



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

**CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS VIA REDE NEURAL
CONVOLUCIONAL COM MECANISMO DE ATENÇÃO LOCAL**

DANIEL DANTAS DO AMARAL RAMOS

DM 27/2025

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

DANIEL DANTAS DO AMARAL RAMOS

**CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS VIA REDE NEURAL
CONVOLUCIONAL COM MECANISMO DE ATENÇÃO LOCAL**

Dissertação submetida à Banca Examinadora
do Programa de Pós-Graduação em Engenharia
Elétrica da UFPA para obtenção do Grau de
Mestre em Engenharia Elétrica na Área de
Inteligência Computacional.

Orientadora: **Prof.^a Dr.^a Adriana Rosa Garcez Castro**

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

2025

Catálogo na Publicação (CIP)

Ficha Catalográfica feita pelo autor

Sobrenome, Nome, 1900-
Título/Nome Sobrenome. - Local: Editora, Ano.
123 p.;21 cm.

ISBN: 12-34567-89-0

1. Assunto geral.2.Assunto específico.
1. Título.

CDD: B123.45
CDU:123.456./(89)-0

Gere a ficha catalográfica da UFPA em: <https://bcficat.ufpa.br>



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

CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS VIA REDE NEURAL CONVOLUCIONAL COM MECANISMO DE ATENÇÃO LOCAL

AUTOR: DANIEL DANTAS DO AMARAL RAMOS

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA APROVADA PARA OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE INTELIGÊNCIA COMPUTACIONAL.

APROVADA EM: 22/09/2025

BANCA EXAMINADORA:

Prof.^a Dr.^a Adriana Rosa Garcez Castro
(Orientadora - PPGEE / UFPA)

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

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

VISTO:

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

Dedico este trabalho àqueles que, ditosamente, continuam sendo meus maiores incentivadores, Ismael, Maria do Socorro, Livia, Diógenes e Angélica.

Agradecimentos

Agradeço aos meus pais, Ismael e Maria do Socorro por todo amor e companheirismo. Sem vocês eu não estaria aqui! Agradeço à minha irmã Livia Dantas, que me acompanhou e acompanha na jornada da vida, e por quem nutro o mais puro carinho. Agradeço também à minha segunda mãe, Ana Alice, por ter me dado amor incondicional e todo carinho do mundo. Agradeço aos meus padrinhos Diógenes e Dione, por quem tenho tamanha admiração, por todas as trocas, risadas, conselhos e incentivos. Tenho certeza de que sou quem sou por causa da influência de vocês! Agradeço à minha noiva Angélica Alcântara, que me dá forças e incentivo todo dia, que me acompanha na jornada da vida como mulher, guia, ponto de partida e destino. Seja como for, estarei por ti. Seja onde for, estarei contigo. Agradeço à minha nova família: Amanda, Adriana, Lindalvo e Allan. Com a Angélica eu ganhei um novo lar cheio de amor e boas risadas. Agradeço aos meus amigos e companheiros Isac, Mateus, Thiago e Pedro. A vida é feita de momentos compartilhados, sou grato por compartilhar os de vocês! Agradeço aos meus sócios Arthur e Iran, que embarcaram em diversos desafios comigo, e diariamente me alegram e incentivam. Agradeço também aos meus sócios Heleno, Amanda, Gabriella, Antonny e Edey, que acreditam e apostam em mim, sou muito grato por compartilhar vitórias e experiências com vocês! Agradeço especialmente aos meus mestres e mentores, amigos e orientadores: Prof. Adriana Castro e Prof. Ítalo Di Paolo. Certamente, sem a dedicação, altruísmo e boa vontade de vocês, nada disso seria possível. Tenho orgulho de fazer parte das suas vidas acadêmicas e ter tido a oportunidade de produzir conhecimento junto de vocês! Por fim, à toda minha família e amigos, aos quais não pude citar nominalmente, saibam que cada um tem um espaço especial em meu coração e influenciam em minhas conquistas.

A todos, meus sinceros agradecimentos.

“O pessimista reclama do vento, o otimista espera que ele mude, o realista ajusta as velas”
(Desconhecido)

Resumo

As arritmias cardíacas representam alterações no ritmo ou frequência dos batimentos do coração e estão associadas a riscos significativos à saúde. O eletrocardiograma (ECG) permanece como o principal exame não invasivo para seu diagnóstico, sendo que a análise manual de longos registros é trabalhosa e sujeita a erros. Considerando as dificuldades provenientes da análise manual dos sinais de ECG, sistemas automáticos de classificação de arritmias baseados em inteligência artificial vêm sendo propostos na literatura como alternativa promissora para apoio ao diagnóstico médico. Dentro deste contexto, este trabalho apresenta a proposta de um classificador automático de arritmias cardíacas baseado em Redes Neurais Convolucionais (CNN - Alexnet), com mecanismo de atenção local integrado a sua arquitetura, desenvolvido considerando o paradigma interpaciente. O estudo se concentrou na investigação e definição da técnica mais adequada, entre Gramian Angular Field (GAF) e Hilbert Space-Filling (HSFC), para a conversão dos sinais temporais de ECG em imagens, para serem utilizadas como entrada para a CNN e, em avaliar o impacto de diferentes configurações internas do mecanismo de atenção local proposto, especificamente no que diz respeito à escolha da função de ativação (Tangente Hiperbólica ou Sigmóide) e do tipo de kernel (fixo ou adaptativo). Para os experimentos foi utilizada a base de dados MIT-BIH Arrhythmia Database, sendo que o desempenho dos modelos treinados foi avaliado segundo medidas de desempenho padronizadas, como acurácia, precisão, sensibilidade, especificidade e F1-score. Os experimentos mostraram que a combinação GAF com função de ativação Tangente Hiperbólica e kernel adaptativo no módulo de atenção apresentou o melhor resultado, alcançando acurácia de 97,88

Palavras-chave: Eletrocardiograma (ECG); Redes Neurais Convolucionais; Mecanismo de atenção; Transformação de sinais em imagens.

Abstract

Cardiac arrhythmias represent alterations in the rhythm or frequency of heartbeats and are associated with significant health risks. The electrocardiogram (ECG) remains the primary non-invasive examination for their diagnosis; however, the manual analysis of long recordings is labor-intensive and prone to errors. Considering the challenges arising from manual ECG signal analysis, automatic arrhythmia classification systems based on artificial intelligence have been proposed in the literature as a promising alternative to support medical diagnosis. Within this context, this work presents the proposal of an automatic cardiac arrhythmia classifier based on Convolutional Neural Networks (CNN – AlexNet), with an integrated local attention mechanism in its architecture, developed under the inter-patient paradigm. The study focused on investigating and defining the most suitable technique, between Gramian Angular Field (GAF) and Hilbert Space-Filling Curve (HSFC), for converting ECG temporal signals into images to be used as input for the CNN, and on evaluating the impact of different internal configurations of the proposed local attention mechanism, specifically regarding the choice of activation function (Hyperbolic Tangent or Sigmoid) and kernel type (fixed or adaptive). For the experiments, the MIT-BIH Arrhythmia Database was used, and the performance of the trained models was evaluated according to standardized metrics, such as accuracy, precision, sensitivity, specificity, and F1-score. The experiments showed that the combination of GAF with the Hyperbolic Tangent activation function and adaptive kernel in the attention module achieved the best result, reaching an accuracy of 97.88% and an F1-score of 0.7678. The model outperformed the baseline CNN (AlexNet) architecture without the attention module and demonstrated competitive performance compared to solutions previously presented in the literature, even without the application of additional data balancing techniques.

Key-words: Electrocardiogram (ECG); Convolutional Neural Networks (CNNs); Attention Mechanism; Signal-to-Image Transformation

Lista de figuras

Figura 1 – Sinal de ECG indicando as ondas do complexo QRS.	5
Figura 2 – Posicionamento dos eletrodos para aquisição de ECG.	8
Figura 3 – Curva de Hilbert	19
Figura 4 – Curva de Hilbert: Ordens	20
Figura 5 – Estrutura de CNN	21
Figura 6 – Operação de convolução	23
Figura 7 – Estrutura <i>Self-Attention</i> local	29
Figura 8 – Funções de ativação: Sigmoid vs. Tanh	31
Figura 9 – Estrutura proposta neste trabalho	37
Figura 10 – Segmentação do sinal ECG	39
Figura 11 – Transformação em imagens de segmentos de sinal ECG para cada classe . .	41
Figura 12 – Classificador com mecanismo de atenção	44
Figura 13 – Matriz de confusão para o melhor cenário com módulo de atenção	53
Figura 14 – Matrizes de confusão - AlexNet base com HSFC e GAF	53

Lista de tabelas

Tabela 1 – Indicadores de desempenho de trabalhos correlatos para o paradigma interpaciente.	36
Tabela 2 – Distribuição por Classes na Base MIT-BIH (Paradigma Interpaciente)	40
Tabela 3 – Estrutura das camadas convolucionais	43
Tabela 4 – Indicadores de desempenho globais na base de teste: Estrutura com mecanismo de atenção	51
Tabela 5 – Indicadores de desempenho por classe na base de teste: Estrutura com mecanismo de atenção	51
Tabela 6 – Indicadores de desempenho globais na base de teste: Modelo base	51
Tabela 7 – Indicadores de desempenho por classe: Modelo base	52
Tabela 8 – Indicadores de desempenho na base de teste de trabalhos correlatos	59

Lista de abreviaturas e siglas

AAMI	Association for the Advancement of Medical Instrumentation
AI	Artificial Intelligence
ANN	Artificial Neural Network
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DL	Deep Learning
DS1	Dataset 1
DS2	Dataset 2
ECG	Electrocardiogram
FN	False Negative
FP	False Positive
GAF	Gramian Angular Field
GNNS	Graph Neural Networks
GPU	Graphics Processing Unit
HSFC	Hilbert Space Filling Curve
IA	Inteligência Artificial
LSTM	Long Short-Term Memory
MIT-BIH	Massachusetts Institute of Technology - Beth Israel Hospital
ML	Machine Learning
MLII	Modified Lead II
PSO	Particle Swarm Optimization
PTB	Physikalisch-Technische Bundesanstalt
QRS	Q-R-S complex

ReLU	Rectified Linear Unit
RNA	Rede Neural Artificial
RNN	Recurrent Neural Network
RP	Recurrence Plot
RR	R-R interval
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UFPA	Universidade Federal do Pará
VP	Verdadeiro Positivo
WGAN-GP	Wasserstein Generative Adversarial Network with Gradient Penalty

Sumário

1	INTRODUÇÃO	1
1.1	Objetivos	2
1.2	Organização do Texto	3
2	FUNDAMENTAÇÃO TEÓRICA	4
2.1	O Coração e o Sinal Cardíaco	4
2.2	Arritmias Cardíacas	9
2.2.1	Arritmias por Frequência	9
2.2.2	Arritmias por Origem Anatômica	9
2.2.3	Arritmias por Condução	9
2.2.4	Outras Arritmias	10
2.3	Sistemas Automáticos para Classificação de Arritmias	10
2.3.1	Diretrizes da AAMI	11
2.3.2	Paradigmas para desenvolvimento dos classificadores	12
2.3.2.1	Paradigma Intrapaciente	12
2.3.2.2	Paradigma Interpaciente	13
2.3.3	Medidas de desempenho	14
2.4	Séries temporais	15
2.4.1	Técnicas para transformação de séries temporais em imagens	16
2.4.1.1	<i>Gramian Angular Fields</i> (GAF)	17
2.4.1.2	Curvas de Hilbert	18
2.5	Redes Neurais Convolucionais	20
2.5.1	A Arquitetura Fundamental de uma CNN	21
2.5.1.1	Camada Convolucional	22
2.5.1.2	Função de Ativação	23
2.5.1.3	Camada de <i>Pooling</i>	24
2.5.1.4	Camada de Achatamento (<i>Flatten</i>)	24
2.5.1.5	Camada Totalmente Conectada	24
2.5.2	Treinamento de redes neurais convolucionais	25
2.5.3	Arquiteturas de CNN para classificação de imagens	26
2.5.4	Mecanismos de atenção em redes CNN	27
2.5.4.1	Taxonomia dos Mecanismos de Atenção	27

2.5.4.2	O Paradigma de Autoatenção (<i>Self-Attention</i>)	28
2.6	A Influência do <i>Kernel</i>	29
2.6.1	Teoria dos <i>Kernels</i>	29
2.6.2	<i>Kernel</i> no Módulo <i>Self-Attention</i>	29
2.6.3	Diferenças entre <i>Kernels</i> Fixo e Adaptativo	30
2.7	A Influência da Função de Ativação	30
2.7.1	Teoria das Funções de Ativação	30
2.7.2	Função de Ativação no Módulo <i>Self-Attention</i>	30
2.7.3	Diferença entre Sigmóide e tanh no Contexto	31
3	ESTADO DA ARTE	33
3.1	Estado da Arte Geral	33
3.2	Trabalhos Correlatos	34
3.2.1	Abordagem Interpaciente	35
4	METODOLOGIA	37
4.1	Banco de Dados e Paradigma Interpaciente	38
4.2	Transformação do Sinal ECG em Imagem	40
4.2.1	Pré-processamento das Imagens	42
4.3	Arquitetura da Rede Neural Proposta	42
4.3.1	Componente de Entrada	42
4.3.2	Componente de Camadas Convolucionais	43
4.3.3	Classificador com Mecanismo de Atenção	44
4.4	Configuração de Experimentos	47
4.5	Medidas de desempenho	48
5	RESULTADOS EXPERIMENTAIS	50
6	ANÁLISE E DISCUSSÃO DOS RESULTADOS	55
6.1	Análise Geral dos Resultados	55
6.1.1	Análise comparativa de modelo base variando transformação de imagem	56
6.1.2	Análise comparativa do modelo base GAF com o melhor modelo com módulo de atenção	57
6.2	Comparação com modelos propostos na literatura	58
7	CONCLUSÃO	61

Referências 64

1 INTRODUÇÃO

O corpo humano é um sistema complexo, com subsistemas interconectados que trabalham em harmonia para manter a saúde e o equilíbrio. Dentre estes, o sistema cardiovascular desempenha um papel crucial ao transportar nutrientes, oxigênio e resíduos metabólicos, sendo vital para a manutenção da homeostase corporal. O coração, atuando 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, em um processo contínuo essencial para as funções celulares e a saúde geral do corpo. No entanto, disfunções nesse sistema podem levar a anomalias e doenças cardíacas, como a doença arterial coronariana, insuficiência cardíaca e cardiomiopatias, com consequências que variam desde a redução da capacidade funcional até, em casos extremos, a morte súbita (HALL, 2021). A compreensão e o desenvolvimento de métodos avançados de diagnóstico e tratamento dessas condições são fundamentais para mitigar seus impactos negativos.

Nesse cenário, o eletrocardiograma (ECG) destaca-se como uma técnica não invasiva e fundamental para registrar os ritmos cardíacos e identificar anomalias, como as arritmias cardíacas, que são caracterizadas por alterações no ritmo ou frequência dos batimentos. Desenvolvido em 1901, o ECG permanece como a abordagem mais prevalente no diagnóstico de enfermidades cardíacas até os dias atuais (GIFFONI; TORRES, 2009). Apesar de sua eficácia, a análise manual de longos registros de ECG, como os obtidos em exames de Holter, é uma tarefa trabalhosa, propensa a erros e que exige alta especialização por parte dos médicos especialistas, devido à necessidade de analisar batimento a batimento ao longo de horas ou dias de registros (Di Paolo; CASTRO, 2024).

Diante desses desafios, sistemas automáticos de classificação surgem como alternativas promissoras para apoiar o diagnóstico médico, oferecendo vantagens como o aumento da eficiência na análise de grandes volumes de dados, a padronização das interpretações e a redução da subjetividade e de erros humanos. Tais sistemas podem, inclusive, identificar padrões sutis que seriam imperceptíveis à análise visual, fornecendo subsídios objetivos para a tomada de decisão clínica e permitindo o monitoramento contínuo (LUZ et al., 2016; HONG et al., 2020).

A evolução da tecnologia e, em particular, da Inteligência Artificial (IA) e do Aprendizado Profundo, tem permitido avanços significativos no desenvolvimento desses sistemas de apoio à decisão em medicina (LUZ et al., 2016; HONG et al., 2020). As Redes Neurais Convolucionais (CNNs) têm se destacado especialmente na análise de sinais de ECG. Projetadas para processar dados com estrutura em grade, como imagens, as CNNs são eficazes tanto para sinais ECG unidimensionais quanto para representações bidimensionais obtidas a partir de transformações de sinais temporais em imagens (ZHOU; FANG, 2024). A principal vantagem das CNNs reside em sua capacidade de aprender automaticamente características relevantes dos dados, sem a

necessidade de extração manual, o que as torna robustas para lidar com a complexidade e as variações presentes nos sinais ECG clínicos reais. A aplicação de CNNs pode levar a diagnósticos mais precisos, monitoramento contínuo, tratamento personalizado, acesso a cuidados de saúde em áreas remotas e redução de custos.

Complementarmente, os mecanismos de atenção surgem como uma técnica inspirada na capacidade humana de focar em partes específicas de uma entrada ao processar informações, permitindo que modelos de aprendizado profundo ponderem dinamicamente a importância de diferentes partes dos dados (Di Paolo, 2025). Quando integrados às arquiteturas de CNNs, esses mecanismos permitem que a rede aprenda a dar maior peso a características ou regiões mais informativas da entrada, ao invés de processar todos os elementos de forma uniforme, potencializando assim o desempenho no processo de classificação. No contexto do diagnóstico de arritmias cardíacas, essa capacidade de atenção seletiva tem se mostrado promissora para destacar padrões discriminativos e refinar a identificação de batimentos anômalos, especialmente em classes minoritárias e complexas como as arritmias supraventriculares (classe S), que são particularmente desafiadoras de classificar, como evidenciado por Di Paolo (2025). A literatura recente e trabalhos correlatos demonstram a eficácia da aplicação desses mecanismos em CNNs para a classificação de arritmias cardíacas, conforme evidenciado por Ahmad et al. (2021), Islam et al. (2023) e Di Paolo e Castro (2024).

Contudo, é fundamental que a interpretação clínica dos resultados desses modelos seja realizada por profissionais de saúde qualificados, uma vez que eles atuam como ferramentas de apoio e não substituem o julgamento médico (Di Paolo; CASTRO, 2024).

1.1 Objetivos

Esta dissertação tem como objetivo geral o estudo e desenvolvimento de um sistema automático de classificação de arritmias cardíacas baseado em uma Rede Neural Convolucional (CNN) com mecanismo de atenção local para auxílio ao diagnóstico médico, considerando a abordagem metodológica interpaciente para a validação do modelo.

Para alcançar este propósito, os objetivos específicos delineados são:

1. Investigar e definir as técnicas mais adequadas para a conversão dos sinais temporais de ECG em representações visuais (imagens) que servirão como dados de entrada para a arquitetura da rede CNN. Para tanto, serão exploradas e avaliadas duas abordagens: *Gramian Angular Field* (GAF) e Curva de Preenchimento do Espaço de Hilbert (HSFC);
2. Avaliar o impacto de diferentes configurações internas do mecanismo de atenção local proposto, especificamente no que diz respeito à escolha da função de ativação (Tangente Hiperbólica ou Sigmóide) e do tipo de kernel (fixo ou adaptativo). Esta análise visa determinar as combinações que otimizam o desempenho do classificador de arritmias;

3. Conduzir uma análise comparativa dos resultados quantitativos obtidos com a arquitetura proposta em relação a trabalhos correlatos. Esta etapa visa posicionar o desempenho do classificador desenvolvido frente a outras soluções contemporâneas, utilizando medidas de desempenho padronizadas, como Acurácia, Precisão, Sensibilidade (*Recall*), Especificidade e *F1-Score*, para garantir a validade e a relevância das contribuições.

1.2 Organização do Texto

Esta dissertação está organizada em seis Capítulos, além deste Capítulo introdutório de contextualização.

O **Capítulo 2** (Fundamentação Teórica) detalha os conceitos essenciais para o entendimento do trabalho, abrangendo a anatomia e fisiologia cardíaca, a caracterização das arritmias, os sistemas automáticos para classificação de arritmias (com foco nas diretrizes da AAMI e no paradigma interpaciente), e as medidas de desempenho empregadas (Acurácia, Precisão, Sensibilidade, Especificidade e *F1-Score*). Além disso, aborda o processamento de séries temporais e as técnicas de transformação em imagens, bem como a arquitetura e funcionamento das CNNs e dos mecanismos de autoatenção local, explorando a influência do kernel e das funções de ativação no seu desempenho.

O **Capítulo 3** (Estado da Arte) apresenta uma revisão abrangente da literatura científica, delineando as tendências atuais e os desafios na classificação de arritmias cardíacas por aprendizado profundo, com especial atenção aos trabalhos correlatos que empregam a abordagem interpaciente.

O **Capítulo 4** (Metodologia) detalha a utilização da base de dados MIT-BIH no paradigma interpaciente, as técnicas de transformação do sinal ECG em imagem (*Gramian Angular Field* - GAF e Curva de Preenchimento do Espaço de Hilbert - HSFC), o pré-processamento das imagens, a arquitetura da rede neural proposta, a configuração dos experimentos e as medidas de desempenho.

O **Capítulo 5** (Resultados Experimentais) apresenta os achados quantitativos obtidos nas configurações testadas, incluindo as comparações com os modelos base, por meio de tabelas e matrizes de confusão, para ilustrar o desempenho do classificador.

O **Capítulo 6** (Análise e Discussão dos Resultados) interpreta o desempenho do modelo proposto, comparando as diferentes configurações e o impacto do mecanismo de atenção, posicionando os resultados frente aos trabalhos da literatura.

Finalmente, o **Capítulo 7** (Conclusão) sintetiza os principais achados, apresenta as contribuições da pesquisa e delineia as direções para trabalhos futuros, com o intuito de aprimorar ainda mais o classificador de arritmias cardíacas.

