

Abordagem de Atenção Local em Rede Neural Convolutacional para Classificação de Arritmias Cardíacas

Daniel Dantas do Amaral Ramos, Adriana Rosa Garcez Castro, Ítalo Flexa Di Paolo
Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Pará, Belém, Brasil
Centro de Ciências Naturais e Tecnologia, Universidade do Estado do Pará, Ananindeua, Brasil
daniel.dantas.a.r@gmail.com, adcastro@ufpa.br, itflexa@uepa.br

Resumo—Este trabalho propõe uma arquitetura baseada em Rede Neural Convolutacional (CNN) com mecanismo de atenção local para a classificação automática de arritmias cardíacas, utilizando sinais de ECG transformados em imagens por meio da técnica Gramian Angular Field (GAF). O estudo investiga como diferentes funções de ativação (tangente hiperbólica e sigmoide) e tipos de kernel (fixo e adaptativo) podem impactar o desempenho do módulo de atenção da arquitetura proposta. Os experimentos foram conduzidos com a base de dados pública MIT-BIH para abordagem interpaciente, considerando as classes N (Normal), S (Supraventricular) e V (Ventricular). Após os diversos experimentos, a melhor configuração encontrada para o módulo de atenção, com tangente hiperbólica e kernel adaptativo, obteve F1-score de 0,7678, superando os demais cenários testados. Os resultados confirmam a relevância das escolhas arquiteturais internas do mecanismo de atenção para o aprimoramento de classificadores de arritmia baseados em aprendizado profundo.

Palavras chave: Eletrocardiograma, Função de ativação, kernel, Rede Neural Convolutacional, Mecanismo de Atenção, Arritmia Cardíaca.

I. INTRODUÇÃO

O eletrocardiograma (ECG) é uma técnica não invasiva amplamente utilizada para monitorar a atividade elétrica do coração e identificar anomalias como as arritmias cardíacas, que se caracterizam por alterações no ritmo ou frequência dos batimentos cardíacos [1]. Embora eficaz, a análise visual de longos registros de ECG é trabalhosa, propensa a erros e exige alta especialização [1].

Nesse cenário, sistemas automáticos de classificação de arritmias cardíacas surgem como alternativas promissoras para apoio ao diagnóstico médico, melhorando a eficiência e reduzindo a subjetividade. Dentre as abordagens de aprendizado de máquina, as Redes Neurais Convolutacionais (CNNs) têm se destacado, sobretudo quando aplicadas sobre representações visuais dos sinais de ECG, obtidas por transformações que convertem os dados temporais em imagens. Trabalhos como o apresentado em [2] demonstram a eficácia dessa abordagem, onde os autores transformaram o sinal ECG em uma imagem de espectrograma frequência-tempo usando a Transformada de Fourier de Tempo Curto (STFT) e alimentaram uma CNN para classificação de arritmias. Outros exemplos incluem [3], que desenvolveram uma CNN 2D para classificação de imagens de ECG transformadas via Gramian Angular Field (GAF), e [4],

onde foi proposta uma CNN 2D com características Wavelet, após conversão do sinal ECG em imagem bidimensional. Além disso, em [5], os autores apresentaram uma abordagem híbrida que combina três diferentes técnicas de conversão de sinais unidimensionais em imagens bidimensionais.

Técnicas híbridas que combinam CNNs e mecanismos de atenção têm mostrado avanços significativos em diversos problemas de classificação. Os mecanismos de atenção permitem que modelos atribuam dinamicamente pesos às partes mais relevantes dos dados, processo inspirado na capacidade humana de focar seletivamente em aspectos específicos enquanto ignora outros. Essa capacidade permite, em particular, um aprimoramento substancial na capacidade discriminativa do modelo, tornando-o especialmente crucial em cenários com dados desbalanceados, além de contribuir para mitigar o problema de overfitting, ao permitir que a rede foque seletivamente em atributos mais informativos e, consequentemente, suprima informações irrelevantes ou ruidosas, atenuando o impacto de características menos importantes ou redundantes, ajudando a melhorar a eficiência da classificação. [6]

CNNs têm sido aprimoradas com esses mecanismos para melhorar a precisão em tarefas de classificação de arritmias, como apresentado por [7], que utilizaram uma rede recorrente bidirecional com atenção hierárquica e convoluções dilatadas, e por [8], que propuseram atenção por canal de frequência para ajuste dinâmico dos mapas de ativação. Além disso, [1] propuseram uma CNN multimodal com atenção local (Self-Attention), aplicada a imagens geradas dos sinais ECG utilizando transformações como Hilbert Space Filling Curve (HSFC) e Recurrence Plot (RP).

No contexto de classificação automática de arritmias cardíacas, destacam-se duas abordagens principais: a intrapaciente, onde os conjuntos de treinamento e teste podem conter dados do mesmo paciente, sendo útil para analisar variações individuais, e a interpaciente, onde os dados de treino e teste são estritamente separados por pacientes diferentes, avaliando a generalização do modelo para novos indivíduos [1]. A abordagem interpaciente é especialmente relevante em aplicações clínicas reais, e diversos estudos evidenciam sua importância e desafios, como [9], que sugerem uma divisão padrão da base MIT-BIH para avaliação interpaciente, e [10], que aplicaram CNN associada à transformada Wavelet nesse paradigma.

