

UNIVERSIDADE FEDERAL DE OURO PRETO ESCOLA DE MINAS COLEGIADO DO CURSO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO - CECAU



ARTHUR CESAR DA SILVA ALMEIDA

CLASSIFICAÇÃO AUTOMÁTICA DE ARRITMIAS CARDÍACAS: UMA ANÁLISE DAS VARIAÇÕES DE DESEMPENHO ENTRE DIFERENTES ARQUITETURAS CNN'S E CONFIGURAÇÕES DO *DATASET*

MONOGRAFIA DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Ouro Preto, 2023

ARTHUR CESAR DA SILVA ALMEIDA

CLASSIFICAÇÃO AUTOMÁTICA DE ARRITMIAS CARDÍACAS: UMA ANÁLISE DAS VARIAÇÕES DE DESEMPENHO ENTRE DIFERENTES ARQUITETURAS CNN'S E CONFIGURAÇÕES DO *DATASET*

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheiro de Controle e Automação.

Orientador: Prof^o. Eduardo José da Silva Luz, Dr. Ciências da Computação. Coorientador: Prof^a. Adrielle de Carvalho Santana, Dr. Engenharia Elétrica.

> Ouro Preto Escola de Minas – UFOP 2023



MINISTÉRIO DA EDUCAÇÃO UNIVERSIDADE FEDERAL DE OURO PRETO REITORIA ESCOLA DE MINAS DEPARTAMENTO DE ENGENHARIA CONTROLE E AUTOMACAO



FOLHA DE APROVAÇÃO

Arthur Cesar da Silva Almeida

Classificação Automática de arritmias cardíacas: Uma análise das variações de desempenho entre diferentes arquiteturas CNN's e configurações do *dataset*

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Engenharia de Controle e Automação

Aprovada em 24 de agosto de 2023

Membros da banca

Dr. Eduardo José da Silva Luz - Orientador (DECOM - Universidade Federal de Ouro Preto) Dra. Adrielle de Carvalho Santana - Coorientadora (DECAT - Universidade Federal de Ouro Preto) Me. Vinícius Antônio de Oliveira Martins - Examinador (DECOM - Universidade Federal de Ouro Preto) Dr. Alan Kardek Rego Segundo - Examinador (DECAT - Universidade Federal de Ouro Preto)

Eduardo José da Silva Luz, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 28/08/2023



Documento assinado eletronicamente por **Eduardo Jose da Silva Luz**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 28/08/2023, às 19:51, conforme horário oficial de Brasília, com fundamento no art. 6°, § 1°, do <u>Decreto n° 8.539, de 8 de outubro de 2015</u>.



A autenticidade deste documento pode ser conferida no site <u>http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0</u>, informando o código verificador **0577984** e o código CRC **B509F686**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.010997/2023-06

Dedico este trabalho a minha família, que tanto admiro, e a todos que acreditaram em mim e me deram suporte para continuar seguindo em frente.

AGRADECIMENTOS

Esta monografia simboliza a realização de uma enorme conquista pessoal e um sentimento de dever cumprido após muito tempo, empenho e uma quantidade imensa de café. Como diz o ditado *"Entre o desejo e o alcance, há apenas dois moradores, o esforço e a determinação."*.

Gostaria de agradecer, primeiramente, a Deus pela vida, e por sempre guiar meus passos em tempos incertos. Por colocar pessoas incríveis no meu caminho, permitindo que eu nunca me sentisse perdido.

A meus pais, Maria Helena e Júlio Cesar, por todo o apoio e dedicação que sempre tiveram comigo. Estendo este agradecimento também a minha irmã, Júlia, que sempre torceu pelo meu sucesso. As minhas avós, Maria Aparecida e Nilza de Aguiar *(in memorian)* que tanto amo. À tia Sandra, Mireli e meus primos que a todo momento me ampararam e me motivaram com suas palavras. Sem vocês não conseguiria chegar onde estou.

Agradeço a todas as instituições de ensino que estudei, começando pela Escola Municipal Bartholomeu Anacleto, Instituição Tia Maninha, Colégio Marechal Floriano Peixoto e a Escola Técnica Pandiá Colágeras. A todos os meus professores que tive a grande oportunidade de conhecer em minha vida acadêmica e que me mostraram que a educação é a arma mais poderosa que você pode usar para mudar o mundo.

Aos meus grandes amigos da Vila dos Remédios - Floriano, José Carlos, Nicolas Dias, Marcelo Mumu e George. Que em toda oportunidade que eu tinha de voltar para casa nas férias, me recebiam de braços abertos. Um agradecimento sincero a vocês que sempre me fizeram ver que desistir não é a solução.

Estendo o agradecimento para Larissa Quintão, Larissa Lara, Henrique Chagas, Matheus Teixeira e Erlane. Amizades que foram consolidadas e posto a prova de fogo ao longo desses seis anos de curso. Obrigado por todo apoio e pelos muitos momentos divertidos que vivemos em cada prova e em cada aula.

A Eterna e Saudosa República Sparta, pelos intensos anos vividos aí e por serem minha primeira casa em Ouro Preto. Em especial ao Escorado, Vovó, Sabonete, Jahnela, Dimoc e Pimpão. Ao Vitor/Sereia e a República Alcatéia e todos seus moradores, sobretudo ao CVerino, 1/2 toza e cinco/keko, por me oferecerem abrigo quando precisei. Vocês foram uma amizade inesperada, uma segunda família para mim mesmo nesse pouco tempo de convívio.

A meus professores e orientadores ao longo desse jornada. Em especial, Eduardo Luz, Vinícios Martins, Sílvia e Tamirez, que me guiaram pelo mundo da ciência, permitindo me desenvolver como acadêmico. Vocês se tornaram uma referência para min.

Gostaria de expressar minha sincera gratidão à Gerdau Ouro Branco e a toda a equipe

de *capex sustain* da coqueria pela oportunidade de estágio. Em particular, gostaria de expressar minha gratidão ao Marco Aurélio, Victor, Stela, Amsterdan, Flávia, Anna, Herbert e Hegberton, meus colegas de trabalho e instrutores, que dedicaram seu tempo e esforço para me ajudar a crescer profissional e pessoalmente. Admiro muito o trabalho e a dedicação que demonstraram e espero poder seguir seus exemplos no futuro. Desejo-lhes todo o sucesso em suas carreiras e na vida pessoal.

Agradeço à Universidade Federal de Ouro Preto (UFOP) e a todos os seus funcionários que me acolheram durante este trabalho. A cidade histórica de Ouro Preto, por ser minha casa e me proporcionar a experiência mais incrível que já tive até o momento. Além disso, sou extremamente grato à Fundação Gorceix pelo apoio financeiro que me foi concedido, pois sem ele não teria sido possível realizar este trabalho com sucesso.

Por fim, à aqueles que leram os agradecimentos até aqui, muito obrigado!

"A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê." (Arthur Schopenhauer)

RESUMO

Doenças cardiovasculares são males que afetam a anatomia ou o funcionamento do coração, veias e artérias, sendo a principal causa de morte no mundo nos últimos 20 anos. Dentre essas doenças destacam-se as arritmias cardíacas, que são alterações dos estímulos elétricos do coração. Um dos principais exames para a detecção de arritmias cardíacas é o eletrocardiograma que mesmo atualmente ainda representa um método de perícia defasada e geralmente suscetível a erros humanos. Considerando esse contexto, o presente trabalho propõe uma investigação de desempenho das arquiteturas de redes neurais convolucionais (CNN's) comumente empregadas na literatura (CNN-1D, CNN-2D), juntamente com uma abordagem alternativa, fundamentada na estratégia de atenção de dois níveis, (CNN Híbrida) para a classificação automática de arritmias cardíacas baseadas em três classes: batimento normal (Classe N), batimento ectópico ventricular (Classe V) e batimento ectópico supraventricular (Classe S). No estudo dos modelos CNN, também é ressaltada a importância da utilização de conjuntos de dados de sinais de ECG devidamente parametrizados. Outro ponto relevante deste estudo é o impacto que a utilização de conjuntos de dados balanceados e desbalanceados pode resultar no desempenho da rede neural. Dentre as arquiteturas propostas sob a perspectiva de um conjunto de dados desbalanceado, a CNN Híbrida se destacou como o modelo de melhor desempenho na classificação entre as classes de arritmias, obtendo as seguintes pontuações no F1-Score: 75% para a classe N, 29% para a classe S e 71% para a classe V. Contudo, se apenas a acurácia for levada em consideração, a arquitetura CNN-1D se destaca por alcançar uma taxa de 68%. Sob a outra perspectiva do treinamento ser realizado a partir de um conjunto de dados balanceado, a CNN Híbrida novamente se sobressai ao alcançar o melhor desempenho, obtendo uma pontuação no F1-Score de 53% para a classe N, 60% para a classe S, 77% para a classe V e uma acurácia total de 63%, a maior dentre os três modelos apresentados. As arquiteturas CNN's 1D e 2D também mostraram-se eficazes, com um desempenho muito próximo a CNN Híbrida, principalmente sob a perspectiva da quantidade de parâmetros utilizados em cada modelo. Tanto a CNN-1D quando a 2D apresentam cerca de 1/3 da quantidade de parâmetros encontrados na CNN Híbrida.

Palavras-chaves: Aprendizado profundo, Redes Neurais Convolucionais, Sinais de Eletrocardiograma, parametrização de dados, Arritmia Cardíaca e Análise de desempenho.

ABSTRACT

Cardiovascular diseases are ailments that affect the anatomy or functioning of the heart, veins, and arteries. They have been the leading cause of death worldwide over the past two decades. Among these conditions, cardiac arrhythmias stand out, which entail disruptions in the heart's electrical stimuli. One of the key examinations for detecting cardiac arrhythmias is the electrocardiogram, which even in the present day remains an outdated method of expertise and is often susceptible to human errors. In light of this context, the current study proposes an investigation into the performance of commonly employed Convolutional Neural Network (CNN) architectures (CNN-1D, CNN-2D) along with an alternative approach grounded in a two-level attention strategy (Hybrid CNN) for the automatic classification of cardiac arrhythmias based on three classes: normal beats (Class N), ventricular ectopic beats (Class V), and supraventricular ectopic beats (Class S). The importance of using appropriately parameterized datasets of ECG signals is also emphasized in the analysis of CNN models. Another significant aspect of this study is the impact that balanced and unbalanced datasets can have on neural network performance. Among the proposed architectures from the perspective of an unbalanced dataset, the Hybrid CNN outperformed as the best-performing model in classifying arrhythmias, achieving the following F1-Score scores: 75% for Class N, 29% for Class S, and 71% for Class V. However, if accuracy alone is taken into account, the CNN-1D architecture stands out with a rate of 68%. From the viewpoint of training with a balanced dataset, the Hybrid CNN once again excels, achieving an F1-Score of 53% for Class N, 60% for Class S, 77% for Class V, and a total accuracy of 63%, the highest among the three presented models. The CNN-1D and 2D architectures also demonstrated effectiveness, with performance closely approaching that of the Hybrid CNN, particularly considering the quantity of parameters used in each model. Both CNN-1D and 2D architectures feature roughly one-third of the parameter count found in the Hybrid CNN.

Key-words: Deep Learning, Convolutional Neural Networks, Electrocardiogram Signals, Data Parameterization, Cardiac Arrhythmia, Performance Analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Registro do primeiro ECG. .	18	
Figura 2 – Ondas e intervalos elementares de um sinal de ECG	19	
Figura 3 – Aprendizado de Máquina vs Aprendizado Profundo. . <th .<="" <="" td=""><td>22</td></th>	<td>22</td>	22
Figura 4 – Neurônio Humano	22	
Figura 5 – Neurônio Artificial.	23	
Figura 6 Operação de Convolução.	24	
Figura 7 – Subamostragem do valor máximo e da média. .	25	
Figura 8 – Camada totalmente conectada. . <th< td=""><td>25</td></th<>	25	
Figura 9 – Arquitetura LeNet-5. .	26	
Figura 10 – Arquitetura CNN Simplificada.	26	
Figura 11 – Sinal de ECG convertida em imagem em escala de cinza	27	
Figura 12 – Arquitetura CNN baseada em atenção de dois níveis.	28	
Figura 13 – Arquitetura CNN - 2ª Abordagem.	30	
Figura 14 – Sinais de ECG segmentados.	33	
Figura 15 – Função de Ativação ReLU	37	
Figura 16 – Ilustração da técnica de Dropout.	38	
Figura 17 – Representação esquemática da arquitetura CNN-1D	39	
Figura 18 – Representação alternativa das ondas de ECG	40	
Figura 19 – Imagem de um sinal de ECG - 32x32 píxeis	41	
Figura 20 – Representação esquemática da arquitetura CNN-2D	41	
Figura 21 – Representação esquemática canal de convolução 2D	42	
Figura 22 – Representação esquemática canal de convolução 1D	43	
Figura 23 – Representação esquemática camada densa	43	
Figura 24 – Decaimento da taxa de aprendizado	45	
Figura 25 – Matriz de confusão para classificação binária - Exemplo	49	
Figura 26 – Matriz de confusão da CNN-1D para dados desbalanceados	51	
Figura 27 – Processo de treinamento CNN-1D para dados desbalanceados	51	
Figura 28 – Matriz de confusão da CNN-1D para dados balanceados.	52	
Figura 29 – Processo de treinamento CNN-1D para dados balanceados	53	
Figura 30 – Matriz de confusão CNN-2D para dados desbalanceado	54	
Figura 31 – Processo de treinamento CNN-2D para dados desbalanceados	54	
Figura 32 – Matriz de confusão CNN-2D para conjunto de dados balanceado	55	
Figura 33 – Processo de treinamento CNN-2D para dados balanceados	56	
Figura 34 – Matriz de confusão CNN-Híbrida para dados desbalanceados	57	
Figura 35 – Processo de treinamento CNN-Híbrida para dados desbalanceados	57	
Figura 36 – Matriz de confusão CNN-Híbrida para dados balanceado	58	

