooling*) para extrair informações de todo um canal e aplicar pesos aprendidos. Exemplos dessa abordagem são: Srivastava *et al.* (2022) e Fan *et al.* (2023).
4. *Mixed Attention (Spatial and Channel)* ou Atenção Mista (Espacial e em Canal): combina *Spatial Attention* e *Channel Attention* para explorar relações espaciais e de canal simultaneamente. Aplica atenção por canal e atenção espacial em sequência ou paralelamente para refinar os mapas de características. Exemplos dessa abordagem são: Wan *et al.* (2023) e Chen *et al.* (2024).

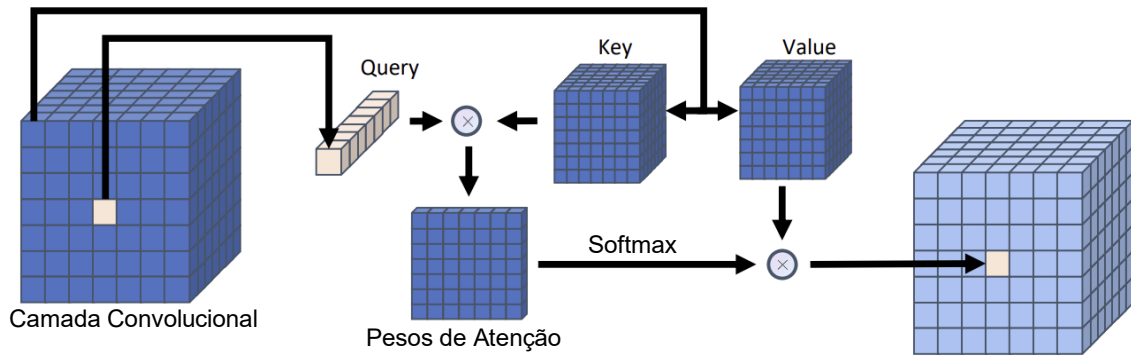
5. *Position Attention* ou Atenção de Posição: similar ao *Spatial Attention*, mas é mais avançado e utiliza dependências de posição para calcular a atenção. Exemplos dessa abordagem são: Zhang *et al.* (2021) e Qiang *et al.* (2024).

Para o caso de classificação de arritmias alguns trabalhos vêm sendo desenvolvidos na área como em Ahmad *et al.* (2021), em que os autores utilizam diferentes técnicas de imagens numa CNN em que cada imagem é processada por um *kernel* de auto atenção; Islam *et al.* (2023), onde os autores apresentam uma rede neural recorrente bidirecional baseada em atenção hierárquica híbrida com método CNN dilatado para classificação de arritmia, atingindo bons resultados na classificação, e em Zhou e Fang, (2024), onde é proposto um mecanismo de atenção em canal de frequência no processamento de imagens para modificar dinamicamente o peso do mapa de recursos dentro da CNN.

Em nossa estrutura proposta, o mecanismo de atenção utilizado é baseado no mecanismo *Self Attention*, sendo este um mecanismo usado para capturar dependências internas dentro de uma mesma entrada (como um mapa de características ou sequência de amostras), ajustando a importância relativa de suas diferentes partes. O mecanismo identifica como cada elemento (ex.: *pixel* se a entrada for 2D ou vetor de recurso se 1D) se relaciona com os outros, permitindo que a rede modele, durante o treinamento, interações globais ou regionais.

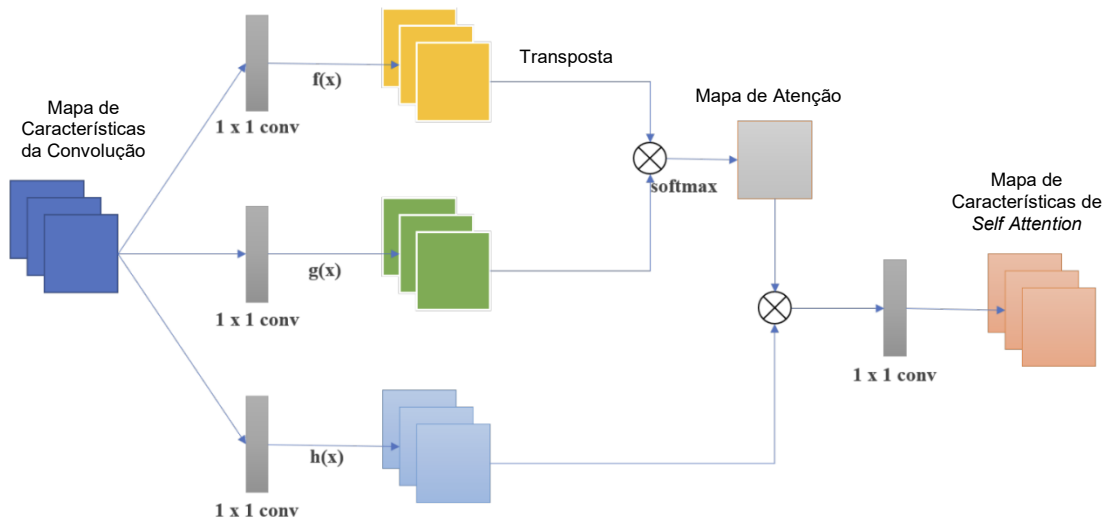
Analisando os trabalhos de Vaswani *et al.* (2017), Chen et a. (2021), Srivastava *et al.* (2022), Ma *et al.* (2023), Tang *et al.* (2023), Silva (2024), Fan *et al.* (2023) e Zhou e Fang (2024), identificam-se dois principais paradigmas de *Self Attention*: o *Self-Attention* Clássico ou Global, e o *Self-Attention* Local.

O *Self-Attention* Clássico ou Global baseia-se nos conceitos de *Queries*, *Keys* e *Values*, como ilustrado na Figura 11, no qual estes elementos se originam de diferentes projeções da mesma entrada. O produto escalar entre *Query* e *Key* é computado gerando os pesos de atenção, que são normalizados por uma função *Softmax* e, finalmente, *Value* é ponderado de acordo com os pesos de atenção que representam a similaridade entre cada elemento e todos os outros. Este mecanismo é útil para capturar interações longas e globais, como em imagens complexas ou sequências textuais, cujas relações entre elementos distantes são importantes.

**Figura 11:** Estrutura de *Self-Attention* clássica com *Query*, *Key* e *Value*.

**Fonte:** Adaptado de Ma *et al.* (2023).

Outra estrutura de implementação do *Self Attention* Clássico ou global é ilustrada na Figura 12, em que  $f(x)$ ,  $g(x)$  e  $h(x)$  são todas convoluções  $1 \times 1$  comuns, gerando três linhas de saída. A saída  $f(x)$  é transposta e multiplicada pela saída de  $g(x)$ , e o mapa de atenção é obtido após a função de ativação *Softmax*. Após multiplicar o mapa de atenção obtido pela saída de  $h(x)$  *pixel por pixel*, o mapa de características de atenção final é obtido.

**Figura 12:** Estrutura de *Self-Attention* clássica com funções  $f$ ,  $g$  e  $h$ .

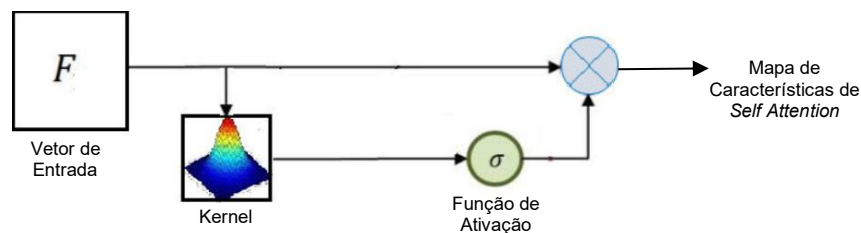
**Fonte:** Adaptado de Chen *et al.* (2021).

Estas duas estruturas apresentadas são interessantes quanto localizadas na camada convolutiva, com estruturas de entrada multidimensionais.

O paradigma de *Self Attention Local* é interessante para dados 1D ou após a camada de achatamento de uma CNN, em que os dados se tornam vetorizados, e foi esta

a estratégia adotada em nosso estudo. Nesta estratégia, o conceito de atenção local é feito usando convoluções para gerar mapas de atenção também achatados, como ilustra a estrutura apresentada na Figura 13, em que uma entrada  $F$ , já achatada, se divide em dois caminhos, sendo um sem alteração, e outro passando por uma convolução com um *kernel* que pode ser fixo ou adaptativo (atualizado durante o treinamento), e a saída desta convolução passando por uma função de ativação, geralmente *Sigmoid* ou *Softmax*, gerando o mapa de atenção achatado. Os resultados são multiplicados para gerar o mapa de características de *Self Attention*. Os pesos do mapa de atenção achatado são aprendidos para enfatizar características mais relevantes, sendo mais eficiente em termos de memória e computação e geralmente integrado em CNNs na camada fortemente conectada, como observado em Ahmad e Khan (2020), Ahmad *et al.* (2021), Tang *et al.* (2023) e Di Paolo e Castro (2024).

**Figura 13:** Estrutura de *Self-Attention* local com convolução.



**Fonte:** Adaptado de Ahmad e Khan. (2020).

## 2.5 CONVERSÃO DE SINAIS TEMPORAIS EM IMAGENS

A conversão de sinais temporais em imagens é uma técnica que consiste em transformar um sinal de sinal temporal em uma imagem, para que possa ser processada por uma rede neural convolucional (CNN). A ideia é tratar cada contorno do sinal temporal como um conjunto de *pixel* de uma imagem, e assim utilizar a capacidade de processamento de imagens das CNNs para realizar a classificação ou previsão do sinal temporal.

Estudos que aplicam conversão de sinais temporais em imagens já são feitos há anos. Wang, Yan e Oates (2015), por exemplo, transformaram cada sinal temporal em uma imagem com um mapa de calor, que representava a intensidade de cada valor da série ao longo do tempo. Outro estudo como Zhang *et al.* (2016) utilizaram várias representações de imagens para representar sinais temporais, como imagens de *greyscale*,

imagens de níveis de cinza com diferentes combinações de janelas temporais e imagens de espectrograma.

A conversão de sinais temporais em imagens tem sido utilizada em diversas aplicações, como na previsão de preços de ações (KHODAEI, ESFAHANIPOUR, TAHERI, 2022), na classificação de diversos tipos de doenças (PHAN *et al.* 2019; WANG *et al.*, 2022), e na detecção de falhas em equipamentos industriais (QIU *et al.*, 2023), entre outras. A técnica apresenta vantagens em relação a métodos tradicionais de processamento de sinais temporais, como a capacidade de detectar padrões complexos e a possibilidade de aproveitar a eficácia das CNNs na extração de características e na classificação.

No estudo aqui desenvolvido, cinco técnicas de conversão de sinais temporais em imagens foram estudadas e implementadas para uso na fase de experimentos: Campo Angular Gramiano (GAF), Gráfico de Recorrência (RP), Campo de Transição de Markov (MTF), Curva de Preenchimento do Espaço de Hilbert (HSFC), e Mapa de Calor (HP). A seguir serão apresentadas somente as técnicas RP e HSFC, que são as 2 técnicas que são usadas na estrutura final consolidada após fase de experimentos para definição da estrutura para classificação de arritmias.

### 2.5.1 Gráfico de Recorrência

Um Gráfico de Recorrência (*Recurrence Plot*, RP) é uma técnica que representa a recorrência de padrões em um sinal temporal, destacando relações entre pontos da série e revelando mudanças na dinâmica do sinal, sendo o RP representado por uma imagem (matriz) obtida a partir da série temporal, representando as distâncias pareadas para cada ponto na série (ECKMANN *et al.*, 1987; CASDAGLI, 1997).

Considerando um sinal temporal representado por  $N$  pontos  $x = [x_1 \ x_2 \ x_3 \ \dots \ x_N]$ , matematicamente, o RP é a representação de uma matriz quadrada  $N \times N$  onde cada elemento da matriz é dado por:

$$R_{ij}(\varepsilon) = \Theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|) \quad (6)$$

com  $i, j = 1, \dots, N$ , onde  $\varepsilon$  é o limiar ou limite de tolerância dentro do qual dois pontos da série são considerados como recorrentes e a função  $\Theta$  é a função de *Heaviside*, que retorna 1 (pontos recorrentes) se  $\varepsilon - \|x_i - x_j\| > 0$  e retorna 0 se  $\varepsilon - \|x_i - x_j\| < 0$ ,

onde  $\|\cdot\|$  é a norma, sendo as mais usadas a norma mínima, a máxima e a Euclidiana. Para este caso o RP gera uma imagem preto e branca.

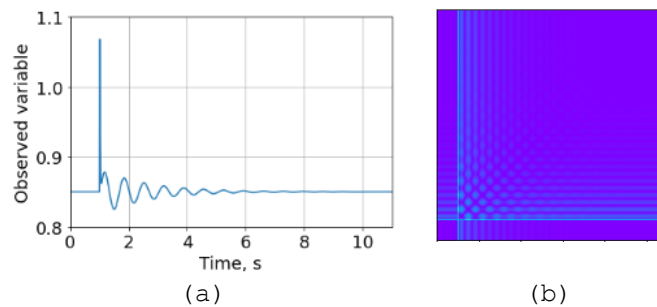
Neste trabalho foi adotada uma versão modificada do RP (FARIA *et al.*, 2016) que gera uma imagem colorida, sendo que cada elemento da matriz é obtido através de:

$$R_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| \quad (7)$$

para  $i, j = 1, \dots, N$ , onde  $\|\cdot\|$  é a norma Euclidiana.

A Figura 14 apresenta um exemplo de transformação de um sinal temporal em imagem usando RP.

**Figura 14:** Conversão de sinal temporal em imagem. a) Gráfico do sinal temporal; b) Imagem RP.



**Fonte:** Próprio Autor.

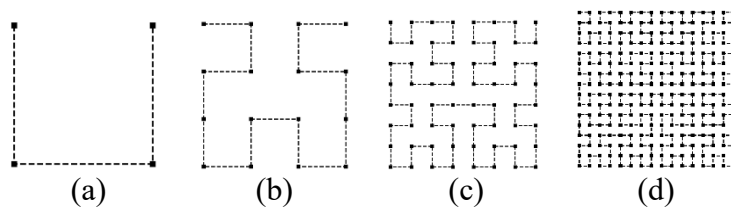
### 2.5.2 Curva de Preenchimento do Espaço de Hilbert

O HSFC é uma técnica que permite mapear uma série temporal (1D) para uma imagem (matriz 2D), sendo que no processo de mapeamento é preservada a proximidade espacial entre as amostras da série temporal. Essa característica facilita a visualização espacial da série visto que, a posição de cada elemento na matriz reflete sua ordem temporal original, onde a intensidade dos *pixels* está de acordo com valores das amostras da série temporal, o que permite que mudanças abruptas do sinal temporal de ECG possam ser identificadas nas imagens geradas. No HSFC uma série temporal é mapeada em um espaço 2D na forma de uma curva de Hilbert, ligando o índice  $i$  de cada amostra da série temporal com uma coordenada correspondente  $(x, y)$  na curva de Hilbert. Neste caso, o valor do ponto na coordenada  $(x, y)$  terá o valor correspondente do sinal temporal, implicando assim na escala de cores da imagem formada.

No HSFC descrito em Hilbert (1891), um sinal temporal é mapeado em um espaço

2D na forma de uma curva de Hilbert, ligando o índice  $i$  de cada amostra do sinal temporal com uma coordenada correspondente  $(x, y)$  na curva de Hilbert. Neste caso, o valor do ponto na coordenada  $(x, y)$  terá o valor correspondente do sinal temporal, formando uma imagem em que a escala de cores representa a intensidade do sinal em cada posição. A Figura 15 ilustra os níveis 1, 2, 3 e 4 da curva de Hilbert. No nível 1, a curva conecta quatro pontos, formando a estrutura mais simples em um padrão em “U”. No nível 2, a curva se refina, conectando 16 pontos e dobrando-se em um padrão mais complexo, que preserva a continuidade espacial e aumenta a densidade de amostragem. No nível 3, a curva atinge uma granularidade ainda maior, conectando 64 pontos em um padrão mais intrincado que destaca mais detalhes na relação entre os valores do sinal temporal e sua representação espacial. Finalmente, no nível 4, a curva se estende para conectar 256 pontos, formando uma estrutura mais detalhada, que permite visualizar variações locais e padrões do sinal temporal com alta resolução. Essa hierarquia dos níveis reflete o aumento progressivo da capacidade da curva de Hilbert de mapear o sinal em um espaço bidimensional, preservando propriedades de proximidade espacial e continuidade.

**Figura 15:** Curvas de Hilbert de níveis 1, 2, 3 e 4.



**Fonte:** Próprio Autor (2024).

Dados de trajetória espaço-temporal são comumente multidimensionais. HSFC também é usado para transformar a trajetória em sinais temporais, preservando a informação espaço-temporal, como tratado em Wang e Oates (2015).

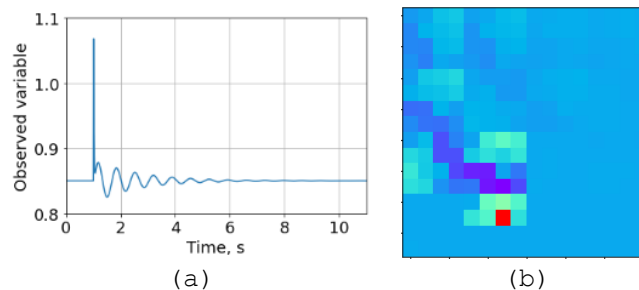
Outra aplicação é de um sinal temporal, como apresentado em Borrell *et al.* (2018) e Bragin e Spitsyn (2020), que pode ser representada em um plano bidimensional na forma de uma curva de Hilbert ligando o índice  $i$  de cada ponto do sinal temporal com a coordenada correspondente  $(x, y)$  na curva. Neste caso, o valor do ponto na coordenada  $(x, y)$  terá o valor correspondente do sinal temporal.

A construção da curva de Hilbert tem, como ponto de partida, um quadrado. Na primeira iteração, o quadrado é dividido em quatro quadrados congruentes e os seus pontos centrais são unidos. Na segunda iteração: cada um dos quatro quadrados é dividido

em quatro novos quadrados congruentes e seus centros são unidos, e assim por diante num processo iterativo. Assim, um sinal temporal pode ser construído num espaço bidimensional na forma de uma curva de Hilbert ligando o índice  $i$  de cada ponto do sinal temporal com a coordenada correspondente  $(x, y)$  na curva de Hilbert. Neste caso, o valor do ponto na coordenada  $(x, y)$  terá o valor correspondente do sinal temporal, implicando assim na escala de cores da imagem formada.

A Figura 16 apresenta um exemplo de transformação de um sinal temporal em imagem usando HSFC, com mais detalhes complementares e de implementação podendo ser consultados em Feng *et al.* (2012), Skilling (2004) e Earl (2013).

**Figura 16:** Conversão de sinal temporal em imagem. a) Gráfico do sinal temporal; b) Imagem HSFC.



**Fonte:** Próprio Autor.

## 2.6 CONSIDERAÇÕES FINAIS

Neste Capítulo de fundamentação teórica foi apresentado o funcionamento do coração, como um sinal cardíaco é captado através do ECG e suas diferentes formas de implementação com as chamadas derivações. As bases de dados públicas MIT-BIH e PTB, recomendadas pela AAMI foram devidamente apresentadas com suas características detalhadas. Como o objetivo desta tese está relacionado à classificação automatizada dos sinais ECG, foi realizada também uma apresentação dos mecanismos de redes neurais convolucionais e suas abordagens multimodais e com módulo de atenção. Estas redes utilizam imagens como entradas, e por isso o mecanismo de conversão de sinais temporais em imagens também foi apresentado.

Com o Capítulo de fundamentação teórica finalizado, o próximo Capítulo apresenta o estado da arte e trabalhos correlatos a esta tese. São apresentados estudos recentes que investigaram este tema de pesquisa e também utilizaram as mesmas bases de dados públicas, o que também será relevante posteriormente na análise dos resultados experimentais obtidos.

### 3 ESTADO DA ARTE DE CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS

Este Capítulo apresenta o estado da arte sobre classificação de arritmias, onde serão apresentados como estão sendo conduzidos de maneira geral os estudos recentes neste domínio de aplicação, incluindo seus desafios atuais e paradigmas adotados. Para finalizar, considerando o objetivo desta tese, serão apresentados trabalhos correlatos já apresentados na literatura considerando sistemas de classificação de arritmias para abordagem intrapaciente e interpaciente especificamente usando as bases MIT-BIH e PTB que são as usadas neste trabalho.

#### 3.1 ESTADO DA ARTE

Luz *et al.* (2016), apresentam um estudo dos métodos de última geração contemporâneos de classificação automatizada de batimentos cardíacos de anormalidades baseadas em ECG, apresentando o pré-processamento do sinal de ECG, as técnicas de segmentação de batimentos cardíacos, os métodos de descrição de recursos e os algoritmos de aprendizado usados. Foi observado que a segmentação de batimentos cardíacos (ou seja, detecção do pico R ou o complexo QRS) foram estudados por mais de três décadas e as gerações desses algoritmos em diferentes tecnologias e métodos recém desenvolvidos refletem a evolução do poder de processamento dos computadores, fazendo com que os trabalhos mais recentes se concentrem na precisão de segmentação de batimentos cardíacos. Nesse sentido, duas medidas são mais consideradas para avaliar a precisão da segmentação do batimento cardíaco: sensibilidade e precisão.

O MIT-BIH tem se apresentado como a base mais popular na apresentação de resultados (LUZ *et al.*, 2016; HONG *et al.*, 2020), sendo uma base extremamente desequilibrada em relação à quantidade de registros de cada classe (CHAZAL; O'DWYER; REILLY, 2004; MAR *et al.*, 2011; LLAMEDO; MARTÍNEZ, 2011). Autores que seguiram a abordagem interpaciente (CHAZAL *et al.*, 2004; SORIA, MARTINEZ, 2009; LLAMEDO, MARTÍNEZ, 2011; LIN, YANG, 2014; ZHANG *et al.*, 2014; GARCIA *et al.*, 2017; WANG *et al.*, 2021; DIAS *et al.*, 2021) relatam grande dificuldade em obter resultados promissores para o batimento cardíaco das classes de arritmia SVEB e VEB, e muitos ignoram a classe F em seus estudos. Dificuldades em classificar as classes SVEB e F são mais relatadas nos trabalhos que utilizam o paradigma

intrapaciente (KACHUEE, FAZELI, SARRAFZADEH, 2018; IZCI *et al.*, 2019; HUANG *et al.*, 2019; ZHAI, TIN, 2018; SHAKER *et al.*, 2020; XU, JEONG, 2020; QUIAO *et al.*, 2020; HE *et al.*, 2021; AHMED *et al.*, 2021). A classe SVEB, em ambos os casos, tem se apresentado mais desafiadora para classificar.

Vários autores também apresentam abordagens semiautomáticas para melhorar os resultados em mais de 40% em relação aos seus testes usando apenas a abordagem automática, mesmo quando um pequeno número de batimentos cardíacos é selecionado (CHAZAL, REILLY, 2006; CHAZAL, 2014; KIRANYAZ, INCE, GABBOUJ, 2016). Na abordagem semiautomática, existe mais uma etapa após a classificação automatizada das arritmias, na qual os resultados são apresentados aos especialistas em ECG, que revisam e validam as classificações propostas, intervindo no modelo com ajustes ou correções nas classificações, se necessário, que deve conter mecanismos computacionais para essa finalidade (CHAZAL, REILLY, 2006; CHAZAL, 2014; KIRANYAZ, INCE, GABBOUJ, 2016; HONG *et al.* 2020).

A abordagem semiautomática combina a eficiência e a consistência da classificação automatizada com a expertise e o conhecimento dos especialistas humanos. Ela permite uma análise mais precisa e confiável dos sinais ECG, além de facilitar a identificação de casos complexos ou ambíguos que requerem a avaliação e intervenção de especialistas. O fato de ser comum a presença de especialistas no ambiente clínico faz com que os autores considerem uma direção de pesquisa promissora, em particular quando há infraestrutura e equipe multidisciplinar colaborando.

Luz *et al.* (2016) apresenta como grande obstáculo para o avanço das pesquisas com foco na classificação automatizada de sinais ECG o reduzido número de bancos de dados disponíveis, sendo este ainda um importante desafio, em particular devido ao investimento financeiro e a necessidade de incorporação aos padrões ANSI/AAMI (2020). Além do uso de novas tendências para captação do sinal de ECG, como abordagens *off-the-person*, para a elaboração de novos bancos de dados. Para estes autores, quem não segue a abordagem interpaciente de divisão de dados de treino e teste, observada apenas na base MIT-BIH devido à grande influência do trabalho de Chazal, O'Dwyer e Reilly (2004), é difícil fazer uma comparação justa entre os métodos. Também é difícil avaliar a real contribuição dos métodos, uma vez que o esquema intrapaciente favorece melhores performances dos modelos. Assim, portanto, também é um desafio estudar e reimplementar métodos intrapacientes publicados na literatura, seguindo um

esquema interpaciente, como observado em Luz e Menotti (2011), avaliando o impacto e comportamento da solução proposta sob uma visão não tendenciosa.

Hong *et al.* (2020) realizaram uma pesquisa sistemática que resume a pesquisa de aprendizado profundo existente usando dados de ECG, analisando artigos que aplicaram modelos de *deep learning* (*deep neural network*) a dados de ECG que foram publicados entre 1º de janeiro de 2010 e 29 de fevereiro de 2020 do *Google Scholar*, *PubMed*, *Digital Bibliography e Library Project*, considerando os aspectos de tarefas, modelos e dados.

Os trabalhos avaliados pelos autores que usam arquiteturas híbridas de uma CNN, rede neural recorrente ou outras técnicas produzem os melhores resultados. Novos desafios e problemas relacionados à interpretabilidade, escalabilidade e eficiência também são abordados.

Redes Neurais Adversariais (GANs) têm sido empregadas para acumular conhecimento sobre a distribuição do ruído de ECG continuamente através de um jogo minimax entre um gerador e um discriminador, onde a qualidade da redução de ruídos de sinais foi avaliada *versus* um algoritmo SVM por Wang *et al.* (2019a). Recentemente, as GANs foram aplicadas para enfrentar o desafio do desequilíbrio de dados nos dados de ECG. Por exemplo, Wang *et al.* (2019a) também propuseram um modelo de detecção de anormalidades para sinais de ECG baseado em uma estrutura que combina CNN e redes neurais recorrentes, usando uma GAN composta de várias CNNs 1D para realizar o aumento de dados. Seu modelo alcançou excelente desempenho para conjuntos de dados de classe desequilibrada, mas seus estudos foram aplicados para avaliar os registros de cada paciente.

Um modelo de rede adversarial generativa personalizado foi proposto por Golany e Radinsky (2019) para realizar a classificação personalizada do ECG, onde os dados rotulados específicos do sujeito são escassos. Especificamente, um GAN foi otimizado usando uma função de perda especializada e treinado para gerar sinais de ECG sintéticos personalizados para diferentes arritmias.

Karras *et al.* (2018) propôs o modelo *Progressive Growing of GANs* (Crescimento Progressivo de GANs, PGAN), sendo uma evolução de GAN para gerar imagens de alta resolução, com crescimento progressivo, que trabalham em etapas. Em cada etapa uma rede neural é treinada para gerar uma imagem que, ao concluir o treinamento, uma nova camada na rede é acrescentada para gerar uma nova imagem com o dobro da resolução.

Assim, as camadas são projetadas para evoluírem as imagens geradas com seu dobro de resolução, iniciando com  $4 \times 4$  pixels, depois  $8 \times 8$ ,  $16 \times 16$ ,  $32 \times 32$ ,  $64 \times 64$ , depois  $128 \times 128$ ,  $256 \times 256$ ,  $1024 \times 1024$  e assim por diante, a um alto custo de recursos computacionais, sendo este ainda um desafio em aberto quanto às aplicações no aumento de dados para sinais ECG.

Na literatura observa-se que as GANs para geração de imagens artificiais tornaram-se um campo de estudo, já tendo aplicações bem avançadas como a StyleGAN desenvolvida por FU *et al.* (2022). No entanto, a PGAN de 2018 foi a última com bons resultados que são suportadas por sistemas computacionais com uma Unidade de Processamento Gráfico (GPU). StyleGAN, embora mais avançada, são recomendadas 8 GPUs para processamento, sendo uma tecnologia que exige muito tempo de processamento e sistemas computacionais mais robustos, e por isso ainda carentes de avaliações no estudo de aumento de dados para sinais ECG desbalanceados.

No que se refere ao aumento de dados para as classes minoritárias, uma técnica mais simples de implementar e já usada com sucesso por Ahmad *et al.* (2021) é a *Synthetic Minority Over-sampling Technique* (ou Técnica de sobreamostragem de minoria sintética, SMOTE), desenvolvida por Chawla *et al.* (2002) que gera novos dados 1D a partir do universo de dados 1D disponíveis.

Já avançando para as CNNs, combinações com redes LSTM têm sido usadas para aprender dependências de longo prazo (por exemplo, classificar ritmo pulsado ou atividade elétrica sem pulso (ELOLA *et al.*, 2018), detectar eventos hipoglicêmicos em indivíduos saudáveis (PORUMB *et al.*, 2020), prever a necessidade de revascularização urgente (GOTO *et al.*, 2019).

Dois tipos de CNN são comumente usados para classificação de ECG, a saber, a CNN 1D e a CNN 2D. Uma CNN 1D opera aplicando *kernels* ao longo da dimensão temporal dos dados brutos de ECG, enquanto uma CNN 2D normalmente opera em dados de ECG transformados, como matrizes de distribuição de distância baseadas em cálculos de entropia, matrizes de co-ocorrência em nível de cinza, ou características combinadas, como morfologias, RR em intervalos e correlações batimento a batimento. No entanto, há alguma controvérsia em relação ao uso de ECG de múltiplos picos ou sinais de ECG no domínio da frequência, como usando a transformada rápida de Fourier (FFT) ou transformada de Fourier de curto prazo. Alguns trabalhos, como Huang *et al.* (2019),

aplicaram diretamente 2D-CNNs a tais espectrogramas no domínio da frequência, mas isso introduz alguns desafios, especialmente quando se considera os comportamentos intracomponentes e intercomponentes dos sinais ECG. Alguns dos problemas observados podem incluir:

- **Variações interpacientes:** Os sinais ECG podem variar significativamente entre diferentes pacientes devido a diferenças na anatomia cardíaca, condições de saúde, idade, sexo e outros fatores. Isso pode tornar a classificação baseada na FFT desafiadora, pois os padrões de frequência podem variar entre os pacientes e exigir um ajuste específico para cada indivíduo.
- **Variações intrapacientes:** Mesmo em um único paciente, os sinais ECG podem apresentar variações intrapacientes, como mudanças na frequência cardíaca, posição do eletrodo, ruídos e artefatos, que podem afetar os padrões de frequência e complicar a classificação baseada na FFT.
- **Comportamentos intracomponentes:** Os componentes do sinal ECG, como as ondas P, QRS, T e U (ver Figuras 1 e 5), podem apresentar variações intracomponentes, como variações na amplitude, duração, morfologia e localização, que podem não ser capturadas adequadamente pela análise da FFT isoladamente. Isso pode levar a uma redução na acurácia da classificação baseada apenas na análise da frequência.
- **Comportamentos intercomponentes:** A relação e a interação entre os diferentes componentes do sinal ECG, como a duração do intervalo PR, o segmento ST e a onda T, também podem influenciar a interpretação correta do sinal. Por exemplo, alterações na duração do intervalo PR podem afetar a morfologia do complexo QRS, e mudanças no segmento ST podem afetar a análise da onda T. Essas interações complexas podem ser difíceis de capturar somente com a análise da frequência pela FFT.
- **Ruídos e artefatos:** Os sinais ECG podem ser afetados por vários tipos de ruídos e artefatos, como ruído muscular, interferência elétrica, artefatos de movimento e outros artefatos técnicos. Esses ruídos e artefatos podem distorcer os padrões de frequência e complicar a análise baseada na FFT.

Para superar esses desafios, podem ser necessárias abordagens complementares à

análise da frequência, como a combinação de múltiplas técnicas de processamento de sinais, o uso de abordagens multicanais, a incorporação de informações temporais e espaciais, o uso de técnicas de pré-processamento adequadas para redução de ruídos e artefatos, e a integração de outros dados clínicos relevantes, como informações clínicas do paciente, dados de imagem cardíaca e dados clínicos longitudinais, entre outros para melhorar a precisão e confiabilidade da classificação automatizada de sinais ECG.

