



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE GRADUAÇÃO EM ENGENHARIA ELETRÔNICA

Bruno Cayres Messias

**REDE NEURAL CONVOLUCIONAL COM DOIS CANAIS PARA CLASSIFICAÇÃO
AUTOMÁTICA DE ARRITMIAS EM SINAIS DE ECG**

Florianópolis, Santa Catarina – Brasil
2022

Bruno Cayres Messias

**REDE NEURAL CONVOLUCIONAL COM DOIS CANAIS PARA CLASSIFICAÇÃO
AUTOMÁTICA DE ARRITMIAS EM SINAIS DE ECG**

Trabalho de Conclusão de Curso submetida ao Programa de Graduação em Engenharia Eletrônica da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Engenharia Eletrônica.
Orientador: Prof. Danilo Silva, Dr.

Florianópolis, Santa Catarina – Brasil
2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Messias, Bruno Cayres
REDE NEURAL CONVOLUCIONAL COM DOIS CANAIS PARA
CLASSIFICAÇÃO AUTOMÁTICA DE ARRITMIAS EM SINAIS DE ECG /
Bruno Cayres Messias ; orientador, Danilo Silva, 2022.
67 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Eletrônica, Florianópolis, 2022.

Inclui referências.

1. Engenharia Eletrônica. 2. ECG. 3. Machine Learning.
4. Arritmia. 5. Classificação. I. Silva, Danilo. II.
Universidade Federal de Santa Catarina. Graduação em
Engenharia Eletrônica. III. Título.

Bruno Cayres Messias

**REDE NEURAL CONVOLUCIONAL COM DOIS CANAIS PARA CLASSIFICAÇÃO
AUTOMÁTICA DE ARRITMIAS EM SINAIS DE ECG**

Este(a) Trabalho de Conclusão de Curso foi julgado adequado(a) para obtenção do Título de Bacharel em Engenharia Eletrônica, e foi aprovado em sua forma final pelo Programa de Graduação em Engenharia Eletrônica do Departamento de Engenharia Elétrica e Eletrônica da Universidade Federal de Santa Catarina.

Florianópolis, Santa Catarina – Brasil, 21 de Março de 2022.

Prof. Fernando Rangel de Sousa, Dr.
Coordenador do Programa de Graduação
em Engenharia Eletrônica

Banca Examinadora:

Prof. Danilo Silva, Dr.
Orientador
Universidade Federal de Santa Catarina –
UFSC

Prof^a. Daniela Ota Hisayasu Suzuki, Dra.
Universidade Federal de Santa Catarina –
UFSC

Prof. Hector Pettenghi Roldan, Dr.
Universidade Federal de Santa Catarina –
UFSC

Aos meus pais, Júlio César e Helzerli, dedico-lhes este trabalho por tudo que compartilharam comigo durante esta jornada. O que eles me deram em amor, entrega e proteção, contribuíram para a minha capacidade intelectual e emotiva, fundamental para concluir este trabalho acadêmico.

AGRADECIMENTOS

Inicialmente, não posso concluir este TCC sem expressar meus sinceros agradecimentos ao meu Orientador Prof. Dr. Danilo Silva a quem, além de orientar este trabalho com dedicação e interesse, contribuiu com seu conhecimento e experiência que me fizeram refletir sobre a validade das minhas conclusões em todas as etapas deste projeto.

Agradeço, também, a todos os professores e colegas de curso que, com seus conhecimentos e experiências, colaboraram para o meu desenvolvimento acadêmico, necessário para concluir este trabalho com êxito.

*“...se não soubermos medir as coisas sobre
as quais estamos falando de maneira a
associar números às nossas afirmações, é
porque o nosso conhecimento ainda é
precário e de má qualidade”.
(William Thomson - "Lorde Kelvin", 1867)*

RESUMO

As Redes Neurais Convolucionais (CNN) são modelos de referência na classificação de imagens clínicas, no entanto, uma nova abordagem de análise de sinais biomédicos de uma dimensão, como os sinais de eletrocardiogramas (ECG), tem se beneficiado de arquiteturas capazes de abstrair melhor os dados em busca de padrões relevantes. Este trabalho desenvolveu arquiteturas CNN com um e dois canais com diferentes níveis de abstração de dados para a classificação de arritmias. A partir da abordagem inter-paciente, executou-se o pré-processamento dos dados de sinais de ECG do *MIT-BIH Arrhythmia Database*, treinou-se os modelos em cenários distintos, com ou sem *data augmentation* e efetuou-se a avaliação comparativa do desempenho de um modelo CNN com dois canais (DC-CNN) em relação ao modelo CNN de referência e outros descritos a literatura. Adotou-se a abordagem inter-paciente por considerá-la mais adequada no diagnóstico de arritmias em cenários da vida real, uma vez que reproduz um ambiente similar ao do cardiologista quando o mesmo realiza o diagnóstico de um novo paciente. Ao investigar a eficácia do uso de algoritmos SMOTE e *Random Over-Sampling* no modelo DC-CNN, verificou-se que ambos apresentaram uma degradação no desempenho, em termos de acurácia, sendo que esta degradação foi maior para o algoritmo SMOTE. No entanto, considerando que a similaridade entre as classes de batimento S e N é um dos principais desafios para os classificadores de arritmia, o modelo DC-CNN com SMOTE apresentou o melhor desempenho geral na detecção da classe S de arritmia em todas as métricas avaliadas. De uma forma geral, o emprego de algoritmos *data augmentation* são instrumentos eficientes na otimização do desempenho de classificadores, no entanto, deve-se atentar para uma exaustiva análise dos efeitos dessas algoritmos na detecção das classes minoritárias antes de incorporá-las aos modelos CNN classificadores de arritmia. Em relação a avaliação do desempenho do modelo DC-CNN empregando uma classificação binária, ou seja, batimentos normal (classe N) e arritmico (classes S, V e F) nas curvas ROC e PR, a arquitetura DC-CNN obteve AUROC de 0,85 e AUPRC de 0,67, um desempenho 6,25% e 39,6% superior ao modelo de referência na detecção das classes de arritmias, respectivamente. Os experimentos executados neste trabalho mostraram que o modelo DC-CNN superou o modelo de referência em todas as métricas aplicadas, com destaque para a métrica de acurácia (85,66%), resultado semelhante ao de modelos CNN empregando a abordagem inter-paciente. Assim, infere-se que o modelo DC-CNN avaliado tem potencial para efetuar a detecção de arritmias morfológicas em sinais de ECG, por meio da análise do complexo QRS.

Palavras-chave: ECG. *Machine Learning*. *Deep Learning*. Arritmia. Classificação. MIT-BIH. *Data Augmentation*. CNN. Inter-paciente.

ABSTRACT

Convolutional Neural Networks (CNN) are benchmark models in clinical image classification, however, a new approach to analyzing one-dimensional biomedical signals, such as electrocardiogram (ECG) signals, has benefited from architectures capable of better abstracting the data in search of relevant patterns. This work developed single and dual-channel CNN architectures with different levels of data abstraction for arrhythmia classification. Using the inter-patient paradigm, ECG signal data from the MIT-BIH Arrhythmia Database were preprocessed, models were trained in different scenarios, with or without data augmentation, and the performance of a two-channel CNN model (DC-CNN) was compared with the reference CNN model and others described in the literature. The inter-patient paradigm was adopted because it is considered to be more appropriate in the diagnosis of arrhythmias in real-life scenarios, since it reproduces an environment similar to that of a cardiologist when diagnosing a new patient. When investigating the effectiveness of using SMOTE and Random Over-Sampling algorithms in the DC-CNN model, it was found that both showed a degradation in performance, in terms of accuracy, with this degradation being greater for the SMOTE algorithm. However, considering that the similarity between S-type and N-type heartbeat constitutes one of the main challenges for arrhythmia classifiers, the DC-CNN model with SMOTE showed the best overall performance in detecting the S-type arrhythmia class in all metrics evaluated. In general, the use of data augmentation algorithms are efficient tools in optimizing the performance of classifiers, however, one should pay attention to a thorough analysis of the effects of these algorithms in the detection of minority classes before incorporating them into the CNN arrhythmia classifier models. In relation to evaluating the performance of the DC-CNN model employing a binary classification, i.e. normal (class N) and arrhythmic heartbeats (classes S, V and F) on the ROC and PR curves, the DC-CNN architecture obtained AUROC of 0.85 and AUPRC of 0.67, a performance 6.25% and 39.6% higher than the reference model in detecting the arrhythmia classes, respectively. The experiments performed in this work showed that the DC-CNN model outperformed the reference model in all the metrics applied, especially in the accuracy metric (85.66%), a result similar to that of CNN models employing the inter-patient paradigm. Thus, we infer that the evaluated DC-CNN model has the potential to perform morphological arrhythmia detection in ECG signals through the analysis of the QRS complex.

Keywords: ECG. Machine Learning. Deep Learning. Arrhythmia. Classification. MIT-BIH. Data augmentation. CNN. Inter-patient.

LISTA DE FIGURAS

Figura 1	– Estrutura de uma unidade neural	19
Figura 2	– Relação entre capacidade e erro do modelo	21
Figura 3	– Exemplo de arquitetura CNN multicanal	23
Figura 4	– Estrutura de um sinal de ECG normal	24
Figura 5	– Amostras de sinais de ECG contendo diferentes categorias de arritmias cardíacas	26
Figura 6	– Exemplo de gravação de ECG extraída do MITDB	27
Figura 7	– Comparação entre as técnicas de <i>undersampling</i> e <i>oversampling</i>	28
Figura 8	– Exemplo de uma curva ROC	32
Figura 9	– Exemplo de uma curva PR	33
Figura 10	– Arquitetura proposta por Xiang, Lin e Meng (2018)	34
Figura 11	– Arquitetura proposta por Habib, Karmakar e Yearwood (2019)	35
Figura 12	– Resultados em diferentes bancos de dados do modelo CNN proposto por Habib, Karmakar e Yearwood (2019)	36
Figura 13	– Arquitetura proposta por Nagata, Candido e Silva (2021)	37
Figura 14	– Etapas da metodologia empregada	42
Figura 15	– Curvas de acurácia e perda obtidas no treinamento	43
Figura 16	– Matriz de confusão dos modelos DC-CNN e CNN <i>baseline</i> sem <i>data augmentation</i>	44
Figura 17	– Curvas ROC dos modelos DC-CNN e CNN <i>baseline</i>	45
Figura 18	– Curvas PR dos modelos DC-CNN e CNN <i>baseline</i>	46
Figura 19	– Curvas de acurácia e perda do modelo DC-CNN obtidas no treinamento	46
Figura 20	– Matriz de confusão do modelo DC-CNN com <i>data augmentation</i>	47
Figura 21	– Comparação das curvas ROC do modelo DC-CNN com <i>data augmentation</i>	49
Figura 22	– Comparação das curvas PR do modelo DC-CNN com <i>data augmentation</i>	49
Figura 23	– Arquitetura do modelo DC-CNN	64

LISTA DE TABELAS

Tabela 1	– Mapeamento das anotações do MITDB em relação as classes da AAMI	25
Tabela 2	– Matriz de confusão para problemas de classificação binária . . .	29
Tabela 3	– Métricas de avaliação baseada na matriz de confusão binária . .	30
Tabela 4	– Matriz de confusão para problemas de classificação multiclasse .	30
Tabela 5	– Métricas de avaliação baseada na matriz de confusão multiclasse	31
Tabela 6	– Número de amostras por classe de arritmia segundo a AAMI . . .	38
Tabela 7	– Relação das gravações dos sinais de ECG por conjunto separado	38
Tabela 8	– Total de amostras categorizadas por conjunto de dados	39
Tabela 9	– Estrutura do modelo DC-CNN	40
Tabela 10	– Estrutura do modelo CNN <i>baseline</i>	40
Tabela 11	– Desempenho geral dos modelos DC-CNN e CNN <i>baseline</i>	43
Tabela 12	– Desempenho dos modelos DC-CNN e CNN <i>baseline</i> por classe de arritmia	44
Tabela 13	– Desempenho geral do modelo DC-CNN com e sem <i>data augmentation</i>	47
Tabela 14	– Desempenho do modelo DC-CNN, com e sem <i>data augmentation</i> , por classe de arritmia	48
Tabela 15	– Resumo dos trabalhos relacionados sobre modelos CNN classificadores de arritmias	50
Tabela 16	– Comparação do desempenho entre o modelo DC-CNN e o de Nagata, Candido e Silva (2021), por classe de arritmia	52
Tabela 17	– Comparação entre o modelo proposto por Habib, Karmakar e Yearwood (2019) e o modelo DC-CNN	65
Tabela 18	– Comparação entre o modelo proposto por Acharya <i>et al.</i> (2017) e o modelo CNN <i>baseline</i>	65
Tabela 19	– Desempenho dos modelos DC-CNN e CNN <i>baseline</i> por classe de arritmia, incluindo valores de TP, TN, FP e FN	66
Tabela 20	– Desempenho do modelo DC-CNN com e sem <i>data augmentation</i> por classe de arritmia, incluindo valores de TP, TN, FP e FN . . .	66

LISTA DE ABREVIATURAS E SIGLAS

AAMI	Associação para o Avanço da Instrumentação Médica
ACC	Acurácia
AI	Inteligência Artificial
AUPRC	<i>Area Under PR Curve</i>
AUROC	<i>Area Under ROC Curve</i>
BIH	<i>Arrhythmia Laboratory of Boston's Beth Israel Hospital</i>
CAP	Contração Atrial Prematura
CNN	Redes Neurais Convolucionais
DC-CNN	Redes Neurais Convolucionais com Dois Canais
DL	<i>Deep Learning</i>
ECG	eletrocardiograma
F1	<i>F1-Score</i>
FN	Falso Negativo
FP	Falso Positivo
FPR	Taxa de Falsos Positivos
INCART	<i>St. Petersburg Institute of Cardiological Technics</i>
KNN	<i>K-Nearest Neighbors</i>
LSTM	<i>Long-Short Term Memory</i>
MITDB	<i>MIT-BIH Arrhythmia Database</i>
ML	<i>Machine Learning</i>
MLII	<i>Modified Limb II</i>
OMS	Organização Mundial da Saúde
PR	<i>Precision-Recall</i>
PRE	Precisão
PTB	<i>Physikalisch-Technische Bundesanstalt</i>
PVC	Contração Ventricular Prematura
QT	<i>European Society of Cardiology ST-T Database</i>
ResNet	<i>Neural Residual Network</i>
RNN	Redes Neurais Recorrentes
ROC	<i>Receiver Operating Characteristic</i>
ROS	<i>Random Over-Sampling</i>
RRI	<i>Inter-Heartbeat Intervals</i>
SEN	<i>Recall</i> ou Sensibilidade
SMOTE	<i>Synthetic Minority Over-Sampling Technique</i>
TN	Verdadeiro Negativo
TP	Verdadeiro Positivo
TPR	Taxa de Verdadeiros Positivos

LISTA DE SÍMBOLOS

a^l	Saída da função de ativação na camada l
y_k	Vetor de saída do modelo
g_l	Função de ativação na camada l
$z^{[l]}$	Somatório dos pesos de cada entrada da camada l
$W^{[l]}$	Matriz de pesos dados a entrada na camada l
$b^{[l]}$	Vetor de <i>bias</i> na camada l
L	Última camada linear do modelo
J	Função custo do modelo
k	Número de unidades neurais para cada camada
θ	Vetor com os parâmetros W e b do modelo
m	Total de saídas do modelo
$L(y^i, \hat{y}^{(i)})$	Perda média entre o vetor real e o previsto pelo modelo
$s(t)$	Saída da função de convolução
$w(t - a)$	Função peso aplicada em todo instante t

SUMÁRIO

	Sumário	13
1	INTRODUÇÃO	15
1.1	OBJETIVOS GERAL E ESPECÍFICOS	16
1.2	TRABALHOS RELACIONADOS	16
1.3	ORGANIZAÇÃO DO TRABALHO	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	REDES NEURAIS	19
2.2	MODELOS <i>DEEP LEARNING</i>	21
2.3	REDES NEURAIS CONVOLUCIONAIS (CNN)	22
2.4	SINAIS DE ELETROCARDIOGRAMAS (ECG)	24
2.5	MIT-BIH <i>ARRHYTHMIA DATABASE</i> (MITDB)	26
2.6	<i>DATA AUGMENTATION</i>	28
2.7	AVALIAÇÃO DO DESEMPENHO DE CLASSIFICADORES	29
2.8	<i>RECEIVER OPERATING CHARACTERISTICS</i> (ROC) E <i>PRECISION RECALL</i> (PR)	31
3	REVISÃO DA LITERATURA	34
4	METODOLOGIA	38
4.1	PRÉ-PROCESSAMENTO	38
4.2	ARQUITETURAS DOS MODELOS CNN	39
4.3	TREINAMENTO DOS MODELOS CNN	41
4.4	AVALIAÇÃO DO DESEMPENHO DOS MODELOS CNN	41
4.5	ESTRUTURA DA METODOLOGIA	42
5	RESULTADOS	43
5.1	EXPERIMENTO 1: AVALIAÇÃO DO DESEMPENHO DOS MODELOS CNN	43
5.2	EXPERIMENTO 2: <i>DATA AUGMENTATION</i> COM SMOTE E ROS	46
5.3	DISCUSSÃO DOS RESULTADOS	49
6	CONCLUSÃO	55
	REFERÊNCIAS	57
	ANEXO A – RECURSOS COMPUTACIONAIS	63

ANEXO B – ARQUITETURA DO MODELO DC-CNN DESENVOLVIDO	64
ANEXO C – COMPARAÇÃO DAS ESTRUTURAS DOS MODELOS CNN DESENVOLVIDOS	65
ANEXO D – DESEMPENHO DOS MODELOS CNN POR CLASSES DE ARRITMIA	66

1 INTRODUÇÃO

Segundo a Organização Mundial da Saúde (OMS), doenças cardíacas são a principal causa de morte dos últimos 20 anos, passando de 2 milhões de mortes decorrentes de doenças cardiovasculares em 2000 para quase 9 milhões em 2019, totalizando 16% de todas as mortes mundiais (OMS, 2020). O diagnóstico das arritmias cardíacas baseado na análise de padrões de registro da atividade elétrica do coração em sinais captados por eletrocardiograma (ECG) podem prever doenças cardiovasculares.