Figura 37 – Processo de treinamento CNN-Híbrida para dados balanceados 59

LISTA DE TABELAS

Tabela 1 Classes MIT-BIH Arritmia Database	32
Tabela 2 – Classes de batimentos cardíacos AAMI	32
Tabela 3 – Distribuição dos dados por classes e pacotes	35
Tabela 4 – Porcentagem de distribuição dos dados por pacotes	35
Tabela 5 – Distribuição padronizada dos dados.	36
Tabela 6 – Estrutura do modelo CNN-1D. . <th< td=""><td>39</td></th<>	39
Tabela 7 – Estrutura CNN-2D.	42
Tabela 8 – Estrutura CNN-Híbrida.	44
Tabela 9 – Relação de perda e acurácia CNN-1D para dados desbalanceados.	50
Tabela 10 – Resultados de performance do teste da CNN-1D para dados desbalanceados.	50
Tabela 11 – Relação de perda e acurácia CNN-1D para dados balanceados.	52
Tabela 12 – Resultados de performance do teste da CNN-1D para dados balanceados	52
Tabela 13 – Relação de perda e acurácia CNN-2D para dados desbalanceados.	53
Tabela 14 – Resultados de desempenho do teste da CNN-2D para dados desbalanceados.	53
Tabela 15 – Relação de perda e acurácia CNN-2D para dados balanceados.	55
Tabela 16 – Resultados de desempenho do teste da CNN-2D para dados balanceados	55
Tabela 17 – Relação de perda e acurácia CNN-Híbrida para dados desbalanceados. .	56
Tabela 18 – Resultados de desempenho do teste da CNN-Híbrida para dados desbalance-	
ados	56
Tabela 19 – Relação de perda e acurácia CNN-Híbrida para dados balanceados.	58
Tabela 20 – Resultados de desempenho do teste CNN-Híbrida para dados balanceados.	58
Tabela 21 – Variações do F1-Score entre as arquiteturas e conjuntos de dados utilizados.	59

LISTA DE ABREVIATURAS E SIGLAS

1D	Unidimensional
2D	Bidimensional
AAMI	Association for the Advancement of Medical Instrumentation ou Associação para Avanção de Instrumentação Médica
AV	Atrioventricular
AVC	Acidente Vascular Celebral
bpm	Batimentos por minuto
CNN	Convolutional Neural network ou Rede Neural Convolucional
DCV	Doenças cardiovasculares
DL	Deep Learning ou Aprendizado Profundo
ECG	Eletrocardiograma
FA	Fibrilação Atrial
GPU	Graphics Processing Unit ou Unidade de Processamento Gráfico
IA	Inteligência Artificial
ms	milissegundo
mV	milivolt
ONU	Organização das Nações Unidas
ReLU	Unidade Linear Retificada
RNA	Redes Neurais Artificiais
UFOP	Universidade Federal de Ouro Preto

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Motivação e Problema de Pesquisa	15
1.2	Objetivos	17
1.2.1	Objetivo Geral	17
1.2.2	Objetivos Específicos	17
1.3	Estrutura da Monografia	17
2	REFERENCIAL TEÓRICO	18
2.1	Eletrocardiograma	18
2.1.1	História	18
2.1.2	Características do sinal ECG	19
2.1.3	Arritmia Cardíaca	19
2.2	Inteligência Artificial	20
2.2.1	Machine Learning	21
2.2.2	Deep Learning	21
2.2.3	Rede Neural Convolucional - CNN	23
2.3	Trabalhos Relacionados	26
3	METODOLOGIA 3	31
3.1	Base de dados	31
3.1.1	Classes dos Sinais de ECG	31
3.2	Pré-Processamento dos dados	33
3.2.1	Remapeamento das classes	33
3.2.2	Parametrização dos dados	34
3.2.2.1	Dataset desbalanceado	34
3.2.2.2	Dataset balanceado	35
3.2.3	Normalização dos dados	36
3.3	Arquiteturas dos modelos CNN's	36
3.3.1	Características	36
3.3.2	Arquitetura CNN 1D	38
3.3.3	Arquitetura CNN 2D	39
3.3.4	Arquitetura CNN Híbrida	42
3.3.5	Ambiente de desenvolvimento	44
3.4	Metodologia de Treinamento e Validação	44
3.4.1	Taxa de aprendizado	45
3.4.2	Método de regularização de penalidades	46

3.4.3	Inicialização dos pesos	46
3.4.4	Tamanho do lote ou batch size	46
3.5	Treinamento	46
3.6	Métricas de avaliação	47
3.6.1	Conjunto de validação	47
3.6.2	Indicadores de desempenho	48
3.6.2.1	Matriz de Confusão	49
4	EXPERIMENTOS E RESULTADOS	50
4.1	Arquitetura CNN 1D	50
4.1.1	Dataset desbalanceado	50
4.1.2	Dataset balanceado	51
4.2	Arquitetura CNN 2D	53
4.2.1	Dataset desbalanceado	53
4.2.2	Dataset balanceado	55
4.3	Arquitetura CNN Híbrida	56
4.3.1	Dataset desbalanceado	56
4.3.2	Dataset balanceado	58
4.4	Discussão	59
5	CONCLUSÃO	51
	REFERÊNCIAS	62
	APÊNDICE A – REPOSITÓRIO	56

1 INTRODUÇÃO

Neste capítulo é apresentado, de forma sucinta, a motivação e os objetivos do trabalho. No primeiro momento é discorrido os problemas que induziram a concepção deste estudo. Em seguida, é feita uma descrição dos objetivos gerais e específicos que o mesmo pretende atingir. Por fim, são apresentados os capítulos que compõem esta monografia.

1.1 Motivação e Problema de Pesquisa

Doenças cardiovasculares (DCV) são males que afetam a anatomia ou o funcionamento do coração, veias e artérias, abrangendo arritmias, doença arterial coronariana, pressão arterial irregular, doença cerebrovascular, parada cardiorrespiratória, dentre outras (SANTANA et al., 2021). Essas enfermidades ocasionam o aumento da morbidade, mortalidade prematura, incapacidades e perda da qualidade de vida, sendo a principal causa de morte no mundo nos últimos 20 anos, representando cerca de 16% do total de mortes por todas as causas (ONU, 2020).

No Brasil o panorama não é diferente. As DCV constituem a primeira causa de óbitos desde a década de 1990, passando de 270 mil para 400 mil óbitos em 2019. Isso representa um aumento de 48% no número de casos em um período de 29 anos e um grande desafio para a saúde pública. Esse aumento está associado ao envelhecimento da população, a condições de saúde como hipertensão, diabetes, obesidade, sedentarismo, fatores sociodemográficos, etc (GOMES et al., 2021).

Dentre as principais DCV existentes, este trabalho se concentrará na investigação das arritmias cardíacas, também chamadas de disritmias cardíacas. Essa condição surge quando há alteração na formação e/ou condução do estímulo elétrico do coração, resultando na redução do fluxo sanguíneo para as células (SZPALHER; BATALHA, 2019). A arritmia mais predominante na população é a fibrilação atrial (FA) que, quando ocorre, ao invés do coração contrair e relaxar, o órgão apenas treme, ou seja, fibrila. A FA está associada ao aumento do risco de acidentes vasculares cerebral (AVC), insuficiência cardíaca, demência e morte (RIENSTRA et al., 2012).

A técnica mais comum e não invasiva para o diagnóstico de arritmias é o eletrocardiograma (ECG), que foi introduzida há mais de 120 anos, no início do século XX, pelos pesquisadores Augustus Waller e Willem Einthoven. O ECG é um método seguro, rápido, de simples realização e baixo custo. Essas características fazem dele um método amplamente difundido em todo o mundo. Atualmente, além das arritmias e distúrbios de condução, ele pode identificar quadro isquêmicos coronarianos, alterações metabólicas, além de efeitos de alguns medicamentos (FERNANDES et al., 2015). Contudo, em alguns casos, as arritmias podem ocorrer de forma eventual e episódica. Nessas circunstâncias somente o ECG pode demonstrar-se insuficiente para a realização do diagnóstico, sendo necessário a execução de outros exames

(ROCHA, 2018a).

O diagnóstico de arritmia depende de profissionais médicos especializados. Por convenção, esses especialistas examinam visualmente as formas de onda do ECG. Entretanto, essa análise visual pode resultar em diagnósticos falhos, seja por problemas de similaridade ou por dificuldade em detectar e categorizar as formas de onda do sinal. Além disso, é comum haver casos onde é necessário a avaliação de sinais de ECG com longos períodos de duração. Para esses casos, a análise visual é muito extensa e exaustiva, o que aumenta significativamente as chances de um diagnóstico falho (SANTANA et al., 2021). Contudo, com os avanços tecnológicos, essas limitações podem ser mitigadas com o emprego de técnicas computacionais voltadas para a detecção e classificação automática de arritmias cardíacas baseadas em sinais de ECG.

Métodos baseados em aprendizado profundo (*deep learning*) vêm sendo amplamente explorados na literatura, em especial as redes neurais convolucionais (CNN's) devido a sua capacidade de extrair melhores representações dos dados (CHEN et al., 2020). Atualmente, as CNN's conseguem alcançar um desempenho ao nível humano, em muitas aplicações relacionadas à visão computacional, inclusive na própria detecção e classificação de arritmias cardíacas (HANNUN et al., 2019).

Na literatura, é possível encontrar diversos trabalhos em que várias arquiteturas de CNN's foram aplicadas com essa finalidade. Alguns estudos fizeram uso de modelos unidimensionais (CNN-1D), visando a extração de características temporais dos sinais de ECG, enquanto outros autores optaram por empregar modelos bidimensionais (CNN 2D). No entanto, é comum encontrar uma falta de padronização na maneira como os estudos abordam a avaliação de desempenho das arquiteturas CNN. Muitas vezes, os resultados apresentados são superficiais, focando apenas na acurácia geral do modelo e negligenciando a análise mais detalhada das métricas de desempenho específicas para cada classe.

Além disso, muitos trabalhos acadêmicos não detalham nem avaliam adequadamente o impacto do pré-processamento dos dados (LUZ et al., 2016). Muitas vezes, esses trabalhos se concentram principalmente na arquitetura e no treinamento da rede neural, deixando de lado a importância do pré-processamento dos dados, que pode ter um impacto significativo no desempenho e nos resultados. Um pré-processamento inadequado dos dados pode originar uma generalização limitada, levando a resultados subótimos, maior tempo de treinamento, dificuldade de reprodução do experimento e validação dos resultados. Por fim, sem avaliar o impacto do pré-processamento, torna-se difícil comparar o desempenho de diferentes abordagens ou modelos em condições justas.

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo geral deste trabalho consiste em investigar o desempenho de diferentes estruturas de redes neurais convolucionais (arquiteturas unidimensionais, bidimensionais e arquiteturas híbridas com dois canais de atenção), na classificação automática de arritmias cardíacas, utilizando conjuntos de dados de sinais de ECG devidamente parametrizados.

1.2.2 Objetivos Específicos

Para que o objetivo geral possa ser alcançado, os seguintes objetivos específicos foram definidos:

- Investigar o estado da arte de CNN's utilizadas para a classificação de arritmias cardíacas a partir de sinais de ECG, para servir como base de comparação neste estudo.
- Realizar uma análise das bases de dados de sinais de ECG presentes na literatura, a fim de selecionar o conjunto de dados mais adequado para este estudo e realizar o préprocessamento e parametrização dos seus dados.
- Elaborar modelos de CNN's a partir de diferentes abordagens e avaliar o desempenho individual de cada proposta.
- Gerar variações no *dataset* visando avaliar o impacto de desempenho que cada arquitetura estudada poderá sofrer durante a etapa de treinamento e inferência.

1.3 Estrutura da Monografia

Este trabalho está estruturado da seguinte forma: o capítulo 2 apresenta uma revisão teórica referente aos principais conceitos abordados nesta monografia. Além da investigação de trabalhos relevantes que fizeram uso de arquiteturas de CNN's para o reconhecimento de sinais de ECG. O capítulo 3 descreve os materiais e os procedimentos metodológicos propostos neste estudo. Os resultados alcançados são apresentados e discutidos no capítulo 4. Finalmente, no capítulo 5, são realizadas reflexões a respeito da metodologia adotada e as possibilidades para trabalhos futuros também são discutidas.

2 REFERENCIAL TEÓRICO

Em um primeiro momento, este capítulo apresenta os conceitos fundamentais necessários para a compreensão da pesquisa como um todo, depois alguns trabalhos relevantes nesta área são explorados com o intuito de fornecer um ponto de referência para comparações ou escolhas.

2.1 Eletrocardiograma

2.1.1 História

O eletrocardiograma é um exame na qual é obtida uma representação gráfica da atividade elétrica que percorre os músculos do coração durante sua contração e expansão. A história do ECG começa em torno de 1887, quando o fisiologista britânico Augustus Waller, usando um eletrômetro capilar de Lippman e eletrodos conectados no tórax anterior e posterior, obteve o primeiro eletrocardiograma humano (Figura 1), demonstrando que o batimento cardíaco era acompanhado por uma atividade elétrica. Contudo, ele não acreditava que tal feito teria alguma aplicação clínica devido as enormes dificuldades técnicas daquela época (GIFFONI; TORRES, 2010).



Figura 1 – Registro do primeiro ECG. Fonte: Adaptado de Waller (1887).

Devido aos resultados alcançados, Waller acabou apresentando sua pesquisa no primeiro congresso internacional de fisiologistas, em Basel, Suíça, em 1889. Esse evento chamou a atenção de diversos outros pesquisadores dessa área, incluindo o médico holandês Willem Eithoven.

Poucos anos depois, em 1901, Eithoven desenvolveu uma nova solução a partir do galvanômetro de corda que superava as limitações da pesquisa de Waller fornecendo registros com qualidade muito semelhante às dos equipamentos atuais. Contudo, tal mecanismo criado pesava cerca de 270 quilos e necessitava de cinco pessoas para opera-la.

Desde então vários cientistas, como Thomas Lewis, destacaram a importância do ECG para a compreensão das arritmias, uma área até então rudimentar sem fundamentação científica (GIFFONI; TORRES, 2010).

2.1.2 Características do sinal ECG

O sinal de ECG, representado na Figura 2, é constituído por uma série de ondas (onda P, complexo QRS, onda T e onda U) e intervalos. Essas ondas são resultados do processo de polarização e despolarização do tecido cardíaco (REGIS; CALDEIRA; GURJÃO, 2016).

