

UNIVERSIDADE FEDERAL DO AMAZONAS  
FACULDADE DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

JOÃO ROBERTO GOMES SANTANA

Classificação de Arritmias Cardíacas em Sinais de ECG  
Utilizando Redes Neurais Profundas

MANAUS

2021

JOÃO ROBERTO GOMES SANTANA

Classificação de Arritmias Cardíacas em Sinais de ECG  
Utilizando Redes Neurais Profundas

Dissertação apresentada ao Curso de  
Mestrado em Engenharia Elétrica, área de  
Concentração Controle e Automação de  
Sistema do Programa de Pós-Graduação  
em Engenharia Elétrica da Universidade  
Federal do Amazonas.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

Co-Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Marly Guimarães Fernandes C o s t a

MANAUS  
2021

### Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S232c Santana, João Roberto Gomes  
Classificação de arritmias cardíacas em sinais de ECG utilizando  
redes neurais profundas / João Roberto Gomes Santana . 2021  
75 f.: il. color; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho  
Coorientadora: Marly Guimarães Fernandes Costa  
Dissertação (Mestrado em Engenharia Elétrica) - Universidade  
Federal do Amazonas.

1. ECG - Eletrocardiograma. 2. Amostragem. 3. Normalização. 4.  
CNN - Convolutional neural network. 5. LSTM - Long short term  
memory. I. Costa Filho, Cícero Ferreira Fernandes. II. Universidade  
Federal do Amazonas III. Título

JOÃO ROBERTO GOMES SANTANA

**CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS EM SINAIS DE  
ECG UTILIZANDO REDES NEURAI PROFUNDAS**

Dissertação apresentada ao Curso de  
Mestrado em Engenharia Elétrica, área de  
Concentração Controle e Automação de  
Sistema do Programa de Pós-Graduação  
em Engenharia Elétrica da Universidade  
Federal do Amazonas.

Aprovado em 10 de dezembro de 2021.

**BANCA EXAMINADORA**

Prof. Dr. Cícero Ferreira Fernandes Costa Filho	
Presidente	
Prof. Dr. Jozias Parente de Oliveira	
Membro Titular 1 - Externo	
Prof. Dr. Celso Barbosa Carvalho	
Membro Titular 2- Interno	
Prof. Dr. Carlos Augusto de Moraes Cruz	
Coordenador	

## AGRADECIMENTOS

Agradeço aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e Prof<sup>a</sup>. Dr<sup>a</sup>. Marly Guimarães Fernandes Costa, pela orientação dada, pelo aprendizado adquirido e, principalmente, pela confiança em mim depositada.

Aos meus pais, que sempre se dedicaram para que eu tivesse a melhor educação.

À Anna Luíza Silva Santana, minha filha, e Ivaneide Maria da Silva, minha esposa, por me apoiarem e insentivarem constantemente. Certamente, não conseguiria sem vocês.

Esta pesquisa, conforme previsto no Art. 48 do decreto nº 6.008/2006, foi financiada pela Samsung Eletrônica da Amazônia Ltda, nos termos da Lei Federal nº 8.387/1991, através de convênio nº 004, firmado com o Centro de P&D em Eletrônica e Tecnologia da Informação da Universidade Federal do Amazonas - CETELI / UFAM.

À Universidade Federal do Amazonas e, em especial, ao Centro de Tecnologia Eletrônica e da Informação – CETELI – pela concessão de toda infraestrutura para realização deste trabalho.

## RESUMO

As doenças cardiovasculares são a principal causa de morte em todo o mundo.

A detecção de doenças cardiovasculares em seus estágios iniciais pode reduzir efetivamente a taxa de mortalidade ao fornecer tratamento oportuno. Neste estudo, propomos uma nova metodologia para detecção de arritmias, utilizando Redes Neurais Convolucionais 2D. A principal característica da metodologia proposta é a utilização de imagens em nível de cinza de 15 x 15 pixels, contendo os valores de um batimento cardíaco do sinal de ECG. Este trabalho tem como objetivo detectar 17 arritmias. Para validar e testar a metodologia proposta, foi utilizado o banco de dados MIT-BIH, principal banco de dados de referência disponível na literatura. Quando comparada a outros resultados publicados anteriormente, a precisão obtida, 92,31%, está no estado da arte.

O trabalho apresentado fornece um método automático para detectar arritmias em sinais de ECG por meio de uma nova metodologia.

Palavras-chave: ECG, Amostragem, Normalização, CNN, LSTM.

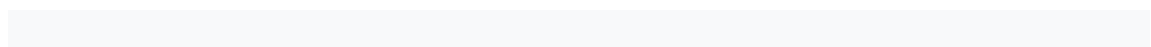
## ABSTRACT

Cardiovascular diseases are the number one cause of death worldwide.

Detecting cardiovascular diseases in its early stages could effectively reduce the mortality rate by providing timely treatment. In this study, we propose a new methodology to detect arrhythmias, using 2D Convolutional Neural Networks. The main characteristic of the proposed methodology is the use of 15 x15 pixels gray-level images, containing the values of a heartbeat of the ECG signal. This work aims to detect 17 arrhythmias. To validate and test the proposed methodology, MIT-BIH database, the main benchmark database available in literature, was used. When compared to other results previously published, the obtained precision, 92.31%, is in the state-of-the-art.

The presented work provides an automatic method to detect arrhythmias in ECG signals by a new methodology.

Keywords: ECG, Sampling, Normalization, CNN, LSTM.



## LISTA DAS FIGURAS

Figura 1– Anatomia do coração humano .....	27
Figura 2– Caminho percorrido pelo sangue dentro do coração .....	28
Figura 3– Ciclo elétrico do coração .....	29
Figura 4– Triângulo de Einthoven .....	30
Figura 5– Sinal registrado pelo eletrocardiograma numa derivação tipo II	30
Figura 6— O neurônio biológico .....	32
Figura 7– O modelo matemático do neurônio .....	33
Figura 8– Modelo de perceptron multicamadas .....	35
Figura 9– Matriz de imagem (RGB).....	38
Figura 10– Exemplo de Operação de Convolução .....	39
Figura 11– À esquerda, mostra-se o local de aplicação do filtro na imagem ou campo receptivo. À direita, é mostrada a matriz resultante da operação de convolução, onde os pixels da imagem filtrada coincidem com a posição dos pixels diferentes de zero na imagem original. ....	39
Figura 12– Visualização do Filtro na Imagem .....	40
Figura 13– Maxpooling .....	41
Figura 14–O resultado da aplicação do dropout na rede da esquerda .....	41
Figura 15– Modelo de redes recorrentes .....	42
Figura 16 – Redes recorrentes. Mostra-se o desmembramento ou “desenrolar” do laço .....	43
Figura 17– Esquema simplificado de uma LSTM.....	45
Figura 18 - Construção dos Registros e da Base de Dados .....	48
Figura 19- Etapas para realizar a construção das imagens .....	49
Figura 20- Sinal Original de ECG .....	49
Figura 21- Sinal de ECG Amostrado.....	50
Figura 22- Matriz dos Elementos Amostrados do Sinal de ECG .....	50
Figura 23- Imagem do Sinal de ECG Normalizado.....	51
Figura 24- Matriz do Sinal de ECG Normalizado .....	51
Figura 25- Imagem Gerada da Matriz Normalizada em Tons de Cinza.....	52
Figura 26– Arquitetura das Redes CNN-1D e CNN-2D .....	53
Figura 27– Arquitetura da Rede LSTM ou GRU .....	54
Figura 28-Desempenho da Arquitetura da Rede CNN.....	57



Figura 29- Matriz de Confusão mostrando os valores de precisão para as 17 arritmias, obtidos com a rede CNN-2D .....	58
Figura 30- Registro de um trecho do sinal de ECG do sinal NSR .....	66
Figura 31- Registro de um trecho do sinal de ECG da arritmia APB .....	66
Figura 32- Registro de um trecho do sinal de ECG da arritmia AFL.....	67
Figura 33- Registro de um trecho do sinal de ECG da arritmia AFIB .....	67
Figura 34- Registro de um trecho do sinal de ECG da arritmia SVTA.....	68
Figura 35- Registro de um trecho do sinal de ECG da arritmia WPW .....	68
Figura 36- Registro de um trecho do sinal de ECG da arritmia PVC .....	69
Figura 37- Registro de um trecho do sinal de ECG da arritmia Bigeminy .	69
Figura 38- Registro de um trecho do sinal de ECG da arritmia Trigeminy	70
Figura 39- Registro de um trecho do sinal de ECG da arritmia VT .....	70
Figura 40- Registro de um trecho do sinal de ECG da arritmia IVR .....	71
Figura 41- Registro de um trecho do sinal de ECG da arritmia VFL.....	71
Figura 42- Registro de um trecho do sinal de ECG da arritmia Fusion .....	72
Figura 43- Registro de um trecho do sinal de ECG da arritmia LBBBB ....	72
Figura 44- Registro de um trecho do sinal de ECG da arritmia RBBBB ....	73
Figura 45- Registro de um trecho do sinal de ECG da arritmia SDHB.....	73
Figura 46- Registro de um trecho do sinal de ECG da arritmia PR .....	74

## LISTA DE QUADRO

Quadro 1– Síntese da revisão bibliográfica .....	25
--------------------------------------------------	----

## LISTA DE TABELAS

Tabela 1- Resultado das 4 arquiteturas de redes profundas avaliadas: CNN-1D, CNN-2D, LSTM e GRU .....	58
Tabela 2- Comparação dos resultados obtidos com a rede CNN-2D com outros resultados publicados.....	59

## Sumário

1	INTRODUÇÃO.....	13
1.2	OBJETIVO GERAL.....	16
1.3	OBJETIVOS ESPECÍFICOS .....	16
1.4	ORGANIZAÇÃO DO TRABALHO .....	16
2	REVISÃO BIBLIOGRÁFICA .....	18
3	REFERENCIAL TEÓRICO .....	27
3.1	ARRITMIAS.....	31
3.2	REDES NEURAIS ARTIFICIAIS (RNA).....	32
3.2.1	FUNÇÕES DE ATIVAÇÕES .....	34
3.2.2	PERCEPTRON MULTICAMADAS ( <i>MULTILAYER PERCEPTRON</i> - MLP) .....	35
3.2.3	ALGORITMO <i>BACKPROPAGATION</i> .....	36
3.3	REDES NEURAIS CONVOLUCIONAIS (CNN) .....	37
3.3.1	CONVOLUÇÕES .....	38
3.3.2	CAMADA DE <i>DROPOUT</i> .....	41
3.4.1	REDE NEURAIS RECORRENTES (RNN).....	42
3.4.2	ALGORITMO <i>BACKPROPAGATION THROUGH TIME</i> (BPTT) .....	43
3.4.3	<i>LONG SHORT TERM MEMORY</i> (LSTM).....	45
4	MATERIAIS E MÉTODOS.....	47
4.1	MATERIAIS.....	47
4.2	MÉTODOS.....	52
4.2.1	ARQUITETURA DAS REDES BIDIMENSIONAIS .....	52
4.2.2	ARQUITETURA DAS REDES LSTM E GRU .....	54
4.2.3	MÉTRICAS DE AVALIAÇÃO .....	55
4.2.4	METODOLOGIA DE TREINAMENTO E VALIDAÇÃO .....	56
5	RESULTADOS .....	57
6	CONCLUSÕES .....	60
	REFERÊNCIAS .....	61
	APÊNDICE I.....	65
	APÊNDICE II.....	66
	APÊNDICE III .....	75

## 1 INTRODUÇÃO

De acordo com a Organização Mundial de Saúde (WHO, 2020), as doenças cardiovasculares (DCVs) são a causa número 1 de morte em todo o mundo, ceifando cerca de 17,9 milhões de vidas a cada ano. As DCVs são um grupo de doenças do coração e dos vasos sanguíneos e incluem doença coronariana, que, conforme (SWEIS e JIVAN, 2020), envolve o comprometimento do fluxo sanguíneo através das artérias coronarianas; bem como a doença cerebrovascular, que de acordo com (FRANÇA, 2018), corresponde a todo tipo de patologia que atinge os vasos do cérebro; além da doença cardíaca reumática, que segundo (THOMAS e KHETRAPAL), é a consequência de uma febre reumática aguda de longo prazo, que ocorre tipicamente na infância e o paciente apresenta uma inflamação nas válvulas cardíacas. Quatro em cada 5 mortes por DCV são devido a ataques cardíacos e derrames, e um terço dessas ocorrem prematuramente em pessoas com menos de 70 anos de idade.

Conforme o relatório de Estatística Cardiovascular – Brasil 2020 (2020), da Sociedade Brasileira de Cardiologia (SBC), as doenças cardiovasculares foram a principal causa de morte no Brasil no ano de 2017. Entre as DCVs, as doenças isquêmicas do coração (DICs) foram a causa número 1. Segundo (PAEZ, 2020), são doenças que ocorrem quando uma parte do coração não recebe sangue suficiente para realizar o bombeamento sanguíneo de forma adequada.

A detecção de uma DCV em seu estágio inicial poderia efetivamente reduzir a taxa de mortalidade ao fornecer um tratamento oportuno (MUSTAQUEEM et al., 2020).

A arritmia cardíaca, que se caracteriza pelo fato de os batimentos cardíacos desviarem-se do seu padrão normal. Um batimento cardíaco normal varia com a idade, tamanho do corpo, atividade e emoções (GUYTON).

Nos casos em que o batimento cardíaco parece muito rápido ou lento, a condição é conhecida como palpitação. Uma arritmia não significa necessariamente que o coração está batendo muito rápido ou lento, ela indica que o coração está seguindo um padrão de batimento irregular. As arritmias do coração podem ser detectadas através da avaliação de sua atividade elétrica, por meio do eletrocardiograma (ECG). O ECG é um sinal gerado por meio da colocação de eletrodos na superfície do corpo humano, permitindo que alterações causadas pela atividade elétrica do coração sejam detectadas (PAO et al., 2019). A utilização do ECG no diagnóstico de arritmias depende

geralmente de profissionais médicos especializados (TADESSE et al., 2019). Convencionalmente, os cardiologistas inspecionam visualmente as formas de onda de ECGs de 12 derivações em um formato de imagem digital. No entanto, é muito comum que os sinais de ECG com duração de muitas horas e até dias precisem ser analisados. Este é um procedimento muito demorado e desgastante, que limita significativamente a imparcialidade do diagnóstico.

Essa limitação pode ser eliminada com o uso de técnicas computacionais para detecção automática de arritmia e classificação de ECG.