Dentro deste contexto, este trabalho propõe um classificador baseado em CNN com mecanismo de atenção local para classificação de arritmias cardíacas, utilizando imagens obtidas pela técnica Gramian Angular Field (GAF) a partir dos sinais ECG, considerando a abordagem interpaciente. O principal objetivo é avaliar o impacto das funções de ativação (tangente hiperbólica e sigmoide) e dos tipos de kernel (fixo e adaptativo) no módulo de atenção, visando aprimorar o desempenho do classificador proposto. Os experimentos foram conduzidos na base pública MIT-BIH, contemplando as classes N (Normal), S (Supraventricular) e V (Ventricular).

II. ARRITMIAS CARDÍACAS

A. Sinal Cardíaco ECG

O coração é um órgão muscular essencial para a circulação sanguínea, atuando no transporte de oxigênio e nutrientes por meio de contrações coordenadas. Estruturalmente, o coração é composto por quatro câmaras: dois átrios e dois ventrículos. A geração rítmica dos batimentos cardíacos é controlada por seu sistema elétrico interno [11].

Um sinal de batimento cardíaco normal, como mostrado na Figura 1, consiste na onda P, no complexo QRS e na onda T, que refletem as atividades elétricas de repolarização e despolarização do átrio e ventrículo [8].

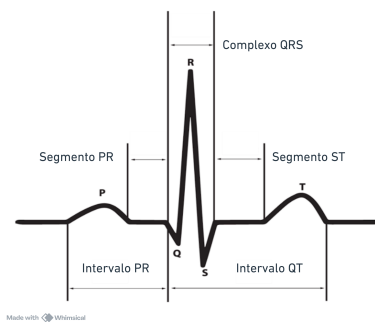


Figura 1. Sinal de ECG indicando as ondas do complexo QRS. Fonte: [12]

A atividade elétrica do coração pode ser monitorada por meio do eletrocardiograma (ECG), exame não invasivo que registra variações de potencial elétrico através de eletrodos fixados na superfície corporal. Os eletrodos são posicionados em diferentes regiões, como tórax, braços e pernas, formando derivações que oferecem visões complementares da atividade cardíaca [8].

A derivação MLII, obtida com eletrodos no braço direito e na perna esquerda, é a mais comumente utilizada na prática clínica e em estudos para a classificação de arritmias devido à sua clareza na representação da atividade elétrica cardíaca [1].

B. Definição de Arritmias Cardíacas

Arritmias cardíacas correspondem a alterações no ritmo, frequência ou sequência dos batimentos do coração, geralmente causadas por falhas na geração ou condução dos impulsos elétricos [11].

O funcionamento elétrico do coração é responsável por sincronizar a despolarização e repolarização das câmaras cardíacas, permitindo a contração eficaz do músculo. Alterações nesse circuito elétrico podem gerar batimentos acelerados (taquicardia), lentos (bradicardia) ou irregulares, caracterizando os diferentes tipos de arritmias.

III. SISTEMAS AUTOMÁTICOS DE CLASSIFICAÇÃO DE ARRITMIAS

A interpretação manual de sinais de ECG para detectar arritmias demanda um alto nível de especialização e atenção, especialmente diante do grande volume de dados gerados em exames prolongados. Nesse cenário, sistemas automáticos têm-se mostrado eficazes, oferecendo vantagens como [6]:

- **Eficiência:** aceleram a análise de grandes volumes de dados;
- **Padronização:** reduzem a variabilidade entre observadores;
- **Deteção precoce:** identificam padrões sutis que podem ser imperceptíveis à análise visual;
- **Redução de erros:** minimizam falhas causadas por cansaço ou subjetividade;
- **Apoio clínico:** fornecem subsídios objetivos para a tomada de decisão médica.

Segundo [6], o fluxo típico para a classificação automática de arritmias cardíacas envolve as seguintes etapas:

- **Pré-processamento:** remoção de artefatos e ruídos, melhorando a qualidade do sinal;
- **Segmentação:** identificação e isolamento de complexos de ECG (como o complexo QRS) e outras ondas;
- **Extração de características:** identificação e quantificação de características relevantes do sinal de ECG;
- **Classificação:** aplicação de um modelo para identificar e classificar os diferentes tipos de arritmias cardíacas.

A. Padrão AAMI EC57

A Associação para o Avanço da Instrumentação Médica (AAMI, *Association for the Advancement of Medical Instrumentation*) estabelece diretrizes e padrões para a classificação automática de sinais ECG. As Diretrizes da AAMI para a Classificação de Eletrocardiogramas (ANSI/AAMI EC57) fornecem orientações sobre terminologia, critérios de desempenho e métricas de avaliação para sistemas de classificação automática de arritmias cardíacas.

As diretrizes da AAMI definem cinco classes principais de arritmias:

- **N (Normal):** Batimentos cardíacos normais, sem nenhuma anormalidade significativa. Esses batimentos podem incluir o complexo QRS normal, ondas P e T normais e intervalos PR e QT dentro dos limites normais.
- **S (Supraventricular):** Arritmias que se originam acima dos ventrículos, como taquicardia sinusal, taquicardia atrial, fibrilação atrial, vibração atrial e bloqueio atrioventricular.