De forma resumida, pode-se citar como desafios atuais para classificação de sinais ECG com aprendizado profundo (HONG *et al.*, 2020):

- **Coleção de dados:** Não há padrão quanto aos procedimentos de coleta e estruturação dos dados, tendo diferentes estudos usando diferentes fontes, pacientes, equipamentos, etc. Isso torna difícil comparar os resultados entre diferentes conjuntos de dados de forma justa. Além disso, dados e anotações de alta qualidade são difíceis de adquirir, por isso muitos trabalhos atuais ainda estão usando o banco de dados de arritmias do MIT-BIH, que foi coletado há mais de 40 anos. O mais recente single-lead PhysioNet Computing em Cardiologia. Challenge 2017 e 12 líder 2018 China Physiological Signal Challenge usaram dados de alta qualidade, mas ambos se concentraram em gravações de ECG de curto prazo, e por isso ainda é pouco utilizado. Esforços para a construção de um novo conjunto de dados de ECG de longo prazo de alta qualidade com anotações e de acordo com ANSI/AAMI (2020) certamente inspiraria novos estudos inovadores.
- **Interpretabilidade:** Os modelos de aprendizado profundo são frequentemente considerados modelos de caixa preta porque normalmente têm muitos parâmetros de modelo ou arquiteturas de modelo complexas, o que torna difícil para um humano entender por que um resultado específico é gerado por esse modelo. Este desafio é muito maior no domínio médico porque diagnósticos sem qualquer explicação não são aceitáveis para médicos especialistas.
- **Eficiência:** Como os modelos profundos são computacionalmente complexos, é difícil implantar modelos grandes em dispositivos portáteis de saúde, o que é um grande obstáculo para a aplicação de modelos de

aprendizado profundo em aplicativos do mundo real. Nesse contexto, direções de pesquisas promissoras são: técnicas de compressão de modelos, quantização, compartilhamento ou codificação de pesos, entre outras.

- **Integração com métodos tradicionais:** A maioria dos modelos de aprendizado profundo existentes são treinados de maneira ponta a ponta, ou seja, o modelo é treinado e validado a cada iteração, gerando assim o modelo final, dificultando a integração com métodos tradicionais baseados em recursos especializados após a conclusão do treinamento do modelo. Pode-se combinar recursos especializados com recursos profundos e construir métodos tradicionais de aprendizado de máquina nesses recursos.
- **Rótulos desequilibrados:** É muito provável que os rótulos de doenças do ECG tenham distribuições muito tendenciosas porque as doenças mais graves ocorrem raramente, mas são muito importantes, e por isso é tão difícil treinar um método eficaz de aprendizado profundo com um grande número de parâmetros de modelo usando pequenos conjuntos de dados de rótulos de doenças. Para isso, o aumento de dados é uma opção promissora, como o pré-processamento de dados usando a técnica *side-and-cut* ou SMOTE ou a geração de conjuntos de dados de treinamento sintéticos usando modelos generativos como DCGAN ou PGAN.
- **Dados multimodais:** Atualmente, a maioria dos trabalhos considera apenas dados de ECG para análise. No entanto, alguns trabalhos consideraram a análise conjunta com outras fontes de dados. Por exemplo, o método proposto por Xu *et al.* (2018) prevê a mortalidade dos pacientes da unidade de terapia intensiva com base em uma combinação de intervenções, testes de laboratório, sinais vitais e dados de ECG. Com base no desenvolvimento de dispositivos médicos e de saúde, muitos sinais vitais, como temperatura, frequência respiratória e pressão arterial, podem ser coletados simultaneamente com dados de ECG. No entanto, esses dados nem sempre são sincronizados com as linhas do tempo de ECG e suas frequências de amostragem variam significativamente, o que significa que podem ser considerados dados multimodais. Daí a

oportunidade potencial para estudar como projetar modelos capazes de utilizar esses dados multimodais simultaneamente para melhorar o desempenho da tarefa em comparação com modelos treinados em qualquer modalidade individual.

- **Estudos interdisciplinares emergentes:** Finalmente, houve alguns estudos interdisciplinares inovadores, alguns dos quais estão listados a seguir:
  - Intensidade de condução segura e avaliação do tempo de reação cardíaca com base em sinais de ECG;
  - Detecção de emoções com base em sinais de ECG;
  - Análise de ECG de mamíferos, ou
  - Estimativa de idade e sexo com base em dados de ECG.

## 3.2 TRABALHOS CORRELATOS

Considerando o objetivo geral e os específicos deste trabalho, a seguir serão apresentados trabalhos correlatos já apresentados na literatura considerando abordagem intrapaciente e interpaciente com as bases MIT-BIH e PTB.

### 3.2.1 *Abordagem intrapaciente*

Nesta seção é apresentada uma breve revisão de resultados atingidos em publicações recentes tanto para a base MIT-BIH quanto para a base PTB. Alguns trabalhos, inclusive, realizam comparações de suas técnicas com essas duas bases como em Kachuee, Fazeli, Sarrafzadeh (2018), onde os autores disponibilizaram estas duas bases de dados já segmentada e pré-processada para estudos futuros. Em seu trabalho, os autores fizeram uso da CNN para classificação de arritmia usando como entrada as imagens dos pulsos 2D, focando no processamento dos sinais ECG a partir da detecção de intervalos RR como forma de melhoria da classificação.

Mais especificamente sobre a MIT-BIH, em Izci *et al.* (2019) os autores fazem uma abordagem de geração de imagens dos sinais ECG em Grayscale na sua forma de batimentos, num estudo focado na busca da melhor acurácia, o que é bastante comum para os estudos intrapaciente.

Em Huang *et al.* (2019) os autores transformam o sinal de ECG em uma imagem

de espectrograma de frequência de tempo usando a Transformada de Fourier de Tempo Curto, as quais alimentam uma CNN para classificação. Zhai, Tin (2018) utiliza também CNN com um método de conversão de sinal 1D em 2D.

Em Shaker *et al.* (2020) os autores utilizam uma técnica de aumento de dados baseada em GANs para restaurar o equilíbrio do conjunto de dados, e uma abordagem de ponta a ponta em conjunto com uma abordagem hierárquica de dois estágios, com base em convolução profunda.

Xu, Jeong e Li (2020) classificam com SVM, *random forests*, e CNN, a partir da geração de imagens com transformada *Wavelet*, comparando seus desempenhos, concluindo que o modelo em CNN foi o que obteve os melhores indicadores de desempenho.

Quiao *et al.* (2020) segmentam os dados com o algoritmo de Pan e Tompkins (1985), ainda muito utilizado pela literatura contemporânea, e posteriormente diminuem a amostragem de cada batimento cardíaco para um determinado comprimento contendo 250 pontos amostrais, finalizando com uma normalização para deixá-los entre um valor máximo e mínimo. Posteriormente apresentam uma metodologia híbrida de abordagem das técnicas de aprendizado extremo baseada em campos receptivos locais, rede bidirecional de memória de tempo longo-curto com uma arquitetura de rede neural recorrente.

He *et al.* (2021) focaram seus estudos num novo algoritmo para detecção automática de complexos QRS usando canais duplos baseados em U-Net e memória bidirecional de longo prazo, aplicando filtragem média e transformada *Wavelet* discreta para remover ruídos, classificando usando métodos tradicionais de SVM e Rede Neural Artificial (RNA).

Ahmed *et al.* (2021) apresentam um estudo de classificação em Matlab usando fusão de imagens e CNN com MGAF em *Grayscale* e aumento de dados com SMOTE, atingindo os melhores indicadores para a base já pré-processada disponível em ECG... (2018). Seitaniadis, Gialelis, Papaconstantinou (2022) também usa a base ECG... (2018) gerando imagens dos pulsos em *Grayscale*, e também usando SMOTE e CNN para aumento de dados e classificação.

Islam *et al.* (2023) fazem suas próprias seleções e organização de dados de treino e teste, desenvolvendo uma nova RNA recorrente bidirecional hierárquica híbrida com

método CNN dilatada para tratar da correlação entre contextos e dispersão de gradiente, usando GANs Condicionais para aumento de dados das classes minoritárias, também atingindo os melhores indicadores deste paradigma.

Mewada (2023) desenvolveu uma CNN 2D, em que o sinal ECG 1D é transformado para 2D usando *Wavelete*, e Zhou e Fang (2024) trabalham no domínio do tempo com a combinação de três técnicas diferentes de conversão 1D para 2D.

O Quadro 4 apresenta um resumo dos indicadores de alguns trabalhos correlatos em publicações qualificadas com a base MIT-BIH, em ordem cronológica.

**Quadro 4:** Indicadores de desempenho de trabalhos correlatos para o paradigma intrapaciente com MIT-BIH.

Resultados Anteriores	Acc. %	Prec. %	Recall %	Spe. %	F1 %
Kachuee, Fazeli, Sarrafzadeh (2018)	95,90	95,20	95,10	-	-
Izci <i>et al.</i> (2019)	92,96	90,08	80,08	98,14	82,17
Huang <i>et al.</i> (2019)	99,00	-	-	-	-
Zhai, Tin (2018)	96,05	65,91	72,06	97,83	68,06
Shaker <i>et al.</i> (2020)	98,35	82,24	93,82	99,01	87,29
Xu, Jeong, Li (2020)	95,90	96,34	95,90	-	95,92
Quiao <i>et al.</i> (2020)	99,30	98,30	97,15	-	97,71
He <i>et al.</i> (2021)	98,29	99,22	98,29	-	-
Ahmed <i>et al.</i> (2021)	99,70	98,00	98,00	-	-
Seitanidis, Gialelis, Papaconstantinou (2022)	95,20	-	95,20	98,80	-
Islam <i>et al.</i> (2023)	99,60	97,66	99,60	-	98,21
Mewada (2023)	99,52	95,12	96,18	-	95,64
Zhou, Fang (2024)	99,66	84,19	97,92	99,70	87,72

**Fonte:** Próprio Autor.

Já para a base PTB, Sharma, Tripathy e Dandapatl (2015) trabalham com sinais 1D usando *Wavelete* multiescala e autovalores de matrizes de covariância multiescala usados como recursos de processamento de sinal e (SVM) com kernel de função de base linear e radial (RBF) e vizinho mais próximo como classificadores.

Acharya *et al.* (2017) também implementam uma CNN para classificação, focando nos dados de pacientes de controle normais e de infarto no miocárdio, de forma similar a Kashuee, Fazeli e Sarrafzadeh (2018). Já Chen *et al.* (2018), Liu *et al.* (2018), Cao *et al.* (2020), e El-Ghaish e Eldele (2024) também focam nesses dados, mas implementam CNN 1D combinando sinais de várias derivações, e Lakhdari e Saeed (2022) e Victor *et al.* (2023) com estratégias de CNN 1D com uma derivação.

Diker *et al.* (2019) apresentam um estudo comparativo com CNN 2D

convencionais usando os sinais ECG da base PTB convertido para imagens usando espectrogramas, e Yousuf *et al.* (2024) desenvolve uma CNN 2D para classificação.

Ahamed *et al.* (2020) fazem um estudo com diferentes técnicas de classificação 1D usando a bases MIT-BIH e PTB. Ahmad *et al.* (2021) também realizam um estudo comparativo com as duas bases de dados utilizando diferentes estratégias de fusão de imagens, numa estratégia de CNN multimodal.

Kim *et al.* (2023) apresentam um modelo de aprendizado de máquina para monitoramento contínuo usando Arduíno, adaptando uma CNN para que seja embarcado num *hardware* de poucos recursos.

O Quadro 5 apresenta um resumi dos indicadores com alguns trabalhos correlatos em publicações qualificadas com a base PTB, em ordem cronológica.

**Quadro 5:** Indicadores de desempenho de trabalhos correlatos para o paradigma intrapaciente com PTB.

Modelo Anteriores	Acc. %	Prec. %	Recall %	Spe %	F1 %
Sharma, Tripathy, Dandapatl (2015)	96,00	-	93,00	99,00	-
Acharya <i>et al.</i> (2017)	95,22	-	95,49	94,19	-
Chen <i>et al.</i> (2018)	96,18	-	93,67	97,32	-
Kashuee, Fazeli, Sarrafzadeh (2018)	95,90	95,20	95,10	-	-
Liu <i>et al.</i> (2018)	96,00	-	95,40	97,37	-
Dicker <i>et al.</i> (2019)	83,82	82,35	95,45	62,50	88,42
Cao <i>et al.</i> (2020)	96,65	-	-	-	-
Ahamed <i>et al.</i> (2020)	97,66	96,90	97,06	-	-
Ahmad <i>et al.</i> (2021)	99,20	98,00	98,00	-	-
Lakhdari, Saeed (2022)	99,00	99,00	98,00	-	99,00
Kim <i>et al.</i> (2023)	96,53	95,54	95,85	95,85	95,69
Victor <i>et al.</i> (2023)	98,45	98,45	98,45	-	98,45
El-Ghaish, Eldele (2024)	-	99,35	99,46	-	99,41
Yousuf <i>et al.</i> (2024)	99,84	-	99,70	99,40	-

Fonte: Próprio Autor.

### 3.2.2 Abordagem interpaciente

Os trabalhos contemporâneos que utilizam a abordagem interpaciente com a base MIT-BIH (MOODY, MARK, 2005) utilizam como principal referência a divisão de dados para treino e teste sugerida no trabalho de Chazal, O'Dwyer e Reilly (2004), denominando cada conjunto de dados de DS1 (*Dataset 1*) e DS2 (*Dataset 2*). A proposta foi organizar os dados para que cada um destes conjuntos de dados possuíssem todas as classes da forma mais equilibrada possível de diferentes pacientes. Esta mesma estratégia de abordagem ainda não foi proposta para a base PTB.

Estes autores também apresentam metodologias de filtros para os sinais ECG e usam como classificador um modelo baseado em discriminante linear, lembrando que as CNNs não estavam no estado da arte neste período. Propostas de melhorias na organização de DS1 e DS2 já citadas apresentadas por Mar *et al.* (2011) e Llamedo e Martínez (2011) também têm sido incorporadas em trabalhos contemporâneos.

Soria e Martinez (2009) usam discriminantes lineares e quadráticos para fins de classificação. Llamedo e Martínez (2011) processam as ondas R filtrando as ondas QRS com auxílio de *Wavelets*, usando classificador quadrático discriminante, metodologia de classificação também usada por Lin e Yang (2014).

Zhang *et al.* (2014) utilizam um classificador SVM com metodologia *one-versus-one*. Garcia *et al.* (2017) exploram uma representação de sinais ECG chamada vetor cardiograma temporal, juntamente com uma rede complexa para extração de características, juntamente com o classificador SVM, utilizando PSO para ajuste de hiper parâmetros, sendo um bom exemplo da abordagem de extração de recursos em várias etapas para este paradigma.

Wang *et al.* (2021) utilizaram na metodologia do seu trabalho transformada *Wavelet* e CNN, e conseguiu uma aparente alta performance nos indicadores, em especial para a classe S, mas em seu trabalho foram considerados também os dados da classe F, na qual os indicadores foram todos abaixo de 2,04%, e por isso o *F1-Score* global de 68,95%, conforme pode-se calcular a partir da sua matriz de confusão final lá disponibilizada. E, por fim, Dias *et al.* (2021) apresenta uma metodologia de segmentação própria a partir do intervalo RR e classifica com um algoritmo discriminante linear.

Observa-se que muitos trabalhos não disponibilizam suas matrizes de confusão final, focando em métricas que estão mais adequadas aos objetivos dos seus estudos, não podendo-se comparar todos os indicadores gerais e por classe, mas podem ser feitas observações mais gerais.

A literatura não costuma usar no paradigma interpaciente apenas os indicadores globais para fins de comparação, sendo tão importantes quanto a comparação com os indicadores por classe. Nesse sentido, foram elaborados os Quadros 6 e 7 que se complementam com estas duas abordagens, também em ordem cronológica de publicação. Todos também utilizam a divisão DS1 e DS2 utilizada neste estudo, mas não com a mesma metodologia de melhorias aqui utilizada, pois são opções de cada autor.

Outra diferença marcante com o paradigma intrapaciente é a ausência total da classe Q, com alguns poucos trabalhos que incluem a classe F, a qual também não foi considerada neste estudo.

**Quadro 6:** Indicadores globais de desempenho de trabalhos correlatos para o paradigma interpaciente.

Resultados Anteriores	Acc. %	Prec. %	Recall %	Spe. %	F1 %
Chazal, O'Dwyer, Reilly (2004)	83,88	45,57	66,00	96,05	59,74
Soria, Martinez (2009)	91,47	69,89	89,98	94,60	76,43
Llamedo, Martínez (2011)	90,62	66,26	86,18	95,83	66,61
Lin, Yang (2014)	93,00	67,60	83,50	-	-
Zhang <i>et al.</i> (2014)	86,66	-	-	-	-
Garcia <i>et al.</i> (2017)	92,38	70,12	81,10	92,12	74,59
Wang <i>et al.</i> (2021)	99,16	71,06	67,57	96,17	68,97
Dias <i>et al.</i> (2021)	80,58	59,37	84,84	94,35	60,77
Oliveira et al (2022)	86,12	76,72	77,04	92,87	76,88
Zahid, Kiranyaz, Gabbouj (2023)	98,19	93,25	91,35	96,37	92,27

Fonte: Próprio Autor.

**Quadro 7:** Indicadores por classe de desempenho de trabalhos correlatos para o paradigma interpaciente.

Resultados Anteriores	Acc. %	Recall %			Prec. %			F1-Score %		
		N	S	V	N	S	V	N	S	V
Chazal, O'Dwyer, Reilly (2004)	83,88	86,9	75,9	77,7	99,2	38,5	81,6	92,6	51,1	79,6
Soria, Martinez (2009)	91,45	91,7	88,3	89,9	98,9	39,5	71,2	95,2	54,6	79,5
Llamedo, Martínez (2011)	90,62	91,8	84,8	81,9	99,5	10,9	88,4	95,5	19,3	85,1
Lin, Yang (2014)	93,00	91,6	81,4	86,2	99,3	31,6	31,6	95,3	45,5	79,5
Zhang <i>et al.</i> (2014)	86,66	88,9	79,1	85,5	99,0	36,0	92,8	-	-	-
Garcia <i>et al.</i> (2017)	92,38	94,0	62,0	87,3	98,0	53,0	59,4	95,9	57,1	70,7
Zhai, Tin (2018)	-	-	76,8	93,8	-	74,0	92,4	-	-	-
Wang <i>et al.</i> (2021)	99,16	99,8	74,6	96,6	99,4	89,5	93,2	99,6	81,4	94,4
Dias <i>et al.</i> (2021)	89,61	91,0	72,1	78,7	98,1	23,6	70,8	94,4	35,6	74,5
Oliveira et al (2022)	86,12	99,8	46,1	85,2	98,6	44,9	86,6	99,2	45,5	85,9
Zahid, Kiranyaz, Gabbouj (2023)	98,19	99,3	83,3	91,4	98,9	83,5	97,4	99,1	83,4	94,3

Fonte: Próprio Autor.

### 3.3 CONSIDERAÇÕES FINAIS

Neste Capítulo foram apresentados estudos recentes que investigaram este tema de pesquisa nos diferentes paradigmas e que também utilizaram as bases de dados públicas aqui trabalhadas de forma a apresentar o campo ativo de pesquisa no qual este trabalho foi conduzido, inclusive apresentando a evolução dos indicadores de classificação no decorrer do tempo, até os dias atuais.

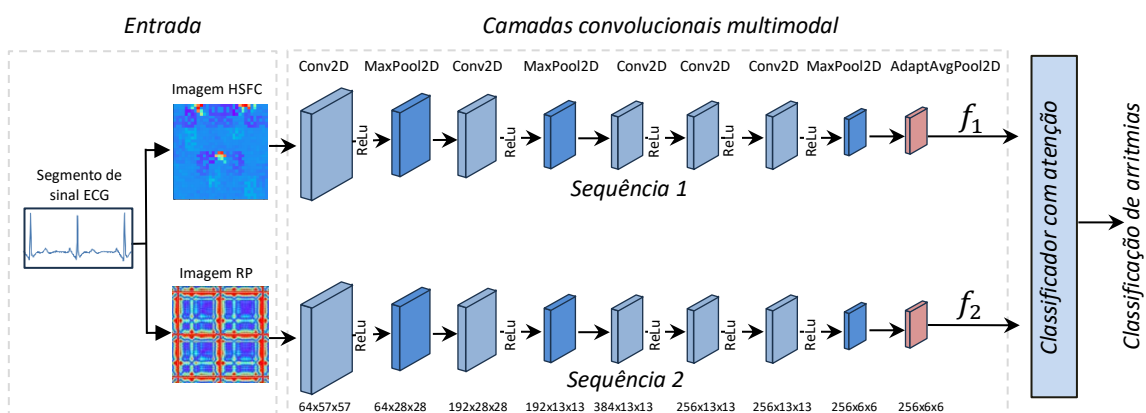
O próximo Capítulo apresenta a estrutura do modelo construído nesta tese para classificação de arritmias, apresentando a metodologia aplicada em seu desenvolvimento, destacando a organização dos bancos de dados de sinais ECG utilizados, seu tratamento

e posterior transformação em imagens. Também são apresentadas as técnicas usadas para geração de dados sintéticos para as classes minoritárias e como foram desenvolvidos e avaliados os classificadores.

## 4 CLASSIFICAÇÃO DE ARRITMIAS COM CNN MULTIMODAL E MECANISMO DE ATENÇÃO

A Figura 17 apresenta uma visão geral da estrutura proposta nesta tese para classificação de arritmias cardíacas baseada em uma CNN Multimodal com mecanismo de atenção (CNN-AM). A estrutura possui três componentes essenciais: entrada, com transformação de sinais temporais de ECG em imagens, camadas convolucionais multimodal para extração de características e classificador com mecanismo de atenção.

**Figura 17:** CNN multimodal com mecanismo de atenção.



Fonte: Di Paolo e Castro (2024).

Nas próximas seções deste Capítulo serão apresentados cada um dos 3 componentes essenciais da estrutura CNN-AM proposta, destacando a funcionalidade de cada componente e a metodologia adotada para seu desenvolvimento. Serão apresentados inicialmente os bancos de dados de sinais ECG utilizados nos experimentos para treinamento e teste da CNN-AM tanto para abordagem interpaciente quanto para abordagem intrapaciente, destacando em uma seção posterior as técnicas de aumento de dados utilizadas em cada abordagem para geração de dados sintéticos para classes minoritárias. Para finalizar, serão apresentadas as métricas usadas para avaliação do desempenho da CNN-AM durante a fase de experimentos.

### 4.1 BANCOS DE DADOS DE SINAIS ECG

Considerando as Diretrizes da AAMI (ANSI/AAMI EC57) relacionadas ao desenvolvimento de sistemas automáticos para classificação de arritmias, para os

experimentos da estrutura CNN-AM proposta neste trabalho foram utilizados três bancos de dados distintos de sinais ECG derivados das bases de dados públicas MIT-BIH e PTB considerando a abordagem interpaciente e intrapaciente.

#### 4.1.1 Banco de dados para abordagem interpaciente

O paradigma interpaciente tem o trabalho de Chazal, O'Dwyer e Reilly (2004) como principal referência da literatura científica na divisão de dados de treino e teste, sendo baseado na base de dados MIT-BIH (MOODY, MARK, 2005). Os autores denominam o conjunto de treino e teste propostos respectivamente de DS1 e DS2 (*dataset* ou conjunto de dados 1 e 2). No entanto, trabalhos posteriores relatam alguns problemas nessa divisão originalmente apresentada e, dessa forma, propõem melhorias que passaram a ser seguidas em trabalhos seguintes.

Em Mar *et al.* (2011) e Llamedo e Martínez (2011) os autores apresentam alguns problemas de desequilíbrio entre as classes dos conjuntos DS1 e DS2 o que levou os autores a proporem o acréscimo de dois prontuários do mesmo paciente nos dois conjuntos, DS1 e DS2. Assim, os autores indicam que os registros 201 e 202, apesar de serem do mesmo paciente, fossem incluídos nos conjuntos DS1 e DS2, respectivamente, resultando em um melhor balanceamento entre as classes.

Outro problema importante da divisão proposta por Chazal, O'Dwyer e Reilly (2004) foi identificada por Mar *et al.* (2011) e Llamedo e Martínez, (2011), relacionado ao uso do registro desbalanceado 232 no DS2. Esse registro contém mais de 75% dos batimentos cardíacos da classe SVEB com uma identificação incorreta de ritmo cardíaco em alguns segmentos do registro. Para manter a integridade e confiabilidade dos resultados do estudo, os autores propuseram a remoção do registro 232 de DS2, para evitar distorções nos resultados de avaliação dos algoritmos.

É importante ressaltar que a remoção desse registro é uma decisão tomada pelos autores do estudo com base em suas análises e julgamentos específicos para garantir a qualidade dos resultados obtidos, pois foi observado por eles que os métodos que atingem a classificação correta dos batimentos cardíacos desse registro 232 estão sendo especializados apenas para os batimentos cardíacos deste único paciente. Devido à importante contribuição, trabalhos recentes já consideram estas características na divisão dos dados para o paradigma interpaciente, com o Quadro 8 resumindo a organização dos

dados, divididos de forma a balancear o melhor possível as classes presentes.

**Quadro 8:** Organização dos dados pelo paradigma interpaciente.

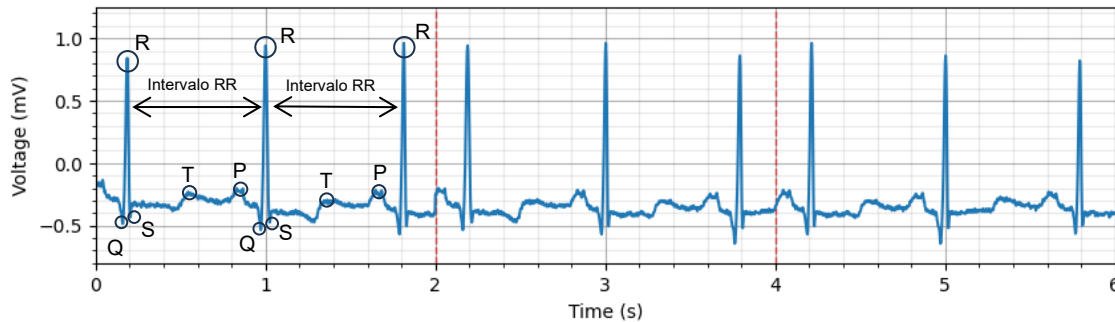
	Registro	N	S	V	F	Q	Batimentos
<b>DS1</b>	101	1.859	3	0	0	2	1.864
	106	1.506	0	520	0	0	2.026
	108	1.739	4	17	2	0	1.762
	109	2.491	0	38	2	0	2.531
	112	2.536	2	0	0	0	2.538
	114	1.819	12	43	4	0	1.878
	115	1.952	0	0	0	0	1.952
	116	2.301	1	109	0	0	2.411
	118	2.165	96	16	0	0	2.277
	119	1.542	0	444	0	0	1.986
	122	2.475	0	0	0	0	2.475
	124	1.535	31	47	5	0	1.618
	201	1.634	128	198	2	0	1.962
	203	2.528	2	444	1	4	2.979
	205	2.570	3	71	11	0	2.655
	207	1.543	106	210	0	0	1.859
	208	1.585	2	992	373	2	2.954
	209	2.620	383	1	0	0	3.004
	215	3.194	3	164	1	0	3.362
	220	1.952	94	0	0	0	2.046
	223	2.044	73	473	25	0	2.604
	230	2.254	0	1	0	0	2.255
	<b>DS2</b>	100	2.238	33	1	0	0
103		2.081	2	0	0	0	2.083
105		2.525	0	41	0	5	2.571
111		2.122	0	1	0	0	2.123
113		1.788	6	0	0	0	1.794
117		1.533	1	0	0	0	1.534
121		1.860	1	1	0	0	1.862
123		1.514	0	3	0	0	1.517
200		1.742	30	826	2	0	2.600
202		2.060	55	19	1	0	2.135
210		2.422	22	195	10	0	2.649
212		2.747	0	0	0	0	2.747
213		2.640	28	220	362	0	3.250
214		2.002	0	256	1	2	2.261
219		2.081	7	64	1	0	2.153
221		2.030	0	396	0	0	2.426
222		2.273	209	0	0	0	2.482
228	2.681	3	362	0	0	2.052	
231	1.567	1	2	0	0	1.570	
233	2.229	7	831	11	0	3.078	
234	2.699	50	3	0	0	2.752	
	DS1	45.844	943	3.788	415	8	50.998
	DS2	44.238	1.836	3.221	388	7	47.911
	Total	90.082	2.779	7.009	803	15	98.909

**Fonte:** Baseado em Chazal, O'Dwyer e Reilly (2004), Mar *et al.* (2011) e Llamedo e Martínez (2011).

Sendo assim, o banco de dados interpaciente, em nossos experimentos, foi construído segmentando os sinais de ECG de cada registro de paciente em intervalos de dois segundos, sendo que essa segmentação foi determinada a partir da observação do tempo necessário, considerando todos os registros de todos os pacientes, para capturar no mínimo dois intervalos RR por segmento, como observado no exemplo da Figura 18,

coletando o máximo de informação possível do sinal, em que se destaca a localização dos intervalos RR e as ondas Q, S, T e P. Essa abordagem de segmentação também é adotada em muitos outros trabalhos como Mathunjwa (2022), Thill *et al.* (2021), além de Jing *et al.* (2021) que ainda apresenta um estudo do efeito de diferentes tamanhos da segmentação nos indicadores de classificação.

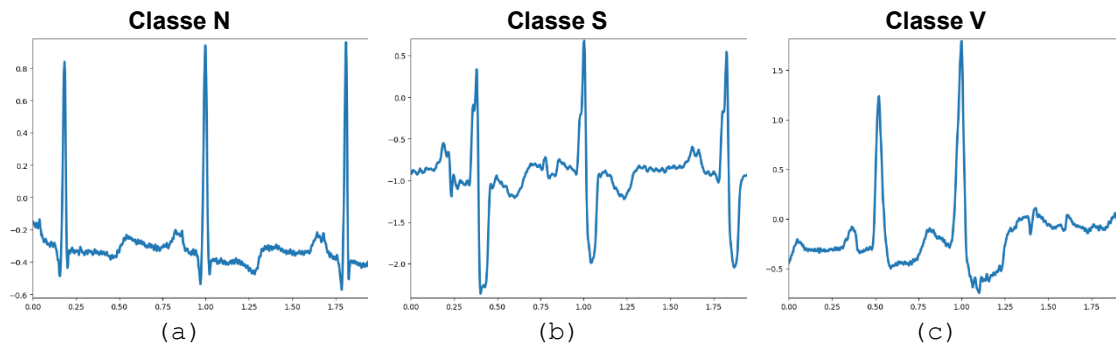
**Figura 18:** Segmentação de um sinal ECG Normal da base MIT-BIH.



**Fonte:** Próprio Autor.

Neste estudo optou-se por selecionar para classificação apenas as classes N, S e V, da mesma forma como Soria, Martinez (2009), Lin, Yang (2014), Garcia *et al.* (2017) e Dias *et al.* (2021), onde os autores consideram, como é bem observado na prática, que a classe F é considerada difícil de classificar devido à natureza complexa dos complexos QRS para esta classe, pois sua morfologia varia significativamente, sendo uma fusão de batimentos cardíacos normais com batimentos cardíacos ventriculares prematuros de forma simultânea, gerando ambiguidade na rotulagem correta dessa classe específica de arritmia, além de haverem poucos dados disponíveis para esta classe.

A derivação MLII é onipresente em todos os registros da base MIT-BIH. Esta derivação é muito utilizada para medições em longos períodos e menos sensível ao movimento dos pacientes, além de gerar sinais onde se pode destacar com mais clareza o complexo QRS e as ondas T e P em relação às demais derivações (FRANCIS, 2016; LUZ *et al.*, 2016), sendo por estes motivos a derivação mais frequentemente utilizada para desenvolvimento dos classificadores de arritmias apresentados na literatura, seja na abordagem intrapaciente ou interpaciente. A Figura 19 apresenta exemplos de segmentos de 2 segundos obtidos de cada classe considerando registros da derivação MLII.