O ciclo cardíaco tem início com a onda P, que caracteriza a despolarização atrial, que sob condições normais, deve ter duração inferior a 0,12 segundos e a amplitude inferior a 0,25 mV. O intervalo PR refere-se ao tempo de condução do estímulo elétrico pelo nó atrioventricular (AV), do início da onda P até o início do complexo QRS. Sua duração normal deve ser entre 0,12 e 0,20 segundos. O complexo QRS é relacionado ao período em que ocorre a despolarização e ativação dos ventrículos. Sua duração normal é entre 0,06 e 0,11 segundos para QRS estreito. O intervalo QT é responsável por medir a duração do processo de despolarização e repolarização nos ventrículos do coração. Por fim, a onda T indica a repolarização ventricular juntamente com o segmento ST (GARCÍA et al., 2015).



Figura 2 – Ondas e intervalos elementares de um sinal de ECG. Fonte: Adaptado de Regis, Caldeira e Gurjão (2016)

2.1.3 Arritmia Cardíaca

As arritmias, também chamadas de disritmias, surgem quando há alteração na formação e/ou condução do estímulo elétrico cardíaco. Elas são frequentes na população em geral, mas em particular nos pacientes com doenças cardiovasculares. Porém, podem surgir em pacientes devido a outros fatores como, tabagismo, excesso de cafeína, alterações emocionais, esforço físico, etc. Além disso, sua gravidade também é variável. Enquanto algumas podem ser benignas, outras podem apresentar risco imediato de vida (SOUZA; FILHO, 2019).

Há vários tipos de arritmias cardíacas que podem ser classificadas, com base em quatro conceitos eletrofisiológicos, a partir de um ECG de 12 derivações (GARCÍA et al., 2015).

Conforme a frequência cardíaca

- Bradiarritmias: Caracterizada por uma frequência cardíaca inferior a 60 batimentos por minuto (bpm).
- Taquiarritmias: Caracterizada por uma frequência cardíaca superior a 100 bpm em pelo menos 3 batimentos consecutivos.

Conforme a amplitude do complexo QRS

- QRS estreito ou normal: Complexos QRS com duração inferior a 120 milissegundos (ms) indicam uma despolarização ventricular normal.
- QRS largo: Complexos QRS com uma duração superior ou igual a 120 ms indicam algum comprometimento ventricular, por exemplo, a presença de um bloco de ramificação estrutural pré-existente.

Conforme a regularidade do complexo QRS

- Regular: Intervalo R-R, dos complexos QRS, for constante.
- Irregular: Intervalo R-R, dos complexos QRS, for variável.

Conforme a origem da arritmia

- Supraventricular: Desencadeada por algum problema localizado na região acima do feixe Atrioventricular, também conhecido como feixe de HIS, podendo ser nos átrios, nó sinusal e/ou nó atrioventricular.
- Ventricular: Desencadeada por algum problema localizado na região abaixo do feixe de HIS, nos ventrículos. Ela possui grande potencial de letalidade e pode ser classificada em quatro categorias: monomórficas, polimórficas, estáveis e instáveis.

2.2 Inteligência Artificial

O ser humano tem capacidade de pensar e raciocinar desde o começo de sua evolução que ocorreu há milhares de anos. Inclusive, é devido a esse fato que nos autodenominamos de *Homo sapiens* (homem sábio). Mas o que é inteligência? Essa é uma pergunta que até hoje não há um consenso. Segundo Kaufman (2019), uma possível definição para inteligência é a "capacidade de aprender ideias complexas, de se adaptar efetivamente ao ambiente, de aprender com a experiência, de se envolver em várias formas de raciocínio, de superar obstáculos".

Com a evolução da sociedade e das tecnologias ao longo do tempo, o homem pôde aprender, prever e controlar um mundo muito maior e mais complexo que ele mesmo. Surge então o campo da Inteligência Artificial (IA), no qual o objetivo do homem não se limita apenas a compreender, mas sim a construir entidades inteligentes e fomentar a interação entre o ser humano e a máquina (GOMES, 2010). Especialistas desse campo, afirmam que nas próximas cinco décadas o nível de IA deve se comparar ao nível humano e o que a humanidade vivencia hoje é somente uma pequena parte do que está por vir (LOPES, 2022).

2.2.1 Machine Learning

O aprendizado de máquina (*machine learning*) é um ramo da área da inteligência artificial responsável por sistemas que podem cumprir uma tarefa sem uma programação explícita. Ou seja, são sistemas que aprendem automaticamente a partir de dados (KOROL, 2019). Sendo a inferência indutiva um dos principais métodos para produzir resultados confiáveis e predizer eventos. Contudo, a capacidade de generalização desses sistemas pode ser comprometida caso haja escassez de dados ou se eles apresentarem baixa qualidade ou falta de representatividade em relação ao que se deseja aprender.

Basicamente, há três tipos de algoritmos de aprendizado de máquina: supervisionado, não supervisionado e por reforço. No primeiro, é necessária uma base de dados (*dataset*) de treino que fornece exemplos ao algoritmo juntamente com seu rótulo informando a classe a qual o dado pertence. A ideia é que o algoritmo, após ser treinado, consiga inferir rótulos para dados que antes eram desconhecidos. No aprendizado não supervisionado, os dados são fornecidos sem rótulos, cabendo ao algoritmo agrupá-los pelas suas similaridades. Nesse tipo de aprendizado, não há necessidade de um *dataset* para treinamento. Por fim, tem-se o aprendizado por reforço, no qual o algoritmo não recebe a resposta correta, mas um sinal de reforço, seja de recompensa ou punição. Esse método assemelha-se ao primeiro por ser capaz de utilizar experiências anteriores e ao segundo por aprender de forma não supervisionada (LUDERMIR, 2021).

Como este trabalho se fundamenta no aprendizado supervisionado para classificação de arritmias, tal método será discutido mais profundamente nos próximos tópicos.

2.2.2 Deep Learning

O aprendizado profundo (*deep learning*) é uma sub área de aprendizado de máquina que busca extrair conceitos complexos de uma maneira hierárquica de dados brutos menos complexos. É uma técnica que tenta imitar o raciocínio que o cérebro humano utiliza para aprender e identificar padrões (ASSIS, 2019).

Para esse objetivo, esse tipo de rede se divide em múltiplas camadas. Cada camada é capaz de compreender uma característica diferente a respeito do objeto em questão, sendo composto por vários nós que transformam os dados em informação que são passadas adiante para a próxima camada a fim de prever a saída final (SCHONS, 2018).

Na Figura 3, é apresentado um exemplo de algoritmo de aprendizado de máquina

tradicional utilizado para o reconhecimento de imagens. Nesse algoritmo é necessário selecionar manualmente os recursos e um classificador para categorizar as imagens. Já um algoritmo de aprendizado profundo, é capaz de extrair os recursos automaticamente e aprender com seus próprios erros. Contudo, técnicas de aprendizado profundo exigem muito mais dados para funcionarem corretamente (SEIF, 2018).



Figura 3 – Aprendizado de Máquina vs Aprendizado Profundo. Fonte: Adaptado de Seif (2018).

Para conseguir tal feito, os algoritmos de *deep learning* funcionam com base em redes neurais artificiais (RNAs). A unidade fundamental dessas redes são os neurônios, inspirados na dinâmica dos neurônios cerebrais. A Figura 4 ilustra a representação do neurônio biológico, sendo formado pelos dendritos, corpo celular e axônio. Os dendritos possuem a função de receber os impulsos nervosos de outros neurônios da vizinhança e direcioná-los até o corpo celular (soma). No corpo da célula a informação é processada e se a soma das entradas elétricas for suficientemente poderosa para ativar o neurônio, pulsos são gerados e transmitidos ao longo do axônio (SANTANA et al., 2021).



Figura 4 – Neurônio Humano Fonte: Adaptado de Borges et al. (2015)

O modelo de neurônio artificial (Figura 5) é formado por três elementos. O primeiro é um conjunto de sinais de entrada ou sinapses, na qual cada entrada é correspondida por um peso específico. O segundo é um somador para realizar a soma ponderada de todas as entradas, podendo também ser adicionado um viés para corrigir um erro sistemático ou uma tendenciosidade da rede. Por última, há uma função de ativação que adiciona um comportamento não-linear aos dados de entrada, fazendo-o capaz de modelar funções mais complexas (KOROL, 2019).



Figura 5 – Neurônio Artificial. Fonte: Adaptado de Godoi (2005).

Os neurônios podem ser conectados de várias maneiras. Algumas das arquiteturas mais populares são: (i) rede neural *feedforward*, na qual os dados fluem em apenas uma direção, da camada de entrada para a camada de saída; (ii) rede neural recorrente, essas redes salvam a saída de uma camada e a devolvem à camada de entrada para ajudar a predizer o resultado; (iii) redes adversárias generativas, que são modelos treinados para a para geração de novos dados que seguem uma distribuição semelhante aos conjuntos de treinamento; (iv) rede neural convolucional, em que os neurônios da camada de extração de características não se conectam a todos os neurônios da camada seguinte, mas apenas a uma pequena região e o resultado é reduzido a um único vetor que segue para as camadas densas, camadas responsáveis pelo processo de classificação. A CNN é uma das técnicas de *deep learning* mais bem sucedidas para processamento de imagens, voz e vídeo (AZURE, 2022).

2.2.3 Rede Neural Convolucional - CNN

As CNN's foram inspiradas na configuração do córtex visual humano. Essa estrutura biológica consiste em um modelo hierárquico formado por camadas em cascata de células que desempenham a função de processamento das informações visuais e está completamente envolvida no processo de reconhecimento de objetos (CASTRO, 2020).

A arquitetura de uma CNN pode ser dividida em três grupos: camada de entrada, camadas de extração de características e camadas de classificação. As camadas de entrada são responsáveis por carregar os dados de entrada na rede. No caso desses dados serem imagens, por exemplo, serão carregadas matrizes bidimensionais, em que cada ponto da grade matricial, chamado de

píxel, possui um valor que representa a intensidade luminosa do ponto observado. Caso a imagem seja monocromática, uma única matriz é suficiente para representá-la, no entanto, caso o sistema de cores seja RGB serão necessárias três matrizes para representar a imagem, cada uma referente a um canal de cor (ALMEIDA, 2022).

As camadas de extração de características possuem um padrão comum, mas não obrigatório, que consiste em uma camada de convolução seguida de uma função de ativação e uma subamostragem. A camada de convolução para dados de duas dimensões é ilustrada na Figura 6, sendo composta por um filtro de *n* x *m* dimensões, também conhecido como kernel. Esse filtro desliza sobre a matriz de entrada, também conhecida como mapa de característica ou *feature map*. No decorrer do seu deslocamento, a passos pré-determinados, a operação de convolução é realizada entre os elementos sobrepostos da janela de entrada com os respectivos pesos do kernel. Após a multiplicação ter sido realizada, os valores são somados resultando numa saída. A operação de convolução acaba após o kernel ter percorrido toda a matriz de entrada, resultando numa nova matriz de saída. A proposta da CNN é que ela também aprenda quais são os filtros ideais para a identificação das características relevantes da imagem (CASTRO, 2020).



Figura 6 – Operação de Convolução.

Fonte: Adaptado de Wang et al. (2020).

A função de ativação tem o objetivo de aprimorar a robustez da rede, retificando o valor de saída da operação de convolução. Esse processo acaba menosprezando informações menos significantes (SCHONS, 2018). Contudo, não há como prever qual função de ativação terá um melhor desempenho para o projeto. O processo de design consiste em tentativa e erro. Normalmente, as funções de ativação mais usadas são a sigmoid, tangente hiperbólica (Tanh), unidade linear retificada (ReLu) e softplus (ROCHA, 2018a).

A subamostragem ou *polling* é uma operação que tem a finalidade de reduzir a dimensão dos mapas de características. Esse processo permite reduzir o número de parâmetros da rede e também ajuda a controlar o *overfitting*. Sua estrutura é parecida com a camada de convolução,

na qual janelas deslizantes também são empregadas para extração de um valor dessa região. Há dois tipos de subamostragem (Figura 7): (i) *Max polling*, na qual é retornado o maior valor; (ii) *Average polling*, na qual é retornado a média de todos os valores (BONATTO et al., 2018).



Figura 7 – Subamostragem do valor máximo e da média.

Fonte: Adaptado de Wang et al. (2020).

Por último existem as camadas totalmente conectadas (*fully connected*) responsáveis pela etapa de classificação. Contudo, antes desse processo iniciar, os mapas de características da camada anterior precisam ser convertidos para um vetor unidimensional. Nesta camada, cada neurônio está conectado a todos os outros neurônios da camada posterior. Por isso, a maioria dos parâmetros a serem aprendidos em uma CNN está nesta camada, o que pode ser bastante custoso do ponto de vista computacional (ASSIS, 2019).



Figura 8 – Camada totalmente conectada.

Fonte: Adaptado de Payne (2019).

Hoje em dia existem diversas arquiteturas de CNN's como a AlexNet, GoogLeNet, VGG-16, Inception-v4 e muitas outras. No entanto, a primeira vez que uma CNN foi implementada e testada com sucesso foi em 1998, com a arquitetura LeNet-5 desenvolvida por Yann Lecun e sua equipe. Essa arquitetura (Figura 9), demonstrou que treinar CNN's por meio de gradiente descendente podia levar a resultados promissores para o problema de reconhecimento de dígitos manuscritos (LECUN et al., 1998).



Figura 9 – Arquitetura LeNet-5.

Fonte: Adaptado de LeCun et al. (1998).

2.3 Trabalhos Relacionados

Nesta seção são investigados alguns trabalhos relevantes que fizeram uso de arquiteturas de CNN's para o reconhecimento de sinais de ECG e suas soluções apresentadas.

No trabalho de Marinho (2021), é apresentado um método para a classificação da qualidade de sinais de ECG fazendo uso de uma CNN unidimensional. Seu objetivo é identificar sinais que, durante o seu processo de gravação, apresentaram ruídos ou outras inconsistências que acabaram por comprometer a qualidade da amostra. O *dataset* utilizado foi o BUT QDB, um banco de dados criado pela equipe de cardiologia do Departamento de Engenharia Biomédica da Universidade de Tecnologia de Brno. Este conjunto de dados é formado por três classes referentes a qualidade do sinal (sinais com boa, média e baixa qualidade) e suas amostras são representadas por vetores de uma dimensão já rotulados.