Existem diversos tipos de arritmias associadas a um padrão de sinal elétrico específico, categorizados em dois grupos: morfológicas, quando o batimento é irregular, ou rítmicas, quando associadas a um ritmo irregular (S. LUZ *et al.*, 2016, p. 144). Neste contexto, diversos estudos estão sendo realizados para desenvolver sistemas eficientes para o diagnóstico de arritmias. Assim será possível diminuir os erros inerentes à inspeção visual realizado pelo cardiologista e reduzir o tempo de diagnóstico (S. LUZ *et al.*, 2016, p. 144–145).

Além do mais, com o advento de novas tecnologias *wearables*, sensores portáteis de coleta de dados de saúde ganharam mais atenção no monitoramento do estado de saúde dos usuários, devido ao design e a miniaturização de tais sensores que continuamente adquirem dados fisiológicos (VEERAVALLI; DEEPU; NGO, 2017, p. 404).

Com a constante e rápida evolução tecnológica na área de Inteligência Artificial (AI), modelos utilizando *Machine Learning* (ML) têm ganhado popularidade nos últimos anos, pois melhoram o diagnóstico de arritmias em sinais de ECG. Dentre as arquiteturas computacionais de ML, as de *Deep Learning* (DL) são destaque na literatura científica. Tais arquiteturas apresentam múltiplos níveis de abstração e são utilizadas nas atividades de detecção, otimização e classificação de dados. Assim sendo, os modelos DL amplificam os aspectos relevantes dos dados de entrada e reduzem a influência de variações de padrão inexpressivas (LECUN; BENGIO; HINTON, 2015, p. 436 v. 521).

Dentre as arquiteturas DL mais utilizadas, destacam-se as Redes Neurais Convolucionais (CNN), modelos de referência na classificação de imagens clínicas. Atualmente, uma nova abordagem com foco na análise de sinais biomédicos de uma dimensão tem se beneficiado de modelos CNN capazes de abstrair melhor os dados em busca de padrões relevantes, reduzir o custo computacional e serem utilizados na classificação de arritmias. Esses modelos aumentam a efetividade tanto na detecção como na classificação das arritmias cardíacas (GANAPATHY; SWAMINATHAN; DESERNO, 2018, p. 1).

1.1 OBJETIVOS GERAL E ESPECÍFICOS

Neste contexto, este trabalho propõe o seguinte objetivo geral: efetuar a análise do desempenho de um modelo de Redes Neurais Convolucionais com Dois Canais (DC-CNN), com ou sem *data augmentation*, na classificação de arritmias morfológicas em sinais ECG utilizando o complexo QRS.

Portanto, para abordar o tema proposto, definiu-se os seguintes objetivos específicos para este trabalho:

1. Executar o pré-processamento dos dados de sinais ECG do *MIT-BIH Arrhythmia Database* (MITDB), disponibilizado pela *PhysioBank* (GOLDBERGER *et al.*, 2000);
2. Desenvolver arquiteturas CNN com um e dois canais para a classificação de arritmias morfológicas;
3. Treinar o modelo DC-CNN em cenários distintos, com ou sem *data augmentation*, no conjunto de treinamento, utilizando a abordagem inter-paciente; e
4. Efetuar a avaliação comparativa do desempenho do modelo DC-CNN na detecção e classificação de arritmias morfológicas em relação aos modelos CNN propostos na literatura.

1.2 TRABALHOS RELACIONADOS

A detecção e classificação de arritmias em sinais ECG desempenha um importante papel no diagnóstico de doenças cardíacas. Entre os classificadores de arritmias, os modelos de redes neurais obtiveram os melhores resultados e tornaram-se bem populares na maioria das aplicações (JAMBUKIA; DABHI; PRAJAPATI, 2015).

Kiranyaz, Ince e Gabbouj (2016) relataram que, no treinamento de modelos classificadores, o emprego de redes neurais baseadas em CNN integram duas etapas: a extração de características e a classificação do sinal. Os autores relataram, ainda, que modelos CNN facilitam o desenvolvimento de melhores classificadores de arritmias morfológicas (KIRANYAZ; INCE; GABBOUJ, 2016, p. 664–665).

Empregando um sensor *wearable*, Rajpurkar *et al.* (2017) utilizaram dados de apenas uma derivação para realizar a classificação de arritmias, superando o desempenho de cardiologistas. Igualmente, Feng, Chen e Hou (2021) desenvolveram um modelo CNN para classificar arritmias cardíacas em dispositivos embarcados com atenção na segurança e privacidade de dados.

Neste contexto Habib, Karmakar e Yearwood (2019) utilizaram detectores de complexos QRS dos sinais de ECG empregando a segmentação dos dados em batimentos cardíacos na avaliação morfológica de diferentes arritmias. Além do mais, os autores

utilizaram arquiteturas CNN para extrair diferentes características dos sinais de ECG em diferentes banco de dados.

Jun *et al.* (2018) transformaram os segmentos de ECG do complexo QRS em imagens empregando um modelo CNN de duas dimensões. Tais modelos possuem um bom histórico na classificação de imagens. Além do mais, os autores concluíram que o aumento de dados através de cortes e translações em imagens não afetou o sinal de ECG.

Xiang, Lin e Meng (2018), utilizando uma abordagem de dois níveis, reduziram o custo computacional e tornou os modelos CNN mais adaptáveis às variações dos sinais ECG, aumentando a precisão na detecção de arritmias cardíacas com a análise do complexo QRS.

Nagata, Candido e Silva (2021) avaliaram o efeito da divisão das gravações dos sinais de ECG de diferentes pacientes entre os conjuntos de treinamento e de teste. Os autores concluíram que, ao considerar os batimentos cardíacos de um mesmo paciente em ambos conjuntos, os resultados tendem a serem parciais e não muito realistas, pois algumas características do sinal de ECG são distintas para cada indivíduo.

Considerando que as irregularidades entre dois *Inter-Heartbeat Intervals* (RRI)¹ podem indicar arritmias cardíacas em sinais de ECG, Wu, Chang e Chu (2018, p. 92) propuseram um modelo CNN para detecção de arritmias a partir de uma abordagem inter-paciente utilizando a análise RRI no MITDB.

Runchuan Li *et al.* (2019) propuseram um modelo de Redes Neurais Recorrentes (RNN) utilizando camadas *Long-Short Term Memory* (LSTM) bidirecionais, aplicação de filtros de remoção de ruídos, extração de características e o algoritmo de atenção *BiLSTM-Attention*. Os autores treinaram o modelo RNN no MITDB em cinco diferentes classes de arritmias. Contudo, os autores alertaram que a complexidade do modelo *BiLSTM-Attention* exige uma grande quantidade de dados para alcançar uma alta acurácia.

Runchuan Li *et al.* (2019) propuseram um modelo de RNN contendo camadas LSTM bidirecionais, filtros de remoção de ruídos, extração de características e algoritmo de atenção *BiLSTM-Attention*. Os autores treinaram o modelo RNN no MITDB em cinco diferentes classes de arritmias. Contudo, eles concluíram que a complexidade do modelo *BiLSTM-Attention* exige uma grande quantidade de dados para alcançar uma alta acurácia (LI, R. *et al.*, 2019, p. 109870).

Por outro lado, Kachuee, Fazeli e Sarrafzadeh (2018) propuseram uma arquitetura *Neural Residual Network* (ResNet) desenvolvida a partir de cinco blocos residuais com camadas convolucionais, de *pooling*, uma camada de saída com cinco unidades e a função de ativação *softmax*. As conexões residuais dos blocos aumentam significativamente o tempo de treinamento do modelo.

¹ RRI é o intervalo entre dois batimentos cardíacos sucessivos.

1.3 ORGANIZAÇÃO DO TRABALHO

A fim de alcançar o objetivo geral do trabalho, o Capítulo 2 - Fundamentação Teórica - apresenta os conceitos gerais sobre o treinamento de redes neurais, destaca a arquitetura CNN multicanal, detalha a estrutura dos sinais de ECG, define os algoritmos de *data augmentation* de interesse deste trabalho e aborda as principais métricas utilizadas na avaliação do desempenho de classificadores de arritmias.

O Capítulo 3 - Revisão Literária - revisa os trabalhos científicos mais relevantes relacionados aos modelos CNN de uma dimensão utilizados na detecção e/ou classificação arritmias cardíacas.

O Capítulo 4 - Metodologia - apresenta os procedimentos utilizados na estruturação, análise e interpretação dos sinais de ECG do MITDB, descreve as arquiteturas dos modelos CNN desenvolvidas, detalha os procedimentos adotados no treinamento e avaliação do desempenho dos modelos estudados.

O Capítulo 5 - Resultados - analisa e interpreta os resultados do desempenho do modelo DC-CNN na classificação das categorias de arritmias em sinais de ECG a partir de dois experimentos: o primeiro, que efetua a análise comparativa do desempenho do modelo DC-CNN e; o segundo, que investiga a eficácia da aplicação dos algoritmos de *data augmentation*.

Por último, o Capítulo 6 - Conclusão - revisa os principais aspectos deste trabalho, apresenta os resultados obtidos e propõe novos estudos sobre o emprego de modelos CNN para a detecção e classificação de arritmias.

2 FUNDAMENTAÇÃO TEÓRICA

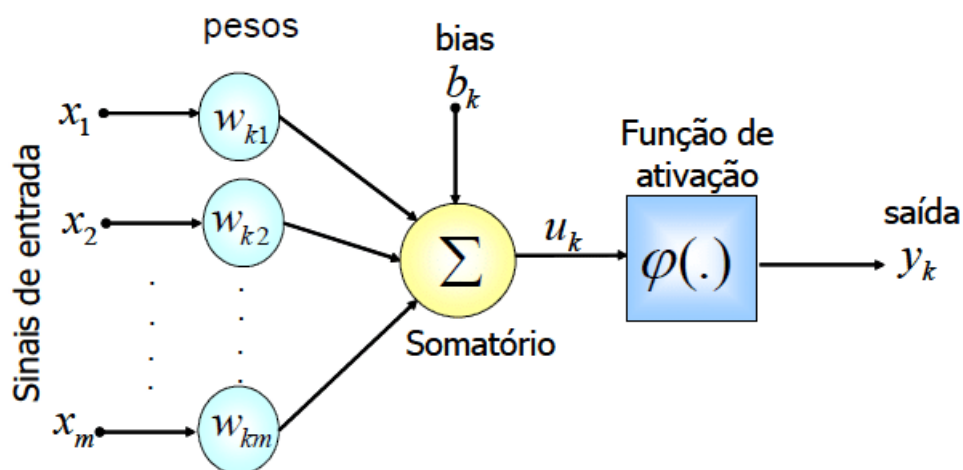
Este capítulo apresenta os conceitos gerais sobre treinamento de redes neurais, apresenta a arquitetura CNN multicanal, detalha a estrutura dos sinais de ECG, define os algoritmos de *data augmentation* de interesse deste trabalho e aborda as principais métricas utilizadas na avaliação do desempenho de classificadores de arritmias.

2.1 REDES NEURAIS

Modelos computacionais que utilizam técnicas de ML permitem a resolução de tarefas complexas como reconhecimento de imagens, detecção de anomalias, etc. Em geral, as tarefas de ML são resolvidas por um modelo que processa um conjunto de dados. Estas amostras possuem um conjunto de características que serão reconhecidas nos modelos treinados utilizando técnicas de otimização (GOODFELLOW; BENGIO; COURVILLE, 2016, p.99).

Uma rede neural, estrutura interconectada que se assemelha às conexões de neurônios biológicos, realiza a combinação linear de suas entradas e emprega a função de ativação, adicionando um comportamento de não-linearidade para modelar funções mais complexas (Figura 1).

Figura 1 – Estrutura de uma unidade neural



Fonte: Soares e Silva (2011, p. 60)

Observando a Figura 1, inferem-se as seguintes Equações (1) e (2), onde: a^l e y_k representam a saída da função de ativação do modelo na camada l ; g_l , a função de ativação na camada l ; $z^{[l]}$, o somatório dos pesos de todas as entradas da camada l ; $W^{[l]}$, a matriz de pesos dados a entrada na camada l e; $b^{[l]}$, o vetor de *bias* na camada l .

$$a^l = g_l(z^{[l]}) = y_k \quad (1)$$

$$z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]} \quad (2)$$

A unidade neural pode ser ampliada para um arranjo linear com múltiplas entradas, saídas e camadas internas. Na saída de cada unidade, a função de ativação adiciona componentes não lineares, possibilitando aproximações de funções mais complexas. Em geral, na classificação multi-classe, utiliza-se a função de ativação *softmax* na camada L e a função custo (J) *categorical crossentropy*¹ apresentada nas Equações (3) e (4), onde: k representa o número de unidades neurais para cada camada e L a última camada neural do modelo.

$$a_k^{[L]} = softmax_k(z^{[L]}) = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}} \quad (3)$$

$$J = - \sum_{k=1}^k y_k \log a_k^{[L]} \quad (4)$$

No treinamento de redes neurais, o pesquisador define ou realiza a busca empregando algoritmos de otimização dos hiperparâmetros, isto é: o número de unidades de camadas ocultas, as funções de ativações, a função custo e outros parâmetros. Assim, minimiza-se a função custo, definindo os pesos necessários para que o modelo se aproxime da função real, conforme a Equação (5), onde: θ representa o vetor que combina os parâmetros $W^{[l]}$ e $b^{[l]}$; m , o total de saídas do modelo e; $L(y^i, \hat{y}^{(i)})$, a perda média entre o vetor real e o previsto pelo modelo.

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m L(y^i, \hat{y}^{(i)}) \quad (5)$$

¹ Perda de entropia cruzada categórica.

A função custo compara o resultado em uma iteração e calcula o erro médio entre o objeto real e resultado retornado. Assim, os pesos são atualizados progressivamente desde a camada de saída até a camada de entrada para retornar um erro menor na próxima iteração. Essa abordagem, conhecida como *backpropagation*, otimiza os pesos de cada unidade neural da rede e minimiza a função custo, permitindo o treinamento de diversas camadas ocultas (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 204).

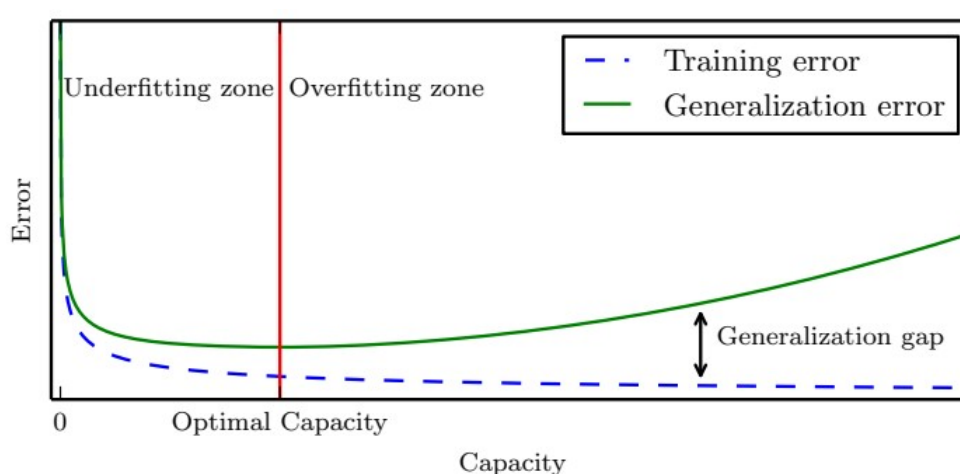
2.2 MODELOS DEEP LEARNING

As arquiteturas DL possuem um número maior de camadas internas capazes de extrair as características da amostra, superando as limitações dos modelos ML. Deste modo, aumenta-se a escalabilidade dos modelos e a generalização de novas variações internas nos dados (LECUN; BENGIO; HINTON, 2015, p. 436 v. 521).

O principal objetivo de treinamento de modelos é prever corretamente o resultado a partir de novas amostras que não foram treinadas, tal capacidade é conhecida como generalização (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 110). Antes do treinamento dos modelos, infere-se os resultados num conjunto de dados separado para determinar o erro de teste do modelo.

Dois problemas afetam a generalização do modelo e dificulta a correta previsão de novas amostras: (1) o *underfitting*, que ocorre quando o modelo não foi treinado por tempo suficiente para a boa previsão dos dados e; (2) o *overfitting*, que ocorre quando o modelo foi treinado por um tempo superior ao necessário e suas previsões ficam muito específicas na classificação no conjunto de treinamento (Figura 2).

Figura 2 – Relação entre capacidade e erro do modelo



Fonte: Goodfellow, Bengio e Courville (2016, p. 115)

Devido a dificuldade em determinar o tempo de treinamento ideal do conjunto

de treinamento em razão do *overfitting*, utilizam-se técnicas de regularizações nos modelos, como a camada *dropout*.

Em geral, esses modelos possuem um grande número de parâmetros, os quais agravam o problema de *overfitting*. Neste sentido, o *dropout* é um método eficaz de regularização menos custoso computacionalmente para os modelos mais complexos. Tal método elimina, aleatoriamente, os pesos e as conexões aprendidas pelas unidades durante o treinamento, prevenindo que os modelos se adaptem excessivamente às especificidades dos dados durante o treinamento (SRIVASTAVA, 2013, p. 1).

Entretanto, devido ao alto número de parâmetros que precisam ser treinados, os modelos tradicionais, com apenas unidades neurais, são lentos e necessitam de uma grande quantidade de dados para evitar o *overfitting*. Uma solução para minimizar essas limitações é a utilização das CNN.

2.3 REDES NEURAIIS CONVOLUCIONAIS (CNN)

Uma CNN é uma arquitetura que permite processar os dados, sem perder a coerência da amostra, utilizando filtros convolucionais que substituem as combinações lineares. Essa arquitetura faz parte de um tipo especial de rede neural utilizada para processar dados de tipo grade, tais como: imagens, séries temporais, etc. A CNN melhora o processamento de uma grande quantidade de dados e auxilia na classificação (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 330).