**Figura 19:** Segmentos de 2 segundos da derivação MLII por classe da base MIT-BIH.

**Fonte:** Próprio Autor.

Seguindo como referência os resultados apresentados por Wang *et al.* (2021) e Zhang *et al.* (2023), que concluíram que a classificação utilizando as derivações V em conjunto com a derivação MLII, para estrutura classificadora proposta em seus trabalhos, poderiam levar a melhoria de resultados, em nosso trabalho optamos por também realizar experimento com dados desta derivação para verificar a influência destes sinais, em conjunto com os sinais da derivação MLII, na classificação de arritmias utilizando nossa estrutura proposta.

Os sinais de derivações V são obtidos de eletrodos posicionados no tórax, que registram os potenciais de ação na musculatura ventricular, podendo favorecer a classificação de arritmias da classe V (Ventricular) (LUZ *et al.*, 2016). Os sinais de derivações V foram extraídos e segmentados da mesma forma como já apresentado para a derivação MLII, com cada derivação extraída dos registros de pacientes conforme apresentado no Quadro 9. Observa-se que os registros dos pacientes 100, 103, 114, 117, 123 e 124 não possuem registros da derivação V1, e por isso foram utilizados os registros disponíveis de V2, V4 e V5, para completar os registros da derivação V, resultando na mesma quantidade de registros da derivação MLII.

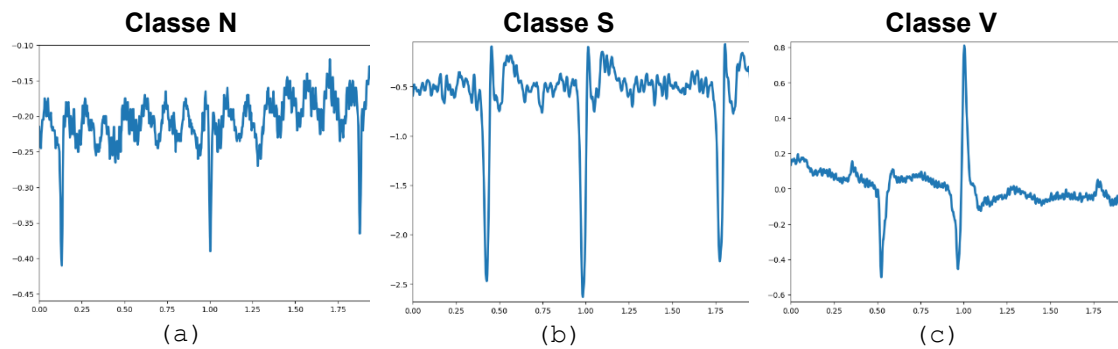
**Quadro 9:** Registros das derivações V utilizados.

Derivação	Pacientes
V1	101, 105, 106, 108, 109, 111, 112, 113, 115, 116, 118, 119, 121, 122, 200, 201, 202, 203, 205, 207, 208, 209, 210, 212, 213, 214, 215, 219, 220, 221, 222, 223, 228, 230, 231, 233, 234
V2	103, 117
V4	124
V5	100, 114, 123

**Fonte:** Próprio Autor.

A Figura 20 apresenta exemplos de sinais segmentados em intervalos de 2 segundos para cada classe, obtidos das derivações V1 extraídas da base MIT-BIH.

**Figura 20:** Exemplos de segmentos obtidos da derivação V1 por classe da base MIT-BIH.



**Fonte:** Próprio Autor.

Assim, cada registro segmentado de sinal ECG resultante, considerando tanto a derivação MLII quanto a derivação V, ficou como um vetor de 720 amostras (sinal original de 360 Hz, com 2 s de segmentação). O Quadro 10 apresenta a distribuição de dados segmentados por classes e anotações presentes na base MIT-BIH, já separados entre DS1 e DS2. O Quadro 10 apresenta o quantitativo original e o quantitativo consolidado, sendo o consolidado os dados onde se aplicou as propostas de ajustes ao trabalho original de Chazal, O'Dwyer e Reilly (2004), considerando as propostas dos trabalhos de Mar *et al.* (2011) e Llamedo e Martínez (2011), onde foi também observado que as anotações “e”, “E” e “S” da classe S possuem registros em DS1, mas não possuem registros em DS2, e por isso estes também foram eliminados deste estudo. Ressalta-se que este quantitativo consolidado apresentado no Quadro 10 refere-se tanto à derivação MLII quanto à derivação V consolidada, mantendo-se a integridade dos dados sincronizados para abordagem multimodal.

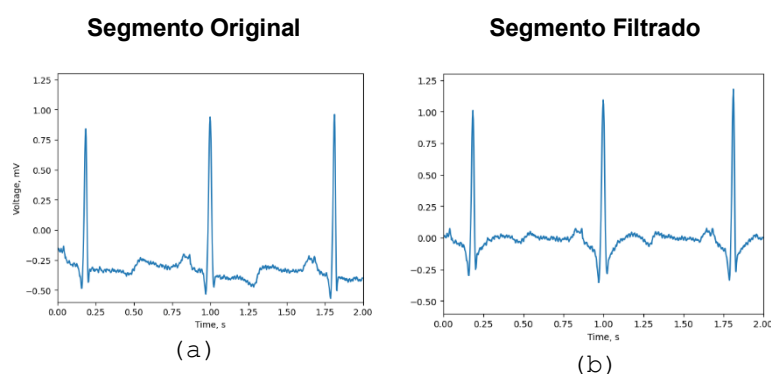
**Quadro 10:** Distribuição dos segmentos extraídos do MIT-BIH.

	Originais				Consolidados		
	DS1	DS2	DS1	DS2	DS1	DS2	
<b>N</b>	45.780	43.597	<b>L</b>	3.945	4.121	45.781	43.598
			<b>N</b>	38.058	36.401		
			<b>R</b>	3.778	3.076		
<b>S</b>	975	667	<b>A</b>	909	404	957	667
			<b>e</b>	16	0		
			<b>j</b>	48	263		
			<b>S</b>	2	0		
<b>V</b>	3.786	3.218	<b>E</b>	105	0	3.681	3.218
			<b>V</b>	3.681	3.218		
<b>Total</b>						50.419	47.483

**Fonte:** Próprio Autor.

Aos segmentos de sinal ECG do banco de dados consolidado das derivações MLII e V, foi aplicado um filtro baseado em Emrich *et al.* (2023), sendo um processo de limpeza utilizando técnica baseada em gráficos de visibilidade que procura destacar os picos R e reduzir os ruídos do sinal, com posterior remoção da linha de base, que pode ser gerada devido respiração ou movimentação durante coleta dos dados do paciente.

A Figura 21 apresenta um exemplo do processo de limpeza de sinal aplicado a um segmento de 2dois segundos da classe N do registro do paciente 10, onde em (a) é observado o segmento de sinal original, ou seja, tal qual observado na base MIT-BIH (MOODY, MARK, 2005), e em (b) o sinal após limpeza.

**Figura 21:** Segmento de sinal ECG original e filtrado.

**Fonte:** Próprio Autor.

Pode-se observar no exemplo da Figura 18 que o sinal filtrado é nivelado sob o eixo 0, com redução de ruídos, com destaques realçados para as ondas Q e S, deixando também mais padronizadas as ondas T, U e P.

Os sinais filtrados serão posteriormente utilizados para geração das imagens que serão entradas da CNN-AM proposta.

#### 4.1.2 Banco de dados para abordagem intrapaciente

No que se refere ao paradigma intrapaciente, foram utilizados três conjuntos de dados diferentes nos experimentos. Primeiramente, tem-se o trabalho de Kachuee, Fazeli e Sarrafzade (2018), que se tornou uma importante referência, por disponibilizar em ECG... (2018) registros de sinais ECG já segmentados e pré-processados provenientes da base MIT-BIH, sendo que esta base, que chamaremos de Base 1, vem servindo para desenvolvimento de classificadores e possibilitando comparação de estruturas propostas na literatura.

O banco de dados disponibilizado pelos autores foi dividido por eles aleatoriamente em 80% para treino e 20% para testes, considerando todos os pacientes da base MIT-BIH. Os sinais foram segmentados para destacar em cada amostra um pico R e sua vizinhança próxima, de forma que cada segmento possua um comprimento igual a 1,2 vezes o período de um batimento, que vai do início da onda, passando pelo pico R até o final do ciclo de batimento cardíaco.

O Quadro 11 apresenta a organização dos dados MIT-BIH disponibilizados em ECG... (2018), já devidamente rotulados e divididos em dados de treino e dados de teste.

**Quadro 11:** Organização dos dados da Base 1 intrapaciente MIT-BIH.

Classes	Tipos de Batimento	Quantidade de Dados
<b>N</b> <b>Normal</b>	Batimento normal Batimento de bloqueio de ramo esquerdo Batimento de bloqueio de ramo direito Batimento de escape atrial Batida de escape nodal (juncional)	Total: 90.589 Treino: 72.471 Teste: 18.118
<b>F</b> <b>Batimento de Fusão</b>	Fusão dos batimentos ventriculares e normais	Total: 803 Treino: 641 Teste: 162
<b>Q</b> <b>Batimento desconhecida</b>	Batimento ritmada Fusão de batimento ritmada e normal Batimento inclassificável	Total: 8.039 Treino: 6.431 Teste: 1.608
<b>S</b> <b>Batimento Ectópico Supraventricular</b>	Batimento atrial prematuro Batimento atrial prematuro aberrado Batimento prematuro nodal (juncional) Batimento prematuro supraventricular	Total: 2.779 Treino: 2.223 Teste: 556
<b>V</b> <b>Batimento Ventricular Ectópico</b>	Contração ventricular prematura Batimento de escape ventricular	Total: 7.236 Treino: 5.788 Teste: 1.448
		Total: 109.446 Treino: 87.554 Teste: 21.892

**Fonte:** Baseado em Kachuee, Fazeli e Sarrafzade (2018) e ECG... (2018).

Para este caso, o classificador desenvolvido neste estudo foi treinado para identificar as 5 classes já disponibilizadas: N, F, Q, S e V, visto que o objetivo dos autores Kachuee, Fazeli e Sarrafzade (2018) também foi o de disponibilizá-la para estudos comparativos futuros, deixando os dados já segmentados e pré-processados e prontos para implementação em sistemas inteligentes de classificação. Observa-se, também, que esta base foi construída usando o paradigma intrapaciente, não podendo reorganizá-la para interpaciente, e já se esperando que os resultados das métricas de classificação sejam mais elevados.

Além da base de dados apresentada no Quadro 11, para nossos experimentos, um segundo conjunto de dados, o qual chamaremos de Base 2, foi construído a partir do conjunto de dados usado para abordagem interpaciente, com base nos dados consolidados das derivações MLII e V. Neste novo conjunto de dados, unimos os dados de DS1 e DS2 para depois dividir os dados, de forma aleatória, em 80% para treino e 20% para teste, sem considerar pacientes. O Quadro 12 apresenta a distribuição resultante para abordagem interpaciente. A estes dados também foram aplicados o filtro baseado em Emrich *et al.* (2023). Para este caso teremos classificação apenas para as classes N, S e V como na interpaciente, o que é interessante para fins de comparação direta das métricas de classificação nas diferentes abordagens.

**Quadro 12:** Distribuição de dados da Base 2 - MIT-BIH intrapaciente.

	DS1	DS2	DS1+DS2	Treino 80%	Teste 20%
<b>N</b>	45.781	43.598	89.379	71.503	17.876
<b>S</b>	957	667	1.624	1.299	325
<b>V</b>	3.681	3.218	6.899	5.519	1.380
				78.321	19.581

**Fonte:** Próprio Autor.

Em ECG... (2018) além da base MIT-BH é disponibilizado o banco de dados PTB que também foi utilizado em experimentos em nosso estudo, e que chamaremos de Base 3. O Quadro 13 apresenta a organização dos dados para esta base, sendo que os autores disponibilizaram um arquivo para cada classe (A=Anormal e N=Normal), cabendo ao usuário a divisão de forma aleatória de 80% para treino e 20% para teste. Para estes dados, foi utilizada apenas a derivação II (MLII), com todos os dados desta derivação dos 52 pacientes de controle saudáveis rotulados como N, e apenas os dados desta derivação dos

pacientes diagnosticados com infarto no miocárdio rotulados como A.

**Quadro 13:** Organização dos dados pelo paradigma intrapaciente da base PTB (Base 3).

	Treino 80%	Teste 20%
A	8.404	2.102
N	3.237	809
	11.641	2.911

**Fonte:** Baseado em Kachuee, Fazeli e Sarrafzade (2018) e ECG... (2018).

## 4.2 AUMENTO DE DADOS

Adequado ao treinamento de bases de dados altamente desbalanceadas, o método de aumento de dados (*data augmentation*) é uma técnica usada para aumentar a quantidade de elementos de classes minoritárias que compõem o conjunto de dados, através da geração de novos dados a partir dos originais. Ao criar variações sintéticas nos dados reais existentes, a rede neural é capaz de generalizar melhor para novos dados, melhorando assim sua capacidade de classificação (HAMMAD, LIU, WANG, 2018; XIAO *et al.*, 2023; EL-GHAISH, ELDELE, 2024; ZHOU, FANG, 2024).

Existem várias metodologias possíveis para o aumento de dados, como a adição de ruído aos sinais, a rotação, espelhamento e translação de imagens, entre outras. A escolha das técnicas depende do tipo de dados a serem aumentados e dos objetivos do problema de classificação em questão.

No caso da classificação de sinais ECG, o aumento de dados é particularmente útil, pois os dados são limitados e a obtenção de novos dados pode ser difícil e custosa. Em particular, a base MIT-BIH apresenta grande desbalanceamento entre as classes, já discutido na seção 4.1.

Além disso, o aumento de dados pode ajudar a lidar com a variabilidade natural dos sinais, que pode dificultar a tarefa de classificação. A variabilidade natural dos sinais é referente à variação natural das características dos sinais adquiridos de diferentes indivíduos ou de um mesmo indivíduo em diferentes momentos. Essa variabilidade pode ser causada por vários fatores, como idade, gênero, condições de saúde, atividade física, entre outros, sendo importante ser considerada na análise de sinais, pois pode afetar a sua interpretação e diagnóstico. Portanto, é importante garantir que os modelos de classificação de sinais ECG sejam robustos o suficiente para lidar com essa variabilidade natural e não produzam resultados incorretos ou imprecisos.

O aumento de dados é uma técnica útil para lidar com essa variabilidade, permitindo que o modelo de classificação aprenda a reconhecer variações sutis nos sinais e aumentando a quantidade de dados disponíveis para treinamento do modelo.

Nesse sentido, duas técnicas se destacaram durante alguns experimentos e foram selecionadas para aumento de dados dos segmentos temporais, sendo elas: SMOTE (CHAWLA *et al.*, 2002) e WGAN-GP (GULRAJANI *et al.*, 2017; ARJOVSKY, CHINTALA, BOTTOU, 2017). A técnica SMOTE foi utilizada no aumento de dados para as classes Q e V, nos conjuntos de dados e derivações utilizados. E para as classes F e S, foi utilizada a técnica WGAN-GP. De forma empírica, a partir dos testes realizados, observou-se que esta combinação proporcionou melhores resultados de classificação da estrutura, proporcionando melhores métricas para os dados de teste, e que o uso da técnica WGAN-GP especificamente para aumento de dados de treinamento das classes F e S proporcionaram uma maior sensibilidade para estas classes mais difíceis de classificar.

Assim, o Quadro 14 apresenta o quantitativo e o aumento percentual por classe e geral em relação aos dados reais de dados sintéticos gerados para as classes F, Q, S e V para a base MIT-BIH em ECG... (2018), e os Quadros 15 e 16 para a base MIT-BIH em Moody e Mark (2005), nos paradigmas inter e intrapaciente, respectivamente, nas derivações MLII e V. Para a base PTB, o aumento de dados foi explorado, sem ganhos observados. Essa característica também foi observada nos artigos correlatos.

**Quadro 14:** Conjunto de dados para o paradigma intrapaciente da Base 1 com aumento de dados.

		Dados Originais	Dados Sintéticos	Aumento Percentual	Total Consolidado
<b>Treino</b>	F	641	11.359	1.772,07%	12.000
	N	72.471	0		72.471
	Q	6.431	3.569	55,50%	10.000
	S	223	17.777	7.971,75%	20.000
	V	5.788	4.212	72,77%	10.000
<b>Teste</b>	F	162	0		162
	N	18.118	0		18.118
	Q	1.608	0		1.608
	S	556	0		556
	V	1.448	0		1.448
<b>Total:</b>		109.446	36.917	33,73%	146.363

**Fonte:** Próprio Autor.

**Quadro 15:** Conjunto de dados para o paradigma interpaciente da base MIT-BIH com aumento de dados.

		Dados Originais	Dados Sintéticos	Aumento Percentual	Total Consolidado
<b>DS1</b>	N	45.781	0		45.781
	S	957	29.043	3.034,80%	30.000
	V	3.681	6.319	171,67%	10.000
<b>DS2</b>	N	43.598	0		43.598
	S	667	0		667
	V	3.218	0		3.218
<b>Total:</b>		97.902	35.362	36,12%	133.264

**Fonte:** Próprio Autor.

**Quadro 16:** Conjunto de dados para o paradigma intrapaciente da base 2 MIT-BIH com aumento de dados.

		Dados Originais	Dados Sintéticos	Aumento Percentual	Total Consolidado
<b>Treino (80%)</b>	N	71.503	0		71.503
	S	1.299	28.701	2.209,47%	30.000
	V	5.519	4.481	81,19%	10.000
<b>Teste (20%)</b>	N	17.876	0		17.876
	S	1.380	0		1.380
	V	325	0		325
<b>Total:</b>		97.902	33.182	33,89%	131.084

**Fonte:** Próprio Autor.

Neste estudo, optou-se pelo aumento de dados diretamente nos sinais temporais de ECG e não o aumento de dados nas imagens que serão geradas a partir destes sinais temporais que serão as entradas para as CNNs extratoras de características. Esta escolha se deve principalmente ao fato de que técnicas de aumento de dados em imagens como de recorte, reflexão ou rotação de imagens não seriam adequadas no nosso contexto de imagens geradas a partir dos segmentos de sinais temporais visto que para sinais ECG, essas operações mudariam a semântica dos sinais que as imagens representam, pois as técnicas utilizadas concentram informações espaciais em pontos específicos de acordo com o observado em sinais temporais.

As técnicas SMOTE e WGAN-GP serão apresentadas a seguir de acordo como foram aplicadas em nosso trabalho.

### 4.2.1 SMOTE

O método SMOTE (*Synthetic Minority Over-sampling Technique*), desenvolvido por Chawla *et al.* (2002), tem sido aplicado em diversas tarefas de classificação, incluindo classificação de sinais ECG. A ideia básica do SMOTE é gerar novos dados sintéticos para as classes minoritárias, a fim de equilibrar a distribuição dos dados. O SMOTE é implementado de acordo com as seguintes etapas:

- (1) Seleção de um ponto de referência: selecionar um ponto de referência da classe minoritária;
- (2) Seleção de um vizinho: escolher um vizinho aleatório do ponto de referência;
- (3) Geração de novos pontos: gerar novos pontos sintéticos ao longo do segmento de linha que une o ponto de referência e o vizinho selecionado.
- (4) Repetição do processo: repetir as etapas 2 e 3 várias vezes para gerar vários pontos sintéticos.

Os pontos sintéticos são gerados a partir de:

$$\text{new\_point} = \text{point\_of\_reference} + \text{random\_value} * (\text{neighbor} - \text{point\_of\_reference}) \quad (8)$$

onde “point\_of\_reference” é o ponto selecionado da classe minoritária, “neighbor” é o vizinho selecionado aleatoriamente, e “random\_value” é um número aleatório entre 0 e 1.

Para aplicar o SMOTE em sinais ECG, cada sinal de uma classe minoritária específica é convertido em uma matriz bidimensional de amplitude e tempo, e os pontos de referência são selecionados aleatoriamente. As etapas de seleção de vizinho e geração de novos pontos são realizadas no espaço da matriz de amplitude-tempo.

O SMOTE tem sido amplamente utilizado em estudos de classificação de sinais ECG, sendo atualmente de fácil implementação, e vários estudos relataram melhoria significativa na acurácia da classificação após o uso do SMOTE, tais como Elreedy e Atiya (2019), Ahmad *et al.* (2021), Victor *et al.* (2023), El-Ghaish e Eldele (2024) e Zhou e Fang (2024).

### 4.2.2 WGAN-GP

O modelo *Wasserstein Generative Adversarial Network with Gradient Penalty* (WGAN-GP) foi proposto por Gulrajani *et al.* (2017), sendo uma extensão do modelo proposto originalmente por Arjovsky, Chintala e Bottou (2017). O WGAN-GP introduz uma técnica de gradiente de penalidade para melhorar a estabilidade e o desempenho do treinamento do modelo original. No estudo, os autores do modelo WGAN-GP destacam os problemas do treinamento do GAN convencional, como a instabilidade e a falta de métricas de avaliação consistentes, e propõem o uso da distância de Wasserstein como uma métrica para avaliar a qualidade dos resultados gerados pela rede geradora e, além disso, introduzem o termo de penalidade de gradiente.

Nesse ponto, cabe um melhor esclarecimento sobre o desenvolvimento das Redes Adversariais Generativas (*Generative Adversarial Network*, GAN). As GANs foram propostas pela primeira vez por Goodfellow *et al.* (2014) e têm sido uma área ativa de pesquisa na IA desde então, sendo originalmente aplicadas para dados 1D, mas também aplicáveis a dados multidimensionais. São um tipo de modelo de aprendizado de máquina que consiste em duas redes neurais, o gerador buscando gerar dados que enganem o discriminador, enquanto o discriminador tenta identificar dados falsos. A abordagem original de GANs foi revolucionária e abriu um novo paradigma de geração de dados sintéticos com qualidade sem precedentes, para gerar dados sintéticos que se assemelham a dados reais.

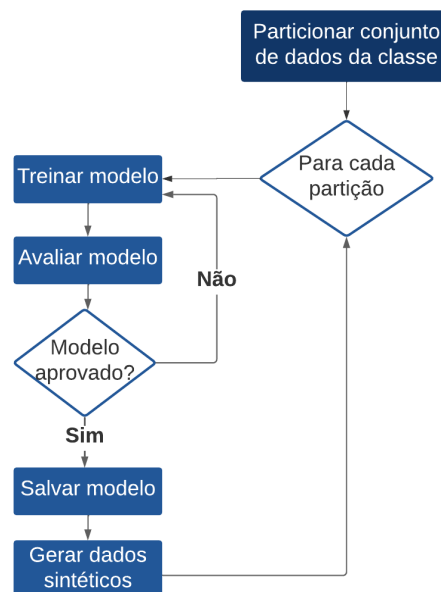
De maneira geral, gerar dados sintéticos com modelos baseados em GAN são mais custosos computacionalmente em relação a modelos estatísticos como SMOTE, exigindo etapas de treino e avaliação dos dados sintéticos gerados. Também observou-se com a prática que um único modelo não gera diversidade de dados sintéticos suficientes para servir como gerador de dados sintéticos. Nesse sentido, foi elaborada e realizada as etapas apresentadas na Figura 22 para buscar maior diversidade de dados.

Primeiramente, para cada classe minoritária que se deseja aumentar, deve-se usar apenas os dados de treino existentes e particioná-lo, por exemplo, em três partes iguais. Para cada partição, ao menos três ou mais modelos WGAN-GP são gerados. A avaliação de cada modelo não foi realizada de forma automatizada, tendo sido avaliados por inspeção visual dos sinais ECG 1D sintéticos, comparando-se os padrões dos sinais temporais reais existentes na partição utilizada para construção deste modelo, sendo

considerados como adequados padrões visualmente regulares de sinais temporais.

Com esses modelos, são gerados dados sintéticos que serão agregados à base de treino da classe minoritária que se deseja aumentar. Com pelo menos nove modelos construídos para cada classe minoritária, garante-se maior diversidade de dados sintéticos produzidos, com cada modelo gerando uma determinada quantidade de sinais ECG sintéticos para uma mesma classe.

**Figura 22:** Etapas para geração de dados sintéticos em modelos baseados em GAN.



**Fonte:** Próprio Autor.

### 4.3 TRANSFORMAÇÃO DE SINAIS ECG PARA IMAGENS

O primeiro componente da estrutura CNN-AM apresentada na Figura 14 recebe um segmento de sinal ECG e transforma este sinal em duas imagens que serão entregues para o componente de camadas convolucionais multimodal. Várias técnicas para transformação de sinais biológicos em imagens vêm sendo apresentadas em trabalhos na literatura com o objetivo de utilizar as imagens geradas como entrada para estruturas de CNN 2D para problemas de classificação.

Em Qi *et al.* (2023), por exemplo, os autores usaram *short-time Fourier Transform* (STFT) para transformar dados de eletroencefalograma (EEG) em imagens 2D para previsão de crises de epilepsia. Em Machida, Nambu e Wada (2023) os autores usaram STFT para gerar imagens de ondas cerebrais, apresentando melhores rendimentos nesta abordagem com imagens em comparação à abordagem 1D. Considerando sinais

ECG, Seitanidis, Gialelis e Papaconstantinou (2022) usaram representações gráficas da linha espacial-temporal para representações gráficas 2D; Singh e Krishnan (2023) trabalharam no domínio da frequência apresentando pipeline com diversos processamentos para geração de imagens com extração de recursos, incluindo *Fast Fourier Transform* (FFT); Ahmad *et al.* (2021) trabalharam com as técnicas *Gramian Angular Field* (GAF), *Recurrence Plot* (RP) e *Markov Transition Field* (MTF) e Mathunjwa *et al.* (2021) também usaram RP para representação de sinais ECG; Adib *et al.* (2023) fizeram uso de MTF e Farag (2022) representou sinais ECG com STFT e Heat Map.

Cada uma das técnicas de transformação de sinais temporais em imagens apresenta características próprias, sendo que cada imagem gerada pode dar mais ênfase para determinados segmentos do sinal temporal, sendo então uma etapa importante do nosso trabalho determinar quais seriam as melhores técnicas a serem utilizadas, considerando que nosso objetivo visou o uso de mais de uma técnica para que as camadas convolucionais multimodal pudessem ser utilizadas para extrair características de alto nível destas imagens para posteriormente estas características serem fundidas com o intuito de se obter características ainda mais robustas para serem usadas pelo classificador.

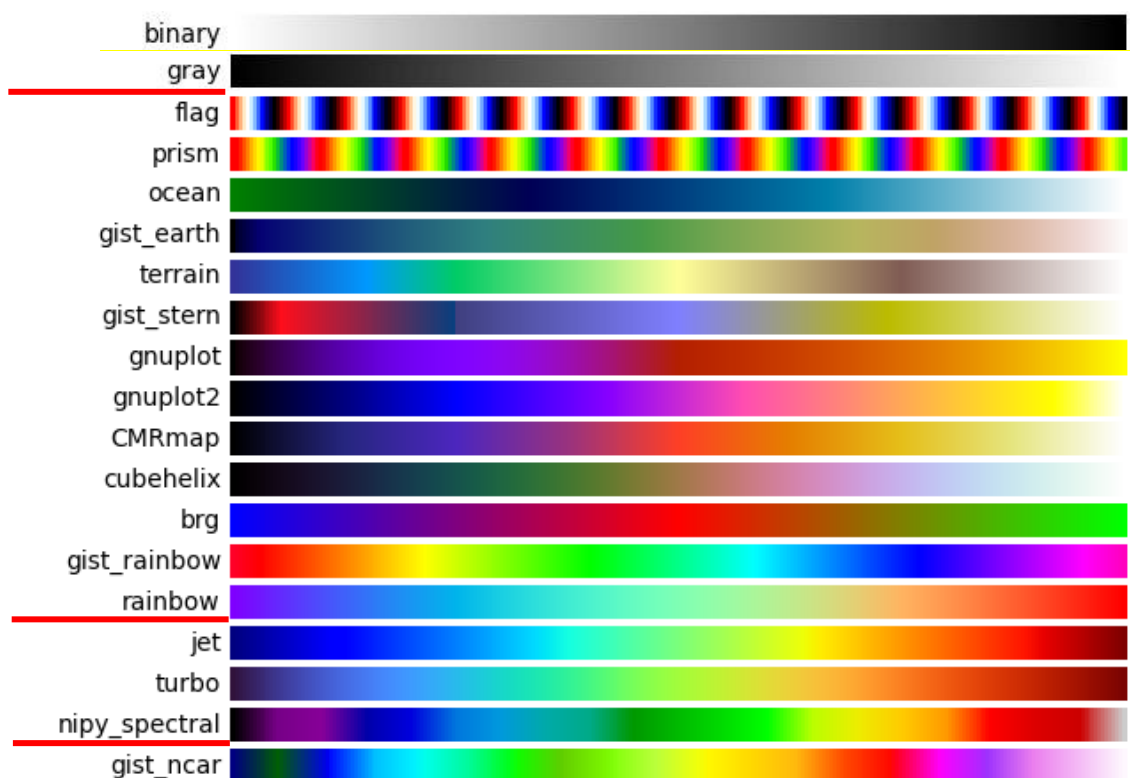
Desta forma, para definição das técnicas que seriam utilizadas para transformação de sinais ECG em imagens, diversos experimentos foram realizados com a estrutura proposta utilizando as técnicas já comumente usadas na literatura para transformação de sinais ECG em imagens 2D, sendo elas as técnicas *Gramian Angular Field* (GAF), *Recurrence Plot* (RP), *Markov Transition Field* (MTF) e *Heat Map* (HM), e utilizando também uma técnica ainda não testada nesta área, sendo ela a técnica de *Hilbert Space Filling Curve* (HSFC).

A técnica HSFC já vem sendo utilizada em outros domínios de aplicação, com bons resultados, como em Zhang *et al.* (2022) onde a técnica foi utilizada para representar gestos humanos captados por eletrodos, ou como em Valdes *et al.* (2023) onde a técnica foi utilizada para visualização e exploração de perfis metagenômicos a partir de dados de sequenciamento de DNA, e também em Yang *et al.* (2024) em que os autores converteram dados de fala humana para curvas de Hilbert com o objetivo de identificar emoções. Assim, na fase de experimentos, em busca da melhor estrutura do classificador, variamos o número técnicas de transformação de sinais temporais em imagens de 1 até 3 (número

de entradas para camadas convolucionais multimodal, considerando aqui a possibilidade de até 3 camadas convolucionais na estrutura CNN-MA), sendo que os melhores resultados obtidos, considerando as métricas adotadas neste trabalho para avaliação do classificador de arritmias durante fase de experimentos, foram obtidos para o caso de uso de duas técnicas no componente de entrada, sendo elas HSFC e RP, sendo então estas técnicas as definidas para a estrutura CNN-MA proposta.

Nos experimentos realizados, foi observado que a escala de cores para geração das imagens influencia severamente nos resultados da classificação. Diante desse fato, diversos experimentos foram conduzidos usando Matplotlib ([2023?]). A Figura 23 apresenta uma relação não exaustiva de escalas de cores possíveis para geração de imagens, com destaque para as que foram utilizadas nesta pesquisa, onde concluímos que o mapa de cores com o melhor resultados nos experimentos realizados foi *Rainbow*. Uma justificativa para a *Rainbow* ter gerado os melhores resultados de classificação é o fato desse mapa de cor armazenar mais informação, pois a menor variação numérica dos *pixels*, muda fortemente a sua cor também, diferente de mapas de cores como *flag* e *ocean*, que apresentam menos cores e menor diversidade na variação dos *pixels*.