A Figura 10 é uma representação esquemática da arquitetura CNN utilizada por Marinho. Ela é constituída por duas camadas de convolução intercalada com a operação de subamostragem *Max Polling*. Para a etapa de classificação, há duas camadas totalmente conectadas com aplicação da função de ativação *LeakyReLU* e uma camada de saída de três canais na qual a função *softmax* é utilizada para força a rede a representar a probabilidade dos dados pertencerem a cada uma das classes.



Figura 10 – Arquitetura CNN Simplificada. Fonte: Elaborada pelo autor.

Como metodologia de validação da arquitetura, o conjunto de dados foi dividido em dados de treino (80%), validação (10%) e teste (10%). Contudo, não foi exposto à porcentagem de representação de cada uma das classes para esses subconjuntos. Além disso, visando melhorar a generalização da rede, também foi empregue a técnica de validação cruzada durante a fase de treinamento. Os resultados experimentais obtidos conseguiram atingir uma precisão média, baseada no F1-Score, de aproximadamente 92% para os dados de treino. No entanto, a matriz de confusão não foi apresentada para que o desempenho efetivo de cada uma das classes possa ser verificada. Por fim, os autores utilizaram algumas combinações de hiperparâmetros a fim de observar o impacto da sua variação no desempenho da rede neural. A melhor combinação de hiperparâmetros para este modelo foi um *batch size* com tamanho 300, uma taxa de 20% de *dropout* aplicada nas camadas densas, um total de 64 filtros e um kernel de tamanho 8 sendo 4 o número de passos a cada convolução.

Em Santana et al. (2021), os autores propõem uma CNN bidimensional para a classificação de 17 tipos de sinais de ECG. Na Figura 11, é possível visualizar uma imagem de entrada dessa rede. Essa imagem possui um formato de 15 x 15 pixels em escala de cinza, contendo os valores de um sinal de ECG convertida do *dataset MIT-BIH Arrythmia Database*. Para cada uma das 17 classes, foram extraídos 150 registros, totalizando 2550 amostras. Essa estratégia permitiu que todas as classes tivessem a mesma representatividade.



Figura 11 – Sinal de ECG convertida em imagem em escala de cinza. Fonte: Adaptado de Santana et al. (2021).

A CNN-2D utilizada é composta por 20 camadas ao todo. A etapa de extração de características é formada por 5 blocos, cada bloco possuindo 3 camadas referentes ao processo de convolução, normalização em lote e aplicação função de ativação *ReLu*, totalizando 15 camadas. A etapa de classificação é formada por dois blocos, cada bloco possuindo uma camada inteiramente conectada de 200 neurônios seguida de uma camada de *dropout*. Por fim, há uma camada de saída de 17 neurônios referentes as tipos de classes possíveis. A metodologia empregada para o treinamento e teste da rede foi a validação cruzada de cinco pastas, sendo que em cada pasta 80% dos dados foram separados para treinamento e 20% para validação.

Os autores obtiveram resultados notáveis, alcançando uma acurácia de 99,13%, precisão de 92,31% e uma pontuação no F1-Score de 92,68%. Esses resultados foram comparados com outras arquiteturas como uma CNN-1D, LSTM e GRU. Dentre todas essas redes, a CNN-2D destacou-se por possuir os melhores resultados. A CNN-1D teve o pior desempenho, com uma pontuação de 49,85% no F1-Score. Já as redes recorrentes, LSTM e GRU, tiveram resultados intermediários. Embora os resultados tenham sido comparados, não houve descrição das particularidades dessas outras arquiteturas neste estudo.

Diferentemente dos trabalhos anteriores, Messias et al. (2022) propõem uma metodologia comparativa de dois modelos de CNN's unidimensionais. Ambas as arquiteturas são voltadas para a classificação de arritmias seguindo recomendações da Associação para o Avanço de Instrumentação Médica (*AAMI*), que reagrupam os sinais de ECG em quatro grandes classes: Batimento normal (N), Ectópico Supraventricular (S), Ectópico Ventricular (V) e Fusão (F). O primeiro modelo, denominado de DC-CNN, foi baseado nos trabalhos anteriores de Habib, Karmakar e Yearwood (2019) (Figura 12) e Xiang, Lin e Meng (2018).



Figura 12 – Arquitetura CNN baseada em atenção de dois níveis.Fonte: Adaptado de Habib, Karmakar e Yearwood (2019).

Essas arquiteturas são compostas por dois canais paralelos de convolução, sendo que um dos canais é voltado para a extração de informações mais genéricas do sinal de ECG, possuindo apenas duas operações de convolução. Enquanto o outro canal é responsável pela extração de informação mais detalhadas, possuindo seis camadas sequenciais dividas em dois grupos. Cada grupo é responsável pelos seguintes processos: Convolução, *Max Polling* e *Batch Normalization*.

Para a etapa de classificação, tal modelo realiza a concatenação entre esses dois canais fazendo uso de uma camada densa de 100 neurônios.

O segundo modelo proposto por Messias, denominado de CNN baseline, foi baseado no trabalho de Acharya et al. (2017). Este modelo possui apenas um canal de fluxo dos dados e foi empregada para realizar uma análise comparativa de desempenho em relação ao modelo DC-CNN. A CNN baseline possui 4 camadas convolucionais seguidas da normalização em lote dos dados. Também apresenta, a cada duas convoluções, uma operação de subamostragem. Por fim, existem 2 camadas densas e uma camada de saída de 4 canais. Os dados utilizados neste estudo são dados da derivação MLII apresentadas no *MIT-BIH Arrythmia Database*.

Os autores utilizaram da abordagem inter-paciente e dividiram os dados em conjuntos de treinamento, validação e teste. Contudo, como o número de amostras de cada classe não são iguais, também avaliou-se o emprego de técnicas de *data augmentation* para reduzir o desbalanceamento entre as classes. A avaliação desses dois modelos foi divido em dois cenários distintos. O primeiro cenário avaliou o desempenho do modelo DC-CNN em relação ao CNN-Baseline. O segundo cenário avaliou a eficácia do emprego da técnica de aumento sintético de dados no modelo DC-CNN. Os resultados obtidos para o primeiro cenário, concluíram que o modelo DC-CNN superou o modelo CNN-baseline em todas as métricas aplicadas. Já os resultados obtidos para o segundo cenário, permitiram demostrar que a aplicação dos algoritmos de *data augmentation* apresentaram um efeito oposto ao que era esperado. Ao invés da rede melhorar sua generalização, acabou por reduzir os valores das métricas de desempenho da rede.

Por fim, há o trabalho de Rocha (2018b), na qual o *dataset* empregado foi o PhysioNet. Esse *dataset* é composto por mais de 12000 gravações que variam em um intervalo de 9 a 61 segundos. Contudo, uma parte desses dados, cerca de 3600 gravações, não foram utilizados por não estarem disponíveis ao público.

Na etapa de pré-processamento dos dados, o autor realizou a conversão dos sinais para o formato de imagens, as quais foram utilizadas como entrada para a CNN 2D. Para todas as gravações, foram geradas imagens do sinal de ECG em escala de cinza, com uma resolução final de 128x128 píxeis. Considerando que as gravações possuem intervalos variados, todas as imagens geradas foram segmentadas em intervalos de 10 segundos, visando garantir uma maior equidade e qualidade dos dados. O conjunto de dados foi dividido em quatro classes: Ritmo normal, fibrilação atrial (FA), outros ritmos e dados ruidosos. No entanto, é importante ressaltar que essas classes apresentam um notável desbalanceamento, com 60% dos dados correspondendo ao ritmo normal, 9% referentes a pacientes com arritmias do tipo AF, 28% relacionados aos demais tipos de arritmias e os 3% restantes são registros ruidosos. Visando corrigir essa desproporcionalidade, o autor optou por empregar a estratégia de redimensionamento do conjunto de dados.

No que diz respeito à arquitetura adotada, foram utilizadas duas abordagens. No entanto, apenas a segunda abordagem proposta será o foco desta revisão. Essa arquitetura possui como entrada imagens de ECG segmentadas, composta por 16 camadas, ilustradas na Figura 13, sendo

14 camadas referentes à etapa de extração de características, formadas por 7 blocos que seguem as seguintes operações: convolução, função de ativação e *maxpoolling*. Por fim, existem 2 camadas totalmente conectadas para a classificação, que incluem a operação de dropout. Na camada de saída, é empregado o algoritmo softmax para fornecer as probabilidades relacionadas às 4 classes definidas.



Figura 13 – Arquitetura CNN - 2ª Abordagem. Fonte: Adaptado de Rocha (2018b).

Para a obtenção dos resultados, foram realizados testes baseados no método de validação cruzada de 10 folds. Em cada uma das 10 combinações possíveis, 9 folds foram utilizadas para treino e 1 pasta para teste. Além disso, foram realizados quatro ensaios variando o número de dados de entrada, sendo o menor com 200 amostras (50 amostras por classe) e o maior com 1200 amostras (300 amostras por classe). Os melhores resultados foram obtidos usando a combinação das funções de ativação ReLu com a função de otimização Adam Optimizer. O menor conjunto de dados, com 200 amostras, obteve uma acurácia de 82%, AUC (área sob a curva ROC) de 88%, uma AVG F-Measure de 81% e uma F-Measure para a classe AF de 82,8%. Para um conjunto de entrada de 1200 amostras, os resultados foram ligeiramente menores.

3 METODOLOGIA

Neste capítulo, é apresentado o conjunto de dados utilizado neste estudo. Em seguida, são explicados os procedimentos adotados para a estruturação, análise e interpretação dos sinais de ECG. Além disso, são descritas as arquiteturas dos modelos CNN desenvolvidos. Por fim, são detalhados os procedimentos utilizados no treinamento dos modelos, assim como o processo de avaliação do desempenho alcançado.

3.1 Base de dados

O banco de dados de arritmia do MIT-BIH foi o primeiro material de ensaio padrão disponível para a comunidade de pesquisa em geral. Seu objetivo é estimular trabalhos nesse campo, impulsionar avaliações reprodutíveis e comparáveis de diferentes algoritmos, arquiteturas e métodos de processamento de dados.

Ele possui 48 segmentos de meia hora de registros de ECG de dois canais, totalizando 24 horas de registros. Esses registros foram obtidos de 47 indivíduos estudados pelo BIH Arrhythmia Laboratory entre os anos de 1975 e 1979. Desses, 23 indivíduos (anotados na "série de 100") foram escolhidos aleatoriamente de um conjunto de mais de 4.000 fitas Holter. Os outros 25 indivíduos (anotados na "série de 200") foram selecionados especificamente para abranger exemplos de arritmias incomuns, porém relevantes, que não estariam adequadamente representados em uma pequena amostra aleatória. Dos 47 indivíduos, 25 são homens com idades entre 32 e 89 anos e 22 são mulheres com idades entre 23 e 89 anos. Além disso, aproximadamente 60% dos indivíduos eram pacientes internados. É importante que um banco de dados destinado a representar sinais do "mundo real" contenha a maior variedade possível de formas de onda.

Em relação à rotulagem, cada registro teve os prontuários entregues a dois cardiologistas, os quais trabalharam adicionando rótulos de ritmo e qualidade do sinal. Ao final, as anotações dos cardiologistas foram comparadas e um rótulo final foi atribuído. Para as anotações divergentes, o sinal foi reavaliado até que os especialistas chegassem a um consenso (MOODY; MARK, 2001).

3.1.1 Classes dos Sinais de ECG

O MIT-BIH Arrhythmia possui 15 tipos de classes, como pode ser observado na Tabela 1. São 13 classes referentes as arritmias cardíacas, 1 classe referente aos batimentos normais e 1 classe relacionada aos sinais não classificados. No entanto, ao avaliar a quantidade de amostras existentes em cada classe, percebe-se que há um desbalanceamento considerável.

A fim de reduzir essa variação, foi adotado o padrão recomendado pela Association for the Advancement of Medical Instrumentation (AAMI), que é uma organização comprometida

MIT-BIH Arritmia Database

- 01 Normal beat (N)
- 02 Left bundle branch block beat (L)
- 03 Right bundle branch block beat (R)
- 04 Atrial premature beat (A)
- 05 Aberrated atrial premature beat (a)
- 06 Nodal (junctional) premature beat (J)
- 07 Supraventricular premature beat (S)
- 08 Premature ventricular contraction (V)
- 09 Fusion of ventricular and normal beat (F)
- 10 Atrial escape beat (e)
- 11 Nodal (junctional) escape beat (j)
- 12 Ventricular escape beat (E)
- 13 Fusion of paced and normal beat (f)
- 14 Paced beat (/)
- 15 Unclassifiable beat (Q)



Fonte: Elaborado pelo autor

com o avanço e a segurança na utilização da tecnologia médica. Esse padrão propõe agrupar as classes existentes do MIT-BIH Arrhythmia em 5 grandes famílias, como pode ser observado na Tabela 2.

The AAMI heartbeat class	Ν	SVEB	VEB	F	Q
Description	Normal beat	Supraventricular ectopic beat	Ventricular ectopic beat	Fusion beat	Unknown beat
MIT-BIH	Normal beat (N)	Atrial premature beat (A)	Premature ventricular contraction (V)	Fusion of ventricular and normal beat (F)	Paced beat (/)
heartbeat types	Left bundle branch block beat (L)	Aberrated atrial premature beat (a)	Ventricular escape beat (E)		Fusion of paced and normal beat (f)
	Right bundle branch block beat (R) Atrial escape beat (e) Nodal (junctional) premature beat (J)	Nodal (junctional) premature beat (J) Supraventricular premature beat (S)			Unclassifiable beat (Q)

Tabela 2 - Classes de batimentos cardíacos AAMI

Fonte: Adaptado de Luz e Menotti (2011).