Na literatura, identificamos dois grupos de trabalhos desenvolvidos para detecção automática de arritmias. O primeiro grupo compreende trabalhos mais antigos e que utilizaram ferramentas clássicas de aprendizado de máquina, enquanto o segundo grupo compreende trabalhos mais recentes e utilizam técnicas de aprendizado profundo, como redes convolucionais e redes recorrentes.

Dentre outros, no primeiro grupo de trabalhos identificamos a utilização das seguintes técnicas clássicas de aprendizado de máquina: Classificadores utilizando máquinas de vetores de suporte (PLAWIAK et al., 2017).

No segundo grupo de trabalhos identificamos duas tendências. A primeira engloba trabalhos que utilizam apenas redes neurais convolutivas, sendo as redes unidimensionais as mais frequentes (SWAPNA, SOMAN, VINAYAKUMAR, 2018).

Apenas poucos trabalhos utilizam redes convolucionais bidimensionais. Esses últimos utilizam a imagem do ECG ao invés do sinal unidimensional. No entanto, a abordagem utilizada nesses trabalhos é diferente daquela proposta neste trabalho.

Enquanto nos primeiros usa-se uma “foto” do ECG, neste trabalho propomos converter o sinal unidimensional em uma imagem. Uma segunda tendência observada consiste em trabalhos que utilizam redes híbridas, uma composição em série de redes convolucionais e redes recorrentes. As redes convolucionais são utilizadas para extração de características. As características extraídas, então, após uma redução de dimensionalidade, através de camadas de amostragem, alimentam a entrada de uma rede recorrente, responsável pela classificação das arritmias. As redes recorrentes são mais usadas para aplicações com sinais unidimensionais, como séries temporais, bem como para detecção de anomalias em sinais biológicos, como o ECG e eletroencefalograma (EEG).

Outra observação que fazemos é que a base de dados predominantemente utilizada para *benchmark* de algoritmos desenvolvidos para detecção de arritmias é a

base de dados do “Beth Israel Hospital”, situado em Boston. É mais conhecida como base de dados MIT-BIH, sendo a abreviatura MIT referente a Massachusetts Institute of Technology, pois foi desenvolvida em conjunto com esse importante instituto de pesquisa. Essa base, publicada no site da Physionet (Physionet, 2021) foi construída com exames de 47 pacientes. O sinal de ECG foi amostrado a uma taxa de 360 Hz. A base é constituída por 48 registros, com uma duração aproximada de 30 minutos de batimentos cardíacos cada um. Ao todo, os registros contêm dados de 17 tipos de arritmias diferentes.

A maioria dos trabalhos existentes na literatura que utilizam a base MIT-BIH destinam-se a classificação de 8 ou 5 classes. Apenas poucos trabalhos foram realizados para classificação das 17 classes de arritmias (YILDIRIM et al., 2018) e (PLAWIAK et al., 2017).

Do exposto anteriormente, identificamos que: 1) Ao se utilizar redes convolucionais bidimensionais, elas utilizam como entrada “fotos” do sinal de ECG, ao invés do sinal unidimensional do ECG. Mesmo assim, apenas poucos trabalhos utilizaram redes convolutivas bidimensionais; 2) Poucos destinaram-se a classificação das 17 classes da base de dados MIT-BIH.

Nesta dissertação, utilizou-se redes convolucionais 2D para a classificação de ECG, utilizando outro tipo de informação, que não uma “foto” do sinal de ECG, como tem sido feito até então. Converteu-se o sinal unidimensional em uma imagem.

Assim, o sinal unidimensional é convertido em uma imagem em nível de cinza que é utilizada na entrada da rede convolucional. No capítulo de metodologia, descreveremos em detalhes o processo de geração dessas imagens.

Uma abordagem semelhante já foi utilizada para análise de ruídos de motores (WEN et al., 2017). Esses autores realizaram a amostragem de sinais relacionada a falhas de motores.

Após um pré-processamento em que os dados foram normalizados, construíram imagens bidimensionais de 64 x 64 pixels, em nível de cinza. Os sinais utilizados no artigo foram extraídos da base de dados “bearing data case”, da Western Reserve University. Para o treinamento e teste foram utilizados 6400 e 1600 amostras, respectivamente. Os autores relatam uma acurácia na detecção de falhas de 99,74%.

Do ponto de vista do reconhecimento de padrões, a vantagem de trabalhar com um sinal 2D ao invés de um sinal 1D, é que novas relações de vizinhança (vizinho acima e abaixo) podem ser exploradas para extração de características através dos filtros

das redes convolucionais 2D, além das relações de vizinhança já existentes nos sinais unidimensionais (vizinho à esquerda e à direita) (WEN et al., 2017).

Comparou-se o desempenho de redes convolucionais bidimensionais, com a abordagem descrita anteriormente, com o desempenho de redes recorrentes, *Long Short Time Memory* (LSTM) ou *Gated Recurrent Unit* (GRU), na classificação de arritmias.

Por último, pretende-se utilizar as arquiteturas profundas para a classificação das 17 classes de arritmias presentes na base de dados MIT-BIH, um nicho pouco explorado na literatura.

## 1.2 OBJETIVO GERAL

Propor uma arquitetura de rede convolucional 2D, utilizando imagens do sinal de ECG convertidas para uma imagem 2D, para classificação das arritmias cardíacas.

## 1.3 OBJETIVOS ESPECÍFICOS

- a) Classificar as 17 arritmias presentes na base de dados MIT-BIH Arrhythmia e descritas no Referencial Teórico;
- b) Avaliar o desempenho de redes convolucionais 2D, na classificação de arritmias cardíacas, propondo a conversão do sinal de ECG 1D em uma imagem 2D em tons de cinza;
- c) Comparar o desempenho da rede convolucional 2D proposta com o desempenho das arquiteturas de redes recorrentes LSTM ou GRU; e
- d) Comparar o desempenho das arquiteturas propostas com o estado da arte.

## 1.4 ORGANIZAÇÃO DO TRABALHO

A organização do trabalho segue as divisões a seguir:

- a) Introdução;
- b) Revisão Bibliográfica;
- c) Referencial Teórico;
- d) Materiais e Métodos;
- e) Conclusões;
- f) Referências;
- g) Apêndice I;



- h) Apêndice II; e
- i) Apêndice III.

O capítulo 2 elenca trabalhos desenvolvidos recentemente na área de classificação de exames de ECG por meio de redes neurais convolucionais, bem como o uso de técnicas de transformações desses exames em imagens.

O capítulo 3 apresenta os fundamentos das redes neurais convolucionais; das Redes LSTM ou GRU e, de forma breve, a teoria sobre anatomia do coração e sobre arritmias. Nessa seção são mostradas as principais características do coração e das arritmias, bem como as arquiteturas das máquinas utilizadas na classificação das arritmias.

O capítulo 4 apresenta os materiais e os métodos utilizados. Descrevendo as características das bases de dados utilizadas, os parâmetros e os hiperparâmetros das redes neurais utilizadas, a metodologia de treinamento, validação e teste, e as métricas utilizadas para avaliação de desempenho.

O capítulo 5 avalia e compara o desempenho dos modelos neurais utilizados nesse trabalho, através das métricas apresentadas na seção anterior.

## 2 REVISÃO BIBLIOGRÁFICA

Nessa seção iremos abordar trabalhos publicados no período entre 2017 e 2021 sobre detecção de arritmias em sinais de ECG. Os trabalhos selecionados, em sua maioria, utilizam aprendizado profundo, baseados em redes convolucionais unidimensionais e bidimensionais, ou na associação de rede convolucionais e redes recorrentes. Os trabalhos selecionados realizam a classificação do sinal de ECG em 2 classes (normal ou anormal), 7 classes ou 8 classes (arritmias mais frequentemente discutidas na literatura) ou 17 classes. Essas últimas classes são constituídas por sinais normais, de marcapasso e de 15 disfunções cardíacas.

(SWAPNA, SOMAN e VINAYAKUMAR, 2018) utilizaram técnicas de aprendizado profundo, com o objetivo de diferenciar sinais com e sem arritmias. A base de dados utilizadas pelos autores foi a do MIT-BIH. A partir das 48 sequências de dados (101.dat a 124.dat e 200.dat a 234.dat, com algumas intermediárias ausentes), os autores extraíram manualmente 93521 batimentos cardíacos isolados, em janelas com extensão de 1s.

Desse total, 18482 (19,7%) batimentos correspondiam a arritmias. Os autores utilizaram tanto redes recorrentes (RNN, LSTM e GRU), quanto redes convolucionais.

As seguintes combinações de arquiteturas de redes foram utilizadas: CNN, RNN, LSTM, GRU, CNN-RNN, CNN-LSTM e CNN-GRU. A arquitetura da CNN foi uma rede 1D, com uma, duas ou três camadas escondidas e uma camada inteiramente conectada com duas saídas. Camadas escondidas com 32, 64 e 128 neurônios foram avaliadas, sendo o melhor desempenho obtido com 64 neurônios na camada escondida.

As redes RNN, GRU e LSTM também foram avaliadas com 1, 2 ou três camadas contendo 32, 64 e 128 blocos de memória. O melhor desempenho foi obtido com camadas contendo 64 blocos de memória. Para a conexão da CNN com as redes recorrentes é feita através de uma camada de *maxpooling*, colocada após a última camada convolucional.

Foram feitas simulações com taxas de aprendizado variando na faixa entre [0.01, 0.5]. Isoladamente, o melhor desempenho foi alcançado pela arquitetura CNN com três camadas, com uma acurácia de 0.781. Entre as redes recorrentes, a que apresentou melhor desempenho foi a rede GRU com 3 camadas, com uma acurácia de 0,776. Sendo que a de melhor desempenho verificado no trabalho foi com a combinação de uma rede CNN de 3 camadas e uma GRU, com acurácia de 0,837. Nas conclusões,

os autores ressaltam que, apesar dos excelentes resultados conseguidos pelas redes profundas, elas fornecem poucas explicações sobre seus mecanismos de decisão.

Em (PETMEZAS et al., 2020) os autores propõem um modelo de rede neural híbrida, CNN-LSTM para detecção de fibrilação atrial. Mais precisamente, na saída do classificador são identificados quatro tipos de sinais: fibrilação atrial, vibração atrial, ritmo do nodo atrioventricular e normal). Os sinais são obtidos a partir da base de dados MIT-BIH. Na etapa de pré-processamento os autores realizam um processo de filtragem com um filtro Butterworth passa alta de 7ª ordem, para remoção do ruído da linha de base, e com outro filtro Butterworth passa baixa de 6ª ordem, para remoção do ruído da rede de alimentação e também para os de alta frequência. Além disso, utilizam uma transformada discreta wavelet de 4 níveis para detecção do pico R do complexo QRS. Após a detecção do pico R, uma amostra do sinal é colhida de uma janela em torno dele, 250ms à esquerda e 500ms à direita.

A rede CNN utilizada é uma rede 1D. A extração de características na rede CNN é feita através de conjuntos de camadas CONV-BATCH-RELU. Após a extração de características, os autores aplicaram uma sequência de camadas *maxpooling* para redução de tamanho e alimentaram a rede LSTM formada por uma camada LSTM e uma Densa.

No treinamento, os autores utilizaram um lote de tamanho 128. No treinamento e teste foi empregada validação cruzada de 10 pastas. Os melhores resultados de classificação nas 4 classes anteriormente citadas foi uma sensibilidade de 97,87% e uma especificidade de 99,29%. Numa comparação com outros trabalhos previamente publicados na literatura, os autores concluíram que a arquitetura profunda proposta se encontra no estado da arte.

Em (YILDIRIM et al., 2018), os autores propõem a utilização de uma rede convolucional uni-dimensional para classificação das arritmias presentes na base de dados MIT-BIH. O número de classes de saída da rede CNN pode ser 13, 15 ou 17. Na composição das 17 classes temos o sinal normal, o sinal de marcapasso e os sinais de 15 arritmias. Na composição das 15 classes foram removidas as seguintes arritmias: fusão de batimento ventricular e normal, e taquicardia supraventricular. Na composição das 13 classes foram removidas as seguintes arritmias: taquicardia ventricular, fusão de batimento ventricular e normal, taquicardia supraventricular e contração ventricular prematura. Uma diferença fundamental desse trabalho em relação a outros publicados na literatura é que o mesmo não utiliza apenas informação de um único batimento do

coração, mas de 10 batimentos contíguos. O sinal de entrada da rede convolucional contém 3600 amostras (360 x 10). Segundo os autores, isso reduz a variabilidade do sinal. A arquitetura da rede utilizada tem 16 camadas. A seção de extração de características é constituída por blocos formados pelas seguintes camadas: CONV-BATCH-MAXPOOL.

Entre a seção de extração de características e a camada *softmax* de saída, temos uma camada densa com 512 neurônios.

Nos resultados, os autores apresentam tabelas de confusão, gráficos de acurácia, sensibilidade e especificidade para cada classe, e faz uma comparação com outros trabalhos previamente publicados na literatura com a mesma base. Os resultados de acurácia obtidos para 13, 15 e 17 classes foram, respectivamente, 95,2%, 92,51% e 91,33%. Numa pesquisa bibliográfica apresentada, embora apareçam resultados melhores, os autores argumentam que os resultados do método apresentado estão próximos dos melhores da literatura, com a vantagem de ser mais simples a implementação do mesmo, não exigindo a extração de características.

(ABDULLAH e AL-ANI, 2020) propuseram a utilização de um modelo híbrido CNN-LSTM para classificação de sinais de ECG em normal ou anormal (sem presença de arritmia) e também para detectar a presença de infarto do miocárdio. Para a primeira aplicação, os autores utilizaram a base de dados MIT-BIH, coletada no período de 1975-1979, de 47 pacientes. Nessa base, a amostragem foi feita a uma taxa de 360Hz. Em cada registro de 30s, temos 10800 amostras. Para a segunda aplicação foi utilizada a base de dados PTP Physionet, coletada na Universidade Benjamin Franklin, Berlim, a partir de 290 pacientes. A arquitetura da rede CNN utilizada é constituída de 18 camadas. Na fase de extração de características temos: dois conjuntos de camadas CONV-BATCH-RELU, um conjunto de camadas CONV-BATCH-RELU-DROPOUT e um outro CONV-BATCH-SOFTMAX-DROPOUT. Na fase de classificação temos duas camadas inteiramente conectadas e uma *softmax* com duas saídas. A arquitetura da LSTM é constituída por duas LSTM, uma de *dropout*, duas camadas inteiramente conectadas e uma *softmax* na saída. A conexão CNN-LSTM é feita a partir da saída da camada CONV4 da rede convolutiva. Os resultados obtidos para a primeira aplicação foram os seguintes: acurácia de 94,7 para a CNN e de 98,66 para CNN-LSTM. Para a segunda aplicação, foram: 97,51% para a CNN e 98% para CNN-LSTM.