**Figura 23:** Alguns mapas de cores.

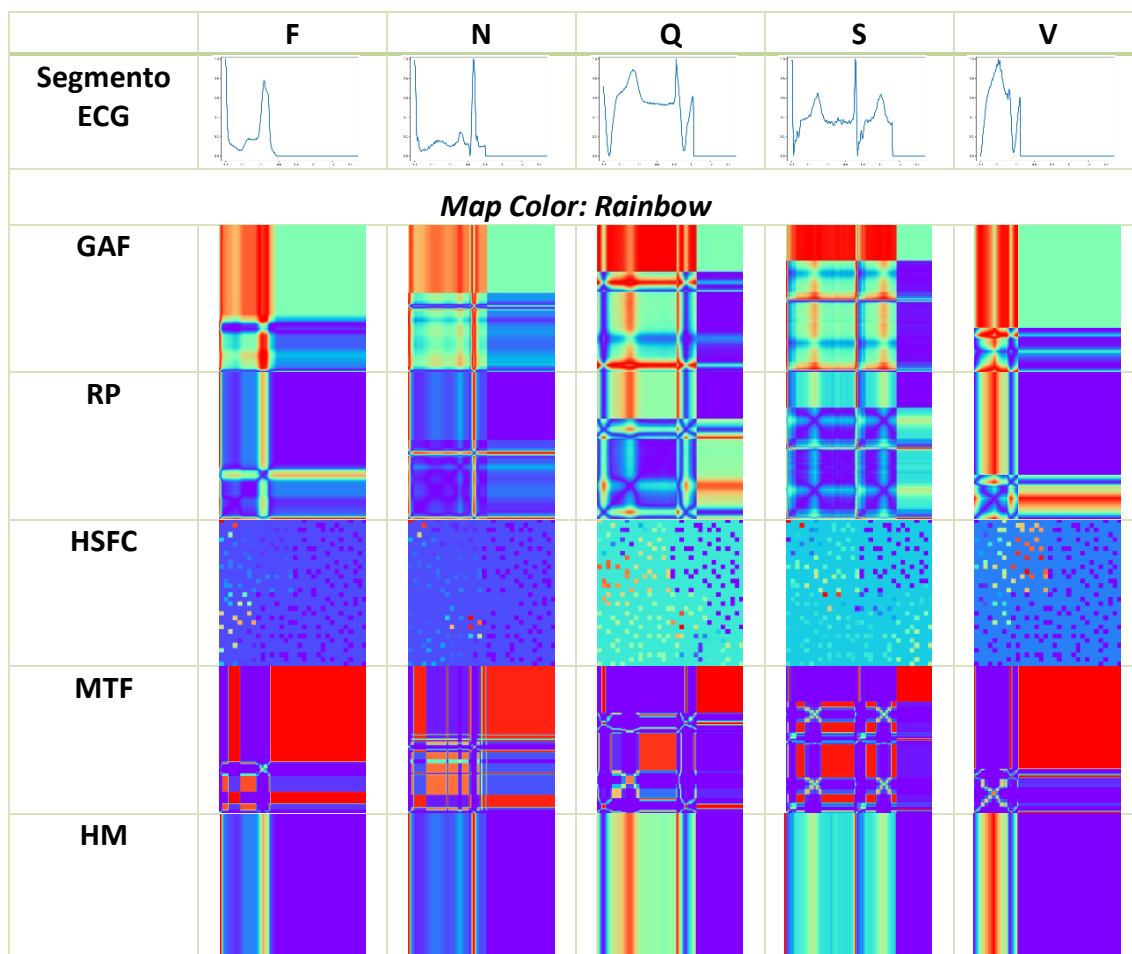


**Fonte:** Matplotlib ([2023?]).

Para implementação da técnica GAF, foi utilizada como referência os estudos de Faouzi e Janati (2020), que desenvolveram a biblioteca Pyts, em Python, aqui utilizada na versão 0.12, com bons resultados apresentados. Foi também desenvolvida uma implementação para RP com base em Eckmann *et al.* (1987), Casdagli (1997), Faria *et al.* (2016) e Ahmad *et al.* (2021). Já para HSFC, a implementação foi baseada nos estudos de Hilbert (1891), Skilling (2004), Earl (2013), Wang e Oates (2015) e Bragin e Spitsyn (2020).

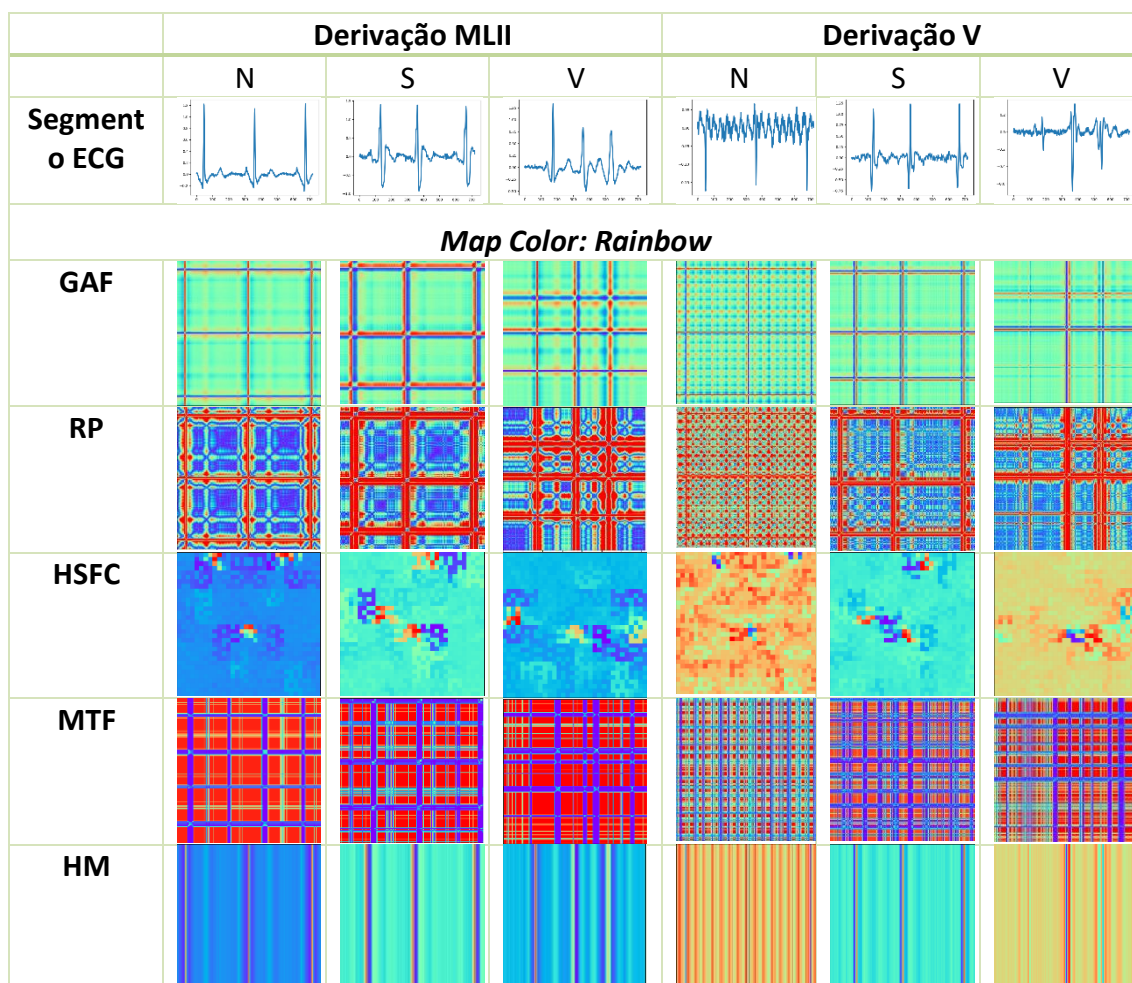
As Figuras 24 a 26 ilustram, como exemplo, um segmento de sinal ECG para cada classe e sua respectiva imagem gerada em RGB no mapa de cores *Rainbow*, considerando cada uma das técnicas usadas em nossos experimentos. Todos os conjuntos de dados de segmentos de sinais ECG usados para os experimentos, tanto para abordagem interpaciente quanto intrapaciente, passaram pelo processo de transformação para imagens. A Figura 24 apresenta exemplos de dados da base MIT-BIH em ECG... (2018), construída por Kachuee, Fazeli e Sarrafzade (2018). A Figura 25 apresenta exemplos da base MIT-BIH segmentados e filtrados, para casos das derivações MLII e V. A Figura 26 apresenta exemplos de dados da base PTB disponibilizada por Kachuee, Fazeli e Sarrafzade (2018).

**Figura 24:** Transformação de segmento de sinal ECG da base MIT-BIH em ECG... (2018) em imagens para cada classe, utilizando os métodos GAF, RP, HSFC, MTF e HM.



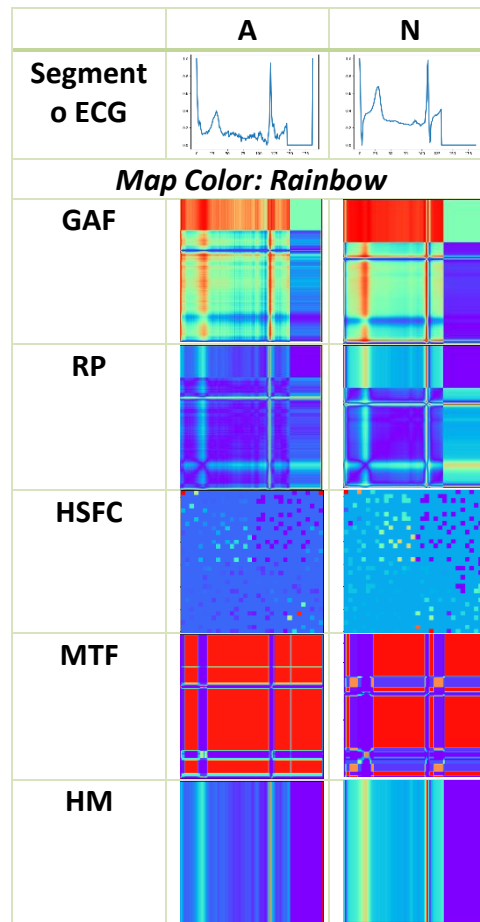
Fonte: Próprio autor.

**Figura 25:** Transformação em imagens de segmento de sinal ECG da base MIT-BIH para cada classe, utilizando os métodos GAF, RP, HSFC, MTF e HM.



Fonte: Próprio autor.

**Figura 26:** Transformação em imagem de segmento de sinal ECG da base PTB para cada classe, utilizando os métodos GAF, RP, HSFC, MTF e HM.



Fonte: Próprio autor.

As técnicas RP e HSFC, como já indicado, foram as técnicas, que após os diversos experimentos, geraram os melhores resultados para estrutura CNN-AM. Das Figuras 24 até 26, considerando as imagens geradas pela técnica RP, podemos observar regiões que representam o sinal temporal nas proximidades dos picos, sendo a imagem bastante sensível às mudanças sutis na frequência ou amplitude do sinal temporal. Variações de amplitude são observadas como diferenças de intensidade de cores dos *pixels* das regiões recorrentes. Altas amplitudes aparecem como áreas vermelhas enquanto baixas amplitudes resultam em áreas azuladas. Variações de frequência são visualizadas através do espaçamento entre a diagonal principal vista nos gráficos, onde frequências altas resultam em linhas próximas e frequências baixas em linhas mais espaçadas. O RP oferece uma visão clara das estruturas repetitivas e padrões globais, destacando regularidades e irregularidades no ritmo cardíaco.

Considerando agora as imagens geradas pelo HSFC, elas fornecem uma visão

global do segmento do sinal temporal, destacando padrões e estruturas em escalas de tempo mais amplas, sendo menos afetada por variações rápidas que ocorrem em intervalos pequenos no domínio (alta frequência num curto tempo). As variações de amplitude aparecem como diferenças na intensidade (cor vermelha são os picos, na escala de cor *Rainbow*) entre diferentes regiões da imagem, enquanto as variações de frequência são vistas como mudanças na densidade e repetição dos padrões de cor no espaço, em que as mesmas cores em sequência representam frequências próximas e a variação de cores também representam a variação de frequência. O Quadro 17 resume as características observadas nas imagens geradas em RP e HSFC.

**Quadro 17:** Características do sinal ECG nas imagens geradas por RP e HSFC.

Sinal ECG	Representação na Imagem RP	Representação na Imagem HSFC
<b>Ritmo dos Batimentos</b>	Linhas perpendiculares escuras sobre a diagonal principal, formando quadrados.	Padrões suaves e repetitivos ao longo na imagem.
<b>Variações Rápidas</b>	Mais sensível. Mudanças súbitas aparecem como rompimentos no padrão.	Menos sensível. Variações rápidas são diluídas na imagem sendo observadas pela variação de cor.
<b>Amplitudes Elevadas (pico R)</b>	Destaque como pontos recorrentes ou áreas densas ao longo da diagonal principal.	Pontos mais intensos na imagem.
<b>Regularidade Temporal</b>	Simetria e espaçamento das linhas ao longo da diagonal principal mostram ritmo.	Distribuição da escala de cores ao longo do espaço.

**Fonte:** Próprio Autor.

### 4.3.1 Pré-processamento das imagens

Após a transformação dos segmentos de sinais ECG em imagens, foi realizado um processo de normalização. A normalização é uma técnica comum usada para pré-processamento das imagens antes de serem usadas como entrada em modelos de aprendizado profundo, incluindo a redes neurais convolucionais (CNNs). A normalização é importante porque as imagens geralmente possuem uma ampla variação de intensidade de *pixel* e distribuição de cor, o que pode levar a problemas de treinamento, como a convergência lenta e a estabilidade inadequada do modelo.

Sem a normalização, os valores de *pixel* da imagem podem variar amplamente, o que pode afetar a escala dos gradientes durante a retropropagação do erro, levando a uma convergência mais lenta ou até instável. A normalização ajuda a reduzir a variação dos valores de *pixel*, o que pode melhorar a estabilidade do treinamento e acelerar a convergência do modelo.

Além disso, a normalização ajuda a reduzir a sensibilidade da CNN a pequenas

variações nos valores de *pixel*, o que pode melhorar a capacidade do modelo de generalizar para novos dados de teste. Isso é particularmente importante em tarefas de classificação de imagens, onde o modelo deve ser capaz de reconhecer padrões semelhantes em imagens de treinamento e teste.

A normalização aplicada neste estudo foi realizada em duas etapas. Primeiramente, foram avaliadas diferentes resoluções e densidade de *pixels*. Nas técnicas de conversão de sinais temporais em imagens aqui estudadas, independentemente do tamanho de amostras de um sinal temporal, a imagem gerada é sempre quadrada, e seu tamanho e resolução são personalizados. Na literatura pesquisada, são mais comuns tamanhos de imagens de  $64 \times 64$ ,  $224 \times 224$  e  $227 \times 227$  *pixels*.

Neste estudo foi considerado o estudo de Krizhevsky, Sutskever e Hinton (2012), original da rede AlexNet, com uso de imagens  $227 \times 227$  *pixels*, e o estudo de Krizhevsky (2014), com atualização da AlexNet para uma GPU usando resolução  $224 \times 224$  *pixels*, por considerar uma resolução não tão pequena para esconder detalhes na imagem que possam diferenciar as classes, e não muito grande que possa implicar em maior robustez e *hardware* para processamento. Ambas as resoluções apresentaram resultados similares, observando-se que para resoluções menores as métricas de avaliação da CNN-AM foram bem inferiores, e para valores acima gerava-se um maior tempo de processamento, sem melhoria nas métricas. Assim, todas as imagens, originais ou sintéticas, foram cuidadosamente geradas para garantir resolução  $227 \times 227$  ou  $224 \times 224$  *pixels* e resolução de 300 dpi.

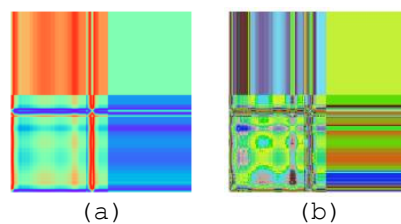
A etapa seguinte para normalização foi utilizar os valores médios (*mean*) e desvio padrão (*std*) de cada canal de cor das imagens (três para RGB e um para escala de cinza). A razão para usar a média e o desvio padrão das imagens para normalização é ajustar a escala dos valores dos *pixels* das imagens de forma que eles fiquem próximos de uma distribuição normal (gaussiana) com média zero e desvio padrão um, forçando os valores dos *pixels* dentro de uma faixa mais restrita, com a propagação do gradiente mais estável e com convergência mais rápida durante o treinamento da CNN-MA (IOFFE, SZEGEDY, 2015; AWAIS, IQBAL, 2021).

A identificação destes parâmetros de normalização foi realizada nas imagens da base de dados de treinamento e, com os parâmetros identificados, estes foram aplicados tanto nas imagens da base de treino quanto nas imagens da base de testes para

normalização. É importante que a penas a base de treino seja usada nesta identificação de parâmetros para evitar transferência de aprendizado, mantendo a influência da base de teste fora do processo de treinamento.

Assim, para cada imagem no conjunto de treino foi calculada a média de cada um dos canais de cor. Posteriormente, com essa média, a partir do conjunto de treino foi calculado o desvio padrão dos *pixels* que formam cada canal de cor. Identificando-se a média e o desvio padrão dos canais de cor, aplicou-se a normalização a todos os dados da base de treino e teste. A Figura 27 ilustra resultado deste processo de normalização para uma única imagem (a) que representa um segmento de sinal E. Observa-se que a borda esquerda mais avermelhada e as áreas quadriculadas nos cantos inferiores esquerdo e direito na imagem originais (a), entre outros elementos, ficam bem mais detalhados na imagem normalizada (b), as quais aparecem com muitos detalhes mais destacados, representando uma maior variabilidade numérica dentro de um intervalo entre zero e um.

**Figura 27:** Normalização de imagem: (a) original; (b) normalizada por média e desvio padrão dos canais de cor.



**Fonte:** Próprio Autor.

### 4.3.2 Armazenamento das imagens

Existem diversas formas de armazenar as imagens em disco para fins de processamento, sendo três exploradas aqui:

- **JPG:** Formato compactado com perdas, baseado em transformações matemáticas, muito utilizado para fotografias, ou quando a compressão ou redução de tamanho de arquivo seja necessária.
- **PNG:** Formato compactado sem perdas, ideal para aplicações com necessidade de alta qualidade, mesmo com compressão.
- **HDF5:** Formato binário generalista, projetado para armazenar grandes volumes de dados de forma eficiente e hierárquica, com

suporte opcional para compressão sem perdas.

O melhor custo/benefício observado foi com o formato PNG, gerando melhores resultados e ocupando espaço em disco dentro dos limites dos recursos computacionais disponíveis. Os Quadros 18 a 20 ilustram o conjunto final de imagens como dados de entrada para a CNN-AM, sendo o Quadro 18 com intrapaciente e a base MIT-BIH em ECG... (2018), e os Quadros 19 e 20 com o paradigma interpaciente com dados da base MIT-BIH em Moody e Mark (2005), para as derivações MLII e V, respectivamente, já segmentados e filtrados, incluindo as imagens de dados sintéticos. Pode-se observar a distribuição de ocupação em espaço em disco por classe e por técnica de geração de imagens.

**Quadro 18:** Características do conjunto de dados no formato PNG, paradigma intrapaciente com ECG... (2018).

		Imagens Originais	Imagens Sintéticas	Espaço em Disco (MB) RP	Espaço em Disco (MB) HSFC
<b>Treino</b>	F	641	11.359	447	27
	N	72.471	0	1566	153
	Q	6.431	3.569	265	24
	S	223	17.777	909	48
	V	5.788	4.212	230	22
<b>Teste</b>	F	162	0	2	0,3
	N	18.118	0	390	39
	Q	1.608	0	43	0,4
	S	556	0	15	2,2
	V	1.448	0	34	0,4
<b>Total:</b>		109.446	36.917	3.901	316,3

**Fonte:** Próprio Autor.

**Quadro 19:** Características do conjunto de dados da derivação MLII no formato PNG, paradigma interpaciente filtrado.

		Imagens Originais	Imagens Sintéticas	Espaço em Disco (MB) RP	Espaço em Disco (MB) HSFC
<b>DS1</b>	N	45.781	0	3.011	171
	S	957	29.043	1.936	120
	V	3.681	6.319	696	39
<b>DS2</b>	N	43.598	0	2.837	167
	S	667	0	37	3
	V	3.218	0	235	13
<b>Total:</b>		97.902	35.362	8.752	513

**Fonte:** Próprio Autor.

**Quadro 20:** Características do conjunto de dados no formato PNG da derivação V, paradigma interpaciente filtrado.

		Imagens Originais	Imagens Sintéticas	Espaço em Disco (MB) RP	Espaço em Disco (MB) HSFC
<b>DS1</b>	N	45.781	0	3.113	188
	S	957	29.043	2.100	125
	V	3.681	6.319	625	40
<b>DS2</b>	N	43.598	0	2.806	173
	S	667	0	36	3
	V	3.218	0	229	13
<b>Total:</b>		97.902	35.362	8.909	542

**Fonte:** Próprio Autor.

As imagens com formato JPG podem ocupar mais ou menos espaço em disco, dependendo da configuração de compressão, mas com espaço em disco equivalente, o formato PNG se destacou. Já no formato HDF5, o espaço em disco ocupado é muito superior e, conseqüente, maior tempo de processamento para desenvolvimento dos classificadores, sem que se tenha observado melhoria nas métricas.

#### 4.4 CAMADAS CONVOLUCIONAIS MULTIMODAIS

O componente de Camadas Convolucionais Multimodais da estrutura CNN-AM apresentado na Figura 14 tem a função de extrair características espaciais das imagens provenientes do componente de entrada, sendo composta de duas seqüências de camadas convolucionais em paralelo baseadas na estrutura da AlexNet (KRIZHEVSKY, 2014), com algumas adaptações para melhor destacar as bordas das imagens de entrada, explicado mais adiante, que para o caso das imagens geradas por HSFC e RP a partir do segmentos de sinal ECG, podem conter informações importantes, conforme pôde ser observado nas Figuras 21 a 23. Estruturas de CNN mais densas como VGG e ResNet foram testadas, sendo que os melhores resultados foram obtidos com a AlexNet.

O Quadro 21 apresenta com detalhes a estrutura de cada uma das seqüências de camadas convolucionais, sendo que na entrada de cada seqüência tem-se uma imagem no formato RGB (três canais), com resolução  $224 \times 224$  e na saída da seqüência tem-se uma matriz  $6 \times 6$ , com 256 canais.

**Quadro 21:** Estrutura de cada uma das sequências de camadas convolucionais.

Names das camadas	Tamanho da saída	Tamanho do Kernel	Padding	Strid
Entrada	$3 \times [224 \times 224]$			
<b>Features:</b>				
Conv2D	$[3 \times 64]$	$[11 \times 11]$	$[6 \times 6]:\text{Reflect}$	$[4 \times 4]$
ReLu				
MaxPool2D		$[3]$	$[0]$	$[2]$
Conv2D	$[64 \times 192]$	$[5 \times 5]$	$[2 \times 2]:\text{Reflect}$	$[1 \times 1]$
ReLu				
MaxPool2D		$[3]$	$[0]$	$[2]$
Conv2D	$[192 \times 384]$	$[3 \times 3]$	$[1 \times 1]:\text{Reflect}$	$[1 \times 1]$
ReLu				
Conv2D	$[384 \times 256]$	$[3 \times 3]$	$[1 \times 1]:\text{Reflect}$	$[1 \times 1]$
ReLu				
Conv2D	$[256 \times 256]$	$[3 \times 3]$	$[1 \times 1]:\text{Reflect}$	$[1 \times 1]$
ReLu				
MaxPool2D		$[3]$	$[0]$	$[2]$
<b>Avgpool:</b>				
AdaptiveAvgPool2d	$256 \times [6 \times 6]$			

**Fonte:** Di Paolo e Castro (2024).

Cada sequência possui uma componente chamada *Features* que tem a função de extrair características de alto nível da imagem apresentada na sua entrada. Ela recebe uma imagem no formato RGB (três canais), com resolução  $224 \times 224$  e realiza uma convolução com 64 filtros de tamanho  $11 \times 11$ , com um passo de 4 e preenchimento de 6, onde é usado o modo de preenchimento reflexivo, o que destaca as bordas, ajudando desta forma a extrair características de baixo nível da imagem. Em seguida, tem-se uma função de ativação *Rectified Linear Unit* (ReLU) para introduzir não-linearidade e, em seguida aplica-se *Max Pooling* com um tamanho de *kernel* de  $3 \times 3$  e um *stride* de 2, reduzindo a resolução espacial da saída (largura e altura) pela metade devido o *stride* 2, selecionando o valor máximo em cada região de um tamanho  $3 \times 3$  e movendo a janela de *pooling* com um passo de 2 *pixels*, mas mantendo os 64 canais da entrada. A seguir tem-se uma segunda camada de convolução com 192 filtros de tamanho  $5 \times 5$  e preenchimento de 2, que usa também o modo de preenchimento reflexivo para destaque das bordas, seguida de uma função de ativação ReLU e *Max Pooling* com um tamanho de *kernel* de  $3 \times 3$  e um passo de 2. Logo a seguir tem-se a terceira, quarta e quinta camadas de convoluções com 384, 256 e 256 filtros respectivamente, todas de tamanho  $3 \times 3$  e preenchimento de 1, cada uma seguida por funções de ativação ReLU. A componente *Features* é então concluída com mais uma função *Max Pooling* com um tamanho de *kernel* de  $3 \times 3$  e um passo de 2. Após a componente *Features* tem-se uma camada

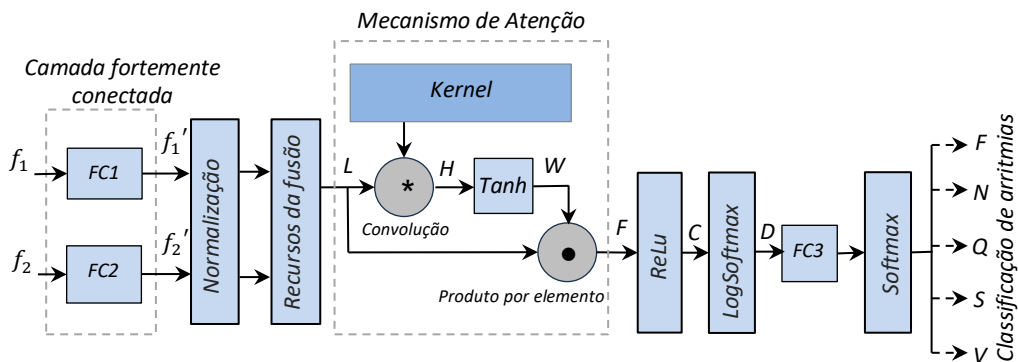
*AdaptiveAvgpool2D* que redimensiona a saída de *Features* de 256 canais numa operação de média adaptativa para criar uma nova representação compacta dos dados numa matriz  $6 \times 6$  com profundidade também de 256 canais.

Antes de ser entregue ao componente Classificador com módulo de atenção, a saída de cada sequência de camadas convolucionais de dimensão  $256 \times 6 \times 6$  passa por uma *flatten layer* para conversão para os vetores  $f_1$  e  $f_2$  de dimensão  $1 \times 9216$ .

#### 4.5 CLASSIFICADOR COM MECANISMO DE ATENÇÃO

O componente classificador com mecanismo de atenção da estrutura CNN-AM tem sua estrutura apresentada na Figura 28, sendo este componente responsável por fornecer a classificação final da estrutura CNN-AM. O componente tem como entrada os vetores de características  $f_1$  e  $f_2$  provenientes do componente de camadas convolucionais multimodais, que são entregues para duas sequências de camadas totalmente conectadas FC1 e FC2. Cada FC é composta por uma camada com 4096 neurônios, com função de ativação ReLu, seguida de *Dropout* de 0,5 para regularização e uma camada com 1024 neurônios com função de ativação ReLu. As duas sequências de camadas totalmente conectadas mapeiam  $f_1$  e  $f_2$ , de dimensão  $1 \times 9216$ , para um novo espaço de características reduzido,  $f_1'$  e  $f_2'$ , com dimensão  $1 \times 1024$ . Os números de neurônios das camadas totalmente conectadas foram definidos durante fase de experimentos a partir de treinamentos fim-a-fim da estrutura CNN-AM buscando sempre o melhor desempenho nas métricas de avaliação do classificador.

**Figura 28:** Componente classificador com mecanismo de atenção.



Fonte: Próprio Autor.

Os vetores de características  $f_1'$  e  $f_2'$  são normalizados no intervalo  $[-1, +1]$  para redução de gradientes instáveis na retropropagação, para então serem fundidos conforme:

$$L = Avg(f_1', f_2') \quad (9)$$

onde  $Avg$  representa a média dos vetores.

O vetor  $L$  de dimensão  $1 \times 1024$  resultante é uma representação global das características extraídas através das sequências de camadas convolucionais e totalmente conectadas, sendo este vetor a entrada para o mecanismo de atenção, que tem a função de indicar a importância relativa de cada elemento do vetor  $L$  a partir da determinação de um vetor pesos.

O mecanismo de atenção aqui proposto, conforme visto na Figura 28, foi inspirado nos trabalhos de Ahmad e Khan (2021) e Ahmad *et al.* (2021) onde os autores apresentam a técnica *Multistage Gated Average Fusion* (MGAF) que utiliza um mecanismo de atenção baseado na convolução do vetor de característica com um *kernel* fixo (*high boost filter*) sendo esta estratégia similar à usada no trabalho de Tang *et al.* (2023).

No mecanismo de atenção aqui proposto, um vetor  $H$  é obtido através da convolução do vetor  $L$  com um *kernel* de dimensão  $1 \times 3$  conforme:

$$H[i]_{i=0}^{1023} = L * kernel = \sum_{k=1}^3 L[i + k - 2].kernel[k] \quad (10)$$

onde o índice  $i$  varia de 0 até 1023 para percorrer as 1024 posições do vetor  $L$ . As dimensões de  $L$  *kernel* são  $1 \times 1024$  e  $1 \times 3$ , respectivamente. Portanto, a convolução é realizada com um somatório de  $L$  nos extremos, onde  $i + k - 2$  representa o valor do vetor de entrada deslocado pelo índice  $k$  do *kernel*. A convolução pondera as informações espaciais do vetor  $L$ , capturando relações entre elementos adjacentes, sendo que o *kernel* tem seus valores ajustados durante o processo de aprendizado fim-a-fim da CNN-AM. Existem duas formas de implementação desta convolução, visto que a sua implementação direta reduz a dimensão de  $H$  em duas posições, em virtude do deslocamento do *kernel*:

1. Pode-se na implementação da convolução, aumentar  $H$  após a convolução de 1022 para 1024 posições diretamente com a operação *padding*, preenchendo com zeros ou por reflexão, no início e no final de  $H$ ;
2. Pode-se na implementação da convolução aumentar o vetor antes da convolução  $L$  de 1024 para 1026 posições com a operação *padding*, preenchendo com zeros ou por reflexão, no início e no final de  $L$ , para

que ao final da convolução, já se obtenha um vetor de dimensão  $1 \times 1024$ .

As duas estratégias foram testadas, e como será visto posteriormente, a primeira implementação apresentou melhores resultados no paradigma interpaciente, e a segunda implementação foi mais eficaz para o paradigma intrapaciente.

O vetor de pesos de atenção  $W$  é obtido então aplicando uma função tangente hiperbólica ao vetor  $H$  conforme:

$$W = \text{Tanh}(H) \quad (11)$$

E o vetor de saída do mecanismo de atenção é obtido conforme:

$$F = W \odot L \quad (12)$$

sendo  $\odot$  a operação de *element-wise product* e  $F$  um vetor de dimensão  $1 \times 1024$  que apresenta as características globais com suas importâncias relativas indicadas a partir do vetor de pesos  $W$ .

O vetor  $F$ , saída do mecanismo de atenção, passa então por uma função ReLU, para introduzir não-linearidade no modelo, onde é gerado o vetor  $C$  que passa por uma função *LogSoftmax*, que empiricamente, após muitos testes durante fase de experimentos, foi observado que a introdução desta função gera melhorias nos indicadores de classificação:

$$D_i = \text{LogSoftmax}(C_i) = \log \left( \frac{e^{C_i}}{\sum_{j=1}^{1024} e^{C_j}} \right) \quad (13)$$

onde  $i$  indica o  $i$ -ésimo elemento do vetor  $C$ , sendo que para cada elemento  $C_i$  é calculado o logaritmo de seu exponencial dividido pelo somatório do exponencial de todos os elementos de  $C$ , o que transforma os valores originais em uma distribuição de probabilidade logarítmica.

Por fim, o vetor resultante  $D$  de características, de dimensão  $1 \times 1024$ , é apresentado na entrada da camada FC3 composta pelos neurônios de saída da estrutura, sendo que cada neurônio tem a função *Softmax* como função de ativação para gerar a probabilidade de cada uma das classes possíveis na saída da CNN-AM.

## 4.6 CONSIDERAÇÕES FINAIS

Este Capítulo abordou os elementos construtivos da CNN-AM proposta nesta tese

assim como foi apresentado a organização dos conjuntos de dados usados para os experimentos conduzidos na CNN-AM para cada abordagem de classificação, seja intrapaciente ou interpaciente. Também foram apresentadas as técnicas de aumento de dados usadas que contribuíram para os melhores resultados obtidos para a estrutura proposta. O Capítulo também apresenta as técnicas utilizadas para converter sinais temporais em imagens e as métricas quantitativas utilizadas para avaliar os resultados durante fase de experimentos.

Com os procedimentos metodológicos definidos, o próximo Capítulo apresenta os resultados alcançados e as análises dos resultados obtidos. São apresentados os recursos de *hardware* e *software* utilizados e resultados alcançados em diferentes experimentos.