- **V (Ventricular):** Arritmias que se originam nos ventrículos, como taquicardia ventricular, fibrilação ventricular e complexos de prematuros ventriculares.
- **F (Fusão de Normal e Ventricular):** Arritmias que têm componentes tanto do complexo QRS normal quanto do complexo QRS ventricular ectópico (anormal).
- **Q (Desconhecido):** Batimento não classificado ou arritmia de ritmo.

A utilização dessas diretrizes promove a padronização dos estudos na área, permitindo comparações entre diferentes métodos e maior confiabilidade na avaliação clínica. A norma também enfatiza a importância do uso de bases de dados representativas e bem anotadas no desenvolvimento dos classificadores automáticos [6].

B. Bases de Dados

A AAMI recomenda o uso de cinco bases de dados para o desenvolvimento e avaliação de algoritmos de classificação de arritmias: AHA, CU, MIT-BIH, NSTDB e PTB. No contexto deste trabalho, a base MIT-BIH (The Massachusetts Institute of Technology - Beth Israel Hospital Arrhythmia Database) foi utilizada, sendo uma das bases de dados mais utilizada na literatura para o desenvolvimento de sistemas classificadores de arritmias. A base contém 48 registros de ECG de 47 pacientes, digitalizados e anotados por especialistas em ECG. Cada registro possui aproximadamente 30 minutos de duração, com duas derivações (MLII e uma precordial) e anotações detalhadas dos complexos QRS e tipos de arritmia. Os registros são categorizados em diversas classes de arritmias.

C. Métricas de Avaliação

A norma ANSI/AAMI EC57:2012 R2020 especifica as métricas de avaliação para sistemas de classificação de arritmias, incluindo Acurácia, Precisão, Sensibilidade, Especificidade e F1-Score.

A acurácia 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. A precisão indica quantas classes de arritmia ou normalidade detectadas pelo modelo de fato estavam corretas. A sensibilidade indica o quanto o modelo é capaz de detectar os casos verdadeiros de arritmia. A especificidade mede o quanto o modelo acerta ao identificar ausência de arritmia. O F1-Score representa o equilíbrio harmônico entre precisão e sensibilidade, sendo especialmente útil em cenários de classes desbalanceadas, pois penaliza bastante valores baixos de qualquer um destes indicadores.

IV. REDES NEURAIS CONVOLUCIONAIS E MECANISMOS DE ATENÇÃO

A. Redes Neurais Convolucionais (CNNs)

Uma Rede Neural Convolucional (CNN – *Convolutional Neural Network*) é uma arquitetura de rede neural artificial, variante do Perceptron de múltiplas camadas (MLP – *Multi Layer Perceptron*) proposto por Rumelhart, Hinton e Williams (1986), sendo comumente usada em aplicações de detecção, classificação e reconhecimento em imagens e vídeos [6].

É uma arquitetura de aprendizado profundo particularmente eficaz no processamento de dados com estrutura em grade, como imagens. Uma CNN típica é composta por diversas camadas, incluindo, entre outras:

- **Camadas Convolucionais:** Aplicam filtros para extrair características locais do dado de entrada.
- **Camadas de Pooling:** Reduzem a dimensionalidade dos mapas de características, tornando a rede mais robusta a pequenas variações na entrada.
- **Funções de Ativação:** Introduzem não-linearidade na rede, permitindo o aprendizado de relações complexas (ex: ReLU e sigmoide).
- **Camadas Totalmente Conectadas (Fully Connected):** Realizam a classificação final com base nas características extraídas, sendo compostas geralmente por camadas de neurônios do Perceptron.

CNNs são classificadores robustos para imagens, dados bidimensionais que guardam padrões espaciais. No contexto da classificação de arritmias, sinais ECG unidimensionais (temporais) podem ser convertidos em imagens para serem processados por CNNs.

B. Conversão de sinais unidimensionais em imagens

A conversão de sinais temporais em imagens é uma técnica que consiste em transformar um sinal temporal em uma imagem, para que possa ser processada por uma CNN. A ideia é tratar cada contorno do sinal temporal como um conjunto de pixels de uma imagem, e assim utilizar a capacidade de processamento de imagens das CNNs para realizar a classificação ou previsão do sinal temporal [6].

Neste estudo, duas técnicas de conversão de sinais temporais em imagens foram estudadas e implementadas para uso na fase de experimentos: Campo Angular Gramiano (GAF) e Curva de Preenchimento de Espaço de Hilbert (HSFC). A seguir será apresentada somente a técnica GAF, que foi a utilizada na estrutura final, pois apresentou melhores resultados nos testes iniciais.

1) *Gramian Angular Field:* O *Gramian Angular Field* (GAF) é uma técnica amplamente utilizada para transformação de séries temporais em imagens, codificando simultaneamente informações de tempo e magnitude em uma representação visual. Conforme proposto por [13], essa abordagem permite preservar a estrutura temporal dos dados em uma matriz, mantendo sua interpretabilidade visual. A Figura 2 ilustra o processo.