Com base numa revisão de outros artigos previamente publicados na literatura, os autores concluíram que a arquitetura CNN-LSTM proposta encontra-se num estado bastante avançado.

(ZHENG et al., 2020) propuseram a utilização de um modelo híbrido 2D CNN-LSTM para classificação de oito classes do sinal de ECG: sinal normal e sete arritmias: bloqueio do ramo direito, bloqueio do ramo esquerdo, batimento atrial prematuro, contração ventricular prematura, batida ritmada, escape ventricular e onda de vibração ventricular. A base de dados de onde os sinais foram extraídos foi a base MIT-BIH. No total, utilizaram 107620 batimentos cardíacos para treinamento e teste do modelo neural profundo. Os dados de entrada do classificador de imagens em nível de cinza com dimensões de 192x128 pixels.

Para o balanceamento do conjunto de imagens foi utilizado o aumento de dados através de operações de *cropping* e *clipping* nas imagens originais. A arquitetura híbrida proposta é composta de 13 camadas. A extração de características na rede convolucional foi feita através de blocos formados pelas seguintes camadas: CONV2D-CONV2D-POOL.

A rede LSTM é constituída por uma camada LSTM, sendo duas inteiramente conectadas e uma camada de saída. A taxa de aprendizado utilizada no treinamento foi de 0.001. Observou-se um treinamento mais bem comportado ao introduzir-se uma camada de regularização de *dropout* na arquitetura da rede. Ao invés de utilizar a função de ativação RELU na camada convolutiva, utilizou-se a função ELU. Com a camada de regularização de *dropout*, os resultados alcançados foram os seguintes: acurácia de 99,01%, sensibilidade de 97,67% e especificidade de 99,57%.

(PLAWIAK et al., 2017) propuseram a utilização de montagem de classificadores para classificação de 17 classes no sinal de ECG: o sinal normal, o sinal de marcapasso e os sinais de 15 arritmias. Na etapa de pré-processamento os autores realizaram um reescalonamento do sinal para a faixa entre -1 e +1 e fizeram a extração do espectro de potência com os métodos de Welch e DCT. A partir desse espectro, foi utilizado algoritmo genético para a remoção de características redundantes. Os classificadores utilizados foram máquinas de vetores de suporte. O processo de montagem dos classificadores, por meio da utilização de algoritmo genético, deu-se em duas camadas. A primeira foi constituída por 17 classificadores, cada um especializado na detecção de uma classe distinta. O ajuste dos parâmetros de cada um deles é efetuado também utilizando algoritmo genético. Na segunda camada tínhamos apenas um

classificador que era alimentado pelos 17 classificadores da primeira camada, com saída binária (pertence ou não pertence a uma classe). A decisão do segundo classificador é efetuada com base na resposta dos clasificadores de entrada e nos valores de limiares de decisão. Para a etapa de treinamento e teste, os autores utilizaram validação cruzada com 10 pastas. O melhor resultado alcançado foi uma acurácia global de 91,40%.

(ROHMANTRI, et al., 2020) propuseram a utilização de redes neurais convolucionais bidimensionais para a detecção de 2 classes (normal e anormal), 7 e 8 do sinal de ECG. Nas oito classes, temos o sinal normal e as seguintes arritmias: Contração atrial prematura, contração ventricular prematura, batimento prematuro, fusão de batimento normal e ritmado, fusão de batimento ventricular e normal, bloqueio do ramo esquerdo e bloqueio do ramo direito. Os autores utilizaram a base de dados MIT-BIH. O sinal de entrada das redes neurais foram imagens com tamanho de 64x64 pixels, 32x32 pixels e 16x16 pixels. Os autores citam que na detecção de arritmias no ECG utilizando sinais 2D, normalmente são utilizadas imagens de 128x128 pixels. A ideia de utilizar imagens menores é tornar o processamento menos pesado. A arquitetura da rede convolucional proposta tinha 8 camadas, sendo 2 convolucionais, uma de *maxpooling*, 2 de *dropout*, 2 densas e uma de linearização. O treinamento e teste foram efetuados utilizando validação cruzada com 10 pastas. Nos resultados, os autores apresentam tabelas de confusão e gráficos de barra. Os melhores resultados obtidos foram com imagens de 64x64 pixels e foram os seguintes: acurácia de 98,91% para duas classes, 98,10% para 7 classes e 98,45% para a de 8 classes. A conclusão dos autores é que a acurácia é melhor com imagens de tamanhos maiores. Os resultados obtidos ficaram um pouco abaixo daqueles obtidos na literatura com imagens de 128x128 pixels.

(ULLAH et al., 2020) propuseram a utilização de uma rede convolucional bidimensional para a detecção de 8 classes no sinal de ECG: sinal normal e 7 arritmias mais frequentes. A base de dados utilizada foi a MIT-BIH. O sinal de entrada das redes convolucionais forma imagens de 256 x 256 pixels, obtidas a partir de espectrogramas extraídos do sinal unidimensional através da transformada de tempo curto de Fourier. A arquitetura da rede convolucional é formada por 10 camadas, sendo 4 conjuntos de camadas CNN-MAXPOOLING, seguidas por uma camada inteiramente conectada e uma camada de saída. Para a melhoria do desempenho da rede convolucional, os autores utilizaram aumento de dados. Os melhores resultados obtidos pelos autores foi com o aumento de dados: acurácia de 99,11%, sensibilidade de 97,91% e especificidade de 99,61%. Através de uma comparação com outros resultados apresentados na literatura,

os autores concluíram que o método proposto encontra-se em um estado muito avançado.

Em (GAO et al., 2019), os autores propõem um método para realizar a classificação de 8 tipos de anomalias do sinal de ECG por meio de aprendizagem profunda. Dessas anomalias, 7 são arritmias amplamente difundidas na literatura, enquanto um tipo que corresponde ao sinal normal. Foi utilizada a base de dados do MIT - BIH. Todos os sinais de ECG foram segmentados em sequências de 250 amostras e centralizadas nos picos R. Na fase de pré-processamento, a transformada discreta wavelet de Daubechies 6 (db6) foi usada para remover ruídos dos sinais de ECG. A arquitetura da rede proposta é composta por uma camada de entrada, uma camada LSTM e duas camadas inteiramente conectadas. Nessa LSTM, foi utilizada 4 camadas de rede. No treinamento utilizou-se um tamanho de lote de 128, otimizador Adam e como função de custo, a *focal loss* com  $\gamma = 2$ . Com esses parâmetros, os melhores resultados obtidos foram: acurácia de 99,26%, revocação de 96,26%, especificidade de 99,14%, precisão de 99,30% e F1-Score de 99,27%. Para generalizar os resultados, os autores propuseram acrescentar mais sinais de ECG em futuras pesquisas, através da ampliação dos dados utilizados nos conjuntos de treinamento e de teste. Além disso, citaram a possibilidade de incorporar diferentes níveis de ruído aos sinais de ECG, com o intuito de avaliar o desempenho da rede LSTM na presença de ruídos.

Em (WU et al., 2021), os autores propuseram uma forma de realizar a classificação de 5 tipos de sinais de ECG utilizando redes neurais convolucionais. Desses sinais, 4 deles são arritmias amplamente discutidas e estudadas na literatura e fazem parte da base do MIT-BIH. O pré-processamento foi feito usando a transformada wavelet. Essa etapa de pré-processamento foi necessária para retirar os ruídos dos sinais de ECG e para tornar a classificação mais precisa.

Após essa fase, todas as amostras utilizadas foram normalizadas. Esse trabalho utilizou uma arquitetura de rede neural convolutiva, composta pelas seguintes camadas: CONV – POOLING - CONV - POOLING - CONV - POOLING - CONV – POOLING - DROPOUT - FULLY CONNECTED - FULLY CONNECTED - SOFTMAX.

Nesse trabalho, foi utilizada uma camada de *average pooling* ao invés da camada de *max pooling*, que é usado na maioria dos trabalhos envolvendo esse tipo de rede. Os autores optaram pela *average pooling* por ela preservar a forma geral dos dados de entrada, sendo assim benéfica à classificação dos sinais. Os melhores

resultados obtidos na classificação das cinco classes foram os seguintes: acurácia de 97,41%, sensibilidade de 97,05%, especificidade de 99,35% e precisão de 97,22%.



Quadro 1– Síntese da revisão bibliográfica

Artigo	Base de Dados/Amostra	Pré-processamento	Classificadores	Tarefa	Resultados
SWAPNA, SOMAN e VINAYAKMUR, 2018	Base de dados MIT-BIH de arritmia: batimentos cardíacos com 361 amostras em torno do pico R.	Extração dos batimentos cardíacos manualmente	1D CNN, CNN-RNN, CNN-LSTM e CNN-GRU	Detectar sinais normais e anormais (com arritmia)	Acurácias: LSTM: 0,774, GRU: 0,776, CNN:0,812, CNN-LSTM: 0,834, CNN-GRU: 0,837
ABDULLAH e AL-ANI, 2020	Base de dados MIT-BIH de arritmia, Base de dados PTB Physionet/ 130 amostras em torno do pico R.	Extração dos batimentos cardíacos manualmente	1D CNN-LSTM	Detectar sinais normais e anormais (com arritmia) no ECG, utilizando a base MIT-BIH . Detectar infarto do miocárdio utilizando a base PTB.	Acurácia: 98,66% na base MIT-BIH e 98,13% na base PTB.
PETMEZAS et al., 2021	Sinais de fibrilação atrial da base MIT-BIH/187 amostras em torno do pico R.	Aplicações de filtros Butterworth para remoção de ruídos de baixa e alta frequência. Detecção do pico R com DWT.	1D CNN-LSTM	Detectar 4 classes no ECG relacionadas a fibrilação atrial: fibrilação atrial, <i>flutter</i> atrial, ritmo do nodo AV e ritmo normal.	Sensibilidade: 97,63% Especificidade: 99,21%
YILDIRIM at al., 2018	Base de dados MIT-BIH de arritmia/3600 amostras (10 batimentos cardíacos)	Extração de 10 batimentos contíguos para cada amostra	1D CNN	Detectar 17 classes no ECG: ritmo normal, marcapasso e 15 arritmias.	Acurácia: 17 classes: 91,33%; 15 classes: 92,51% e 13 classes: 95,2%.
ZHENG et al., 2020	Base de dados MIT-BIH de arritmia/imagens com dimensões de 192x128 pixels.	Conversão de 360 amostras do sinal de ECG em uma imagem bidimensional em nível de cinza de tamanho 192x128	2D CNN – LSTM	Detectar 8 classes no ECG: sinal normal e 7 tipos de arritmias	Acurácia de 99,01%, sensibilidade de 97,67% e especificidade de 99,57%
PLAWIAK, 2017	Base de dados MIT_BIH de arritmia/3600 amostras (10 batimentos cardíacos).	Extração de 10 batimentos contíguos para cada amostra. Reescalonamento do sinal. Extração do espectro de potência através dos métodos de Welch e DCT. Eleição de características com algoritmo genético.	Montagem genética de 17 classificadores SVM utilizando algoritmo genético.	Detectar 17 classes no ECG: ritmo normal, marcapasso e 15 arritmias.	Acurácia de 91,40% para 17 classes
ROHMANTRI e SURANTHA, 2020	Base de dados MIT_BIH de arritmia/ imagens de 64x64, 32x32 e 16x16 pixels.	Obtenção de uma imagem bidimensional em nível de cinza de tamanho 64x64, 32x32 e 16x16, a partir de 360 amostras do sinal de ECG.	2D CNN	Detectar 8 classes no ECG: sinal normal e 7 tipos de arritmias. Detectar 2 classes: normal e anormal. Detectar 7 classes.	Acurácia de 98,91% para duas classes; 98,10% para 7 classes e 98,45% para a de 8 classes

ULLAH et al., 2020	Base de dados MIT-BIH de arritmia/imagens de 256 x 256 pixels.	Obtenção de espectrogramas de sinais, utilizando a transformada de tempo curto de Fourier e a sua conversão numa imagem de 256x256 pixels.	2D CNN	Detectar 8 classes no ECG: sinal normal e 7 arritmias.	Acurácia: 98,59%; sensibilidade: 99,61% e especificidade: 99,11%
GAO et al., 2019	Base de dados MIT-BIH de arritmia: batimentos cardíacos com 250 amostras em torno do pico R.	Foi usada a transformada discreta wavelet de Daubechies 6 (db6) no intuito de remover ruídos dos sinais de ECG	LSTM	Detectar 8 classes no ECG: sinal normal e 7 arritmias	Acurácia de 99,26%, revocação de 96,26%, especificidade de 99,14%, precisão de 99,30% e F1-Score de 99,27%
WU et al., 2021	Base de dados MIT-BIH de arritmia: batimentos cardíacos com 360 amostras em torno do pico R.	Sendo utilizada a transformada de wavelet para retirar os ruídos dos sinais de ECG	1D CNN	Detectar 5 classes no ECG: sinal normal e 4 arritmias	Acurácia de 97,41%, sensibilidade de 97,05%, especificidade de 99,35% e taxa de precisão de 97,22%

### 3 REFERENCIAL TEÓRICO

O principal objetivo desse trabalho é utilizar redes profundas para a classificação de arritmias cardíacas. Para facilitar o entendimento do que foi realizado, apresentaremos a seguir, com brevidade, conceitos fundamentais a respeito da anatomia e fluxo sanguíneo no coração e dos sinais elétricos responsáveis pelo batimento cardíaco.

Segundo (MAGALHÃES, 2020), esse importante músculo está localizado na caixa torácica, entre os pulmões. Ele funciona semelhantemente a uma bomba hidráulica dupla, de tal forma que, enquanto o lado esquerdo bombeia o sangue arterial para diversas partes do corpo, o venoso é bombeado pelo lado direito para os pulmões. O coração impulsiona o sangue por meio de dois movimentos: contração ou sístole e relaxamento ou diástole. Conforme ilustrado na Figura 1, o coração possui duas cavidades localizadas na parte superior, os átrios, e duas cavidades localizadas na parte inferior, os ventrículos.