2 FUNDAMENTAÇÃO TEÓRICA

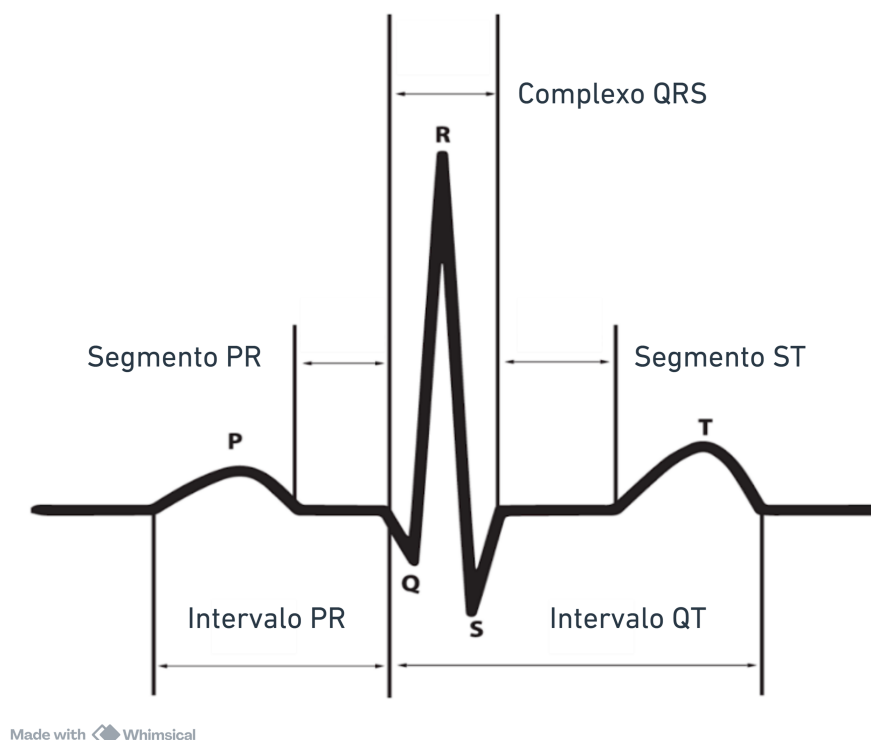
2.1 O Coração e o Sinal Cardíaco

O corpo humano é um sistema complexo composto por subsistemas interconectados que atuam em harmonia para preservar a saúde e o equilíbrio. Dentre esses, o sistema cardiovascular desempenha uma função primordial ao transportar nutrientes, oxigênio e resíduos metabólicos por todo o organismo, sendo vital para a manutenção da homeostase corporal, das funções celulares e do metabolismo geral. O coração, operando como uma bomba muscular, é responsável por impulsionar o sangue rico em oxigênio para os tecidos e órgãos e por direcionar o sangue desoxigenado de volta aos pulmões para reoxigenação, em um processo contínuo essencial para a vida. Disfunções nesse sistema podem originar diversas anomalias e doenças cardíacas, desde arritmias comuns até condições mais severas, como a doença arterial coronariana, insuficiência cardíaca e cardiomiopatias, cujas consequências podem variar desde a redução da capacidade funcional até eventos graves como acidente vascular cerebral e, em situações extremas, morte súbita (HALL, 2021).

Anatomicamente, o coração é estruturado em quatro câmaras: dois átrios (ou aurículas) e dois ventrículos. Os átrios são as câmaras superiores que recebem o sangue: o átrio direito recebe o sangue desoxigenado do corpo pela veia cava superior, enquanto o átrio esquerdo recebe o sangue oxigenado dos pulmões pela veia pulmonar. Os ventrículos são as câmaras inferiores que bombeiam o sangue para fora do coração: o ventrículo direito impulsiona o sangue para os pulmões através da artéria pulmonar, e o ventrículo esquerdo bombeia o sangue oxigenado para o restante do corpo via aorta (HALL, 2021).

A atividade elétrica do coração é registrada graficamente no ECG por meio de ondas e complexos, que representam as diferenças de potencial entre os eletrodos positivos e negativos em cada derivação. O ECG também exibe importantes intervalos e segmentos que oferecem informações diagnósticas adicionais. A Figura 1 apresenta as características de um sinal cardíaco que podem ser visualizadas através do sinal de ECG.

Figura 1 – Sinal de ECG indicando as ondas do complexo QRS.



Fonte: Autor. Adaptado de (HSIEH et al., 2020)

As características do sinal cardíaco são:

- **Ondas:** São deflexões, sejam positivas ou negativas, da linha de base do traçado, cada uma correspondendo a eventos elétricos específicos durante o ciclo cardíaco (BETTS et al., 2022):
 - **Onda P:** Representa a despolarização atrial, que é a contração dos átrios.
 - **Onda Q:** É a primeira deflexão negativa que ocorre após a onda P.
 - **Onda R:** Caracteriza-se como a primeira deflexão positiva, geralmente alta e proeminente, após a onda Q, e representa a despolarização ventricular.
 - **Onda S:** É a deflexão negativa que se segue à onda R e antecede a onda T, também associada à despolarização ventricular.
 - **Onda T:** Indica a repolarização ventricular, fase em que as câmaras cardíacas se preparam para a próxima contração.
 - **Onda U:** Uma pequena onda que pode aparecer após a onda T e antes da onda P.
- **Complexos:** São conjuntos de ondas que ocorrem em conjunto e representam fases específicas do ciclo cardíaco (HALL, 2021; BETTS et al., 2022):

- **Complexo QRS:** É formado pelas ondas Q, R e S, e reflete a despolarização ventricular (a contração dos ventrículos). A análise do complexo QRS é essencial para avaliar a função cardíaca, diagnosticar arritmias e identificar possíveis distúrbios na condução elétrica do coração.
- **Intervalos e segmentos:** Assim como os complexos e ondas, os intervalos e segmentos são caracterizações que oferecem informações diagnósticas adicionais (Di Paolo, 2025):
 - **Intervalo RR:** Corresponde à medida da distância entre dois complexos QRS consecutivos, representando o tempo transcorrido entre dois batimentos cardíacos sucessivos. É um indicador crucial para a análise da frequência cardíaca e para a identificação de arritmias e outros distúrbios do ritmo.
 - **Intervalo PR:** Este período 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 utilizado para avaliar atrasos ou bloqueios na transmissão elétrica entre os átrios e os ventrículos.
 - **Intervalo QT:** Este período abrange 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.
 - **Segmento PR:** Este segmento do ECG é observado após a onda P até a onda R, e representa o tempo de condução elétrica do nódulo atrioventricular. Ele indica a duração da condução do impulso cardíaco entre as câmaras atriais e ventriculares e pode ser empregado para identificar anormalidades na condução cardíaca.
 - **Segmento ST:** Este segmento do ECG se localiza após a onda S e antecede a onda T. Ele representa o período de repolarização ventricular (a recuperação dos ventrículos após a contração). Alterações no segmento ST podem indicar isquemia ou lesão cardíaca e são utilizadas na avaliação de infarto do miocárdio e outras condições cardíacas.

A atividade elétrica do coração é fundamental para coordenar suas contrações rítmicas e regulares, conhecidas como batimentos cardíacos. Esse sistema elétrico compreende células especializadas que geram e conduzem os sinais elétricos por todo o músculo cardíaco (CLIFFORD; AZUAJE; MCSHARRY, 2006; HALL, 2021). As estruturas-chave desse processo são:

- **Nó Sinoatrial (SA):** Localizado no átrio direito, funciona como o marcapasso natural do coração, iniciando os impulsos elétricos que se propagam pelos átrios direito e esquerdo, causando sua contração e o bombeamento de sangue para os ventrículos. No eletrocardiograma (ECG), a onda P reflete essa despolarização atrial.
- **Nó Atrioventricular (AV):** Posicionado entre os átrios e os ventrículos, este nó atrasa levemente a propagação do impulso elétrico antes de transmiti-lo aos ventrículos. Esse retardo é crucial para garantir que os ventrículos se encham completamente de sangue antes de serem estimulados a contrair. O segmento PR no ECG representa esse tempo de condução entre átrios e ventrículos.
- **Sistema de His-Purkinje (Fibras de Purkinje):** Após atravessar o nó AV, os sinais elétricos são rapidamente conduzidos por uma rede de fibras especializadas que se ramificam pelos ventrículos. Essa rápida condução promove a despolarização e contração simultânea e eficiente dos ventrículos (sístole ventricular), ejetando o sangue para as circulações pulmonar e sistêmica. No ECG, o complexo QRS reflete a despolarização ventricular, e a onda T representa a repolarização ventricular, preparando o coração para o próximo ciclo. Alterações nesse circuito elétrico podem dar origem a diferentes tipos de arritmias, como batimentos acelerados (taquicardia), lentos (bradicardia) ou irregulares.

O Eletrocardiograma (ECG) é uma técnica não invasiva e fundamental para monitorar a atividade elétrica do coração e identificar anomalias como arritmias cardíacas, que são caracterizadas por alterações no ritmo ou frequência dos batimentos. Desenvolvido em 1901 pelo Dr. Willem Einthoven, o ECG permanece a abordagem mais prevalente no diagnóstico de enfermidades cardíacas até os dias atuais (GIFFONI; TORRES, 2009). Ele registra as variações de potencial elétrico por meio de eletrodos fixados na superfície corporal, geralmente no peito, braços e pernas. Esses eletrodos detectam as variações nos sinais elétricos que ocorrem durante as contrações e relaxamentos do coração, e essas informações são registradas graficamente, permitindo aos médicos avaliar a atividade elétrica cardíaca para identificar normalidade ou possíveis problemas.

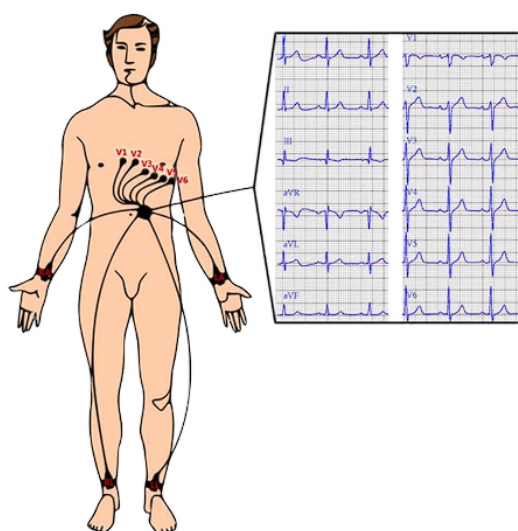
Um esquema simplificado para a aquisição de sinais de ECG envolve eletrodos que realçam a diferença de potencial entre pontos marcados na pele com o auxílio de um amplificador instrumental com isolamento óptica. Em seguida, o sinal é submetido a um filtro passa-alta, um segundo estágio de amplificação, um filtro passa-baixa, e finalmente é amostrado por um conversor analógico-digital, gerando o ECG pré-processado (Di Paolo, 2025).

Os pontos onde os eletrodos são posicionados na pele são denominados derivações, e cada uma oferece uma perspectiva única do sinal cardíaco, possibilitando uma análise detalhada da atividade elétrica do coração. As derivações do ECG, como pode-se observar na Figura 2, são categorizadas em duas classes principais:

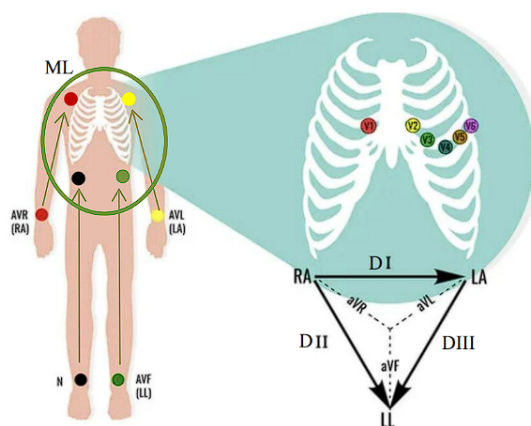
- **Derivações bipolares ou periféricas:** Obtidas por meio de eletrodos positivos e negativos localizados nos membros superiores e inferiores do corpo, como as derivações DI, DII e DIII. As derivações modificadas (ML) são variações das derivações bipolares padrão, adaptadas para monitoramento contínuo de longa duração, proporcionando maior conforto ao paciente. Os eletrodos são realocados para posições equivalentes no tórax, sendo então denominadas MLI, MLII e MLIII.
- **Derivações unipolares ou precordiais:** Utilizam um eletrodo explorador positivo posicionado em pontos específicos do tórax (V1, V2, V3, V4, V5, V6). Elas fornecem uma visão frontal do coração e são valiosas na detecção de problemas cardíacos nas regiões anterior e lateral.

Figura 2 – Posicionamento dos eletrodos para aquisição de ECG.

(a) Pontos onde os eletrodos são posicionados na pele



(b) Pontos específicos do tórax para posicionamento de eletrodos



Fonte: (a) ECG Lead Positioning, 2024. [Online]. Disponível em: <https://litfl.com/ecg-lead-positionin>. (b) Adaptado de (Di Paolo, 2025)

A derivação MLII (*Modified Lead II*) é amplamente empregada para medições de longo período por ser menos sensível ao movimento dos pacientes e por permitir que o complexo QRS e as ondas P e T sejam destacados com maior clareza em comparação com outras derivações (FRANCIS, 2016; LUZ et al., 2016). As derivações V são obtidas a partir de eletrodos posicionados no tórax, registrando os potenciais de ação na musculatura ventricular, o que pode ser vantajoso para a identificação de arritmias da classe V (Ventricular). As derivações são sempre medidas em relação a uma referência neutra, a derivação "zero" ou "terra", usualmente localizada no tornozelo direito (Di Paolo, 2025).

2.2 Arritmias Cardíacas

As arritmias cardíacas representam alterações no ritmo, frequência ou sequência dos batimentos do coração. Essas anomalias são geralmente causadas por falhas na geração ou condução dos impulsos elétricos que coordenam a atividade cardíaca. Disfunções no sistema elétrico interno do coração, que é responsável pela geração rítmica dos batimentos, podem levar a diversas anomalias e doenças cardíacas, com consequências que variam desde a redução da capacidade funcional até, em casos extremos, a morte súbita (HALL, 2021). As arritmias podem surgir tanto nos átrios quanto nos ventrículos.

As arritmias cardíacas podem ser classificadas de diversas formas, dependendo de sua frequência, origem anatômica ou problemas de condução (HALL, 2021):

2.2.1 Arritmias por Frequência

- **Taquicardia:** Caracterizada por uma frequência cardíaca rápida, geralmente acima de 100 batimentos por minuto em adultos. Pode ser causada por fatores como aumento da temperatura corporal, estimulação simpática ou condições tóxicas do coração.
- **Bradicardia:** Ocorre quando a frequência cardíaca se torna lenta, tipicamente abaixo de 60 batimentos por minuto.
- **Arritmia Sinusal:** Um distúrbio do ritmo cardíaco que causa batidas irregulares, mais rápidas ou mais lentas que o normal, com variações frequentemente ligadas ao ciclo respiratório.

2.2.2 Arritmias por Origem Anatômica

- **Supraventriculares:** Originam-se acima dos ventrículos, ou seja, nos átrios ou no nó atrioventricular. São caracterizadas por um ritmo irregular e rápido devido à ativação desorganizada dos átrios, podendo reduzir a eficácia do bombeamento atrial e causar palpitações, cansaço e falta de ar.
- **Ventriculares:** Originam-se nos ventrículos. São consideradas mais graves, pois podem comprometer diretamente a função de bombeamento de sangue do coração e aumentar o risco de evoluírem para fibrilação ventricular, que pode levar à parada cardíaca. Seus sintomas incluem tontura, desmaios e colapso circulatório.

2.2.3 Arritmias por Condução

- **Bloqueios Atrioventriculares:** Ocorrem quando há atraso ou interrupção na condução do impulso elétrico entre os átrios e os ventrículos. No ECG, é caracterizado por um intervalo PR prolongado. Podem ser assintomáticos ou causar tonturas em casos avançados.

- **Bloqueios de Ramo:** Falha na condução dos impulsos nos ramos direito ou esquerdo dos ventrículos, podendo causar desconforto no peito, fadiga e falta de ar.

2.2.4 Outras Arritmias

- **Extrasístoles:** São batimentos cardíacos "extras" ou prematuros, que podem se originar nos átrios (extrasístoles atriais) ou nos ventrículos (extrasístoles ventriculares), frequentemente causando a sensação de batimentos fortes no peito.
- **Síndrome do QT Longo:** Um distúrbio que prolonga o tempo de repolarização ventricular, refletido por um intervalo QT prolongado no ECG, e que pode causar palpitações e desmaios.

2.3 Sistemas Automáticos para Classificação de Arritmias

A interpretação manual de sinais de eletrocardiograma (ECG) para a detecção de arritmias cardíacas é uma tarefa que demanda alto nível de especialização e atenção, tornando-se exaustiva e propensa a erros em face do vasto volume de dados gerado por exames prolongados. Nesse cenário desafiador, os sistemas automáticos de classificação emergem como soluções promissoras, oferecendo uma série de vantagens cruciais que beneficiam tanto os pacientes quanto os profissionais de saúde. Estes sistemas não apenas aceleram a análise de grandes volumes de dados, mas também padronizam as interpretações, reduzem a subjetividade e erros humanos, e são capazes de identificar padrões sutis que poderiam passar despercebidos à análise visual. Tal apoio é fundamental para a tomada de decisão clínica, potencializando o monitoramento contínuo e a detecção precoce de anomalias cardíacas (Di Paolo; CASTRO, 2024).

Um fluxo típico para a classificação automática de arritmias cardíacas envolve uma sequência de etapas interconectadas, que garantem a preparação e análise eficaz do sinal de ECG. Primeiramente, o pré-processamento é crucial para a remoção de artefatos e ruídos indesejados do sinal, melhorando sua qualidade por meio de técnicas como filtragem passa-baixa e passa-alta, remoção da linha de base, cancelamento de artefatos musculares, normalização e digitalização. Em seguida, a segmentação foca na identificação e isolamento de componentes chave do ECG, como o complexo QRS, ondas P e T, e intervalos, que são representações vitais da atividade elétrica cardíaca. Essa etapa é um desafio devido à variabilidade dos sinais, mas é essencial para avaliar a função cardíaca e diagnosticar arritmias (Di Paolo, 2025). Após a segmentação, a extração de características visa identificar e quantificar atributos relevantes do sinal que permitam diferenciar os diversos tipos de arritmias. Estes podem ser analisados tanto no domínio do tempo, através de medidas como a duração do complexo QRS, intervalos PR e QT, e intervalos RR, que fornecem informações sobre frequência cardíaca e condução elétrica, ou no domínio da frequência, explorando propriedades espectrais do sinal, como frequências dominantes, banda de energia e coeficientes de transformada de Fourier (SINGH; KRISHNAN, 2023). A etapa

final é a classificação, onde um modelo ou algoritmo é aplicado para identificar e categorizar os diferentes tipos de arritmias com base nas características extraídas.

Diversos métodos de classificação podem ser empregados para essa finalidade, cada um com suas particularidades (Di Paolo, 2025). As Máquinas de Vetores de Suporte (SVM), por exemplo, buscam otimizar a separação entre classes através de um hiperplano que maximiza a margem entre os pontos de dados. Estudos como os de Zhang et al. (2014), Garcia et al. (2017) e Xu, Jeong e Li (2020) têm demonstrado sua aplicação bem-sucedida. As Redes Neurais Recorrentes (RNNs) são especialmente adequadas para lidar com sequências de dados como sinais temporais, capturando dependências de longo prazo. Islam et al. (2023) aplicaram RNNs para a classificação de arritmias. Já as Redes Neurais Convolucionais (CNNs), um dos focos deste estudo, são classificadores robustos para imagens e dados com estrutura espacial, destacando-se pela capacidade de aprender características relevantes automaticamente a partir dos dados de treinamento, sem a necessidade de extração manual. Sua eficácia na análise de sinais ECG, mesmo com variações e ruídos, as torna uma escolha promissora. Trabalhos de Dias et al. (2021), Ahmad et al. (2021), Mewada (2023) e Zhou e Fang (2024) ilustram a ampla utilização das CNNs. Outras abordagens incluem Lógica Fuzzy, que lida com incertezas e imprecisões nos dados, como abordado em Frahan, Li e Ahmed (2018) e Redes Bayesianas, que modelam relações e fornecem probabilidades de classificação, estudadas por Casas et al. (2018).

2.3.1 Diretrizes da AAMI

A aplicação de sistemas automáticos para a classificação de arritmias cardíacas é um campo de pesquisa ativo, visando auxiliar profissionais de saúde no diagnóstico. Para garantir a confiabilidade e comparabilidade desses sistemas, é fundamental a adoção de padrões e diretrizes reconhecidas. Nesse contexto, a Associação para o Avanço da Instrumentação Médica (AAMI) é uma organização de destaque que estabelece diretrizes e padrões para a classificação automática de sinais de ECG (Di Paolo, 2025).

A AAMI desenvolveu o conjunto de recomendações conhecido como Diretrizes da AAMI para a Classificação de Eletrocardiogramas (ANSI/AAMI EC57). Este padrão fornece orientações sobre a terminologia, os critérios de desempenho e as medidas de desempenho para sistemas de classificação automática de arritmias cardíacas. A sua criação e adoção visam, portanto, padronizar a terminologia e a abordagem de classificação, o que é essencial para permitir a comparação de resultados entre diferentes sistemas e facilitar a interpretação clínica dos resultados, contribuindo para o desenvolvimento de sistemas mais confiáveis e precisos (Di Paolo, 2025).

Essas diretrizes de padrões e terminologias para a classificação automática de sinais de ECG, define cinco classes principais de arritmias. Essas classes são amplamente utilizadas para padronizar o desenvolvimento e a avaliação de sistemas de classificação automática de arritmias cardíacas (Di Paolo, 2025), sendo elas:

- **Classe N (Normal):** Inclui batimentos cardíacos normais, sem anormalidades significativas. Isso abrange o complexo QRS normal, ondas P e T normais, e intervalos PR e QT dentro dos limites esperados.
- **Classe S (Supraventricular Ectópica - SVEB):** Compreende arritmias que se originam acima dos ventrículos, como taquicardia sinusal, taquicardia atrial, fibrilação atrial, vibração atrial e bloqueio atrioventricular.
- **Classe V (Ventricular Ectópica - VEB):** Abrange arritmias que se originam nos ventrículos, incluindo taquicardia ventricular, fibrilação ventricular e complexos ventriculares prematuros.
- **Classe F (Fusão):** Refere-se a arritmias que apresentam componentes tanto do complexo QRS normal quanto do complexo QRS ventricular ectópico (anormal), resultando em uma aparência intermediária entre as duas formas.
- **Classe Q (Desconhecido):** Reservada para casos em que a classificação não pode ser definitivamente determinada ou quando há dúvidas na interpretação do sinal de ECG.

A adoção dessas diretrizes é fundamental para a padronização dos estudos e a comparação de resultados entre diferentes sistemas, contribuindo para o desenvolvimento de classificadores mais confiáveis e precisos.

2.3.2 Paradigmas para desenvolvimento dos classificadores

Como discutido anteriormente, a interpretação de longos registros de sinais de eletrocardiograma (ECG) é uma tarefa visualmente cansativa e desafiadora para especialistas médicos. Neste cenário de necessidade do uso de sistemas automáticos para auxiliar no diagnóstico, a escolha do paradigma de avaliação é uma decisão estratégica, pois impacta diretamente a capacidade de generalização do modelo (Di Paolo, 2025). Dois paradigmas principais são amplamente empregados para orientar o desenvolvimento e a validação desses sistemas: o paradigma intrapaciente e o paradigma interpaciente.

2.3.2.1 Paradigma Intrapaciente

No paradigma intrapaciente, os dados de treinamento e teste podem incluir registros do mesmo paciente, embora divididos aleatoriamente. Isso significa que todos os dados são embaralhados e separados em conjuntos de treino e teste, de forma que registros do mesmo paciente estão presentes em ambos os conjuntos.

Isso permite que o modelo seja avaliado em sua capacidade de generalização para diferentes condições dentro do mesmo indivíduo, levando em consideração a variabilidade intra-individual do sinal de ECG, como variações de ritmo cardíaco e alterações na morfologia

das ondas. Modelos avaliados sob este paradigma tendem a apresentar melhor desempenho (e.g., acurácia), pois são treinados e testados em dados que contêm as características únicas dos pacientes que também estarão no conjunto de teste (Di Paolo, 2025).

Embora geralmente resulte em melhores indicadores de desempenho, este método pode superestimar a capacidade de generalização para novos pacientes. O fato de registros do mesmo paciente estarem presentes em ambos os conjuntos (treino e teste), mesmo que não sejam os mesmos segmentos, pode levar a uma avaliação otimista da capacidade de generalização. Consequentemente, não é a abordagem mais adequada para modelos que visam generalização para pacientes novos ou populações distintas.