O método parte da normalização da série temporal $X = \{x_1, x_2, \dots, x_n\}$, seguida da conversão para coordenadas polares, onde a amplitude é representada como o cosseno de um ângulo e o tempo como raio. Em seguida, é aplicada a função arco-cosseno, gerando uma matriz chamada Matriz de Gram, cuja composição angular traduz o relacionamento entre os pontos da série. Essa matriz é, então, convertida em uma imagem que carrega padrões relevantes para tarefas de classificação [13].

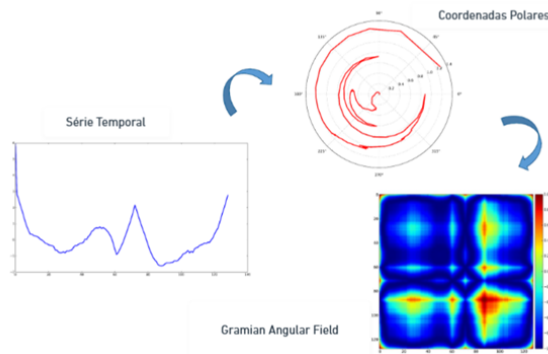


Figura 2. Processo de conversão de uma série temporal para imagem utilizando Gramian Angular Field (GAF). Adaptado de [13].

C. Mecanismos de Atenção

Mecanismos de atenção, originalmente propostos no contexto do Processamento de Linguagem Natural, operam com base no princípio de que modelos devem atribuir maior peso e focar seletivamente nas partes mais relevantes da entrada para a tarefa em questão. Esses mecanismos podem ser amplamente categorizados como globais ou locais. São ditos globais quando aplicados nas camadas mais próximas da entrada ou dentro de camadas convolucionais, permitindo capturar dependências de longo alcance. Trabalhos como o de Vaswani et al. [14], que introduziram a arquitetura Transformer com mecanismos de autoatenção global, e Chen et al. [15], que exploram estruturas de autoatenção clássica em camadas convolucionais, exemplificam essa abordagem.

Por outro lado, mecanismos de atenção podem ser locais, sendo mais comuns em camadas próximas à saída, como nas camadas totalmente conectadas (FC), focando em características específicas aprendidas em estágios anteriores, como no caso de Ahmad e Khan [16], Di Paolo [6] e Ahmad et al. [1].

Existem diferentes tipos de mecanismos de atenção, entre eles:

- **Self-Attention (Autoatenção):** Este mecanismo permite que o modelo aprenda dependências e relacionamentos dentro da própria sequência de entrada. Ele identifica e pondera a importância de diferentes partes da sequência ao atender a si mesmo. O objetivo é capturar dependências de longo alcance entre diferentes posições de um mapa de características, considerando a interação entre todas as localizações. Existem dois paradigmas principais de Self-Attention:
 - **Self-Attention Clássico (Global):** Baseia-se nos conceitos de Query, Key e Value, originários de diferentes projeções da mesma entrada, para calcular pesos de atenção que indicam a similaridade entre cada elemento e todos os outros. É útil para capturar interações longas e globais. Exemplo desta abordagem pode ser visto em [17].
 - **Self-Attention Local:** Utiliza convoluções para gerar mapas de atenção também achatados e é mais eficiente em

termos de memória e computação, sendo geralmente integrado em CNNs na camada totalmente conectada. Exemplos desta abordagem podem ser vistos em [1], [6], [16].

- **Spatial Attention (Atenção Espacial):** Este tipo de atenção foca em identificar regiões importantes no espaço, destacando onde a atenção deve ser aplicada. Ele utiliza informações espaciais, como padrões em pixels adjacentes, para gerar um mapa de atenção. Exemplos dessa abordagem podem ser vistos em [18], [19].
- **Channel Attention (Atenção de Canal):** Concentra-se em ajustar a importância dos canais no mapa de características, permitindo que a rede destaque canais mais relevantes. Geralmente, utiliza uma operação global, como global average pooling, para extrair informações de todo um canal e aplicar pesos aprendidos. Exemplos dessa abordagem estão em [18], [20].
- **Mixed Attention (Spatial and Channel) ou Atenção Mista (Espacial e em Canal):** Combina Spatial Attention e Channel Attention para explorar relações espaciais e de canal simultaneamente. Pode aplicar atenção por canal e atenção espacial em sequência ou paralelamente para refinar os mapas de características. Exemplos dessa abordagem podem ser vistos em [21]–[23].

Na estrutura proposta neste trabalho, um mecanismo de atenção baseado em *self-attention local* é integrado ao componente classificador, conforme explorado em [6]. Essa abordagem é particularmente eficaz para dados vetorizados, como os obtidos após o achatamento das saídas convolucionais.

O mecanismo local aplica uma operação de convolução sobre o vetor de características global para gerar um mapa de atenção achatado, seguido por uma função de ativação (sigmoide ou tangente hiperbólica). Esse mapa é utilizado para ponderar os elementos do vetor de entrada por multiplicação elemento a elemento, destacando as características mais relevantes para a tarefa de classificação [6].

O kernel de convolução pode ser fixo ou adaptativo. No caso adaptativo, seus parâmetros são aprendidos durante o treinamento via retropropagação, permitindo que a própria rede defina os padrões de atenção mais adequados. Essa estrutura oferece vantagens computacionais por operar em vetores compactados, sendo especialmente apropriada para integração em camadas totalmente conectadas de redes CNN [6].