## 5 RESULTADOS EXPERIMENTAIS

Neste Capítulo serão apresentados os resultados obtidos através dos experimentos realizados na CNN-AM proposta considerando os paradigmas intrapaciente e interpaciente.

As implementações realizadas para obtenção dos resultados obtidos da CNN-AM foram realizadas em Python 3.9.15 utilizando, principalmente, as bibliotecas Tensorflow 2.11, PyTorch 1.12 e CUDA 11.6. O hardware utilizado foi um processador Intel Core i7 12<sup>a</sup> geração 14 núcleos 4,7 GHz com RAM 32 GB 4.800 MHz e GPU NVIDIA GeForce RTX 3070 Ti 8GB GDDR6, e também o ambiente Google Colab Pro+, que no momento da condução dos experimentos apresentava configuração de processador Intel Xeon 12 núcleos 2,2 GHz com RAM 83 GB e GPU NVIDIA A100 40 GB.

A consolidação da estrutura final da CNN-AM proposta ocorreu a partir de diversos experimentos. De forma resumida, segue o caminho trilhado para definição e consolidação da topologia de CNN multimodal com módulo de atenção adaptativo proposta neste estudo:

- a) Exploração de modelos de classificação com CNN consolidadas (AlexNet e ResNet, principalmente), avaliando o desempenho do ambiente computacional de *hardware* e *software*, considerando as técnicas para transformação de sinais ECG em imagens (GAF, RP, HSFC, MTF e HM) em separado e em conjunto e avaliando as técnicas de aumento de dados (SMOTE, WGAN-GP, ACGAN e PGAN) em diversas situações;
- b) Avaliação da técnica de fusão de imagens, unindo diferentes representações numa imagem única em cada canal, como com GAF, RP e HSFC consolidando uma imagem com 3 canais, entre outras combinações, ou GAF, RP, HSFC e MTF consolidando uma imagem com 4 canais. Neste caso, cada representação foi gerada em cores e escala de cinza, bem como exploradas técnicas de armazenamento de arquivos nos formatos JPG, PNG e HDF5;
- c) Exploração de técnica de classificação complementar à CNN, com uso de Máquina de Vetores de Suporte (SVM) e otimização de hiper parâmetros com uso de enxame de partículas (PSO);

- d) Exploração de modelos de CNN multimodal com módulo de atenção fixo e adaptativo.

Durante a fase de treinamentos e experimentos para consolidação da CNN-AM proposta, a função custo minimizada utilizada foi baseada na *Hinge Loss*, introduzida por Cortes e Vapnik (1995) e adaptada por Crammer e Singer (2001) para o caso de classificação multiclasses conforme:

$$\text{Hinge Loss} = \max(0, 1 - \hat{y}_y + \max_{i \neq y}(\hat{y}_i)) \quad (14)$$

onde  $y \in 0,1,2$  é a saída desejada para o caso de 3 classes,  $\hat{y} \in \mathbb{R}^3$  é a saída prevista por classe e  $\hat{y}_y$  se refere aos elementos que são da classe  $y$  onde  $\max_{i \neq y}(\hat{y}_i)$  se refere ao valor máximo dos elementos de  $\hat{y}$  que não são da classe  $y$ .

A *Hinge Loss* é comumente usada em problemas de classificação binária, especialmente em contextos de aprendizado de máquina linear, como em máquinas de vetores de suporte (SVMs). A ideia principal é penalizar as previsões incorretas do modelo, onde a penalização aumenta à medida que a distância entre a previsão correta e a previsão do modelo aumenta. A perda é zero quando a previsão do modelo está correta e positiva quando a previsão do modelo está incorreta. Como neste trabalho a classificação é multiclasse, foi utilizada a função *MulticlassHingeLoss* da biblioteca *TorchMetrics* que é a função *Hinge Loss* adaptada para o caso de múltiplas classes.

O otimizador utilizado foi o Adam (KINGMA; BA, 2015), com taxa de aprendizado inicial de  $1e-4$  para as bases em ECG... (2018),  $4e-6$  nos demais casos, blocos de aprendizado (*batch size*) com 128 imagens, número máximo de 300 épocas e decaimento de taxa de aprendizado de 50% a cada 10 épocas de treinamento sem redução do erro.

As seções que seguem focam na apresentação apenas dos melhores resultados obtidos para abordagem interpaciente e intrapaciente com a estrutura da CNN-AM consolidada após os diversos experimentos exploratórios. Resultados de experimentos com outras configurações testadas durante fase exploratória para se chegar a CNN-AM consolidada serão apresentados apenas na seção de discussão dos resultados para comparações de desempenhos.

## 5.1 RESULTADOS COM O PARADIGMA INTRAPACIENTE

Como já apresentado na seção 4.1.2, para os experimentos no paradigma intrapaciente, foram utilizados nos experimentos três conjuntos de bases de dados, como apresentados nos Quadros 11, 12 e 13, que são indicados como Base 1, Base 2 e Base 3 respectivamente. Para determinar o melhor modelo, foi observado durante o treinamento o indicador de *F1-Score* geral ao classificar os dados de teste, guardando-se o modelo com maior valor deste indicador. O *F1-Score* é uma métrica especialmente interessante para avaliar o desempenho de modelos em bases de dados fortemente desbalanceadas como as de sinais ECG aqui utilizadas, pois considera tanto a precisão quanto o *recall* em um único valor harmônico, permitindo um equilíbrio entre as taxas de verdadeiros positivos, falsos positivos e falsos negativos. Em bases desbalanceadas, um modelo pode obter uma alta acurácia apenas prevendo consistentemente a classe majoritária, mesmo que falhe em identificar corretamente as instâncias da classe minoritária. O *F1-Score* evita esse problema, pois penaliza falsos negativos (que impactam o *recall*) e falsos positivos (que impactam a precisão).

### 5.1.1 Experimentos com CNN-AM com Base 1 – MIT-BIH

A Base 1 apresenta apenas dados obtidos dos registros da derivação MLII, sendo que os dados estão divididos em dados de treino e dados de teste na proporção 80/20, tendo sido aplicado o aumento de dados que resultou na base de dados apresentada no Quadro 14. Os Quadros 22 e 23 apresentam, respectivamente, os indicadores globais e por classe para a CNN-AM obtidos para o caso da base de teste da Base 1 (com aumento de dado) após treino da estrutura com a base de treino da Base 1.

**Quadro 22:** Indicadores globais para CNN-AM, com paradigma intrapaciente e base de teste da Base 1

Indicadores	Acc	Prec	Recall	Spe	F1
	98,91	95,11	92,40	99,18	93,71

Fonte: Próprio Autor.

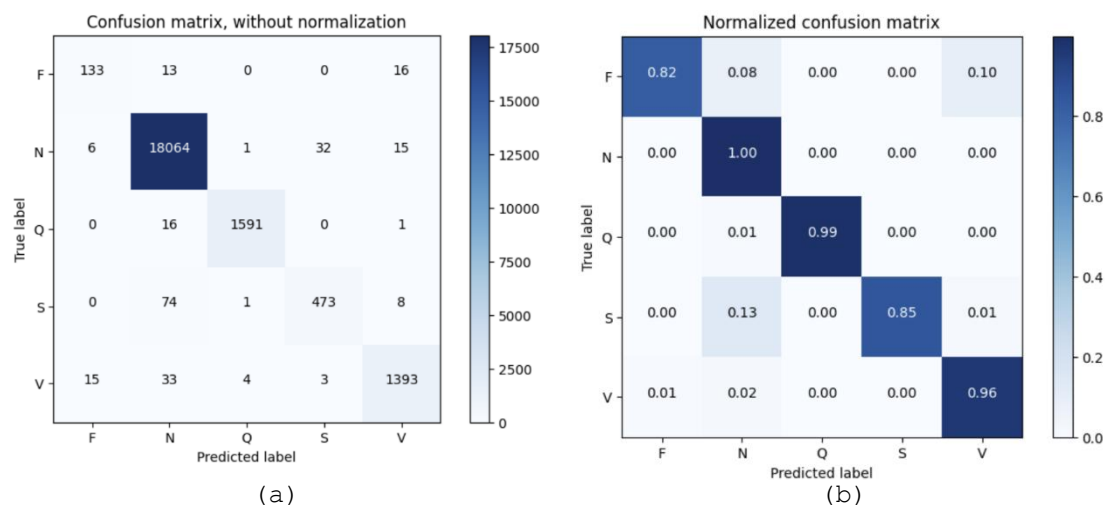
**Quadro 23:** Indicadores por classe para CNN-AM, com paradigma intrapaciente e base de teste da Base 1

Indicadores.	F					N				Q				S				V			
	AccT	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	f1	Recall	Prec	Spe	F1
	98,91	82,10	86,36	82,10	84,18	99,70	99,25	96,40	99,48	98,94	99,62	99,97	99,28	85,07	93,11	99,84	88,91	96,20	97,21	99,80	96,70

Fonte: Próprio Autor.

A Figura 29 apresenta a matriz de confusão considerando a base de teste da Base 1.

**Figura 29:** Matriz de confusão para base de teste da Base 1 MIT-BIH.



Fonte: Próprio Autor.

### 5.1.2 Experimentos com CNN-AM com Base 2 – MIT-BIH

O quadro 24 e 25 apresentam, respectivamente, os indicadores gerais e por classe obtidos para a CNN-AM para os dados da base de teste da Base 2, sendo que para este caso temos resultados do experimento considerando somente a derivação MLII e resultados considerando a derivação MLII+V, com a Base 2 com aumento de dados como apresentado no quadro 16. A estrutura CNN-AM foi treinada com dados de treino da Base 2 (caso MLII e caso MLII+V)

**Quadro 24:** Indicadores globais para CNN-AM com paradigma intrapaciente e base de teste da Base 2 MIT-BIH.

Indicadores	Acc	Prec	Recall	Spe	F1
MLII	99,70	97,96	97,21	99,29	97,58
MLII+V	99,72	98,19	97,26	99,34	97,72

Fonte: Próprio Autor.

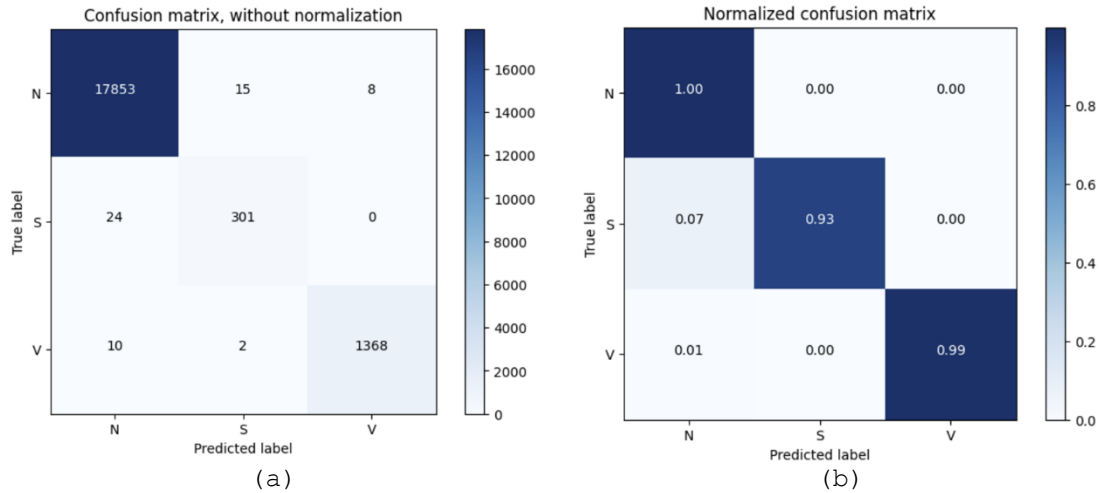
**Quadro 25:** Indicadores por classe para CNN-AM com paradigma intrapaciente e base de teste da Base 2 MIT-BIH.

Indicadores	AccT	N				S				V			
		Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1
MLII	99,70	99,87	99,81	98,01	99,84	92,62	94,65	99,91	93,62	99,13	99,42	99,96	99,27
MLII+V	99,72	99,89	99,82	98,12	99,85	92,62	95,25	99,92	93,92	99,28	99,49	99,96	99,38

Fonte: Próprio Autor.

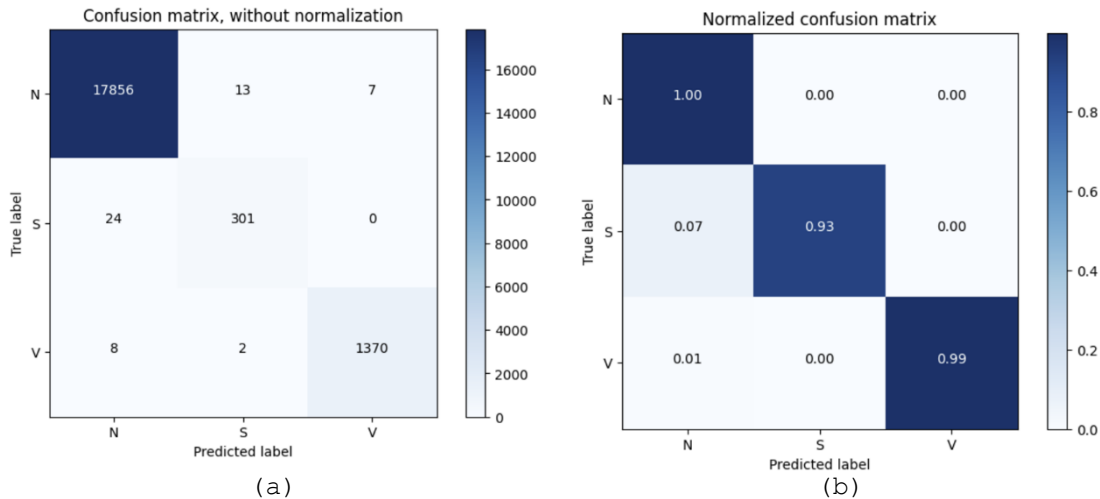
As matrizes de confusão para o caso de uma derivação e duas derivações são apresentadas nas Figuras 30 e 31, respectivamente.

**Figura 30:** Matriz de confusão para caso de uma derivação (MLII) e base de teste da Base 2 MIT-BIH.



Fonte: Próprio Autor.

**Figura 31:** Matriz de confusão para caso de duas derivações (MLII+V) e base de teste da Base 2 MIT-BIH.



Fonte: Próprio Autor.

### 5.1.3 Experimentos com CNN-AM com Base 3 – PTB

Também foram conduzidos experimentos com a base PTB, disponível em ECG... (2018). Nesta base, o aumento de dados não trouxe melhorias de resultados nos indicadores de teste e, portanto, não serão apresentados resultados para este caso. Trabalhos correlatos também não apontam benefícios para estes dados com esta

estratégia. Para este caso temos a classificação apenas de 2 classes, Normal e Anormal, em que a classe Normal possui o rótulo de indivíduos saudáveis e a classe Anormal refere-se ao rótulo de infarto do miocárdio, sendo esta a classe majoritária. Artigos correlatos também usam estas mesmas classes, permitindo melhor comparação dos resultados.

Os Quadros 26 e 27 apresentam os indicadores globais e por classe para a base de teste da base PTB.

**Quadro 26:** Indicadores globais para CNN-AM para base de teste da Base 3 PTB.

<b>Indicadores</b>	<b>Acc</b>	<b>Prec</b>	<b>Recall</b>	<b>Spe</b>	<b>F1</b>
	99,73	99,73	99,58	99,58	99,66

**Fonte:** Próprio Autor.

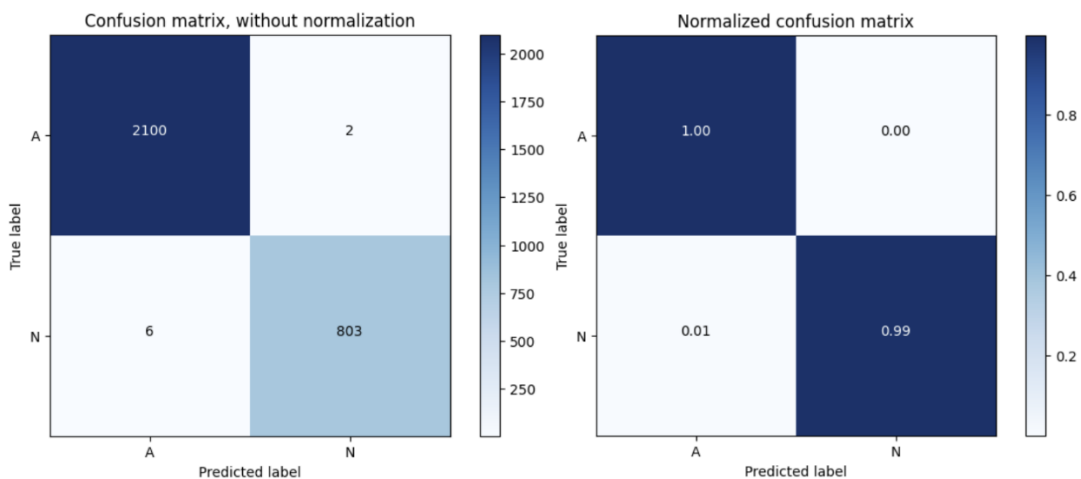
**Quadro 27:** Indicadores por classe para CNN-AM para base de teste da Base 3 PTB.

<b>Indicadores</b>	<b>A</b>					<b>N</b>			
	<b>AccT</b>	<b>Recall</b>	<b>Prec</b>	<b>Spe</b>	<b>F1</b>	<b>Recall</b>	<b>Prec</b>	<b>Spe</b>	<b>F1</b>
	99,73	99,90	99,72	99,26	99,81	99,26	99,75	99,90	99,50

**Fonte:** Próprio Autor.

A Figura 32 apresenta a matriz de confusão para base de teste da Base 3 PTB.

**Figura 32:** Matriz de confusão considerando base de teste da Base 3 PTB.



**Fonte:** Próprio Autor.

#### 5.1.4 Considerações sobre a abordagem intrapaciente

Considerações sobre experimentos para abordagem intrapaciente:

1. O aumento de dados com SMOTE em classes mais fáceis de classificar, como N, Q e V, e agregar dados sintéticos gerados por WGAN-GP em classes mais difíceis, como F e S, melhoraram as métricas na base de teste;
2. O aumento vetorial reflexivo na convolução do módulo de atenção aumentou significativamente as métricas na base de teste em relação à convolução direta com redimensionamento vetorial, tanto para as bases 1 e 2 MIT-BIH quanto para a base 3 PTB;
3. Arquivar as imagens no formato PNG na resolução  $224 \times 224$  gerou melhores resultados e de forma computacionalmente mais eficiente em relação aos formatos JGP e HDF5.
4. A abordagem multimodal com duas derivações (MLII e V) levou ao aumento de todas as métricas de teste do classificador, embora a dificuldade na classificação das classes F e S continuem bem aparentes.

## 5.2 RESULTADOS COM O PARADIGMA INTERPACIENTE

Como detalhado na seção 4.1.1, a abordagem interpaciente foi aplicada diretamente na base MIT-BIH disponibilizada por Moody e Mark (2005), devidamente segmentadas e filtradas utilizando Emrich *et al.* (2023), e incluindo o aumento de dados apresentado na seção 4.2, Quadro 15.

Da mesma forma como feito na abordagem intrapaciente, para definição do melhor modelo classificador foi utilizada a métrica de *F1-Score* geral considerando a base de teste. No entanto, visando comparações com trabalhos correlatos, também foram guardados os modelos que apresentaram maior sensibilidade para a classe S, por ser a mais difícil de classificar.

Os Quadros 28 e 29 apresentam os indicadores gerais e por classe para os melhores resultados obtidos para cada um dos seguintes casos de experimento:

- 1) CNN-AM com derivação MLII e escolha do melhor resultado na base de teste considerando *F1-Score*;
- 2) CNN-AM com derivação MLII e escolha do melhor resultado na base de teste considerando sensibilidade para a classe S;
- 3) CNN-AM com derivações MLII e V, e escolha do melhor resultado na base de teste considerando *F1-Score*;
- 4) CNN-AM com derivações MLII e V, e escolha do melhor resultado na

base de teste considerando sensibilidade para a classe S;

**Quadro 28:** Indicadores globais para o paradigma interpaciente considerando a base de teste.

	Acc	Prec	Recall	Spe	F1
MLII-F1	97,29	74,41	78,33	95,42	75,96
MLII-S	83,89	59,51	84,86	93,24	61,86
(MLII+V)-F1	98,48	94,15	80,23	96,34	81,91
(MLII+V)-S	95,47	72,28	87,21	96,62	75,25

Fonte: Próprio Autor.

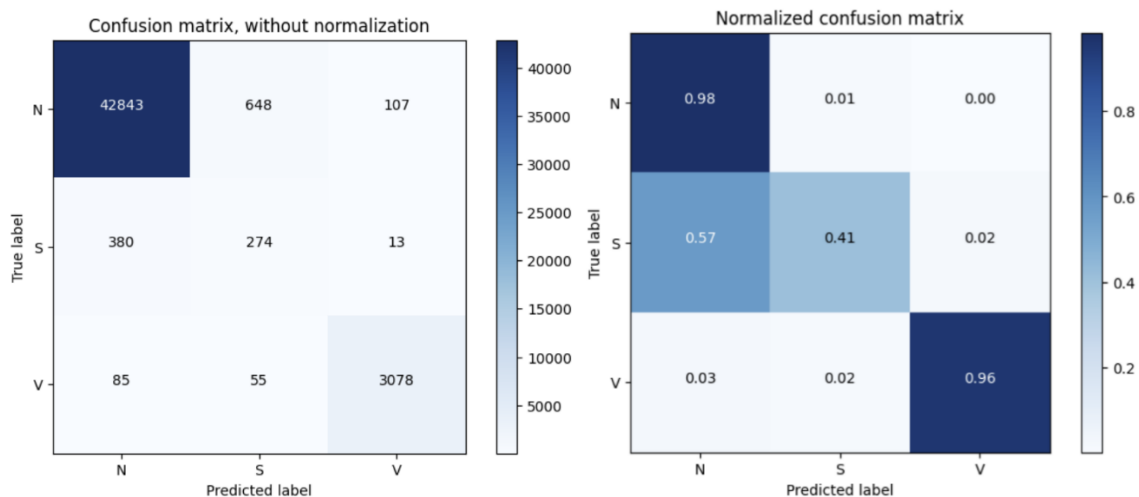
**Quadro 29:** Indicadores por classe para o paradigma interpaciente considerando a base de teste.

	AccT	N				S				V			
		Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1
MLII-F1	97,29	98,27	98,93	88,03	98,60	41,08	28,05	98,50	33,33	95,65	96,25	99,73	95,95
MLII-S	83,89	83,33	99,56	95,86	90,68	77,51	7,61	86,59	13,86	93,82	71,35	97,26	81,06
(MLII+V)-F1	98,48	99,36	99,09	89,73	99,22	43,33	56,34	99,52	48,98	98,01	97,02	99,78	97,51
(MLII+V)-S	95,47	95,79	99,44	93,95	97,58	69,27	20,21	96,10	31,29	96,58	97,19	99,80	96,88

Fonte: Próprio Autor.

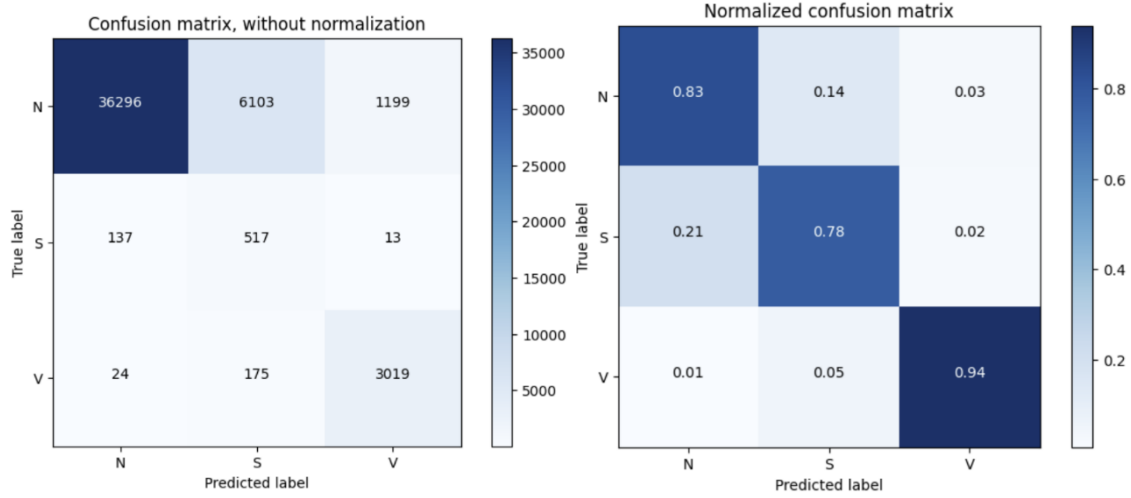
As matrizes de confusão dos experimentos são apresentadas nas Figuras 33 a 36 considerando base de teste.

**Figura 33:** Matriz de confusão do experimento 1, com melhor resultado de *F1-Score* interpaciente com uma derivação com a base MIT-BIH.



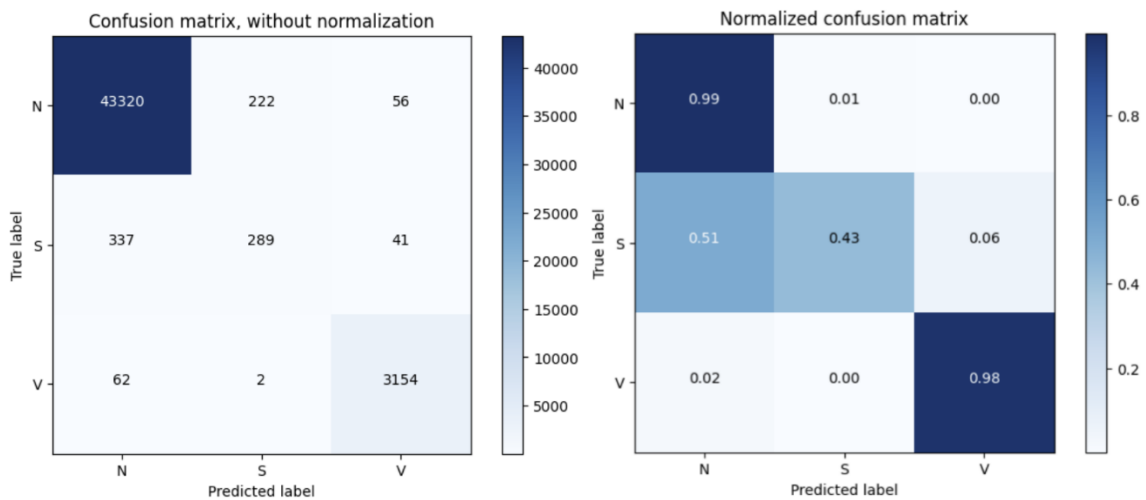
Fonte: Próprio Autor.

**Figura 34:** Matriz de confusão do experimento 2, com melhor resultado de sensibilidade de S interpaciente com uma derivação com a base MIT-BIH.



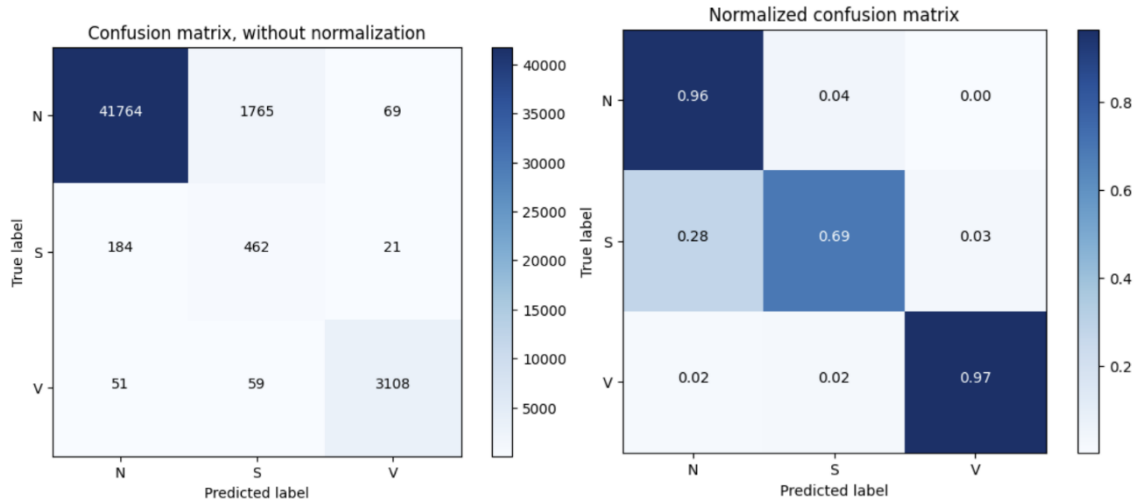
Fonte: Próprio Autor.

**Figura 35:** Matriz de confusão do experimento 3, com melhor resultado de *F1-Score* interpaciente com duas derivações com a base MIT-BIH.



Fonte: Próprio Autor.

**Figura 36:** Matriz de confusão do experimento 4, com melhor resultado de sensibilidade de S interpaciente com duas derivações com a base MIT-BIH.



Fonte: Próprio Autor.

### 5.2.1 Considerações sobre a abordagem interpaciente

Considerações sobre experimentos para abordagem interpaciente:

- O uso de duas derivações levou ao aumento de todas as métricas de teste do modelo classificador.
- Melhores resultados foram obtidos usando as técnicas SMOTE e WGAN-GP para aumento de dados, com WGAN-GP sendo usada apenas na classe S, tanto para a base proveniente da derivação MLII quanto para a base proveniente da derivação V unificadas.
- Melhores resultados foram obtidos com o ajuste no módulo de atenção para aumento vetorial após a convolução com o *kernel*, diferentemente do que ocorreu com a abordagem intrapaciente.

## 5.3 CONSIDERAÇÕES FINAIS

Este Capítulo apresentou resultados experimentais alcançados, informando a estrutura de *hardware* e *software* utilizada e os detalhes de implementação de treino e teste, incluindo as métricas de avaliação de todas as bases utilizadas neste estudo, dividindo na abordagem intra e interpaciente. O próximo Capítulo apresenta a análise e discussão destes resultados alcançados para cada abordagem e comparando-os, apresentando comparação com diferentes modelos implementados na fase exploratório

deste estudo, e também dos melhores resultados aqui alcançados em cada paradigma com resultados apresentados na literatura atualizada, de forma a deixar claro o alcance do estado da arte na pesquisa aqui conduzida.

## 6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste Capítulo serão apresentadas as análises e discussões sobre os resultados alcançados da CNN-AM proposta, sendo apresentado posteriormente um estudo de ablação para avaliar a importância de diferentes componentes da CNN-AM consolidada. A seguir será apresentado um estudo comparativo de resultados da CNN-AM com resultados obtidos durante toda fase exploratória, finalizando com um estudo comparativo com trabalhos já apresentados na literatura contemporânea.

### 6.1 ANÁLISE GERAL DOS RESULTADOS

#### 6.1.1 *Abordagem Interpaciente*

Considerando os resultados apresentados nos Quadros 28 e 29 para derivação MLII, focando no classificador baseado no *F1-Score* (MLII-f1), o modelo atingiu 97,29% de acurácia geral e 75,96% de *F1-Score* sendo que o modelo apresentou uma forte dificuldade em classificar a classe S, com 41,08% de sensibilidade e 28,05% de precisão, mas atingindo 98,27% e 98,93% de acurácia e precisão para a classe N, e 95,65% e 96,25% para classe V. A dificuldade de aprender a classe S pode vir do fato que a base de dados para classe V inclui vários subtipos de arritmias supraventricular, com características distintas, o que pode tornar mais difícil o aprendizado e classificação, considerando também o número reduzido de padrões desta classe para treino (MORADY, 2004).

Considerando a os resultados do classificador com foco na sensibilidade da classe S (MLII-S), tem-se o melhor modelo obtido com 77,51% de sensibilidade a uma precisão de 7,61% para classe S, totalizando 83,89% e 61,86% de acurácia e *F1-Score* gerais.