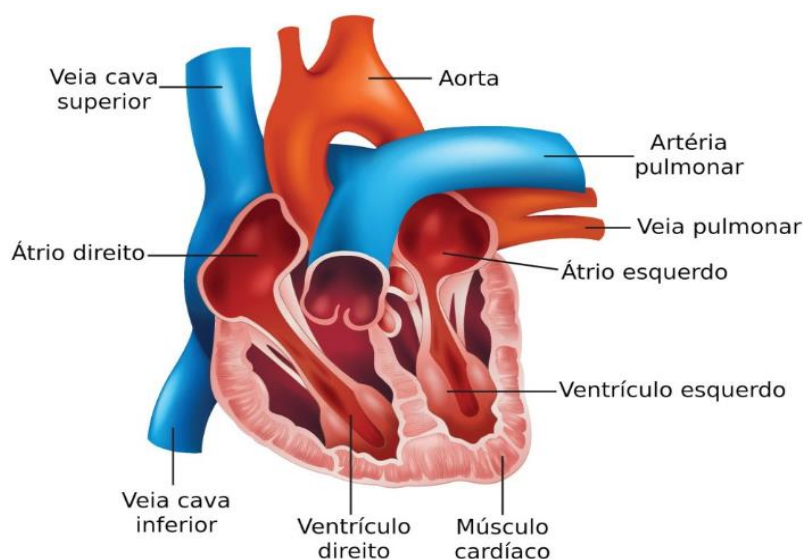


Figura 1– Anatomia do coração humano

Fonte: Dextro

(DOS SANTOS) descreve que o sangue proveniente das diversas partes do corpo chega no átrio direito. A partir dele, flui para o ventrículo direito e é conduzido para os pulmões, através da artéria pulmonar onde ocorre o processo de hematose, ou seja, onde o sangue rico em gás carbônico é oxigenado. Quatro veias pulmonares levam o sangue do pulmão de volta ao coração, ingressando no átrio esquerdo através da

válvula mitral. A partir do ventrículo esquerdo, através da aorta, para o corpo. Na Figura 2 ilustramos o fluxo do sangue pelas cavidades cardíacas.

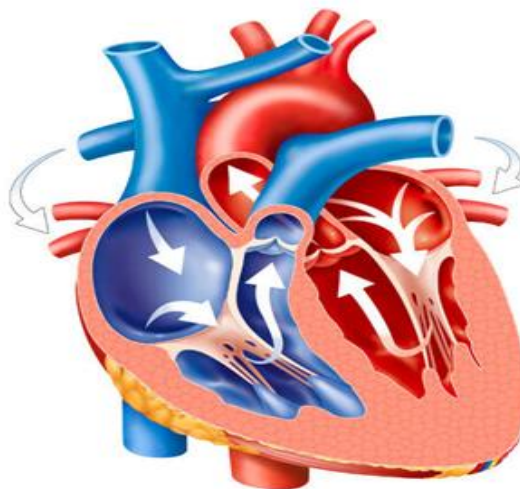


Figura 2– Caminho percorrido pelo sangue dentro do coração  
Fonte: Dos Santos

O bombeamento do sangue no coração é coordenado por uma complexa malha elétrica, que se assemelha a um circuito pulsado. O sinal responsável pela atividade cardíaca é gerado no Nodo Sinoatrial (nodo SA). As células do nodo sinoatrial têm características diferenciadas das outras células cardíacas, sendo eletricamente autoexcitáveis, gerando um potencial de ação característico (WEBSTER, 2009). O nodo sinoatrial localiza-se na junção da veia cava superior com o átrio direito. Através de uma rede de tecidos especializados, o sinal gerado no nodo SA, propaga-se pelos átrios e atinge o nodo atrioventricular (nodo AV), em que o sinal elétrico sofre um retardo antes de passar para os ventrículos. A partir do nodo AV, o sinal cruza a barreira isolante que existe entre os átrios e os ventrículos através do feixe-átrio-ventricular (feixe AV), de onde segue, através do feixe de HIS e das fibras de Purkinje, para alimentar o endocárdio e o epicárdio, dos ventrículos esquerdo e direito. A Figura 3 ilustra os diversos componentes do circuito elétrico do coração e a sequência temporal de propagação do impulso elétrico.

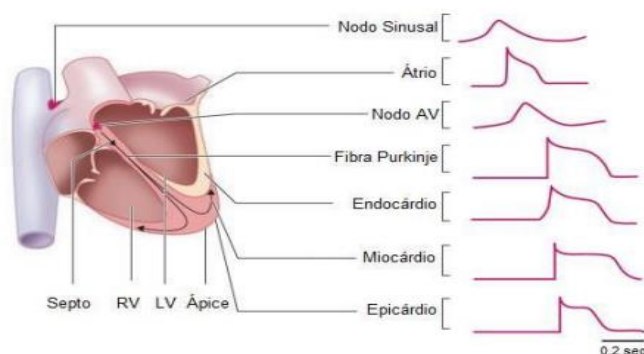


Figura 3– Ciclo elétrico do coração

Fonte: Pecharki, Stempi e Witt (2013)

Os atrasos gerados pela propagação do sinal elétrico, associados aos processos de polarização e repolarização do músculo cardíaco, irá ocasionar o surgimento de um dipolo elétrico. Esse é representado por um vetor que parte da base (negativo) para o ápice do coração (positivo). A direção e a amplitude desse dipolo varia com o tempo.

Em virtude da dependência espacial do campo elétrico do coração, posições distintas dos eletrodos colocados na pele, resultarão em diferentes sinais a serem medidos (WEBSTER, 2009). Na Figura 4, ilustramos o triângulo de Einthoven, que é um triângulo envolvendo o músculo cardíaco. Ao percorrê-lo, devido a Lei das tensões de Kirchhoff, a soma dos potenciais é zero. Nessa Figura, são mostradas três localizações principais para efetuar medição dos potenciais cardíacos, denominadas de derivações I, II e III. Na derivação I, o eletrodo positivo é localizado no ombro esquerdo. No direito, temos a derivação II e na perna esquerda, a III. Eventualmente, um outro eletrodo poderá ser colocado na perna direita e será adequadamente aterrado.

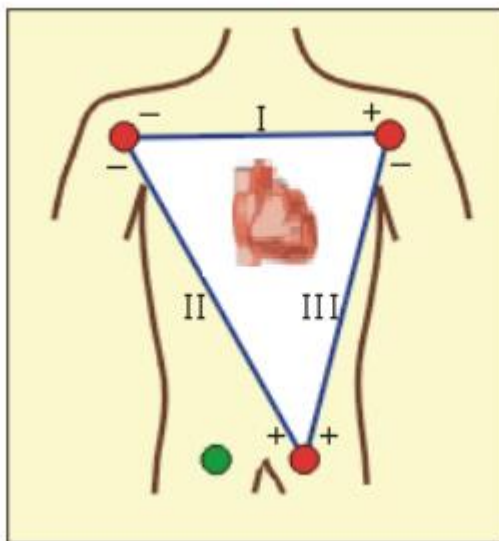


Figura 4– Triângulo de Einthoven  
 Fonte: Pecharki, Stempi e Witt (2013)

Na Figura 5 ilustramos o sinal registrado por um eletrocardiograma na derivação tipo II. Observa-se a presença de três ondas principais. A onda P, que está relacionada com a despolarização atrial, o complexo QRS, que está associado com a despolarização ventricular e a onda T, que está associada com a repolarização dos ventrículos. Já a repolarização dos átrios é normalmente ofuscada pelo complexo QRS.



Figura 5– Sinal registrado pelo eletrocardiograma numa derivação tipo II  
 Fonte: Pecharki, Stempi e Witt (2013)

### 3.1 ARRITMIAS

De acordo com a Sociedade Brasileira de Arritmias Cardíacas (SOBRAC), as arritmias ou disritmias cardíacas são alterações que ocorrem possivelmente na geração ou na condução do estímulo elétrico do coração e provocam modificações no ritmo dos batimentos cardíacos. Ainda de acordo com essa instituição, elas são responsáveis em média por cerca de trezentas mil mortes súbitas por ano no Brasil.

As arritmias possuem e podem acometer crianças, jovens, adultos e idosos e são resultantes de anomalias no ritmo cardíaco, podendo ser resultante da desaceleração do ritmo cardíaco, da aceleração do ritmo cardíaco, do bloqueio da passagem do sinal pelo nodo átrio-ventricular ou pelo feixe átrio-ventricular, por sinais reentrantes, que ocorrem depois da passagem do sinal por áreas isquêmicas, por efeitos de drogas como cafeína e nicotina, por calcificações no tecido cardíaco etc.

Existem diversos tipos de arritmias. Podemos citar: 1) Fibrilação atrial (*Atrial Fibrillation* - AFIB), segundo (MITCHELL, 2019) essa arritmia ocorre quando o ritmo atrial é rápido e irregular; 2) Pré-excitação (*Pre-excitation* - WPW), conforme (RODRIGUES, 2018) é uma arritmia em que o estímulo elétrico atinge o ventrículo pela via acessória antes do nódulo AV, gerando uma pré-excitação ventricular (Onda Delta) e encurtando o intervalo PR (intervalo entre a onda P e o pico do complexo QRS); e 3) Batimento atrial prematuro (*Atrial Premature Beat* - APB), segundo (MITCHELL, 2021) essa arritmia é um batimento atrial prematuro, ou seja, é um batimento cardíaco causado pela ativação elétrica dos átrios a partir de uma zona anormal, antes de ocorrer o batimento cardíaco normal.

As 17 arritmias, cujos sinais encontram-se disponíveis na base de dados do MIT-BIH utilizadas neste trabalho, são: 1) Ritmo sinusal normal (*Normal Sinus Rhythm* – NSR); 2) Batimento atrial prematuro (*Atrial Premature Beat* - APB); 3) Vibração atrial (*Atrial Flutter* – AFL); 4) Fibrilação atrial (*Atrial Fibrillation* - AFIB); 5) Taquiarritmias supraventricular (*Supraventricular Tachyarrhythmia* - SVTA); 6) Pré-excitação (*Pre-excitation* - WPW); 7) Contração ventricular prematura (*Premature Ventricular Contraction* - PVC); 8) Bigeminia ventricular (*Ventricular Bigeminy* - Bigeminy); 9) Trigeminismo ventricular (*Ventricular Trigeminy* - Trigeminy); 10) Taquicardia ventricular (*Ventricular Tachycardia* - VT); 11) Ritmo idioventricular (*Idioventricular Rhythm* - IVR); 12) Vibração

ventricular (*Ventricular Flutter* - VFL); 13) Fusão de batimento ventricular e normal (*Fusion of Ventricular and Normal Beat* - Fusion); 14) Batida do ramo esquerdo (*Left Bundle Branch Block Beat* - LBBBB); 15) Batida do ramo direito (*Right Bundle Branch Block Beat* - RBBBB); 16) Bloqueio cardíaco de segundo grau (*Second Degree Heart Block* - SDHB); e 17) Ritmo de marcapasso (*Pacemaker Rhythm* - PR).

### 3.2 REDES NEURAIS ARTIFICIAIS (RNA)

Para compreender a lógica de funcionamento das redes neurais, faz-se necessário compreender a dinâmica de funcionamento do cérebro humano. De acordo com (COSTA e LUDWIG JÚNIOR, 2007), o neurônio é a unidade básica do cérebro humano. Nele, ocorre a transmissão de informações e a condução de mensagens nervosas. O neurônio é constituído por 3 partes principais: o soma ou corpo celular, dendritos e axônio. Nas extremidades dos axônios se localizam os terminais, onde são realizadas as transmissões das informações para outros neurônios. Essas são conhecidas como sinapses. A Figura 6 mostra a estrutura da célula neuronal.

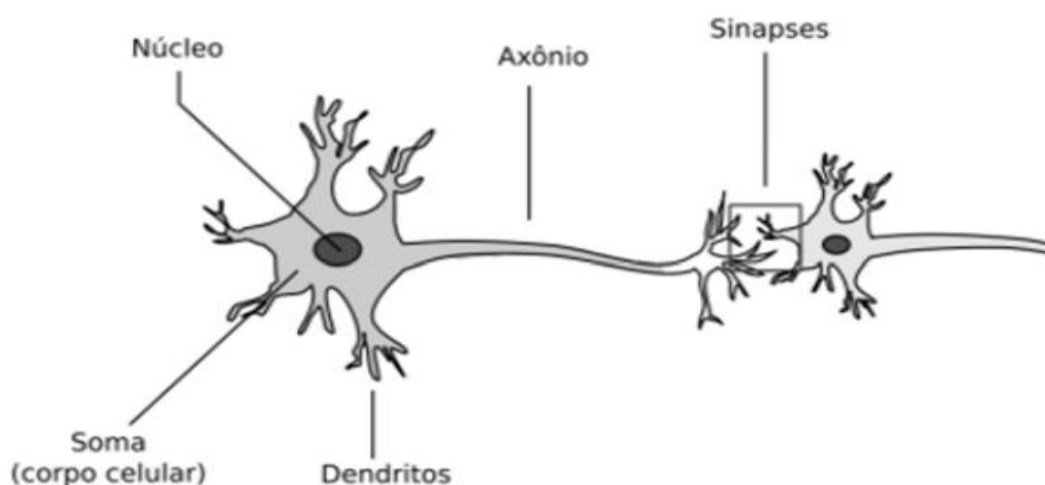


Figura 6— O neurônio biológico  
Fonte: Chagas (2019)

O corpo da célula processa a informação proveniente de outras células nervosas e a envia para outros neurônios. Os dendritos têm como função receber os impulsos nervosos provenientes de outros neurônios e conduzi-los até o corpo celular. Assim, a informação coletada pelos dendritos é enviada e processada no corpo, que gera pulsos em função do valor do potencial elétrico resultado do processo de soma. O pulso



só é gerado, se esse potencial é superior a um limiar de -50mV (COSTA e LUDWIG JÚNIOR, 2007).

Esses pulsos gerados pelo soma são conhecidos pelo nome de “potências de ação”. Através do axônio, o mesmo é transmitido para os dendritos dos neurônios seguintes.

O modelo de um neurônio artificial, inspirado no neurônio biológico, tornou-se a base da Inteligência Artificial.

(COSTA e LUDWIG JÚNIOR, 2007) informam que o modelo mais bem aceito foi o proposto por Warren McCulloch e Walter Pitts em 1943, o qual implementava de maneira simplificada os componentes e o funcionamento de um neurônio biológico.