V. METODOLOGIA

A estrutura proposta neste trabalho para a classificação de arritmias cardíacas é baseada em uma CNN AlexNet com Mecanismo de Atenção Local (*Self-Attention*). A arquitetura foi projetada com o objetivo de avaliar o impacto de diferentes funções de ativação e da variação entre kernel fixo e adaptativo no módulo de atenção proposto. As Figuras 3 e 4 ilustram os principais componentes da estrutura proposta, sendo eles:

- 1) **Entrada:** Segmentos de sinais de ECG, que passam por um filtro baseado em [24], seguindo o apresentado em [6],

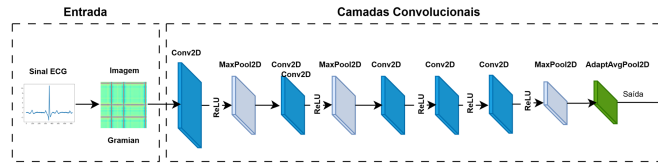


Figura 3. Camada convolutiva. Autor. Adaptado de [6].

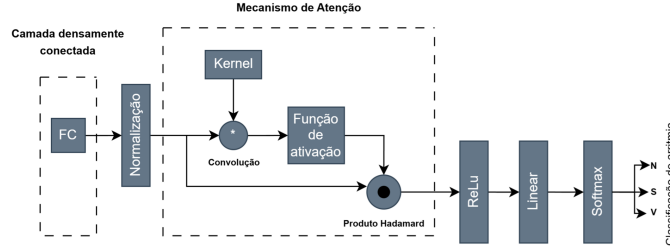


Figura 4. Camada classificadora com Mecanismo de Atenção. Autor. Adaptado de [6].

para realizar um processo de limpeza. Após isso, os sinais são convertidos em imagens bidimensionais utilizando a técnica GAF ou HSFC.

- 2) **Camadas convolucionais:** Este componente recebe a imagem de entrada e captura características espaciais, sendo baseado na arquitetura AlexNet discutida em [6].
- 3) **Camada classificadora com Mecanismo de Atenção:** O vetor extraído das convoluções é passado por uma camada densamente conectada (FC), composta por: *Dropout*, *Linear*, *ReLU*, *Dropout* e *Linear*. Em seguida, o vetor de características é normalizado no intervalo $[-1, 1]$, e é utilizado como entrada para o módulo de atenção, conforme proposto em [6].

O módulo de atenção aplica uma convolução 1D com kernel fixo ou adaptativo, seguida por uma função de ativação (sigmoide ou tangente hiperbólica), gerando um vetor de pesos. Esse vetor é então multiplicado elemento a elemento com o vetor original de características, produzindo uma representação refinada. Por fim, a saída do módulo de atenção é processada por uma sequência (*ReLU*, *Linear* e *Softmax*), resultando nas probabilidades de classificação para as classes N, S e V.

- Apenas a derivação **MLII** foi considerada para a classificação. Como destacado por [6], esta derivação é onipresente em todos os registros da base MIT-BIH. Esta derivação é amplamente utilizada para medições em longos períodos e é menos sensível ao movimento dos pacientes, além de gerar sinais nos quais se pode destacar com mais clareza o complexo QRS e as ondas T e P em relação às demais derivações.
- Para fins deste trabalho, não foi usado nenhum tipo de balanceamento de dados.

A Tabela I apresenta a distribuição quantitativa dos segmentos por classe para os conjuntos de treino e teste, conforme adotado neste trabalho.

Tabela I
DISTRIBUIÇÃO POR CLASSES NA BASE DS1 E DS2, COM ABORDAGEM INTERPACIENTE

Classe	DS1 (Treino)	DS2 (Teste)
N (Normal)	45.781	43.598
S (Supraventricular)	975	667
V (Ventricular)	3.786	3.219
Total	50.542	47.484

A. Bancos de Dados de Sinais ECG

Neste trabalho, a divisão da base MIT-BIH em conjuntos de treino e teste seguiu o paradigma interpaciente, conforme proposto por [25], com ajustes descritos em estudos subsequentes [6]. Os registros foram separados em dois subconjuntos: DS1 (treinamento) e DS2 (teste). De maneira específica:

- **MIT-BIH:** Os sinais de ECG foram segmentados em intervalos de dois segundos. Assim como em [6], apenas as classes N, S e V foram consideradas para classificação, visto que, como apontado pelo autor e pela literatura, a classe F é considerada difícil de classificar devido à natureza complexa dos complexos QRS para esta classe.

VI. EXPERIMENTOS

Os experimentos foram realizados em ambiente computacional local, com processador AMD Ryzen 5 5500U (6 núcleos, 12 threads, 2,1GHz), 8 GB de RAM, SSD de 250 GB, sistema operacional Windows 11 Pro, e execução via Anaconda com kernel do Jupyter Notebook, com Visual Studio Code. A estrutura proposta foi implementada em Python 3.10.15, com uso da biblioteca PyTorch v2.6.0 e backend CPU. Para o treinamento dos modelos, utilizou-se o otimizador Adam, com taxa de aprendizado inicial de $4e^{-6}$, tamanho de batch de 128 amostras, 200 épocas de treinamento e estratégia de decaimento da taxa de aprendizado de 50% a cada 10 épocas sem melhora na função custo.