Considerando os resultados dos classificadores com dados das derivações MLII e V apresentados nas Tabelas 24 e 28, com os indicadores gerais intra e interpaciente, para o classificador com foco na métrica *F1-Score*, observa-se claramente um aumento de todas as métricas gerais em relação aos resultados sem derivação V, com um aumento de 1,19%, 19,74%, 1,90%, 0,92% e 5,95% para as métricas de acurácia, precisão, sensibilidade/recall, especificidade e *F1-Score* respectivamente. Entretanto, apesar do aumento nas métricas para a classe S, ainda continuamos com dificuldade para aprendizado desta classe devido os mesmos motivos já indicados para o classificador

somente com derivação MLII.

Considerando agora o classificador com foco na sensibilidade da classe S, observou-se um aumento de 12,60% da precisão desta classe em relação ao uso de apenas uma derivação, mas com redução de sua sensibilidade para 69,27%. No entanto, a diferença do *F1-Score* entre estes dois cenários também é significativa, com um aumento de 13,39% para este indicador, em especial devido ao aumento da acurácia e especificidade das classes N e V com a abordagem com duas derivações. Além disso, a maior sensibilidade obtida para a classe S nestes experimentos foi de 77,51% com uma precisão de 7,61%, totalizando 83,89% e 61,86% de acurácia e *F1-Score* gerais. Ou seja, o aumento da sensibilidade à classe S também reduziu sua precisão, e também reduziu a sensibilidade para a classe N.

### 6.1.2 Abordagem Intrapaciente

O desempenho do modelo proposto para abordagem intrapaciente apresentou, como observado nos Quadros 24 a 29, um resultado que pode ser considerado promissor para este tipo de abordagem nas 3 bases avaliadas.

A Base 1 MIT-BIH já é disponibilizada dividida em conjuntos de treino e teste, segmentada com destaque para um pico R e pré-processada, possuindo 5 classes severamente desbalanceadas. O uso de WGAN-GP no aumento de dados das classes F e S, mais difíceis de classificar, contribuiu para maior sensibilidade a elas e melhoria dos indicadores globais.

A Base 2 também MIT-BIH foi construída neste estudo a partir da organização dos dados interpaciente, unindo DS1 e DS2, e posteriormente dividindo de forma aleatória para formar os conjuntos de treino e teste, com destaque a um intervalo RR em cada segmento, além de ser realizado experimento com uma e duas derivações: MLII e V consolidada. Embora com menos classes, o modelo classificador atingiu bons resultados, em especial para duas derivações, atingindo acurácia geral de 99,72%, precisão de 98,19% e *F1-Score* de 97,72%. A dificuldade em classificar a classe S é menos observada nesta abordagem devido ao uso de dados para teste dos mesmos pacientes usados para treino, atingindo 92,62% de sensibilidade para a classe S nos dois casos, mas com maior precisão para duas derivações com 95,25% em relação a 94,65% com uma derivação.

Ainda sobre a Base 2, a diferença entre os classificadores com uma e duas

derivações também pode ser observada, com aumento do classificador com duas derivações de 0,02%, 0,23%, 0,05%, 0,05% e 0,14% para as métricas gerais de acurácia, precisão, sensibilidade/*recall*, especificidade e *F1-Score* respectivamente.

Para a Base 3 PTB, com 2 classes e 1 derivação, os resultados também foram expressivos, superando 99% em todos os indicadores gerais e por classe avaliados.

## 6.2 ESTUDO DE ABLAÇÃO

Para avaliar as contribuições de determinados componentes individuais dentro da estrutura da CNN-AM proposta, foi realizado um estudo onde novos classificadores foram treinados retirando da CNN-AM determinados componentes. Os Quadros de 30 a 34 apresentam, respectivamente, os resultados obtidos das métricas globais e por classe para abordagem interpaciente e intrapaciente de cada novo modelo desenvolvido, assim como os resultados da CNN-AM proposta para comparação, considerando métricas obtidas na base de teste MIT-BIH nas derivações MLII e MLII +V, e para a base PTB na derivação DII, sendo os modelos a serem analisados:

- M1: AlexNet com transformação de sinal ECG para imagem com RP;
- M2: AlexNet com transformação de sinal ECG para imagem com HSFC;
- M3: CNN-AM retirando o módulo de atenção e para uma derivação MLII/DII;
- **M4: CNN-AM com o módulo de atenção e para uma derivação MLII/DII;**
- M5: CNN-AM retirando o módulo de atenção e para duas derivações MLII + V;
- **M6: CNN-AM com o módulo de atenção e para duas derivações MLII + V.**

**Quadro 30:** Indicadores globais interpaciente MIT-BIH.

		<b>Acc</b>	<b>Prec</b>	<b>Recall</b>	<b>Spe</b>	<b>F1</b>
<b>M1</b>	MLII	96,74	73,02	77,54	94,64	74,40
<b>M2</b>	MLII	97,14	73,42	76,28	94,92	74,69
<b>M3</b>	MLII-f1	97,32	74,72	78,39	95,37	76,02
	MLII-S	85,32	62,39	83,98	93,43	64,01
<b>M4</b>	MLII-f1	97,29	74,41	78,33	95,42	75,96
	MLII-S	83,89	59,51	84,86	93,24	61,86
<b>M5</b>	MLII+V-f1	98,15	81,17	80,18	96,30	80,56
	MLII+V-S	92,79	63,84	69,17	91,90	62,70
<b>M6</b>	<b>MLII+V-f1</b>	<b>98,48</b>	<b>94,15</b>	<b>80,23</b>	<b>96,34</b>	<b>81,91</b>
	MLII+V-S	95,47	72,28	87,21	96,62	75,25

**Fonte:** Próprio Autor.

**Quadro 31:** Indicadores por classe interpaciente MIT-BIH.

		N				S				V			
		Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1
<b>M1</b>	MLII	97,9	98,8	86,1	98,3	42,0	23,5	98,1	30,1	92,8	96,8	99,8	94,7
<b>M2</b>	MLII	98,3	98,8	86,6	98,5	36,3	28,5	98,7	31,9	94,3	93,0	99,5	93,6
<b>M3</b>	MLII-f1	98,3	98,9	97,8	98,6	40,8	26,2	98,4	31,9	96,1	99,1	99,9	97,6
	MLII-S	84,8	99,5	94,9	91,6	72,1	7,4	87,1	13,4	95,1	80,3	98,3	87,0
<b>M4</b>	MLII-f1	98,3	98,9	88,0	98,6	41,1	28,0	98,5	33,3	95,6	96,2	99,7	95,9
	MLII-S	83,3	99,6	95,9	90,7	77,5	7,6	86,6	13,9	93,8	71,4	97,3	81,1
<b>M5</b>	MLII+V-f1	99,0	99,1	90,0	99,0	43,2	50,5	99,4	46,6	98,4	93,9	99,5	96,1
	MLII+V-S	95,3	98,3	81,5	97,0	48,4	12,8	95,3	20,2	63,5	80,5	98,9	71,0
<b>M6</b>	<b>MLII+V-f1</b>	<b>99,4</b>	<b>99,1</b>	<b>88,4</b>	<b>99,2</b>	<b>43,3</b>	<b>56,3</b>	<b>99,7</b>	<b>49,0</b>	<b>98,0</b>	<b>97,0</b>	<b>99,7</b>	<b>97,5</b>
	MLII+V-S	95,8	99,4	97,0	97,6	69,3	20,2	96,1	31,3	96,6	97,2	99,8	96,6

Fonte: Próprio Autor.

**Quadro 32:** Indicadores globais intrapaciente com a Base 2 MIT-BIH.

		Acc	Prec	Recall	Spe	F1
<b>M1</b>	MLII	98,81	94,37	91,85	99,06	93,03
<b>M2</b>	MLII	98,53	94,04	91,33	98,88	92,62
<b>M3</b>	MLII	99,58	97,02	95,89	99,02	96,44
<b>M4</b>	MLII	99,70	97,96	97,21	99,29	97,58
<b>M5</b>	MLII+V	99,73	98,21	97,26	99,34	97,73
<b>M6</b>	MLII+V	99,72	98,19	97,26	99,34	97,72

Fonte: Próprio Autor.

**Quadro 33:** Indicadores por classe intrapaciente com a Base 2 MIT-BIH.

		N				S				V			
		Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1
<b>M1</b>	RP	99,7	99,1	95,8	99,4	82,0	93,6	98,8	87,4	98,0	97,7	99,8	96,4
<b>M2</b>	HSFC	99,5	99,0	95,1	99,3	79,7	90,0	99,8	84,5	95,2	96,4	99,8	95,8
<b>M3</b>	MLII	99,8	99,7	97,2	99,8	88,9	92,0	99,9	90,5	98,9	99,3	99,9	99,1
<b>M4</b>	MLII	99,9	99,8	98,0	99,8	92,6	92,7	99,9	93,6	99,1	99,4	100	99,3
<b>M5</b>	MLII+V	99,9	99,8	98,1	99,9	92,6	95,3	99,9	93,9	99,3	99,6	100	99,4
<b>M6</b>	MLII+V	99,9	99,8	98,1	99,8	92,6	95,3	99,9	93,9	99,3	99,5	100	99,4

Fonte: Próprio Autor.

O Quadro 34 apresenta dados da base 3 PTB, que não disponibiliza derivações V. Portanto, os modelos M5 e M6 não se aplicam.

**Quadro 34:** Indicadores globais intrapaciente com a Base PTB.

	Acc	Prec	Recall	Spe	F1
<b>M1</b>	99,55	99,58	99,31	99,31	99,44
<b>M2</b>	98,97	98,83	98,60	98,60	98,71
<b>M3</b>	99,59	99,52	99,45	99,45	99,49
<b>M4</b>	<b>99,73</b>	<b>99,73</b>	<b>99,58</b>	<b>99,58</b>	<b>99,66</b>

Fonte: Próprio Autor.

Considerando os resultados para caso interpaciente e intrapaciente, é possível observar que, para o caso dos modelos M1 e M2, usando apenas a AlexNet com uma única imagem HSFC ou RP, temos resultados inferiores em relação os modelos M3 e M5, que utiliza abordagem multimodal com duas imagens, o que comprova a eficácia desta abordagem. Já os modelos M4 (MLII-f1) e M6 (MLII+V-f1), apresentaram resultados globais superiores, unindo a abordagem multimodal com módulo de atenção, apresentando melhorias em todas as métricas, indicando que a estrutura proposta, com os componentes de sequência de camadas convolucionais multimodal e mecanismo de atenção, podem trazer melhorias nos resultados do modelo baseado na AlexNet convencional, apesar do aumento de complexidade computacional.

Para o caso dos modelos em que foi retirado o mecanismo de atenção da estrutura da CNN-AM (modelos M3 e M5) podemos observar que para casos de abordagem interpaciente e derivações MLII + V os resultados alcançados pela CNN-AM foram um pouco superiores. Embora o modelo M3 para uma derivação tenha apresentado desempenho superior nos indicadores globais, observa-se que a sensibilidade para a classe S no modelo M4 com módulo de atenção aumentou, o que é importante para o caso interpaciente. Para o caso intrapaciente, o modelo M5 apresentou maiores indicadores globais, enquanto o modelo com módulo de atenção M4 foi superior no caso com uma derivação MLII.

Em todos os casos, a inclusão do mecanismo de atenção trouxe maior sensibilidade para a classe S, sendo este resultado um bom indicativo da importância do mecanismo de atenção na estrutura, visto que para abordagem interpaciente, que é a abordagem mais realista em relação a prática clínica, a classificação para este tipo de arritmias tem se mostrado desafiadora, com trabalhos sendo apresentados na literatura voltados especificamente para melhoria desta classe (ZHANG *et al.*, 2014; ZHAI, TIN, 2018; DIAS *et al.*, 2021; WANG *et al.*, 2021).

### 6.3 COMPARAÇÃO COM ESTRUTURAS DA FASE EXPLORATÓRIA

Para avaliar os resultados da estrutura da CNN-AM consolidada em relação as outras estruturas testadas durante a fase exploratória, esta seção apresenta resultados obtidos durante a fase de experimentos para comparação entre os melhores resultados obtidos com a CNN-consolidada e os melhores resultados obtidos com outras estruturas

testadas com variação da técnica de transformação de imagens, uso ou não de aumento de dados e aplicação ou não do módulo de atenção.

### 6.3.1 Explorações na abordagem intrapaciente

#### 6.3.1.1 Comparação com outras estruturas na Base 1 MIT-BIH

Para a Base 1, foi construído o Quadro 35 comparando a CNN-AM com estruturas dos seguintes experimentos:

- N1: AlexNet com transformação de sinal ECG para imagem com GAF e aumento de dados;
- N2: AlexNet com transformação de sinal ECG para imagem com RP e aumento de dados;
- N3: AlexNet com transformação de sinal ECG para imagem com HSFC e aumento de dados;
- N4: AlexNet com transformação de sinal ECG para imagem com MTF e aumento de dados;
- N5: AlexNet com transformação de sinal ECG para imagem com HM e aumento de dados;
- N6: ResNet18 com transformação de sinal ECG para imagem com GAF e aumento de dados;
- N7: ResNet18 com transformação de sinal ECG para imagem com RP e aumento de dados;
- N8: ResNet18 com transformação de sinal ECG para imagem com HSFC e aumento de dados;
- N9: ResNet18 com transformação de sinal ECG para imagem com MTF e aumento de dados;
- N10: ResNet18 com transformação de sinal ECG para imagem com HM e aumento de dados;
- N11: AlexNet com fusão numa única imagem contendo GAF, RP e HSFC;
- N12: CNN multimodal com imagens GAF, RP, HSFC e MTF, com extração de recursos e posterior classificação via SVM, com parâmetros otimizados via PSO;
- N13: CNN-AM com imagens GAF e RP sem módulo de atenção, e
- N14: CNN-AM com imagens GAF e RP com módulo de atenção adaptativo.
- **N15: CNN-AM com imagens RP e HSFC com módulo de atenção adaptativo (estrutura consolidada).**

**Quadro 35:** Comparação de indicadores globais para a base de teste entre diferentes estruturas para o paradigma intrapaciente com a Base 1.

	<b>Acc</b>	<b>Prec</b>	<b>Recall</b>	<b>Spe</b>	<b>F1</b>
<b>N1</b>	98,92	94,77	91,38	99,15	92,99
<b>N2</b>	98,81	94,37	91,85	99,06	93,03
<b>N3</b>	98,53	94,04	91,33	98,88	92,62
<b>N4</b>	97,99	90,55	88,87	98,49	89,61
<b>N5</b>	98,71	94,47	91,76	99,06	93,07
<b>N6</b>	98,85	95,11	91,70	99,10	93,31
<b>N7</b>	98,90	93,76	93,12	99,22	93,40
<b>N8</b>	98,32	95,10	87,46	98,59	90,85
<b>N9</b>	97,38	89,56	84,73	97,83	86,98
<b>N10</b>	98,87	94,01	92,07	99,16	93,02
<b>N11</b>	98,85	94,95	91,91	99,09	93,35
<b>N12</b>	98,70	94,84	91,44	98,89	93,02
<b>N13</b>	98,96	95,49	92,28	99,16	93,82
<b>N14</b>	<b>99,03</b>	<b>96,09</b>	<b>93,11</b>	<b>99,20</b>	<b>94,54</b>
<b>N15</b>	<b>98,91</b>	<b>95,11</b>	<b>92,40</b>	<b>99,18</b>	<b>93,71</b>

**Fonte:** Próprio Autor.

Observa-se que para este conjunto de dados, a abordagem multimodal em conjunto com o módulo de atenção melhorou todos os indicadores gerais. No entanto, os melhores indicadores foram obtidos com esta Base 1 foram observados com a CNN-AM consolidada usando imagens GAF e RP. Uma possível razão seria a influência do processamento do sinal temporal aplicado nesta base que destaca apenas um pico R, ou a influência das classes F e Q no treinamento. Embora maiores estudos sejam necessários para confirmar estas hipóteses, o objetivo desta tese está relacionado em apresentar uma estrutura de CNN-AM com melhores resultados em diferentes paradigmas, e devido ao melhor desempenho com HSFC e RP nos experimentos com o paradigma interpaciente, esta foi a estrutura consolidada.

### 6.3.1.2 Comparação com outras estruturas na Base 2 MIT-BIH

Para a Base 2, foi construído o Quadro 36 comparando a CNN-AM consolidada com estruturas dos seguintes experimentos:

- O1: AlexNet com transformação de sinal ECG da derivação MLII para imagem com GAF;
- O2: AlexNet com transformação de sinal ECG da derivação MLII para imagem com RP;
- O3: AlexNet com transformação de sinal ECG da derivação MLII para imagem com HSFC;
- O4: CNN-AM com o módulo de atenção e para uma derivação MLII, usando GAF e RP;
- O5: CNN-AM com o módulo de atenção e para duas derivações MLII + V, usando GAF e RP.
- **O6: CNN-AM com o módulo de atenção e para uma derivação MLII, usando**

**RP e HSFC (estrutura consolidada);**

- **O7: CNN-AM com o módulo de atenção e para duas derivações MLII + V, usando RP e HSFC (estrutura consolidada).**

**Quadro 36:** Comparação de indicadores globais para a base de teste da Base 2.

		Acc	Prec	Recall	Spe	F1
<b>O1</b>	MLII	98,92	94,77	91,38	99,15	92,99
<b>O2</b>	MLII	98,81	94,37	91,85	99,06	93,03
<b>O3</b>	MLII	98,53	94,04	91,33	98,88	92,62
<b>O4</b>	MLII	99,63	97,23	97,09	99,20	97,16
<b>O5</b>	MLII+V	99,67	97,09	97,70	99,37	97,39
<b>O6</b>	<b>MLII</b>	<b>99,70</b>	<b>97,96</b>	<b>97,21</b>	<b>99,29</b>	<b>97,58</b>
<b>O7</b>	<b>MLII+V</b>	<b>99,72</b>	<b>98,19</b>	<b>97,26</b>	<b>99,34</b>	<b>97,72</b>

Fonte: Próprio Autor.

Para este conjunto de dados, observa-se que todas as métricas nos cenários O6 e O7 com módulo de atenção foram as maiores obtidas. Diante disso, fazendo um detalhamento por classe destes cenários no Quadro 37, observa-se maior precisão, especificidade e *F1-Score* para a classe S, que é a mais difícil de classificar, quando se usa duas derivações.

**Quadro 37:** Comparação de indicadores por classe para a base de teste entre diferentes estruturas para o paradigma intrapaciente com a Base 2.

		N				S				V			
		Recall	Prec	Spe	F1	Recall	Prec	Spe	F1	Recall	Prec	Spe	F1
<b>O6</b>	MLII+V	99,87	99,81	98,01	99,84	92,62	94,65	99,91	93,62	99,13	99,42	99,96	99,27
<b>O7</b>	MLII+V	99,89	99,82	98,12	99,85	<b>92,62</b>	<b>95,25</b>	<b>99,92</b>	<b>93,92</b>	99,28	99,49	99,96	99,38

Fonte: Próprio Autor.

### 6.3.1.3 Comparação com outras estruturas na Base 3 - PTB

Para a Base 3, foi construído o Quadro 38 comparando os seguintes cenários:

- P1: AlexNet com transformação de sinal ECG para imagem com GAF;
- P2: AlexNet com transformação de sinal ECG para imagem com RP;
- P3: AlexNet com transformação de sinal ECG para imagem com HSFC;
- P4: AlexNet com transformação de sinal ECG para imagem com MTF;
- P5: AlexNet com transformação de sinal ECG para imagem com HM;
- P6: CNN-AM com o módulo de atenção adaptativo para imagens GAF e RP;
- **P7: CNN-AM com o módulo de atenção adaptativo para imagens RP e HSFC (estrutura consolidada);**

**Quadro 38:** Comparação de indicadores globais para a base de teste entre diferentes estruturas para o paradigma intrapaciente com a Base 3.

	<b>Acc</b>	<b>Prec</b>	<b>Recall</b>	<b>Spe</b>	<b>F1</b>
<b>P1</b>	99,45	99,39	99,24	99,24	99,31
<b>P2</b>	99,55	99,58	99,31	99,31	99,44
<b>P3</b>	98,97	98,83	98,60	98,60	98,71
<b>P4</b>	96,12	95,15	95,18	95,18	95,17
<b>P5</b>	99,11	99,04	98,74	98,74	98,89
<b>P6</b>	99,59	99,60	99,37	99,37	99,49
<b>P7</b>	<b>99,73</b>	<b>99,73</b>	<b>99,58</b>	<b>99,58</b>	<b>99,66</b>

**Fonte:** Próprio Autor.

Para este conjunto de dados, observa-se que a abordagem convencional de uma CNN com uma imagem se mostrou inferior à abordagem com duas imagens, e a estrutura proposta da CNN-AM com imagens RP e HSFC se mostrou superior em todos os indicadores avaliados.

### 6.3.2 Exploração na abordagem interpaciente

Para a base MIT-BIH, foi construído o Quadro 39 comparando os seguintes cenários ( outros cenários foram apresentados na ablação):

- Q1: AlexNet com derivação MLII e transformação de sinal ECG para imagem com RP;
- Q2: AlexNet com derivação MLII e transformação de sinal ECG para imagem com HSFC;
- **Q3: CNN-AM com o módulo de atenção e para uma derivação MLII com imagens RP e HSFC (estrutura consolidada);**
- **Q4: CNN-AM com o módulo de atenção e para duas derivações MLII + V com imagens RP e HSFC (estrutura consolidada).**

Para os cenários Q3 e Q4, em que foram utilizados a CNN-AM em abordagem multimodal, são apresentados os resultados das estruturas que obtiveram o maior *F1-Score* geral, e a estrutura com maior acurácia para a classe S, devidamente identificados, para fins de comparação.

**Quadro 39:** Comparação de indicadores globais para a base de teste entre diferentes estruturas para o paradigma interpaciente com a base MIT-BIH.

		<b>Acc</b>	<b>Prec</b>	<b>Recall</b>	<b>Spe</b>	<b>F1</b>
<b>Q1</b>	MLII	96,74	73,02	77,54	94,64	74,40
<b>Q2</b>	MLII	97,14	73,42	76,28	94,92	74,69
<b>Q3</b>	MLII-f1	97,29	74,41	78,33	95,42	75,96
	MLII-S	83,89	59,51	84,86	93,24	61,86
<b>Q4</b>	<b>MLII+V-f1</b>	<b>98,48</b>	<b>94,15</b>	<b>80,23</b>	<b>96,34</b>	<b>81,91</b>
	MLII+V-S	95,47	72,28	87,21	96,62	75,25

**Fonte:** Próprio Autor.

Observa-se que para este conjunto de dados a abordagem convencional com uma imagem se mostrou inferior à abordagem multimodal em conjunto com o mecanismo de atenção. Com uma imagem, o maior *F1-Score* observado foi de 74,69%, enquanto com a CNN-AM consolidada obteve-se *F1-Score* de 81,91%, um aumento de 7,22%. A abordagem multimodal com mecanismo de atenção trouxe melhorias em todos os indicadores, sendo o melhor resultado obtido para o paradigma interpaciente neste estudo.

#### 6.4 COMPARAÇÃO COM TRABALHOS CORRELATOS

No contexto da comparação de resultados deste estudo em relação ao estado da arte, uma busca por trabalhos correlatos que também tenham utilizado os dados do MIT-BIH foi realizada, considerando abordagem interpaciente e intrapaciente.

No que se refere aos indicadores de desempenho a serem utilizados para comparação, é preciso analisar cada paradigma em separado. Para o paradigma interpaciente a literatura não costuma usar apenas os indicadores globais para fins de comparação, sendo tão importante quanto a comparação também com os indicadores por classe. Nesse sentido, para abordagem interpaciente, foram elaborados os Quadros 40 e 41, sendo que todos os trabalhos analisados adotaram a mesma divisão DS1 e DS2 utilizada neste estudo, com algumas diferenças na metodologia de organização, o que não interfere para fins de comparação geral.

**Quadro 40:** Comparação com trabalhos anteriores para o paradigma interpaciente dos indicadores globais.

<b>Resultados Anteriores</b>	<b>Acc. %</b>	<b>Prec. %</b>	<b>Recall %</b>	<b>Spe. %</b>	<b>F1 %</b>
<i>Chazal, O'Dwyer, Reilly (2004)</i>	83,88	45,57	66,00	96,05	59,74
<i>Soria, Martínez (2009)</i>	91,47	69,89	89,98	94,60	76,43
<i>Llamedo, Martínez (2011)</i>	90,62	66,26	86,18	95,83	66,61
<i>Lin, Yang (2014)</i>	93,00	67,60	83,50	-	-
<i>Zhang et al. (2014)</i>	86,66	-	-	-	-
<i>Garcia et al. (2017)</i>	92,38	70,12	81,10	92,12	74,59
<i>Wang et al. (2021)</i>	97,48	70,75	67,47	96,01	68,76
<i>Dias et al. (2021)</i>	80,58	59,37	84,84	94,35	60,77
<i>Oliveira et al (2022)</i>	86,12	76,72	77,04	92,87	76,88
<i>Zahid, Kiranyaz, Gabbouj (2023)</i>	98,19	93,25	91,35	96,37	92,27
<b>CNN-AM MLII (S)</b>	<b>83,89</b>	<b>59,51</b>	<b>84,86</b>	<b>93,24</b>	<b>61,86</b>
<b>CNN-AM MLII+V (S)</b>	<b>95,47</b>	<b>72,28</b>	<b>87,21</b>	<b>96,62</b>	<b>75,25</b>
<b>CNN-AM MLII (F1)</b>	<b>97,29</b>	<b>74,41</b>	<b>78,33</b>	<b>95,42</b>	<b>75,96</b>
<b>CNN-AM MLII+V (F1)</b>	<b>98,48</b>	<b>94,15</b>	<b>80,23</b>	<b>96,34</b>	<b>81,91</b>

Fonte: Próprio Autor.

**Quadro 41:** Comparação com trabalhos anteriores para o paradigma interpaciente dos indicadores por classe.

<b>Resultados Anteriores</b>	<b>Acc. %</b>	<b>Recall %</b>			<b>Prec. %</b>			<b>F1-Score %</b>		
		<b>N</b>	<b>S</b>	<b>V</b>	<b>N</b>	<b>S</b>	<b>V</b>	<b>N</b>	<b>S</b>	<b>V</b>
Chazal, O'Dwyer, Reilly (2004)	83,88	86,9	75,9	77,7	99,2	38,5	81,6	92,6	51,1	79,6
Soria, Martinez (2009)	91,45	91,7	88,3	89,9	98,9	39,5	71,2	95,2	54,6	79,5
Llamedo, Martínez (2011)	90,62	91,8	84,8	81,9	99,5	10,9	88,4	95,5	19,3	85,1
Lin, Yang (2014)	93,00	91,6	81,4	86,2	99,3	31,6	31,6	95,3	45,5	79,5
Zhang et al. (2014)	86,66	88,9	79,1	85,5	99,0	36,0	92,8	-	-	-
Garcia et al. (2017)	92,38	94,0	62,0	87,3	98,0	53,0	59,4	95,9	57,1	70,7
Zhai, Tin (2018)	-	-	76,8	93,8	-	74,0	92,4	-	-	-
Wang et al. (2021)	99,16	99,8	74,6	96,6	99,4	89,5	93,2	99,6	81,4	94,4
Dias et al. (2021)	89,61	91,0	72,1	78,7	98,1	23,6	70,8	94,4	35,6	74,5
Oliveira et al (2022)	86,12	99,8	46,1	85,2	98,6	44,9	86,6	99,2	45,5	85,9
Zahid, Kiranyaz, Gabbouj (2023)	98,19	99,3	83,3	91,4	98,9	83,5	97,4	99,1	83,4	94,3
<b>CNN-AM MLII (S)</b>	<b>83,89</b>	<b>98,3</b>	<b>77,5</b>	<b>93,8</b>	<b>99,6</b>	<b>7,6</b>	<b>96,3</b>	<b>90,7</b>	<b>13,9</b>	<b>81,1</b>
<b>CNN-AM MLII+V (S)</b>	<b>95,47</b>	<b>95,8</b>	<b>69,3</b>	<b>96,6</b>	<b>99,4</b>	<b>20,2</b>	<b>97,2</b>	<b>97,6</b>	<b>31,3</b>	<b>96,9</b>
<b>CNN-AM MLII (F1)</b>	<b>97,29</b>	<b>98,3</b>	<b>41,1</b>	<b>95,7</b>	<b>98,9</b>	<b>28,1</b>	<b>96,3</b>	<b>98,6</b>	<b>33,3</b>	<b>96,0</b>
<b>CNN-AM MLII+V (F1)</b>	<b>98,48</b>	<b>99,4</b>	<b>43,3</b>	<b>98,0</b>	<b>99,0</b>	<b>56,3</b>	<b>97,0</b>	<b>99,2</b>	<b>49,0</b>	<b>97,5</b>

Fonte: Próprio Autor.

Considerando os indicadores por classe apresentados no Quadro 41, a dificuldade em classificar a classe S é notória em todos os trabalhos, sendo que a tendência de uma maior acurácia geral implica numa menor sensibilidade para a classe N, como pode ser observado em Soria, Martinez (2009) e em Dias *et al.* (2021) com 91,3% e 40,3%. Existe uma grande variação de foco no desenvolvimento dos trabalhos, o que também deve ser considerado para realizar comparações, pois trabalhos como Wang *et al.* (2021) e Dias *et al.* (2021) exploram efetivamente a melhoria de acurácia da classe S, enquanto outros como Zhang *et al.* (2014), Garcia *et al.* (2017) e Oliveira *et al.* (2022) focam mais nos indicadores gerais.

O trabalho de Zahid, Kiranyaz e Gabbouj (2023) merece destaque nesse contexto pois avançou em metodologias de segmentação e extração de recursos nos dados 1D, revelando características que se tornaram decisivas para apresentar uma das melhores métricas da atualidade para este paradigma, com 92,27% de *F1-Score* geral e 83,3% e 83,5% de acurácia e precisão para a classe S, sem reduzir significativamente a sensibilidade para a classe N.

Em relação aos nossos resultados pode ser observado que a CNN-AM proposta, em especial no cenário de duas derivações com maior *F1-Score* (MLII+V-f1), atingiu os maiores valores de acurácia e precisão gerais, com 98,48% e 94,15%, além de um valor alto de *F1-Score*, de 81,91%. No cenário com uma derivação e maior *F1-Score* (MLII-f1) se observa que os resultados se aproximam dos valores dos trabalhos correlatos, com

melhor desempenho nos indicadores gerais, precisão e *F1-Score*, demonstrando um bom desempenho do modelo.

Observando especificamente a classe S e as dificuldade na sua classificação, pode ser observado uma forte flutuação nos valores das métricas obtidas em todos os trabalhos, tendo o nosso modelo com uma derivação obtido uma baixa precisão, de 7,5% para esta classe, mesmo com 77,5% de acurácia. O desempenho para a classe S só se mostra superior com duas derivações, com 43,3% e 56,3% de acurácia no cenário com foco no maior *F1-Score* geral, e 69,3% e 20,2% para o foco na maior sensibilidade de S. Nesse contexto, o modelo com duas derivações e maior *F1-Score* se torna o mais competitivo em relação aos resultados apresentados nos trabalhos comparados.

Para o caso do paradigma intrapaciente, a maioria dos trabalhos na literatura têm apresentado apenas resultados considerando os indicadores globais, sendo que desta forma o Quadro 42 foi elaborado, onde todos os trabalhos relacionados utilizaram a base de dados MIT-BIH mas com divisão da base de treino e teste diferentes, impossibilitando uma análise comparativa direta dos resultados.

**Quadro 42:** Comparação com trabalhos anteriores para o paradigma intrapaciente com a base MIT-BIH.

<b>Resultados Anteriores</b>	<b>Acc. %</b>	<b>Prec. %</b>	<b>Recall %</b>	<b>Spe %</b>	<b>F1 %</b>
<i>Kachuee, Fazeli, Sarrafzadeh (2018)</i>	95,90	95,20	95,10	-	-
<i>Izci et al. (2019)</i>	92,96	90,08	80,08	98,14	82,17
<i>Huang et al. (2019)</i>	99,00	-	-	-	-
<i>Zhai, Tin (2018)</i>	96,05	65,91	72,06	97,83	68,06
<i>Shaker et al. (2020)</i>	98,35	82,24	93,82	99,01	87,29
<i>Xu, Jeong, Li (2020)</i>	95,90	96,34	95,90	-	95,92
<i>Quiao et al. (2020)</i>	99,32	98,30	97,15	-	97,71
<i>He et al. (2021)</i>	98,29	99,22	98,29	-	-
<i>Ahmed et al. (2021)</i>	99,70	98,00	98,00	-	-
<i>Seitanidis, Gialelis, Papaconstantinou (2022)</i>	95,20	-	95,20	98,80	-
<i>Islam et al. (2023)</i>	99,60	97,66	99,60	-	98,21
<i>Mewada (2023)</i>	99,52	95,12	96,18	-	95,64
<i>Zhou, Fang (2024)</i>	99,59	84,19	97,92	99,70	87,72
<b>CNN-AM MLII</b>	<b>99,70</b>	<b>97,96</b>	<b>97,21</b>	<b>99,29</b>	<b>97,58</b>
<b>CNN-AM MLII+V</b>	<b>99,72</b>	<b>98,19</b>	<b>97,26</b>	<b>99,34</b>	<b>97,72</b>

**Fonte:** Próprio Autor.

Observando os resultados apresentados no Quadro 42, podemos considerar que os resultados obtidos neste estudo para paradigma são promissores, com resultados competitivos em todas as métricas, com maior destaque para os resultados obtidos para o classificador proposto com duas derivações.