3.2 Pré-Processamento dos dados

3.2.1 Remapeamento das classes

Para o remapeamento das classes para o padrão AAMI, foi necessário realizar a segmentação de todos os sinais presentes nas gravações da base de dados MIT-BIH. A etapa inicial do processo de segmentação tem início com a detecção do pico R do complexo QRS. Uma vez detectado, são escolhidos 300 pontos para representar o sinal de ECG, sendo 150 pontos anteriores ao pico R e 150 pontos posteriores a ele. Esses 300 pontos são suficientes para representar um sinal individual de ECG, o qual é o foco desta pesquisa. Além disso, nesta fase também é realizada a atribuição do rótulo correspondente à classe do AAMI à qual o sinal pertence. Na Figura 14, pode ser observado três exemplos de sinais segmentados e seus rótulos.



Figura 14 – Sinais de ECG segmentados. Fonte: Elaborado pelo autor.

Ao todo foram segmentados mais de 100 mil sinais. Contudo, ao realizar uma análise quantitativa das 5 classes resultantes, notava-se que o problema da variação entre as classes persistia. A classe dos batimentos normais ainda representava cerca de 89% de todos os dados. Em segundo lugar estava a classe do batimento ectópico ventricular (VEB) com 7% de representatividade. A classe do batimento ectópico supraventricular (SVEB) seguia em terceiro lugar com cerca de 3%. Por último, estavam as classes de batimento de fusão e batimentos não classificados que juntos não chegam a representar 1% dos dados.

3.2.2 Parametrização dos dados

Ter uma base de dados parametrizada é de extrema importância por várias razões. Primeiramente, uma base parametrizada ajuda a melhorar a capacidade de generalização da rede neural, pois fornece dados relevantes e representativos do problema em questão. Isso evita contratempos como *overfitting* (sobreajuste) ou *underfitting* (subajuste), que podem comprometer a capacidade de generalização. Também permite que a rede neural seja eficiente em termos computacionais, reduzindo o tempo e os recursos necessários para o treinamento e evitando o consumo excessivo de memória ou processamento desnecessário. Por fim, possibilita a criação de conjuntos de treinamento, validação e teste com proporções específicas, permitindo avaliar e comparar diferentes arquiteturas ou algoritmos de rede neural de forma consistente e justa.

Percebendo a importância desse processo, neste trabalho são propostas duas abordagens em relação aos conjuntos de dados. Além das análises das arquiteturas de CNN's para a classificação automática de arritmias, também será investigado o impacto que diferentes parametrizações do *dataset* podem resultar no processo de inferência da rede neural.

3.2.2.1 Dataset desbalanceado

O intuito desta primeira abordagem é de avaliar o impacto que as classes desbalanceadas poderiam resultar nas CNN's. No entanto, esse desequilíbrio foi amenizado para não ser algo extremo, visto que um desbalanceamento exorbitante pode resultar um viés de aprendizado e uma precisão artificialmente inflada, pois a rede neural passa a acertar a classe majoritária com maior frequência. Feito que resulta numa avaliação enganosa, pois o modelo pode obter uma alta acurácia ao prever consistentemente a classe majoritária, mas seu desempenho real em todas as classes pode ser ruim. Para isso, as classes F e Q não foram utilizadas devido à sua baixa representatividade. Em outras palavras, este estudo trabalhou apenas com as classes {N, S, V}.

Para a classe N, realizou-se uma amostragem aleatória visando reduzir seu tamanho e promover uma maior equalização entre as classes. A partir dessa amostragem aleatória, a classe N foi reduzida de aproximadamente 90.000 para 8.000 sinais. A classe S permaneceu inalterada, por ser a classe com menor representatividade dentre as três. Por último, a classe V também passou por uma amostragem aleatória, resultando em uma redução de 7.000 para 5.000 sinais. A distribuição final dos dados entre as classes e os conjuntos de treinamento, validação e teste pode ser visualizada na Tabela 3.

CLASSES/PACOTES	TREINO	VALIDAÇÃO	TESTE	TOTAL
Ν	3500	1900	2600	8000
S	1770	190	810	2770
V	2400	800	1800	5000
TOTAL	7670	2890	5210	15770

Tabela 3 – Distribuição dos dados por classes e pacotes

Fonte: Elaborado pelo autor.

A definição das proporções de cada classe nos conjuntos de treinamento, validação e teste não foi fixa. A estratégia adotada baseou-se na quantidade de amostras das classes N e V em relação à classe S, a qual é a classe com menor representatividade. Quanto maior fosse a quantidade de amostras em relação à classe S, maior seria a redistribuição dos dados nos conjuntos de teste e validação. Essa distribuição pode ser visualizada com mais facilidade na Tabela 4.

CLASSES/PACOTES	TREINO	VALIDAÇÃO	TESTE	TOTAL
N	44%	23%	33%	100%
S	64%	7%	29%	100%
V	48%	16%	36%	100%

Tabela 4 – Porcentagem de distribuição dos dados por pacotes

Fonte: Elaborado pelo autor.

3.2.2.2 Dataset balanceado

A segunda abordagem visa avaliar o desempenho das arquiteturas CNN quando treinadas a partir de um conjunto de dados em que todas as classes possuem a mesma representatividade. Para isso, foi necessário redimensionar as classes N e V para terem a mesma quantidade de amostras da classe S. A estratégia utilizada para o redimensionamento foi a mesma empregada na primeira abordagem, que foi uma amostragem randômica para as classes N e V até que as mesmas apresentassem um total de 2770 amostras. A distribuição percentual dos dados em cada pacote pode ser observada na Tabela 5, na qual 56,1% dos dados de cada classe são para o conjunto de treinamento, 8,7% para validação e 35,2% para o conjunto de teste.

Não há uma regra rígida sobre as porcentagens exatas a serem usadas, mas existem algumas diretrizes que ajudam a determinar esses valores, como: (i) o conjunto de treinamento deve ser o maior, pois é aonde o modelo aprende os dados e uma proporção comum é alocar de 60-80% dos dados. (ii) o conjunto de validação é usado para o ajuste dos hiperparâmetros e alocar de 10-20% dos dados nele é uma prática usual. (iii) o conjunto de teste é usado para avaliar o desempenho do modelo após o treinamento e a porcentagem restante é atribuída a ele, normalmente de 10-20% dos dados. Contudo, essas diretrizes são apenas orientações gerais e o mais importante é garantir que você tenha conjuntos de dados representativos para cada fase

PACOTES/CLASSES	TREINO (56,1%)	VALIDAÇÃO (8,7%)	TESTE (35,2%)	TOTAL(100%)
N	1555	242	973	2770
S	1555	242	973	2770
\mathbf{V}	1555	242	973	2770
TOTAL	4665	726	2920	

(treinamento, validação e teste) e que o desempenho do modelo seja avaliado de maneira justa e robusta.

Tabela 5 – Distribuição padronizada dos dados.

Fonte: Elaborado pelo autor.

3.2.3 Normalização dos dados

A normalização de dados é um procedimento para transformar os dados em uma escala comum, de forma que todos os valores estejam em um determinado intervalo. Sua importância para a rede neural está no fato de que a escala de dados pode afetar o desempenho do treinamento, pois valores muito grandes ou muito pequenos que podem dificultar a convergência da rede (NISSILA, 2023).

Existem várias técnicas de normalização, sendo as mais comuns a normalização min-max e a normalização por desvio padrão. Neste trabalho, foi utilizado a normalização min-max para a CNN-1D, na qual os dados são dimensionados para o intervalo entre 0 e 1. Para a CNN-2D e Híbrida, foi utilizada a técnica L2. A normalização tipo L2 refere-se a uma abordagem na qual as amostras de dados são ajustadas de forma que a norma L2 (norma euclidiana) do vetor seja igual a 1. Isso significa que cada amostra de dados é dividida pelo seu comprimento Euclidiano, o que resulta em uma representação em que a soma dos quadrados dos valores é igual a 1.

3.3 Arquiteturas dos modelos CNN's

Nesta seção, serão aprofundadas algumas características e técnicas inerentes às arquiteturas CNN's, tais como o *padding*, funções de ativação, normalização de lote, *dropout* e *flattening*/aplanamento. Em seguida, serão apresentadas e detalhadas as três arquiteturas de redes neurais convolucionais propostas neste estudo: unidimensional, bidimensional e uma abordagem híbrida com dois canais em paralelo para a etapa de extração de características.

3.3.1 Características

As técnicas abaixo visam aprimorar o desempenho da rede neural, evitar o *overfitting* e tornar o processo de aprendizado mais eficiente diante de uma ampla variedade de dados.

• *Padding*: O *padding* é uma técnica que lida com o tamanho dos mapas de características (*feature maps*) resultantes após a aplicação das operações de convolução. Quando utilizado,

ele evita a diminuição dos mapas de características após a convolução, preenchendo bordas à imagem para manter o tamanho dos mapas de características após a convolução.

 Funções de Ativação: A função de ativação desempenha um papel importante, especialmente nas camadas ocultas. Ela introduz não-linearidades na rede, permitindo que esta aprenda relações complexas e representações não-lineares dos dados de entrada. Neste estudo, durante a etapa de ajuste dos hiperparâmetros, várias funções de ativação foram testadas em todas as camadas, com exceção da camada de saída, na qual foi aplicada a função softmax. Entre as funções testadas, a função ReLU demonstrou o melhor desempenho em comparação com as demais.

ReLU (*Rectified Linear Unit*): Essa função mapeia qualquer valor de entrada negativo para zero e deixa passar os valores positivos diretamente. Essa função foi amplamente utilizada nas arquiteturas propostas, tanto nas camadas de extração de características quanto nas camadas ocultas de classificação, seu comportamento pode ser observado na Figura 15.



Figura 15 – Função de Ativação ReLU Fonte: Adaptado de Camacho (2018).

Softmax: A função Softmax foi empregada somente na camada de saída, pois ela é capaz de normalizar um conjunto de valores de entrada, transformando-os em uma distribuição de probabilidade, onde cada valor de saída representa a probabilidade estimada de uma classe específica. A soma de todas as saídas softmax é igual a 1.

Normalização em lote ou *Batch Normalization*: O objetivo dessa técnica é normalizar as ativações de cada camada durante o treinamento. Isso significa que, para cada conjunto de dados (lote ou mini lote) que passa pela rede, as ativações de cada camada são ajustadas de modo que sua média seja próxima de zero e sua variância, seja próxima de um. Neste trabalho, esse processo é aplicado apenas após a primeira camada densa de cada arquitetura proposta

Dropout: É uma técnica que desativa aleatoriamente um conjunto de neurônios durante o treinamento, sendo definida uma probabilidade de desativação para cada neurônio. Durante o treinamento, a rede é treinada em diferentes subconjuntos de neurônios desativados a cada iteração, forçando a rede a ser mais robusta e a evitar a dependência excessiva de neurônios específicos para a classificação dos dados. O funcionamento dessa técnica é demonstrado na Figura 16.



Figura 16 – Ilustração da técnica de *Dropout*. Fonte: Adaptado de Roffo (2017).

• Aplanamento ou *Flattening*: O *flattening* é uma operação usada para transformar as representações multidimensionais dos dados em um vetor unidimensional. Essa técnica é comumente usada para transformar imagens 2D ou tensores 3D em vetores 1D antes de passá-los para as camadas densas (totalmente conectadas) da rede neural.

3.3.2 Arquitetura CNN 1D

O primeiro modelo foi desenvolvido sob uma perspectiva unidimensional, em que os dados de entrada são representados como vetores de uma dimensão de 300 unidades que representam um sinal segmentado de EGC com valores previamente normalizados. Como pode ser observado na Figura 17, a parte relacionada a extração de características é composta por 3 blocos em sequência. Sendo cada bloco formado por uma operação de convolução seguida da operação de subamostragem, conhecida também por maxpooling. De acordo com os trabalhos relacionados discutidos e outros estudos como da Hannun et al. (2019), o uso de 3 camadas convolucionais é suficiente para a tarefa de classificação de sinas de ECG. Para a etapa de classificação, há 2 camadas ocultadas totalmente conectadas e uma camada de saída de 3 canais.

A Tabela 7 apresenta, de maneira mais detalhada, as configurações do modelo, incluindo os tamanhos dos kernels de convolução e subamostragem, além da estrutura das camadas densas, com a inclusão das operações de normalização em lote e *dropout*, com uma taxa de aplicação de 30%.



Figura 17 – Representação esquemática da arquitetura CNN-1D.

	TIPO DE	FREATURE	TAMANHO	Nº KEDNEI S
LSIKUIUKA	CAMADA	MAP	KERNEL	IN NENINELS
	Entrada	300	-	-
	Convolucional 1D	271	30	16
Extração	MaxPooling	135	2	16
Extração	Convolucional 1D	121	15	32
ue Comostamísticos	MaxPooling	60	2	32
Características	Convolucional 1D	53	8	64
	MaxPooling	26	2	64
	Flattening	1664	-	-
	TIPO DE	Nº DE		
LSIKUIUKA	CAMADA	NEURÔNIOS		
	Densa	80		
Classificação	BatchNormalization	-		
	Dropout (0.3)	-		
	Densa	50		
	Saída	3		

Fonte: Elaborado pelo autor.

Tabela 6 – Estrutura do modelo CNN-1D.

Fonte: Elaborado pelo autor.

3.3.3 Arquitetura CNN 2D

A segunda arquitetura foi baseada numa abordagem bidimensional, mesma abordagem adota por Santana et al. (2021). Essa técnica visa criar novas características relacionadas a dependência espacial dos dados, ao invés de somente as características temporais presentes

no espaço unidimensional da arquitetura anterior. Para isso, primeiramente, foi necessário redimensionar os dados, modificando-os de vetores de tamanho 300 para matrizes no formato 15x20.

A título de exemplo, é possível visualizar essas matrizes em forma de imagens na Figura 18. Para isso, os dados dessa grade matricial tiveram seus valores normalizados para o intervalo de 0-255, somente para fins de demonstração. Dessa forma, cada ponto se comporta como um pixel.



Figura 18 – Representação alternativa das ondas de ECG. Fonte: Elaborado pelo autor.

Embora não exista um padrão rígido para o tamanho de entrada de imagens em redes neurais, o tamanho 15x20 resultante da reconstrução do vetor mostra-se deficiente mesmo para o problema de classificação simples. Dependendo da aplicação e da arquitetura da rede, os tamanhos mais comuns de entrada de imagens são 32x32, 64x64, 128x128, etc. Para este trabalho, foi adotado o tamanho 32x32 como padrão, visando uma melhor detecção das características espaciais.