Para a avaliação final, foram realizados experimentos exploratórios para definir a arquitetura ideal da rede, assim como a melhor técnica de transformação de sinais de ECG em imagens, entre GAF e HSFC. Estratégias de aumento de dados e balanceamento de classes 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 (tangente hiperbólica ou sigmoide) e o tipo de kernel (fixo ou adaptativo) no módulo de atenção. Entende-se por kernel fixo o valor constante que não muda durante o aprendizado da rede, sendo igual a $[-1, 9, -1]$, um vetor com valores seguindo o trabalho de [6], [1] e [16]. Já o kernel adaptativo é entendido como mais um parâmetro da rede que deve ser ajustado durante o treinamento.

A. Resultados

A Tabela II apresenta os resultados quantitativos globais obtidos 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 módulo 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 II
PRINCIPAIS INDICADORES DE DESEMPENHO GLOBAL PARA OS CENÁRIOS TESTADOS

Transformação	Ativação	Kernel	Acurácia	Precisão	Recall	Especificidade	F1-score
GAF	Tanh	Adaptativo	0,9788	0,7745	0,7614	0,9539	0,7678
GAF	Tanh	Fixo	0,9793	0,7735	0,7385	0,9501	0,7544
GAF	Sigmoide	Adaptativo	0,9744	0,7676	0,7329	0,9495	0,7442
GAF	Sigmoide	Fixo	0,9733	0,7565	0,7473	0,9391	0,7512
HSFC	Tanh	Adaptativo	0,9740	0,7338	0,7396	0,9486	0,7366
HSFC	Tanh	Fixo	0,9758	0,7373	0,7006	0,9399	0,7121
HSFC	Sigmoide	Adaptativo	0,9690	0,6956	0,7065	0,9410	0,7004
HSFC	Sigmoide	Fixo	0,9704	0,7096	0,7031	0,9382	0,7040

Para fins de comparação, uma rede CNN AlexNet, sem nenhum tipo de mecanismo de atenção, foi treinada utilizando o mesmo conjunto de dados, tendo obtido os resultados de acurácia (0,9751), precisão (0,6893), recall (0,6469), especificidade (0,9246) e F1-score (0,6615). Observa-se que esses resultados são inferiores aos obtidos em todas as estruturas apresentadas na Tabela 3, com inclusão do mecanismo de atenção local, o que destaca a importância da capacidade do modelo proposto com mecanismo de atenção em focar seletivamente nas regiões mais relevantes das representações imagéticas. A comparação reforça que a atenção local, ao ponderar dinamicamente as características extraídas, contribui de forma decisiva para o aumento da sensibilidade e do F1-score, métricas cruciais em contextos clínicos com classes desbalanceadas, como o de arritmias cardíacas.

Adicionalmente, ao comparar os dois métodos de conversão de sinais temporais em imagens, o GAF e o HSFC, observou-se desempenho superior consistente nos cenários com GAF, onde todas as configurações apresentaram maiores valores de F1-score quando comparadas às respectivas variações com HSFC, indicando que essa representação foi mais eficaz na

preservação de padrões discriminativos relevantes para a tarefa de classificação. Com isso, como observado na Tabela II, o melhor resultado se deu com F1-score de 0,7678 e foi obtido na configuração: imagens de GAF, função tangente hiperbólica e kernel adaptativo. Para fins de comparação deste melhor resultado com o modelo base AlexNet, a Tabela III apresenta os indicadores de desempenho obtidos por classe.

Portanto, na comparação por classe entre a AlexNet base e a melhor estrutura proposta, observa-se que embora a AlexNet comece apresentando excelente detecção de batimentos normais (F1-score = 98,78%) e ventriculares (F1-score = 92,88%), ela praticamente não reconhece batimentos supraventriculares (F1-score = 6,80%), refletindo o viés do modelo para a classe majoritária. Ao incorporar o módulo de atenção adaptativo com função tangente hiperbólica e a transformação da imagem GAF, nosso modelo eleva a acurácia global de 97,51% para 97,88%, mas, mais importante, melhora significativamente as métricas de S (supraventriculares), passando de 6,8% para 36% no F1-score, de 12,7% para 37% na precisão e de 4,6% para 35% na sensibilidade. Além disso, alcançamos F1-score de 99% em N e 95% em V. Esses ganhos mostram que o módulo de atenção consegue realçar as características mais relevantes de cada batimento, tornando o classificador muito mais equilibrado e capaz de identificar os ritmos supraventriculares e ventriculares sem perder desempenho na detecção de batimentos normais.

VII. RESULTADOS E DISCUSSÕES

A. Análise Geral dos Resultados

Os resultados obtidos demonstraram que todas as variações de arquitetura dos experimentos realizados alcançaram desempenho elevado, com acurácia superior a 96%. Entretanto, a métrica de F1-score é a mais relevante nesse contexto, uma vez que a base MIT-BIH é desbalanceada e o interesse clínico está na detecção eficaz das classes minoritárias (S e V). O melhor desempenho geral foi alcançado pela configuração com tangente hiperbólica e kernel adaptativo, que obteve o maior F1-Score (0,7678), além de valores elevados de precisão (0,7745) e recall (0,7614). Esse resultado corrobora a hipótese de que a combinação entre a expressividade da função tangente hiperbólica e a flexibilidade do kernel adaptativo potencializa a capacidade discriminativa do mecanismo de atenção, conforme defendido na tese de [6].