De forma similar, o Quadro 43 apresenta um comparativo com publicações qualificadas em ordem cronológica dos resultados atingidos com a base PTB. Em todos os casos, também são usadas as classes de infarto no miocárdio e de indivíduos saudáveis, organizando os dados em treino e teste de forma aleatória na proporção 80/20, o que permite uma comparação objetiva. Observa-se que os resultados alcançados com a estrutura consolidada neste estudo apresentaram maior *F1-Score* entre os estudos comparados.

**Quadro 43:** Comparação com trabalhos anteriores para o paradigma intrapaciente com a base PTB.

<b>Resultados Anteriores</b>	<b>Acc. %</b>	<b>Prec. %</b>	<b>Recall %</b>	<b>Spe %</b>	<b>F1 %</b>
<i>Sharma, Tripathy, Dandapatl (2015)</i>	96,00	-	93,00	99,00	-
<i>Acharya et al. (2017)</i>	95,22	-	95,49	94,19	-
<i>Chen et al. (2018)</i>	96,18	-	93,67	97,32	-
<i>Kachuee, Fazeli, Sarrafzadeh (2018)</i>	95,90	95,20	95,10	-	-
<i>Liu et al. (2018)</i>	96,00	-	95,40	97,37	-
<i>Diker et al. (2019)</i>	83,82	82,35	95,45	62,50	88,42
<i>Cao et al. (2020)</i>	96,65	-	-	-	-
<i>Ahamed et al. (2020)</i>	97,66	96,90	97,06	-	-
<i>Ahmad et al. (2021)</i>	99,20	98,00	98,00	-	-
<i>Lakhdari, Saeed (2022)</i>	99,00	99,00	98,00	-	99,00
<i>Kim et al. (2023)</i>	96,53	95,54	95,85	95,85	95,69
<i>Victor et al. (2023)</i>	98,45	98,45	98,45	-	98,45
<i>El-Ghaish, Eldele (2024)</i>	-	99,35	99,46	-	99,41
<i>Yousuf et al. (2024)</i>	99,84	-	99,70	99,40	-
<b>CNN-AM RP+HSFC</b>	<b>99,73</b>	<b>99,73</b>	<b>99,58</b>	<b>99,58</b>	<b>99,66</b>

Fonte: Próprio Autor.

## 6.5 CONSIDERAÇÕES FINAIS

Observa-se que os resultados obtidos com a CNN-AM proposta estão acompanhando os melhores resultados da literatura contemporânea tanto para a base MIT-BIH quanto para a base PTB, comprovando a robustez da metodologia aqui desenvolvida, atingindo o estado da arte.

O paradigma interpaciente é importante por ser mais realista, visto que um classificador no mundo real para avaliação clínica será utilizado em indivíduos cujos dados não foram coletados para treinamento do modelo. No entanto, comparar o desempenho de modelos nos dois paradigmas continua a ser um campo ativo de pesquisa para classificação de sinais ECG.

Destacaram-se os resultados utilizando imagens RP e HSFC, que atingiram os

melhores indicadores globais, utilizando uma abordagem mais complexa com CNN e módulo de atenção com *kernel* adaptativo, ou seja, que é atualizado na fase de treinamento. Observou-se, no entanto, uma diferença na estrutura de cálculo da convolução com o *kernel* que obteve resultados superiores no paradigma intrapaciente com ajuste de dimensão vetorial após a convolução com o *kernel*, enquanto a estrutura com aumento reflexivo do vetor de características antes da convolução se mostrou superior no paradigma interpaciente, sendo que, nos dois casos, comprovou-se que a abordagem *Self Attention* se mostrou eficaz na classificação, como observado no estudo de ablação.

## 7 CONCLUSÃO

Nesta tese, uma estrutura baseada em uma CNN Multimodal com Mecanismo de Atenção para classificação de arritmias cardíacas foi construída. Ela foi desenvolvida e avaliada através das bases de dados MIT-BIH e PTB, considerando tanto o paradigma interpaciente quanto o intrapaciente. A CNN-AM consolidada agrega diversos métodos e técnicas, dentre os quais: abordagem multimodal de diferentes dados e de diferentes representações do mesmo dado, filtro de sinais ECG, transformação de sinais temporais em imagens, fusão de características, implementação de *Self Attention* com *kernel adaptativo* e aumento de dados.

Na fase exploratória, diversas técnicas de imagens foram estudadas de forma individual e conjugadas com diversas estratégias de fusão, e duas delas se destacaram: RP e HSFC. Essas técnicas oferecem perspectivas complementares na análise de sinais temporais, com RP destacando estruturas repetitivas e padrões globais, sendo sensível a variações sutis de frequência e amplitude, e HSFC fornece uma visão global do sinal em escalas de tempo mais amplas, sendo menos influenciado por variações rápidas, fazendo com que mudanças de amplitude sejam refletidas na intensidade entre diferentes regiões da imagem, enquanto variações de frequência se manifestam na densidade e repetição dos padrões de cor, onde sequências de cores similares indicam frequências próximas, e transições de cores refletem alterações de frequência. Juntas, essas técnicas permitiram uma análise complementar de padrões no sinal temporal, alcançando os resultados experimentais apresentados.

Com relação às metodologias de aumento que permitiram a geração de dados sintéticos, observou-se que são essenciais para melhores indicadores de classificação. Nesta tese, foram feitos estudos com diferentes técnicas de forma isolada e conjunta, e pode-se concluir que para classes com dinâmicas mais padronizadas, como a classe V, o SMOTE se mostrou suficiente, sendo também uma técnica de rápida implementação e baixo custo computacional, além de ser baseada em método estatístico. Mas para classes com dinâmicas mais difíceis de serem aprendidas pelo classificador, como a classe S, o uso conjunto com a técnica WGAN-GP mostrou-se mais eficaz, mas é uma técnica que exige um processamento computacional mais elevado, sendo baseada em GAN.

Diante dos resultados apresentados e do estudo de ablação, observou-se que a presença do mecanismo de atenção na estrutura aumentou os indicadores globais e a

sensibilidade para a classe S, realizando o que já se esperava pela fundamentação teórica: os pesos de atenção amplificam as contribuições de características relevantes, como padrões específicos ou elementos com maior impacto para o modelo, e também suprimindo informações irrelevantes, atenuando o impacto de características menos importantes ou redundantes, ajudando a melhorar a eficiência da classificação.

Portanto, pode-se concluir que a CNN-AM consolidada nesta tese alcançou o estado da arte de classificação de sinais ECG em ambos os paradigmas, com destaque para os indicadores no paradigma intrapaciente na base MIT-BIH, usando duas derivações, de 99,72%, 98,19%, 97,26%, 99,34% e 97,72% de acurácia global, precisão, sensibilidade, especificidade e *F1-Score*, na base PTB com 99,73%, 99,73%, 99,58%, 99,58% e 99,66%, e no paradigma interpaciente apenas com a base MIT-BIH, também com duas derivações, de 98,48%, 94,15%, 80,23%, 96,34% e 81,91%.

A abordagem de dois paradigmas com o mesmo procedimento metodológico não é comum na literatura contemporânea, sendo que este estudo apresenta as diferenças de tratamento com as duas abordagens, e os desafios que se apresentam em cada uma delas, mostrando de forma eficaz que um bom modelo considerado em um paradigma pode também ser robusto em outro, mas com indicadores de referência diferentes, pois são esperados altos indicadores para uma abordagem intrapaciente que, embora tenha sua relevância para avaliação de variabilidade intra-individual, não é uma avaliação justa para aplicações clínicas em que um modelo automatizado é aplicado em pacientes cujos dados não foram observados no treinamento do modelo. Para essa finalidade, o paradigma interpaciente se torna a referência de avaliação, com indicadores bem mais baixos para avaliação de sua eficácia.

Por fim, a segmentação e o processamento de sinais ECG antes da geração das imagens são elementos ainda importante a serem trabalhados futuramente com as metodologias aqui propostas, bem como inovações metodológicas na abordagem multimodal, incluído diferentes estruturas convolucionais, ou complementando com técnicas de redes neurais 1D, além da introdução de novos mecanismos de atenção em camadas convolucionais ou fortemente conectadas.

## 7.1 PONTOS DE INOVAÇÃO

Esta tese apresenta os seguintes pontos de inovação dentro do que já vem sendo

apresentado na literatura para desenvolvimento de sistemas automáticos de classificação de arritmia:

1. Estudo comparativo de resultados entre abordagem interpaciente e intrapaciente para demonstrar a importância da abordagem interpaciente para desenvolvimento de sistemas de classificação automáticos.
2. Indicação de técnicas de transformação de sinais ECG em imagens que possam gerar melhores resultados com classificação multimodal.
3. Indicação sobre melhoria de resultados de classificação através da convolução aplicada junto com a fusão na camada fortemente conectada com diferentes imagens que representem o mesmo sinal de entrada da CNN (*Self Attention local*).
4. Indicação sobre o uso de diferentes derivações, além das normalmente utilizadas, para definição da base de dados de treinamento dos modelos de classificação, para obtenção de melhorias de resultados na classificação multimodal.
5. Indicação de metodologias de geração de dados sintéticos 1D quando se utilizar abordagem multimodal.
6. Indicação de combinação de técnicas de aumento de dados pouco explorada na literatura contemporânea para melhorar a variabilidade de classes mais difíceis de classificar.

## REFERÊNCIAS

ABEDIN, Zainul. Differential diagnosis of wide QRS tachycardia: A review. *Journal of Arrhythmia*, Vol. 37, n. 5; 2021. Disponível em: <https://doi.org/10.1002/joa3.12599>. Acesso em: 17 out. 2024.

ACHARYA, U. Rajendra *et al.* Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, Vol. 415–416, 2017. Disponível em: <https://doi.org/10.1016/j.ins.2017.06.027>. Acesso em: 19 out. 2024.

ADEWOLE, Kayode S. *et al.* Expert System and Decision Support System for Electrocardiogram Interpretation and Diagnosis: Review, Challenges and Research Directions. *Applied Sciences*, Vol. 12, n. 23, 2022. Disponível em: <https://doi.org/10.3390/app122312342>. Acesso em: 17 out. 2024.

ADIB, Edmond *et al.* Synthetic ECG Signal Generation Using Probabilistic Diffusion Models. *IEEE Access* 2023, 11, 75818–75828. Disponível em: <https://doi.org/10.1109/ACCESS.2023.3296542>. Acesso em: 17 out. 2024.

AHAMED, Atik *et al.* ECG heartbeat classification using ensemble of efficient machine learning approaches on imbalanced datasets. In: *Proceedings of the 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, Dhaka, Bangladesh, nov. 2020. Disponível em: <https://doi.org/10.1109/ICAICT51780.2020.9333534>. Acesso em: 19 out. 2024.

AHMAD, Zeeshan *et al.* ECG Heartbeat Classification Using Multimodal Fusion. *IEEE Access*, Vol. 9, 2021. Disponível em: <https://doi.org/10.1109/ACCESS.2021.3097614>. Acesso em: 17 out. 2024.

AHMAD, Zeeshan; KHAN, Naimul. CNN-Based Multistage Gated Average Fusion (MGAF) for Human Action Recognition Using Depth and Inertial Sensors. *IEEE Sensors Journal*, Vol. 21, n. 3, 2021. Disponível em: <https://doi.org/10.1109/JSEN.2020.3028561>. Acesso em: 17 out. 2024.

AHMED, Adel A. *et al.* 1D CNNs can be used for tasks such as classifying cardiac arrhythmias, detecting specific events on the ECG. *Mathematics*, Vol. 11, n.3, 2023. Disponível em: <https://doi.org/10.3390/math11030562>. Acesso em: 17 out. 2024.

ALOTAIBY, Turkey N. *et al.* ECG-Based Subject Identification Using Statistical Features and Random Forest. *Journal of Sensors*, Vol. 2019, 2019. Disponível em: <https://doi.org/10.1155/2019/6751932>. Acesso em: 17 out. 2024.

ANSI/AAMI EC57:2012 (R2020). Testing And Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. AAMI, 2020. Disponível em: <https://webstore.ansi.org/Standards/AAMI/ANSIAAMIEC572012R2020>. Acesso em: 17 out. 2024.

ARJOVSKY, Martin; CHINTALA, Soumith; BOTTOU, Léon. Wasserstein generative adversarial networks. *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*, Vol. 70, 2017. Disponível em: <https://dl.acm.org/doi/10.5555/3305381.3305404>. Acesso em: 17 out. 2024.

ARPEN Brasil – Associação dos Registradores de Pessoas Naturais do Brasil. Portal da Transparência, 2022. Disponível em: <https://transparencia.registrocivil.org.br>. Acesso em: 17 out. 2024.

ASFAND-E-YAR, Muhammad *et al.* Multimodal CNN-DDI: Using multimodal CNN for drug to drug interaction associated events. *Scientific Reports*, Vol. 14, 2024. Disponível em: <https://doi.org/10.1038/s41598-024-54409-x>. Acesso em: 17 out. 2024.

AWAIS, Muhammad; IQBAL, Md. Tauhid Bin; BAE, Sung-Ho. Revisiting Internal Covariate Shift for Batch Normalization. *IEEE Transactions on Neural Networks and Learning Systems*, Volume: 32, n. 11, 2021. Disponível em: <https://doi.org/10.1109/TNNLS.2020.3026784>. Acesso em: 17 out. 2024.

BELKADI, Mohamed Amine; DAAMOUCHE, Abdelhamid; MELGANI, Farid. A deep neural network approach to QRS detection using autoencoders. *Expert Systems with Applications*, Vol. 184, 2021. Disponível em: <https://doi.org/10.1016/j.eswa.2021.115528>. Acesso em: 17 out. 2024.

BENALCAZAR, Daniel P. *et al.* A 3D Iris Scanner From a Single Image Using Convolutional Neural Networks. *IEEE Access*, Vol. 8, 2020. Disponível em: <https://doi.org/10.1109/ACCESS.2020.2996563>. Acesso em: 17 out. 2024.

BETTS, J. Gordon *et al.* *Anatomy and Physiology*. 2ed. Houston, Texas, USA: OpenStax, Rice University, 2022. Disponível em: <https://openstax.org/k12/anatomy-and-physiology>. Acesso em: 17 out. 2024.

BORRELL, R. *et al.* Parallel mesh partitioning based on space filling curves. *Computers & Fluids*, Elsevier, Vol. 173, n. 15, 2018, p. 264-272. Disponível em: <https://doi.org/10.1016/j.compfluid.2018.01.040>. Acesso em: 17 out. 2024.

BOUSSELJOT, R., KREISELER, D.; SCHNABEL, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik (Biomedical Engineering)*, Vol. 40, n. s1, 1995. Disponível em: <https://doi.org/10.1515/bmte.1995.40.s1.317>. Acesso em: 19 out. 2024.

BOUSSELJOT, Ralf. Aufbau der EKG-Datenbank CARDIODAT der PTB. *Biomedizinische Technik (Biomedical Engineering)*, Vol. 39, n. s1, 1994. Disponível em: <https://doi.org/10.1515/bmte.1994.39.s1.250>. Acesso em: 19 out. 2024.

BOUSSELJOT, Ralf-Dieter. PTB Diagnostic ECG Database, Version 1.0.0. Physikalisch-Technische Bundesanstalt, Berlin, Germany (Instituto Nacional de Metrologia da Alemanha). *PhysioNet*: Cabridge-MA, USA, 2004. Disponível em: <https://doi.org/10.13026/C28C71>. Acesso em: 17 out. 2024.

BRAGIN, Aleksandr; SPITSYN, Vladimir. Neural Network Analysis of Electroencephalograms Graphical Representation. *Anais do 30th International Conference on Computer Graphics and Machine Vision*, Saint Petersburg, Russia, 2020. Disponível em: <http://dx.doi.org/10.51130/graphicon-2020-2-4-52>. Acesso em: 17 out. 2024.

CAO, Yangjie *et al.* Multi-channel lightweight convolutional neural network for remote myocardial infarction monitoring. In: IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Seoul, Korea (South), Apr. 2020. Disponível em: <https://doi.org/10.1109/WCNCW48565.2020.9124860>. Acesso em: 19 out. 2024.

CASAS, Manuel M. *et al.* Bayesian Classification Models for Premature Ventricular Contraction Detection on ECG Traces. *Journal of Healthcare Engineering*, Vol. 2018; 2018. Disponível em: <https://doi.org/10.1155%2F2018%2F2694768>. Acesso em: 17 out. 2024.

CASDAGLI, M. C. Recurrence plots revisited. *Physica D: Nonlinear Phenomena*, Vol. 108, n. 1–2, 1997. Disponível em: [https://doi.org/10.1016/S0167-2789\(97\)82003-9](https://doi.org/10.1016/S0167-2789(97)82003-9). Acesso em: 17 out. 2024.

CHAWLA, N. V. *et al.* SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 2002. Disponível em: <https://doi.org/10.1613/jair.953>. Acesso em: 17 out. 2024.

CHAZAL, Philip de. Heartbeat classification system using adaptive learning from selected beats, in: *Computing in Cardiology Conference (CinC)*, 2014. Disponível em: <https://ieeexplore.ieee.org/document/7043146>. Acesso em: 17 out. 2024.

CHAZAL, Philip de; O'DWYER, Maria; REILLY, Richard B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, Vol. 51, n. 7, 2004. Disponível em: <https://doi.org/10.1109/tbme.2004.827359>. Acesso em: 17 out. 2024.

CHAZAL, Philip de; REILLY, Richard B. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features, *IEEE Transactions on Biomedical Engineering*, Vol. 53, n. 12, 2006. Disponível em: <https://doi.org/10.1109/tbme.2006.883802>. Acesso em: 17 out. 2024.

CHEN Aiyun *et al.* A Real Time QRS Detection Algorithm Based on ET and PD Controlled Threshold Strategy. *Sensors*, Vol. 20, n. 14, 2020. Disponível em: <https://doi.org/10.3390/s20144003>. Acesso em: 17 out. 2024.

CHEN, Long *et al.* Cross-modal multiscale multi-instance learning for long-term ECG classification. *Information Sciences*, Vol. 643, 2023. Disponível em: <https://doi.org/10.1016/j.ins.2023.119230>. Acesso em: 22 out. 2024.

CHEN, Yufei *et al.* Multi-channel lightweight convolution neural network for anterior myocardial infarction detection. In: *Proceedings of the IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*, Guangzhou, China, Oct. 2018. Disponível em: <https://doi.org/10.1109/SmartWorld.2018.00119>. Acesso em: 19 out. 2024.

CHEN, Zhikang *et al.* A novel imbalanced dataset mitigation method and ECG classification model based on combined 1D\_CBAM-autoencoder and lightweight CNN model. *Biomedical Signal Processing and Control*, Vol. 87, Part B, 2024. Disponível em: <https://doi.org/10.1016/j.bspc.2023.105437>. Acesso em: 15 jan. 2025.

- CHEN, Zhitao *et al.* Self-Attention-Based Conditional Variational AutoEncoder Generative Adversarial Networks for Hyperspectral Classification. *Remote Sensing*, Vol. 13, 2021. Disponível em: <https://doi.org/10.3390/rs13163316>. Acesso em: 21 jan. 2025.
- CLARO, Maíla *et al.* Utilização de Técnicas de Data Augmentation em Imagens: Teoria e Prática. In: TELES, Ariel Soares; CALÇADA, Dario Brito; VERAS, Nécio Lima (Eds.). *Minicursos da ERCEMAPI 2020*, SBC: Porto Alegre, 2020. Disponível em: <https://doi.org/10.5753/sbc.11.5.3>. Acesso em: 17 out. 2024.
- CLIFFORD, Gari D.; AZUAJE, Francisco; MCSHARRY, Patrick. *Advanced Methods And Tools for ECG Data Analysis*. Artech House Publishers, 1 ed., 2006
- CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. *Mach Learn.* 1995, 20, 273–297. Disponível em: <https://doi.org/10.1007/BF00994018>. Acesso em: 17 out. 2024.
- CRAMMER, Koby; SINGER, Yoram. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.* 2001, 2, 265–292. Disponível em: <https://dl.acm.org/doi/10.5555/944790.944813>. Acesso em: 17 out. 2024.
- CUI, Yin *et al.* Class-Balanced Loss Based on Effective Number of Samples. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019. Disponível em: <https://doi.org/10.1109/CVPR.2019.00949>. Acesso em: 17 out. 2024.
- DI PAOLO, Ítalo Flexa; CASTRO, Adriana Rosa Garcez. Intra- and Interpatient ECG Heartbeat Classification Based on Multimodal Convolutional Neural Networks with an Adaptive Attention Mechanism. *Applied Sciences*, Vol. 14, n. 20, 2024. Disponível em: <https://doi.org/10.3390/app14209307>. Acesso em: 17 out. 2024.
- DIAS, Felipe Meneguitti *et al.* Arrhythmia classification from single-lead ECG signals using the inter-patient paradigm. *Computer Methods and Programs in Biomedicine*, Vol. 202, 2021. Disponível em: <https://doi.org/10.1016/j.cmpb.2021.105948>. Acesso em: 17 out. 2024.
- DIKER, Aykut *et al.* A novel application based on spectrogram and convolutional neural network for ECG classification. In: *Proceedings of the 1st International Informatics and Software Engineering Conference (UBMYK)*, Ankara, Turkey, nov. 2019. Disponível em: <https://doi.org/10.1109/UBMYK48245.2019.8965506>. Acesso em: 19 out. 2024.
- EARL, Dent. Script to plot 1D data in 2D using the Hilbert Curve. Honestly a pretty terrible visualization technique for conveying information, but it looks cool. Santa Cruz, CA, Estados Unidos, 2013. Disponível em: <https://github.com/dentearl/simpleHilbertCurve>. Acesso em: 17 out. 2024.
- EBERHART, R.; KENNEDY, J. A new optimizer using particle swarm theory. *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan, 1995. Disponível em: <https://doi.org/10.1109/MHS.1995.494215>. Acesso em: 17 out. 2024.

EBERHART, R.; SHI, Yuhui. Particle swarm optimization: developments, applications and resources. Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), Seoul, South Korea, 2001. Disponível em: <https://doi.org/10.1109/CEC.2001.934374>. Acesso em: 17 out. 2024.

ECG Heartbeat Categorization Dataset. 2018. [Online]. Disponível em: <https://www.kaggle.com/shayanfazeli/heartbeat>. Acesso em: 17 out. 2024.

ECKMANN, J.-P.; KAMPHORST, S. Oliffson; RUELLE, D. Recurrence Plots of Dynamical Systems. Europhysics Letters, Vol. 4, n. 9, 1987. Disponível em: <https://doi.org/10.1209/0295-5075/4/9/004>. Acesso em: 17 out. 2024.

ELGENDI, Mohamed. Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases. PLoS ONE, Vol. 8, n. 9, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0073557>. Acesso em: 17 out. 2024.

EL-GHAISH, Hany; ELDELE, Emadeldeen. ECGTransForm: Empowering adaptive ECG arrhythmia classification framework with bidirectional transformer. Biomedical Signal Processing and Control, Vol. 89, 2024. Disponível em: <https://doi.org/10.1016/j.bspc.2023.105714>. Acesso em: 20 out. 2024.

ELOLA, Andoni *et al.* Deep Learning for Pulse Detection in Out-of-Hospital Cardiac Arrest Using the ECG. Computing in Cardiology Conference (CinC), Maastricht, Holanda: IEEE, 2018. Disponível em: <https://doi.org/10.22489/CinC.2018.093>. Acesso em: 17 out. 2024.

ELREEDY, Dina; ATIYA, Amir F. A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. Information Sciences, Vol. 505, 2019. Disponível em: <https://doi.org/10.1016/j.ins.2019.07.070>. Acesso em: 17 out. 2024.

EMRICH, Jonas *et al.* Accelerated Sample-Accurate R-Peak Detectors Based on Visibility Graphs. In: Proceedings of the 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finlandia, 4–8 September 2023. Disponível em: <https://doi.org/10.23919/EUSIPCO58844.2023.10290007>. Acesso em: 17 out. 2024.

FAN, Tianqi *et al.* A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition. Computers in Biology and Medicine, Vol. 159, 2023. Disponível em: <https://doi.org/10.1016/j.compbimed.2023.106938>. Acesso em: 15 jan. 2025.

FAOUZI, Johann (Org.). A Python Package for Time Series Classification. 2021. URL: <https://pyts.readthedocs.io/en/stable/index.html>. Acesso em: 17 out. 2024.

FAOUZI, Johann; JANATI, Hicham. pyts: A python package for time series classification. Journal of Machine Learning Research, Vol. 21, n.46, 2020. Disponível em: <http://jmlr.org/papers/v21/19-763.html>. Acesso em: 17 out. 2024.

FARAG, Mohammed M. A Self-Contained STFT CNN for ECG Classification and Arrhythmia Detection at the Edge. IEEE Access, Vol. 10, 2022. Disponível em: <https://doi.org/10.1109/ACCESS.2022.3204703>. Acesso em: 17 out. 2024.

- FARIA, Fabio A. *et al.* Fusion of time series representations for plant recognition in phenology studies. *Pattern Recognition Letters*, Vol. 83, Parte 2, 2016. Disponível em: <https://doi.org/10.1016/j.patrec.2016.03.005>. Acesso em: 17 out. 2024.
- FENG, Chunsheng *et al.* The parallel generation of 2-D Hilbert Space-filling Curve on GPU. In *Proceedings of the 5th International Conference on BioMedical Engineering and Informatics*, Chongqing, China, 16–18 October 2012. Disponível em: <https://doi.org/10.1109/BMEI.2012.6512968>. Acesso em: 17 out. 2024.
- FRADI, Marwa *et al.* A. Automatic heart disease class detection using convolutional neural network architecture-based various optimizers-networks. *IET Smart Cities*, Vol. 3, n. 1, 2021. Disponível em: <https://doi.org/10.1049/smc2.12003>. Acesso em: 17 out. 2024.
- FRAHAN, Ahmed; LI, Chen; AHMED, Md Toukir. A qualitative overview of fuzzy logic in ECG arrhythmia classification. *International Journal of Engineering Works*, Vol. 5, n. 11, 2018. Disponível em: <https://doi.org/10.5281/zenodo.1486134>. Acesso em: 17 out. 2024.
- FRANCIS, Johnson. ECG monitoring leads and special leads. *Indian Pacing and Electrophysiology Journal*, Vol. 16, n. 3, 2016. Disponível em: <https://doi.org/10.1016/j.ipej.2016.07.003>. Acesso em: 15 jan. 2025.
- FU, Jianglin *et al.* StyleGAN-Human: A Data-Centric Odyssey of Human Generation. *Computer Vision and Pattern Recognition*, 2022. Disponível em: <https://doi.org/10.48550/arXiv.2204.11823>. Acesso em: 17 out. 2024.
- GARCIA, Gabriel *et al.* Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO. *Scientific Reports*, Vol. 7, 2017. Disponível em: <https://doi.org/10.1038/s41598-017-09837-3>. Acesso em: 17 out. 2024.
- GIFFONI, Rodrigo Tobias; TORRES, Rosália Morais Torres. Breve história da eletrocardiografia. *Revista Médica de Minas Gerais RMMG*, Vol. 20, n. 2, 2009. Disponível em: <https://www.rmmg.org/artigo/detalhes/323>. Acesso em: 17 out. 2024.
- GOLANY, Tomer; RADINSKY, Kira. PGANs: Personalized Generative Adversarial Networks for ECG Synthesis to Improve Patient-Specific Deep ECG Classification, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019. Disponível em: <https://doi.org/10.1609/aaai.v33i01.3301557>. Acesso em: 17 out. 2024.
- GOLDBERGER, A. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*. Vol. 101, n. 23, 2000. Disponível em: <https://doi.org/10.1161/01.cir.101.23.e215>. Acesso em: 17 out. 2024.
- GOODFELLOW, Ian J. *et al.* Generative Adversarial Networks. *Machine Learning*, 2014. Disponível em: <https://doi.org/10.48550/arXiv.1406.2661>. Acesso em: 17 out. 2024.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. Massachusetts, EUA: MIT Press. 2016.

- GOTO, Shinichi *et al.* Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients, *PloS One* Vol. 14, n. 1, 2019. Disponível em: <https://doi.org/10.1371/journal.pone.0210103>. Acesso em: 17 out. 2024.
- GU, Jiuxiang *et al.* Recent advances in convolutional neural networks. *Pattern Recognition*, Vol. 77, 2018. Disponível em: <https://doi.org/10.1016/j.patcog.2017.10.013>. Acesso em: 17 out. 2024.
- GULRAJANI, Ishaan *et al.* Improved Training of Wasserstein GANs. In *Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS)*, Long Beach, California, USA, 4–9 December 2017. Disponível em: <https://doi.org/10.48550/arXiv.1704.00028>. Acesso em: 17 out. 2024.
- HALEEM, Muhammad Salman *et al.* Time adaptive ECG driven cardiovascular disease detector. *Biomedical Signal Processing and Control*. Elsevier, 2021. Disponível em: <https://doi.org/10.1016/j.bspc.2021.102968>. Acesso em: 17 out. 2024.
- HALL John E. *Guyton and Hall Textbook of Medical Physiology*. Elsevier: EUA, 14 ed., 2021.
- HAMMAD, Mohamed; LIU, Yashu; WANG, Kuanquan. Multimodal Biometric Authentication Systems Using Convolution Neural Network Based on Different Level Fusion of ECG and Fingerprint. *IEEE Access*, Vol. 6, 2018. Disponível em: <https://doi.org/10.1109/ACCESS.2018.2886573>. Acesso em: 12 out. 2022.
- HANNUN, Awni Y. *et al.* Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, Vol. 25, 2019. Disponível em: <https://doi.org/10.1038/s41591-018-0268-3>. Acesso em: 17 out. 2024.
- HAO, Jingyu *et al.* Fetal Electrocardiogram Signal Extraction Based on Fast Independent Component Analysis and Singular Value Decomposition. *Sensors (Basel)*, Vol. 22, n. 10, 2022. Disponível em: <https://doi.org/10.3390/s22103705>. Acesso em: 17 out. 2024.
- HE, Kaiming *et al.* Deep Residual Learning for Image Recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, EUA, 2016. Disponível em: <https://doi.org/10.1109/CVPR.2016.90>. Acesso em: 17 out. 2024.
- HE, Kaiming *et al.* Deep Residual Learning for Image Recognition. *Computer Science*, 2015. Disponível em: <https://doi.org/10.48550/arXiv.1512.03385>. Acesso em: 17 out. 2024.
- HE, Runnan *et al.* Automatic detection of QRS complexes using dual channels based on U-Net and bidirectional long short-term memory. *IEEE Journal of Biomed Health Informatics*, Vol. 25, n. 4, 2021. Disponível em: <https://doi.org/10.1109/jbhi.2020.3018563>. Acesso em: 17 out. 2024.
- HILBERT, David. Ueber die stetige Abbildung einer Line auf ein Flächenstück. *Mathematische Annalen*, Vol. 38, 1891. Disponível em: <https://doi.org/10.1007/BF01199431>. Acesso em: 17 out. 2024.