Para ser possível redimensionar as imagens de 15x20 para 32x32 píxeis (Figura 19), foi empregado o método da interpolação do vizinho mais próximo. Essa técnica envolve a atribuição do valor do píxel mais próximo da posição estimada na imagem original para o píxel na nova posição na imagem redimensionada. Esse método não envolve cálculos complexos e é bastante direto.



Figura 19 – Imagem de um sinal de ECG - 32x32 píxeis. Fonte: Elaborado pelo autor.

Conforme demonstrado na Figura 20, a parte relacionada a extração de características compreende 3 operações de convolução e 2 operações de subamostragem, alternadas com as convoluções. A etapa de classificação é composta por duas camadas ocultas completamente conectadas e uma camada de saída com três canais.



Figura 20 – Representação esquemática da arquitetura CNN-2D. Fonte: Elaborado pelo autor.

A Tabela 8 apresenta de forma mais detalhada algumas configurações, incluindo o tamanho dos kernels de convolução e de subamostragem, bem como a estrutura das camadas densas com a incorporação das operações de normalização em lote e *dropout* com uma taxa de aplicação também de 30%.

ESTDUTUDA	TIPO DE	FREATURE	TAMANHO	Nº VEDNEI S
ESIKUIUKA	CAMADA	MAP	KERNEL	IN KENINELS
	Entrada	32x32	-	-
	Convolucional 2D	28x28	5x5	64
Extração	MaxPooling	14x14	2x2	64
de	Convolucional 2D	06x12	9x3	64
Características	MaxPooling	03x06	2x2	64
	Convolucional 2D	03x06	3x2	128
	Flattening	2304	-	-
	TIPO DE	Nº DE		
ESTRUTURA	CAMADA	NEURÔNIOS		
	Densa	100		
Classificação	BatchNormalization	-		
	Dropout (0.3)	-		
	Densa	30		
	Saída	3		

Tabela 7 – Estrutura CNN-2D.

Fonte: Elaborado pelo autor.

3.3.4 Arquitetura CNN Híbrida

A arquitetura CNN Híbrida, foi baseada nos trabalhos anteriores de Santana et al. (2021) e Habib, Karmakar e Yearwood (2019), na qual foi empregado dois canais em paralelo de convolução (Figuras 21 e 22) na qual os autores chamaram de atenção em dois níveis,



Figura 21 – Representação esquemática canal de convolução 2D. Fonte: Elaborado pelo autor.



Figura 22 – Representação esquemática canal de convolução 1D. Fonte: Elaborado pelo autor.

O diferencial dessa arquitetura é que um canal ficou responsável pela extração das características espaciais, enquanto o outro canal foi encarregado pela extração das características temporais dos sinais de ECG. A saída desses canais, ao final do processo de convolução, são concatenadas para formar um único vetor para servir de entrada para as camadas densas (Figura 23). A estrutura de classificação da CNN Híbrida é composta por duas camadas densas ocultas de 180 e 120 neurônios e uma camada de saída de 3 canais.



Figura 23 – Representação esquemática camada densa. Fonte: Elaborado pelo autor.

Na Tabela 9, é apresentado os detalhes das configurações do modelo, incluindo o tamanho dos kernels de convolução e de subamostragem, dimensão do mapa de característica, quantidade de filtros, bem como a estrutura das camadas densas com a incorporação as operações de

	TIPO DE	FREATURE	TAMANHO	
ESIKUIUKA	CAMADA	MAP	KERNEL	N KERNELS
	Entrada	1024	-	-
	MaxPooling	341	3	-
	Convolucional 1D	312	30	16
CANAL	MaxPooling	156	2	16
	Convolucional 1D	139	18	32
1D	MaxPooling	69	2	32
	Convolucional 1D	59	11	64
	MaxPooling	29	2	64
	Flattening	1856	-	-
	Entrada	32x32	-	-
	Convolucional 2D	28x28	5x5	32
CANAL	MaxPooling	14x14	2x2	32
	Convolucional 2D	6x12	9x3	64
2D	MaxPooling	3x6	2x2	64
	Convolucional 1D	3x6	3x6	128
	Flattening	2304	-	-
	TIPO DE	Nº DE		
ESIKUIUKA	CAMADA	NEURÔNIOS		
	Densa	180		
	BatchNormalization	-		
Classificação	Dropout (0.3)	-		
	Densa	120		
	Densa	3		

normalização em lote e aplicação do *dropout* a uma taxa de 0.3% ocorrem apenas após a primeira camada densa (180 neurônios).

Tabela 8 – Estrutura CNN-Híbrida.

Fonte: Elaborado pelo autor.

3.3.5 Ambiente de desenvolvimento

Todas as arquiteturas foram construídas por meio da linguagem de programação Python, utilizando os *frameworks* Tensorflow e Keras. O desenvolvimento ocorreu no ambiente do Google Colab, uma plataforma baseada em nuvem, com o auxílio de aceleradores de GPU para otimizar o processo computacional.

3.4 Metodologia de Treinamento e Validação

A definição de hiperparâmetros em uma rede neural não é algo trivial, sendo uma tarefa desafiadora. Eles são parâmetros que não são aprendidos durante o processo de treinamento da rede, mas que precisam ser definidos antes do início do treinamento. Para encontrar uma configuração adequada de hiperparâmetros, foi realizado uma busca exaustiva mediante várias

combinações de valores possíveis para os hiperparâmetros. Além disso, análises cuidadosas dos resultados do treinamento foram realizadas para compreender como cada configuração afeta o desempenho da rede.

3.4.1 Taxa de aprendizado

Um desses hiperparâmetros é a taxa de aprendizado, também conhecida como *"learning rate"* em inglês. Ela controla o tamanho dos passos dados durante o treinamento para ajustar os pesos do modelo. Uma taxa alta acelera o treinamento, mas pode levar à instabilidade. Uma taxa baixa é mais precisa, mas pode tornar o treinamento lento.

A técnica utilizada neste trabalho para definição da taxa de aprendizado foi a redução gradual, também conhecida como "*step decay*". Nessa técnica, a taxa de aprendizado é reduzida a cada certo número de épocas ao longo do treinamento. O objetivo é permitir que a taxa comece alta para acelerar a convergência inicial e, à medida que o treinamento avança, reduza gradualmente para permitir ajustes mais precisos aos mínimos locais da função de perda. A taxa inicial utilizada foi de 0,00005 e seu decaimento não segue um padrão linear, como demonstrado na Figura 24. Esse padrão foi aplicado para o treinamento das três arquiteturas apresentadas.

No entanto, antes de atingir esse valor, foram conduzidos ensaios experimentais com valores maiores. O resultado foi que o treinamento não apresentou uma boa convergência e a rede não foi capaz de aprender. Portanto, visando alcançar de maneira mais eficaz os mínimos locais, a taxa inicial foi reduzida para o valor apresentado.



Figura 24 – Decaimento da taxa de aprendizado Fonte: Elaborado pelo autor.

3.4.2 Método de regularização de penalidades

Essa técnica adiciona termos de penalização ao cálculo da função de perda durante o treinamento da rede. Neste trabalho, foi empregada somente a penalização L2 que adiciona uma penalidade proporcional à soma dos quadrados dos pesos. Após diversos ajustes durante os ensaios, o valor do parâmetro que controla a força da penalidade foi estabelecido em 0.001, sendo aplicado em todas as camadas densas ocultas em todos os modelos.

3.4.3 Inicialização dos pesos

Refere-se a um parâmetro que permite definir a forma como os pesos do kernel (ou filtros) de uma camada serão inicializados. Os pesos do kernel são parâmetros que a camada de convolução aprende durante o treinamento para detectar padrões específicos nos dados de entrada. A inicialização desses pesos é crucial, pois determina os valores iniciais antes do início do treinamento, podendo impactar significativamente o desempenho e a velocidade de convergência do modelo.

Neste estudo, os pesos foram inicializados usando uma técnica conhecida como inicialização de Xavier ou *glorot uniform*. A ideia por trás dessa inicialização é ajustar os pesos do kernel de forma a manter a variância constante em todas as camadas da rede. Essa abordagem visa evitar problemas como o desvanecimento ou a explosão do gradiente durante o treinamento.

3.4.4 Tamanho do lote ou *batch size*

O termo *batch size* refere-se ao número de amostras de treinamento que são usadas em cada iteração. Durante o treinamento, os dados são divididos em lotes (*batches*) e o modelo é atualizado com base no erro médio calculado para as amostras dentro de cada lote. *Batch sizes* muito pequenos podem aumentar a variância nas atualizações dos pesos, enquanto *batch sizes* muito grandes podem exigir mais memória e desacelerar o processo de treinamento.

Após conduzir uma série de testes experimentais em todos os modelos, com o *batch size* variando entre 32, 64, 80, 128 e 256, determinou-se um tamanho de lote (*batch size*) de 80. Esse tamanho proporcionou um desempenho superior durante o processo de treinamento das redes.

3.5 Treinamento

Para o treinamento das redes neurais, foi utilizado o algoritmo de otimização Adam (*Adaptive Moment Estimation*). Esse algoritmo é uma combinação dos conceitos do algoritmo RMSProp com o método de momento (momentum). A ideia principal por trás do Adam é ajustar a taxa de aprendizado para cada parâmetro individualmente com base na média móvel dos gradientes do passado e suas segundas ordens. Dessa forma, ele pode se adaptar automaticamente a diferentes taxas de aprendizado para diferentes parâmetros, tornando-se bastante robusto para

a escolha dos hiperparâmetros. Embora o parâmetro relativo ao *learning rate* às vezes precise ser diferente do padrão convencional (KINGMA; BA, 2014).

Outro parâmetro é a função de custo, conhecida também como função de perda. Ela uma métrica que quantifica o quão bem as previsões da rede se aproximam dos valores reais dos dados de treinamento. O objetivo é minimizar essa função de custo, ajustando os parâmetros da rede (pesos e vieses) para melhorar a precisão das previsões. Neste estudo a função de perda empregada foi a entropia cruzada categórica (*Categorical Cross-Entropy*) que é uma função de custo frequentemente usada em problemas de classificação multiclasse, onde cada exemplo de treinamento pertence a uma única classe dentre várias possíveis. O algoritmo de otimização Adam é usado para ajustar os parâmetros da rede de modo a reduzir essa perda. Dessa forma, a medida que o treinamento progride, espera-se que o valor da perda diminua, indicando que a rede está melhorando suas previsões em relação aos dados de treinamento. No entanto, o valor absoluto da perda pode variar dependendo do problema, do conjunto de dados e da arquitetura da rede.

Por fim, a estratégia adotada durante o processo de treinamento da rede, foi de utilizar funções do tipo *callbacks* para acompanhar métricas de desempenho durante o treinamento, como a precisão ou a perda do modelo em cada época/iteração. Toda vez que o modelo atingia o melhor resultado da acurácia em relação aos dados de validação, todo o modelo era salvo juntamente com os seus pesos. Dessa forma, era possível restaurar o modelo com o seu melhor desempenho em vez de utilizar o modelo final após todas as épocas. Além disso, foi estabelecido um limite máximo de 200 épocas para o treinamento, contudo o treinamento podia ser interrompido antecipadamente caso as métricas de desempenho estabilizassem.

3.6 Métricas de avaliação

Nesta seção, serão apresentados e discutidos os conceitos de validação e teste. Além dos indicadores utilizados na avaliação de desempenho e eficácia das arquiteturas CNN's propostas.

3.6.1 Conjunto de validação

Como já mencionado neste trabalho, os dados foram divididos em 3 grupos: conjuntos de treino, validação e teste. Ao final do processo de treinamento de uma rede neural, espera-se que o modelo tenha aprendido a fazer previsões precisas e acuradas em relação aos dados de treinamento e, idealmente, também em relação a novos dados não vistos durante o treinamento (dados de teste e dados de validação).

Diferentemente do teste da rede neural, que ocorre após o processo de treinamento do modelo, o processo de validação é feito durante o treinamento, sendo de extrema importância para garantir que o modelo CNN tenha um bom desempenho em dados que não foram usados no treinamento. Os dados de validação permitem avaliar a capacidade de generalização do modelo,

pois se o desempenho nos dados de validação piorar enquanto o desempenho nos dados de treinamento melhora, isso pode indicar overfitting. Ademais, a validação também permite tomar decisões visando a otimização de hiperparâmetros e mudanças nas estratégias de treinamento para obtenção de melhores resultados.

3.6.2 Indicadores de desempenho

Ao final do processo de classificação, cada sinal de ECG possui suas probabilidades de pertencerem a cada uma das três possíveis classes (N, V, S) devido ao emprego da função *softmax* na última camada da rede. A partir dessas probabilidades, o sinal recebe o rótulo da classe que apresentar a maior probabilidade. Ao final desse processo, tanto as informações das classes verdadeiras quanto das classes preditas estão à disposição, sendo possível atribuir os seguintes indicadores:

Acúracia: Esse indicador mensura a proporção de previsões corretas em relação ao total de previsões feitas, considerando os verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN). Sua fórmula é dada pela Equação 3.1. Essa é a métrica mais básica e amplamente utilizada para problemas de classificação binária e multiclasse.

$$Acuracia = \frac{VP + VN}{VP + FN + VN + FP}$$
(3.1)

Precisão: A precisão indica a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo (Equação 3.2). Esse indicador é relevante quando o foco é na minimização dos falsos positivos, ou seja, reduzir o número de casos em que o modelo classifica erroneamente uma amostra como positiva quando, na verdade, é negativa.

$$Precisao = \frac{VP}{VP + FP} \tag{3.2}$$

Recall ou Sensibilidade: A sensibilidade representa a proporção de exemplos positivos corretamente identificados em relação ao total de exemplos positivos no conjunto de teste (Equação 3.3). Esse indicador é relevante quando o foco é na minimização dos falsos negativos.

$$Recall = \frac{VP}{VP + FN}$$
(3.3)

F1 Score: Esse indicador combina a precisão e o *recall* em um único valor. Ele é especialmente útil quando há um desequilíbrio entre as classes em um problema de classificação, ou seja, quando uma classe é muito mais frequente do que a outra. O F1 Score varia de 0 a 1, onde 0 representa o pior desempenho e 1 representa o melhor desempenho. Ele fornece uma medida equilibrada do desempenho do modelo, considerando tanto a capacidade de evitar falsos

positivos (alta precisão) quanto a capacidade de evitar falsos negativos (alto *recall*), sua fórmula é dada pela Equação 3.4.

$$F1Score = 2 * \frac{Recall * Precisao}{Recall + Precisao}$$
(3.4)

3.6.2.1 Matriz de Confusão

A matriz de confusão é uma tabela usada para avaliar o desempenho de um modelo de classificação. Essa matriz exibe as contagens das previsões realizadas pelo modelo para cada classe, comparadas aos rótulos verdadeiros das amostras presentes no conjunto de teste. Essa dinâmica é ilustrada na Figura 25.