Em geral, toda operação de convolução consiste numa média ponderada com maior peso alocado ao dados mais recentes, conforme a Equação (6), onde: $s(t)$ representa a função decorrente a convolução sobre a entrada $x(a)$ e $w(t - a)$ a função peso aplicado no instante t .

$$s(t) = \int x(a)w(t - a) da \quad (6)$$

Por tratar-se de meios computacionais, os dados estão discretizados no tempo de amostragem, assim, define-se sua forma discreta na Equação (7).

$$s(t) = (x \otimes w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a) \quad (7)$$

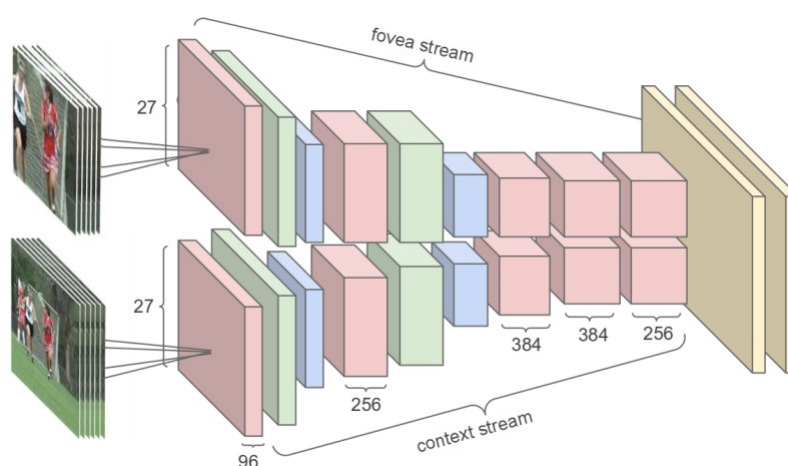
Em uma CNN os pesos são tratados como parâmetros a serem treinados, utilizando o algoritmo *backpropagation*. Assim, o modelo reduz a quantidade de parâmetros treináveis, otimizando o desempenho da memória e sua eficiência estatística (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 335).

Essas redes melhoram a capacidade de compartilhar o treinamento realizado ao calcular apenas um filtro convolucional por camada, reduzindo o *overfitting*. No processamento de dados de séries temporais, o treinamento do modelo CNN produz uma linha temporal indicando a posição de diferentes características dos dados treinados (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 339).

Outra estrutura no desenvolvimento de modelos CNN são as camadas *pooling*, que proporcionam a redução da dimensionalidade dos dados de entrada e aceleram o treinamento. Tais camadas otimizam a classificação de algumas características presentes nos dados sem o treinamento de parâmetros. As camadas *pooling* permitem que o modelo torne-se invariante para pequenas translações dos dados (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 342).

De um modo geral, as CNN possuem uma única entrada de dados, no entanto, existem arquiteturas convolucionais com múltiplos canais paralelos de processamento de sinais. A Figura 3 apresenta um exemplo de arquitetura com dois canais de processamento.

Figura 3 – Exemplo de arquitetura CNN multicanal



Fonte: Karpathy *et al.* (2014, p. 4)

Ao identificar diferentes padrões em sinais temporais em paralelo, uma DC-CNN extrai características distintas em cada canal de processamento, reduzindo a dimensionalidade do dado e o custo computacional do treinamento. Apesar de ser geralmente aplicada em processamento de imagens, os modelos DC-CNN podem ser utilizados na extração de diferentes características em dados de uma dimensão, como sinais de ECG (XIANG; LIN; MENG, 2018, p. 2–3).

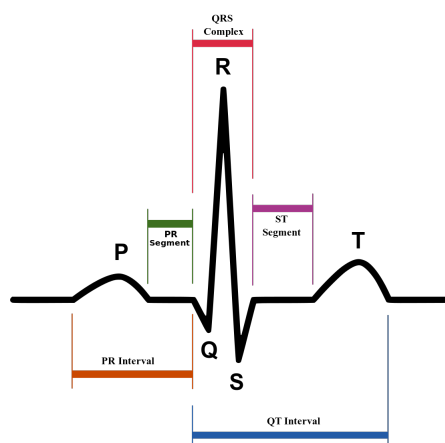
2.4 SINAIS DE ELETROCARDIOGRAMAS (ECG)

O coração é um músculo que se contrai de forma rítmica, bombeando sangue por todo o corpo. Esta contração tem seu início a partir do nó sinoatrial que atua como um marcapasso natural e propaga-se através do coração. Essa atividade gera as correntes elétricas na superfície do corpo humano, provocando variações no potencial elétrico na superfície da pele (S. LUZ *et al.*, 2016, p. 145).

A análise dos impulsos elétricos do coração, por meio de eletrodos colocados em locais específicos no corpo do paciente, possibilita inferir sobre o estado de saúde de um indivíduo (GUYTON; HALL, 2011, p. 129). Os ECG são utilizados por serem métodos acessíveis e não invasivos para avaliar o estado fisiológico do coração (MOODY; MARK, 2001, p. 45).

Um sinal de ECG normal é composto pela onda P, complexo QRS e a onda T (Figura 4). A onda P é produzida pelos potenciais elétricos gerados quando os átrios despolarizam-se antes do início da contração atrial. O complexo QRS é produzido pelos potenciais gerados quando os ventrículos despolarizam-se antes de sua contração. A onda T é produzida pelos potenciais gerados, enquanto os ventrículos retornam ao estado de despolarização (GUYTON; HALL, 2011, p. 129).

Figura 4 – Estrutura de um sinal de ECG normal



Fonte: Atkielski (2007)

Segundo a recomendação da Associação para o Avanço da Instrumentação Médica (AAMI), os sinais de ECG podem ser divididos em quatro classes de arritmias cardíacas e uma classe de batimento cardíaco normal, ou seja, batimento normal (N), batimento ectópico supraventricular (S), batimento ectópico ventricular (V), fusão de batimento ventricular (F). A Tabela 1 apresenta as classes de arritmias recomendadas pela AAMI correspondentes às anotações presentes no MITDB.

Tabela 1 – Mapeamento das anotações do MITDB em relação as classes da AAMI

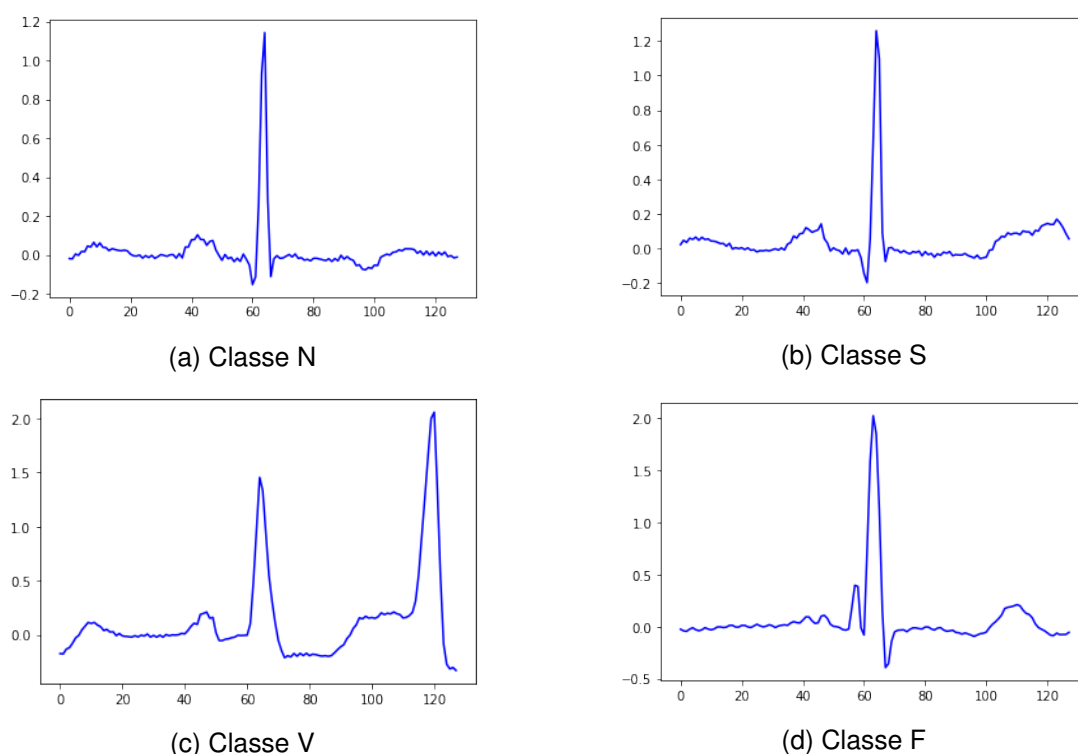
Classes AAMI	Anotações presentes no MITDB
N	<i>Normal beat</i> <i>Left bundle branch block beat</i> <i>Right bundle branch block beat</i> <i>Bundle branch block beat (unspecified)</i>
S	<i>Atrial premature beat</i> <i>Aberrated atrial premature beat</i> <i>Nodal (junctional) premature beat</i> <i>Supraventricular premature or ectopic beat (atrial or nodal)</i> <i>Atrial escape beat</i> <i>Nodal (junctional) escape beat</i> <i>Supraventricular escape beat (atrial or nodal)</i>
V	<i>Premature ventricular contraction</i> <i>R-on-T premature ventricular contraction</i> <i>Ventricular escape beat</i>
F	<i>Fusion of ventricular and normal beat</i>
Q	<i>Paced beat</i> <i>Fusion of paced and normal beat</i> <i>Unclassifiable</i> <i>Beat not classified during learning</i>

Os batimentos ectópicos supraventriculares (classe S) consistem na Contração A-trial Prematura (CAP). Vários ritmos resultam de focos supraventriculares (em geral nos átrios). Muitos ritmos são assintomáticos e não necessitam de tratamento. Os ritmos ectópicos supraventriculares incluem: extrassístoles supraventriculares, taquicardia a-trial, taquicardia atrial multifocal, taquicardia juncional não paroxística, marca-passo atrial mutável (MITCHELL, 2019b).

Os batimentos ectópicos ventriculares (classe V) consistem na Contração Ventricular Prematura (PVC) provocados pela reentrada no ventrículo ou atividade anormal das células ventriculares (MITCHELL, 2019a). Em corações normais, sua ocorrência está associada a nenhuma significância clínica, no entanto, em alguns pacientes, a PVC é acionada pelo mesmo mecanismo que dá origem à taquicardia ventricular (NG, 2006, p. 1707). Os batimentos de fusão (classe F) são complexos QRS formados a partir da fusão de um impulso originário do ventrículo com outro originário dos átrios (MOORE; KULKARNI; JOGLAR, 2017).

As amostras de sinais de ECG das diferentes classes de batimento cardíaco estão apresentadas na Figura 5. Observa-se que o batimento da classe V apresenta uma grande diferença morfológica em relação às outras classes, enquanto as classes N e S são similares em formato. A diferenciação entre os batimentos da classe S e os normais é um dos maiores desafios dos métodos de detecção de arritmias, pois é pouco provável obter a identificação de classe S com base em apenas sua morfologia (HE, 2020, p. 2).

Figura 5 – Amostras de sinais de ECG contendo diferentes categorias de arritmias cardíacas



Fonte: Autor (2022)

Considerando que as irregularidades entre dois RRI podem indicar arritmias cardíacas, He (2020, p. 6) comenta que, clinicamente, conhecer o RRI seria necessário para auxiliar na detecção da classe S, pois a mesma apresenta batimentos prematuros e normalmente mais curtos que o normal.

2.5 MIT-BIH *ARRHYTHMIA DATABASE* (MITDB)

Como resultado dos trabalhos de Holter (1961 apud MOODY; MARK, 2001) realizou-se gravações de ECG de longa duração de pacientes internados. O *Long-Term ECG*, com durações típicas de 24 horas, consolidou-se como procedimento padrão para a observação dos aspectos das atividades elétricas cardíacas (HOLTER, 1961, p. 45 apud MOODY; MARK, 2001).

Em 1975, devido a necessidade de estabelecer um padrão nas gravações de *Long-Term ECG*, os dados do *Arrhythmia Laboratory of Boston's Beth Israel Hospital* (BIH)² foram coletados, digitalizados e anotados (MOODY; MARK, 2001, p. 45). Posteriormente, com base nos dados adquiridos pelo BIH, o MITDB foi desenvolvido para avaliar detectores de arritmias cardíacas e, também, como suporte à pesquisas relacionadas às dinâmicas cardíacas (MOODY; MARK, 2001, p. 45).

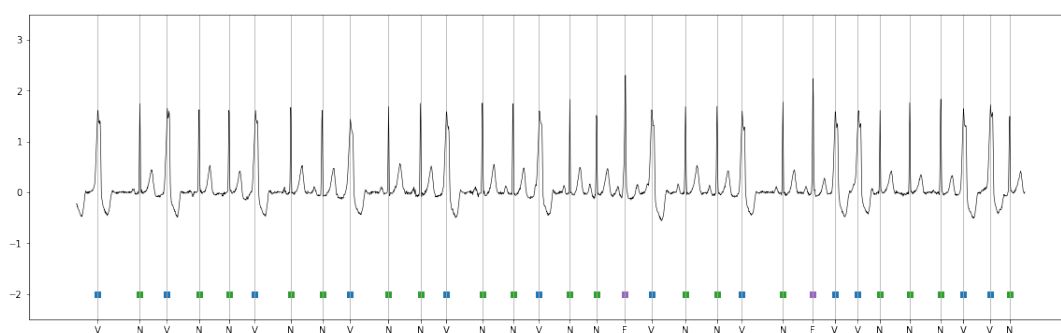
Em síntese, o MITDB consta de 48 recortes de meio hora de sinais de ECG

² Atualmente conhecido como *Beth Israel Deaconess Medical Center*.

extraídos de 47 indivíduos³ pesquisados pelo laboratório BIH, entre os anos de 1975 e 1979. Ele inclui 25 homens com idades de 32 anos a 82 anos e 22 mulheres com idades entre 23 e 89 anos. Na maioria das gravações de ECG obtidas foram coletados dados de apenas do canal, o *Modified Limb II* (MLII), onde os eletrodos são colocados no tórax do indivíduo, prática padrão para gravações de ECG ambulatoriais (MOODY; MARK, 2001, p. 46).

As anotações sobre essas gravações foram realizadas por dois cardiologistas independentes, totalizando cerca de 110.000 anotações, sendo que cerca de 109.000 foram anotados para o complexo QRS (MOODY; MARK, 2001, p. 47). Um exemplo de uma gravação extraída do MITDB é apresentada na Figura 6.

Figura 6 – Exemplo de gravação de ECG extraída do MITDB



Fonte: Autor (2022)

De um modo geral, o MITDB possibilita que as gravações de ECG por indivíduo possam ser divididas em conjuntos de treinamento, validação e teste. Assim, o pesquisador pode utilizar duas abordagens no processamento das diferentes gravações: a intra-paciente, onde os conjuntos de treinamento e de teste provêm de um mesmo grupo de indivíduos e; a inter-paciente, onde o conjunto de treinamento provém de um grupo de indivíduos e o conjunto de teste de outro.

Na abordagem inter-paciente os conjuntos de treinamento e teste não possuem sinais dos mesmos pacientes e o modelo treinado poderá generalizar os padrões encontrados no grupo de treinamento e inferir sobre o grupo de teste (YE, 2018, p. 19). Por outro lado, na abordagem intra-paciente, os sinais de um grupo de indivíduos são aleatoriamente combinados antes da separação dos conjuntos de treinamento e validação (YE, 2018, p. 19).

O uso dos batimentos do mesmo paciente no treinamento e teste torna tendencioso o processo de avaliação do desempenho dos modelos classificadores. Esse problema ocorre porque tais modelos tendem a aprender as particularidades dos batimentos de um indivíduo durante o treinamento, favorecendo a sua detecção na etapa

³ As gravações 201 e 202 são de um mesmo indivíduo.

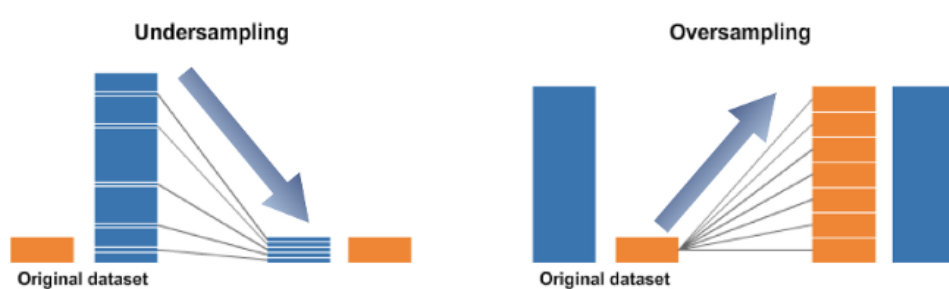
de testes (GARCIA *et al.*, 2017, p. 1). Assim, para obter resultados mais alinhados aos cenários do mundo real, recomenda-se seguir os protocolos da AAMI, ou seja, os sinais são divididos em quatro classes de arritmias cardíacas e uma classe normal, e empregando a abordagem inter-paciente (CHAZAL; O'DWYER; REILLY, 2004, p. 1197–1198).

2.6 DATA AUGMENTATION

Num conjunto de dados desbalanceados as classes podem não estar representadas igualmente. No caso de bancos de dados derivados de diagnósticos médicos predominam as amostras normais com uma pequena percentagem de amostras anormais (CHAWLA *et al.*, 2002, p. 321).

Assim, antes do treinamento de modelos classificadores pode-se minimizar o problema de desbalanceamento das classes utilizando as seguintes técnicas: (1) *undersampling*, que retira os dados de classes majoritárias; e (2) *oversampling*, que adiciona novas amostras artificiais das classes minoritárias ao conjunto de dados (Figura 7).

Figura 7 – Comparação entre as técnicas de *undersampling* e *oversampling*



Fonte: Mars (2020)

O *oversampling*, ou *data augmentation*, baseia-se no aumento artificial de amostras a partir dos dados existentes para a uniformização de banco de dados desbalanceados. Essa técnica auxilia na generalização do modelo e melhora a classificação das amostras minoritárias. No entanto, a inserção artificial de cópias de amostras minoritárias, pode acelerar o *overfitting* do modelo. Uma abordagem para evitar o *overfitting*, em modelos classificadores de séries temporais de uma dimensão seria adicionar ruído ao sinal a fim de gerar novas amostras. Contudo, a aplicação desta técnica pode descaracterizar o sinal original, reduzindo o desempenho do classificador (JUN *et al.*, 2018, p. 3).

Entre os algoritmos de *data augmentation* descritos na literatura, destacam-se o *Synthetic Minority Over-Sampling Technique* (SMOTE) e o *Random Over-Sampling* (ROS). O algoritmo SMOTE adiciona novas amostras sintéticas utilizando estratégias

de *K-Nearest Neighbors* (KNN)⁴ (CHAWLA *et al.*, 2002). Por outro lado, o ROS consiste de um método não heurístico que replica, aleatoriamente, amostras das classes minoritárias. Uma vez que este algoritmo insere cópias exatas das classes minoritárias, diversos autores concordam que este método pode aumentar a probabilidade de *overfitting* (BATISTA; PRATI; MONARD, 2004, p. 4).