A função tangente hiperbólica, ao permitir pesos de atenção em um intervalo simétrico $[-1, 1]$, oferece maior nuance na ponderação das características extraídas. Isso possibilita destacar atributos relevantes. Já o kernel adaptativo, por ser treinável, ajusta-se dinamicamente aos padrões dos dados, permitindo que o mapa de atenção seja aprendido ao longo do processo de treinamento. Essa flexibilidade foi essencial para alcançar o equilíbrio entre sensibilidade e precisão observado no melhor cenário.

Embora a configuração tangente hiperbólica e kernel fixo também tenha apresentado desempenho competitivo (F1-score = 0,7544), ela foi superada nos principais indicadores pela versão adaptativa, sugerindo que a capacidade de aprendizado

Tabela III
INDICADORES DE DESEMPENHO POR CLASSE

Modelo	Acurácia (%)	F1-score (%)			Precisão (%)			Sensibilidade (%)		
		N	S	V	N	S	V	N	S	V
AlexNet base	97,51	98,78	6,80	92,88	98,08	12,70	96,02	99,49	4,65	89,93
Melhor estrutura AlexNet com módulo de atenção	97,88	99	36	95	98,87	37,24	96,26	99,11	34,93	94,37

do kernel traz ganhos adicionais, mesmo quando já se utiliza uma função de ativação que permite mais valores possíveis dentro do seu intervalo.

As configurações com sigmoide, apesar de robustas, mostraram desempenho inferior nos cenários testados, especialmente no F1-score. Com isso observa-se que a limitação no intervalo [0,1] imposta pela sigmoide numa abordagem convencional é superada por uma maior generalização aplicada na função tangente hiperbólica, em especial após a etapa de normalização que foi aplicada neste estudo, conforme observado na Figura 4.

B. Comparação com outras estruturas

A Tabela IV apresenta, para comparação com nosso melhor resultado, métricas obtidas para estruturas propostas em trabalhos correlatos que adotaram a mesma divisão DS1 e DS2 utilizada em nosso trabalho.

A Tabela IV mostra que o modelo proposto neste trabalho apresenta desempenho competitivo frente a diferentes abordagens da literatura. Com F1-score de 76,78%, nossa estrutura supera os resultados de Dias et al. [9] e Wang et al. [10]. No caso de Dias et al. [9], os autores apresentaram 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 [26], foi proposto um modelo multimodal baseado em aprendizado por múltiplas instâncias com fusão de características via mecanismo de atenção, combinando sinais ECG e imagens GAF, alcançando F1-score de 69,10% na abordagem interpaciente. Esse desempenho superior da abordagem proposta neste trabalho 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.

Nosso trabalho teve desempenho similar ao apresentado em Oliveira et al. [27], 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 aos resultados obtidos com o trabalho de [28], que obteve F1-score de 0,8191 (para interpaciente), ao utilizar uma arquitetura mais complexa de redes neurais convolucionais multimodais, combinando as técnicas de transformação de imagens Hilbert Space Filling Curve (HSFC) e Recurrence Plot (RP), além de empregar aumento de dados com SMOTE e WGAN-GP. A diferença de desempenho entre os dois modelos 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 derivadas combinadas;

Apesar disso, 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 com arquiteturas mais complexas, mostrando sua aplicabilidade em cenários clínicos reais, mesmo sem recursos computacionais avançados.

VIII. CONCLUSÕES

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) 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;
- 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.

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. Além disso, a utilização de diferentes técnicas de transformação pode viabilizar um modelo mais generalizável. Tais aprimoramentos

Tabela IV
INDICADORES DE DESEMPENHO DE TRABALHOS CORRELATOS

Referência	Abordagem chave	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)	F1-score (%)
Dias et al. (2021)	Segmentação baseada no intervalo RR e classificação com algoritmo discriminante linear; uso de ECG unidimensional	80,58	59,37	84,84	94,35	60,77
Wang et al. (2021)	Modelo multimodal com sinais ECG e imagens GAF, combinando transformada Wavelet, CNN e mecanismo de atenção via aprendizado por múltiplas instâncias	97,48	70,75	67,47	96,01	68,76
Oliveira et al. (2022)	Cada batimento é representado como grafo com o algoritmo Visibility Graph; classificação feita com GraphSAGE	86,12	76,72	77,04	92,87	76,88
Han et al. (2023)	Abordagem multimodal similar à de Wang, utilizando ECG e GAF com atenção em aprendizado por múltiplas instâncias	–	56,65	66,08	–	69,10
Di Paolo; Castro (2024)	Modelo multimodal com imagens HSFC e RP; usa técnicas de aumento de dados com SMOTE e WGAN-GP	98,48	94,15	80,23	96,34	81,91
Modelo proposto neste trabalho	CNN com atenção local baseada em função Tanh e kernel adaptativo; entrada via imagens GAF extraídas de sinais MLII	97,88	77,45	76,14	95,39	76,78

visam expandir as capacidades e potencialização do desempenho alcançado neste trabalho. Isso amplia a aplicabilidade clínica dos classificadores de arritmia com redes neurais convolucionais e atenção local, contribuindo diretamente para a transformação da inteligência clínica da solução desenvolvida.