CLASSIFICAÇÃO DO MODELO

Figura 25 – Matriz de confusão para classificação binária - Exemplo Fonte: Adaptado de Ruback, Avila e Cantero (2021).

A importância da matriz de confusão reside no fato de que ela oferece uma visão detalhada do desempenho do modelo em cada classe, permitindo a análise das métricas já mencionadas. Além disso, ela ajuda a identificar os principais erros do modelo, fornecendo informações valiosas para melhorar seu desempenho, ajustar limiares de decisão, reequilibrar classes, selecionar métricas mais adequadas e, em geral, aprimorar o modelo para que ele se comporte de maneira mais eficaz em diferentes situações.

4 EXPERIMENTOS E RESULTADOS

Neste capítulo, apresentam-se os resultados das experiências para cada uma das arquiteturas CNN propostas, considerando os dois tipos de conjuntos de dados (balanceado e desbalanceado). O primeiro objetivo é avaliar o impacto que as diferentes arquiteturas CNN podem exercer no desempenho da rede para um mesmo conjunto de dados. A outra finalidade consiste em investigar as variações de desempenho que a utilização de um conjunto de dados balanceado e desbalanceado pode gerar em uma mesma arquitetura CNN. Por fim, serão realizadas comparações e discussões dos resultados obtidos.

4.1 Arquitetura CNN 1D

4.1.1 *Dataset* desbalanceado

Nesta seção, são apresentados os resultados da arquitetura CNN-1D aplicada ao conjunto de dados desbalanceado. Os valores de perda e acurácia para os conjuntos de treino e validação, bem como as métricas de avaliação referentes aos dados de teste, são detalhados nas Tabelas 10 e 11, respectivamente.

DADOS/INDICADORES	PERDA	ACURÁCIA
TREINO	0.88	0.67
VALIDAÇÃO	0.85	0.73
TESTE	0.91	0.68

Tabela 9 – Relação de perda e acurácia CNN-1D para dados desbalanceados.

Fonte: Elaborado pelo autor.

CLASSES	PRECISÃO	RECALL	F1-SCORE	DADOS DE TESTE
Ν	0.66	0.79	0.72	2600
S	0.37	0.08	0.13	810
\mathbf{V}	0.72	0.78	0.75	1800
ACURÁCIA TOTAL CONJUNTO DE TESTE			0.68	

Tabela 10 – Resultados de performance do teste da CNN-1D para dados desbalanceados.

Fonte: Elaborado pelo autor.

Na Figura 26, encontra-se a representação da matriz de confusão, permitindo uma investigação mais detalhada das previsões e equívocos do modelo.



Figura 26 – Matriz de confusão da CNN-1D para dados desbalanceados. Fonte: Elaborado pelo autor.

O gráfico na Figura 27 ilustra o progresso do desempenho da rede neural durante o processo de treinamento, empregando os conjuntos de dados de treinamento e validação. Esse gráfico oferece uma representação visual nítida da evolução do modelo à medida que é aprimorado por meio das épocas de treinamento. Conforme previamente mencionado, os dados de validação são instrumentais para detectar indícios de *overfitting* no modelo.



Figura 27 – Processo de treinamento CNN-1D para dados desbalanceados. Fonte: Elaborado pelo autor.

4.1.2 Dataset balanceado

Nesta seção, são apresentados os resultados da arquitetura CNN-1D aplicada ao conjunto de dados balanceado. Os valores de perda e acurácia para os conjuntos de treino, validação

e teste, bem como as métricas de avaliação referentes aos dados de teste, são detalhadas nas Tabelas 12 e 13, respectivamente.

DADOS/INDICADORES	PERDA	ACURÁCIA
TREINO	0.90	0.64
VALIDAÇÃO	1.01	0.58
TESTE	0.99	0.60

Tabela 11 – Relação de perda e acurácia CNN-1D para dados balanceados.

Fonte: Elaborado pelo autor.

CLASSES	PRECISÃO	RECALL	F1-SCORE	DADOS DE TESTE
N	0.56	0.44	0.49	973
S	0.53	0.61	0.57	973
V	0.73	0.77	0.75	973
ACURÁCIA TOTAL CONJUNTO DE TESTE				0.60

Tabela 12 - Resultados de performance do teste da CNN-1D para dados balanceados.

Fonte: Elaborado pelo autor.

Na Figura 28, encontra-se a representação da matriz de confusão.



Figura 28 – Matriz de confusão da CNN-1D para dados balanceados. Fonte: Elaborado pelo autor.

O gráfico na Figura 29 ilustra o progresso do desempenho da rede neural durante o processo de treinamento, empregando os conjuntos de dados de treinamento e validação.



Figura 29 – Processo de treinamento CNN-1D para dados balanceados. Fonte: Elaborado pelo autor.

4.2 Arquitetura CNN 2D

4.2.1 Dataset desbalanceado

Nesta seção, são apresentados os resultados da arquitetura CNN-2D aplicada ao conjunto de dados desbalanceado. Os valores de perda e acurácia para os conjuntos de treino e validação, bem como as métricas de avaliação referentes aos dados de teste, são detalhados nas Tabelas 14 e 15, respectivamente.

DADOS/INDICADORES	PERDA	ACURÁCIA
TREINO	0.82	0.71
VALIDAÇÃO	0.79	0.71
TESTE	0.90	0.65

Tabela 13 – Relação de perda e acurácia CNN-2D para dados desbalanceados.

CLASSES	PRECISÃO	RECALL	F1-SCORE	DADOS DE TESTE
Ν	0.62	0.83	0.71	2600
S	0.22	0.01	0.03	810
V	0.72	0.68	0.70	1800
ACURÁCIA TOTAL CONJUNTO DE TESTE				0.65

Fonte: Elaborado pelo autor.

Tabela 14 – Resultados de desempenho do teste da CNN-2D para dados desbalanceados. Fonte: Elaborado nalo autor

Fonte: Elaborado pelo autor.

Na Figura 30, encontra-se a representação da matriz de confusão, permitindo uma investigação mais detalhada das previsões e equívocos do modelo.



Figura 30 – Matriz de confusão CNN-2D para dados desbalanceado. Fonte: Elaborado pelo autor.

O gráfico na Figura 31 ilustra o progresso do desempenho da rede neural durante o processo de treinamento, empregando os conjuntos de dados de treinamento e validação. Esse gráfico oferece uma representação visual nítida da evolução do modelo à medida que é aprimorado por meio das épocas de treinamento. Conforme previamente mencionado, os dados de validação são instrumentais para detectar indícios de *overfitting* no modelo.



Figura 31 – Processo de treinamento CNN-2D para dados desbalanceados Fonte: Elaborado pelo autor.

4.2.2 *Dataset* balanceado

Nesta seção, são apresentados os resultados da arquitetura CNN-2D aplicada ao conjunto de dados balanceado. Os valores de perda e acurácia para os conjuntos de treino, validação e teste, bem como as métricas de avaliação referentes aos dados de teste, são detalhadas nas Tabelas 16 e 17, respectivamente.

DADOS/INDICADORES	PERDA	ACURÁCIA
TREINO	0.94	0.77
VALIDAÇÃO	1.03	0.58
TESTE	0.98	0.62

Tabela 15 – Relação de perda e acurácia CNN-2D para dados balanceados.

CLASSES	PRECISÃO	RECALL	F1-SCORE	DADOS DE TESTE
Ν	0.57	0.47	0.51	973
S	0.58	0.77	0.66	973
\mathbf{V}	0.71	0.61	0.66	973
ACURÁCIA TOTAL CONJUNTO DE TESTE				0.62

Fonte: Elaborado pelo autor.

Tabela 16 – Resultados de desempenho do teste da CNN-2D para dados balanceados.

Fonte: Elaborado pelo autor.



Matriz de Confusão

Figura 32 – Matriz de confusão CNN-2D para conjunto de dados balanceado. Fonte: Elaborado pelo autor.

O gráfico na Figura 33 ilustra o progresso do desempenho da rede neural durante o processo de treinamento, empregando os conjuntos de dados de treinamento e validação.



Figura 33 – Processo de treinamento CNN-2D para dados balanceados. Fonte: Elaborado pelo autor.

4.3 Arquitetura CNN Híbrida

4.3.1 Dataset desbalanceado

Nesta seção, são apresentados os resultados da arquitetura CNN-Híbrida aplicada ao conjunto de dados desbalanceado. Os valores de perda e acurácia para os conjuntos de treino e validação, bem como as métricas de avaliação referentes aos dados de teste, são detalhados nas Tabelas 18 e 19, respectivamente.

DADOS/INDICADORES	PERDA	ACURÁCIA
TREINO	0.49	0.86
VALIDAÇÃO	0.92	0.69
TESTE	0.97	0.66

Tabela 17 – Relação de perda e acurácia CNN-Híbrida para dados desbalanceados.

CLASSES	PRECISÃO	RECALL	F1-SCORE	DADOS DE TESTE
Ν	0.69	0.81	0.75	2600
S	0.29	0.28	0.29	810
\mathbf{V}	0.82	0.62	0.71	1800
ACURÁCIA TOTAL CONJUNTO DE TESTE				0.66

Fonte: Elaborado pelo autor.

Tabela 18 – Resultados de desempenho do teste da CNN-Híbrida para dados desbalanceados.

Fonte: Elaborado pelo autor.

Na Figura 34, encontra-se a representação da matriz de confusão, permitindo uma investigação mais detalhada das previsões e equívocos do modelo.



Figura 34 - Matriz de confusão CNN-Híbrida para dados desbalanceados.

Fonte: Elaborado pelo autor.

O gráfico na Figura 35 ilustra o progresso do desempenho da rede neural durante o processo de treinamento, empregando os conjuntos de dados de treinamento e validação. Esse gráfico oferece uma representação visual nítida da evolução do modelo à medida que é aprimorado por meio das épocas de treinamento.



Figura 35 – Processo de treinamento CNN-Híbrida para dados desbalanceados Fonte: Elaborado pelo autor.

4.3.2 *Dataset* balanceado

Nesta seção, são apresentados os resultados da arquitetura CNN-Híbrida aplicada ao conjunto de dados balanceado. Os valores de perda e acurácia para os conjuntos de treino e validação, bem como as métricas de avaliação referentes aos dados de teste, são detalhadas nas Tabelas 20 e 21, respectivamente.

DADOS/INDICADORES	PERDA	ACURÁCIA
TREINO	0.56	0.83
VALIDAÇÃO	1.01	0.62
TESTE	0.95	0.63

Tabela 19 – Relação de perda e acurácia CNN-Híbrida para dados balanceados.

CLASSES	PRECISÃO	RECALL	F1-SCORE	DADOS DE TESTE
Ν	0.66	0.44	0.53	973
S	0.50	0.76	0.60	973
\mathbf{V}	0.86	0.69	0.77	973
ACURÁCIA TOTAL CONJUNTO DE TESTE				0.63

Fonte: Elaborado pelo autor.

Tabela 20 - Resultados de desempenho do teste CNN-Híbrida para dados balanceados.

Fonte: Elaborado pelo autor.

Na Figura 36, encontra-se a representação da matriz de confusão, permitindo uma investigação mais detalhada das previsões e equívocos do modelo.



Figura 36 – Matriz de confusão CNN-Híbrida para dados balanceado.

Fonte: Elaborado pelo autor.

O gráfico na Figura 36 ilustra o progresso do desempenho da rede neural durante o processo de treinamento, empregando os conjuntos de dados de treinamento e validação.



Figura 37 – Processo de treinamento CNN-Híbrida para dados balanceados Fonte: Elaborado pelo autor.

4.4 Discussão

Na Tabela 22, são apresentadas as comparações de desempenho individuais de cada modelo nos experimentos executados. O indicador F1-Score foi selecionado devido à sua capacidade de unificar a precisão e o *recall* em um único valor.

ARQUITETURA	CONJUNTO DE DADOS	F1-SCORE			
		CLASSES			ACURÁCIA
		Ν	S	V	
CNN-1D	Desbalanceado	0.72	0.13	0.75	0.68
	Balanceado	0.49	0.57	0.75	0.60
CNN-2D	Desbalanceado	0.71	0.03	0.70	0.65
	Balanceado	0.51	0.66	0.66	0.62
CNN-HÍBRIDA	Desbalanceado	0.75	0.29	0.71	0.66
	Balanceado	0.53	0.60	0.77	0.63

Tabela 21 – Variações do F1-Score entre as arquiteturas e conjuntos de dados utilizados. Fonte: Elaborado pelo autor. A análise da tabela em relação ao desempenho das redes neurais diante da utilização do conjunto de dados desbalanceado revela um padrão recorrente em todas as arquiteturas. Grupos com maior volume de dados (classes N e V) apresentaram pontuações consideravelmente superiores em relação à classe com menor número de amostras (classe S). Isso ocorre devido ao viés enraizado na rede durante seu treinamento. Quando uma classe possui um número maior de amostras em comparação com outras classes, a rede neural tem mais exposição e oportunidades para aprender os padrões e características específicas dessa classe dominante. Como resultado, quando a rede encontra um novo exemplo dessa classe durante o processo de inferência, é provável que ela faça uma previsão precisa. No entanto, essa vantagem pode ser prejudicial quando se trata de classes minoritárias, pois a rede neural pode não ter tido exposição suficiente para aprender adequadamente os padrões dessas classes.