2.7 AVALIAÇÃO DO DESEMPENHO DE CLASSIFICADORES

Não existe apenas um caminho para a seleção do melhor modelo classificador. Por essa razão, diversas métricas são geralmente consideradas, permitindo uma visualização multifacetada das características desses modelos. Assim sendo, o uso de métricas de avaliação do desempenho de classificadores a partir dos valores obtidos nas matrizes de confusão é um dos métodos mais comuns para avaliar o desempenho de modelos classificadores (LUQUE *et al.*, 2019, p. 216–217).

A matriz de confusão empregada em problemas de classificação binária está apresentada na Tabela 2. Os elementos desta matriz são caracterizados: (1) Verdadeiro Positivo (TP) e Falso Positivo (FP), número de predições positivas da classe atual e; (2) Verdadeiro Negativo (TN) e Falso Negativo (FN), número de predições negativas em relação à classe atual (POWERS, 2020, p. 37).

Tabela 2 – Matriz de confusão para problemas de classificação binária

		Classe Prevista	
		Positivo (P)	Negativo (N)
Classe Atual	Verdadeiro (T)	TP	FN
	Falso (F)	FP	TN

Fonte: Markoulidakis *et al.* (2021, p. 4)

A partir das combinações da matriz de confusão binária, define-se as métricas de avaliação de interesse deste trabalho, ou seja: Acurácia (ACC), *Recall* ou Sensibilidade (SEN), Precisão (PRE) e *F1-Score* (F1) (Tabela 3).

Na matriz de confusão multiclasse, as colunas representam os valores de cada classe prevista ($C_1, C_2, C_3, \dots, C_N$), enquanto as linhas representam os valores da classe atual. Assim, um elemento da matriz de confusão multiclasse da linha i e coluna j ($C_{i,j}$) fornece o valor na qual a classe prevista é j e a classe atual é i (MARKOULIDAKIS *et al.*, 2021, p. 4) (Tabela 4).

⁴ Classificador onde o aprendizado é baseado “no quão similar” é um dado do outro.

Tabela 3 – Métricas de avaliação baseada na matriz de confusão binária

Métrica	Fórmula
Acurácia	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
Precisão	$PRE = \frac{TP}{TP+FP}$
Sensibilidade	$SEN = \frac{TP}{TP+FN}$
F1-Score	$F1 = 2 \cdot \frac{PRE \cdot SEN}{PRE+SEN}$

Fonte: Powers (2020, p. 38–39)

Tabela 4 – Matriz de confusão para problemas de classificação multiclasse

		Classe Prevista			
		C_1	C_2	...	C_N
Classe Atual	C_1	$C_{1,1}$	FP	...	$C_{1,N}$
	C_2	FN	TP	...	FN

	C_N	$C_{N,1}$	FP	...	$C_{N,N}$

Fonte: Markoulidakis *et al.* (2021, p. 4)

Contudo, em uma classificação multiclasse, as métricas de avaliação obtidas da matriz de confusão binária não podem ser aplicadas diretamente, pois a matriz de confusão multiclasse tem dimensão $N \times N$ ⁵. Assim, realiza-se uma análise para cada classe específica.

Como exemplo, utilizando a classe C_2 da matriz de confusão multiclasse da Tabela 4 os valores da diagonal principal da matriz representam o número de amostras classificadas corretamente, isto é, o valor de $C_{2,2}$ representa a grandeza TP.

Selecionando a linha da classe C_2 , o número total de amostras excluindo TP fornece a grandeza FN, ou seja, o número de amostras classificadas erroneamente como pertencentes a uma classe diferente de C_2 . Assim, deduz-se a métrica de sensibilidade (SEN) ou *recall* para a classe C_2 . Igualmente, selecionando a coluna da classe C_2 , o número total de amostras excluindo TP fornece a grandeza FP, ou seja, que representa número de amostras classificadas erroneamente como pertencentes à classe C_2 . Assim, deduz-se a métrica de Precisão (PRE) para a classe C_2 . A média harmônica entre os valores de sensibilidade e precisão obtidas na matriz de confusão multiclasse fornece o valor da métrica F1.

O somatório de todos os valores na diagonal principal da matriz de confusão multiclasse ($C_{1,1} + C_{2,2} + C_{3,3} + \dots + C_{N,N}$) fornece o total de amostras classificadas

⁵ Onde N é o número dos rótulos das classes $C_1, C_2, C_3, \dots, C_N$.

corretamente em todas as classes. Assim, deduz-se a acurácia global do modelo classificador como sendo o total de amostras classificadas como TP em relação ao total de amostras analisadas. A Tabela 5 apresenta uma visão geral de métricas definidas a partir da matriz de confusão multiclasse, em particular as de interesse deste trabalho.

Tabela 5 – Métricas de avaliação baseada na matriz de confusão multiclasse

Métrica	Fórmula
Acurácia	$ACC (geral) = \frac{\sum_{i=1}^N TP(C_i)}{\sum_{i=1}^N \sum_{j=1}^N C_{i,j}}$
Precisão da Classe C_i	$PRE (C_i) = \frac{TP(C_i)}{TP(C_i)+FP(C_i)}$
Sensibilidade da Classe C_i	$SEN (C_i) = \frac{TP(C_i)}{TP(C_i)+FN(C_i)}$
F1-Score da Classe C_i	$F1 (C_i) = 2 \cdot \frac{PRE(C_i) \cdot SEN(C_i)}{PRE(C_i) + SEN(C_i)}$

Fonte: Markoulidakis *et al.* (2021, p. 5)

2.8 RECEIVER OPERATING CHARACTERISTICS (ROC) E PRECISION RECALL (PR)

Um método avançado para avaliar o desempenho de algoritmos de classificação binária é conhecido como *Receiver Operating Characteristic* (ROC), um dos estimadores estatísticos mais utilizados na avaliação de modelos classificadores. A ROC interpreta graficamente o desempenho de um modelo a partir da relação entre suas taxas de sensibilidade e especificidade e são alternativas robustas e intuitivas na comparação de modelos classificatórios (POLO; MIOT, 2020, p. 1–3).

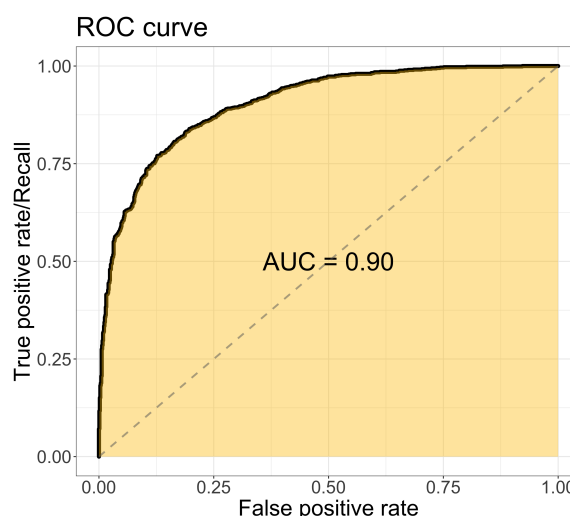
Nas curvas ROC os eixos estão identificados da seguinte forma: (1) o eixo vertical, Taxa de Verdadeiros Positivos (TPR) definida na Equação (8); e o eixo horizontal, Taxa de Falsos Positivos (FPR) definida na Equação (9) (CHAWLA *et al.*, 2002) (Figura 8).

$$TPR = \frac{TP}{TP + FN} = SEN \quad (8)$$

$$FPR = \frac{FP}{TN + FP} \quad (9)$$

A partir da curva ROC extrai-se a *Area Under ROC Curve* (AUROC), métrica empregada na avaliação do desempenho de modelos classificadores (PROVOST; FAWCETT, 2000). A AUROC é o resultado da integração de todos os pontos durante o

Figura 8 – Exemplo de uma curva ROC



Fonte: Chou (2020)

trajeto da curva ROC e estima o comportamento da acurácia do teste do modelo classificador. No gráfico da curva ROC, sua *baseline*⁶ é representada por uma linha tracejada na diagonal com AUROC de 0,5 (Figura 8).

Neste sentido, Hosmer, Lemeshow e Sturdivant (2013, p. 177) sugere a seguinte diretriz para avaliar o desempenho de modelos classificadores, segundo os valores obtidos da métrica AUROC, nos seguintes intervalos: 0,5 – 0,7 (ruim); 0,7 – 0,8 (aceitável); 0,8 – 0,9 (excelente); e > 0,9 (perfeito). Assim sendo, verifica-se que quanto mais a curva ROC aproxima-se do canto superior esquerdo, melhor é a qualidade do teste quanto à capacidade para determinar a classe de uma amostra.

Enquanto curvas ROC são visualmente interessantes para realizar comparações fornecendo uma perspectiva global do desempenho de classificadores, o seu uso em conjuntos de dados desbalanceados pode conduzir a conclusões equivocadas sobre o verdadeiro desempenho do modelo classificador (SAITO; REHMSMEIER, 2015, p. 1).

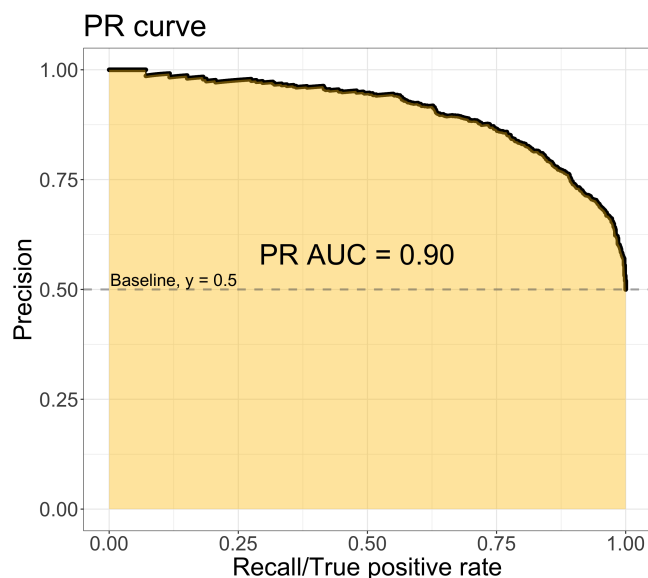
Uma solução para minimizar este problema seria a utilização da curva *Precision-Recall* (PR) e sua respectiva métrica derivada, a *Area Under PR Curve* (AUPRC). Diferentemente das curvas ROC, o objetivo da curva PR é maximizar os valores de sensibilidade e precisão ao aproximar-se da extrema direita superior do gráfico (Figura 9).

Numa classificação binária, o *baseline* da curva PR é determinada pela razão entre positivos (P) e negativos (N) como $y_{AUPRC} = P/(P + N)$. Por exemplo, $y_{AUPRC} = 0,5$ ⁷ para uma distribuição balanceada, mas $y_{AUPRC} = 0,09$ numa distribuição com uma

⁶ A linha de referência diagonal do gráfico ROC representa a classificação de um modelo aleatório.

⁷ A $y_{AUPRC} = x$ descreve o *baseline* de um classe no gráfico da curva PR.

Figura 9 – Exemplo de uma curva PR



Fonte: Chou (2020)

razão de 1 : 10. Em razão desta *baseline* móvel, a AUPRC altera-se com a razão $P : N$. Por exemplo, a AUPRC de classificadores aleatórios é 0,5 apenas para distribuições de classes balanceadas, enquanto, no caso geral, $AUPRC = P/(P + N)$ é válida para as distribuições balanceadas e desbalanceadas (SAITO; REHMSMEIER, 2015, p. 7).

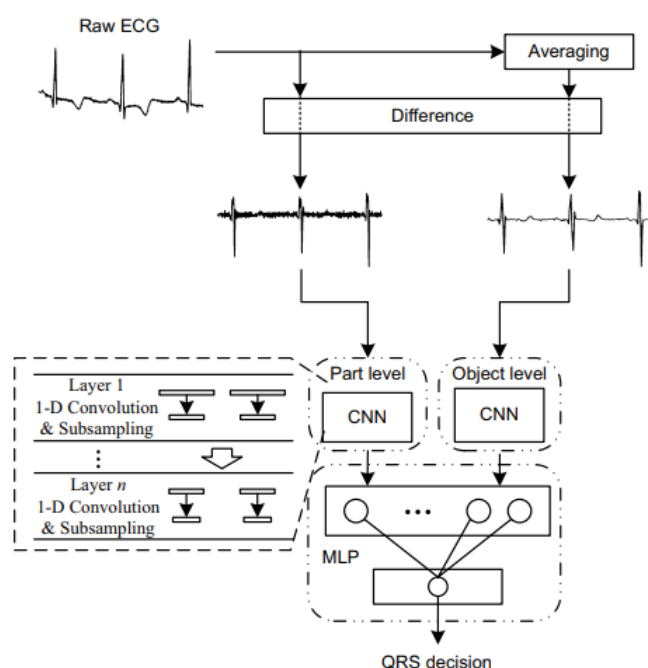
Assim, uma vez que a *baseline* da curva PR varia de acordo com a proporção de amostras de cada classe, emprega-se a AUPRC métrica mais efetiva na comparação do desempenho de modelos em banco de dados desbalanceados (SAITO; REHMSMEIER, 2015, p. 7).

3 REVISÃO DA LITERATURA

Este capítulo revisa os trabalhos científicos mais relevantes sobre modelos CNN de uma dimensão utilizados na detecção e classificação arritmias cardíacas. Além do mais, descreve os aspectos mais relevantes das arquiteturas CNN, descreve a abordagem utilizada no processamento dos sinais de ECG e apresenta os resultados alcançados pelos modelos propostos na literatura.

Nesse sentido, Xiang, Lin e Meng (2018, p. 4) propõem um modelo CNN de uma dimensão, constituído de dois canais de abstrações granulares de dados para a análise do complexo QRS, utilizando o MITDB (Figura 10). Sua arquitetura inclui um canal de abstração menos detalhado e um segundo canal capaz de extrair outras características relevantes da amostra.

Figura 10 – Arquitetura proposta por Xiang, Lin e Meng (2018)



Fonte: Xiang, Lin e Meng (2018, p. 4)

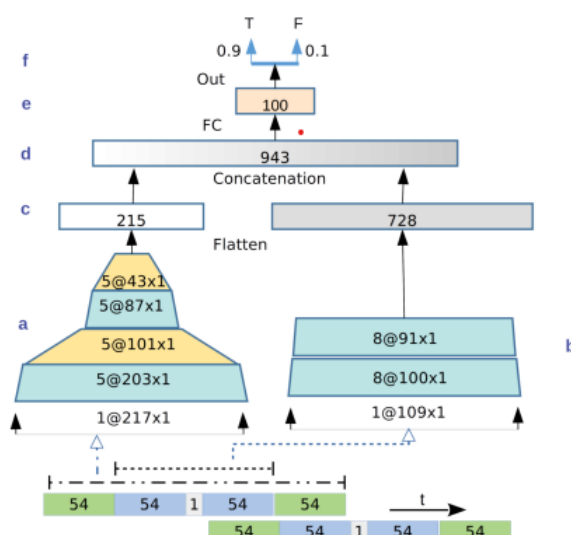
Nesse modelo, os dados utilizados no treinamento do modelo foram redimensionados para uma amostragem de 360Hz e as amostras foram segmentadas em 56 pontos com 22 antes do pico R e 33 após o pico R. Na entrada do modelo foi realizada uma diferenciação entre duas amostras adjacentes (XIANG; LIN; MENG, 2018, p. 4).

Com este modelo CNN, Xiang, Lin e Meng (2018) alcançaram acurácia de 99.68%, sensibilidade de 99.77%, taxa de predição positiva de 99.91% e taxa de detecção de erros de 0.32%. A arquitetura proposta apresentou maior robustez ao ruído e melhor desempenho na detecção de arritmias. Apesar do ruído reduzir o desempenho geral do modelo, eles concluíram que o modelo poderia ser utilizado na detecção de anor-

malidades cardíacas com ou sem a presença do ruído. No entanto, os autores não realizaram a divisão inter-paciente nos conjuntos de treinamento e teste.

Do mesmo modo Habib, Karmakar e Yearwood (2019, p. 93277) apresentaram um modelo CNN elaborado a partir de camadas convolucionais de uma dimensão, utilizando uma rede de dois canais de abstração. Os dados treinados foram obtidos por segmentação das amostras originais com base na detecção do pico R do complexo QRS (Figura 11). Os autores realizaram a diferenciação das amostras adjacentes segmentadas em 109 pontos e redimensionadas para $360Hz$, sem efetuar a divisão dos dados pela abordagem inter-paciente.

Figura 11 – Arquitetura proposta por Habib, Karmakar e Yearwood (2019)

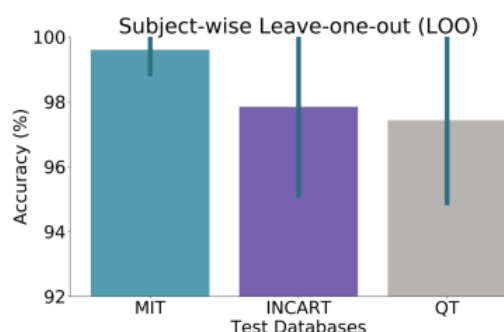


Fonte: Habib, Karmakar e Yearwood (2019, p.93276)

Além do mais, Habib, Karmakar e Yearwood (2019, p. 93281) investigaram a capacidade da generalização do modelo CNN na detecção do complexo QRS em diferentes bancos de dados: MITDB, *St. Petersburg Institute of Cardiological Technics* (INCART) e *European Society of Cardiology ST-T Database* (QT). O modelo proposto atingiu a acurácia de 99.22% (MITDB), 97.13% (INCART) e 96.25% (QT) (Figura 12).

Habib, Karmakar e Yearwood (2019, p. 93284) concluíram que, utilizando diferentes bancos de dados, a capacidade de generalização do modelo CNN não aumentou com a adição de mais amostras no treinamento para obter melhor acurácia na detecção de arritmias.