Nesse modelo, o neurônio calcula a soma ponderada de diversas entradas, e o resultado é submetido à entrada de uma função, denominada de função de ativação do neurônio. A saída dessa função é a saída do neurônio. Na Figura 7, mostramos a arquitetura do neurônio artificial.

No neurônio biológico, os estímulos na entrada podem ser tanto excitatórios quanto inibitórios. No neurônio de Warren McCulloch e Walter Pitts a excitação ou inibição é representada pelo valor algébrico de pesos sinápticos na entrada. Valores positivos do peso representam estímulos excitatórios, enquanto que pesos negativos representam estímulos inibitórios. Quanto maior for o valor de um determinado peso, mais importante e relevante ele será.

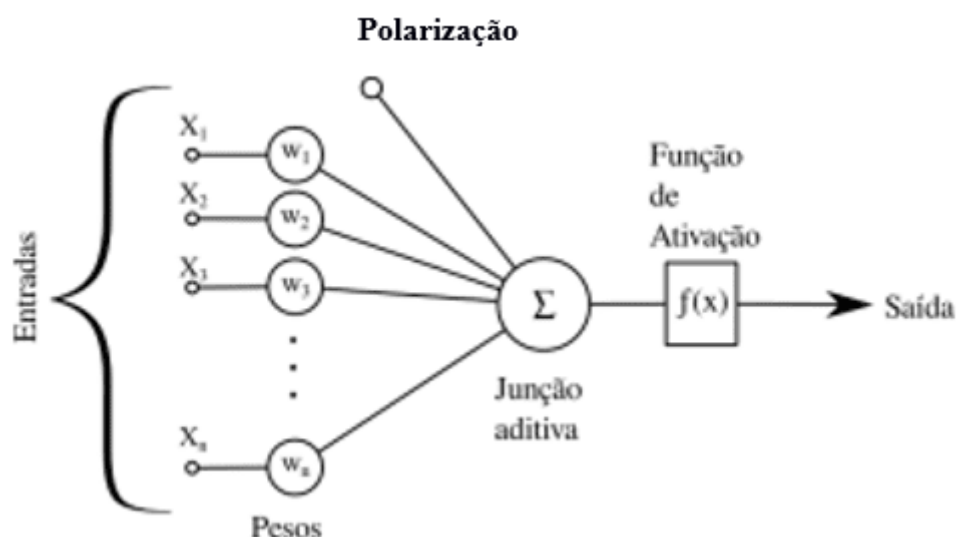


Figura 7– O modelo matemático do neurônio

Fonte: Adaptado de Azevedo (2016)

No neurônio artificial a relação entre a entrada e a saída é mostrada através da equação (1) e no formato matricial, na (2):

$$X w = y \quad (1)$$

Em que:

$X$  = Vetor das entradas;

$w$  = Matriz dos pesos e polarizações; e

$y$  = Vetor das saídas.

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} \end{bmatrix} \times \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_d \end{bmatrix} \quad (2)$$

O uso de funções de ativação não-lineares torna a rede neural uma ferramenta muito poderosa.

O modelo inclui também uma polarização ou *bias*. Esta nova variável é incluída ao somatório da função de ativação, no intuito de aumentar o grau de liberdade dessa função, bem como elevar a capacidade de aproximação da rede. O seu valor é ajustado da mesma forma que os pesos sinápticos. Ele possibilita que um neurônio apresente uma saída não nula, ainda que todas as suas entradas sejam nulas. Assim, caso ele não fosse usado e todas as entradas de um neurônio fossem iguais a zero, o valor da função de ativação também iria se anular.

Desta forma; não conseguiríamos implementar, por exemplo, a operação lógica do ou exclusivo sem utilizar o recurso do *bias* (COSTA e LUDWIG JÚNIOR, 2007).

Um dos modelos de um neurônio artificial de uma única camada mais bem sucedido foi o Perceptron, proposto por Roseblatt em meados de 1955 (BOSCARIOLI, DA SILVA e PERES, 2016). O perceptron não é nada mais que vários neurônios semelhantes a esse mostrado na Figura 7 em uma única camada, conectados às mesmas entradas.

### 3.2.1 FUNÇÕES DE ATIVAÇÕES

A função de ativação  $f(\cdot)$  precisa ser não linear. Assim sendo, a rede irá adquirir uma capacidade de representação bem mais generalizada. Para que o treinamento seja possível, ela deverá ser diferenciável. Por isso; geralmente, utiliza-se a tangente hiperbólica ou a sigmóide, bem como a relu (*rectified linear unit*) ou

até mesmo a *softmax*, entre outras. Definidas respectivamente na equação (3), (4), (5) e (6):

$$f(x) = \frac{e^{+x} - e^{-x}}{e^{+x} + e^{-x}} \quad (3)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

$$f(x) = \max(0, x) \quad (5)$$

$$S_j = \frac{z_j}{\sum_{k=1}^K e^{z_k}} \quad (6)$$

Em que:

k= Índice dos neurônios do mesmo nível;

z= Vetor dos neurônios de saída; e

j= Índice do neurônio de saída.

### 3.2.2 PERCEPTRON MULTICAMADAS (*MULTILAYER PERCEPTRON* - MLP)

Em virtude da simplicidade do modelo do perceptron, ele apresentou algumas limitações, sendo a maior delas o fato de não classificar classes que não sejam linearmente separáveis (MINSKEY e PAPERT, 1969). Uma maneira de superar essa limitação é por meio da MLP. De forma resumida, o MLP é uma rede neural composta por uma ou mais camadas ocultas ou escondidas, as quais são constituídas por um determinado número de neurônios. Com intuito de facilitar o entendimento desse tipo de rede e melhorar a sua visualização, a Figura 8 ilustra uma MLP com apenas uma camada oculta.

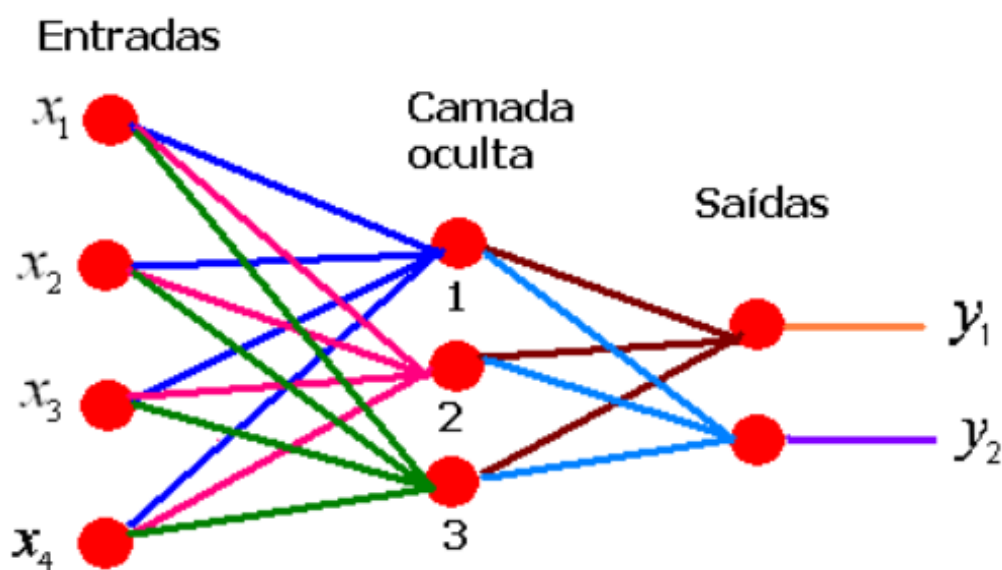


Figura 8– Modelo de perceptron multicamadas

Fonte: Lanhellas (2013)

### 3.2.3 ALGORITMO BACKPROPAGATION

O *backpropagation* é um dos algoritmos mais utilizados no treinamento de redes MLP. Ele é constituído de duas etapas, propagação direta e reversa (HAYKIN, 2008) e utiliza o algoritmo do gradiente descendente para a atualização dos pesos e polarizações, objetivando a minimização de uma função de custo da rede, calculada em função da saída calculada e da desejada da rede.

A propagação direta é usada para calcular a saída de todas as camadas da rede, inclusive da camada de saída, em função de uma entrada apresentada à rede. A propagação reversa, por sua vez, é utilizada para o cálculo da derivada da função de custo da rede em relação aos pesos e polarizações. Com o valor dessas derivadas, o algoritmo do gradiente descendente atualiza os pesos e polarizações da rede (COSTA e LUDWIG JÚNIOR, 2007). A seguir descrevemos essas etapas:

- 1) A partir do vetor de entrada aplicado à rede, calcula-se o valor de saída de todas as camadas, inclusive da camada de saída, através da propagação direta;
- 2) Calcula-se os gradientes da função de custo em relação a todos os parâmetros dos neurônios da camada de saída:

$$\delta_j(n) = -e_j(n) \cdot \varphi'_j(v_j(n)) \quad (7)$$

- 3) Atualiza-se os pesos e as polarizações dos parâmetros da última camada:

$$\Delta W_{ij} = -\eta \cdot \delta_j(n) \cdot y_i(n) \quad (8)$$

$$\Delta b_j = -\eta \cdot \delta_j(n) \quad (9)$$

- 4) Encontra-se o gradiente da função de custo para todos os neurônios da camada oculta:

$$\delta_i(n) = \varphi'_i(v_i(n)) \cdot \sum_{j=1}^J (\delta_j(n) \cdot \Delta W_{ij}(n)) \quad (10)$$

- 5) Atualiza-se os pesos e as polarizações da camada oculta:

$$\Delta W_{kj}(n) = -\eta \cdot \delta_j(n) \cdot y_k(n) \quad (11)$$

$$\Delta b_j = -\eta \cdot \delta_j(n) \quad (12)$$

O processo prosseguirá de forma semelhante para as demais camadas ocultas até chegarmos à camada de entrada da rede. Onde, o valor de  $y_k(n)$  será substituído por  $x_k(n)$ .

Em que:

$\delta_j(n)$  = Derivada da função de custo local;

$\eta$  = Taxa de aprendizagem;

$e_j(n)$  = Erro cometido na saída  $n$  da rede;

$W$  = Peso;

$b$  = Polarização;

$y_k$  = Saída da camada  $k$ ;

$\varphi_i$  = Função de transferência;

$\varphi'_j$  = **Derivada da** Função de transferência; e

$v_j(n)$  = Saída da camada anterior.

$$v_j = \sum_{i=1}^J (W_{ij} \cdot x_j) \quad (13)$$

### 3.3 REDES NEURAIIS CONVOLUCIONAIS (CNN)

As redes neurais convolucionais são geralmente utilizadas para classificação de imagens apresentadas à sua entrada. De maneira semelhante às RNA que foram tratadas anteriormente neste texto, são constituídas também por neurônios com ligações compostas por pesos e polarizações, que vão sendo ajustados em cada interação.

Diferentemente das RNA, no entanto, os neurônios das redes convolucionais só se conectam a poucos neurônios da camada anterior. Assim os neurônios das redes convolucionais se especializam em apenas uma pequena região da imagem. Isso permite que elas lidem com um número muito elevado de entradas, devido a um número reduzido de conexões (LECUN; KAVUKCUOGLU e FARABET, 2010).

(HAYKIN, 2008) informa que foi realizado em 1959 por Hubel e Wiesel experimentos demonstrando que neurônios específicos do cérebro de um gato foram ativados ao serem expostos a determinados estímulos. Esses pesquisadores conseguiram efetuar o reconhecimento visual de uma imagem por meio dessa técnica.

(ALVES, 2018) afirma que essa ideia é o cerne de uma Rede Convolucional:

Aplicando-se filtros a pequenas regiões da imagem, consegue-se extrair características da imagem, como linhas e curvas. As camadas das redes convolucionais são constituídas por matrizes tridimensionais de neurônios, caracterizadas pelos parâmetros altura, largura e profundidade. Nos dados de entrada, esse último parâmetro corresponde a quantidade de canais de cores da imagem.

Assim sendo, as imagens coloridas possuem três canais (RGB), conforme mostrado na Figura 9. No caso de tons de cinza, teríamos apenas um canal.

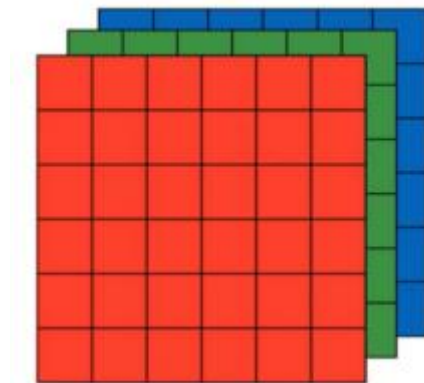


Figura 9– Matriz de imagem (RGB)

Fonte: Autor

### 3.3.1 CONVOLUÇÕES

Conforme documentado na literatura, uma das principais camadas das redes CNN é a camada de convolução. (DALVI, 2018) afirma que a convolução é um operador linear, sendo um dos mais utilizados nas arquiteturas das CNN. Ela é implementada ao se deslizar sobre cada pixel da imagem um elemento estruturante ou máscara. Dessa forma, gera-se um mapa de características. Ela é definida matematicamente pela equação (14).

Nessa equação;  $I(x, y)$  representa a imagem,  $f(x, y)$  é o filtro ou *kernel*, o qual tem  $a$  e  $b$  como suas dimensões.

$$g(x, y) = I(x, y) * f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b I(s, t) * f(x - s, y - t) \quad (14)$$

Na Figura 10, é mostrada uma operação de convolução. Nela, temos uma matriz que corresponde a uma imagem em tons de cinza com dimensões de  $5 \times 5$  e também um filtro de tamanho  $2 \times 2$ . Na saída, obtemos o mapa de características com uma dimensão de  $4 \times 4$ . Além disso, podemos notar que cada valor obtido com essa operação é o resultado do somatório das multiplicações dos valores do filtro pelos respectivos valores da imagem da entrada.