Dentro desse contexto, a arquitetura que apresentou o desempenho mais desfavorável em relação à classificação entre as classes foi a CNN-2D. Nesta arquitetura, a classe S obteve uma pontuação de 3% no F1-Score, enquanto as classes N e V alcançaram 71% e 70%, respectivamente. A arquitetura CNN Híbrida se destacou como o modelo de melhor desempenho, obtendo as seguintes pontuações: 75% para a classe N, 29% para a classe S e 71% para a classe V. Se apenas a acurácia for levada em consideração, a arquitetura CNN-1D se destaca por alcançar uma taxa de 68%. No entanto, é importante notar que sua pontuação no F1-Score para a classe S foi de apenas 13% (intermediária).

Considerando as arquiteturas treinadas a partir do conjunto de dados balanceado, é possível notar um maior equilíbrio nos índices de classificação entre as classes. Tal comportamento se manifesta porque o balanceamento proporciona à rede neural oportunidades equitativas para aprender com cada classe, permitindo-lhe desenvolver uma compreensão mais abrangente e precisa das características de todas as classes envolvidas.

Nessa perspectiva, a CNN Híbrida novamente se sobressai ao alcançar o melhor desempenho, obtendo uma pontuação de 53% para a classe N, 60% para a classe S, 77% para a classe V e uma acurácia total de 63%, a maior dentre os três modelos apresentados. Era esperado que esta arquitetura, por ser uma fusão das outras duas, alcançasse um desempenho superior. No entanto, seu desempenho foi apenas ligeiramente superior em relação às outras abordagens. As arquiteturas CNN-1D e CNN-2D também demonstraram desempenho satisfatório, mesmo com um número substancialmente menor de parâmetros.

5 CONCLUSÃO

O presente trabalho propôs uma avaliação das arquiteturas CNN's comumente empregadas na literatura, juntamente com uma abordagem alternativa fundamentada na estratégia de atenção de dois níveis, todas elas considerando um *dataset* devidamente parametrizado. Essa perspectiva muitas vezes não é o foco em outras pesquisas acadêmicas, as quais se concentram principalmente na arquitetura e no treinamento da rede neural, deixando de lado a importância do pré-processamento e qualidade dos dados.

Outro ponto relevante deste estudo é a análise do impacto da utilização de conjuntos de dados balanceados e desbalanceados no treinamento da rede neural. Isso levanta questionamentos sobre até que ponto a desproporcionalidade dos dados é viável e quais técnicas que poderiam ser empregadas em conjunto para alcançar melhores resultados. Analisando os resultados obtidos, pode-se concluir que todas as arquiteturas mostraram-se eficazes e conseguiram cumprir com o objetivo, contudo a CNN híbrida obteve resultados ligeiramente melhores do que as demais redes neurais.

Como trabalho futuro, uma abordagem adicional relacionada aos dados de entrada pode ser considerada. Em vez de utilizar um único sinal de ECG segmentado como entrada para a rede neural, uma alternativa viável é selecionar um segmento mais extenso que englobe dois ou mais sinais de ECG, tendo em vista que as arritmias cardíacas podem se manifestar em sequência. Outra linha de pesquisa potencial pode envolver a aplicação de estratégias como o ajuste de pesos conforme a representatividade de cada classe, aumento sintético de dados (*data augmentation*) para a classe S, bem como a validação cruzada para permitir uma avaliação abrangente do modelo em diferentes subconjuntos de dados.

Por último, a intenção é que esta pesquisa contribua para a evolução de diferentes metodologias e estratégias capazes de classificar arritmias cardíacas a partir de sinais de ECG. Dessa maneira, tais recursos poderiam ser aplicados por profissionais de saúde, fornecendo suporte na realização de diagnósticos.

REFERÊNCIAS

ACHARYA, U. R. et al. Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. *Information sciences*, 2017. Elsevier, v. 405, p. 81–90, 2017. Citado na página 29.

ALMEIDA, A. C. da S. Ampliação da base minds-libras: um estudo de aplicação de técnicas de aumento sintético de dados e da inclusão de novos conjuntos de vídeos disponíveis na literatura. 2022. p. 7, 2022. Citado na página 24.

ASSIS, M. M. d. *Aplicação de Deep Learning no reconhecimento de sinais de libras: Aspextos técnicos e sociais.* 63 p., 2019. Engenharia de Sistemas. Citado 2 vezes nas páginas 21 e 25.

AZURE. Deep learning vs. machine learning in azure machine learning. *Azure Machine Learning Documentation*, 2022. Microsoft, v. 1, p. 1630, 2022. Citado na página 23.

BONATTO, L. V. M. et al. Avaliação de arquiteturas de hardware para acelerar as operações aritméticas da camada convolucional em redes neurais convolucionais. 2018. 2018. Citado na página 25.

BORGES, R. R. et al. Sincronização de disparos em redes neuronais com plasticidade sináptica. *Revista Brasileira de Ensino de Física*, 2015. SciELO Brasil, v. 37, p. 2310–1, 2015. Citado na página 22.

CAMACHO, C. *Convolutional Neural Networks*. 2018. Disponível em: https://cezannec-.github.io/Convolutional_Neural_Networks/. Citado na página 37.

CASTRO, G. Z. *Reconhecimento de Línguas de Sinais Utilizando Redes Neurais Convolucionais e Transferência de Aprendizado*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Minas Gerais - UFMG, Dezembro 2020. Citado 2 vezes nas páginas 23 e 24.

CHEN, C. et al. Low-power fpga implementation of convolution neural network accelerator for pulse waveform classification. *Algorithms*, 2020. Multidisciplinary Digital Publishing Institute, v. 1, p. 17, 2020. Citado na página 16.

FERNANDES, L. S. et al. Conhecimento teórico-prático de enfermeiras sobre eletrocardiograma. *Revista Baiana de Enfermagem*, 2015. v. 29, n. 2, 2015. Citado na página 15.

GARCÍA, E. G. et al. Arritmias cardiacas en urgencias. *Medicine-Programa de Formación Médica Continuada Acreditado*, 2015. Elsevier, v. 11, n. 87, p. 5175–5184, 2015. Citado na página 19.

GIFFONI, R. T.; TORRES, R. M. Breve história da eletrocardiografia. *Revista Médica de Minas Gerais*, 2010. v. 20, n. 2, p. 263–270, 2010. Citado na página 18.

GODOI, W. C. Detecc ao de defeitos em isoladores polim ericos por radiografia digital. 2005. 2005. Citado na página 23.

GOMES, C. S. et al. Fatores associados às doenças cardiovasculares na população adulta brasileira: Pesquisa nacional de saúde, 2019. *Revista Brasileira de Epidemiologia*, 2021. SciELO Brasil, v. 24, 2021. Citado na página 15.

GOMES, D. d. S. Inteligência artificial: conceitos e aplicações. *Olhar Científico. v1*, 2010. n. 2, p. 234–246, 2010. Citado na página 21.

HABIB, A.; KARMAKAR, C.; YEARWOOD, J. Impact of ecg dataset diversity on generalization of cnn model for detecting qrs complex. *IEEE access*, 2019. IEEE, v. 7, p. 93275–93285, 2019. Citado 2 vezes nas páginas 28 e 42.

HANNUN, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 2019. Nature Publishing Group US New York, v. 25, n. 1, p. 65–69, 2019. Citado 2 vezes nas páginas 16 e 38.

KAUFMAN, D. A inteligência artificial irá suplantar a inteligência humana? [S.l.]: ESTAÇÃO DAS LETRAS E CORES EDI, 2019. Citado na página 20.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2014. Citado na página 47.

KOROL, G. d. S. *An FPGA Implementation for Convolutional Neural Network*. 71 p., 2019. Engenharia da Computação. Citado 2 vezes nas páginas 21 e 23.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998. Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado 2 vezes nas páginas 25 e 26.

LOPES, R. Martha Gabriel, especialista em inteligência artificial: "Em 30 anos, a humanidade será uma nova coisa, não mais o que a gente é". 2022. Urlhttps://gauchazh.clicrbs.com.br/tecnologia/noticia/2022/05/martha-gabriel-especialistaem-inteligencia-artificial-em-30-anos-a-humanidade-sera-uma-nova-coisa-nao-mais-o-que-agente-e-cl333qra6002j019ilocaur1h.html. Citado na página 21.

LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, 2021. SciELO Brasil, v. 35, p. 85–94, 2021. Citado na página 21.

LUZ, E.; MENOTTI, D. An x-ray on methods aiming at arrhythmia classification in ecg signals. In: THE STEERING COMMITTEE OF THE WORLD CONGRESS IN COMPUTER SCIENCE, COMPUTER Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP). [S.1.], 2011. p. 1. Citado na página 32.

LUZ, E. J. d. S. et al. Ecg-based heartbeat classification for arrhythmia detection: A survey. *Computer methods and programs in biomedicine*, 2016. Elsevier, v. 127, p. 144–164, 2016. Citado na página 16.

MARINHO, A. S. Classificação automática da qualidade de sinal de ecg com redes neurais convolucionais. 2021. 2021. Citado na página 26.

MESSIAS, B. C. et al. Rede neural convolucional com dois canais para classificação automática de arritmias em sinais de ecg. 2022. Florianópolis, SC, 2022. Citado na página 28.

MOODY, G. B.; MARK, R. G. The impact of the mit-bih arrhythmia database. *IEEE engineering in medicine and biology magazine*, 2001. IEEE, v. 20, n. 3, p. 45–50, 2001. Citado na página 31.

NISSILA, L. A. Análise comparativa de métodos computacionais para a classificação de estoque: redes neurais, knn e svm. 2023. 2023. Citado na página 36.

ONU. *OMS revela principais causas de morte e incapacidade em todo o mundo entre 2000 e 2019.* 2020. Url: https://brasil.un.org/pt-br/104646-oms-revela-principais-causas-de-morte-e-incapacidade-em-todo-o-mundo-entre-2000-e-2019. Acesso em: 18 mai. 2022. Citado na página 15.

PAYNE, J. A Beginner's Guide to Convolutional Neural Networks. 2019. Urlhttps://medium.datadriveninvestor.com/a-beginners-guide-to-convolutional-neuralnetworks-49384c75d1a. Citado na página 25.

REGIS, C. D. M.; CALDEIRA, L. G.; GURJÃO, E. C. Avaliação da amostragem compressiva em sinais de ecg e imagens digitais. *Revista Principia-Divulgação Científica e Tecnológica do IFPB*, 2016. v. 1, n. 29, p. 95, 2016. Citado na página 19.

RIENSTRA, M. et al. Symptoms and functional status of patients with atrial fibrillation: state of the art and future research opportunities. *Circulation*, 2012. Am Heart Assoc, v. 125, n. 23, p. 2933–2943, 2012. Citado na página 15.

ROCHA, G. M. d. *Deteção de arritmias cardíacas em eletrocardiogramas usando deep learning*. Tese (Doutorado), 2018. Citado 2 vezes nas páginas 16 e 24.

ROCHA, G. M. da. *Deteção de arritmias cardíacas em eletrocardiogramas usando deep learning*. Tese (Doutorado) — Instituto Politecnico do Porto (Portugal), 2018. Citado 2 vezes nas páginas 29 e 30.

ROFFO, G. Ranking to learn and learning to rank: On the role of ranking in pattern recognition applications. *arXiv preprint arXiv:1706.05933*, 2017. 2017. Citado na página 38.

RUBACK, L.; AVILA, S.; CANTERO, L. Vieses no aprendizado de máquina e suas implicações sociais: Um estudo de caso no reconhecimento facial. In: SBC. *Anais do II Workshop sobre as Implicações da Computação na Sociedade*. [S.1.], 2021. p. 90–101. Citado na página 49.

SANTANA, J. R. G. et al. Classificação de arritmias cardíacas em sinais de ecg utilizando redes neurais profundas. 2021. Universidade Federal do Amazonas, 2021. Citado 6 vezes nas páginas 15, 16, 22, 27, 39 e 42.

SCHONS, T. *Rede de Convoluçao para Sistema Biométrico baseado em EEG*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DE OURO PRETO, 2018. Citado 2 vezes nas páginas 21 e 24.

SEIF, J. *I'll tell you why Deep Learning is so popular and in demand.* 2018. Urlhttps://medium.com/swlh/ill-tell-you-why-deep-learning-is-so-popular-and-in-demand-5aca72628780. Citado na página 22.

SOUZA, J. W. M. d.; FILHO, P. P. R. Desenvolvimento de sistema de detecção arritmia cardíaca em sinais de eletrocardiograma utilizando métodos de reconhecimento de padrões e extrator de sinal baseados na transformada de fourier. 2019. 2019. Citado na página 19.

SZPALHER, A. S.; BATALHA, M. C. Arritmias cardíacas: Diagnósticos de enfermagem baseados na taxonomia da nanda-i (2018-2020). *Revista Eletrônica Acervo Saúde*, 2019. v. 11, n. 17, p. e1447–e1447, 2019. Citado na página 15.

WALLER, A. D. A demonstration on man of electromotive changes accompanying the heart's beat. *The Journal of physiology*, 1887. Wiley-Blackwell, v. 8, n. 5, p. 229, 1887. Citado na página 18.

WANG, X. et al. Intelligent detection and recovery of missing electric load data based on cascaded convolutional autoencoders. *Scientific Programming*, 2020. Hindawi, v. 2020, 2020. Citado 2 vezes nas páginas 24 e 25.

XIANG, Y.; LIN, Z.; MENG, J. Automatic qrs complex detection using two-level convolutional neural network. *Biomedical engineering online*, 2018. BioMed Central, v. 17, n. 1, p. 1–17, 2018. Citado na página 28.

APÊNDICE A – REPOSITÓRIO

Todos os códigos e documentação referente à implementação deste estudo podem ser acessados pelo link abaixo:

Link: TCC - Arthur C. S. Almeida - Eng^a Controle e Automação - UFOP¹.

Dentro deste repositório, é possível encontrar documentos, imagens da fase de construção e treinamento das arquiteturas CNN's, arquiteturas ótimas, bem como técnicas de processamento de dados utilizadas. Isso possibilita uma visão mais completa e abrangente do projeto em sua totalidade.

https://drive.google.com/drive/folders/luIrnq2z0noe_u-m8grpzgx0_HH13f2Az? usp=sharing