2.3.2.2 Paradigma Interpaciente

O paradigma interpaciente envolve o treinamento do modelo em um conjunto de dados de um grupo de pacientes e a testagem em outro conjunto de dados de pacientes completamente diferentes. Não há sobreposição de indivíduos entre as fases de treinamento e teste. Isso garante que o modelo não teve acesso ao comportamento dinâmico dos registros dos pacientes presentes no conjunto de teste. A divisão da base de dados MIT-BIH em conjuntos de treino (DS1) e teste (DS2), uma prática estabelecida por Chazal, O'Dwyer e Reilly (2004), aprimorada por Mar et al. (2011) e Llamedo e Martínez (2011), e adotada neste trabalho, é um exemplo proeminente dessa abordagem.

Esta abordagem oferece uma avaliação mais realista da capacidade de generalização do modelo para novos indivíduos, sendo crucial para o desenvolvimento de classificadores que serão aplicados em diversas populações. Esta característica é vital para o desenvolvimento de classificadores que buscam aplicabilidade em contextos clínicos reais, onde o sistema deve ser capaz de diagnosticar arritmias em pacientes que nunca foram vistos durante o treinamento (Di Paolo, 2025).

Uma desvantagem potencial é que essa abordagem geralmente requer um volume maior de dados para permitir a divisão adequada entre os conjuntos, o que pode ser difícil de obter na prática. Além disso, variações entre os pacientes (como diferenças de idade, sexo, etnia ou comorbidades) podem afetar o desempenho do modelo. Tipicamente, modelos avaliados sob este paradigma podem resultar em menor desempenho, pois os dados de treinamento podem ser limitados na representatividade da população geral de pacientes (Di Paolo, 2025).

Em suma, a escolha do paradigma de avaliação é um passo metodológico importante. Embora o paradigma intrapaciente possa oferecer medidas de desempenho mais elevadas, é o paradigma interpaciente que fornece uma avaliação mais robusta e representativa da aplicabilidade clínica real do classificador.

2.3.3 Medidas de desempenho

A norma ANSI/AAMI EC57:2012 R2020 especifica os indicadores que devem ser usados para avaliação de desempenho de classificadores de arritmias, sendo eles acurácia, precisão ou acertos positivos, sensibilidade (*recall*), especificidade e *F1-Score*. As medidas de desempenho são essenciais para quantificar o desempenho dos classificadores, seus valores advêm da comparação entre o conjunto de teste e a predição do sistema, que computa todos os acertos e erros. As medidas de desempenho utilizam cada um dos valores abaixo:

TP (*True Positive*) significa verdadeiro positivo, no contexto da classificação de arritmias cardíacas indica que o modelo classificou corretamente uma amostra como pertencente a uma determinada classe de arritmia (ou batimento normal).

TN (*True Negative*) significa verdadeiros negativo, ocorre quando o modelo classificou corretamente uma amostra como não pertencente a uma determinada classe de arritmia (ou batimento normal).

FP (*False Positive*) significa falso positivo, acontece quando o modelo classifica erroneamente uma amostra como pertencente a uma determinada classe de arritmia (ou batimento normal), quando, na realidade, ela não pertencia àquela classe.

FN (*False Negative*) significa falso negativo, ocorre quando o modelo classifica erroneamente uma amostra como não pertencente a uma determinada classe de arritmia (ou batimento normal), quando, na realidade, ela pertencia àquela classe.

A medida de desempenho acurácia é definida como:

- **Acurácia (*Accuracy*)**: A acurácia é uma das medidas de desempenho mais utilizadas, pois divide o total de acertos pelo total de valores presentes na base de dados e indica o quão bem o modelo acerta nas classificações possíveis. É a medida de desempenho com interpretação mais intuitiva, sendo calculada por:

$$Acuracia = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

- **Precisão (*Precision*)**: Esta medida de desempenho responde à pergunta: Qual a proporção de identificações positivas que foram realmente corretas? No diagnóstico de arritmias, indica quantas classes de arritmia ou normalidade detectadas pelo modelo de fato estavam corretas, sendo calculada por:

$$Precisao = \frac{TP}{TP + FP} \quad (2.2)$$

- **Sensibilidade (*Recall*)**: Esta medida de desempenho responde à seguinte pergunta: qual proporção de positivos foi identificada corretamente? Em outras palavras, ela indica o quanto o modelo é capaz de detectar os casos reais (verdadeiros) de arritmia, podendo ser

calculada por:

$$Sensibilidade = \frac{TP}{TP + FN} \quad (2.3)$$

- **Especificidade (*Specificity*):** A especificidade é o complemento da sensibilidade. No caso de diagnósticos de arritmia, enquanto a sensibilidade mede o quanto o modelo identifica corretamente a presença da condição, a especificidade mede o quanto ele acerta ao identificar sua ausência, sendo calculada por:

$$Especificidade = \frac{TN}{TN + FP} \quad (2.4)$$

- ***F1-Score*:** Essa medida de desempenho representa o equilíbrio entre precisão e sensibilidade, sendo especialmente útil em cenários de classes desbalanceadas. Embora não seja citada na AAMI, é especialmente útil em cenários de classes desbalanceadas, como as arritmias cardíacas, pois penaliza tanto os falsos positivos quanto os falsos negativos, fornecendo uma visão mais realista do desempenho. Por isso, é bastante recomendada na literatura para bases desbalanceadas (Di Paolo, 2025). A base MIT-BIH, por exemplo, é intrinsecamente desbalanceada, tornando o *F1-Score* uma medida de desempenho preferencial para a avaliação de modelos. A medida de desempenho pode ser calculada por:

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (2.5)$$

2.4 Séries temporais

As séries temporais, constituídas por uma sequência de pontos de dados coletados em intervalos regulares de tempo, representam uma parte fundamental na análise estatística contemporânea. Através da aplicação de técnicas estatísticas e de modelagem, a análise de séries temporais busca prever futuros comportamentos com base em observações passadas (PRIYAMVADA; WADHVANI, 2017). Segundo Shterev, Metchkarski e Koparanov (2022), a maioria dos processos físicos ocorre em uma sequência natural de passado-presente-futuro. Muitos desses processos podem ser representados como sequências temporais unidimensionais contínuas ou discretas. Sob a perspectiva da informação, um sinal contínuo pode ser transformado em um sinal discreto, mesmo com perdas de dados. Dessa forma, problemas cotidianos como o consumo de eletricidade, tráfego de informações e taxas de mercado podem ser apresentados como sequências discretas unidimensionais, ou seja, séries temporais.

Diferente de outros tipos de dados, as séries temporais apresentam uma dependência temporal inerente. Em uma série temporal, a independência entre os pontos de dados é quase inexistente, sendo os valores em um determinado momento geralmente correlacionados com os anteriores. Segundo Shumway e Stoffer (2010), a presença óbvia de correlação ao amostrar pontos próximos no tempo pode limitar significativamente o uso de muitos métodos estatísticos tradicionais que assumem que essas observações adjacentes são independentes e têm a mesma

distribuição. O processo sistemático para responder às questões matemáticas e estatísticas decorrentes dessas correlações temporais é conhecido como análise de séries temporais.

Séries temporais e suas análises desempenham um papel crucial em uma variedade de campos de conhecimento. A busca por identificar relações causais em contextos específicos amplia os horizontes das técnicas de análise de séries temporais. Contribuições nesse sentido emanam em diversos domínios, como em Economia (como produção industrial, crescimento econômico e taxas de desemprego), em Medicina (como sinais de eletrocardiograma e encefalograma), Epidemiologia (como o número semanal de casos de COVID-19), Meteorologia (como temperatura diária e precipitação pluviométrica), Astronomia (como o deslocamento de satélites observáveis), entre outros (NIELSEN, 2019).

Há diversas abordagens tradicionais para a análise de séries temporais, focando nas relações dos valores em diferentes tempos da série, entre elas a análise de estacionaridade, a análise de autocorrelação e a decomposição de séries temporais.

Uma série temporal não é nada mais do que um processo estocástico. Seja $x = x_1, x_2, \dots, x_T$, uma série temporal qualquer; x é a realização do processo estocástico $X = \{X_1, X_2, \dots\} = \{X_t\}_{t=1}^{\infty}$, nesse sentido, um processo estocástico X é estacionário se suas propriedades estatísticas forem independentes do tempo (DAMA; SINOQUET, 2021).

2.4.1 Técnicas para transformação de séries temporais em imagens

A análise de séries temporais tem sido uma ferramenta essencial em vários campos, incluindo econometria, finanças, meteorologia e ciência da computação (WANG; OATES, 2015). Em muitos desses domínios, a transformação de séries temporais em imagens está se tornando uma técnica cada vez mais adotada, possibilitando o emprego de abordagens de visão computacional e aprendizado profundo para a exploração de dados temporais (QUAN et al., 2023). Esta técnica de transformação, como indicado por Wang e Oates (2015) e Quan et al. (2023), oferece o potencial de simplificar a análise de dados complexos, facilitando a identificação de padrões e a detecção de anomalias, visto que convertem dados unidimensionais em imagens bidimensionais, podendo revelar com eficácia informações numéricas e temporais de séries temporais, bem como o conhecimento prévio das informações de correlação, similaridade e quantidade de informações em diferentes pontos de tempo, fornecendo informações adicionais para melhorar a precisão de um modelo classificador. A informação encapsulada em imagens se mostra mais intuitiva e fácil de compreender, transformando uma série temporal complexa em uma estrutura visualmente acessível.

Adicionalmente, técnicas de aprendizado profundo, como as Redes Neurais Convolucionais (CNNs), têm mostrado grande eficácia na extração de características informativas dessas imagens de séries temporais. As CNNs têm a capacidade de aprender hierarquias de características, desde bordas e cores até padrões mais complexos, tornando-as adequadas para a análise de imagens convertidas de séries temporais.

A transformação de séries temporais em representações gráficas leva em consideração que cada ponto de dados numa sequência temporal pode ser convertido num pixel em uma imagem. Essas abordagens apresentam propriedades importantes que possibilitam preservar características e relações temporais das séries originais na forma de imagens (WANG; OATES, 2015).

De forma geral, a transformação de séries temporais em imagens oferece o potencial de revelar informações numéricas e temporais, bem como o conhecimento prévio das informações de correlação, similaridade e quantidade de informações em diferentes pontos de tempo, fornecendo dados adicionais para aprimorar a precisão de modelos classificadores (WANG; OATES, 2015; QUAN et al., 2023). Existem diversos métodos propostos para a conversão de séries temporais em imagens e, entre eles, destacam-se o *Gramian Angular Fields* (GAF), *Recurrence Plot* (RP), *Markov Transition Field* (MTF) e *Hilbert Curve* (HC). Conforme descrito por Wang e Oates (2015), esses métodos têm características próprias e adequadas para tarefas específicas.

O GAF, baseado na transformação de coordenadas polares, preserva a informação de correlação temporal na imagem resultante, conforme demonstrado por seus idealizadores em (WANG; OATES, 2015) e argumentado por Quan et al. (2023). O RP se destaca por representar a recorrência de padrões dentro de um sinal temporal a partir das distâncias pareadas entre cada ponto da série temporal (Di Paolo, 2025), sendo um método importante para analisar a periodicidade e o caos das séries temporais, bem como a não estacionariedade, e para revelar a estrutura interna das séries temporais (QUAN et al., 2023). O MTF, por outro lado, é uma representação de imagem que captura a dinâmica de transição de uma série temporal, conforme apontado por seus idealizadores também em Wang e Oates (2015). Outro método importante, a Curva de Hilbert (também conhecida como curva de preenchimento do espaço de Hilbert), descrita pela primeira vez pelo matemático alemão David Hilbert em 1891, é uma curva de preenchimento de espaço fractal contínuo, ou seja, uma curva que passa por todos os pontos de um espaço d-dimensional sequencialmente (TSINGANOS et al., 2021).

A seguir será apresentado com mais detalhes as duas técnicas que foram utilizadas neste trabalho, sendo elas GAF e HSFC.

2.4.1.1 *Gramian Angular Fields* (GAF)

O *Gramian Angular Fields* (GAF), popular na transformação de séries temporais em imagens, consegue codificar a informação de tempo e magnitude na coloração e ângulos da imagem resultante. Isso foi ilustrado pelo trabalho de Wang e Oates (2015), onde enfatizaram a habilidade do GAF em preservar a estrutura temporal dos dados, mantendo simultaneamente a interpretabilidade da imagem.

A GAF funciona com a prerrogativa de que é possível realizar uma transformação dos dados de séries temporais para coordenadas polares ao aplicar a função inversa do cosseno (arcocosseno) nos valores. No trabalho de Wang e Oates (2015), dada uma série temporal

$X = \{x_1, x_2, \dots, x_n\}$, os autores executam uma normalização dos dados, dada por:

$$x'_i = \frac{(x_i - \max(X) + (x_i - \min(X)))}{\max(X) - \min(X)} \quad (2.6)$$

Em seguida aplicam a função arcocosseno para representar a amplitude como o cosseno angular e a informação temporal como o raio, conforme a equação:

$$\phi = \arccos(x'_i), \quad -1 \leq x'_i \leq 1, \quad x'_i \in X, \quad r = \frac{t_i}{N}, \quad t_i \in N \quad (2.7)$$

Essa abordagem baseada em coordenadas polares oferece uma nova perspectiva para a interpretação das séries temporais, que pode ser utilizada para definir a relação temporal de cada intervalo de tempo específico, após a transformação da série temporal para o sistema de coordenadas polares, levando em consideração a soma trigonométrica entre cada ponto (K) (SREENIVAS; GANESAN; LAVANYA, 2021). Essa transformação resulta em uma matriz conhecida como matriz de Gram, representada por G .

$$G = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} \quad (2.8)$$

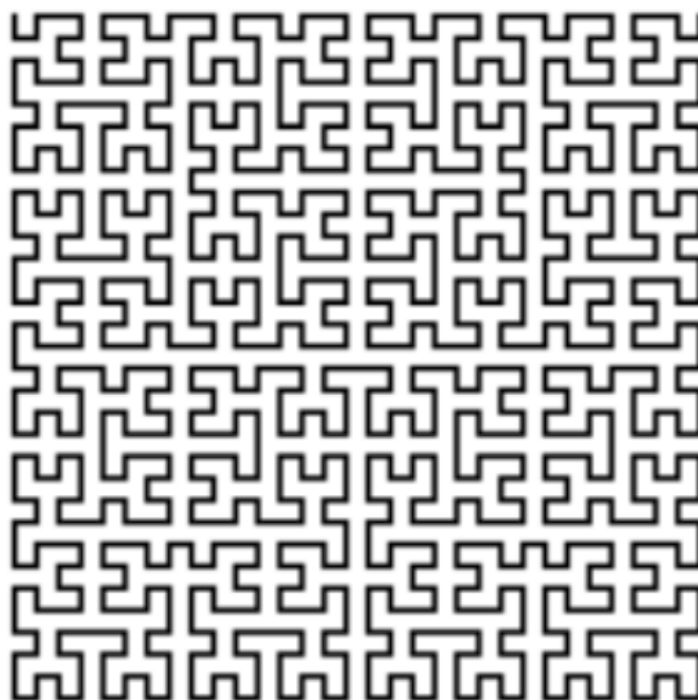
E como argumentado por Jiang et al. (2022), Os GAFs possuem várias vantagens. Primeiramente, eles oferecem uma maneira de preservar a dependência temporal, uma vez que o tempo aumenta à medida que a posição se move de cima para baixo e da esquerda para a direita. Os GAFs contêm correlações temporais, pois $G|i - j| = k$ representa a correlação relativa por meio da sobreposição/diferença de direções em relação ao intervalo de tempo k . A diagonal principal $G(i,i)$ é o caso especial quando $k = 0$, que contém a informação original do valor.

2.4.1.2 Curvas de Hilbert

A curva de Hilbert, caracterizada por sua habilidade em preencher todo o espaço, tem sido amplamente empregada em processamento de imagem e análise de dados. O matemático italiano Peano e o matemático alemão Hilbert apresentaram a curva, que preenche uma grade quadrada, respectivamente, em 1890 e 1891, e propuseram um método para percorrer cada nó na grade usando essa curva contínua, que é chamada de curva de Hilbert.

Essas curvas estão localizadas em um espaço euclidiano com dimensão maior que 1 e têm interiores não vazios dentro do espaço. A Figura 3 apresenta um exemplo de curva de Hilbert.

Figura 3 – Curva de Hilbert

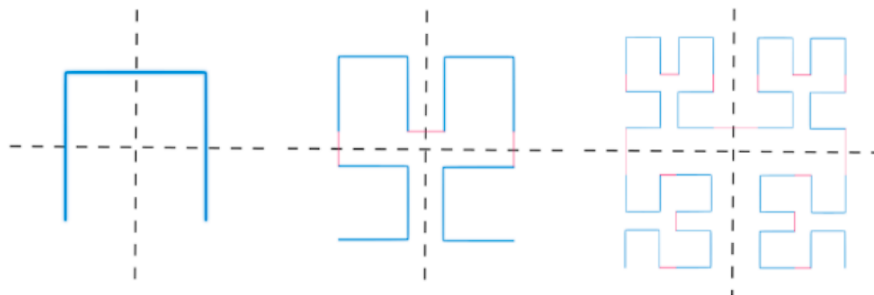


Fonte: Zhang et al. (2019)

A curva de Hilbert é uma curva contínua fractal que preenche o espaço. Se tal curva for estendida ao longo de uma linha reta, pontos localizados próximos uns dos outros em uma representação bidimensional também tendem a estar próximos uns dos outros em uma sequência linear.

A Figura 4 ilustra as ordens da curva de Hilbert. A curva de Hilbert de primeiro grau (H1) é semelhante a um "U" invertido, posicionado em três dos lados do quadrado. Para se obter curvas de ordens superiores, substituem-se os vértices superiores pelas curvas da ordem anterior. No caso do vértice inferior esquerdo, utiliza-se uma curva rotacionada 90 graus à direita, e para o vértice inferior direito, uma curva rotacionada 90 graus à esquerda. A curva de Hilbert de segunda ordem (H2) é composta por curvas de primeira ordem (H1) que se conectam através de três segmentos de reta, também conhecidos como feixes.

Figura 4 – Curva de Hilbert: Ordens



Fonte: (Di Paolo, 2025)

A curva de Hilbert de terceira ordem (H3) é formada por quatro curvas H2 e três feixes. Toda curva de uma dada ordem inicia no canto inferior esquerdo e finaliza no canto inferior direito. A curva de Hilbert de ordem zero corresponde a um ponto. Portanto, uma curva de Hilbert de grau n , denominada H_n , engloba 4^n pontos e se distribui numa área de dimensões $2^n \times 2^n$ (BRAGIN; SPITSYN, 2020).

A série temporal $f(t)$, composta por p pontos, pode ser representada em um plano bidimensional na forma de uma curva de Hilbert associando o índice i de cada ponto na série temporal com a respectiva coordenada (x, y) na curva. Neste caso, o valor do ponto na coordenada (x, y) terá o valor correspondente da série temporal (BRAGIN; SPITSYN, 2020).

A gama de aplicações práticas decorrentes da transformação de séries temporais em imagens é ampla e diversificada. Por exemplo, na esfera financeira, tais técnicas podem ser empregadas para visualizar e analisar comportamentos de mercado e identificar anormalidades, conforme destacado por Sezer e Ozbayoglu (2018) em sua análise dos mercados de ações.

No campo médico, a utilização dessas transformações tem proporcionado avanços significativos. Uma aplicação chave tem sido no diagnóstico de condições de saúde com base em dados de sinais vitais. Fawaz et al. (2019) evidenciaram isso em sua pesquisa, onde a transformação de séries temporais em imagens demonstrou ser um meio eficaz para classificar dados de eletrocardiograma.

Essas diversas aplicações ilustram a versatilidade e o potencial da transformação de séries temporais em imagens, evidenciando que tal técnica está longe de ser uma mera ferramenta teórica, apresentando um amplo espectro de aplicabilidade em uma variedade de setores e disciplinas.

2.5 Redes Neurais Convolucionais

Uma Rede Neural Convolutiva ou *Convolutional Neural Network* (CNN) é um tipo especial de redes neurais de múltiplas camadas, com sua arquitetura inspirada no sistema visual de seres vivos. A CNN é um tipo de rede neural que utiliza a operação de convolução para

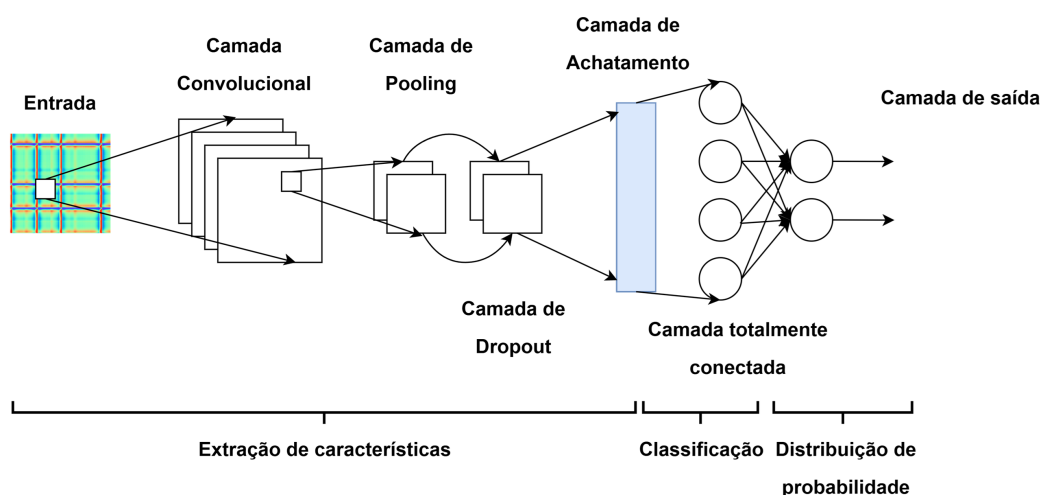
extrair características relevantes de dados como imagens, sons ou textos, sendo muito usada em aplicações de visão computacional, reconhecimento de fala e processamento de linguagem natural. Ela foi introduzida como uma variação da *Perceptron* de Múltiplas Camadas (MLPs) por Rumelhart, Hinton e Williams (1986), e desde então vêm sendo amplamente empregadas no processamento de dados com estrutura espacial ou temporal local (Di Paolo, 2025).

A operação de convolução é o bloco de construção básico das CNNs e é responsável por sua grande eficácia em tarefas de visão computacional. Assim, as CNNs são especialmente úteis na análise de dados com estrutura espacial, como imagens e sinais de áudio, sendo particularmente eficazes em tarefas como classificação de imagens, detecção de objetos e reconhecimento de padrões, sendo também aplicáveis em domínios biomédicos onde a complexidade dos dados exige modelos capazes de capturar variações sutis em morfologias fisiológicas.

2.5.1 A Arquitetura Fundamental de uma CNN

A estrutura básica de uma CNN é organizada em duas partes principais: extratora de características e classificadora. A primeira parte é geralmente composta por uma sequência de camadas convolucionais e de *pooling*, enquanto a segunda parte é composta por camadas totalmente conectadas. A Figura 5 ilustra a arquitetura típica de uma CNN.

Figura 5 – Estrutura de CNN



Fonte: Autor. Adaptada de (Di Paolo, 2025)

A estrutura básica de uma CNN, conforme Figura 5, inclui:

- **Camadas Convolucionais:** Responsáveis por aplicar filtros treináveis que percorrem o dado de entrada, identificando padrões locais como bordas, curvas ou picos.
- **Camadas de Pooling:** Utilizadas para reduzir a dimensionalidade das ativações, preservando os elementos mais relevantes.