Figura 12 – Resultados em diferentes bancos de dados do modelo CNN proposto por Habib, Karmakar e Yearwood (2019)



Fonte: Habib, Karmakar e Yearwood (2019, p. 93281)

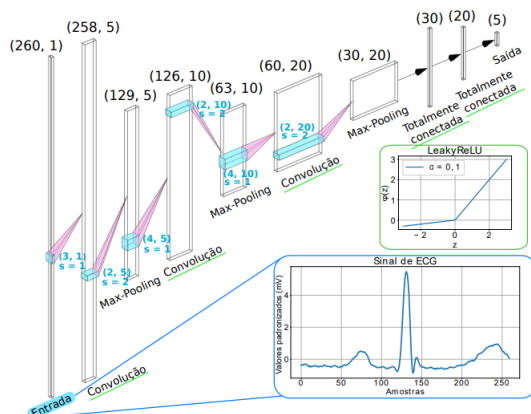
Por outro lado, Acharya *et al.* (2017, p. 85) propuseram o emprego de modelos CNN para detecção automática empregando diferentes segmentos de sinais de ECG. A arquitetura consistia de 11 camadas convolucionais de uma dimensão com quatro saídas para detectar diferentes classes de fibrilação cardíaca. Os autores utilizaram sinais segmentados em dois e cinco segundos de duração sem a detecção do complexo QRS, utilizando amostras redimensionadas para $250Hz$ e normalizadas pela estratégia *Z-Score* do MITDB e da *Creighton University*.

Ao contrário de outros autores, Acharya *et al.* (2017, p. 88) analisaram sinais de ECG em modelos CNN utilizando segmentos temporais reduzidos, ao invés de classificação no batimento cardíaco. Segundo os autores, os cardiologistas analisam segmentos de pequenas durações de sinais de ECG em vez dos batimentos cardíacos no diagnóstico da fibrilação cardíaca. Os autores concluíram que seria mais realista utilizar segmentos de ECG de curta duração em modelos CNN para a detecção automática de fibrilação cardíaca.

O modelo de Acharya *et al.* (2017, p. 88), alcançou os seguintes desempenhos: (1) acurácia (92.50%), sensibilidade (98.09%) e especificidade (93.13%) em segmentos de dois segundos e; (2) acurácia (94.90%), sensibilidade (99.13%) e especificidade (81.44%) em segmentos de cinco segundos. Os autores concluíram que o modelo proposto, baseado em sinais de curta duração, dispensa o processamento de detecção do complexo QRS.

Nagata, Candido e Silva (2021, p.1) empregaram um modelo CNN de uma dimensão para avaliar os sinais de ECG na classificação de arritmias morfológicas, extraídos do MITDB. Os autores segmentaram os dados em 260 amostras em torno do complexo QRS e efetuaram a normalização *Z-Score* (Figura 13).

Figura 13 – Arquitetura proposta por Nagata, Candido e Silva (2021)



Fonte: Nagata, Candido e Silva (2021, p. 2)

Neste estudo, Nagata, Candido e Silva (2021, p. 1) avaliaram o efeito da divisão de dados, intra e inter-paciente. Os autores obtiveram uma acurácia de 98.8%, sem efetuar a divisão de dados entre indivíduos (abordagem intra-paciente) e 88.2%, com a divisão de dados (abordagem inter-paciente).

Uma vez que os modelos classificadores possuem dificuldade no aprendizado quando uma classe é dominante em relação às outras, Mousavi e Afghah (2019, p. 1308) propuseram uma arquitetura RNN utilizando o algoritmo de *data augmentation* SMOTE aplicado ao MITDB para reduzir o desbalanceamento das classes minoritárias, considerando as abordagens intra e inter-paciente. Os autores obtiveram acurácia de 99,53%, na abordagem inter-paciente, e 99,92%, na abordagem intra-paciente.

4 METODOLOGIA

Este capítulo apresenta os procedimentos utilizados na estruturação, análise e interpretação dos sinais de ECG, utilizando o MITDB e descreve as arquiteturas dos modelos CNN desenvolvidos neste trabalho. Além do mais, detalha os procedimentos adotados no treinamento e o processo de avaliação do desempenho dos modelos.

4.1 PRÉ-PROCESSAMENTO

Utilizou-se os dados da derivação MLII do MITDB, retirando-se as gravações 102 e 104 por não conterem os dados da derivação e a gravação 107 por conter apenas dados não classificados. De acordo com AAMI, empregou-se as classes de arritmias cardíacas descritas na Tabela 6. Os dados da classe Q, classificados como desconhecidos e/ou não-classificados, foram rejeitados.

Tabela 6 – Número de amostras por classe de arritmia segundo a AAMI

Classes AAMI	Total de amostras MITDB
Batimento Normal (N)	89924
Batimento Ectópico Supraventricular (S)	3023
Batimento Ectópico Ventricular (V)	7052
Batimento de Fusão (F)	802
Não Classificados (Q)	263

Fonte: MITDB

A partir da abordagem inter-paciente, as gravações de ECG por indivíduo foram divididas nos conjuntos de treinamento, validação e teste. Assim, selecionou-se a seguinte proporção de gravações: 29 gravações para o conjunto de treinamento, 7 para validação, e 9 para teste (Tabela 7).

Tabela 7 – Relação das gravações dos sinais de ECG por conjunto separado

Conjunto de Dados	Gravações
Treinamento	103, 106, 108, 111, 112, 114, 115, 117, 118, 119, 121, 122, 123, 200, 202, 207, 209, 210, 212, 213, 214, 215, 219, 220, 222, 223, 230, 231, 233
Validação	101, 116, 124, 203, 205, 217, 228
Teste	100, 105, 109, 113, 201, 208, 221, 232, 234

Fonte: Autor (2022)

Todas as gravações foram redimensionadas para uma amostragem de $128Hz$, retirando a linha de base com duas filtragens medianas em janelas temporais de $200ms$ seguido por uma janela de $600ms$ (SHARMA, 2020). As amostras foram segmentadas em 63 pontos antes e 64 pontos após o pico R, totalizando um tamanho fixo de $63 + 1 + 64 = 128$ pontos. A Tabela 8 apresenta o total de amostras categorizadas entre os conjuntos de treinamento, validação e teste.

Tabela 8 – Total de amostras categorizadas por conjunto de dados

Classe	Treinamento	Validação	Teste
N	59932	12630	17362
S	1365	48	1610
V	4303	1081	1668
F	409	17	376

Fonte: Autor (2022)

4.2 ARQUITETURAS DOS MODELOS CNN

Para desenvolver as arquiteturas de modelos CNN deste trabalho, utilizou-se a linguagem de programação *Python 3.7* em conjunto com a plataforma *Tensorflow* e *Keras*, ambos na versão 2.7, no sistema de computação em nuvem *Google Colab* com aceleradores de *GPU*. O Anexo A detalha os recursos computacionais empregados neste trabalho.

O primeiro modelo, baseado numa arquitetura CNN com dois canais (DC-CNN), foi desenvolvido a partir dos modelos propostos por Habib, Karmakar e Yearwood (2019, p. 93277) e Xiang, Lin e Meng (2018, p. 4). Esse modelo apresenta uma arquitetura com dois canais de estruturas convolucionais, em paralelo, de uma dimensão especializado na extração de diferentes características com: um canal mais raso de camadas convolucionais que extrai informações mais generalistas dos sinais de ECG; e um canal mais profundo que extrai informações mais detalhadas. O Anexo B apresenta o diagrama da arquitetura do modelo DC-CNN.

O modelo DC-CNN realiza a concatenação entre os dois canais utilizando uma camada densa de 100 neurônios interconectados, uma camada de *dropout*, uma função de ativação *softmax*, a função custo *categorical crossentropy*. Diferentemente de Habib, Karmakar e Yearwood (2019, p. 93277), o modelo DC-CNN utiliza uma saída de quatro unidades (Tabela 9). O Anexo C detalha as diferenças entre o modelo DC-CNN e o proposto por Habib, Karmakar e Yearwood (2019).

Tabela 9 – Estrutura do modelo DC-CNN

Estrutura	Camada	N° de neurônios	Kernel/Pool Size
1° Canal	<i>Convolutacional 1D</i>	217	15
	<i>MaxPooling</i>	-	2
	<i>BatchNormalization</i>	-	-
	<i>Convolutacional 1D</i>	101	15
	<i>MaxPooling</i>	-	2
	<i>BatchNormalization</i>	-	-
2° Canal	<i>Convolutacional 1D</i>	109	10
	<i>Convolutacional 1D</i>	100	10
Concatenação	Densa	100	-
	<i>BatchNormalization</i>	-	-
	<i>Dropout (0.5)</i>	-	-
	Densa	4	-

Fonte: Autor (2022)

O segundo modelo, baseado numa arquitetura CNN de uma dimensão e um canal de processamento, foi utilizado como referência para a análise comparativa do desempenho do modelo DC-CNN. Desenvolveu-se o modelo de referência ou *baseline* a partir de um modelo proposto por Acharya *et al.* (2017, p. 85), composto por quatro camadas convolucionais e de *pooling* intercaladas. Possui, ainda, uma camada de *dropout*, camadas densas (30, 10 neurônios), uma saída de quatro unidades com uma função de ativação *softmax* e a função custo *categorical crossentropy*. Diferentemente do modelo apresentado por Acharya *et al.* (2017, p. 85), o modelo CNN *baseline* apresenta camadas de normalização (Tabela 10). O Anexo C detalha as diferenças entre o modelo CNN *baseline* e o proposto por Acharya *et al.* (2017).

Tabela 10 – Estrutura do modelo CNN *baseline*

Camadas	N° de neurônios	Kernel/Pool Size
<i>Convolutacional 1D</i>	474	27
<i>BatchNormalization</i>	-	-
<i>Convolutacional 1D</i>	224	14
<i>BatchNormalization</i>	-	-
<i>MaxPooling</i>	-	2
<i>Convolutacional 1D</i>	110	3
<i>BatchNormalization</i>	-	-
<i>Convolutacional 1D</i>	52	4
<i>BatchNormalization</i>	-	-
<i>MaxPooling</i>	-	2
<i>Dropout (0,5)</i>	-	-
Densa	30	-
Densa	10	-
Densa	4	-

Fonte: Autor (2022)

4.3 TREINAMENTO DOS MODELOS CNN

Empregou-se o algoritmo *Hyperband*¹ (LI, L. *et al.*, 2018) e os algoritmos disponibilizados pela biblioteca *Keras Tuner*² (O'MALLEY *et al.*, 2019) para encontrar a melhor taxa de aprendizagem dos modelos entre as opções selecionadas de 0.01, 0.005, 0.001, 0.0005 e 0.0001.

Visto que as amostras do MITDB não são uniformemente distribuídas, isto é, o número de amostras da classe N (89924 amostras) é muito maior que as outras categorias, especialmente em relação à classe F (802 amostras), empregou-se estratégias de *data augmentation* para reduzir o desbalanceamento entre as classes.

Para avaliar o desempenho dos modelos DC-CNN e CNN *baseline*, realizou-se o treinamento em dois experimentos em cenários distintos, com e sem *data augmentation*, nos conjuntos de treinamento e validação:

- Experimento 1: Avaliou o desempenho do modelo DC-CNN em relação ao CNN *baseline*, sem o uso de algoritmos de *data augmentation*; e
- Experimento 2: Investigou a eficácia da aplicação de ferramentas de *data augmentation*, utilizando os algoritmos SMOTE e ROS, no modelo DC-CNN.

No aplicação do algoritmo SMOTE utilizou-se o método *3-nearest neighbors* e, no algoritmo ROS, um gerador de números aleatórios como 42 para base para reprodução dos resultados, com as bibliotecas oferecidas por Lemaître, Nogueira e Aridas (2017).

Durante os experimentos, treinou-se os modelos por 20 épocas utilizando amostras de treinamento e validação em conjunto, um *batch size* de 32 amostras por época e otimizador *Adam* para ajustar os pesos aprendidos durante o treinamento. Utilizou-se a abordagem inter-paciente durante todas as simulações. Ao final dos treinamentos, os modelos foram testados utilizando as nove gravações separadas anteriormente no conjunto de testes.

4.4 AVALIAÇÃO DO DESEMPENHO DOS MODELOS CNN

Nos experimentos realizados empregou-se as métricas de avaliação de interesse para na análise comparativa entre os modelos DC-CNN e CNN *baseline* sem *data augmentation* e avaliação do desempenho do modelo DC-CNN com *data augmentation*.

A partir dos resultado dos testes, extraiu-se as respectivas matrizes de confusão para as grandezas de interesse: TP, TN, FP³, e FN. A partir destas, determinou-se as seguintes métricas: (1) ACC, (2) SEN, (3) PRE e (4) F1. Do mesmo modo, construiu-se

¹ Algoritmo que acelera a busca através de uma variedade de otimizações.

² Biblioteca especializada na busca de hiperparâmetros.

³ Foi desconsiderado o número FP devido a classe F em relação à classe V.

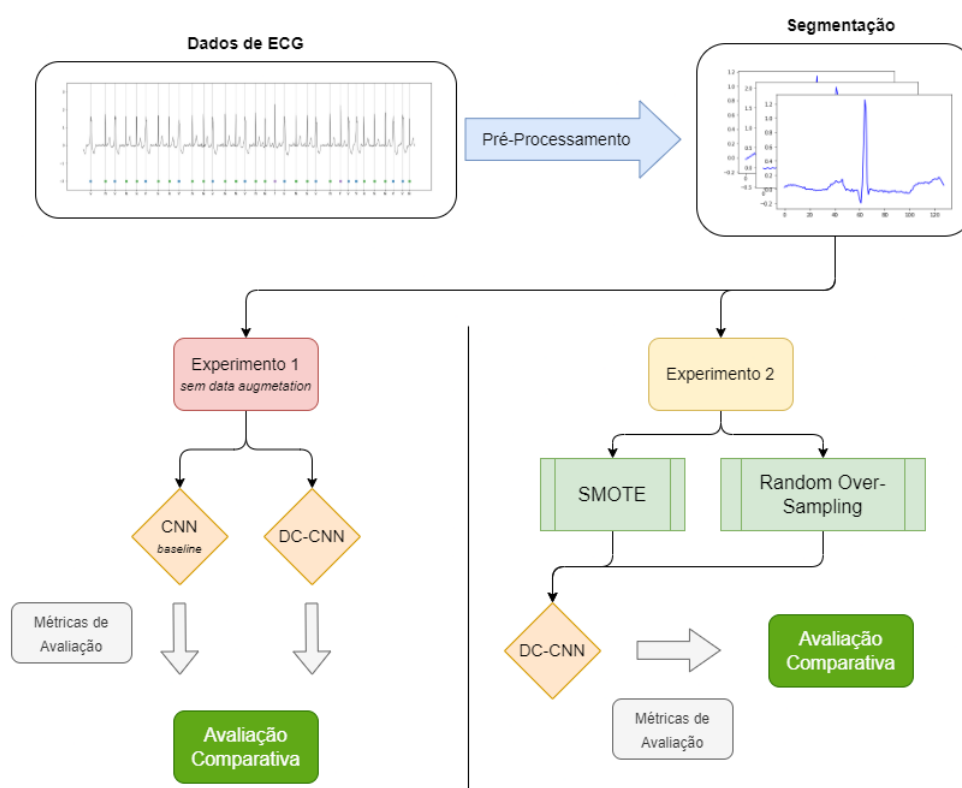
as curvas ROC e PR para determinar as métricas de avaliação AUROC e AUPRC. Na avaliação dos modelos a partir dessas curvas, utilizou-se uma classificação binária para as arritmias, isto é, separou-se os sinais de ECG em duas macro-classes: batimentos normais (classe N) e arritmicos (classes S, V e F).

Na análise comparativa dos modelos CNN avaliados, para obter o valor geral das métricas de sensibilidade e precisão, considerou-se a média ponderada calculada a partir dos valores de sensibilidade e precisão de cada classe em relação aos pesos com percentual de 82,6% (classe N), 7,9% (classe V), 7,7% (classe S) e 1,8% (classe F), extraídos do conjunto de teste selecionado do MITDB.

4.5 ESTRUTURA DA METODOLOGIA

As principais etapas da metodologia empregada neste trabalho: pré-processamento dos dados extraídos do MITDB, segmentação das gravações de ECG, treinamento dos modelos nos dois experimentos, emprego dos algoritmos de *data augmentation* e análise comparativa do desempenho dos modelos CNN estão detalhadas na Figura 14.

Figura 14 – Etapas da metodologia empregada



Fonte: Autor (2022)

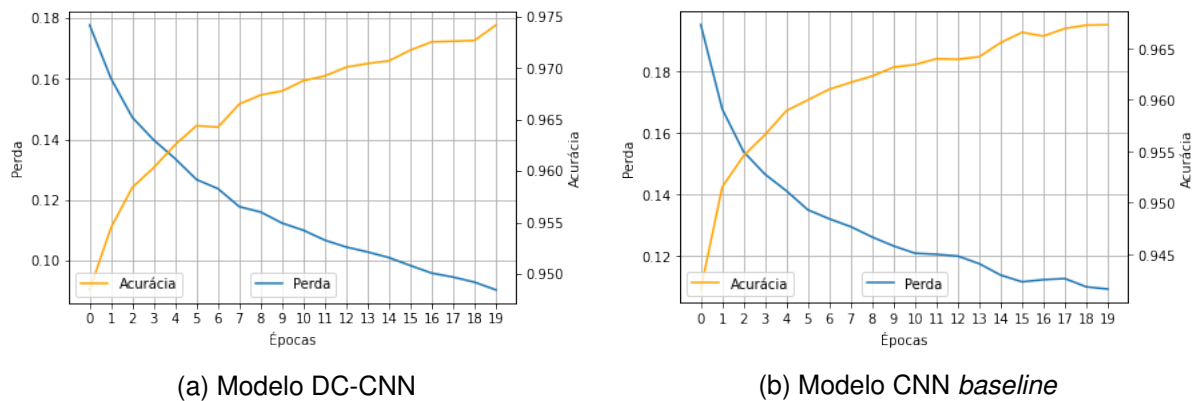
5 RESULTADOS

Este capítulo avalia o desempenho do modelo DC-CNN na classificação de arritmias em sinais de ECG. A partir da abordagem inter-paciente, realizou-se dois experimentos: o primeiro, que efetua a análise comparativa do desempenho do modelo DC-CNN em relação ao CNN *baseline* e; o segundo que investiga a eficácia da aplicação dos algoritmos SMOTE e ROS no modelo DC-CNN.