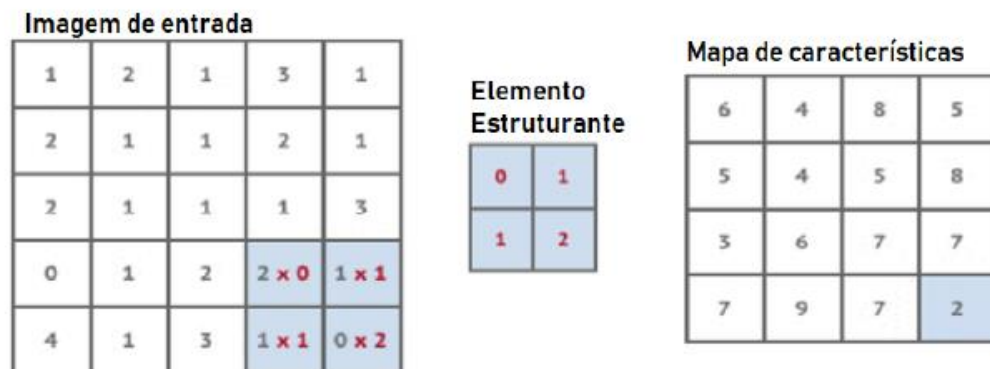


Figura 10– Exemplo de Operação de Convolução

Fonte: Dalvi (2018)

No trecho destacado na Figura 10, realizamos os seguintes cálculos:

$$2 * 0 + 1 * 1 + 1 * 1 + 0 * 2 = 2,$$

De forma semelhante, fizemos os cálculos para os demais elementos da matriz da imagem de entrada, multiplicando os pixels das imagens em outras posições pelos respectivos pesos do elemento estruturante.

O filtro ou *kernel* é formado por pesos que são inicializados aleatoriamente.

Esses pesos serão atualizados pelo método de propagação reversa, a cada etapa do treinamento, ao se apresentar um novo padrão na entrada da rede. A pequena região da entrada onde ele é aplicado é denominado de campo receptivo (*receptive field*). Nas Figuras 11 e 12, podemos observar melhor esse conceito.

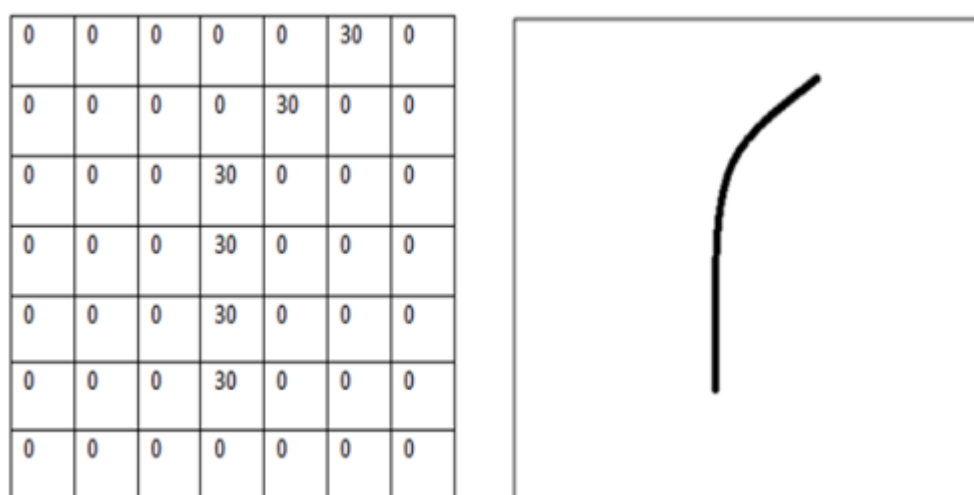


Figura 11– À esquerda, mostra-se o local de aplicação do filtro na imagem ou campo receptivo. À direita, é mostrada a matriz resultante da operação de convolução, onde os pixels da imagem filtrada coincidem com a posição dos pixels diferentes de zero na imagem original.

Fonte: Alves (2018)

Na imagem destacada na Figura 12, o campo receptivo é multiplicado pelo filtro ilustrado anteriormente.



Figura 12– Visualização do Filtro na Imagem

Fonte: Alves (2018)

Ao aplicar-se o filtro mostrado na Figura 11 na imagem da Figura 12, a operação resultará em um número bem alto, indicando uma grande semelhança entre o filtro e a região da imagem onde ele foi aplicado.

Caso não houvesse semelhança, a operação de filtragem resultaria em um valor muito baixo, próximo a zero.

De maneira semelhante às RNA, as CNN também utilizam funções não lineares de ativação.

Para conseguir realizar a simplificação das informações das camadas anteriores, as CNN fazem uso de uma camada de amostragem (*pooling*). Ela é estruturada de maneira similar à camada de convolução, sendo a operação de *pooling* aplicada a uma pequena região de neurônios de uma camada anterior. O resultado da operação produz um único valor real, que pode ser o maior valor (*maxpooling*), o valor médio (*average pooling*) etc.

Na Figura 13 mostramos a operação de *maxpooling*. O principal objetivo da operação de *pooling* é diminuir as dimensões de largura e altura da representação. Além disso, produz operações de detecção de características invariantes à translação e rotação.





Figura 13– *Maxpooling*

Fonte: Autor

As últimas camadas da rede CNN são as camadas inteiramente conectada (*Full Connected* - FC). Essas camadas correspondem a estruturas lineares de neurônios. Pode existir várias camadas FC. Se a camada FC corresponder à última camada da rede, o seu número de neurônios corresponde a quantidade de classes do problema de reconhecimento de padrões a ser resolvido. A camada FC é idêntica às camadas de neurônios utilizadas nas redes rasas.

### 3.3.2 CAMADA DE *DROPOUT*

As técnicas de *dropout*, regularização  $L_1$  e  $L_2$  e a parada antecipada são maneiras de tentar se mitigar os efeitos do *overfitting* (GOODFELLOW, BENGIO e COURVILLE, 2016), que ocorre quando a rede se adapta excessivamente aos dados na fase de treinamento. Com a técnica de *dropout*, algumas unidades da rede serão desativadas e não participarão efetivamente do treinamento, conforme mostrado na Figura 14. Dessa forma, foi usada a técnica de *dropout* no intuito de diminuir o *overfitting* nas arquiteturas CNN e LSTM ou GRU utilizadas neste trabalho.

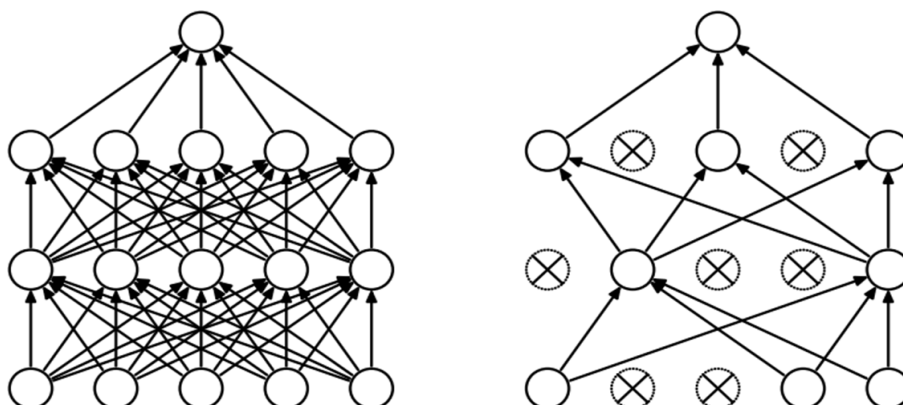


Figura 14–O resultado da aplicação do dropout na rede da esquerda

Fonte: SRIVASTAVA et al. (2014)

### 3.4 REDES RECORRENTES

#### 3.4.1 REDE NEURAIIS RECORRENTES (RNN)

(FERREIRA JÚNIOR, 2019) afirma que ao iniciar a abordagem de um novo tema ou assunto, os seres humanos começam a entendê-lo com base em suas experiências pretéritas.

Assim sendo, não começam completamente do zero a aprender um novo assunto.

Fazem continuamente uso de suas vivências que estão contidas nas suas lembranças ou memórias. Na verdade, essas informações persistem e provavelmente serão usadas posteriormente.

Redes neurais tradicionais não têm grande facilidade em reter dados e informações a longo prazo. Porém, as RNN superam esse tipo de limitação das RNA.

Elas são redes compostas por *loops*, os quais permitem que as informações persistam no sistema, conforme mostrado na Figura 15.

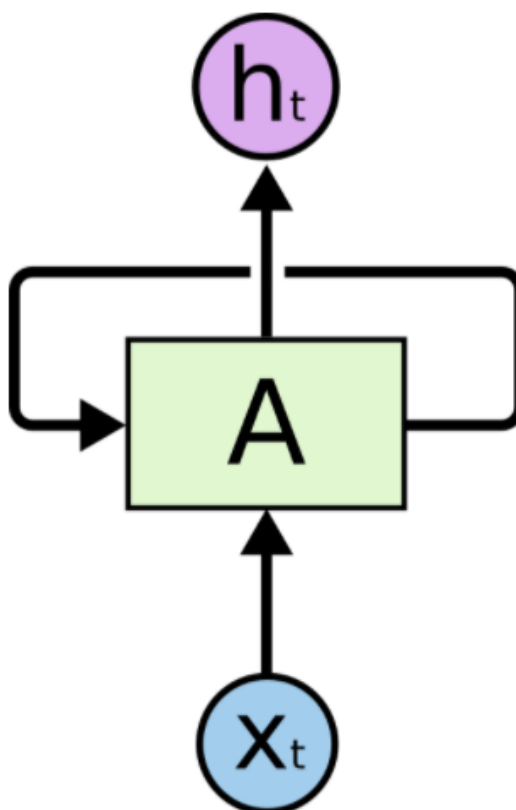


Figura 15– Modelo de redes recorrentes

Fonte: Ferreira Júnior (2019)

Na figura 15, uma parte da rede neural recebe uma entrada  $X_t$ , a qual será processada por A, que representa o estado interno atual da rede, gerando um valor  $h_t$ . O *loop* permite que as informações sejam passadas de uma etapa do treinamento ou da aplicação da rede, para a seguinte. (FERREIRA JÚNIOR, 2019) afirma que poderíamos abstrair uma RNN como múltiplas cópias dela mesma, com cada uma repassando uma mensagem a sua sucessora, como mostrado na Figura 16. Nessa representação, é como se o *loop* tivesse sido desmembrado ou até mesmo, fosse desenrolado.

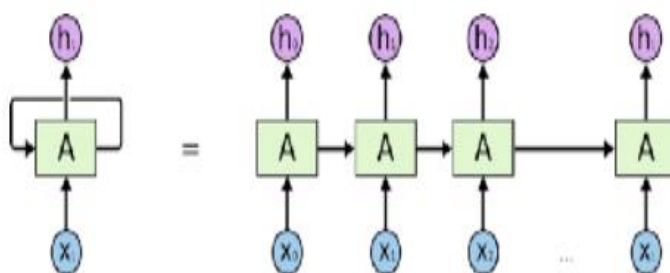


Figura 16– Redes recorrentes. Mostra-se o desmembramento ou “desenrolar” do laço

Fonte: Ferreira Júnior (2019)

Devido ao fato de possuir uma natureza de cadeia, RNN estão relacionadas com o processamento de sequências e listas. Elas são as opções naturais de escolha de arquitetura de redes neurais a serem usadas com esses tipos de dados. Assim, as RNN são utilizadas para o processamento dos seguintes sinais:

- 1) Reconhecimento de fala;
- 2) Modelagem de idiomas;
- 3) Traduções; e
- 4) Legendas de filmes.

### 3.4.2 ALGORITMO *BACKPROPAGATION THROUGH TIME* (BPTT)

O algoritmo BPTT é utilizado para realizar o treinamento da RNN e assim, minimizar a função de custo. Ele é uma adaptação do *backpropagation* (WERBOS,1990).

Assim, os pesos vão sendo ajustados ao longo de todo o intervalo de tempo, acrescentando os gradientes da função de custo em cada instante  $t$ . Em sequências muito longas ocorrem o problema do desvanecimento ou da explosão do gradiente, o que dificulta o processo de convergência do algoritmo de otimização (GOODFELLOW, BENGIO e COURVILLE, 2016). No intuito de superar essas limitações, Hochreiter e

Schmidhuber (1997) propuseram um tipo de neurônio denominado de *Long Short Term Memory*. Que será descrito detalhadamente na próxima seção.

### 3.4.3 LONG SHORT TERM MEMORY (LSTM)

A LSTM possui uma estrutura em cadeia com três blocos, conforme mostrado na Figura 17. Conforme mostrado no bloco central dessa Figura, as funções  $\sigma$  e  $\tanh$  (sigmóide e tangente hiperbólica) são utilizadas como funções de ativação.

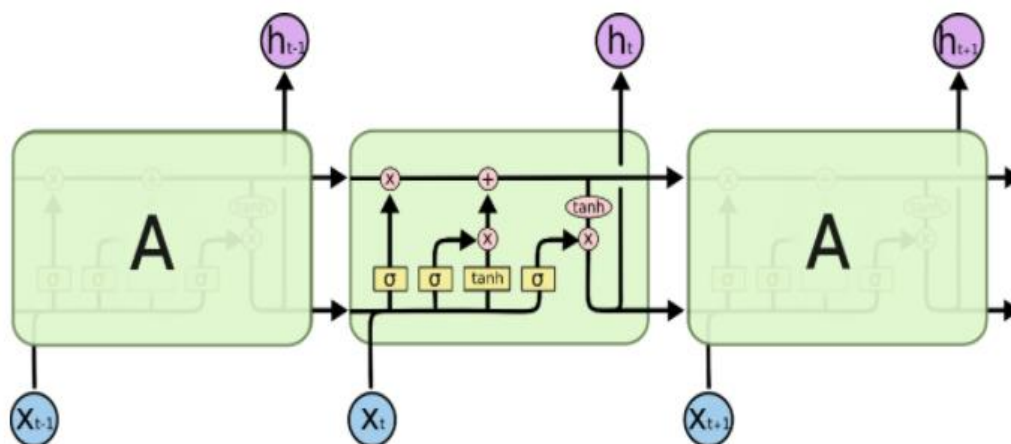


Figura 17– Esquema simplificado de uma LSTM

Fonte: Colah (2015)

Segundo (FERREIRA JÚNIOR, 2019), nas LSTM, a informação é retida pelas células e as manipulações de memória são feitas pelas portas (*gates*). Existem três tipos de portas:

- 1) *Forget Gate*: Essa porta remove as informações que não são mais úteis. As entradas  $x_t$  (no momento atual) e  $h_{t-1}$  (saída de célula anterior) são alimentadas na entrada da porta e são multiplicadas por matrizes de peso, seguidas pela adição da polarização.