- **Funções de Ativação:** Introduzem não-linearidades, sendo a ReLU (*Rectified Linear Unit*) a mais comumente utilizada, embora variantes como *LeakyReLU* também possam ser aplicadas para evitar o problema de gradientes mortos.
- **Camadas Totalmente conectadas ou *Fully Connected* (FC):** posicionadas nas etapas finais. Realizam a classificação com base nas características extraídas nas camadas anteriores.

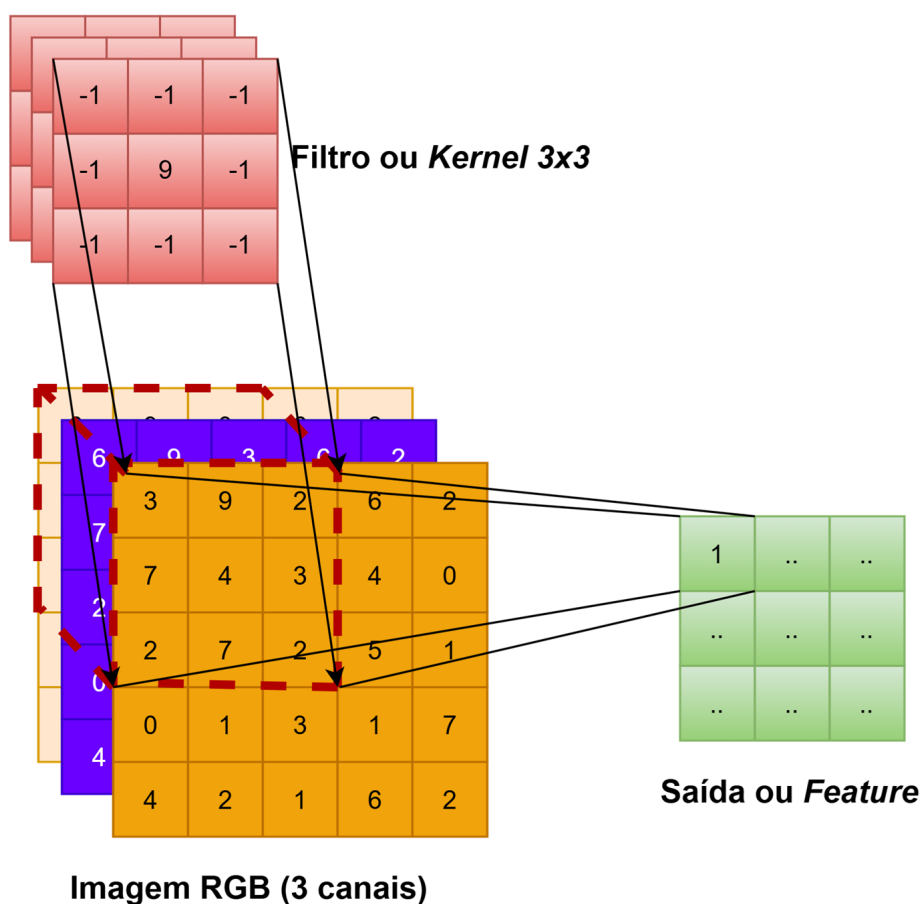
A seguir, cada um dos componentes fundamentais dessa arquitetura será detalhado.

2.5.1.1 Camada Convolucional

As camadas convolucionais são responsáveis pela leitura de propriedades e recursos dos dados de entrada. A CNN recebe seu nome em função da operação de convolução realizada na camada convolucional, sendo fundamental para o funcionamento da rede. É importante destacar que a CNN se destaca por seu desempenho com imagens digitais ao utilizar o formato 2D, guardando os valores de cada pixel, porém ela também pode ser utilizada com estruturas de dados com múltiplas dimensões, com CNNs que aceitam vídeos de três dimensões com o intuito de extrair o tempo como uma variável.

A Figura 6 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 6) se desloca a cada iteração.

Figura 6 – Operação de convolução



Fonte: Autor. Adaptado de (Di Paolo, 2025)

A camada convolucional é a mais importante na CNN. Nela, existe um grid de números com múltiplas dimensões (no caso de imagens digitais, duas dimensões) chamado *kernel* ou filtro, com cada valor sendo denominado peso do *kernel*. Em geral, são utilizados *kernels* com pesos de valores randômicos, mas é possível encontrar valores pré-definidos para a tentativa de identificação ou realce de um algum elemento da imagem, como bordas ou formas de N lados. Para o *kernel* realizar o filtro desejado na imagem, ele realiza a convolução em conjuntos de pixels de entrada. As camadas de convolução de uma CNN são responsáveis por identificar características e padrões nas imagens de entrada e aplicar a operação com os filtros (ou *kernels*) para extrair características como bordas, curvas e texturas da imagem, gerando um mapa de características como saída. Dessa forma, o *kernel* desliza pela imagem para aplicar a operação a todos os pixels presentes e salvar os resultados na imagem resultante.

2.5.1.2 Função de Ativação

Na arquitetura de uma CNN, após cada camada convolucional, é posicionada uma camada de ativação, onde uma função é aplicada para cada elemento da matriz obtida na camada prévia,

gerando como resultado uma matriz de iguais dimensões. O propósito da camada de ativação é introduzir não linearidade dentro da rede, de forma que o modelo seja capaz de aprender padrões complexos nos dados de entrada, ajustando um limite máximo de valor para a camada de convolução entre um valor de 0 a N.

2.5.1.3 Camada de *Pooling*

Em seguida, tem-se a camada de *Pooling*, que é responsável por diminuir a quantidade de pixels na imagem de entrada (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 o *Max Pooling* e *Average Pooling* (Di Paolo, 2025):

- ***Max Pooling***: É o tipo de *pooling* mais utilizado em arquiteturas de redes neurais convolucionais. Durante essa operação, uma janela de tamanho fixo percorre a matriz de entrada e seleciona o valor máximo dentro dessa região. Essa abordagem ajuda a preservar as características mais importantes enquanto reduz a resolução espacial dos dados.
- ***Average Pooling***: Calcula o valor médio dos elementos dentro da região do filtro.

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.

2.5.1.4 Camada de Achatamento (*Flatten*)

Após a fase de extração de características, os mapas de características resultantes, que são bidimensionais (ou tridimensionais), precisam ser preparados para a etapa de classificação. A camada de achatamento (*Flatten*) realiza essa tarefa, transformando a matriz de dados em um vetor unidimensional, que servirá como entrada para a camada seguinte.

2.5.1.5 Camada Totalmente Conectada

Por fim, tem-se a camada totalmente conectada ou densa, que segue a estrutura de uma Rede Neural *Perceptron* de múltiplas camadas padrão, sendo responsável por receber as características aprendidas pelas camadas de convolução e *pooling* e utilizá-las para realizar a classificação ou predição desejada. Essa camada é marcada pela interconexão total entre seus nós ou neurônios e geralmente é utilizada como camada de saída da rede CNN. Nela, é utilizada uma função de ativação, de forma a incluir a não-linearidade ao sistema.

Por fim, após o processamento das características extraídas, emprega-se a função *Softmax* como função de ativação. Seu propósito primordial é transformar os valores brutos (*logits*)

gerados pelos neurônios da camada anterior em uma distribuição de probabilidade sobre as classes possíveis (Di Paolo, 2025). Isso significa que a *Softmax* garante que cada saída será um valor entre 0 e 1, e que a soma de todas as saídas será igual a 1, interpretando-se esses valores como a probabilidade de a entrada pertencer a cada uma das classes. Por exemplo, no contexto deste trabalho, ela é utilizada para gerar a distribuição de probabilidade para as classes "Normal"(N), "Supraventricular"(S) e "Ventricular"(V), permitindo que o modelo efetue a classificação ao indicar a classe com maior probabilidade.

2.5.2 Treinamento de redes neurais convolucionais

O treinamento e aprendizado de CNNs ocorre sob o paradigma de aprendizado supervisionado, no qual os modelos aprendem a partir de um conjunto de exemplos rotulados de entrada e sua correspondente saída, ajustando iterativamente seus parâmetros internos para minimizar o erro de predição. Este treinamento consiste na captura das informações e características de interesse da imagem inicial a partir das etapas descritas na estrutura da rede. Para melhorar a acurácia desse processo, ajustes são realizados nos blocos de convolução e *pooling* e para calcular o erro, são usadas diferentes funções de custo, como a entropia cruzada (*Cross Entropy*) e perda de dobradiça (*Hinge Loss*).

O mecanismo de aprendizado envolve três componentes principais:

- **Funções de Custo (ou Perda):** Medem o quão distante a saída da rede está do valor verdadeiro. As funções mais utilizadas neste contexto são:
 - **Entropia Cruzada (*Cross Entropy*):** Ideal para classificação multiclasse, mede a distância entre a distribuição prevista pela rede e a distribuição verdadeira.
 - **Perda de dobradiça (*Hinge Loss*):** Comum em classificadores como SVMs, penaliza fortemente classificações incorretas.
- **Complexidade Computacional e *Overfitting*:** Quanto maior a quantidade de parâmetros da rede, maior é a sua capacidade de aprendizado, mas também o risco de *overfitting*, ou seja, de memorizar os dados de treino em vez de generalizar para dados novos. Técnicas como regularização, *Dropout* e estratégias de validação são essenciais para mitigar esse problema e garantir um bom desempenho do modelo.

Para o treinamento de CNNs, primeiro é necessária uma etapa de preparação dos dados, onde ocorre a operação de normalização dos dados, redimensionamento e rotulação das imagens de entrada. Nessa etapa também dividimos os dados em três subconjuntos: treino, validação e teste. O conjunto de treino é utilizado unicamente para o treinamento da CNN, já o conjunto de validação é utilizado para a otimização do modelo, por fim, o conjunto de teste é utilizado para a avaliação da performance do modelo após o treino.

Após a preparação dos dados, a arquitetura da CNN é projetada, definindo a organização e quantidade de camadas. A estrutura da rede é planejada com base na complexidade do problema a ser resolvido e na disponibilidade de recursos computacionais.

Em seguida, temos o treinamento do modelo, onde os dados de entrada são divididos em lotes, pequenos conjuntos de dados e então alimentados na CNN. O treinamento ocorre através do ajuste iterativo dos pesos da rede de acordo com uma função de perda. Esse processo envolve a propagação dos dados para frente (em inglês *forward pass*), o cálculo do gradiente da função de perda em relação aos pesos da rede e a atualização dos pesos usando algoritmos de otimização, como o Adam discutido em (KINGMA; BA, 2015).

2.5.3 Arquiteturas de CNN para classificação de imagens

Ao longo das últimas décadas, diversas arquiteturas de CNN foram propostas e evoluíram consideravelmente em profundidade, eficiência e desempenho, como:

- **LeNet-5 (1998):** Desenvolvida por LeCun et al. (1998) para reconhecimento de dígitos manuscritos. Apresenta uma estrutura simples, com duas camadas convolucionais e três camadas totalmente conectadas.
- **AlexNet (2012):** Foi proposta em 2012 por Krizhevsky, Sutskever e Hinton (2012). Revolucionou a área ao vencer o *ImageNet Challenge* com grande margem. Utiliza 8 camadas (5 convolucionais e 3 FC), ativação ReLU, *Max Pooling*, e *Dropout*. Tornou-se base para muitas redes subsequentes.
- **VGGNet (2014):** Proposta em 2014 por Simonyan e Zisserman (2014). Caracteriza-se pela utilização de pequenas convoluções (3x3) empilhadas em profundidade (até 19 camadas), o que melhora a extração hierárquica de características.
- **GoogLeNet / Inception (2014):** Introduziu o conceito de módulos *Inception*, que executam convoluções paralelas de diferentes tamanhos, promovendo eficiência no uso de parâmetros.
- **ResNet (2015):** Proposta por He et al. (2016), apresentou as conexões residuais (*skip connections*), permitindo o treinamento de redes extremamente profundas (até 152 camadas) sem degradar o desempenho.
- **MobileNet (2017):** Proposta em 2017 por Howard et al. (2017). Projetada para dispositivos com capacidade computacional limitada, utiliza convoluções separáveis em profundidade para reduzir a complexidade.

2.5.4 Mecanismos de atenção em redes CNN

O mecanismo de atenção é uma técnica inspirada pela capacidade humana de focar em partes específicas de uma entrada ao processar informações, melhorando assim o desempenho e a eficiência do modelo. Inicialmente popularizada em tarefas de processamento de linguagem natural (NLP), como tradução automática, o mecanismo de atenção permite que modelos de aprendizado profundo ponderem dinamicamente a importância de diferentes partes dos dados. Por exemplo, em modelos de sequência, a atenção ajuda a focar em palavras relevantes em um contexto dado para a geração de uma resposta adequada.

A técnica foi adaptada para redes convolucionais (CNNs) e outras arquiteturas, onde ajuda a destacar características importantes em dados espaciais ou sequenciais, resultando em modelos mais precisos e adaptativos às características essenciais da tarefa em questão. O mecanismo de atenção permite que a rede aprenda a dar mais importância a características ou regiões mais informativas da entrada. Em vez de processar todos os elementos de um mapa de características de maneira uniforme, a rede pode ponderar dinamicamente a relevância de cada um, potencializando o desempenho no processo de classificação (Di Paolo, 2025).

2.5.4.1 Taxonomia dos Mecanismos de Atenção

Os mecanismos de atenção em CNNs podem ser categorizados com base em como e onde a ponderação de características é aplicada. A literatura destaca as seguintes abordagens principais:

- **Atenção de Canal (*Channel Attention*)**: Foca em ajustar a importância de cada canal no mapa de características, permitindo que a rede destaque os canais mais relevantes para a tarefa. Essencialmente, responde à pergunta: "o que é mais importante?". Exemplos dessa abordagem são encontrados nos trabalhos de Srivastava et al. (2022) e Fan et al. (2023).
- **Atenção Espacial (*Spatial Attention*)**: Concentra-se em identificar as regiões espaciais mais informativas dentro de um mapa de características, respondendo à pergunta: "onde a rede deve focar?". Exemplos notáveis incluem os trabalhos de Fan et al. (2023) e Huang et al. (2024).
- **Atenção Mista (Espacial e de Canal)**: Combina as duas abordagens anteriores para explorar simultaneamente as relações espaciais e de canal, refinando os mapas de características de forma mais completa. Exemplos podem ser vistos em Wan et al. (2023) e Chen et al. (2024).
- **Autoatenção (*Self-Attention*)**: Este mecanismo captura as dependências e relacionamentos dentro da própria sequência de entrada, permitindo que o modelo pondere a importância de diferentes partes da informação ao "atender a si mesmo". Ele é capaz de modelar interações de longo alcance entre diferentes posições de um mapa de características. Esta

é a abordagem adotada neste trabalho, seguindo exemplos de Ahmad et al. (2021), Tang et al. (2023) e Di Paolo e Castro (2024).

2.5.4.2 O Paradigma de Autoatenção (*Self-Attention*)

O mecanismo de *Self-Attention* captura dependências internas em uma mesma entrada, que pode ser um mapa de características ou um vetor, ajustando a importância relativa de suas diferentes partes. A rede aprende a modelar como cada elemento se relaciona com os outros, permitindo a captura de interações globais ou regionais. Conforme Di Paolo (2025), este paradigma pode ser dividido em duas abordagens principais, sendo elas:

- **Autoatenção Clássica/Global:** Baseia-se nos conceitos de *Queries*, *Keys* e *Values* (QKV), onde os pesos de atenção são calculados a partir do produto escalar entre *Queries* e *Keys* para ponderar os *Values*. Esta abordagem é útil para capturar interações complexas e de longo alcance, sendo frequentemente aplicada diretamente nas camadas convolucionais sobre mapas de características multidimensionais.
- **Autoatenção Local:** Esta abordagem é particularmente interessante para dados unidimensionais (1D) ou para vetores de características após a camada de achatamento (*flatten*) de uma CNN. Nela, a atenção é gerada localmente, muitas vezes usando uma operação de convolução para criar um mapa de atenção achatado que é então aplicado ao vetor de características original. Esta estratégia é mais eficiente em termos computacionais e de memória, sendo ideal para aplicação na camada totalmente conectada, como observado em Ahmad et al. (2021), Tang et al. (2023) e Di Paolo e Castro (2024).

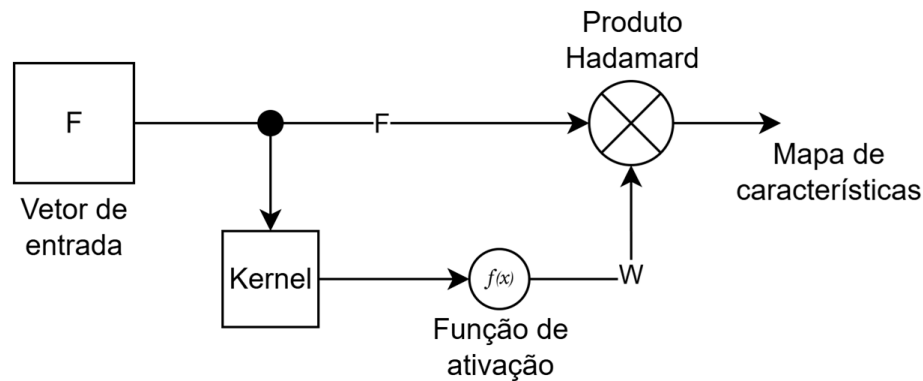
O mecanismo de *Self-Attention* local, conforme apresentado na Figura 7, se baseia em um fluxo duplo. O vetor de características de entrada (F) é processado simultaneamente em duas vias paralelas:

- **Via de Ponderação:** Na primeira via, o vetor passa por uma operação de convolução 1D, cujo *kernel* (que pode ser fixo ou aprender durante o treinamento) transforma as características. A saída desta convolução é então submetida a uma função de ativação (como Sigmóide ou Tanh), resultando em um vetor de pesos de atenção (W). Este vetor quantifica a relevância aprendida para cada elemento no vetor de características original.
- **Via de Características Originais:** Na segunda via, o vetor de características original (F) é simplesmente preservado sem alterações.

Em seguida, os dois caminhos convergem em uma operação de multiplicação elemento a elemento (Produto Hadamard), onde o vetor de pesos (W) modula o vetor de características original (F). O resultado é um novo vetor de características ponderado, no qual a rede aprendeu

a amplificar os atributos mais discriminativos e atenuar os menos importantes para a tarefa de classificação. Essa abordagem, que integra o mecanismo de atenção diretamente na camada totalmente conectada da CNN, é uma estratégia observada em trabalhos como os de Ahmad et al. (2021), Tang et al. (2023) e Di Paolo e Castro (2024) e foi a abordagem adotada neste trabalho.

Figura 7 – Estrutura *Self-Attention* local



Fonte: Autor. Adaptado de (Di Paolo, 2025)

2.6 A Influência do *Kernel*

2.6.1 Teoria dos *Kernels*

Os *kernels* são funções que permitem mapear dados de entrada para um espaço de maior (ou menor) dimensão de forma a revelar padrões ou relações que não são facilmente perceptíveis no espaço original. Em aprendizagem de máquina, o conceito de *kernel* é amplamente utilizado, por exemplo, em métodos como SVM (*Support Vector Machine*), para realizar transformações não lineares dos dados. Em redes neurais, embora o termo *kernel* seja comumente associado aos filtros das camadas convolucionais, sua aplicação pode ser estendida para módulos de atenção, onde os *kernels* atuam na modulação das ativações das *features*.

2.6.2 *Kernel* no Módulo *Self-Attention*

No contexto do *self-attention*, o *kernel* desempenha um papel crucial no cálculo dos pesos de atenção que serão aplicados às *features* extraídas. Em uma implementação típica, o *kernel* é aplicado sobre o vetor de características para gerar um mapa que determina a importância relativa de cada componente do sinal. Essa operação é, na essência, uma convolução 1D que transforma o vetor de entrada e, assim, gera um vetor de "máscaras" que irão modular (por meio de multiplicação elemento a elemento) as ativações provenientes das camadas anteriores. Dessa forma, o módulo de *self-attention* consegue concentrar a "atenção" da rede em certas partes do sinal, realçando padrões discriminativos essenciais para a classificação.

2.6.3 Diferenças entre *Kernels* Fixo e Adaptativo

A implementação do *kernel* pode assumir duas abordagens fundamentais:

- **Kernel Fixo:** Nesta abordagem, o *kernel* é definido como um tensor constante (por exemplo, $[[[-1, 9, -1], [-1, 9, -1], [-1, 9, -1]]]$) que não sofre atualização durante o treinamento. Essa configuração impõe uma transformação pré-determinada às *features*. Embora possa ajudar a reduzir ruídos e impor uma regularização, essa limitação pode não capturar adequadamente a variabilidade inerente aos dados, uma vez que o modelo não aprende a ajustar a transformação para realçar os padrões relevantes.
- **Kernel Adaptativo:** Quando o *kernel* é definido como um parâmetro treinável, ele é inicializado com um valor pré-determinado mas passa a ser ajustado durante o treinamento. Isso permite à rede adaptar o *kernel* às características específicas dos dados, possibilitando uma modulação mais refinada e, potencialmente, melhorias no desempenho. Essa flexibilidade, contudo, pode aumentar a complexidade do treinamento e exigir cuidados com a regularização para evitar sobreajuste.

O estudo comparativo entre essas abordagens fornece *insights* importantes sobre a robustez, a generalização e a capacidade de extração de características do módulo de *self-attention* local.

2.7 A Influência da Função de Ativação

2.7.1 Teoria das Funções de Ativação

Funções de ativação são componentes essenciais em redes neurais, responsáveis por introduzir não linearidade no modelo. Sem elas, uma rede composta apenas por operações lineares não poderia aproximar funções complexas. As funções de ativação transformam a saída dos neurônios, permitindo que a rede capture relações não lineares entre as variáveis de entrada e as saídas desejadas. Exemplos comuns incluem a função Sigmóide, a Tangente Hiperbólica (*tanh*) e a ReLU.

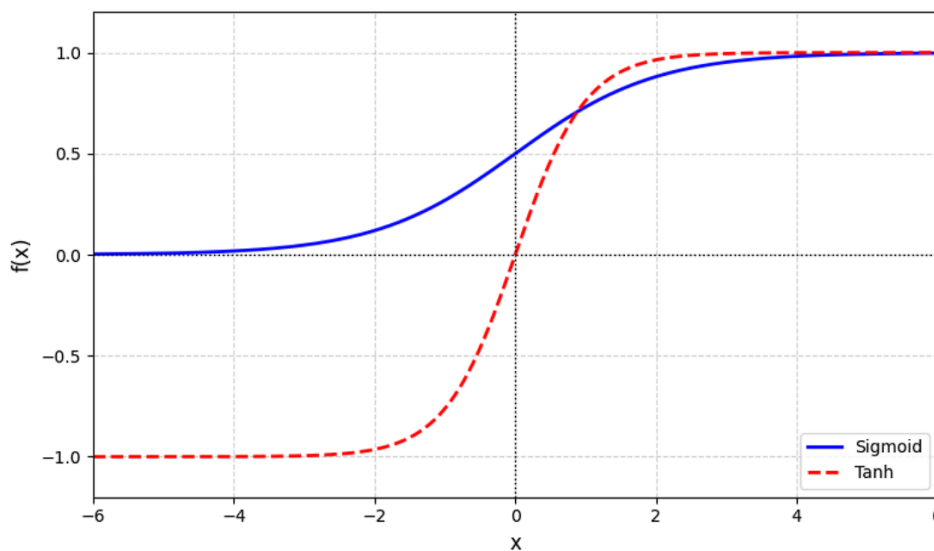
2.7.2 Função de Ativação no Módulo *Self-Attention*

No módulo de *self-attention* local, após a aplicação do *kernel*, a função de ativação é utilizada para converter a saída da operação de convolução em um conjunto de pesos que, posteriormente, serão aplicados às características extraídas. Esses pesos determinam a intensidade ou a importância dada a cada elemento da representação extraída (ALHAZMI; ALTAHHAN, 2023). Dessa forma, a escolha da função de ativação dentro do módulo de *self-attention* é um aspecto crítico que determina como a rede atribui importância e modula as características e pode

afetar diretamente a forma como o mecanismo de atenção opera, influenciando a capacidade de destacar ou suprimir certas informações.