5.1 EXPERIMENTO 1: AVALIAÇÃO DO DESEMPENHO DOS MODELOS CNN

Inicialmente, para verificar o comportamento dos modelos estudados quanto à ocorrência de *overfitting*, extraiu-se as curvas de acurácia e perda do treinamento (Figura 15). Ambas as curvas do modelo apresentam uma tendência de estabilização a partir de 20 épocas.

Figura 15 – Curvas de acurácia e perda obtidas no treinamento



Fonte: Autor (2022)

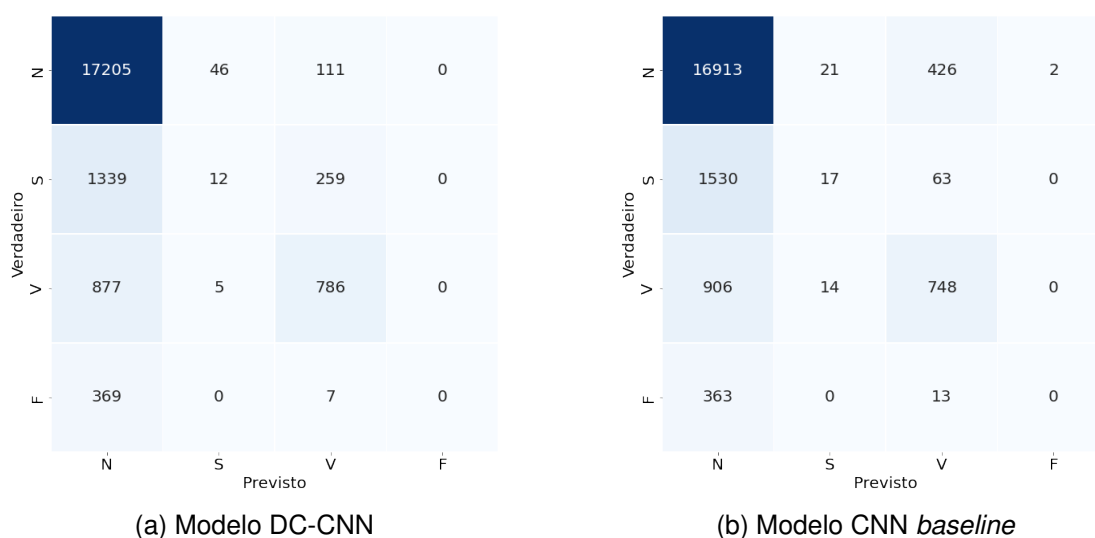
Para estimar as métricas de avaliação do desempenho de classificadores, obteve-se, durante os testes, as matrizes de confusão dos modelos estudados em relação às quatro classes de arritmia separadamente (Figura 16).

A partir dos dados dessas matrizes, a Tabela 11 apresenta o desempenho geral dos modelos DC-CNN e CNN *baseline*. Observa-se que o modelo DC-CNN superou o modelo de referência em todas as métricas aplicadas, com destaque para a métrica de acurácia (85,66% vs 84,11%).

Tabela 11 – Desempenho geral dos modelos DC-CNN e CNN *baseline*

Modelo	ACC (%)	SEN (%)	PRE (%)	F1 (%)
DC-CNN	85,66	86	79	81
CNN (<i>baseline</i>)	84,11	84	78	80

Fonte: Autor (2022)

Figura 16 – Matriz de confusão dos modelos DC-CNN e CNN *baseline* sem *data augmentation*

Fonte: Autor (2022)

Do mesmo modo, a partir dos dados das matrizes de confusão, extraiu-se os valores das métricas de sensibilidade, precisão e *F1-Score* em relação as quatro classes de arritmia (Tabela 12).

Tabela 12 – Desempenho dos modelos DC-CNN e CNN *baseline* por classe de arritmia

Modelo	Classe	SEN (%)	PRE (%)	F1 (%)
DC-CNN	N	99,09	86,94	92,62
	S	0,75	19,05	1,44
	V	47,12	67,76	55,59
	F	0	0	0
CNN <i>baseline</i>	N	84,54	97,41	90,52
	S	1,05	32,69	2,03
	V	44,84	59,84	51,27
	F	0	0	0

Fonte: Autor (2022)

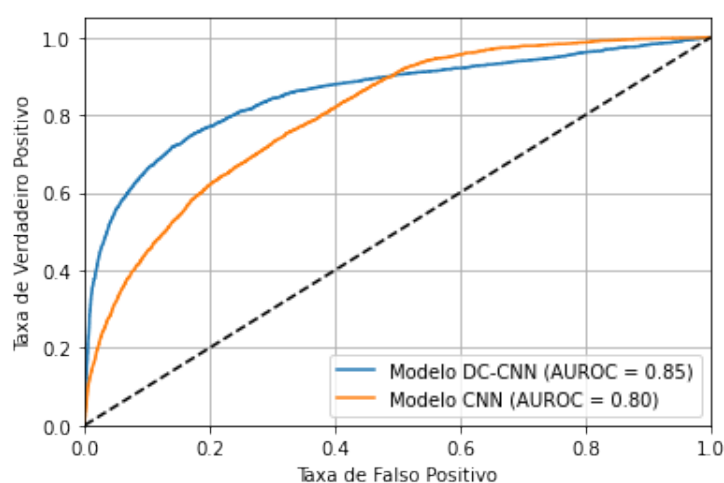
Observando os resultados apresentados na Tabela 12 verifica-se que, em relação a métrica F1, o modelo DC-CNN apresenta desempenho 2,1% superior na detecção da classe N (92,62% vs 90,52%) e 4,32% superior na detecção da classe V (55,59% vs 51,27%). Por outro lado, ambos os modelos apresentaram um baixo desempenho na classificação do tipo S (1,44% vs 2,03%). Em geral, a distinção entre as classes S e N é uma tarefa complexa para os modelos classificadores de arritmia devido ao padrão muito semelhante entre as mesmas (HE, 2020, p. 6), fato este que pode ter refletido nos baixos valores de F1 alcançados pelos modelos avaliados.

Em relação aos valores de sensibilidade, o modelo DC-CNN apresenta desempenho 14,55% superior na classe N (99,09% vs 84,54%) e 2,28% superior na classe V (47,12% vs 44,84%). Em relação à classe S, ambos os modelos apresentam uma sensibilidade de 0,75% e 1,05%, respectivamente, confirmando a baixa sensibilidade desta classe ao utilizar-se o MITDB (MOUSAVI; AFGHAH, 2019, p. 1308).

Em relação aos valores de precisão, o modelo DC-CNN apresenta desempenho 10,47% inferior na classe N (86,94% vs 97,41%) e 13,64% inferior na classe S (19,05% vs 32,69%) em relação ao modelo de referência. No entanto, o modelo DC-CNN apresenta desempenho 7,92% superior na classe V (67,79% vs 59,84%).

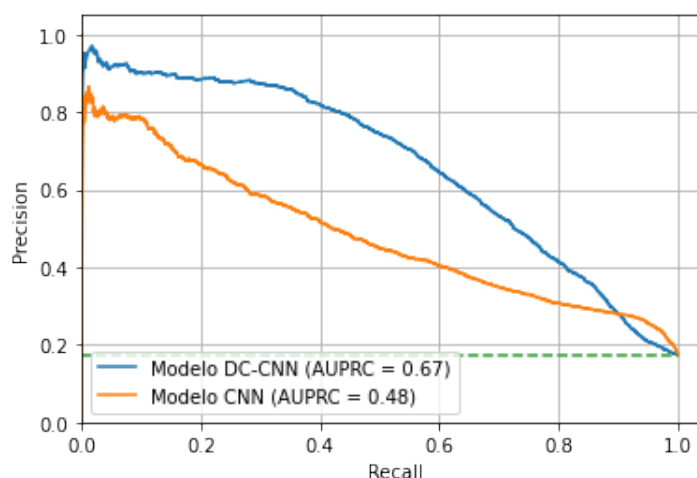
Na avaliação do desempenho dos modelos utilizando as curvas ROC e PR, efetuou-se uma análise binária, ou seja, o conjunto de dados foi dividido em duas macro-classes: normal (N) e arritmico (S, V e F). Como resultado, a curva ROC obtida revelou que, em relação aos valores da AUROC, o modelo DC-CNN apresentou melhor desempenho geral (0,85 vs 0,80), isto é, um ganho de 6,25% em relação ao modelo de referência (Figura 17).

Figura 17 – Curvas ROC dos modelos DC-CNN e CNN *baseline*



Fonte: Autor (2022)

No entanto, a curva ROC pode não ter discriminado corretamente as classes minoritárias de arritmia, pois as mesmas são derivadas de um conjunto de teste desbalanceado. Para minimizar este problema, utilizou-se a curva PR para avaliar o desempenho desses modelos (Figura 18). Observando os valores de AUPRC, constata-se que o modelo DC-CNN apresenta desempenho 39,6% superior em relação ao CNN *baseline* (0,67 vs 0,48).

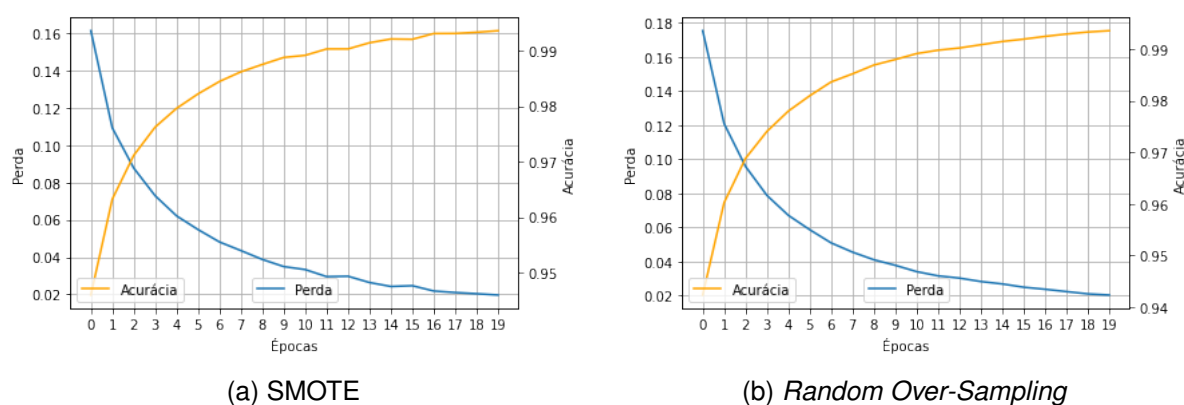
Figura 18 – Curvas PR dos modelos DC-CNN e CNN *baseline*

Fonte: Autor (2022)

5.2 EXPERIMENTO 2: DATA AUGMENTATION COM SMOTE E ROS

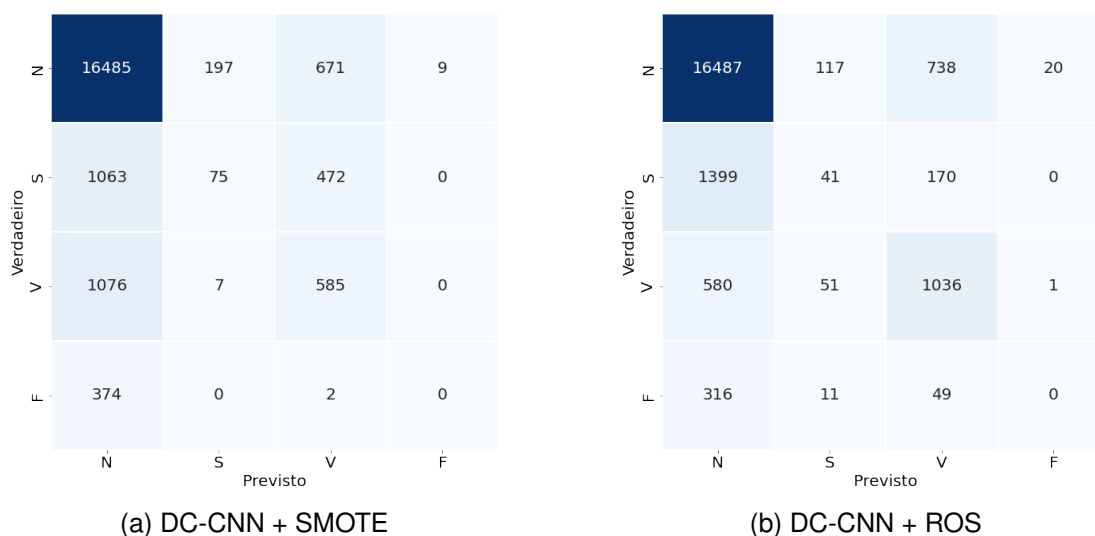
Como no experimento anterior, extraiu-se as curvas de acurácia e perda durante treinamento para verificar a existência de possível *overfitting*. Ambas as curvas do modelo DC-CNN apresentam uma tendência de estabilização a partir de 20 épocas (Figura 19).

Figura 19 – Curvas de acurácia e perda do modelo DC-CNN obtidas no treinamento



Fonte: Autor (2022)

Para uma avaliação mais detalhada dos efeitos da inclusão de *data augmentation* no desempenho do modelo DC-CNN na classificação de arritmia, extraiu-se as matrizes de confusão obtidas durante o teste (Figura 20).

Figura 20 – Matriz de confusão do modelo DC-CNN com *data augmentation*

Fonte: Autor (2022)

A partir dos dados das matrizes de confusão, a Tabela 13 apresenta os resultados com o emprego dos algoritmos SMOTE e ROS no modelo DC-CNN. Observando os resultados obtidos, verifica-se que a utilização dos algoritmos SMOTE e ROS reduziram o desempenho geral do modelo DC-CNN. Aparentemente a aplicação desses algoritmos falha na geração de novas amostras das classes minoritárias no conjunto de treinamento, refletindo numa redução dos valores das métricas de avaliação.

Tabela 13 – Desempenho geral do modelo DC-CNN com e sem *data augmentation*

Modelo	ACC (%)	SEN (%)	PRE (%)	F1 (%)
Sem <i>data augmentation</i>	85,66	86	79	81
SMOTE	81,58	82	76	78
ROS	83,57	84	78	80

Fonte: Autor (2022)

Para uma investigação mais detalhada da eficácia da aplicação dos algoritmos SMOTE e ROS, a Tabela 14 apresenta os resultados do desempenho do modelo DC-CNN para cada uma das classes minoritárias de arritmia, com e sem *data augmentation*.

Em relação a F1, a classe minoritária S apresentou desempenho superior com SMOTE (7,94% vs 1,44%) e ROS (4,48% vs 1,44%). A similaridade entre as classes S e N é um dos principais problemas para os classificadores de arritmias (HE, 2020, p. 5), no entanto, o emprego do algoritmo SMOTE melhorou a detecção da classe S em 6,5%.

Em termos de precisão, a aplicação do algoritmo SMOTE melhorou a detecção da classe S em 7,83% (26,88% vs 19,05%) e manteve um desempenho similar ao do modelo DC-CNN sem *data augmentation* (86,77% vs 86,94%).

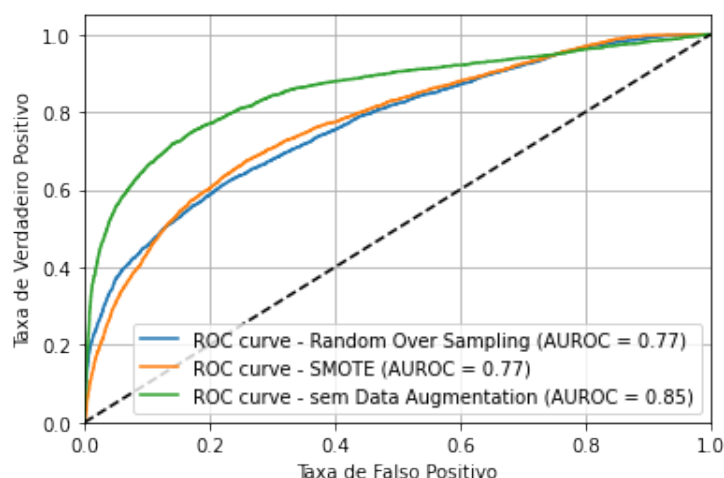
Tabela 14 – Desempenho do modelo DC-CNN, com e sem *data augmentation*, por classe de arritmia

Modelo DC-CNN	Classe	SEN(%)	PRE (%)	F1 (%)
Sem <i>data augmentation</i>	N	99,09	86,94	92,62
	S	0,75	19,05	1,44
	V	47,12	67,76	55,59
	F	0	0	-
SMOTE	N	94,95	86,77	90,68
	S	4,66	26,88	7,94
	V	35,07	33,81	34,43
	F	0	0	0
Random Over-Sampling	N	94,96	87,78	91,23
	S	2,55	18,64	4,48
	V	62,11	51,98	56,60
	F	0	0	-

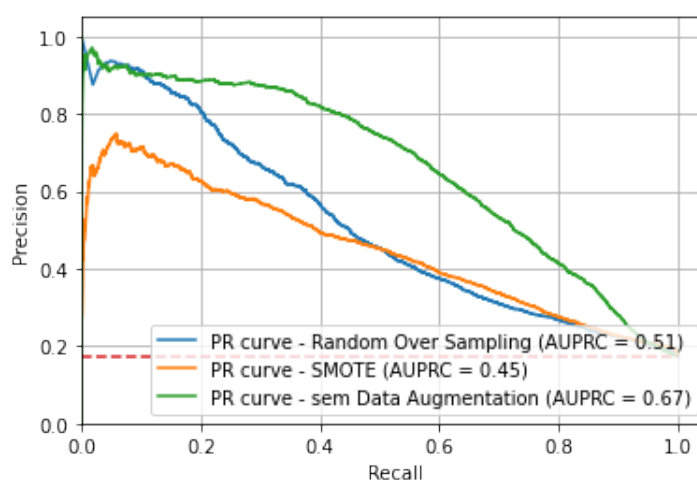
Fonte: Autor (2022)

Em termos de sensibilidade, o modelo DC-CNN sem *data augmentation* apresenta o melhor desempenho na classe N (99,09%) e o modelo DC-CNN com SMOTE apresenta um desempenho 3,91% superior na classe S (4,66% vs 0,75%). Por outro lado, a aplicação do algoritmo ROS aumentou a detecção da classe V em 14,99% (62,11% vs 47,12%) em relação ao modelo DC-CNN sem *data augmentation*.