Esse resultado é passado por uma função de ativação que resultará numa saída binária. Caso seja 0; a informação é esquecida e se for igual a 1, ela será mantida. A equação (15) resume essas informações.

$$f_t = \delta(W_f[h_{t-1}, x_t] + b_f) \quad (15)$$

Em que:

$f_t$  = Porta *forget gate*

$\delta$  = Função de ativação sigmóide

$W_f$  = Matriz dos pesos sinápticos

$b_f$  = Vetor de polarizações

$h_{t-1}$ =Saída de célula anterior

$x_t$  = Entrada no momento atual

- 2) *Input Gate*: Essa porta realiza a adição das informações úteis ao estado da célula. Ela utiliza a função sigmoide, que filtra os valores a serem lembrados. Essa porta utiliza as entradas  $h_{t-1}$  e  $x_t$  de forma semelhante ao *forget gate*. Na sua saída, um vetor é criado utilizando a função tangente hiperbólica, devido as características dessa função, a sua saída limita-se entre -1 a +1, conforme ilustrado pelas equações (16) e (17).

$$i_t = \delta(W_i[h_{t-1}, x_t] + b_i) \quad (16)$$

$$\hat{C}_t = \tanh(W_{\hat{C}}[h_{t-1}, x_t] + b_{\hat{C}}) \quad (17)$$

Em que:

$i_t$  = Porta *input gate*

$W_i$  = Matriz dos pesos sinápticos

$b_i$  = Vetor de polarizações

$h_{t-1}$  = Saída de célula anterior

$x_t$  = Entrada no momento atual

$\hat{C}_t$  = Vetor das possibilidades dos valores da entrada

$W_{\hat{C}}$  = Matriz dos pesos sinápticos

$b_{\hat{C}}$  = Vetor de polarizações

- 3) *Output Gate*: Essa porta tem como tarefa extrair informações relevantes a respeito do estado da célula atual para ser apresentada em sua saída. Primeiro, um vetor é gerado, aplicando na célula a função tangente hiperbólica. Então, a informação é ajustada usando a função sigmoide, que filtra os valores a serem lembrados usando as entradas  $h_{t-1}$  e  $x_t$ . Os valores do vetor e dos regulados são multiplicados para serem enviados como uma entrada para a próxima célula, conforme mostrado na equação (18).

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (18)$$

Em que:

$C_t$  = Vetor de estado atual

$C_{t-1}$  = Vetor de estado anterior

## 4 MATERIAIS E MÉTODOS

### 4.1 MATERIAIS

Relembramos que o objetivo principal dessa dissertação de mestrado é propor uma arquitetura de rede convolucional 2D, utilizando imagens do sinal de ECG convertida para uma imagem 2D, para classificação de arritmias.

Os sinais de ECG utilizados, quer sejam sinais normais ou sinais de arritmias, foram extraídos da base de dados “MIT-BIH Arrhythmia Database”. Nessa base, os sinais de ECG foram obtidos de 47 pacientes, sendo amostrados a uma taxa de 360 Hz.

A base é composta por 48 registros anotados, com duração aproximada de 30 minutos de batimento cardíaco. As 17 classes estão presentes no banco de dados. Essas classes foram citadas no item 3.1 desse trabalho.

Para o treinamento e validação de aprendizado de máquina propostas nesse trabalho, redes CNN 1D, CNN-2D, LSTM e GRU, um conjunto de 2550 registros foi utilizado. Esse número é obtido dos seguintes sinais: o sinal normal, o sinal de marcapasso e os sinais de 15 arritmias, totalizando 17 tipos de sinais de ECGs diferentes.

Para cada tipo de sinal de ECG foram extraídos 150 registros, totalizando  $150 \times 17 = 2550$  registros. A Figura 18 mostra as etapas necessárias para realizar essas extrações.

Embora alguns tipo de sinais de ECG oferecessem a possibilidade de extração de mais registros, o número de 150 registros uniformizou todos os tipos de sinais de ECG, já que algumas arritmias têm menos registros disponíveis para extração. Cada registro de um tipo de sinal de ECG contém 225 amostras. No Apêndice II, nas Figuras de 30 a 46 são mostrados registros, com 225 amostras, referentes aos sinais de ECG de cada arritmia utilizada neste trabalho. Como observado nessas Figuras, as 225 amostras extraídas de cada arritmia são suficientes para a representação das mesmas. As 225 amostras de um tipo de sinal de ECG são extraídas atendendo as seguintes condições:

- 1) As 225 amostras são extraídas em torno de um pico R, não necessariamente com ele centralizado;
- 2) As amostras extraídas devem conter a arritmia. Cada registro com 225 amostras foi utilizado como entrada de uma das arquiteturas utilizadas nesse trabalho: CNN-2D, LSTM e GRU. Na primeira, o sinal unidimensional é convertido em uma imagem. Já as

redes recorrentes, LSTM e GRU, utilizaram um sinal unidimensional com 225 amostras.

A Figura 19 ilustra o processo de conversão de um registro em uma imagem com formato jpg, para ser utilizada na rede CNN-2D. Conforme mostrado, os primeiros 15 valores de um registro constituem a primeira linha da matriz em nível de cinza (0 – 255), a partir da qual é gerada a imagem.

Os 15 valores seguintes constituem a segunda linha dessa matriz, e assim por diante. Todos os registros de sinais de ECG foram normalizados na faixa entre 0 e 255, através da equação (19).

$$F(x) = \frac{I - I_{min}}{I_{max} - I_{min}} * 255 \quad (19)$$

Em que:

$I$  = Valor a ser normalizado do Sinal de ECG;

$I_{min}$  = Mínimo do Sinal de ECG;

$I_{max}$  = Máximo do Sinal de ECG;

$F_{max} = 255$ ; e

$F(x)$  = Valor normalizado do Sinal de ECG.

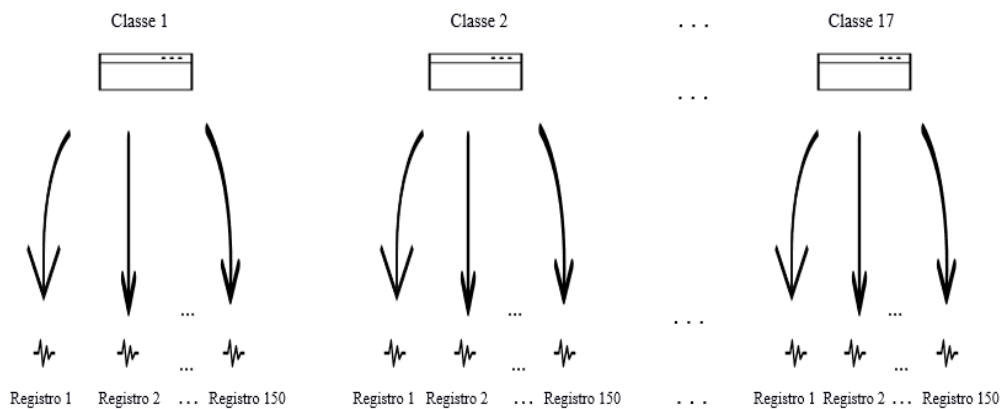


Figura 18 - Construção dos Registros e da Base de Dados

Fonte: Autor



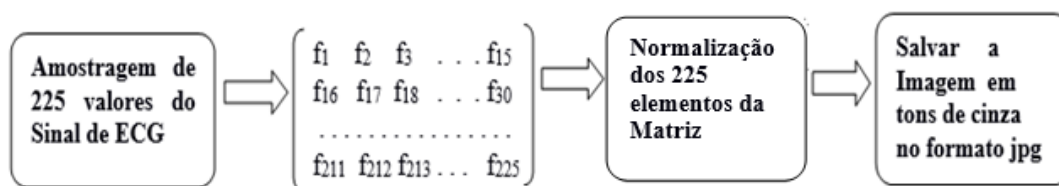


Figura 19- Etapas para realizar a construção das imagens

Fonte: Autor

O processo de conversão de um registro de 225 coordenadas em uma imagem é ilustrado através das Figuras 20, 21, 22, 23, 24 e 25. Na Figura 20, mostra-se um sinal de ECG com 13 picos QRS. É importante ressaltar que todas as imagens dos sinais ou registros de ECG utilizados neste trabalho estão apresentados em *adu* (*analog to digital units*); correspondendo a uma unidade física, que nesse caso é o mV (MOODY). A partir da Figura 20, extraiu-se um registro com 225 coordenadas, conforme mostrado na Figura 21. Na Figura 22, mostra-se a matriz 15x15 com os valores desse registro ainda não normalizados. Na faixa de 0-255. A Figura 23 mostra o registro extraído, já normalizado na faixa de 0-255.

A Figura 24 mostra a matriz 15x15 com os valores normalizados, mostrados na Figura 23. Finalmente, a Figura 25 mostra a imagem jpg, obtida através da conversão da matriz mostrada na Figura 24 em uma imagem. É importante observar que os picos QRS estão relacionados com as regiões mais claras das imagens geradas; enquanto os vales do sinal de ECG amostrado e normalizado, com as áreas mais escuras.

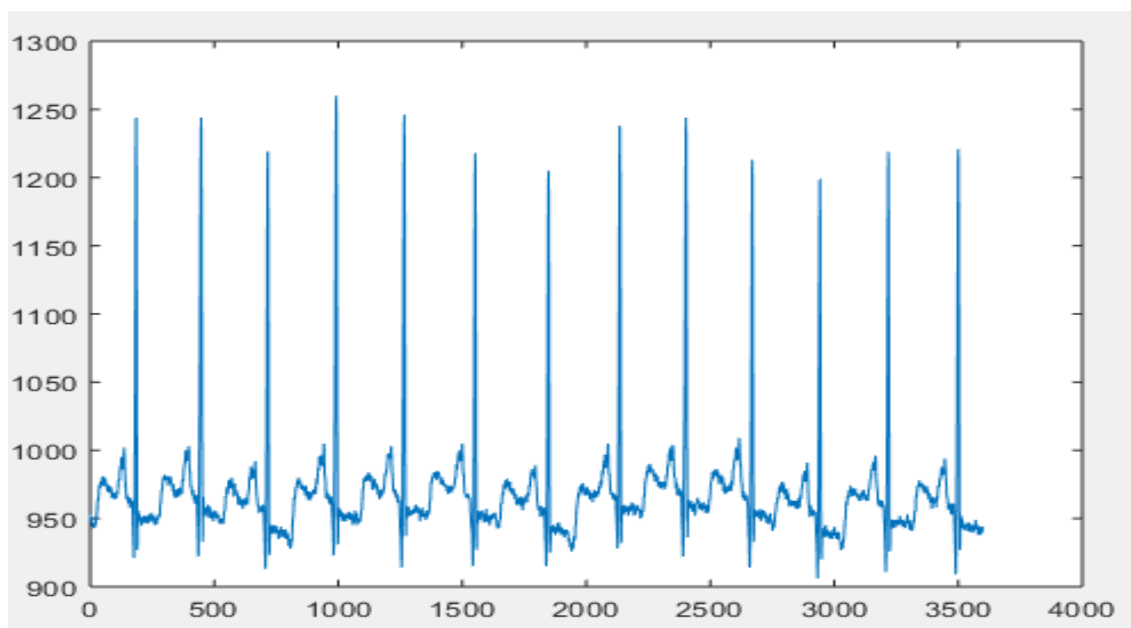


Figura 20- Sinal Original de ECG

Fonte : Autor

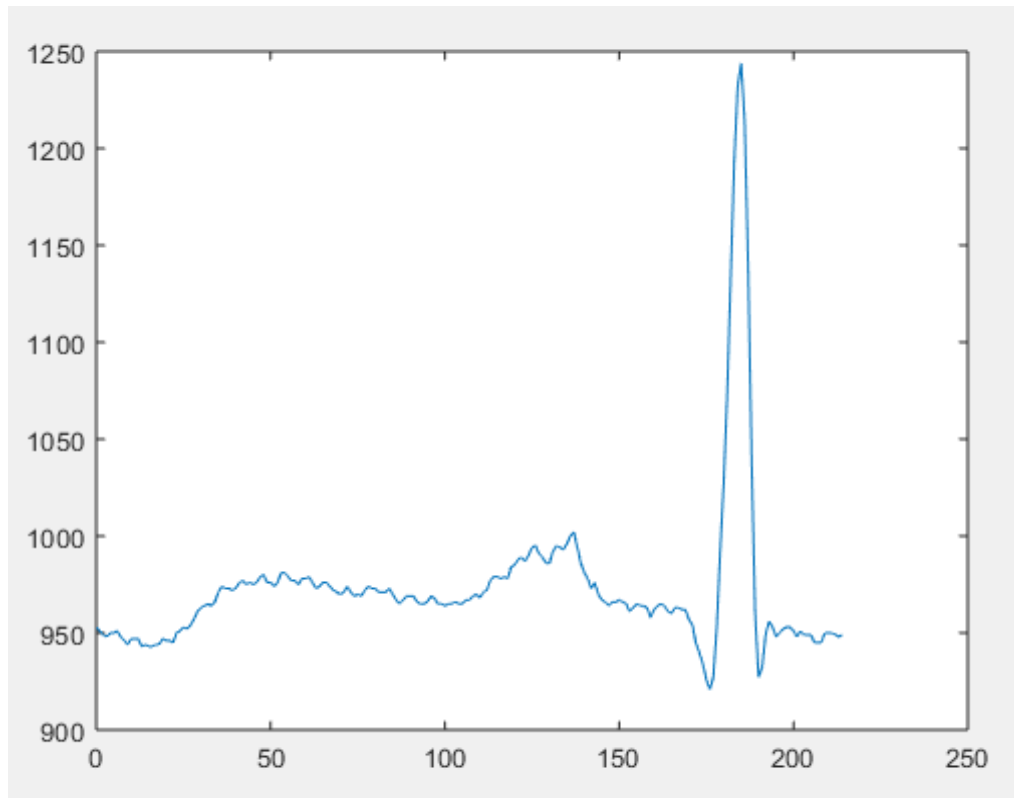


Figura 21- Sinal de ECG Amostrado

Fonte: Autor

953	951	949	948	950	950	951	948	946	944	947	947	947	943	944
943	943	944	944	947	946	946	945	950	951	953	952	954	957	961
963	964	965	964	966	971	974	973	973	972	973	976	977	975	976
975	976	979	980	976	976	974	976	981	981	979	977	977	975	978
978	979	976	973	974	976	976	974	972	971	970	971	974	971	969
970	969	972	974	973	973	971	971	971	973	970	967	965	967	969
969	969	966	965	965	966	969	968	965	965	964	965	965	966	965
965	967	967	969	970	968	971	972	977	979	979	978	979	978	984
985	988	989	987	990	994	995	991	989	986	986	992	995	994	993
996	1000	1002	994	986	981	978	973	976	970	967	966	964	966	966
967	966	965	961	963	965	964	964	963	958	962	964	965	964	961
960	963	963	962	962	957	954	945	940	934	926	921	927	953	993
1028	1074	1134	1193	1232	1244	1214	1139	1039	960	927	932	949	956	953
948	950	952	953	953	951	948	951	949	949	949	945	945	945	950
950	950	949	948	949	951	952	951	950	947	947	950	952	951	951