Embora ambas as funções, Sigmóide e Tangente Hiperbólica (Tanh), introduzam a não-linearidade essencial para o aprendizado de padrões complexos, suas faixas de saída distintas conferem-lhes propriedades únicas que podem impactar diretamente a capacidade do mecanismo de atenção de realçar ou suprimir informações, conforme pode ser observado na Figura 8. A compreensão dessas diferenças é fundamental para otimizar o desempenho do classificador em tarefas sensíveis, como o diagnóstico de arritmias cardíacas, onde a nuance na ponderação dos sinais pode ser decisiva.

Figura 8 – Funções de ativação: Sigmoid vs. Tanh



Fonte: Autor

2.7.3 Diferença entre Sigmóide e tanh no Contexto

- **Sigmóide:** A função sigmóide mapeia as entradas para o intervalo (0, 1). Esse comportamento atua como um mecanismo de *gating* suave, onde as ativações podem ser interpretadas como probabilidades ou "pesos de atenção" que restringem os valores a somente atenuá-los ou mantê-los. Por não produzir valores negativos, a sigmoide não permite a inversão dos sinais, o que pode ser útil para preservar a interpretação original das *features* mas pode limitar a expressividade em alguns contextos. A função sigmoide é dada por:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.9)$$

- **Tanh (Tangente Hiperbólica):** A função Tanh mapeia as entradas para o intervalo (-1, 1), permitindo não somente atenuação, mas também a amplificação ou inversão das ativações. Essa característica pode ser vantajosa quando se deseja que o mecanismo de atenção também possa cancelar ou inverter determinados padrões indesejados, proporcionando

uma modulação mais rica das informações. Por outro lado, a capacidade de gerar valores negativos pode aumentar a complexidade da interpretação e, se não controlada, propiciar instabilidades durante o treinamento. A função Tanh é dada por:

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (2.10)$$

Um estudo comparativo entre o uso de sigmoide e tanh no módulo de *self-attention* permite determinar qual função melhor se adapta ao problema específico de classificação, avaliando medidas de desempenho como acurácia, *f1 score*, e robustez do modelo sob diferentes condições de treinamento.

3 ESTADO DA ARTE

3.1 Estado da Arte Geral

A classificação automática de arritmias cardíacas usando aprendizado profundo é um campo de pesquisa em constante evolução, impulsionado pela necessidade de aprimorar o diagnóstico médico, dada a natureza trabalhosa e propensa a erros da análise manual de longos registros de eletrocardiogramas (ECG) (Di Paolo; CASTRO, 2024).

Atualmente, as tendências de pesquisa na classificação de arritmias cardíacas com aprendizado profundo incluem:

- **Uso de Redes Neurais Convolucionais (CNNs):** As CNNs são amplamente utilizadas devido à sua eficácia no processamento de dados com estrutura de grade, como imagens. Elas se destacam pela capacidade de aprender automaticamente características relevantes dos dados, sem a necessidade de extração manual, o que as torna robustas para lidar com a complexidade e as variações presentes nos sinais de ECG clínicos reais (LUZ et al., 2016; HONG et al., 2020).
- **Conversão de Sinais 1D (Temporais) em Imagens 2D:** Uma técnica proeminente é a transformação de sinais de ECG unidimensionais (temporais) em representações visuais (imagens), que servem como entrada para as CNNs. Métodos como *Gramian Angular Field* (GAF) e Curva de Preenchimento do Espaço de Hilbert (HSFC) são empregados para codificar informações de tempo e magnitude, preservando a estrutura temporal dos dados em uma representação visual (WANG; OATES, 2015).
- **Mecanismos de Atenção:** A integração de mecanismos de atenção em arquiteturas de CNNs tem demonstrado avanços significativos. Esses mecanismos permitem que o modelo aprenda a focar seletivamente em regiões mais relevantes da entrada, atribuindo maior peso a partes mais importantes. Tipos explorados incluem *Self-Attention* (autoatenção), Atenção Espacial e Atenção de Canal. A autoatenção local, em particular, aplica uma convolução 1D sobre o vetor de características global para gerar um mapa de atenção que pondera os elementos, destacando características relevantes (Di Paolo; CASTRO, 2024).
- **Arquiteturas Híbridas e Multimodais:** Trabalhos que combinam CNNs com outras técnicas, como Redes Neurais Recorrentes (RNNs) e LSTM, ou que utilizam abordagens multimodais (combinando diferentes derivações do ECG ou diferentes representações do mesmo sinal) têm mostrado os melhores resultados. A multimodalidade pode envolver a fusão de características extraídas de diferentes entradas em paralelo (Di Paolo, 2025).

Apesar dos avanços, a classificação automática de arritmias cardíacas enfrenta vários desafios:

- **Desbalanceamento de Dados:** Bases de dados como a MIT-BIH são intrinsecamente desbalanceadas, com a classe Normal (N) representando a vasta maioria dos registros, enquanto classes minoritárias como S (Supraventricular) e V (Ventricular) são significativamente menos frequentes. Classes como F (Fusão) e S são consideradas particularmente difíceis de classificar devido à sua complexidade morfológica e ao número reduzido de dados disponíveis (CHAZAL; O'DWYER; REILLY, 2004; LLAMEDO; MARTÍNEZ, 2011).
- **Interpretabilidade (Modelos "Caixa Preta"):** Modelos de aprendizado profundo, com sua complexidade e grande número de parâmetros, são muitas vezes considerados "caixas pretas". É um desafio entender por que um resultado específico é gerado, o que é um problema significativo no domínio médico, onde diagnósticos sem explicação não são aceitáveis para especialistas (Di Paolo, 2025; HONG et al., 2020).
- **Coleta de Dados e Padronização:** Há uma falta de padronização nos procedimentos de coleta e estruturação de dados, com diferentes estudos usando diversas fontes, pacientes e equipamentos. Isso dificulta a comparação justa dos resultados entre diferentes conjuntos de dados. Bases de dados de alta qualidade e bem anotadas são difíceis de adquirir e caras (HONG et al., 2020).
- **Dados Multimodais e Sincronização:** Embora a combinação de diferentes fontes de dados (como ECG, temperatura, frequência respiratória e pressão arterial) seja promissora, um desafio é garantir que todos os sinais vitais estejam sincronizados na linha do tempo e lidar com suas frequências de amostragem variáveis (Di Paolo, 2025; HONG et al., 2020).

3.2 Trabalhos Correlatos

Este trabalho insere-se em um campo de pesquisa ativo, onde a classificação automática de arritmias cardíacas a partir de sinais de eletrocardiograma (ECG) tem sido amplamente investigada. Diversos estudos têm contribuído para o avanço da área, utilizando diferentes metodologias e abordagens, principalmente com foco nas bases de dados MIT-BIH (*The Massachusetts Institute of Technology - Beth Israel Hospital Arrhythmia Database*) e PTB (*Physikalisch-Technische Bundesanstalt*). Essas bases são recomendadas pela AAMI (*Association for the Advancement of Medical Instrumentation*) e são amplamente utilizadas para o desenvolvimento e avaliação de algoritmos de classificação de arritmias.

Para fins de avaliação de desempenho, a literatura adota principalmente dois paradigmas: intrapaciente e interpaciente. Neste estudo, optamos pela utilização do paradigma interpaciente.

3.2.1 Abordagem Interpaciente

O paradigma interpaciente envolve o treinamento do modelo em um conjunto de dados de um grupo de pacientes e a testagem em outro conjunto de dados de pacientes completamente diferentes. Esta abordagem oferece uma avaliação mais realista da capacidade de generalização do modelo para novos indivíduos, sendo crucial para classificadores aplicados em diversas populações, embora geralmente resulte em desempenho inferior devido às variações entre pacientes. A base MIT-BIH é a principal referência neste paradigma, utilizando a divisão de dados em DS1 (*Dataset 1*) para treino e DS2 (*Dataset 2*) para teste, proposta por Chazal, O'Dwyer e Reilly (2004).

A seguir, uma revisão de estudos que utilizaram paradigma interpaciente com a base MIT-BIH:

- **Chazal, O'Dwyer e Reilly (2004):** O trabalho seminal que estabeleceu a divisão DS1 e DS2 para a base MIT-BIH, utilizando um modelo baseado em discriminante linear.
- **Soria e Martínez (2009):** Empregaram discriminantes lineares e quadráticos para classificação.
- **Llamedo e Martínez (2011):** Processaram as ondas R, filtrando as ondas QRS com o auxílio de *Wavelets*, e utilizaram um classificador quadrático discriminante. É importante notar que este trabalho identificou e propôs melhorias na divisão original dos conjuntos DS1 e DS2, incluindo registros 201 e 202 para melhor balanceamento de classes e removendo o registro 232 do DS2 devido a anotações incorretas e desbalanceamento na classe SVEB.
- **Lin e Yang (2014):** Também utilizaram uma metodologia de classificação discriminante quadrática.
- **Zhang et al. (2014):** Aplicaram um classificador SVM com metodologia *one-versus-one*.
- **Garcia et al. (2017):** Exploraram uma representação de sinais ECG chamada vetor cardiograma temporal, combinada com uma rede complexa para extração de características e um classificador SVM, otimizado por PSO (*Particle Swarm Optimization*).
- **Wang et al. (2021):** Utilizaram a transformada *Wavelet* e CNN em sua metodologia. Embora tenham reportado alta performance para a classe S, o *F1-Score* global foi impactado negativamente devido a um desempenho muito baixo na classe F (abaixo de 2,04%).
- **Dias et al. (2021):** Apresentaram uma metodologia de segmentação própria a partir do intervalo RR e classificaram com um algoritmo discriminante linear.
- **Oliveira et al. (2022):** Acc 86.12%, Prec 76.72%, Recall 77.04%, Spe 92.87%, F1 76.88%.

- **Zahid, Kiranyaz e Gabbouj (2023):** Este trabalho se destaca por suas metodologias avançadas de segmentação e extração de recursos em dados 1D, que foram cruciais para alcançar uma das melhores medidas de desempenho atuais para este paradigma, com 92,27% de *F1-Score* geral e alta acurácia e precisão para a classe S sem comprometer significativamente a sensibilidade da classe N.
- **Xia, Xiong e Wang (2023):** Este estudo propõe um modelo híbrido baseado em *Transformer*, que combina CNN e *denoising autoencoder* (DAE) para *embedding* de características. A CNN é usada para extrair características locais de um único batimento. Suas medidas de desempenho globais foram Acurácia de 97,93%, *F1-score* de 52,87%, Precisão de 56,29%, Sensibilidade de 50%, e Especificidade de 93%.
- **Xu et al. (2022):** Os autores apresentam uma arquitetura que integra um mecanismo de atenção e conexões de salto residuais (*residual skip connection*) no modelo U-Net. O bloco de atenção é aplicado após cada *downsampling* e *upsampling*, aprendendo atenção de canal e atenção espacial em paralelo. Suas medidas de desempenho globais foram Acurácia de 98,5%, *F1-score* de 82%, Precisão de 92,62%, Sensibilidade de 78,32%, e Especificidade de 96,86%.

O panorama atual da pesquisa em classificação automática de arritmias cardíacas está em constante evolução, impulsionado pelos avanços na Inteligência Artificial e no Aprendizado Profundo, especialmente as CNNs. Os trabalhos correlatos revisados demonstram progressos significativos na capacidade de detectar e classificar arritmias com alta precisão. A Tabela 1 resume os resultados de medidas de desempenho obtidos nestes trabalhos.

Tabela 1 – Indicadores de desempenho de trabalhos correlatos para o paradigma interpaciente.

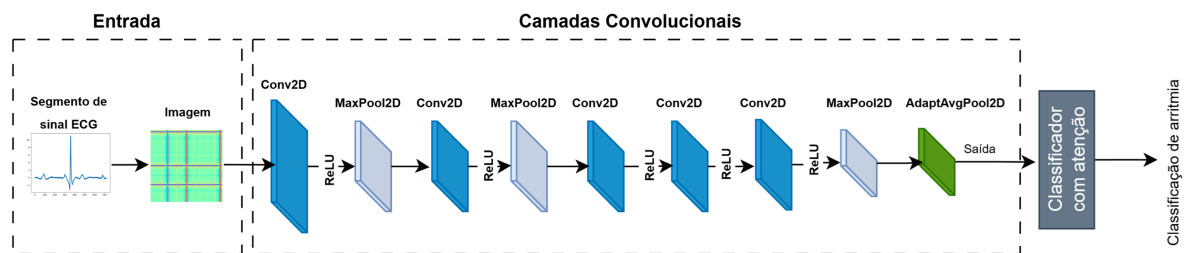
Resultados Anteriores	Acurácia (%)	Precisão (%)	Recall (%)	Especificidade (%)	F1-Score (%)
Chazal, O'Dwyer, Reilly (2004)	83,88	45,57	66,00	96,05	59,74
Soria, Martinez (2009)	91,47	69,89	89,98	94,60	76,43
Llamedo, Martínez (2011)	90,62	66,26	86,18	95,83	66,61
Lin, Yang (2014)	93,00	67,60	83,50	-	-
Zhang et al. (2014)	86,66	-	-	-	-
Garcia et al. (2017)	92,38	70,12	81,10	92,12	74,59
Wang et al. (2021)	99,16	71,06	67,57	96,17	68,97
Dias et al. (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
Xia et al (2023)	97,93	56,29	50	93	52,87
Xu et al (2022)	98,5	92,62	78,32	96,86	82

Fonte: Próprio Autor.

4 METODOLOGIA

A estrutura proposta para a classificação automática de arritmias cardíacas, conforme apresentado na Figura 9, baseia-se em uma Rede Neural Convolutiva (CNN) que incorpora um mecanismo de atenção local (*Self-Attention*). A arquitetura foi desenvolvida com o objetivo de classificar arritmias cardíacas automaticamente a partir de sinais de eletrocardiograma (ECG) transformados em imagem. O estudo procurou focar na investigação do impacto no desempenho do classificador, de diferentes funções de ativação (Tanh e Sigmoid) e tipos de *kernel* (fixo ou adaptativo) na estrutura do mecanismo de atenção e na investigação da melhor técnica de transformação de sinais ECG em imagem, considerando a técnica *Gramian Angular Field* (GAF) e a técnica da Curva de Preenchimento do Espaço de Hilbert (*Hilbert Space Filling Curve* - HSFC). Os experimentos foram conduzidos com a base de dados MIT-BIH, utilizando a abordagem interpaciente e considerando as classes N (Normal), S (Supraventricular) e V (Ventricular).

Figura 9 – Estrutura proposta neste trabalho



Fonte: Autor.

Nas próximas seções deste capítulo serão apresentados os componentes essenciais da estrutura proposta e a metodologia adotada para seu desenvolvimento. Primeiramente, será apresentado o banco de dados utilizado para treinamento e teste do modelo considerando o paradigma interpaciente. Em seguida, será apresentada a metodologia de transformação de sinais ECG em imagem baseada na técnica GAF e HSFC. Então, será apresentada com maiores detalhes a arquitetura da rede neural proposta, finalizando com apresentação da configuração de experimentos e medidas de desempenho usadas para avaliação do modelo proposto.

4.1 Banco de Dados e Paradigma Interpaciente

A seleção e o tratamento criterioso dos bancos de dados são etapas fundamentais para o desenvolvimento e a avaliação de sistemas de classificação automática de arritmias cardíacas. Para este trabalho, a Base de Dados MIT-BIH (*The Massachusetts Institute of Technology - Beth Israel Hospital Arrhythmia Database*) foi utilizada, seguindo a recomendação das diretrizes da AAMI para desenvolvimento de classificadores de arritmias.

A base de dados MIT-BIH é uma das mais amplamente utilizadas e antiga na literatura para o desenvolvimento de sistemas classificadores de arritmias. Contém 48 registros de ECG de 47 pacientes, com aproximadamente 30 minutos de duração cada, digitalizados a 360 Hz e anotados detalhadamente por especialistas. Cada registro inclui duas derivações (MLII e uma precordial), e as anotações indicam os complexos QRS e os tipos de arritmia. A diversidade de arritmias presente na MIT-BIH permite o desenvolvimento de algoritmos abrangentes.

Contudo, é imperativo destacar que a base MIT-BIH apresenta um desbalanceamento severo de classes, com a classe N (Normal) contendo mais de 90 mil registros, enquanto as demais classes (S - Supraventricular, V - Ventricular, F - Fusão, Q - Desconhecido) possuem significativamente menos de 10 mil registros. Para este estudo, apenas as classes N, S e V foram selecionadas para classificação. A classe F (Fusão) foi desconsiderada devido à sua complexidade morfológica e ao número reduzido de dados disponíveis, uma prática consistente com diversos trabalhos na literatura (Di Paolo, 2025).

A derivação MLII (*Modified Lead II*) é onipresente e amplamente utilizada devido à sua clareza para destacar o complexo QRS e as ondas P e T, além de ser menos sensível ao movimento do paciente em medições de longo período, sendo então a derivação selecionada neste estudo para desenvolvimento do classificador.

No que tange à organização dos dados, adotou-se o paradigma interpaciente, uma abordagem que oferece uma avaliação mais realista da capacidade de generalização do modelo para novos indivíduos. Neste paradigma, os dados utilizados para treinamento e teste provêm de pacientes distintos, garantindo que o modelo não tenha sido exposto, durante a fase de treinamento, ao comportamento dinâmico dos registros dos pacientes presentes no conjunto de testes.

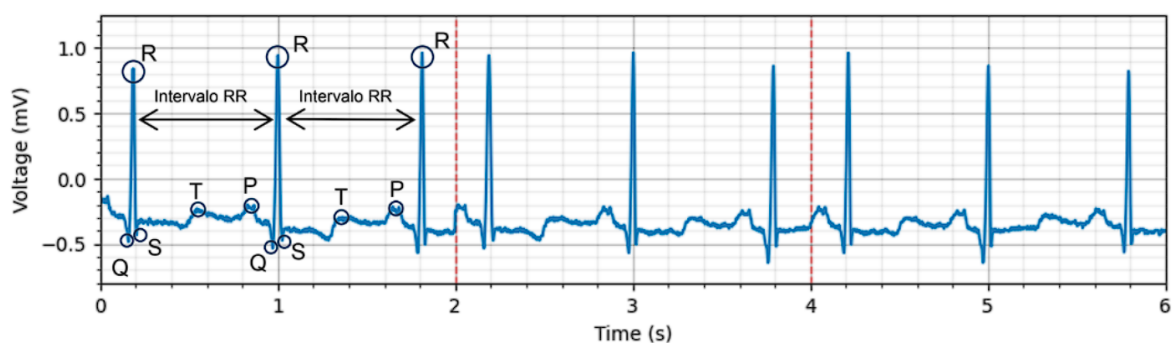
A divisão da base MIT-BIH em conjuntos de treino e teste foi realizada levando-se em consideração a principal referência na literatura científica para a divisão de dados de treino e teste, o trabalho de Chazal, O'Dwyer e Reilly (2004). Eles denominaram os conjuntos de treino e teste como DS1 (*Dataset 1*) e DS2 (*Dataset 2*), de tal maneira que os dados de um paciente específico são integralmente destinados a um dos conjuntos (DS1 ou DS2), organizando os dados para que cada conjunto contivesse as classes de forma mais equilibrada possível de diferentes pacientes. Após essa alocação de pacientes nos conjuntos, os sinais ECG dentro dos registros de cada paciente são então segmentados em intervalos específicos. No caso deste trabalho, em intervalos de dois segundos, de tal forma que houvesse, no mínimo, dois intervalos RR por

segmento, mesma metodologia adotada por (Di Paolo, 2025).

Trabalhos posteriores, como Mar et al. (2011) e Llamedo e Martínez (2011) relataram problemas de desequilíbrio na divisão original proposta por Chazal, O'Dwyer e Reilly (2004), propondo melhorias que foram seguidas neste estudo. Assim, os seguintes procedimentos foram adotados seguindo a metodologia adotada em (Di Paolo, 2025):

- Os registros 201 e 202, apesar de serem do mesmo paciente, foram incluídos nos conjuntos DS1 e DS2, respectivamente, para um melhor balanceamento entre as classes. Ainda por conta de Mar et al. (2011) e Llamedo e Martínez (2011), o registro 232 foi removido de DS2 para evitar distorções nos resultados, pois ele continha mais de 75% dos batimentos da classe SVEB com identificação incorreta em alguns segmentos, e métodos que o classificaram corretamente se especializaram apenas para esse paciente.
- As anotações 'e', 'E' e 'S' da classe S foram descartadas deste estudo, pois possuíam registros em DS1, mas não em DS2.
- **Segmentação de 2 segundos e Classes N, S, V:** O banco de dados interpaciente foi construído seguindo a mesma metodologia de Di Paolo (2025), segmentando os sinais de ECG de cada registro de paciente em intervalos de dois segundos, como pode-se observar na Figura 10. Essa duração foi escolhida para capturar no mínimo dois intervalos RR por segmento, coletando o máximo de informação possível do sinal. Apenas as classes N (Normal), S (Supraventricular) e V (Ventricular) foram selecionadas para classificação, seguindo a prática de diversos trabalhos da literatura. A classe F (Fusão) foi desconsiderada devido à sua natureza complexa e ao número reduzido de dados disponíveis.

Figura 10 – Segmentação do sinal ECG



Fonte: (Di Paolo, 2025)

É fundamental ressaltar que, para a avaliação da arquitetura proposta neste trabalho, nenhuma estratégia de aumento de dados (*data augmentation*) ou balanceamento de classes foi aplicada. O objetivo foi avaliar a robustez direta do classificador sobre a distribuição natural e

inerentemente desbalanceada dos dados originais da base MIT-BIH, refletindo um cenário real de aplicação clínica.

Os quantitativos detalhados dos dados utilizados nos conjuntos de treino (DS1) e teste (DS2) para o paradigma interpaciente, na sua composição original e desbalanceada (sem aumento de dados) são apresentados na Tabela 2. Esses dados dizem respeito a cada registro segmentado de sinal ECG resultante em cada conjunto DS1 e DS2.

Tabela 2 – Distribuição por Classes na Base MIT-BIH (Paradigma Interpaciente)

Classe	DS1 (Treino)	DS2 (Teste)
N (Normal)	45.781	43.598
S (Supraventricular)	975	667
V (Ventricular)	3.786	3.218
Total	50.542	47.483

Fonte: Próprio Autor.

Esta tabela demonstra claramente o desbalanceamento intrínseco da base de dados, com a classe N representando a maioria das amostras, enquanto as classes S e V são significativamente minoritárias. Ao total, obteve-se 50.542 imagens para o treino (DS1) e 47.483 imagens para o teste (DS2).