Na avaliação do desempenho do modelo DC-CNN com algoritmos de *data augmentation*, as curvas ROC sugerem que, em relação aos valores da AUROC, o emprego tanto do algoritmo SMOTE como o ROS degradaram o desempenho geral do modelo DC-CNN, refletindo na redução dos valores de AUROC em até 9,41% (0,85 vs 0,77) (Figura 21). Igualmente, em relação às curvas PR, o emprego do algoritmo SMOTE reduziu o valor da AUPRC do modelo DC-CNN em 32,8% (0,67 vs 0,45) e o ROS reduziu a AUPRC em 23,8% (0,67 vs 0,51) (Figura 22).

Figura 21 – Comparação das curvas ROC do modelo DC-CNN com *data augmentation*

Fonte: Autor (2022)

Figura 22 – Comparação das curvas PR do modelo DC-CNN com *data augmentation*

Fonte: Autor (2022)

5.3 DISCUSSÃO DOS RESULTADOS

A avaliação comparativa do modelo DC-CNN com os modelos CNN apresentados na Revisão da Literatura (Tabela 15) revelou que os autores que utilizaram a abordagem intra-paciente alcançaram valores mais elevados nas métricas de acurácia, sensibilidade e precisão. Entretanto, o emprego desta abordagem, ou seja, quando os conjuntos de treinamento e teste provém de um mesmo grupo de indivíduos, os resultados alcançados não são clinicamente realistas, pois as características dos sinais de ECG são distintas para cada indivíduo (NAGATA; CANDIDO; SILVA, 2021, p. 1).

Tabela 15 – Resumo dos trabalhos relacionados sobre modelos CNN classificadores de arritmias

Autores	Banco de Dados	N° de Classes	Abordagem	Deteccção QRS	ACC (%)	SEN (%)	PRE (%)
Xiang, Lin e Meng (2018)	MITDB	2	Intra-paciente	SIM	99,68	97,7	99,91
	INCART	2	Intra-paciente	SIM	99,75	99,86	99,89
Habib, Karmakar e Yearwood (2019)	MITDB	2	Intra-paciente	SIM	99,60	99,22	99,38
	INCART	2	Intra-paciente	SIM	97,85	97,13	97,70
	QT	2	Intra-paciente	SIM	97,42	96,25	96,81
Acharya <i>et al.</i> (2017)	MITDB	4	Intra-paciente	NÃO	92,50	98,09	93,13
Nagata, Candido e Silva (2021)	MITDB	5	Inter-paciente	SIM	88,2	-	-
		5	Intra-paciente	SIM	98,8	-	-
Autor, 2022	MITDB	4	Inter-paciente	SIM	85,66	86,0	79,0

Fonte: Autor (2022)

Em termos gerais, utilizando a abordagem inter-paciente, os resultados experimentais deste trabalho, mostram que a arquitetura do modelo DC-CNN superam o modelo CNN *baseline* em todas as métricas gerais, alcançando acurácia de 85,66%, resultado semelhante ao do modelo CNN inter-paciente proposto por Nagata, Candido e Silva (2021) (Tabela 15).

Entretanto, a maioria dos modelos CNN revisados neste trabalho empregaram a abordagem intra-paciente, ou seja, utilizaram amostras dos mesmos pacientes tanto no treinamento como no teste, como foi o caso de Habib, Karmakar e Yearwood (2019) e Xiang, Lin e Meng (2018). Por esta razão, a Tabela 16 apresenta a análise comparativa entre o modelo DC-CNN, com e sem *data augmentation*, e o proposto por Nagata, Candido e Silva (2021) por classe de arritmia, os quais utilizaram a abordagem inter-paciente.

Os resultados das classes Q e F foram omitidos pois, segundo a AAMI, a classe Q representa os sinais não-classificados e os modelos não detectaram a classe F no conjunto de teste.

Tabela 16 – Comparação do desempenho entre o modelo DC-CNN e o de Nagata, Candido e Silva (2021), por classe de arritmia

Autores	Modelo	ACC (%)	N			S			V		
			SEN (%)	PRE (%)	F1 (%)	SEN (%)	PRE (%)	F1 (%)	SEN (%)	PRE (%)	F1 (%)
Autor (2022)	DC-CNN	85,66	99,09	86,94	92,62	0,75	19,05	1,44	47,12	67,76	55,59
	DC-CNN + SMOTE	81,58	94,95	86,77	90,68	4,66	26,88	7,94	35,07	33,81	34,43
	DC-CNN + ROS	83,57	94,96	87,78	91,23	2,55	18,64	4,48	62,11	51,98	56,60
Nagata, Candido e Silva (2021)	CNN	88,2	93,2	93,8	93,50	2,9	9,0	4,39	79,6	66,7	72,58

Fonte: Autor (2022)

Observando a Tabela 16, constata-se que na detecção dos batimentos normais N, o modelo DC-CNN apresenta o melhor desempenho global em termos de sensibilidade, ou seja, 99,09%, superando em 5,89% o modelo de Nagata, Candido e Silva (2021). Na métrica *F1-Score*, o modelo DC-CNN, sem *data augmentation*, apresenta desempenho similar ao de Nagata, Candido e Silva (2021) na classe N (92,62 vs 93,50).

Considerando que a similaridade entre as classes de batimento S e N é um dos principais desafios para os classificadores de arritmia, o modelo DC-CNN com SMOTE apresenta o melhor desempenho geral na detecção da classe S, superando os resultados do modelo CNN de Nagata, Candido e Silva (2021) nas métricas de *F1-Score* (+3,55%), precisão (+17,88%) e sensibilidade (+1,76%).

De uma forma geral, conhecer a acurácia do modelo classificador não é suficiente para ter-se uma ideia completa do desempenho de um classificador quando treinados em bancos de dados desbalanceados. O conjunto de teste selecionado neste estudo, a partir dos dados extraídos do MITDB é extremamente desbalanceado pois cerca de 82,6% dos batimentos cardíacos pertencem à classe N, enquanto os restantes 7,9%, 7,7% e 1,8% pertencem as classes minoritárias V, S e F, respectivamente. Por esta razão, em modelos classificadores de arritmias utilizando conjunto de dados desbalanceados, onde as classes minoritárias são mais importantes, recomenda-se o uso das curvas ROC para efetuar a comparação de desempenho de classificadores.

Considerando uma classificação binária, ou seja, batimentos normais (classe N) e arritmicos (classes S, V e F), a análise das curvas ROC revelam que a arquitetura DC-CNN apresenta desempenho 6,25% superior ao modelo de referência, com AU-ROC de 0,85. Assim, conforme a diretriz de avaliação do desempenho de modelos de classificação sugerida por Hosmer, Lemeshow e Sturdivant (2013, p. 177) infere-se que modelo DC-CNN pode ser considerado um excelente classificador de arritmias.

Uma vez que as curvas ROC podem não ter discriminado corretamente as classes minoritárias de arritmia, pois as mesmas derivam de um conjunto de teste desbalanceado, a análise das curvas PR, considerando uma classificação binária, revelam que o modelo DC-CNN apresenta desempenho 39,6% superior ao modelo de referência, com AUPRC de 0,67.

Ao investigar a eficácia do uso de algoritmos de *data augmentation* no modelo DC-CNN revela uma redução na acurácia de até 4,08%. Além do mais, o algoritmo SMOTE reduziu o valor da AUPRC em 32,8% e o ROS reduziu a AUPRC em 23,8%. Contudo, analisando o desempenho do modelo DC-CNN, o algoritmo SMOTE melhorou o desempenho na detecção da classe S nas métricas de *F1-Score* (+6,5%), precisão (+7,83%) e sensibilidade (+3,91%). O emprego do ROS melhorou a detecção da classe V em 14,99% na métrica de precisão.

Em resumo, a partir dos resultados da análise comparativa dos modelos desenvolvidos neste trabalho e a investigação da eficácia na aplicação de algoritmos de *data augmentation*, destacam-se os seguintes desafios para a avaliação do modelo DC-CNN:

1. O primeiro, relaciona-se ao problema do desbalanceamento dos dados. O conjunto de teste selecionado neste estudo, a partir dos dados extraídos do MITDB, é extremamente desbalanceado, pois cerca de 82,6% dos batimentos cardíacos pertencem à classe N e, portanto, a ocorrência de batimentos com arritmias, especialmente a classe F (1,8%), é um evento raro. Assim, a escassez de dados de classes minoritárias de arritmias é um obstáculo para os modelos classificadores reconhecerem os padrões arrítmicos;
2. O segundo, refere-se às divergências entre as abordagens intra-paciente e inter-paciente em relação aos resultados obtidos nas métricas de acurácia, sensibilidade e precisão. Na abordagem intra-paciente os modelos classificadores tendem a aprender as particularidades dos batimentos de um indivíduo do conjunto de teste durante o treinamento, favorecendo sua detecção pelo classificador. Na abordagem inter-paciente, o classificador generaliza os padrões encontrados no grupo de treinamento e infere sobre um grupo de teste desconhecido pelo classificador. Essas divergências prejudicam a execução de uma avaliação comparativa mais precisa entre os modelos classificatórios propostos na literatura e;
3. O terceiro, trata-se da diferenciação da classe S de arritmia em relação aos batimentos normais. A detecção dessa classe é uma tarefa complexa para os modelos classificadores devido seu padrão muito semelhante aos batimentos normais, sendo um dos maiores problemas dos métodos de detecção de arritmias (HE, 2020, p. 2). No entanto, ao investigar a eficácia do uso de algoritmos de *data augmentation* no modelo DC-CNN, constatou-se que o emprego de algoritmos SMOTE melhorou a detecção da classe S em todas as métricas de avaliação.

6 CONCLUSÃO

Diversos estudos vêm sendo desenvolvidos sobre modelos e metodologias para a detecção de classes de arritmias cardíacas utilizando as arquiteturas DL em diferentes bancos de dados. Neste sentido, este trabalho revisou estudos no estado-da-arte sobre o desenvolvimento e aplicação de modelos CNN de uma dimensão para classificar e/ou detectar arritmias em sinais de ECG.

No entanto, pouca pesquisa tem sido conduzida na classificação de arritmias utilizando arquiteturas CNN multicanais. Assim sendo, este trabalho avaliou o desempenho de um modelo CNN com dois canais para a classificação automática de arritmias, o *Dual-Channel CNN* (DC-CNN) construído a partir de uma arquitetura com estruturas convolucionais de uma dimensão que extraem de características distintas dos sinais de ECG.

Assim, executou-se dois experimentos para avaliar o desempenho do modelo DC-CNN desenvolvido, incluindo as seguintes contribuições: aplicou a abordagem inter-paciente nos conjuntos de treinamento e teste; introduziu o uso do algoritmo *Hyperband* na busca da melhor taxa de aprendizagem; incorporou estratégias de *data augmentation* nas classes minoritárias do MITDB e; utilizou métricas de avaliação para a análise comparativa do modelo DC-CNN com os modelos CNN da literatura.

A análise dos modelos CNN apresentados na Revisão Literária revelou que os autores que utilizaram a abordagem intra-paciente nos conjuntos de treinamento e teste alcançaram resultados mais elevados nas métricas de acurácia, precisão e sensibilidade. Contudo, este trabalho adotou uma abordagem inter-paciente por considerá-la mais adequada nos diagnósticos de arritmias em cenários da vida real. Apesar de ser mais desafiador alcançar desempenhos superiores com essa abordagem, os resultados experimentais obtidos são mais confiáveis e significativos, uma vez que reproduz um ambiente similar ao do cardiologista quando o mesmo realiza o diagnóstico de um novo paciente.

Ao investigar a eficácia do uso de algoritmos SMOTE e ROS no modelo DC-CNN verificou-se que ambos apresentaram uma degradação no desempenho, sendo que esta degradação foi maior para o algoritmo SMOTE. Entretanto, considerando que a similaridade entre as classes de batimento cardíaco S e N é um dos principais desafios para os classificadores de arritmia, o uso do algoritmo SMOTE melhorou o desempenho geral do modelo DC-CNN na detecção da classe S em todas as métricas avaliadas.

De uma forma geral, o emprego de algoritmos de *data augmentation* são instrumentos eficientes na otimização do desempenho de modelos classificadores. Portanto, deve-se atentar para uma exaustiva análise dos efeitos desses algoritmos na detecção das classes minoritárias antes de incorporá-las aos modelos CNN classificadores de arritmia.

Em relação ao desempenho do modelo DC-CNN, empregando uma classificação binária, nas curvas ROC e PR, a arquitetura DC-CNN obteve AUROC de 0,85 e AUPRC de 0,67, desempenho 6,25% e 39,6% superior ao modelo de referência, respectivamente.

Em um sentido mais amplo, os experimentos executados neste trabalho mostraram que o modelo DC-CNN superou o modelo de referência em todas as métricas aplicadas, com destaque para a métrica de acurácia (85.66%), resultado semelhante ao de modelos CNN que empregam a abordagem inter-paciente. Assim, infere-se que o modelo DC-CNN avaliado neste trabalho tem potencial para efetuar a detecção de arritmias morfológicas em sinais de ECG, por meio da análise do complexo QRS.

Por fim, sugere-se a realização de estudos futuros sobre classificação automática de arritmias em sinais de ECG, incluindo:

- O desenvolvimento de outras arquiteturas DL, de um ou mais canais, tais como: CNN, RNN, etc.;
- A análise comparativa do desempenho de modelos DL utilizando outros bancos de dados de sinais de ECG, tais como: INCART, QT, *Physikalisch-Technische Bundesanstalt* (PTB), etc.;
- A avaliação comparativa do efeito de outros algoritmos de *data augmentation* nos modelos DL na detecção das classes minoritárias de arritmias;
- O emprego de outros hiperparâmetros para melhorar o desempenho de modelos DL, tais como: número de épocas, *batch size*, taxa de aprendizagem, etc.; e
- A análise de outras formas de segmentação e pré-processamento de sinais de ECG, como RRI e aplicação de técnicas de normalização.

REFERÊNCIAS

- ACHARYA, U. Rajendra *et al.* Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. **Information Sciences**, Elsevier Inc., v. 405, p. 81–90, set. 2017. ISSN 00200255. DOI: 10.1016/j.ins.2017.04.012. Citado nas pp. 36, 40, 50, 65.
- ATKIELSKI, Anthony. **Electrocardiography - Wikipedia**. [S.l.: s.n.], 2007. Disponível em: <<https://en.wikipedia.org/wiki/Electrocardiography>>. Citado na p. 24.
- BATISTA, Gustavo E. A. P. A.; PRATI, Ronaldo C.; MONARD, Maria Carolina. A study of the behavior of several methods for balancing machine learning training data. **ACM SIGKDD Explorations Newsletter**, Association for Computing Machinery (ACM), v. 6, p. 20–29, 1 jun. 2004. ISSN 1931-0145. DOI: 10.1145/1007730.1007735. Disponível em: <<https://dl.acm.org/doi/10.1145/1007730.1007735>>. Citado na p. 29.
- CHAWLA, N. V. *et al.* SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, jun. 2002. ISSN 1076-9757. DOI: 10.1613/jair.953. Disponível em: <<https://www.jair.org/index.php/jair/article/view/10302>>. Citado nas pp. 28, 29, 31.
- CHAZAL, Philip de; O'DWYER, Maria; REILLY, Richard B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. **IEEE transactions on bio-medical engineering**, v. 51, p. 1196–206, 7 jul. 2004. ISSN 0018-9294. DOI: 10.1109/TBME.2004.827359. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15248536>>. Citado na p. 28.
- CHOLLET, Francois *et al.* **Keras**. 2015. Disponível em: <<https://github.com/fchollet/keras>>. Citado na p. 63.
- CHOU, Sin-Yi. **Precision - Recall Curve, a Different View of Imbalanced Classifiers – Sin-Yi Chou – Data Science Enthusiast**. [S.l.: s.n.], abr. 2020. Disponível em: <<https://sinyi-chou.github.io/classification-pr-curve/>>. Citado nas pp. 32, 33.
- FENG, Hsin Yu; CHEN, Po Ying; HOU, Janpu. SR-ScatNet Algorithm for On-device ECG Time Series Anomaly Detection. *In*: DOI: 10.1109/SoutheastCon45413.2021.9401872. Citado na p. 16.
- GANAPATHY, Nagarajan; SWAMINATHAN, Ramakrishnan; DESERNO, Thomas M. Deep Learning on 1-D Biosignals: a Taxonomy-based Survey. **Yearbook of Medical Informatics**, Thieme Medical Publishers, v. 27, p. 98, 1 ago. 2018. ISSN 23640502. DOI: 10.1055/S-0038-1667083. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/30115218/>>. Citado na p. 15.
- GARCIA, Gabriel *et al.* Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO. **Scientific reports**, Nature Publishing Group, v. 7, p. 10543, 1 dez. 2017. ISSN 2045-2322. DOI: 10.1038/s41598-017-09837-3. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pubmed/28874683>><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5585360>>. Citado na p. 28.

GOLDBERGER, A L *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. **Circulation**, v. 101, e215–20, 23 jun. 2000. ISSN 1524-4539. DOI: 10.1161/01.cir.101.23.e215. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/10851218>>. Citado na p. 16.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado nas pp. 19, 21–23.

GUYTON, Arthur C.; HALL, John Edward. **Tratado de Fisiologia Médica**. 12. ed. [S.l.]: Elsevier, 2011. ISBN 9788535249804. Citado na p. 24.

HABIB, Ahsan; KARMAKAR, Chandan; YEARWOOD, John. Impact of ECG Dataset Diversity on Generalization of CNN Model for Detecting QRS Complex. **IEEE Access**, Institute of Electrical e Electronics Engineers Inc., v. 7, p. 93275–93285, 2019. ISSN 21693536. DOI: 10.1109/ACCESS.2019.2927726. Citado nas pp. 16, 35, 36, 39, 50, 51, 65.

HE, Jinyuan. **Automated Heart Arrhythmia Detection from Electrocardiographic Data**. 2020. Victoria University. Disponível em: <<https://vuir.vu.edu.au/41284/>>. Citado nas pp. 25, 26, 44, 47, 54.

HOLTER, N. J. **New methods for heart studies**. v. 134. [S.l.: s.n.], 1961. P. 1214–1220. Citado na p. 26.

HOSMER, David W.; LEMESHOW, Stanley.; STURDIVANT, Rodney X. **Applied logistic regression**. 3. ed. [S.l.]: Wiley-Blackwell, abr. 2013. ISBN 9780470582473. Citado nas pp. 32, 53.