REFERÊNCIAS

- [1] Z. Ahmad *et al.*, “ECG heartbeat classification using multimodal fusion,” *IEEE Access*, vol. 9, 2021.
- [2] J. Huang *et al.*, “ECG arrhythmia classification using stft-based spectrogram and convolutional neural network,” *IEEE Access*, vol. 7, 2019.
- [3] A. Yousuf *et al.*, “Inferior myocardial infarction detection from lead ii of ECG: A gramian angular field-based 2d-cnn approach,” *Sensor Applications*, vol. 8, no. 10, 2024.
- [4] H. Mewada, “2d-wavelet encoded deep CNN for image-based ECG classification,” *Multimedia Tools and Applications*, vol. 82, 2023.
- [5] F. Zhou and D. Fang, “Multimodal ECG heartbeat classification method based on a convolutional neural network embedded with fca,” *Scientific Reports*, vol. 14, p. 8804, 2024.
- [6] Í. F. Di Paolo, “Classificação automática de arritmias cardíacas através de redes neurais convolucionais multimodais com mecanismo de atenção,” Tese de Doutorado, Universidade Federal do Pará, 2025.
- [7] M. S. Islam *et al.*, “Hard: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured rnn with dilated CNN,” *Neural Networks*, vol. 162, 2023, DOI: <https://doi.org/10.1016/j.neunet.2023.03.004>.
- [8] G. D. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods And Tools for ECG Data Analysis*. Artech House, 2006.
- [9] F. M. Dias *et al.*, “Arrhythmia classification from single-lead ECG signals using the inter-patient paradigm,” *Computer Methods and Programs in Biomedicine*, vol. 202, 2021.
- [10] T. Wang *et al.*, “Automatic ECG classification using continuous wavelet transform and convolutional neural network,” *Entropy*, vol. 23, no. 1, 2021.
- [11] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, 14th ed. Elsevier, 2021.
- [12] C.-H. Hsieh, Y.-S. Li, B.-J. Hwang, and C.-H. Hsiao, “Detection of atrial fibrillation using 1d convolutional neural network,” *Sensors (Basel)*, vol. 20, no. 7, p. E2136, 2020.
- [13] Z. Wang and T. Oates, “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*, 2015.
- [14] A. Vaswani *et al.*, “Attention is all you need,” in *NeurIPS Proceedings*, 2017, neural Information Processing Systems 30, Curran Associated Inc.
- [15] Z. Chen *et al.*, “Self-attention-based conditional variational autoencoder generative adversarial networks for hyperspectral classification,” *Remote Sensing*, vol. 13, 2021.
- [16] Z. Ahmad and N. Khan, “CNN-based multistage gated average fusion (MGAF) for human action recognition using depth and inertial sensors,” *IEEE Sensors Journal*, vol. 21, no. 3, 2021.
- [17] H. Ma *et al.*, “Change detection needs neighborhood interaction in transformer,” *Remote Sensing*, vol. 15, 2023, DOI: <https://doi.org/10.3390/rs15235459>.
- [18] T. Fan *et al.*, “A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition,” *Computers in Biology and Medicine*, vol. 159, 2023.
- [19] Y. Huang *et al.*, “ECG classification based on guided attention mechanism,” *Computer Methods and Programs in Biomedicine*, vol. 257, 2024, DOI: <https://doi.org/10.1016/j.cmpb.2024.108454>.
- [20] A. Srivastava *et al.*, “A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG,” *Physiological Measurement*, vol. 43, no. 6, 2022.
- [21] S. Woo *et al.*, “CBAM: Convolutional block attention module,” *arXiv preprint arXiv:1807.06521*, 2018.
- [22] D. Wan *et al.*, “Mixed local channel attention for object detection,” *Engineering Applications of Artificial Intelligence*, vol. 123, 2023.
- [23] Z. Chen *et al.*, “A novel imbalanced dataset mitigation method and ECG classification model based on combined 1d CBAM autoencoder and lightweight CNN model,” *Biomedical Signal Processing and Control*, vol. 87, 2024.
- [24] J. Emrich *et al.*, “Accelerated sample-accurate r-peak detectors based on visibility graphs,” in *Proceedings of the 31st European Signal Processing Conference (EUSIPCO)*, Helsinki, Finland, September 4–8 2023.
- [25] P. de Chazal, M. O’Dwyer, and R. B. Reilly, “Automatic classification of heartbeats using ECG morphology and heartbeat interval features,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, 2004.
- [26] H. Han, C. Lian, Z. Zeng, B. Xu, J. Zang, and C. Xue, “Multimodal multi-instance learning for long-term ECG classification,” *Knowl. Based Syst.*, vol. 270, 2023.
- [27] R. F. Oliveira *et al.*, “Explorando redes neurais de grafos para classificação de arritmias,” in *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, 2022.
- [28] Í. F. Di Paolo and A. R. G. Castro, “Intra- and interpatient ECG heartbeat classification based on multimodal convolutional neural networks with an adaptive attention mechanism,” *Applied Sciences*, vol. 14, no. 20, p. 9307, 2024, DOI: <https://doi.org/10.3390/app14209307>.