4.2 Transformação do Sinal ECG em Imagem

Como proposto neste trabalho, para aproveitar a capacidade das CNNs na extração de padrões espaciais complexos, uma estratégia metodológica eficaz foi adotada para converter os segmentos de sinais temporais 1D em representações de imagens 2D. Como evidenciado na Figura 9, o primeiro componente da estrutura (sua entrada) é a transformação do segmento de sinal ECG em uma imagem. Duas técnicas de transformação foram investigadas na fase experimental: *Gramian Angular Field* (GAF) e a Curva de Preenchimento do Espaço de Hilbert (*Hilbert Space Filling Curve* - HSFC).

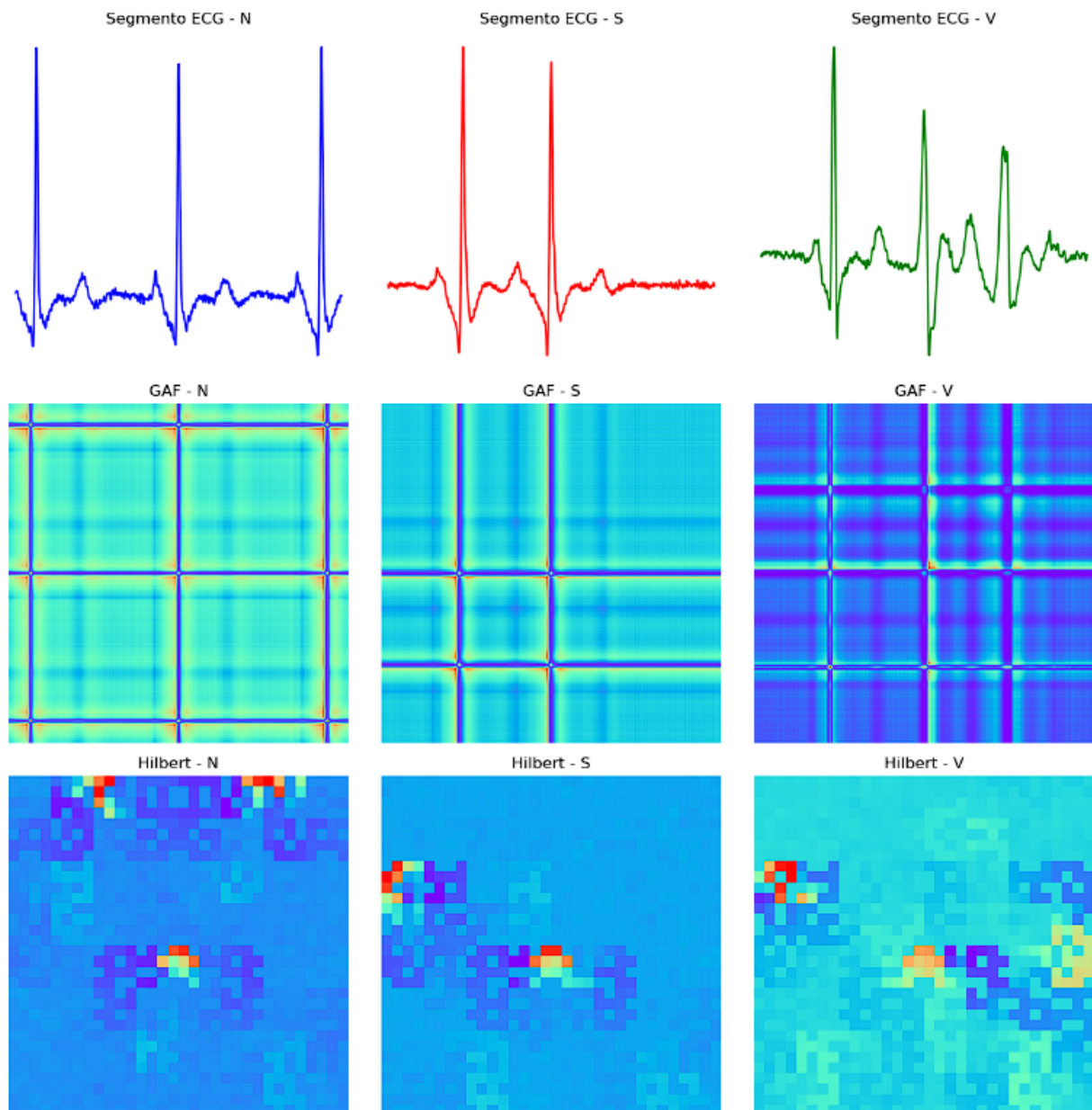
A técnica GAF é um método que codifica uma série temporal em uma imagem, preservando a estrutura temporal dos dados. O processo funciona transformando os dados da série temporal para um sistema de coordenadas polares, onde a amplitude é representada como o cosseno angular e a informação temporal como o raio. Essa abordagem gera uma matriz (imagem) que contém correlações temporais, oferecendo uma maneira de preservar a dependência temporal à medida que a posição se move na imagem. Essa capacidade de manter a interpretabilidade temporal em um formato 2D foi fundamental para a escolha desta técnica.

A técnica HSFC é classificada como uma curva contínua fractal de preenchimento de espaço, o que significa que ela tem a capacidade de passar por todos os pontos de um espaço d-dimensional de forma sequencial. Seu funcionamento se dá ao associar o índice temporal de cada

ponto da série a uma coordenada específica (x, y) na curva de Hilbert no plano bidimensional. O valor numérico dessa amostra da série temporal é então traduzido na intensidade do pixel correspondente na imagem, o que define a coloração da imagem.

A Figura 11 apresenta exemplos de segmentos de sinais ECG para cada classe (N, S, e V) transformados em imagem pela técnica GAF e HSFC.

Figura 11 – Transformação em imagens de segmentos de sinal ECG para cada classe



Fonte: Autor.

As transformações de séries temporais em imagens, como o *Gramian Angular Field* (GAF) e a Curva de Preenchimento do Espaço de Hilbert (HSFC) são, portanto, estratégias sofisticadas que permitem que Redes Neurais Convolucionais (CNNs), que são intrinsecamente

projetadas para dados de imagens, possam processar sinais unidimensionais de ECG de forma eficaz. Elas preservam a estrutura temporal original dos dados em um formato 2D, que é o que as CNNs conseguem aprender automaticamente, sem a necessidade de extração manual de características.

4.2.1 Pré-processamento das Imagens

Após o processo de geração das imagens, foi realizado um pré-processamento para normalizar a entrada da CNN. A normalização é uma etapa crucial, pois as imagens podem apresentar uma ampla variação na intensidade dos *pixels*, o que pode levar a instabilidades e convergência lenta durante o treinamento do modelo.

Primeiramente, foi definida a resolução das imagens. Com base na arquitetura AlexNet que serve de referência para o extrator de características deste trabalho, como será apresentado posteriormente, e em estudos como o de Krizhevsky, Sutskever e Hinton (2012), foi adotada a resolução de 224×224 *pixels*. Esta dimensão representa um equilíbrio eficaz entre preservar detalhes relevantes nas imagens para uma melhor diferenciação das classes e manter um custo computacional viável, visto que resoluções menores testadas apresentaram resultados inferiores, enquanto resoluções maiores testadas aumentaram o tempo de processamento sem ganho de performance.

Em seguida, foi aplicado um processo de normalização utilizando a média e o desvio padrão de cada canal de cor (RGB) das imagens. O objetivo desta etapa é ajustar a escala dos valores dos *pixels* para que sigam uma distribuição com média próxima de zero e desvio padrão unitário. Este ajuste torna a propagação do gradiente mais estável e acelera a convergência do modelo durante o treinamento. É fundamental ressaltar que os valores de média e desvio padrão foram calculados utilizando apenas o conjunto de dados de treinamento, e estes mesmos parâmetros foram então aplicados para normalizar tanto as imagens de treino quanto as de teste, garantindo que nenhuma informação do conjunto de teste influenciasse o processo de treinamento.

4.3 Arquitetura da Rede Neural Proposta

4.3.1 Componente de Entrada

O componente de entrada é responsável por preparar os sinais de ECG para o processamento pela rede. Para isso, segmentos de sinais de ECG de 2 segundos são convertidos em imagens bidimensionais. O estudo explorou duas diferentes técnicas de transformação, sendo elas GAF e HSFC.

4.3.2 Componente de Camadas Convolucionais

Este componente é a parte da rede neural convolucional (CNN) encarregada de receber a imagem de entrada e, progressivamente, capturar e refinar as características espaciais presentes nela. A estrutura dessas camadas é baseada na arquitetura das camadas convolucionais da AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A Tabela 3 apresenta os elementos desta componente.

Tabela 3 – Estrutura das camadas convolucionais

Nomes das camadas	Tamanho da saída	Tamanho do <i>Kernel</i>	<i>Padding</i>	<i>Stride</i>
Entrada	$3 \times [224 \times 224]$	—	—	—
<i>Features</i>				
Conv2D	$[3 \times 64]$	$[11 \times 11]$	$[2 \times 2]$	$[4 \times 4]$
ReLU	—	—	—	—
MaxPool2D	—	$[3]$	$[0]$	$[2]$
Conv2D	$[64 \times 192]$	$[5 \times 5]$	$[2 \times 2]$	$[1 \times 1]$
ReLU	—	—	—	—
MaxPool2D	—	$[3]$	$[0]$	$[2]$
Conv2D	$[192 \times 384]$	$[3 \times 3]$	$[1 \times 1]$	$[1 \times 1]$
ReLU	—	—	—	—
Conv2D	$[384 \times 256]$	$[3 \times 3]$	$[1 \times 1]$	$[1 \times 1]$
ReLU	—	—	—	—
Conv2D	$[256 \times 256]$	$[3 \times 3]$	$[1 \times 1]$	$[1 \times 1]$
ReLU	—	—	—	—
MaxPool2D	—	$[3]$	$[0]$	$[2]$
<i>Avgpool</i>				
AdaptiveAvgPool2D	$256 \times [6 \times 6]$	—	—	—

Fonte: Autor

Em essência, as camadas convolucionais aplicam filtros para identificar padrões locais, enquanto as camadas de *pooling* reduzem a dimensionalidade dos mapas de características, conferindo robustez à rede contra pequenas variações nos dados de entrada.

A estrutura inicia com a camada de entrada, ela recebe as imagens que representam os segmentos de sinais de ECG. Ela tem o formato $3 \times [224 \times 224]$, onde cada imagem possui 3 canais de cor (RGB) e uma resolução de 224 por 224 *pixels*. A extração de características começa com a primeira camada convolucional (Conv2D), que emprega 64 filtros de 11 por 11 *pixels*, com um passo de 4 e *padding* de 2 *pixels*. A esta, segue-se uma função de ativação ReLU, que introduz a não-linearidade essencial para o aprendizado de padrões complexos, e uma camada de *Max Pooling* 2D, que reduz a dimensionalidade do mapa de características ao selecionar o valor máximo em janelas de 3 por 3 *pixels* com passo de 2, conferindo robustez à rede contra pequenas variações na entrada.

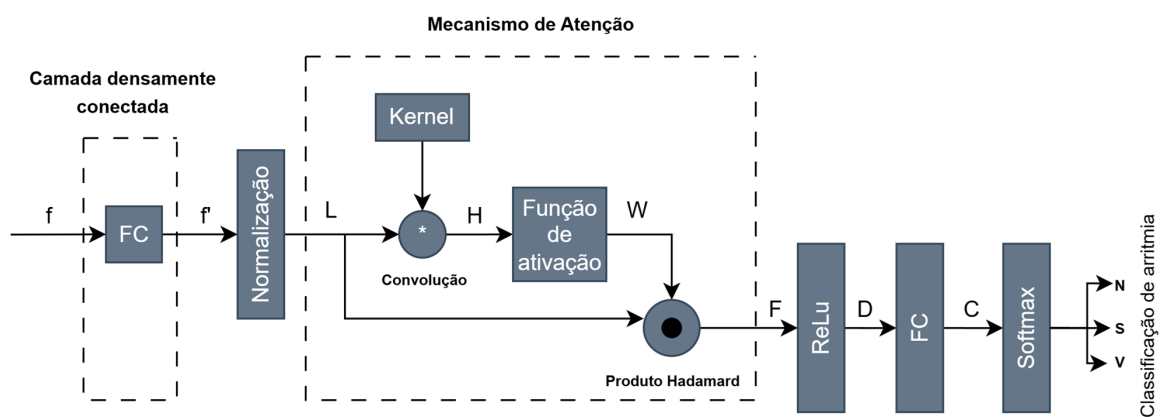
A rede prossegue com uma sequência de camadas semelhantes, aprofundando a extração de características. A segunda camada Conv2D utiliza 192 filtros de 5 por 5 *pixels* com *padding* de 2 e passo de 1, seguida por ReLU e outro *Max Pooling* 2D de *kernel* 3 e passo 2. Em seguida, três camadas convolucionais consecutivas refinam ainda mais os padrões: a terceira com 384 filtros de 3 por 3 *pixels*, a quarta com 256 filtros de 3 por 3 *pixels* e a quinta também com 256 filtros de 3 por 3 *pixels*, todas aplicando *padding* de 1 e passo de 1. Cada uma dessas camadas convolucionais é sucedida por uma função ReLU, e após a quinta, uma última camada de *Max Pooling* 2D (*kernel* 3, passo 2) é aplicada. A finalidade dessas camadas mais profundas é aprender características mais abstratas e complexas, com filtros menores e passo unitário, permitindo uma análise mais detalhada dos dados.

Finalmente, uma camada de *Adaptive Average Pooling* 2D redimensiona os mapas de características resultantes para uma dimensão espacial fixa de 6 por 6 *pixels*, mantendo 256 canais. Esta camada é crucial para criar uma representação compacta e consistente que servirá como entrada padronizada para o componente classificador com mecanismo de atenção, onde a decisão final sobre o tipo de arritmia será tomada com base nas características hierárquicas aprendidas.

4.3.3 Classificador com Mecanismo de Atenção

Após as camadas convolucionais extraírem as características da imagem, o vetor resultante é passado para o componente classificador com mecanismo de atenção apresentado na Figura 12.

Figura 12 – Classificador com mecanismo de atenção



Fonte: Autor. Adaptado de (Di Paolo, 2025).

O vetor de características proveniente das camadas convolucionais é inicialmente achatado (*flatten*) para formar um vetor unidimensional f . Este vetor é então processado por um camada densamente conectada (FC) com 4096 neurônios, responsável por extrair e aprender

padrões dos dados de entrada, que inclui camadas de *Dropout*, Linear e funções de ativação ReLU. Essa camada mapeia o vetor de características para um novo espaço de representação reduzido f' com dimensão 1×1024 .

Após o processamento pelas camadas totalmente conectadas, o vetor de características f' é normalizado no intervalo $[-1, 1]$ através da operação:

$$L = \frac{f'}{\max(|f'|)} \quad (4.1)$$

onde $\max(|f'|)$ representa o valor absoluto máximo do vetor f' . Esta normalização é fundamental para estabilizar o processo de treinamento, evitando gradientes instáveis durante a retropropagação, especialmente considerando que as imagens podem apresentar grande variação de intensidade de *pixel* e distribuição de cor. O vetor normalizado L possui dimensão 1×1024 e serve como entrada para o mecanismo de atenção, que aprende a dar maior peso às características ou regiões mais informativas da entrada, e simultaneamente atenuar a influência de atributos menos relevantes.

O mecanismo de atenção proposto baseia-se em autoatenção (*self-attention*) local, aplicando uma operação de convolução unidimensional sobre o vetor de características normalizado L para gerar um mapa de atenção. O vetor H é obtido através da convolução do vetor L com um *kernel* de dimensão 1×3 , conforme:

$$H[i] = L * kernel = \sum_{k=1}^3 L[i + k - 2] \cdot kernel[k] \quad (4.2)$$

onde o índice i varia de 0 até 1023 para percorrer as 1024 posições do vetor L . A convolução pondera as informações espaciais do vetor, capturando relações entre elementos adjacentes. O *kernel* utilizado no mecanismo de atenção pode ser configurado de duas formas distintas:

- **Kernel Fixo:** Os valores do *kernel* são pré definidos como $[-1, 9, -1]$, mantendo-se constantes durante todo o processo de treinamento. O uso de um kernel com estes valores em mecanismos de atenção tem precedentes na literatura, com Ahmad et al. (2021) e Tang et al. (2023) empregando a propriedade desses valores atuarem como "high boost filters" em seus módulos, conforme citado evidenciado e discutido por Di Paolo (2025).
- **Kernel Adaptativo:** Os parâmetros do *kernel* são inicializados como $[-1, 9, -1]$ e ajustados durante o processo de aprendizado via retropropagação, permitindo que a rede defina os padrões de atenção mais adequados para a tarefa específica.

Como discutido por Di Paolo (2025), a implementação direta da convolução reduziria o vetor L de 1024 para 1022 posições devido ao deslocamento do *kernel*. Para preservar as dimensões originais, é aplicada uma estratégia de *padding* antes da convolução, aumentando

temporariamente o vetor L com valores de preenchimento nas extremidades. Seguindo esta estratégia, mantém-se as 1024 dimensões após a convolução. Esta abordagem garante que o vetor H resultante preserve a mesma dimensionalidade do vetor de entrada L .

O vetor de pesos de atenção W é obtido aplicando uma função de ativação ao vetor H . Duas funções de ativação foram investigadas:

$$W = \text{Sigmoid}(H) \quad (4.3)$$

$$W = \text{Tanh}(H) \quad (4.4)$$

A função Sigmóide mapeia os valores para o intervalo $[0, 1]$, enquanto a função Tanh mapeia para o intervalo $[-1, 1]$. Ambas as funções permitem que a rede atribua diferentes níveis de importância às características extraídas.

O vetor de saída do mecanismo de atenção F é obtido através do produto elemento a elemento entre o vetor de pesos W e o vetor de características normalizado L :

$$F = W \odot L \quad (4.5)$$

onde $\odot$ representa a operação de multiplicação elemento a elemento (Produto Hadamard). O vetor resultante F mantém a dimensão 1×1024 e apresenta as características globais com suas importâncias relativas indicadas pelos pesos de atenção, produzindo uma representação refinada que destaca as características mais relevantes para a tarefa de classificação.

O vetor F , enriquecido pelo mecanismo de atenção, é direcionado para a sequência final de classificação composta por:

- **Função ReLU:** Introduce não-linearidade adicional ao modelo, permitindo a modelagem de padrões complexos:

$$D = \text{ReLU}(F) \quad (4.6)$$

- **Camada Linear Totalmente Conectada (FC):** Mapeia o vetor D de 1024 dimensões para o número de classes de saída (3 classes: N, S, V):

$$C = \text{FC}(D) \quad (4.7)$$

- **Função Softmax:** Gera a distribuição de probabilidade para cada classe, efetuando a classificação final das arritmias cardíacas.

A arquitetura proposta permite que o mecanismo de atenção local identifique e priorize automaticamente as características mais discriminativas para a classificação de arritmias, procu-

rando melhorar a capacidade do modelo de distinguir entre os padrões das classes Normal (N), Supraventricular (S) e Ventricular (V).

4.4 Configuração de Experimentos

Os experimentos foram conduzidos em ambiente computacional controlado, com configuração específica de *hardware* e *software*. O treinamento da rede utilizou o otimizador Adam (*Adaptive Moment Estimation*) (KINGMA; BA, 2015), devido à sua eficiência na atualização adaptativa dos pesos e convergência estável mesmo em redes profundas, com taxa de aprendizado inicial definida e estratégia de decaimento baseada no desempenho durante as épocas. Os hiperparâmetros, como tamanho do *batch* e número máximo de épocas, foram ajustados previamente. As configurações para treinamento empregadas incluíram:

- Taxa de aprendizado inicial: 4×10^{-4}
- Tamanho do *batch*: 128 amostras
- Número de épocas: 200
- Taxa de aprendizado: *scheduler* "ReduceLROnPlateau" configurado para reduzir a taxa de aprendizado em 50% a cada 10 épocas de treinamento sem redução do erro de treinamento, com critério de parada após 4 decaimentos consecutivos da taxa de aprendizagem.

Durante a fase de treinamentos e experimentos a função custo a ser minimizada foi baseada na *Cross-Entropy Loss*. A *Cross-Entropy Loss* pode ser matematicamente expressa como:

$$CrossEntropyLoss = - \sum (y_{true} * \log(softmax(y_{pred}))) \quad (4.8)$$

onde $y_{true} \in \{0, 1, 2\}$ é a saída desejada para o caso de 3 classes, $y_{pred} \in \mathbb{R}^3$ é a saída prevista por classe e *softmax* se refere à função de ativação *softmax* aplicada às predições brutas do modelo.

A *Cross-Entropy Loss* é comumente usada em problemas de classificação multiclasse, especialmente em contextos de aprendizado profundo com redes neurais convolucionais. A ideia principal é penalizar severamente as predições incorretas do modelo, onde a penalização aumenta exponencialmente à medida que a confiança na predição incorreta aumenta. A perda é zero quando a predição do modelo está correta e com máxima confiança, e aumenta drasticamente quando a predição do modelo está incorreta. Para lidar com o desbalanceamento das classes N, S e V, foram aplicados pesos de [1.0, 2.0, 1.0] respectivamente, priorizando a classe minoritária S (Supraventricular) através da *WeightedCrossEntropyLoss* disponível na biblioteca PyTorch.

Para possibilitar uma avaliação abrangente da contribuição isolada do mecanismo de atenção local proposto, bem como das diferentes técnicas de transformação de sinais em imagens, uma arquitetura de rede neural baseada na AlexNet padrão também foi treinada sob as mesmas condições experimentais. Esta versão de linha de base, referida nos resultados como "AlexNet base", tem a saída das camadas convolucionais conectada a três camadas totalmente conectadas, onde a classificação das classes N, S e V é realizada através da função de perda *CrossEntropyLoss* que incorpora implicitamente a função *softmax*. Importante ressaltar que os parâmetros de treinamento, como o otimizador Adam, a taxa de aprendizado inicial, o tamanho do *batch* e o número de épocas, foram mantidos idênticos aos utilizados para a arquitetura proposta com mecanismo de atenção. Essa estratégia garante uma comparação equitativa e permite quantificar os ganhos de desempenho atribuíveis especificamente à integração do mecanismo de atenção local e à escolha das técnicas de transformação de imagem (GAF e HSFC).

Foram realizados experimentos exploratórios para definir a melhor técnica de transformação de sinais de ECG em imagens, considerando GAF e HSFC. Estratégias de aumento de dados e balanceamento não foram aplicadas, para avaliar a robustez da arquitetura de forma direta. A estrutura final da rede foi mantida fixa ao longo dos testes, sendo variada apenas a função de ativação (Tanh ou Sigmoid) e o tipo de *kernel* (fixo ou adaptativo) no mecanismo de atenção.

As implementações foram realizadas em Python 3.9.15, utilizando a biblioteca PyTorch v2.6.0. O *hardware* utilizado consistiu em ambiente computacional local: Ryzen 5 5500U, 8 GB RAM.