JAMBUKIA, Shweta H.; DABHI, Vipul K.; PRAJAPATI, Harshadkumar B. Classification of ECG signals using machine learning techniques: A survey. *In*: p. 714–721. DOI: 10.1109/ICACEA.2015.7164783. Citado na p. 16.

JUN, Tae Joon *et al.* ECG arrhythmia classification using a 2-D convolutional neural network, abr. 2018. Disponível em: <<http://arxiv.org/abs/1804.06812>>. Citado nas pp. 17, 28.

KACHUEE, Mohammad; FAZELI, Shayan; SARRAFZADEH, Majid. ECG Heartbeat Classification: A Deep Transferable Representation. *In*: DOI: 10.1109/ICHI.2018.00092. Disponível em: <<https://ieeexplore.ieee.org/document/8419425/>>. Citado na p. 17.

KARPATHY, Andrej *et al.* Large-scale Video Classification with Convolutional Neural Networks. *In*: CVPR. [S.l.: s.n.], 2014. Disponível em: <<http://cs.stanford.edu/people/karpathy/deepvideo>>. Citado na p. 23.

KIRANYAZ, Serkan; INCE, Turker; GABBOUJ, Moncef. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. **IEEE Transactions on Biomedical Engineering**, IEEE Computer Society, v. 63, p. 664–675, 3 mar. 2016. ISSN 15582531. DOI: 10.1109/TBME.2015.2468589. Citado na p. 16.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. **Deep learning**. v. 521. [S.l.]: Nature Publishing Group, mai. 2015. P. 436–444. DOI: 10.1038/nature14539. Citado nas pp. 15, 21.

LEMAÎTRE, Guillaume; NOGUEIRA, Fernando; ARIDAS, Christos K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. **Journal of Machine Learning Research**, v. 18, n. 17, p. 1–5, 2017. Disponível em: <<http://jmlr.org/papers/v18/16-365.html>>. Citado na p. 41.

LI, Lisha *et al.* Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. **Journal of Machine Learning Research**, v. 18, n. 185, p. 1–52, 2018. Disponível em: <<http://jmlr.org/papers/v18/16-558.html>>. Citado na p. 41.

LI, Runchuan *et al.* Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network. **IEEE Access**, Institute of Electrical e Electronics Engineers Inc., v. 7, p. 109870–109883, 2019. ISSN 2169-3536. DOI: 10.1109/ACCESS.2019.2933473. Disponível em: <<https://ieeexplore.ieee.org/document/8790681/>>. Citado na p. 17.

LUQUE, Amalia *et al.* The impact of class imbalance in classification performance metrics based on the binary confusion matrix. **Pattern Recognition**, Elsevier Ltd, v. 91, p. 216–231, jul. 2019. ISSN 00313203. DOI: 10.1016/j.patcog.2019.02.023. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0031320319300950>>. Citado na p. 29.

MARKOULIDAKIS, Ioannis *et al.* Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. **Technologies**, MDPI AG, v. 9, p. 81, 4 nov. 2021. ISSN 2227-7080. DOI: 10.3390/technologies9040081. Disponível em: <<https://www.mdpi.com/2227-7080/9/4/81>>. Citado nas pp. 29–31.

MARS. **A Study of Imbalanced Data. An study of imbalanced data: the... | by Mars | Medium**. [S.l.: s.n.], fev. 2020. Disponível em: <<https://medium.com/analytics-vidhya/the-ultimate-guide-to-imbalanced-data-primer-1a15f7c895ee>>. Citado na p. 28.

MITCHELL, L. Brent. **Extrassístole ventricular (ESV) - Doenças cardiovasculares - Manuais MSD edição para profissionais**. [S.l.: s.n.], jul. 2019. Disponível em: <<https://www.msmanuals.com/pt-br/profissional/doen%C3%A7as-cardiovasculares/arritmias-e-doen%C3%A7as-de-condu%C3%A7%C3%A3o/extrass%C3%ADstole-ventricular-esv>>. Citado na p. 25.

MITCHELL, L. Brent. **Ritmos ectópicos supraventriculares - Doenças cardiovasculares - Manuais MSD edição para profissionais**. [S.l.: s.n.], jul. 2019.

Disponível em: <<https://www.msmanuals.com/pt-br/profissional/doen%C3%A7as-cardiovasculares/arritmias-e-doen%C3%A7as-de-condu%C3%A7%C3%A3o/ritmos-ect%C3%B3picos-supraventriculares>>. Citado na p. 25.

MOODY, G B; MARK, R G. The impact of the MIT-BIH arrhythmia database. en. **IEEE Eng. Med. Biol. Mag.**, Institute of Electrical e Electronics Engineers (IEEE), v. 20, n. 3, p. 45–50, mai. 2001. Citado nas pp. 24, 26, 27.

MOORE, Curtiss; KULKARNI, Nitin; JOGLAR, Jose A. Supraventricular tachycardia. **Cardiology Secrets**, Elsevier, p. 330–336, jan. 2017. DOI: 10.1016/B978-0-323-47870-0.00036-2. Citado na p. 25.

MOUSAVI, Sajad; AFGHAH, Fatemeh. Inter- and Intra- Patient ECG Heartbeat Classification for Arrhythmia Detection: A Sequence to Sequence Deep Learning Approach. *In*: p. 1308–1312. DOI: 10.1109/ICASSP.2019.8683140. Disponível em: <<https://ieeexplore.ieee.org/document/8683140/>>. Citado nas pp. 37, 45.

NAGATA, Natália; CANDIDO, Renato; SILVA, Magno T. M. O efeito da divisão de dados na classificação de arritmias usando redes neurais convolucionais. *In*: p. 1–2. Disponível em: <<https://biblioteca.sbrt.org.br/articlefile/2871.pdf>>. Citado nas pp. 17, 36, 37, 49–53.

NG, G André. Treating patients with ventricular ectopic beats. **Heart (British Cardiac Society)**, v. 92, p. 1707–12, 11 nov. 2006. ISSN 1468-201X. DOI: 10.1136/hrt.2005.067843. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17041126>><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1861260>>. Citado na p. 25.

O'MALLEY, Tom *et al.* **KerasTuner**. [S.l.: s.n.], 2019. <https://github.com/keras-team/keras-tuner>. Citado na p. 41.

OMS. **OMS revela principais causas de morte e incapacidade em todo o mundo entre 2000 e 2019 | As Nações Unidas no Brasil**. [S.l.: s.n.], 2020. Disponível em: <<https://brasil.un.org/pt-br/104646-oms-revela-principais-causas-de-morte-e-incapacidade-em-todo-o-mundo-entre-2000-e-2019>>. Citado na p. 15.

POLO, Tatiana Cristina Figueira; MIOT, Hélio Amante. Aplicações da curva ROC em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, Sociedade Brasileira de Angiologia e Cirurgia Vascular, v. 19, p. 1–4, 2020. ISSN 1677-7301. DOI: 10.1590/1677-5449.200186. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1677-54492020000100206&tlng=pt>. Citado na p. 31.

POWERS, David M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, out. 2020. Disponível em: <<http://arxiv.org/abs/2010.16061>>. Citado nas pp. 29, 30.

PROVOST, Foster; FAWCETT, Tom. Robust Classification for Imprecise Environments, set. 2000. Disponível em: <<http://arxiv.org/abs/cs/0009007>>. Citado na p. 31.

RAJPURKAR, Pranav *et al.* Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks, jul. 2017. Disponível em: <<http://arxiv.org/abs/1707.01836>>. Citado na p. 16.

S. LUZ, Eduardo José da *et al.* ECG-based heartbeat classification for arrhythmia detection: A survey. **Computer Methods and Programs in Biomedicine**, Elsevier Ireland Ltd, v. 127, p. 144–164, abr. 2016. ISSN 18727565. DOI: 10.1016/j.cmpb.2015.12.008. Citado nas pp. 15, 24.

SAITO, Takaya; REHMSMEIER, Marc. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. Edição: Guy Brock. **PloS one**, Public Library of Science, v. 10, e0118432, 3 mar. 2015. ISSN 1932-6203. DOI: 10.1371/journal.pone.0118432. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25738806>><<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4349800>>. Citado nas pp. 32, 33.

SHARMA, Nelson. **ECG Lead 2 Dataset PhysioNet (Open Access) | Kaggle**. [S.l.: s.n.], 2020. Disponível em: <<https://www.kaggle.com/nelsonsharma/ecg-lead-2-dataset-physionet-open-access>>. Citado na p. 39.

SOARES, Pablo Luiz Braga; SILVA, José Patrocínio Da. Aplicação de Redes Neurais Artificiais em Conjunto com o Método Vetorial da Propagação de Feixes na Análise de um Acoplador Direcional Baseado em Fibra Ótica. **Revista Brasileira de Computação Aplicada**, UPF Editora, v. 3, p. 58–72, 2 2011. ISSN 2176-6649. DOI: 10.5335/rbca.2011.1803. Citado na p. 19.

SRIVASTAVA, Nitish. **Improving Neural Networks with Dropout**. 2013. Disponível em: <http://www.cs.toronto.edu/~nitish/msc_thesis.pdf>. Citado na p. 22.

VEERAVALLI, Bharadwaj; DEEPU, Chacko John; NGO, DuyHoa. **Real-Time, Personalized Anomaly Detection in Streaming Data for Wearable Healthcare Devices**. [S.l.: s.n.], 2017. P. 403–426. DOI: 10.1007/978-3-319-58280-1_15. Citado na p. 15.

WU, Meng-Hsi; CHANG, Emily J.; CHU, Tzu-Hsuan. Personalizing a Generic ECG Heartbeat Classification for Arrhythmia Detection: A Deep Learning Approach. *In*: p. 92–99. DOI: 10.1109/MIPR.2018.00024. Disponível em: <<https://ieeexplore.ieee.org/document/8396980/>>. Citado na p. 17.

XIANG, Yande; LIN, Zhitao; MENG, Jianyi. Automatic QRS complex detection using two-level convolutional neural network. **BioMedical Engineering Online**, BioMed Central Ltd., v. 17, 1 jan. 2018. ISSN 1475925X. DOI: 10.1186/s12938-018-0441-4. Citado nas pp. 17, 23, 34, 39, 50, 51.

YE, Kun. **Inter-Patient Electrocardiogram Heartbeat Classification with 2-D Convolutional Neural Network**. 2018. University of Vtctoria. Disponível em: <<http://hdl.handle.net/1828/12586>>. Citado na p. 27.

ANEXO A – RECURSOS COMPUTACIONAIS

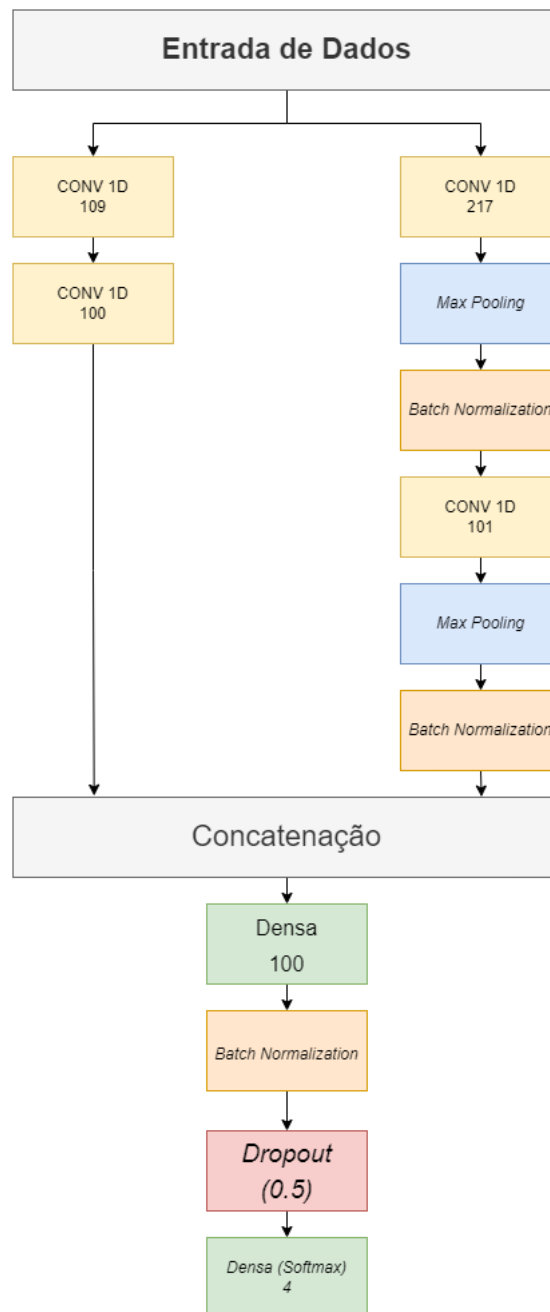
Tensorflow, trata-se de uma interface para a construção de modelos ML e implementação de para a execução destes modelos. Esta estrutura facilita o desenvolvimento de diversas técnicas para as mais variadas aplicações, escalando em diversos equipamentos, desde celulares e *tablets* até sistemas de larga escala que utilizam tecnologias de GPU e TPU para as inferências. A vantagem da utilização do *Tensorflow*, é a sua flexibilidade de se aplicar e expressar uma grande variedade de técnicas como treinamento e inferência de modelos DL, entre as mais variadas aplicações podemos citar a utilização para a criação de modelos para classificação de imagens, reconhecimento de voz, processamento de dados geográficos, etc.

A plataforma foi disponibilizada para o grande público em Novembro de 2015, como uma tecnologia *Open-Source* que facilitou a sua utilização e também cooperou para o rápido crescimento dela.

A biblioteca *Keras* (CHOLLET *et al.*, 2015), foi desenvolvida usando o *Tensorflow* como base, trata-se de uma maior abstração das estruturas ML, desenvolvida com o foco em rápida experimentação de diversos modelos. Esta biblioteca facilita o uso do *Tensorflow* para o melhor desempenho de todas as capacidades de escalabilidade para diversos sistemas DL.

ANEXO B – ARQUITETURA DO MODELO DC-CNN DESENVOLVIDO

Figura 23 – Arquitetura do modelo DC-CNN



Fonte: Autor (2022)

ANEXO C – COMPARAÇÃO DAS ESTRUTURAS DOS MODELOS CNN DESENVOLVIDOS

Tabela 17 – Comparação entre o modelo proposto por Habib, Karmakar e Yearwood (2019) e o modelo DC-CNN

Estrutura	Habib, Karmakar e Yearwood (2019)		Modelo DC-CNN	
	Camada	N° de neurônios	Camada	N° de neurônios
1° Canal	Convolutacional 1D	217	Convolutacional 1D	217
	<i>MaxPooling</i>	-	<i>MaxPooling</i>	-
	<i>BatchNormalization</i>	-	<i>BatchNormalization</i>	-
	Convolutacional 1D	101	Convolutacional 1D	101
	<i>MaxPooling</i>	-	<i>MaxPooling</i>	-
	<i>BatchNormalization</i>	-	<i>BatchNormalization</i>	-
2° Canal	Convolutacional 1D	109	Convolutacional 1D	109
	Convolutacional 1D	100	Convolutacional 1D	100
Concatenação	Densa	100	Densa	100
	<i>BatchNormalization</i>	-	<i>BatchNormalization</i>	-
	<i>Dropout</i> (0.5)	-	<i>Dropout</i> (0.5)	-
	Densa	1	Densa	4

Fonte: Autor (2022)

Tabela 18 – Comparação entre o modelo proposto por Acharya *et al.* (2017) e o modelo CNN *baseline*

Estrutura	Acharya <i>et al.</i> (2017)		Modelo CNN <i>baseline</i>	
	Camadas	N° de neurônios	Camadas	N° de neurônios
Canal Único	Convolutacional 1D	474	Convolutacional 1D	474
	<i>MaxPooling</i>	-	<i>BatchNormalization</i>	-
	Convolutacional 1D	224	Convolutacional 1D	224
	<i>MaxPooling</i>	-	<i>BatchNormalization</i>	-
	Convolutacional 1D	110	<i>MaxPooling</i>	-
	<i>MaxPooling</i>	-	Convolutacional 1D	110
	Convolutacional 1D	52	<i>BatchNormalization</i>	-
			Convolutacional 1D	52
			<i>BatchNormalization</i>	-
	<i>MaxPooling</i>	-	<i>MaxPooling</i>	-
	<i>Dropout</i> (0,5)	-	<i>Dropout</i> (0,5)	-
	Densa	30	Densa	30
	Densa	10	Densa	10
	Densa	4	Densa	4

Fonte: Autor (2022)

ANEXO D – DESEMPENHO DOS MODELOS CNN POR CLASSES DE ARRITMIA

Tabela 19 – Desempenho dos modelos DC-CNN e CNN *baseline* por classe de arritmia, incluindo valores de TP, TN, FP e FN

Modelo	Classe	SEN (%)	PRE (%)	F1 (%)	TP	TN	FP	FN
DC-CNN	N	99,09	86,94	92,62	17205	1066	157	2585
	S	0,75	19,05	1,44	12	19355	1595	51
	V	47,12	67,76	55,59	786	18978	882	374
	F	0	0	0	0	20637	376	0
CNN <i>baseline</i>	N	84,54	97,41	90,52	16913	855	449	2799
	S	1,05	32,69	2,03	17	19371	1593	35
	V	44,84	59,84	51,27	748	18846	920	502
	F	0	0	0	0	20638	376	2

Fonte: Autor (2022)

Tabela 20 – Desempenho do modelo DC-CNN com e sem *data augmentation* por classe de arritmia, incluindo valores de TP, TN, FP e FN

Modelo DC-CNN	Classe	SEN (%)	PRE (%)	F1 (%)	TP	TN	FP	FN
Sem <i>Data Augmentation</i>	N	99,09	86,94	92,62	17205	1066	157	2585
	S	0,75	19,05	1,44	12	19355	1595	51
	V	47,12	67,76	55,59	786	18978	882	374
	F	0	0	0	0	20637	376	0
SMOTE	N	94,95	86,77	90,68	16485	1141	2513	877
	S	4,66	26,88	7,94	75	19202	204	1535
	V	35,07	33,81	34,43	585	19286	1145	1083
	F	0	0	0	0	20631	9	376
ROS	N	94,96	87,78	91,23	16487	1359	2295	875
	S	2,55	18,64	4,48	41	19227	179	1569
	V	62,11	51,98	56,60	1036	18391	957	632
	F	0	0	0	0	20619	21	376

Fonte: Autor (2022)