- HONG, Shenda *et al.* Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine*, n. 122, 2020. Disponível em: <https://doi.org/10.1016/j.compbiomed.2020.103801>. Acesso em: 17 out. 2024.
- HOWARD, Andrew G. *et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *Computer Science*, 2017. Disponível em: <https://doi.org/10.48550/arXiv.1704.04861>. Acesso em: 17 out. 2024.
- HUANG, Jingshan *et al.* ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access*, Vol. 7, 2019. Disponível em: <https://doi.org/10.1109/ACCESS.2019.2928017>. Acesso em: 17 out. 2024.
- HUANG, Yangcheng *et al.* ECG classification based on guided attention mechanism. *Computer Methods and Programs in Biomedicine*, Vol. 257, 2024. Disponível em: <https://doi.org/10.1016/j.cmpb.2024.108454>. Acesso em: 15 jan. 2025.
- Imbalanced-Learn, Version 0.9. 2022. Disponível em: <https://imbalanced-learn.org>. Acesso em: 17 out. 2024.
- IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: accelerating deep network training by reducing internal covariate shift. *ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Vol. 37, Lille, France, 2015. Disponível em: <https://dl.acm.org/doi/10.5555/3045118.3045167>. Acesso em: 17 out. 2024.
- ISLAM, Md Shofiqul *et al.* HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN. *Neural Network*, Vol. 162, 2023. Disponível em: <https://doi.org/10.1016/j.neunet.2023.03.004>. Acesso em: 17 out. 2024.
- IZCI, E. *et al.* Cardiac arrhythmia detection from 2D ECG images by using deep learning technique. In: *Proceedings of the Medical Technologies Congress (TIPTEKNO)*, İzmir, Turquia, 2019. Disponível em: <http://doi.org/10.1109/tiptekno.2019.8895011>. Acesso em: 17 out. 2024.
- JIANG, Weiwei *et al.* Mobile Traffic Prediction in Consumer Applications: A Multimodal Deep Learning Approach. *IEEE Transactions on Consumer Electronics*, Vol. 70, n. 1, 2024. Disponível em: <https://doi.org/10.1109/TCE.2024.3361037>. Acesso em: 17 out. 2024.
- JING, Enbiao *et al.* ECG Heartbeat Classification Based on an Improved ResNet-18 Model. *Computational and Mathematical Methods in Medicine*, Vol. 2021, 2021. Disponível em: <https://doi.org/10.1155/2021/6649970>. Acesso em 23 out. 2024.
- JUN, Wen; LIYUAN, Zheng. Brain Tumor Classification Based on Attention Guided Deep Learning Model. *International Journal of Computational Intelligence Systems*, Vol. 15, 2022. Disponível em: <https://doi.org/10.1007/s44196-022-00090-9>. Acesso em: 17 out. 2024.

KACHUEE, Mohammad; FAZELI, Shayan; SARRAFZADEH, Majid. ECG Heartbeat Classification: a deep transferable representation. In: Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 4–7 June 2018. Disponível em: <https://doi.org/10.1109/ICHI.2018.00092>. Acesso em: 17 out. 2024.

KARRAS, Tero *et al.* Progressive Growing of GANs for Improved Quality, Stability, and Variation. 6th International Conference on Learning Representations, Vancouver, Canada, 2018. Disponível em: <https://doi.org/10.48550/arxiv.1710.10196>. Acesso em: 17 out. 2024.

KHODAEI, Pouya; ESFAHANIPOUR, Akbar; TAHERI, Hassan Mehtari. Forecasting turning points in stock price by applying a novel hybrid CNN-LSTM-ResNet model fed by 2D segmented images. Engineering Applications of Artificial Intelligence, Vol. 116, 2022. Disponível em: <https://doi.org/10.1016/j.engappai.2022.105464>. Acesso em: 17 out. 2024.

KIM, Eunchan *et al.* TinyML-Based Classification in an ECG Monitoring Embedded System. Computers, Materials & Continua, Vol. 75, n. 1, 2023. Disponível em: <https://doi.org/10.32604/cmc.2023.031663>. Acesso em: 20 out. 2024.

KINGMA, Diederik P.; BA, Jimmy. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015. Disponível em: <https://doi.org/10.48550/arXiv.1412.6980>. Acesso em: 17 out. 2024.

KIRANYAZ, Serkan *et al.* 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing, Vol. 151, 2021. Disponível em: <https://doi.org/10.1016/j.ymssp.2020.107398>. Acesso em: 22 out. 2024.

KIRANYAZ, Serkan; INCE, Turker; GABBOUJ, Moncef. Real-time patient-specific ECG classification by 1D convolutional neural networks, IEEE Transactions on Biomedical Engineering, Vol. 63, n. 3, 2016. Disponível em: <https://doi.org/10.1109/TBME.2015.2468589>

KOHONEN, Teuvo. Self-organized formation of topologically correct feature maps. Biological Cybernetics, Vol. 43, 1982. Disponível em: <https://doi.org/10.1007/BF00337288>. Acesso em: 17 out. 2024.

KREISELER, D.; BOUSSELJOT, R. Automatisierte EKG-Auswertung mit Hilfe der EKG-Signaldatenbank CARDIODAT der PTB. Biomedizinische Technik (Biomedical Engineering), Vol. 40, n. 1, 1995. Disponível em: <https://doi.org/10.1515/bmte.1995.40.s1.319>. Acesso em: 19 out. 2024.

KRIZHEVSKY, Alex. One weird trick for parallelizing convolutional neural networks. Neural and Evolutionary Computing, 2014. Disponível em: <https://doi.org/10.48550/arXiv.1404.5997>. Acesso em: 17 out. 2024.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, Vol. 60, n. 6, 2017. Disponível em: <https://doi.org/10.1145/3065386>. Acesso em: 17 out. 2024.

- KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. Proceedings in Advances in Neural Information Processing Systems 25 (NIPS 2012), Nevada, EUA, 2012. Disponível em: [https://papers.nips.cc/paper\\_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html](https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html). Acesso em: 17 out. 2024.
- LAKHDARI, Kheira; SAEED, Nagham. A new vision of a simple 1D Convolutional Neural Networks (1D-CNN) with Leaky-ReLU function for ECG abnormalities classification. Intelligence-Based Medicine, Vol. 6, 2022. Disponível em: <https://doi.org/10.1016/j.ibmed.2022.100080>. Acesso em: 20 out. 2024.
- LECUN, Yann *et al.* Gradient-based learning applied to document recognition. Proceedings of the IEEE, Vol. 86, n. 11, 1998. Disponível em: <https://doi.org/10.1109/5.726791>. Acesso em: 17 out. 2024.
- LEMAÎTRE, Guillaume; NOGUEIRA, Fernando; ARIDAS, Christos K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. Journal of Machine Learning Research, Vol. 18, n. 17, 2017. Disponível em: <http://jmlr.org/papers/v18/16-365.html>. Acesso em: 17 out. 2024.
- LI, Taiyong; ZHOU, Min. ECG classification using wavelet packet entropy and random forests. Entropy, Vol. 18, n. 8, 2016. Disponível em: <https://doi.org/10.3390/e18080285>. Acesso em: 17 out. 2024.
- LIN, Chun-Cheng; YANG, Chun-Min. Heartbeat classification using normalized RR intervals and morphological features. Mathematical Problems in Engineering, Vol. 2014, 2014. Disponível em: <https://doi.org/10.1155/2014/712474>. Acesso em: 17 out. 2024.
- LIN, Kang-Ping; CHANG, W.H. QRS feature extraction using linear prediction. IEEE Transactions on Biomedical Engineering, Vol. 36, n. 10, 1989. Disponível em: <https://doi.org/10.1109/10.40806>. Acesso em: 17 out. 2024.
- LIU, Hiu; ZHANG, Wei. Spatial and temporal variation and convergence in the efficiency of high-standard farmland construction: Evidence in China. Journal of Cleaner Production, Vol. 452, 2024. Disponível em: <https://doi.org/10.1016/j.jclepro.2024.142200>. Acesso em: 17 out. 2024.
- LIU, Mengxin; YANG, Jun. Image Classification of Brain tumor based on Channel Attention Mechanism. Journal of Physics: Conference Series, Vol. 2035, 2021. Disponível em: <https://www.doi.org/10.1088/1742-6596/2035/1/012029>. Acesso em: 17 out. 2024.
- LIU, Wenhan *et al.* Real-time multilead convolutional neural network for myocardial infarction detection. IEEE Journal of Biomedical and Health Informatics, Vol. 22, n. 5, 2018. Disponível em: <https://doi.org/10.1109/JBHI.2017.2771768>. Acesso em: 19 out. 2024.

- LLAMEDO, Mariano; MARTÍNEZ, Juan Pablo. Heartbeat Classification Using Feature Selection Driven by Database Generalization Criteria. *IEEE Transactions on Biomedical Engineering*, Vol. 58, n. 3, 2011. Disponível em: <https://doi.org/10.1109/TBME.2010.2068048>. Acesso em: 17 out. 2024.
- LOKA, Nasrulloh R. B. S.; KAVITHA, Muthusubash; KURITA, Takio. Hilbert Vector Convolutional Neural Network: 2D Neural Network on 1D Data. *Proceedings of 28th International Conference on Artificial Neural Networks, Munich, Germany, 2019*. Disponível em: [https://doi.org/10.1007/978-3-030-30487-4\\_36](https://doi.org/10.1007/978-3-030-30487-4_36). Acesso em: 17 out. 2024.
- LONG, Xi *et al.* Effects of Between- and Within-Subject Variability on Autonomic Cardiorespiratory Activity during Sleep and Their Limitations on Sleep Staging: A Multilevel Analysis. *Comput Intell Neurosci*, Vol. 2015, 2015. Disponível em: <https://doi.org/10.1155%2F2015%2F583620>. Acesso em: 17 out. 2024.
- LUZ, Eduardo José da S. *et al.* ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer methods and programs in biomedicine*, n. 127, 2016. Disponível em: <http://dx.doi.org/10.1016/j.cmpb.2015.12.008>. Acesso em: 17 out. 2024.
- LUZ, Eduardo; MENOTTI, David. How the choice of samples for building arrhythmia classifiers impact their performances. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society, 2011*. Disponível em: <https://doi.org/10.1109/iembs.2011.6091236>. Acesso em: 17 out. 2024.
- MA, Hangling *et al.* Change Detection Needs Neighborhood Interaction in Transformer. *Remote Sensing*, Vol. 15, 2023. Disponível em: <https://doi.org/10.3390/rs15235459>. Acesso em: 21 jan. 2025.
- MADEIRO, João Paulo do Vale *et al.* Evaluation of mathematical models for QRS feature extraction and QRS morphology classification in ECG signals. *Measurement*, Vol. 156, 2020. Disponível em: <https://doi.org/10.1016/j.measurement.2020.107580>. Acesso em: 17 out. 2024.
- MAR, Tanis *et al.* Optimization of ECG classification by means of feature selection. *IEEE Transactions on Biomedical Engineering*, Vol. 58, n. 8, 2011. Disponível em: <https://doi.org/10.1109/tbme.2011.2113395>. Acesso em: 17 out. 2024.
- MATHUNJWA, Bhukumuzi M. *et al.* ECG arrhythmia classification by using a recurrence plot and convolutional neural network. *Biomedical Signal Processing and Control*, Vol. 64, 2021. Disponível em: <https://doi.org/10.1016/j.bspc.2020.102262>. Acesso em: 17 out. 2024.
- Matplotlib, Version 3.7.1. Choosing Colormaps in Matplotlib. [2023?]. Disponível em: <https://matplotlib.org/stable/tutorials/colors/colormaps.html>. Acesso em: 17 out. 2024.
- MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, Vol. 5, n. 4, 1943. Disponível em: <https://doi.org/10.1007/BF02478259>. Acesso em: 17 out. 2024.

MEWADA, Hiren. 2D-wavelet encoded deep CNN for image-based ECG classification. *Multimedia Tools and Applications*, Vol. 82, 2023. Disponível em: <https://doi.org/10.1007/s11042-022-14302-z>. Acesso em: 17 out. 2024.

MODAK, Sudipta; TAHA, Luay Yassin; ABDEL-RAHEEM, Esam. A novel adaptive multilevel thresholding based algorithm for QRS detection. *Biomedical Engineering Advances*, Vol. 2, 2021a. Disponível em: <https://doi.org/10.1016/j.bea.2021.100016>. Acesso em: 17 out. 2024.

MODAK, Sudipta; TAHA, Luay Yassin; ABDEL-RAHEEM, Esam. A novel method of QRS Detection using Time and Amplitude Thresholds with Statistical False Peak Elimination. *IEEE Access*, Vol. 9, 2021b. Disponível em: <https://doi.org/10.1109/ACCESS.2021.3067179>. Acesso em: 17 out. 2024.

MOODY, George B.; MARK, Roger G. The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, Vol. 20, n. 3, 2001. Disponível em: <https://doi.org/10.1109/51.932724>. Acesso em: 17 out. 2024.

MOODY, George, MARK, Roger. MIT-BIH Arrhythmia Database, Version: 1.0.0. Massachusetts Institute of Technology, Laboratory for Computational Physiology. *PhysioNet*: Cambridge-MA, USA, 2005. Disponível em: <https://doi.org/10.13026/C2F305>. Acesso em: 17 out. 2024.

MORADY, Fred. Catheter Ablation of Supraventricular Arrhythmias: State of the Art. *Journal of Cardiovascular Electrophysiology*, Vol. 15, n. 1, 2004. Disponível em: <https://doi.org/10.1046/j.1540-8167.2004.03516.x>. Acesso em: 17 out. 2024.

MUTHULAKSHMI, M.; KAVITHA, G. Deep CNN with LM learning based myocardial ischemia detection in cardiac magnetic resonance images. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019. Disponível em: <https://doi.org/10.1109/EMBC.2019.8856838>. Acesso em: 17 out. 2024.

NAROTAMO, Hemaxi *et al.* Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches. *Biomedical Signal Processing and Control*, Vol. 93, 2024. Disponível em: <https://doi.org/10.1016/j.bspc.2024.106141>. Acesso em: 22 out. 2024.

NAZ, Mahwish *et al.* From ECG signals to images: a transformation based approach for deep learning. *From ECG signals to images: a transformation based approach for deep learning*. *PeerJ Computer Science*, Vol. 7, 2021. Disponível em: <https://doi.org/10.7717/peerj-cs.386>. Acesso em: 22 out. 2024.

O'BRIEN, James A. *Sistemas de Informação e as Decisões Gerenciais na Era da Internet*. Sarava: São Paulo, 3 ed., 2011. Acesso em: 17 out. 2024.

ODENA, Augustus; OLAH, Christopher; SHLENS, Jonathon. Conditional image synthesis with Auxiliary Classifier GANs. In: *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*, Vol. 70, 2017. Disponível em: <https://doi.org/10.48550/arXiv.1610.09585>. Acesso em: 17 out. 2024.

OLIVEIRA, Rafael F. *et al.* Explorando Redes Neurais de Grafos para Classificação de Arritmias. In Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS); Sociedade Brasileira de Computação (SBC): Teresina, Brasil, 2022. Disponível em: <https://doi.org/10.5753/sbcas.2022.222510>. Acesso em: 17 out. 2024.

PAN, Jiapu; TOMPKINS, Willis J. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, Vol. BME-32, n. 3, 1985. Disponível em: <https://doi.org/10.1109/TBME.1985.325532>. Acesso em: 17 out. 2024.

PETMEZAS, Georgios *et al.* State-of-the-Art Deep Learning Methods on Electrocardiogram Data: Systematic Review. *JMIR Medical Informatics*, Vol. 10, n. 8, 2022. Disponível em: <https://doi.org/10.2196/38454>. Acesso em: 17 out. 2024.

PHAM, Tuan *et al.* Classification of Short Time Series in Early Parkinson's Disease With Deep Learning of Fuzzy Recurrence Plots, *IEEE/CAA Journal of Automatica Sinica*, Vol. 6, n. 6, 2019. Disponível em: <https://doi.org/10.1109/JAS.2019.1911774>. Acesso em: 17 out. 2024.

PORUMB, Mihaela *et al.* Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG. *Scientific Reports*, Vol. 10, 2020. Disponível em: <https://doi.org/10.1038/s41598-019-56927-5>. Acesso em: 17 out. 2024.

QIANG, Yupeng *et al.* Conv-RGNN: An efficient Convolutional Residual Graph Neural Network for ECG classification. *Computer Methods and Programs in Biomedicine*, Vol. 257, 2024. Disponível em: <https://doi.org/10.1016/j.cmpb.2024.108406>. Acesso em: 15 jan. 2025.

QIAO, Fengjuan *et al.* A fast and accurate recognition of ECG signals based on ELM-LRF and BLSTM algorithm. *IEEE Access*, Vol. 8, 2020. Disponível em: <https://doi.org/10.1109/ACCESS.2020.2987930>. Acesso em: 17 out. 2024.

QIU, Shaohua *et al.* Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review. *Sensors*, Vol. 23, n. 3, 2023. Disponível em: <https://doi.org/10.3390/s23031305>. Acesso em: 17 out. 2024.

RADFORD, Alec; METZ, Luke, CHINTALA, Soumith. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *Machine Learning*, 2016. Disponível em: <https://doi.org/10.48550/arXiv.1511.06434>. Acesso em: 17 out. 2024.

RAWAL, Viraj; PRAJAPATI, Priyank; DARJI, Anand. Hardware implementation of 1D-CNN architecture for ECG arrhythmia classification. *Biomedical Signal Processing and Control*, Vol. 85, 2023. Disponível em: <https://doi.org/10.1016/j.bspc.2023.104865>. Acesso em: 17 out. 2024.

RODRÍGUEZ, R. *et al.* Feature Extraction of Electrocardiogram Signals by Applying Adaptive Threshold and Principal Component Analysis. *Journal of Applied Research and Technology*, Vol. 13, n. 2, 2015. Disponível em: <https://doi.org/10.1016/j.jart.2015.06.008>. Acesso em: 17 out. 2024.

ROSENBLATT, Frank. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, Vol. 65, n. 6, 1958. Disponível em: <https://doi.org/10.1037/h0042519>. Acesso em: 17 out. 2024.

RUMELHART, David E., HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. *Nature*, 323, 1986. Disponível em: <https://doi.org/10.1038/323533a0>. Acesso em: 17 out. 2024.

SEENA V.; YOMAS, Jerrin. A review on feature extraction and denoising of ECG signal using wavelet transform. *Proceedings of 2nd International Conference on Devices, Circuits and Systems (ICDCS)*, Coimbatore, India, 2014. Disponível em: <https://doi.org/10.1109/ICDCSyst.2014.6926190>. Acesso em: 17 out. 2024.

SEITANIDIS, Panagiotis; GIALELIS, John; PAPACONSTANTINO, Georgia. Identifying heart arrhythmias through multi-level algorithmic processing of ECG on edge devices. *Procedia Computer Science*, Vol. 203, 2022. Disponível em: <https://doi.org/10.1016/j.procs.2022.07.104>. Acesso em: 17 out. 2024.

SHAKER, Abdelrahman M. *et al.* Generalization of convolutional neural networks for ECG classification using generative adversarial networks. *IEEE Access*, Vol. 8, 2020. Disponível em: <https://doi.org/10.1109/ACCESS.2020.2974712>. Acesso em: 17 out. 2024.

SHARMA, L. N.; TRIPATHY, R. K.; DANDAPAT, S. Multiscale energy and eigenspace approach to detection and localization of myocardial infarction. *IEEE Transactions on Biomedical Engineering*, Vol. 62, n. 7, 2015. Disponível em: <https://doi.org/10.1109/TBME.2015.2405134>. Acesso em: 19 out. 2024.

SILVA, Ulrich Kauê Mendes Alencar da. Classificação de Tumor Cerebral: um estudo comparativo entre Rede Neural Convolutiva e Rede Neural Convolutiva com mecanismo de atenção. 2024. Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Pará, Belém, 2024.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science*, 2014. Disponível em: <https://doi.org/10.48550/arXiv.1409.1556>. Acesso em: 17 out. 2024.

SINGH, Anupreet Kaur; KRISHNAN, Sridhar. ECG signal feature extraction trends in methods and applications. *BioMedical Engineerin*, Vol. 22, n. 22, 2023. Disponível em: <https://doi.org/10.1186/s12938-023-01075-1>. Acesso em: 17 out. 2024.

SKILLING, John. Programming the Hilbert curve. In: *AIP Conference Proceedings*, Vol. 707, 2004. Disponível em: <https://doi.org/10.1063/1.1751381>. Acesso em: 17 out. 2024.

SNOW, John. *On the Mode of Communication of Cholera*. 2 ed. Londres: John Churchill, 1855. Disponível em: <https://johnsnow.matrix.msu.edu/work.php?id=15-78-52>. Acesso em 22 out. 2024.

SORIA, M. Llamedo; MARTÍNEZ, JP. Analysis of multidomain features for ECG classification. In: 36th Annual Computers in Cardiology Conference (CinC), Park City, UT, USA, 2009. Disponível em: <https://ieeexplore.ieee.org/document/5445344>. Acesso em: 17 out. 2024.

SOUTO, Bernardino Geraldo Alves. Manual para profissionais da atenção primária de saúde e material de apoio para estudantes de cursos de eletrocardiografia. 1 ed. UFSC: São Carlos, 2016.

SRIVASTAVA, Apoorva *et al.* A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG. *Physiological Measurement*, Vol. 43, n. 6, 2022. Disponível em: <https://doi.org/10.1088/1361-6579/ac6f40>. Acesso em: 15 jan. 2025.

SZEGEDY, Christian *et al.* Rethinking the Inception Architecture for Computer Vision. *Computer Science*, 2014. Disponível em: <https://doi.org/10.48550/arXiv.1409.4842>. Acesso em: 17 out. 2024.

SZNAJDER, Michał; ŁUKOWSKA, Marta. Python Online and Offline ECG QRS Detector based on the Pan-Tompkins algorithm. Zenodo, 2017. Disponível em: <https://doi.org/10.5281/zenodo.826614>. Acesso em: 17 out. 2024.

TANG, Chaosheng *et al.* GAM-SpCaNet: Gradient awareness minimization-based spinal convolution attention network for brain tumor classification. *Journal of King Saud University - Computer and Information Sciences*, Vol. 35, n. 2, 2023. Disponível em: <https://doi.org/10.1016/j.jksuci.2023.01.002>. Acesso em: 17 out. 2024.

TANIOKA, Satoru *et al.* Prediction of hematoma expansion in spontaneous intracerebral hemorrhage using a multimodal neural network. *Scientific Reports*, Vol. 14, 2024. Disponível em: <https://doi.org/10.1038/s41598-024-67365-3>. Acesso em: 17 out. 2024.

THILL, Markus *et al.* Temporal convolutional autoencoder for unsupervised anomaly detection in time series. *Applied Soft Computing*, Vol. 112, 2021. Disponível em: <https://doi.org/10.1016/j.asoc.2021.107751>. Acesso em 23 out. 2024.

TING, Kai Ming. Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA, 2011. Disponível em: [https://doi.org/10.1007/978-0-387-30164-8\\_157](https://doi.org/10.1007/978-0-387-30164-8_157). Acesso em: 17 out. 2024.

TOĞAÇAR, Mesut; ERGEN, Burhan; CÖMERT, Zafer. BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Medical Hypotheses*, Vol. 134, 2020. Disponível em: <https://doi.org/10.1016/j.mehy.2019.109531>. Acesso em: 17 out. 2024.

TSAO, Connie W. *et al.* Heart Disease and Stroke Statistics—2023 Update: A Report From the American Heart Association. *Circulation*, Vol. 145, n. 8, 2023. Disponível em: <https://doi.org/10.1161/CIR.0000000000001052>. Acesso em: 17 out. 2024.

ULLAH, Amin *et al.* Classification of Arrhythmia by Using Deep Learning with 2-D ECG Spectral Image Representation. *Remote Sensing*, Vol. 12, 2020, 12. Disponível em: <https://doi.org/10.3390/rs12101685>. Acesso em: 22 out. 2024.

VANEGHI, Fatemeh Molaei *et al.* A Comparative Approach to ECG Feature Extraction Methods. Proceedings of Third International Conference on Intelligent Systems Modelling and Simulation, Kota Kinabalu, Malaysia, 2012. Disponível em: <https://doi.org/10.1109/ISMS.2012.35>. Acesso em: 17 out. 2024.

VASWANI, Ashish *et al.* Attention is All you Need. In: NeurIPS Poceedings, 2017, Long Beach. Neural Information Processing Systems 30, Long Beach: Curran Associated Inc., 2017. ISBN 978-1-51086096-4. Disponível em: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf). Acesso em: 21 jan. 2025.

VICTOR, Johnson Olanrewaju *et al.* A Cost-Based Dual ConvNet-Attention Transfer Learning Model for ECG Heartbeat Classification. Journal of Informatics and Web Engineering, Vol. 2, n. 2, 2023. Disponível em: <https://doi.org/10.33093/jiwe.2023.2.2.7>. Acesso em: 20 out. 2024.

WAJID, Mohd Anasa *et al.* Neutrosophic-CNN-based image and text fusion for multimodal classification. Journal of Intelligent & Fuzzy Systems, Vol. 45, n. 1, 2023. Disponível em: <https://doi.org/10.3233/JIFS-223752>. Acesso em: 17 out. 2024.

WAN, Dahang *et al.* Mixed local channel attention for object detection Author links open overlay panel. Engineering Applications of Artificial Intelligence, Vol. 123, Part C, 2023. Disponível em: <https://doi.org/10.1016/j.engappai.2023.106442>. Acesso em: 15 jan. 2025.

WANG, Delu *et al.* Forecasting power demand in China with a CNN-LSTM model including multimodal information. Energy, Vol. 263, Part E, 2023. Disponível em: <https://doi.org/10.1016/j.energy.2022.126012>. Acesso em: 17 out. 2024.

WANG, Jilong *et al.* Adversarial de-noising of electrocardiogram. Neurocomputing, Vol. 349, n. 15, 2019a. Disponível em: <https://doi.org/10.1016/j.neucom.2019.03.083>. Acesso em: 17 out. 2024.

WANG, Pu *et al.* ECG Arrhythmias detection using auxiliary classifier generative adversarial network and residual network. IEEE Access, Vol. 7, 2019b. Disponível em: <https://doi.org/10.1109/ACCESS.2019.2930882>. Acesso em: 17 out. 2024.

WANG, Tao el al. Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network. Entropy (Basel), Vol. 23, n. 1, 2021. Disponível em: <https://doi.org/10.3390/e23010119>. Acesso em: 17 out. 2024.

WANG, Will Ke *et al.* A Systematic Review of Time Series Classification Techniques Used in Biomedical Applications. Sensors, Vol. 22, n. 20, 2022. Disponível em: <https://doi.org/10.3390/s22208016>. Acesso em: 17 out. 2024.

WANG, Yue *et al.* Arrhythmia classification algorithm based on multi-head self-attention mechanism. Biomedical Signal Processing and Control, Vol. 79, Part 2, 2023b. Disponível em: <https://doi.org/10.1016/j.bspc.2022.104206>. Acesso em: 15 jan. 2025.

WANG, Zhiguang; OATES, Tim. Spatially Encoding Temporal Correlations to Classify Temporal Data Using Convolutional Neural Networks. *Journal of Computer and Systems Sciences*, 2015. Disponível em: <https://doi.org/10.48550/arxiv.1509.07481>. Acesso em: 17 out. 2024.

WANG, Zhiguang; YAN, Weizhong; OATES, Tim. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. *Computer Science*, 2015. Disponível em: <https://doi.org/10.48550/arXiv.1611.06455>. Acesso em: 17 out. 2024.

WEN, Zeyi *et al.* ThunderSVM: A Fast SVM Library on GPUs and CPUs. *Journal of Machine Learning Research*, Vol. 19, 2018. Disponível em: <https://www.jmlr.org/papers/v19/17-740.html>. Acesso em: 17 out. 2024.

WHO – World Health Organization. Cardiovascular diseases, 2021. Disponível em: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Acesso em: 9 out. 2024.

XIA, Yong *et al.* Detecting atrial fibrillation by deep convolutional neural networks. *Computers in Biology and Medicine*, Vol. 93, n. 1, 2018. Disponível em: <https://doi.org/10.1016/j.compbiomed.2017.12.007>. Acesso em: 17 out. 2024.

XIANG, Yande; LIN, Zhitao; MENG, Jianyi. Automatic QRS complex detection using two-level convolutional neural network. *BioMedical Engineering OnLine*, Vol. 17, n. 13, 2018. Disponível em: <https://doi.org/10.1186/s12938-018-0441-4>. Acesso em: 17 out. 2024.

XIAO, Qiao *et al.* Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review. *Applied Sciences*, Vol. 13, 2023, 13. Disponível em: <https://doi.org/10.3390/app13084964>. Acesso em: 10 mar. 2024.

XIE, Qiyang *et al.* ECG Signal Detection and Classification of Heart Rhythm Diseases Based on ResNet and LSTM. *Advances in Mathematical Physics*, Vol. 2021, 2021. Disponível em: <https://doi.org/10.1155/2021/5862017>. Acesso em: 22 out. 2024.

XU, Xue; JEONG, Sohyun; LI, Jianqiang. Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM. *IEEE Access*, Vol. 8, 2020. Disponível em: <https://doi.org/10.1109/ACCESS.2020.3006707>. Acesso em: 17 out. 2024.

XU, Yanbo *et al.* Recurrent Attentive and intensive model of multimodal patient monitoring data, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, 2018. Disponível em: <https://doi.org/10.48550/arXiv.1807.08820>. Acesso em: 17 out. 2024.

YANG, Kaiyu *et al.* A Study of Face Obfuscation in ImageNet. *Computer Science*, 2021. Disponível em: <https://doi.org/10.48550/arXiv.2103.06191>. Acesso em: 17 out. 2024.

YOUSUF, Asim *et al.* Inferior Myocardial Infarction Detection From Lead II of ECG: A Gramian Angular Field-Based 2D-CNN Approach Sensor applications, Vol. 8, n. 10, 2024. Disponível em: <https://doi.org/10.1109/LSSENS.2024.3450176>. Acesso em: 20 out. 2024.

- YUN, Sangseok *et al.* Deep Learning-Based Ground Vibration Monitoring: Reinforcement Learning and RNN–CNN Approach. *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, 2021. Disponível em: <https://doi.org/10.1109/LGRS.2021.3067974>. Acesso em: 17 out. 2024.
- ZAHID, Muhammad Uzair; KIRANYAZ, Serkan; GABBOUJ, Moncef. Global ECG Classification by Self-Operational Neural Networks with Feature Injection. *IEEE Transactions on Biomedical Engineering*, Vol. 70, n. 1, 2023. Disponível em: <https://doi.org/10.1109/TBME.2022.3187874>. Acesso em: 17 out. 2024.
- ZHAI, Xiaolong; TIN, Chung. Automated ECG Classification using Dual Heartbeat Coupling based on Convolutional Neural Network. *IEEE Access*, Vol. 6, 2018. Disponível em: <https://doi.org/10.1109/ACCESS.2018.2833841>. Acesso em: 17 out. 2024.
- ZHANG, Fuchun *et al.* Multi-classification method of arrhythmia based on multi-scale residual neural network and multi-channel data fusion. *Frontiers in Physiology*, Vol. 14, 2023. Disponível em: <https://doi.org/10.3389/fphys.2023.1253907>. Acesso em: 17 out. 2024.
- ZHANG, Pengfei *et al.* Psychological Stress Detection According to ECG Using a Deep Learning Model with Attention Mechanism. *Applied Sciences*, Vol. 11, n. 6, 2021. Disponível em: <https://doi.org/10.3390/app11062848>. Acesso em: 15 jan. 2025.
- ZHANG, Zhancheng *et al.* Heartbeat classification using disease-specific feature selection. *Computers in Biology and Medicine*, Vol. 46, 2014. Disponível em: <https://doi.org/10.1016/j.compbiomed.2013.11.019>
- ZHAO, Kunkun *et al.* Intra-Subject and Inter-Subject Movement Variability Quantified with Muscle Synergies in Upper-Limb Reaching Movements. *Biomimetics (Basel)*, 2021. Disponível em: <https://doi.org/10.3390/biomimetics6040063>. Acesso em: 17 out. 2024.
- ZHENG, Yi *et al.* Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks. In: *Web-Age Information Management – WAIM 2014*, Macao, China. *Lecture Notes in Computer Science*, Vol. 8485. Springer, Cham, 2014. Disponível em: [https://doi.org/10.1007/978-3-319-08010-9\\_33](https://doi.org/10.1007/978-3-319-08010-9_33). Acesso em: 17 out. 2024.
- ZHOU, Feiyan; FANG, Duanshu. Multimodal ECG heartbeat classification method based on a convolutional neural network embedded with FCA. *Scientific Reports*, Vol. 14, 2024. Disponível em: <https://doi.org/10.1038/s41598-024-59311-0>. Acesso em: 17 out. 2024.
- ZIDELMAL, Zahia *et al.* QRS detection based on wavelet coefficients. *Computer Methods and Programs in Biomedicine*, Vol. 107, n. 3, 2012. Disponível em: <https://doi.org/10.1016/j.cmpb.2011.12.004>. Acesso em: 17 out. 2024.