4.5 Medidas de desempenho

A avaliação do desempenho de um classificador, especialmente em um domínio tão sensível e complexo como a classificação de arritmias cardíacas, exige a utilização de um conjunto de medidas de desempenho que reflitam a robustez e a aplicabilidade clínica do modelo. Em consonância com as diretrizes estabelecidas pela norma ANSI/AAMI EC57:2012 R2020, que especifica os indicadores de desempenho para sistemas automáticos de classificação de arritmias, este trabalho empregou um conjunto de indicadores padrão: Acurácia (*Accuracy*), Precisão (*Precision*), Sensibilidade (*Recall*), Especificidade (*Specificity*) e *F1-Score*.

A Acurácia é uma medida de desempenho intuitiva que mede a proporção de classificações corretas em relação ao total de amostras no conjunto de dados. Para enriquecer a avaliação, utiliza-se a Precisão, que indica a proporção de identificações positivas que foram realmente corretas. A Sensibilidade, por sua vez, foca na capacidade do modelo de detectar corretamente todos os casos positivos reais e, complementar à Sensibilidade, a Especificidade mede o quão bem o modelo acerta ao identificar a ausência da condição. Por fim, o *F1-score* representa o equilíbrio entre a Precisão e a Sensibilidade, sendo calculado como a média harmônica dessas duas medidas de desempenho.

Como apresentado na seção 4.1, o banco de dados é intrinsecamente desbalanceado. Visto a decisão pela não utilização de aumento de dados (*data augmentation*), a avaliação do modelo sob estas condições fornece uma medida direta de sua capacidade de lidar com a variabilidade e a escassez de dados para as classes de arritmia, um desafio para a criação de banco de dados balanceados que evidencia o ambiente clínico real.

A decisão metodológica do uso da medida de desempenho *F1-Score* é, assim, reforçada, já que esta medida de desempenho considera tanto a precisão quanto o *recall* em um único valor harmônico, e assim modelos com baixa Precisão (muitos falsos positivos) ou baixa Sensibilidade (muitos falsos negativos) para a classe minoritária resultam em um *F1-Score* significativamente baixo, oferecendo uma visão mais realista do desempenho do classificador (Di Paolo, 2025).

5 RESULTADOS EXPERIMENTAIS

Neste capítulo serão apresentados os resultados obtidos por meio dos experimentos realizados com a arquitetura proposta na Figura 12 para classificação de arritmias via sinais ECG, considerando paradigma interpaciente.

A arquitetura da CNN é baseada na AlexNet e incorpora um mecanismo de atenção local (*Self-Attention*). Os experimentos se concentraram principalmente na avaliação do impacto de diferentes funções de ativação (Tanh e Sigmoid) e tipos de *kernel* (fixo ou adaptativo) no mecanismo de atenção e na determinação da melhor técnica de transformação do segmento de sinal ECG em imagens para servir de entrada para o modelo proposto, tendo sido analisadas as técnicas GAF e HSFC.

O estudo foi conduzido utilizando a base de dados MIT-BIH (*The Massachusetts Institute of Technology - Beth Israel Hospital Arrhythmia Database*). Adotou-se a abordagem interpaciente, na qual os dados para treinamento e teste provêm de pacientes distintos. As classes de arritmias consideradas para classificação foram N (Normal), S (Supraventricular) e V (Ventricular). Apenas a derivação MLII (*Modified Lead II*) do ECG foi considerada para a classificação, visto ser uma derivação onipresente em todos os registros da base MIT-BIH e amplamente utilizada para medições de longo período. Nenhuma estratégia de aumento ou balanceamento de dados foi aplicada, a fim de avaliar a robustez da arquitetura de forma direta.

O treinamento da rede utilizou o otimizador Adam, com uma taxa de aprendizado inicial e uma estratégia de decaimento baseada no desempenho ao longo das épocas. Os hiperparâmetros, como o tamanho do *batch* (128 imagens) e o número máximo de épocas (300), foram previamente ajustados.

A Tabela 4 apresenta os resultados quantitativos globais obtidos, considerando a base de teste, nos experimentos conduzidos com oito variações arquiteturais, combinando dois tipos de transformação de sinais temporais em imagens (GAF e HSFC) e duas funções de ativação (tangente hiperbólica e sigmoide) com dois tipos de *kernel* (fixo e adaptativo) no mecanismo de atenção da rede. O conjunto de dados utilizado seguiu a abordagem interpaciente da base MIT-BIH, considerando as classes N, S e V.

Tabela 4 – Indicadores de desempenho globais na base de teste: Estrutura com mecanismo de atenção

Cenário	Transformação para Imagem	Função de ativação	Kernel	Acurácia	Precisão	Recall	Especificidade	F1-score
1	GAF	Tangente	Adaptativo	0,9788	0,7745	0,7614	0,9539	0,7678
2	GAF	Tangente	Fixo	0,9793	0,7735	0,7385	0,9501	0,7544
3	GAF	Sigmoide	Adaptativo	0,9744	0,7676	0,7329	0,9495	0,7442
4	GAF	Sigmoide	Fixo	0,9733	0,7565	0,7473	0,9391	0,7512
5	HSFC	Tangente	Adaptativo	0,9740	0,7338	0,7396	0,9486	0,7366
6	HSFC	Tangente	Fixo	0,9758	0,7373	0,7006	0,9399	0,7121
7	HSFC	Sigmoide	Adaptativo	0,9690	0,6956	0,7065	0,9410	0,7004
8	HSFC	Sigmoide	Fixo	0,9704	0,7096	0,7031	0,9382	0,7040

Fonte: Próprio Autor.

A Tabela 5 apresenta o desempenho detalhado por classe (N, S e V) para todos os cenários testados conforme apresentado na Tabela 4, permitindo uma análise granular da capacidade de classificação de cada configuração para as diferentes classes de arritmia.

Tabela 5 – Indicadores de desempenho por classe na base de teste: Estrutura com mecanismo de atenção

Cenário	Acurácia (%)	F1-score (%)			Precisão (%)			Sensibilidade (%)			Especificidade (%)		
		N	S	V	N	S	V	N	S	V	N	S	V
1	97,88	99	36,67	95	98,87	37,24	96,26	99,11	34,93	94,37	87,29	99,16	99,73
2	97,93	99,04	32,44	94,84	98,76	36,67	96,61	99,33	29,09	93,13	85,98	99,28	99,76
3	97,44	98,87	30,51	93,86	98,76	39,86	91,67	98,98	24,74	96,15	86,02	99,47	99,36
4	97,33	98,66	33,66	93,02	98,49	31,99	96,46	98,83	35,52	89,84	83,03	98,92	99,76
5	97,40	98,66	27,98	94,32	98,76	28,71	92,68	98,57	27,28	96,02	86,10	99,04	99,45
6	97,58	98,76	20,99	93,86	98,50	29,92	92,78	99,02	16,19	94,96	83,07	99,46	99,46
7	96,90	98,42	20,31	91,41	98,59	21,71	88,39	98,26	19,04	94,65	84,19	99,02	99,10
8	97,04	98,48	21,74	91,02	98,49	26,05	88,33	98,47	18,59	93,88	83,11	99,25	99,10

Fonte: Autor

Para estabelecer um ponto de referência comparativo robusto e avaliar o impacto da integração do mecanismo de atenção local, a AlexNet padrão também foi treinada sob as mesmas condições experimentais, conforme detalhado na Seção 4.4. A Tabela 6 apresenta os melhores indicadores de desempenho globais (Acurácia, Precisão, *Recall*, Especificidade e *F1-score*) obtidos com esses modelos AlexNet base, utilizando as técnicas de transformação de imagem GAF e HSFC na entrada da rede.

Tabela 6 – Indicadores de desempenho globais na base de teste: Modelo base

Transformação para Imagem	Modelo base	Acurácia	Precisão	Recall	Especificidade	F1-score
GAF	AlexNet	0,9774	0,7631	0,7120	0,9397	0,7337
HSFC	AlexNet	0,9593	0,6943	0,7337	0,9366	0,7054

Fonte: Próprio Autor.

A Tabela 7 apresenta os indicadores de desempenho detalhados por classe (N, S e V)

para os modelos base AlexNet (com GAF e HSFC), a fim de permitir uma comparação direta com os modelos propostos com mecanismo de atenção.

Tabela 7 – Indicadores de desempenho por classe: Modelo base

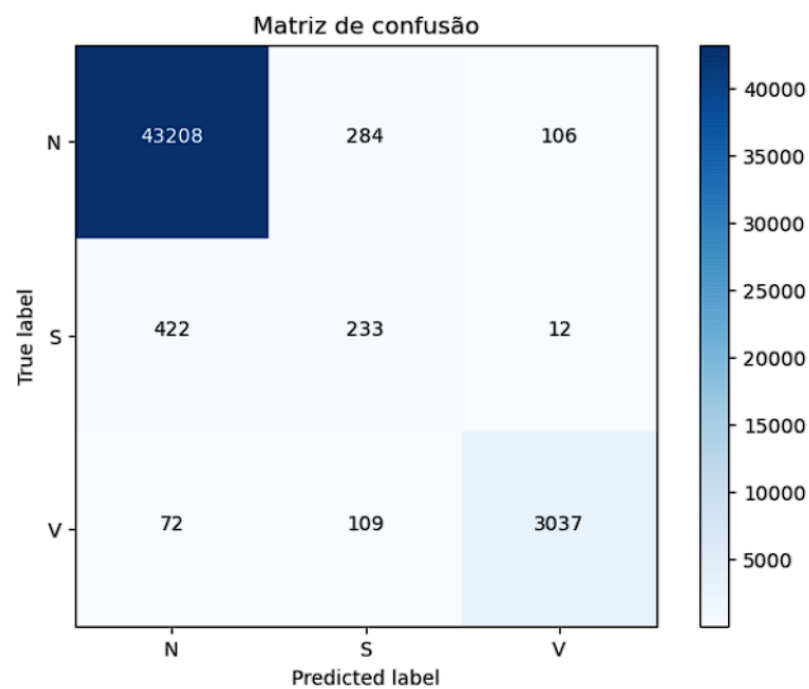
Modelo	Acurácia (%)	F1-score (%)			Precisão (%)			Sensibilidade (%)			Especificidade (%)		
		N	S	V	N	S	V	N	S	V	N	S	V
AlexNet base - HSFC	95,94	97,92	21,93	91,78	98,54	16,58	93,18	97,31	32,38	90,43	83,78	97,68	99,52
AlexNet base - GAF	97,74	98,92	27,26	93,94	98,48	33,93	96,52	99,36	22,79	91,49	82,81	99,37	99,76

Fonte: Autor

A partir da observação dos resultados apresentados na Tabela 4 e 5, foi determinado o melhor resultado, sendo obtido a partir da estrutura do cenário 1: Alexnet com técnica *Gramian Angular Field* (GAF) para transformação do sinal ECG em imagem, mecanismo de atenção com função de ativação Tangente Hiperbólica (Tanh) e *kernel* adaptativo. A seleção desta arquitetura como a de melhor desempenho baseou-se no *F1-score* geral, uma medida de desempenho priorizada devido à natureza intrinsecamente desbalanceada da base de dados MIT-BIH e ao crítico interesse clínico na detecção precisa das classes minoritárias (S e V), já que o *F1-score*, ao oferecer um equilíbrio entre precisão e sensibilidade, proporciona uma avaliação mais fidedigna do desempenho do classificador em cenários com distribuição de classes desigual. A granularidade da Tabela 5 permite uma avaliação precisa da contribuição do mecanismo de atenção, especialmente na classificação das classes minoritárias.

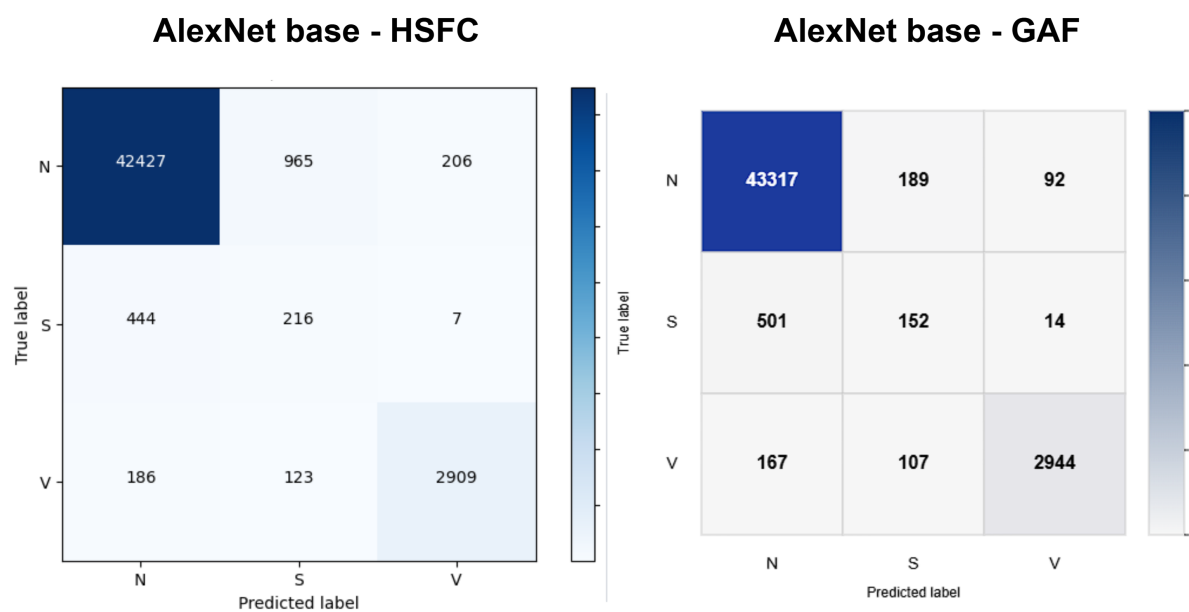
Para complementar a análise, as matrizes de confusão fornecem uma visualização direta do desempenho. A Figura 13 ilustra a distribuição das classificações para a melhor estrutura do modelo proposto com mecanismo de atenção, enquanto a Figura 14 apresenta as matrizes de confusão para os modelos AlexNet base que utilizam as transformações GAF e HSFC, respectivamente.

Figura 13 – Matriz de confusão para o melhor cenário com módulo de atenção



Fonte: Autor

Figura 14 – Matrizes de confusão - AlexNet base com HSFC e GAF



Fonte: Autor

Ao comparar estes cenários, percebemos melhorias no equilíbrio e capacidade de classificação dentre as classes, quando analisamos a melhor estrutura com módulo de atenção, em

relação aos modelos AlexNet base. No próximo capítulo será apresentada uma análise detalhada dos resultados obtidos na fase de experimentos.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Este capítulo apresenta a análise e discussão dos resultados obtidos nos experimentos realizados com as arquiteturas de CNN analisadas (com módulo de atenção e sem módulo de atenção) para classificação de arritmias cardíacas.

6.1 Análise Geral dos Resultados

Os resultados demonstraram que todas as variações da arquitetura CNN analisadas durante a fase de experimentos alcançaram um desempenho elevado em termos de acurácia, com valores superiores a 97%. No entanto, a medida de desempenho de *F1-score* é considerada a mais relevante neste contexto, dada a natureza desbalanceada da base de dados MIT-BIH e o interesse clínico na detecção eficaz das classes minoritárias (S e V). O *F1-score* fornece um equilíbrio entre precisão e sensibilidade, sendo mais indicativo do desempenho real em cenários com classes desproporcionais.

A melhor configuração encontrada, conforme apresentado na Tabela 4, utilizou a técnica GAF para transformação do sinal ECG em imagem e o mecanismo de atenção com função de ativação Tanh e com o *kernel* adaptativo, alcançando o *F1-Score* de 0,7678. Esta configuração também apresentou valores elevados de precisão (0,7745) e *recall* (0,7614). Este resultado corrobora a hipótese de que a combinação de uma representação visual eficaz como a GAF, que preserva as características temporais do ECG, com a expressividade da função Tanh (que permite modulações de pesos em $[-1, 1]$) e a flexibilidade do *kernel* adaptativo (que se ajusta dinamicamente aos padrões dos dados durante o treinamento), potencializa significativamente a capacidade discriminativa do mecanismo de atenção.

A função Tanh, ao permitir pesos de atenção em um intervalo simétrico de $[-1, 1]$, oferece uma maior nuance na ponderação das características extraídas. Isso possibilita que o modelo destaque atributos relevantes de forma mais precisa. Por outro lado, o *kernel* adaptativo, por ser treinável, ajusta-se dinamicamente aos padrões dos dados. Essa capacidade de aprendizado permite que o mapa de atenção seja refinado ao longo do processo de treinamento, o que foi essencial para atingir o equilíbrio entre sensibilidade e precisão observado no melhor cenário. A GAF, ao transformar para um sistema de coordenadas polares, codifica a informação de tempo e magnitude do sinal ECG na coloração e ângulos da imagem resultante. Essa abordagem é crucial, pois permite à GAF preservar a estrutura temporal e as dependências intrínsecas dos dados do ECG, traduzindo-as para um formato bidimensional interpretável.

Embora a configuração usando GAF para transformação do sinal ECG em imagem e o mecanismo de atenção com função de ativação Tanh e *kernel* fixo também tenha apresentado um desempenho competitivo (*F1-score* de 0,7544), ela foi superada nos principais indicadores

pela versão com *kernel* adaptativo. Isso sugere que a capacidade de aprendizado do *kernel* contribui com ganhos adicionais, mesmo quando uma função de ativação expressiva como a Tanh é utilizada.

As configurações com GAF que empregaram a função Sigmoid, embora robustas, mostraram um desempenho inferior nos cenários testados, especialmente no *F1-score*. Isso é esperado, uma vez que a função Sigmoid limita os pesos ao intervalo $[0, 1]$, o que restringe a capacidade da rede de aplicar penalizações mais severas a características menos relevantes.

Em relação aos resultados usando HSFC, para todos os experimentos, considerando todas as combinações do mecanismo de atenção, pode ser observado na Tabela 4, que os resultados foram inferiores para os modelos usando a GAF, indicando desta forma a superioridade da técnica na representação dos sinais ECG, destacando características importantes do sinal para o processo de classificação.

6.1.1 Análise comparativa de modelo base variando transformação de imagem

Analisando as Tabelas 6 e 7 e a Figura 14, observamos que o modelo de AlexNet base com imagens de GAF teve melhores resultados gerais se comparado com o modelo com imagens de HSFC, mas ainda piores do que os modelos com os mecanismos de atenção.

A matriz de confusão da técnica HSFC (Figura 14 - esquerda) mostra um desempenho robusto na classificação de batimentos normais (N), com um alto número de verdadeiros positivos ($VP = 42.427$). No entanto, ela apresenta uma alta taxa de falsos negativos (FN) para a classe N ($965 + 206 = 1.171$), o que indica que um grande número de batimentos normais foi incorretamente classificado como de outras classes. Em contrapartida, a matriz de confusão da técnica GAF (Figura 14 - direita) exibe uma melhoria notável na classificação da classe N, com mais verdadeiros positivos ($VP = 43.317$) e um número significativamente menor de falsos negativos ($189 + 92 = 281$). Isso sugere que o GAF preserva as características da classe N de forma mais eficaz para o modelo AlexNet.

Essa diferença de desempenho é refletida nas medidas de desempenho globais da Tabela 6. O modelo que utiliza a transformação GAF alcança uma acurácia de 97,74%, superando a acurácia de 95,93% obtida com a transformação HSFC. Adicionalmente, o GAF demonstra uma precisão e um *F1-score* superiores, com valores de 0,7631 e 0,7337, respectivamente, em comparação com os valores de 0,6943 e 0,7054 do HSFC. A única medida de desempenho em que o HSFC supera o GAF é a sensibilidade (*recall*). O modelo HSFC atinge 0,7337 de sensibilidade, enquanto o GAF alcança 0,7120. Essa maior sensibilidade do HSFC, no entanto, vem acompanhada de uma precisão mais baixa, o que sugere que o modelo HSFC está classificando um maior número de batimentos supraventriculares (S) e ventriculares (V) como positivos, mas com uma taxa maior de erros.

Em resumo, a análise demonstra que, para a arquitetura AlexNet, a transformação GAF

é mais adequada para a classificação de arritmias, pois, embora apresente uma sensibilidade ligeiramente inferior, proporciona um equilíbrio geral superior entre precisão e *recall*, resultando em um *F1-score* e acurácia mais elevados.

6.1.2 Análise comparativa do modelo base GAF com o melhor modelo com módulo de atenção

Para fins de comparação e explanação de ganhos, vamos focar nossa análise no modelo base que utiliza a transformação GAF, não somente pelo fato de utilizar a mesma transformação de imagem que nosso melhor modelo com módulo de atenção, como também por ter apresentado as melhores medidas de desempenho dentre os modelos base.

Analizando os resultados da Tabela 6, a comparação entre a AlexNet base com transformação GAF e a melhor estrutura (AlexNet com módulo de atenção, transformação GAF, tangente hiperbólica e *kernel* adaptativo) revela ganhos significativos. Embora a AlexNet base com GAF apresente uma acurácia global elevada de 97,74% e um excelente desempenho na detecção de batimentos normais (*F1-score* = 98,92%) e ventriculares (*F1-score* = 93,94%), ela ainda demonstra uma performance limitada na identificação de batimentos supraventriculares (*F1-score* = 27,26%), refletindo a tendência do modelo em focar nas classes majoritárias.

Ao incorporar o módulo de atenção, a acurácia global da melhor estrutura proposta aumenta ligeiramente para 97,88%. Contudo, o mais importante é a melhora considerável nas medidas de desempenho da classe S (supraventricular), na qual o *F1-score* salta de 27,26% para 36%, a precisão aumenta de 33,93% para 37,24%, e a sensibilidade passa de 22,79% para 34,93%, demonstrando ganhos significativos na nossa medida de desempenho alvo. Para a classe N (Normal), observou-se que o *F1-score* teve um aumento marginal de aproximadamente 0,08% (de 98,92% para 99%), a precisão aumentou em cerca de 0,40% (de 98,48% para 98,87%), e a sensibilidade apresentou uma leve variação negativa de aproximadamente 0,27% (de 96,52% para 96,26%). Para a classe V (Ventricular), o *F1-score* registrou um aumento de aproximadamente 1,13% (de 93,94% para 95%), a precisão teve uma pequena variação negativa de cerca de 0,25% (de 99,36% para 99,11%), e a sensibilidade cresceu em aproximadamente 3,15% (de 91,49% para 94,37%).

Esses resultados indicam que o mecanismo de atenção consegue realçar características relevantes de cada batimento, tornando o classificador mais equilibrado e eficaz na identificação de ritmos supraventriculares e ventriculares sem comprometer o desempenho na detecção de batimentos normais.

A matriz de confusão da melhor estrutura também evidencia um classificador mais equilibrado. Por exemplo, na classe S, o modelo demonstra uma capacidade melhorada de identificar corretamente os batimentos S, com 233 verdadeiros positivos. Embora os Falsos Positivos da classe S ($284 + 109 = 393$) e os Falsos Negativos ($422 + 12 = 434$), o modelo consegue uma boa classificação. Os ganhos nas classes minoritárias evidenciam que o modelo de atenção mitigou o viés para a classe majoritária, resultando em um desempenho mais robusto.

Em termos de falsos positivos (FP) e falsos negativos (FN), as medidas de desempenho da melhor estrutura evidenciam um classificador mais equilibrado. Enquanto o modelo AlexNet base com GAF tem um número robusto de verdadeiros positivos na classe N ($VP = 43.317$), ele falha significativamente na classe S ($VP = 152$), com muitos Falsos Negativos ($501 + 14 = 515$) e Falsos Positivos ($189 + 107 = 296$). Isso demonstra que, na sua forma original, o modelo tem uma forte tendência a classificar batimentos minoritários como normais. Em contraste, a melhor estrutura aumenta os VPs da classe S para 233 e, na classe V, de 2944 para 3037, melhorando a detecção das classes minoritárias. Embora a melhor estrutura classifique um pouco mais de batimentos de forma incorreta para a classe S, com Falsos Positivos ($109 + 284 = 393$), essa distribuição mais homogênea de erros demonstra que o mecanismo de atenção consegue mitigar o viés do modelo original, tornando-o mais capaz de realçar características relevantes de cada batimento e, consequentemente, mais apto a identificar ritmos supraventriculares e ventriculares de forma mais precisa, sem comprometer substancialmente o desempenho geral na detecção de batimentos normais, o que é uma característica importante para o cenário clínico real.

6.2 Comparação com modelos propostos na literatura

A Tabela 8 apresenta, para comparação com nosso melhor resultado, medidas de desempenho obtidas para modelos propostos em trabalhos já apresentados na literatura, sendo que para uma comparação justa, foram escolhidos trabalhos que adotaram a mesma divisão DS1 e DS2 utilizada em nosso trabalho (abordagem interpaciente).

Tabela 8 – Indicadores de desempenho na base de teste de trabalhos correlatos

Referência	Abordagem Chave	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)	F1-score (%)
Chazal, O'Dwyer, Reilly (2004)	Modelo de classificador baseado em discriminante linear	83,88	45,57	66,00	96,05	59,74
Soria, Martinez (2009)	Classificação com discriminantes lineares e quadráticos	91,47	69,89	89,98	94,60	76,43
Llamedo, Martínez (2011)	Processamento de ondas R a partir de um filtro da onda QRS, com auxílio de Wavelets, com classificador quadrático discriminante	90,62	66,26	86,18	95,83	66,61
Lin, Yang (2014)	Classificador quadrático discriminante	93,00	67,60	83,50	-	-
Zhang et al. (2014)	Classificador SVM com metodologia one-versus-one	86,66	-	-	-	-
Garcia et al. (2017)	Extração de recursos em várias etapas: Representação do sinal ECG, rede complexa para extração de características, classificador SVM e PSO para ajuste de hiperparâmetros	92,38	70,12	81,10	92,12	74,59
Dias et al. (2021)	Segmentação própria por intervalo RR e classificador com algoritmo discriminante linear; foco em sinais ECG de uma única derivação (MLII)	80,58	59,37	84,84	94,35	60,77
WANG et al. (2021)	Transformada Wavelet Contínua e Rede CNN; Modelo multimodal com fusão de características via mecanismo de atenção; Combinação de sinais ECG e imagens GAF	97,48	70,75	67,47	96,01	68,76
Oliveira et al. (2022)	Representação de batimento como grafo (Visibility Graph) e classificação com GraphSAGE (GNNS)	86,12	76,72	77,04	92,87	76,88
DI PAOLO, Í. F.; CASTRO, A. R. G. (2024)	Modelo Multimodal; Técnicas de transformação de imagens Hilbert Space Filling Curve (HSFC) e Recurrence Plot (RP); Aumento de dados com SMOTE e WGAN-GP.	98,48	94,15	80,23	96,34	81,91
Nossa melhor estrutura	CNN com Mecanismo de Atenção Local (tangente hiperbólica, kernel adaptativo); Técnica de transformação de imagens Gramian Angular Field (GAF).	97,88	77,45	76,14	95,39	76,78

Fonte: Autor

A Tabela 8 mostra que o nosso melhor modelo apresenta desempenho competitivo frente a diferentes abordagens da literatura. Com *F1-score* de 76,78%, nossa estrutura supera os resultados de Dias et al. (2021) e Wang et al. (2021). No caso de Dias et al. (2021), os autores apresentam uma metodologia de segmentação própria a partir do intervalo RR e classificação com um algoritmo discriminante linear, atingindo *F1-score* de 60,77%. Já no caso de Chen et al. (2024) foi aplicada a Transformação Contínua de Wavelet (CWT) e uma CNN, reportando um *F1-score* global de 0,6876. O desempenho superior do modelo proposto pode ser atribuído ao uso de atenção local, à função de ativação tangente hiperbólica e ao *kernel* adaptativo, que contribuíram para melhorar a discriminação entre batimentos normais e anômalos mesmo sem

técnicas de balanceamento ou aumento de dados.

Nossa estrutura teve desempenho similar ao apresentado em Oliveira et al. (2022), onde os autores representaram cada batimento como um grafo via algoritmo *Visibility Graph* e empregaram GraphSAGE para classificação, conseguindo *F1-score* de 76,88%. Por outro lado, nosso trabalho apresentou desempenho inferior ao estudo de Zahid, Kiranyaz e Gabbouj (2023), que propôs uma rede operacional 1D (*Self-ONN*) com injeção de características de intervalo RR, alcançando *F1-score* de 92,27%.

Considerando o trabalho de (Di Paolo; CASTRO, 2024), que obteve *F1-score* de 0,8191 (para interpaciente), nossos resultados foram inferiores. A diferença de resultados pode ser explicada por fatores em nosso trabalho como:

- Ausência de aumento de dados (*data augmentation*);
- Não utilização de técnicas de balanceamento ou ponderação adaptativa por classe;
- Diferente função custo usada durante treinamento;
- Ausência de abordagem multimodal com uso de diferentes técnicas de imagens e diferentes derivações combinadas;

Apesar de nosso modelo apresentar medidas de desempenho menores em relação a alguns dos modelos comparados, os resultados obtidos demonstram robustez e coerência com os padrões da literatura, reforçando que o uso do mecanismo de atenção local, aliado à configuração tangente hiperbólica e *kernel* adaptativo, representa uma solução eficaz e de bom custo-benefício. A estrutura proposta consegue, inclusive, superar diversos trabalhos que apresentam arquiteturas mais complexas, mostrando sua aplicabilidade em cenários clínicos reais, mesmo sem recursos computacionais avançados.

7 CONCLUSÃO

Este trabalho apresentou uma abordagem para a classificação automática de arritmias cardíacas utilizando Redes Neurais Convolucionais (CNNs) com mecanismo de atenção local, aplicada sobre imagens derivadas de sinais de ECG transformados por *Gramian Angular Field* (GAF) ou *Hilbert Curve* (HSFC). O objetivo central foi avaliar como a função de ativação (tangente hiperbólica ou sigmoide), o tipo de transformação de imagem (GAF ou HSFC) e o tipo de *kernel* (fixo ou adaptativo) no módulo de atenção impactam o desempenho do classificador. Os experimentos demonstraram que:

- A função tangente hiperbólica, ao permitir ponderações simétricas (negativas e positivas), produziu os melhores resultados de *F1-score*, superando a sigmoide em todos os cenários;
- A técnica *Gramian Angular Field* (GAF) demonstrou consistentemente melhores resultados de desempenho na transformação dos sinais de ECG em imagens em comparação com a Curva de Preenchimento do Espaço de Hilbert (HSFC), fator atribuído à sua capacidade de preservar a estrutura temporal e as dependências intrínsecas do sinal de ECG;
- O *kernel* adaptativo mostrou desempenho superior ao fixo quando combinado com tangente hiperbólica, confirmando a importância de permitir que o modelo aprenda padrões específicos de atenção;
- A configuração tangente hiperbólica e *kernel* adaptativo obteve o maior *F1-score* (0,7678), além de elevados índices de acurácia e sensibilidade, evidenciando a eficácia da atenção local aprendida;
- A arquitetura geral alcançou desempenho robusto mesmo sem técnicas adicionais de aumento de dados ou balanceamento, o que ressalta sua eficiência em cenários clínicos reais.

Em síntese, os experimentos realizados confirmam a importância dos componentes internos do módulo de atenção proposto. Comparando a melhor configuração da arquitetura proposta (com GAF para transformação do sinal, função de ativação Tanh e *kernel* adaptativo) com o modelo AlexNet base utilizando a mesma técnica de transformação GAF, observou-se uma melhoria substancial no desempenho do classificador.

Especificamente, enquanto o modelo AlexNet base com GAF alcançou um *F1-score* global de 0,7337, a incorporação do mecanismo de atenção elevou esse indicador para 0,7678. Mais notavelmente, para a desafiadora classe S (supraventricular), o *F1-score* saltou de 27,26% para 36%. Essa melhoria substancial, em particular para a classe S, demonstra que o mecanismo

de atenção não só otimiza o desempenho geral, mas também mitiga o viés em relação à classe majoritária.

De forma diferenciada de outros trabalhos da literatura, como Ahmad et al. (2021) e Tang et al. (2023) que empregaram a função de ativação sigmoide em seus mecanismos de atenção, esta pesquisa demonstra a vantagem do uso da função tangente hiperbólica. Esta função, ao permitir ponderações simétricas (negativas e positivas), produziu os melhores resultados de *F1-score*, superando a sigmoide em todos os cenários avaliados. A combinação da tangente hiperbólica com um *kernel* adaptativo obteve o maior *F1-score* (0,7678), além de elevados índices de acurácia e sensibilidade, evidenciando a eficácia da atenção local aprendida com essa configuração específica. A arquitetura geral alcançou desempenho robusto mesmo sem técnicas adicionais de aumento de dados, balanceamento ou uso de novas bases de dados (como a "PTB", também indicada pela AAMI), o que ressalta sua eficiência em cenários clínicos reais. Portanto, a exploração da função tangente hiperbólica e do *kernel* adaptativo no módulo de atenção local representa uma contribuição significativa para o avanço da área de diagnóstico assistido por computador na cardiologia.

Em relação à comparação com modelos previamente apresentados na literatura, o desempenho da arquitetura proposta, especialmente em sua melhor configuração (mecanismo de atenção com tangente hiperbólica e *kernel* adaptativo, e transformação GAF), demonstrou ser altamente competitivo. Nosso modelo superou abordagens como a de Dias et al. (2021) (*F1-score* de 60,77%) e Wang et al. (2021) (*F1-score* de 68,97%), além de apresentar desempenho similar ao de Oliveira et al. (2022) (*F1-score* de 76,88%).

Adicionalmente, em comparação com o estudo de Di Paolo e Castro (2024), que reportou um *F1-score* de 0,8191 para o paradigma interpaciente, nossos resultados foram ligeiramente aquém, o que pode ser atribuído a diferenças metodológicas significativas. O trabalho de Di Paolo e Castro (2024) empregou estratégias de aumento de dados, técnicas de balanceamento de classes, uma função de custo diferente durante o treinamento e uma abordagem multimodal mais abrangente, combinando diferentes técnicas de imagem e múltiplas derivações, o que contribuiu para ganhos adicionais em cenários mais complexos.

Contudo, a robustez da nossa proposta é evidenciada pela capacidade de alcançar resultados expressivos, mantendo-se dentro do estado da arte, mesmo na ausência de técnicas adicionais de aumento ou balanceamento de dados. A eficácia do mecanismo de atenção local, aliado à escolha da função de ativação tangente hiperbólica e do *kernel* adaptativo, demonstrou ser uma solução potente para realçar características discriminativas, especialmente na classificação da classe S (supraventricular), sem comprometer o desempenho geral.

Como trabalhos futuros, sugere-se a aplicação de estratégias de aumento de dados e balanceamento de classes, a fim de mitigar os efeitos do desbalanceamento característico das bases de ECG e melhorar a generalização do modelo. Também se propõe a investigação de novas estruturas de mecanismos de atenção, incluindo módulos como o CBAM, que podem

ampliar a capacidade de foco seletivo da rede sobre padrões discriminativos. Além disso, a utilização de diferentes derivações simultaneamente pode viabilizar uma abordagem multimodal, enriquecendo as representações extraídas e potencializando o desempenho da arquitetura. Por fim, sugere-se a adoção de outras medidas de desempenho, tais como testes estatísticos, para aumentar a intuição e comprovação dos resultados. Tais aprimoramentos podem contribuir para alcançar níveis mais altos de sensibilidade e precisão, aproximando os resultados de modelos de referência na literatura e fortalecendo a aplicabilidade clínica da solução desenvolvida.

Referências

- AHMAD, Z. et al. ECG heartbeat classification using multimodal fusion. *IEEE Access*, v. 9, 2021. Disponível em: <<https://doi.org/10.1109/ACCESS.2021.3097614>>.
- ALHAZMI, M.; ALTAHHAN, A. Best fit activation functions for attention mechanism: Comparison and enhancement. In: *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2023. Disponível em: <<https://doi.org/10.1109/IJCNN54540.2023.10191885>>.
- BETTS, J. G. et al. *Anatomy and Physiology*. 2. ed. Houston, Texas, USA: OpenStax, Rice University, 2022. Disponível em: <<https://openstax.org/k12/anatomyand-physiology>>.
- BRAGIN, A.; SPITSYN, V. Neural network analysis of electroencephalograms graphical representation. In: *Anais do 30th International Conference on Computer Graphics and Machine Vision*. Saint Petersburg, Russia: [s.n.], 2020. Disponível em: <<http://dx.doi.org/10.51130/graphicon-2020-2-4-52>>.
- CASAS, M. M. et al. Bayesian classification models for premature ventricular contraction detection on ECG traces. *Journal of Healthcare Engineering*, v. 2018, 2018. Disponível em: <<https://doi.org/10.1155%2F2018%2F2694768>>.
- CHAZAL, P. de; O'DWYER, M.; REILLY, R. B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, v. 51, n. 7, 2004. Disponível em: <<https://doi.org/10.1109/tbme.2004.827359>>.
- CHEN, Z. 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*, v. 87, p. Part B, 2024. Disponível em: <<https://doi.org/10.1016/j.bspc.2023.105437>>.
- CLIFFORD, G. D.; AZUAJE, F.; MCSHARRY, P. *Advanced Methods And Tools for ECG Data Analysis*. 1. ed. [S.l.]: Artech House Publishers, 2006.
- DAMA, F.; SINOQUET, C. *Time series analysis and modeling to forecast: A survey*. 2021. Disponível em: <<http://arxiv.org/abs/2104.00164>>.
- Di Paolo Ítalo F. *Classificação automática de arritmias cardíacas através de redes neurais convolucionais multimodais com mecanismo de atenção*. Tese (Tese (Doutorado em Engenharia Elétrica)) — Universidade Federal do Pará, Belém, 2025.
- Di Paolo Ítalo F.; CASTRO, A. R. G. Intra- and interpatient ECG heartbeat classification based on multimodal convolutional neural networks with an adaptive attention mechanism. *Applied Sciences*, v. 14, n. 20, 2024. Disponível em: <<https://doi.org/10.3390/app14209307>>.
- DIAS, F. M. et al. Arrhythmia classification from single-lead ECG signals using the inter-patient paradigm. *Computer Methods and Programs in Biomedicine*, v. 202, 2021. Disponível em: <<https://doi.org/10.1016/j.cmpb.2021.105948>>.
- FAN, T. et al. A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition. *Computers in Biology and Medicine*, v. 159, 2023. Disponível em: <<https://doi.org/10.1016/j.combiomed.2023.106938>>.

- FAWAZ, H. I. et al. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, v. 33, n. 4, p. 917–963, 2019. Disponível em: <<https://doi.org/10.48550/arXiv.1809.04356>>.
- FRAHAN, A.; LI, C.; AHMED, M. T. A qualitative overview of fuzzy logic in ECG arrhythmia classification. *International Journal of Engineering Works*, v. 5, n. 11, 2018. Disponível em: <<https://doi.org/10.5281/zenodo.1486134>>.
- FRANCIS, J. ECG monitoring leads and special leads. *Indian Pacing and Electrophysiology Journal*, v. 16, n. 3, 2016. Disponível em: <<https://doi.org/10.1016/j.ipej.2016.07.003>>.
- GARCIA, G. et al. Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO. *Scientific Reports*, v. 7, 2017. Disponível em: <<https://doi.org/10.1038/s41598-017-09837-3>>.
- GIFFONI, R. T.; TORRES, R. M. T. Breve história da eletrocardiografia. *Revista Médica de Minas Gerais RMMG*, v. 20, n. 2, 2009. Disponível em: <<https://www.rmmg.org/artigo/detalhes/323>>.
- HALL, J. E. *Guyton and Hall Textbook of Medical Physiology*. 14. ed. EUA: Elsevier, 2021.
- HE, K. et al. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, EUA: [s.n.], 2016. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>.
- HONG, S. 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.combiomed.2020.103801>>.
- HOWARD, A. G. et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017. Disponível em: <<https://doi.org/10.48550/arXiv.1704.04861>>.
- HSIEH, C.-H. et al. Detection of atrial fibrillation using 1D convolutional neural network. *Sensors*, v. 20, n. 7, p. E2136, abr 2020. Disponível em: <<https://doi.org/10.3390/s20072136>>.
- HUANG, Y. et al. ECG classification based on guided attention mechanism. *Computer Methods and Programs in Biomedicine*, v. 257, 2024. Disponível em: <<https://doi.org/10.1016/j.cmpb.2024.108454>>.
- ISLAM, M. S. et al. HARD-C: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN. *Neural Networks*, v. 162, 2023. Disponível em: <<https://doi.org/10.1016/j.neunet.2023.03.004>>.
- JIANG, W. et al. Time series classification based on image transformation using feature fusion strategy. *Neural Processing Letters*, v. 54, p. 3727–3748, 2022. Disponível em: <<https://doi.org/10.1007/s11063-022-10783-z>>.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. San Diego, CA, USA: [s.n.], 2015. Disponível em: <<https://doi.org/10.48550/arXiv.1412.6980>>.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems 25 (NIPS 2012)*. Nevada, EUA: [s.n.], 2012. Disponível em: <https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, 1998. Disponível em: <<https://doi.org/10.1109/5.726791>>.

LIN, C.-C.; YANG, C.-M. Heartbeat classification using normalized RR intervals and morphological features. *Mathematical Problems in Engineering*, v. 2014, 2014. Disponível em: <<https://doi.org/10.1155/2014/712474>>.

LLAMEDO, M.; MARTÍNEZ, J. P. Heartbeat classification using feature selection driven by database generalization criteria. *IEEE Transactions on Biomedical Engineering*, v. 58, n. 3, 2011. Disponível em: <<https://doi.org/10.1109/TBME.2010.2068048>>.

LUZ, E. J. 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>>.

MAR, T. et al. Optimization of ECG classification by means of feature selection. *IEEE Transactions on Biomedical Engineering*, v. 58, n. 8, 2011. Disponível em: <<https://doi.org/10.1109/tbme.2011.2113395>>.

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

NIELSEN, A. *Practical time series analysis: Prediction with statistics and machine learning*. Sebastopol, CA, USA: O'Reilly Media, 2019.

OLIVEIRA, R. F. et al. Explorando redes neurais de grafos para classificação de arritmias. In: *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. Teresina, Brasil: Sociedade Brasileira de Computação (SBC), 2022. Disponível em: <<https://doi.org/10.5753/sbcas.2022.222510>>.

PRIYAMVADA; WADHVANI, R. Review on various models for time series forecasting. In: *2017 International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2017. Disponível em: <<https://doi.org/10.1109/ICICI.2017.8365383>>.

QUAN, S. et al. Time series classification based on multi-dimensional feature fusion. *IEEE Access*, v. 11, p. 11066–11077, 2023. Disponível em: <<https://doi.org/10.1109/ACCESS.2023.3241013>>.

SEZER, O. B.; OZBAYOGLU, A. M. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing*, v. 70, p. 525–538, 2018. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494618302151>>.

SHTEREV, V. A.; METCHKARSKI, N. S.; KOPARANOV, K. A. Time series prediction with neural networks: a review. In: *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*. Ohrid, North Macedonia: [s.n.], 2022. p. 1–4.

SHUMWAY, R. H.; STOFFER, D. S. *Time series analysis and its applications: With R examples*. 3. ed. Nova Iorque, NY, USA: Springer, 2010.

SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. Disponível em: <<https://doi.org/10.48550/arXiv.1409.1556>>.

- SINGH, A. K.; KRISHNAN, S. ECG signal feature extraction trends in methods and applications. *BioMedical Engineering OnLine*, v. 22, n. 22, 2023. Disponível em: <<https://doi.org/10.1186/s12938-023-01075-1>>.
- SORIA, M. L.; MARTÍNEZ, J. P. Analysis of multidomain features for ECG classification. In: *36th Annual Computers in Cardiology Conference (CinC)*. Park City, UT, USA: [s.n.], 2009. Disponível em: <<https://ieeexplore.ieee.org/document/5445344>>.
- SREENIVAS, K. V.; GANESAN, M.; LAVANYA, R. Classification of arrhythmia in time series ECG signals using image encoding and convolutional neural networks. In: *2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*. Chennai, India: [s.n.], 2021. p. 1–6. Disponível em: <<https://doi.org/10.1109/ICBSII51839.2021.9445177>>.
- SRIVASTAVA, A. et al. A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG. *Physiological Measurement*, v. 43, n. 6, 2022. Disponível em: <<https://doi.org/10.1088/1361-6579/ac6f40>>.
- TANG, C. 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*, v. 35, n. 2, 2023. Disponível em: <<https://doi.org/10.1016/j.jksuci.2023.01.002>>.
- TSINGANOS, P. et al. Hilbert sEMG data scanning for hand gesture recognition based on deep learning. *Neural Computing and Applications*, v. 33, p. 2645–2666, 2021. Disponível em: <<https://doi.org/10.1007/s00521-020-05128-7>>.
- WAN, D. et al. Mixed local channel attention for object detection. *Engineering Applications of Artificial Intelligence*, v. 123, p. Part C, 2023. Disponível em: <<https://doi.org/10.1016/j.engappai.2023.106442>>.
- WANG, T. et al. Automatic ECG classification using continuous wavelet transform and convolutional neural network. *Entropy*, v. 23, n. 1, 2021. Disponível em: <<https://doi.org/10.3390/e23010119>>.
- WANG, Z.; OATES, T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In: *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*. [s.n.], 2015. Disponível em: <<https://cdn.aaai.org/ocs/ws/ws0115/10179-46015-1-PB.pdf>>.
- XIA, Y.; XIONG, Y.; WANG, K. A transformer model blended with CNN and denoising autoencoder for inter-patient ECG arrhythmia classification. *Biomedical Signal Processing and Control*, v. 86, 2023. Disponível em: <<https://doi.org/10.1016/j.bspc.2023.105271>>.
- XU, P. et al. Interpatient ECG arrhythmia detection by residual attention CNN. *Computational and Mathematical Methods in Medicine*, v. 2022, 2022. Disponível em: <<https://doi.org/10.1155/2022/2323625>>.
- XU, X.; JEONG, S.; LI, J. Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM. *IEEE Access*, v. 8, 2020. Disponível em: <<https://doi.org/10.1109/ACCESS.2020.3006707>>.
- ZAHID, M. U.; KIRANYAZ, S.; GABBOUJ, M. Global ECG classification by self-operational neural networks with feature injection. *IEEE Transactions on Biomedical Engineering*, v. 70, n. 1, 2023. Disponível em: <<https://doi.org/10.1109/TBME.2022.3187874>>.

ZHANG, Z. et al. Heartbeat classification using disease-specific feature selection. *Computers in Biology and Medicine*, v. 46, 2014. Disponível em: <<https://doi.org/10.1016/j.compbiomed.2013.11.019>>.

ZHOU, F.; FANG, D. Multimodal ECG heartbeat classification method based on a convolutional neural network embedded with FCA. *Scientific Reports*, v. 14, 2024. Disponível em: <<https://doi.org/10.1038/s41598-024-59311-0>>.