Figura 22- Matriz dos Elementos Amostrados do Sinal de ECG

Fonte: Autor

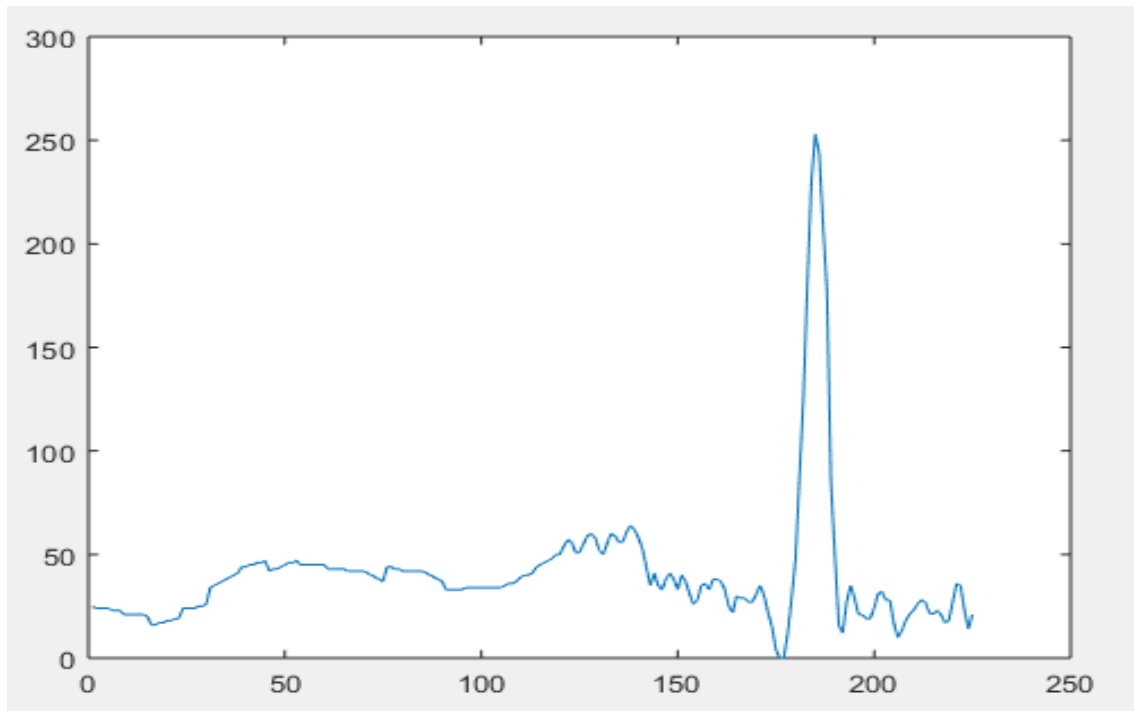


Figura 23- Imagem do Sinal de ECG Normalizado

Fonte: Autor

26	24	23	22	23	23	24	22	20	19	21	21	21	18	19
18	18	19	19	21	20	20	19	23	24	26	25	27	29	32
34	34	35	34	36	40	42	42	42	41	42	44	45	43	44
43	44	46	47	44	44	42	44	48	48	46	45	45	43	46
46	46	44	42	42	44	44	42	41	40	39	40	42	40	38
39	38	41	42	42	42	40	40	40	42	39	37	35	37	38
38	38	36	35	35	36	38	38	35	35	34	35	35	36	35
35	37	37	38	39	38	40	41	45	46	46	46	46	46	50
51	53	54	53	55	58	59	56	54	52	52	57	59	58	57
60	63	64	58	52	48	46	42	44	39	37	36	34	36	36
37	36	35	32	34	35	34	34	34	30	33	34	35	34	32
31	34	34	33	33	29	27	19	16	11	4	0	5	26	57
85	121	169	215	246	255	232	173	94	31	5	9	23	28	26
22	23	25	26	26	24	22	24	23	23	23	19	19	19	23
23	23	23	22	23	24	25	24	23	21	21	23	25	24	24

Figura 24- Matriz do Sinal de ECG Normalizado

Fonte: Autor

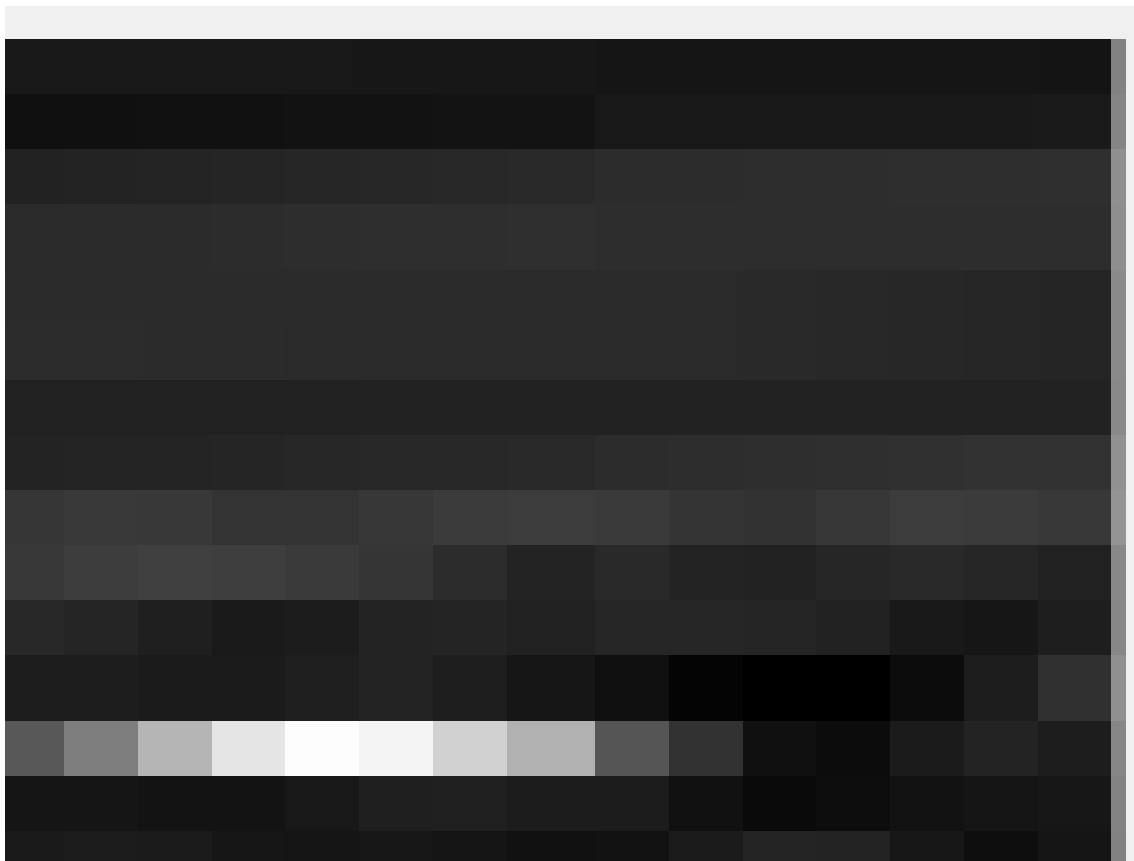


Figura 25- Imagem Gerada da Matriz Normalizada em Tons de Cinza

Fonte: Autor

## 4.2 MÉTODOS

### 4.2.1 ARQUITETURA DAS REDES BIDIMENSIONAIS

As redes convolucionais propostas para classificação das 17 arritmias são constituída por 20 camadas. A arquitetura dessas redes é mostrada na Figura 26. A extração de características é efetuada através de CONV-NORMALIZAÇÃO EM LOTE – RELU. As camadas de convolução da rede CNN-2D têm dimensão de 96x96 e utilizam filtros de tamanho 2x2, aplicados com passo igual a 1. Vários tamanhos de camada foram experimentados, sendo os melhores resultados obtidos com o tamanho de 96x96. O filtro de tamanho 2x2 também foi o que apresentou o melhor resultado, possivelmente pela razão do tamanho da imagem ser muito pequena. O passo de 1 faz com que não haja redução do mapa de características durante o processo convolutivo. Optamos por reduzir o tamanho do mapa de características apenas com as camadas de amostragem. As camadas de convolução da rede CNN-1D são constituídas por 96 neurônios e filtros de tamanho 2x1

aplicados com passo igual a 1. A camada de normalização em lote foi utilizada para evitar o problema de explosão ou desvanecimento do gradiente, que ocorrem em redes muito profundas.

A camada RELU é a camada de ativação, responsável pelo processamento não linear e está definida na equação (5).

Após os 5 blocos utilizados para extração de características, temos as camadas responsáveis pela classificação da rede convolutivas, 2 camadas inteiramente conectadas (*Full Connected* - FC), constituída por 200 neurônios cada. Após a primeira camada inteiramente conectada, temos uma camada de *dropout*. A camada de *dropout* foi introduzida com o objetivo de melhorar o poder de generalização da rede. Por fim, temos a camada *Softmax* e a camada de classificação, com 17 neurônios de saída. Com essa arquitetura de rede, obtivemos um excelente desempenho.



Figura 26– Arquitetura das Redes CNN-1D e CNN-2D  
Fonte: Autor

#### 4.2.2 ARQUITETURA DAS REDES LSTM E GRU

A arquitetura das redes LSTM e GRU utilizadas é mostrada na Figura 27. Trata-se de uma arquitetura semelhante para as duas redes, a exceção da primeira camada de neurônios, que em uma delas é uma camada LSTM, enquanto que na outra é uma camada GRU. Trata-se de uma arquitetura muito simples, constituída por 6 camadas.

Inicialmente, temos uma camada LSTM ou GRU, com 4096 neurônios. Outros tamanhos de camada LSTM e GRU foram avaliados, como 1024 e 2048, porém os melhores resultados foram obtidos com 4096 neurônios. Essa camada é seguida por uma outra inteiramente conectada (*Full Connected* - FC), uma camada de *dropout*, uma camada inteiramente conectada com 17 unidades, uma camada *softmax* e, por último, a camada de classificação. A Figura 27 mostra essa arquitetura.

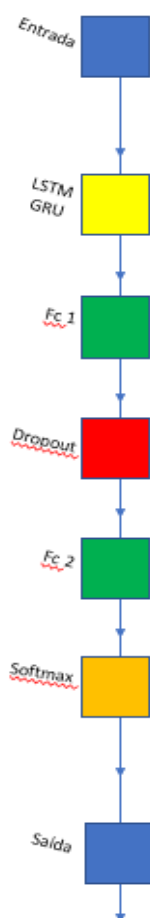


Figura 27– Arquitetura da Rede LSTM ou GRU

Fonte: Autor

#### 4.2.3 MÉTRICAS DE AVALIAÇÃO

As seguintes métricas foram usadas na avaliação do desempenho das redes profundas: acurácia (*Acc*), sensibilidade (*Sens*) ou *recall*, especificidade (*Spec*), precisão (*P*) e F1-Score. De acordo com (RODRIGUES, 2019), a acurácia fornece uma visão geral do desempenho da arquitetura da rede neural utilizada; indicando, entre todas as classificações que foram realizadas, quantas classes foram corretamente classificadas por ela. Essa métrica está descrita na equação 20.

$$Acc = \frac{1}{17} \sum_{c=1}^{17} \frac{T_p^c + T_n^c}{T_p^c + T_n^c + F_p^c + F_n^c} \quad (20)$$

Conforme (RODRIGUES, 2019), a sensibilidade nos informa quantos positivos que eram esperados de fato; foram corretamente classificados pela rede, entre todos os valores positivos obtidos, ou seja, verdadeiros positivos e falsos negativos. Essa métrica está descrita na equação 21.

$$Sens = \frac{1}{17} \sum_{c=1}^{17} \frac{T_p^c}{T_p^c + F_n^c} \quad (21)$$

(MARIANO, 2021) afirma que a especificidade avalia a capacidade que a rede neural possui em detectar resultados negativos, os quais foram corretamente previstos. A especificidade está descrita na equação 22.

$$Spec = \frac{1}{17} \sum_{c=1}^{17} \frac{T_n^c}{T_n^c + F_p^c} \quad (22)$$

(RODRIGUES, 2019) informa que a precisão nos fornece quantas classificações positivas que foram corretamente previstas pela rede. A precisão está descrita na equação 23.

$$P = \frac{1}{17} \sum_{c=1}^{17} \frac{T_p^c}{T_p^c + F_p^c} \quad (23)$$

(RODRIGUES, 2019) afirma que a F1-Score representa a média harmônica entre a precisão e a sensibilidade. Ele explica, ainda, que essa métrica é usada no intuito de substituí-las por um único valor. A F1-Score está descrita na equação 24.

$$F1 - Score = 2 * \frac{(P * Sens)}{(P + Sens)} \quad (24)$$

Em que:

$T_p^c$  representa os verdadeiros positivos: todas as instâncias da classe  $c$  que foram classificadas corretamente como sendo da classe  $c$ ;

$T_n^c$  representa os verdadeiros negativos: todas as instâncias que não pertencem à classe  $c$  que não foram classificadas como sendo da classe  $c$ ;

$F_p^c$  representa os falsos positivos: todas as instâncias que não pertencem à classe  $c$  que foram classificadas como sendo da classe  $c$ ; e

$F_n^c$  representa os falsos negativos: todas as instâncias da classe  $c$  que não foram classificadas como sendo da classe  $c$ .

#### 4.2.4 METODOLOGIA DE TREINAMENTO E VALIDAÇÃO

A metodologia usada para treinamento e teste foi a validação cruzada de 5 pastas.

Em cada pasta, 80% dos dados são usados para treinamento e 20%, para validação. No treinamento, após uma busca exaustiva, foram escolhidos os seguintes parâmetros: o tamanho do lote utilizado foi igual a 512; para tentar mitigar o *overfitting* foi usado o fator de regularização  $L_2$  igual a 0,1 e a taxa de *dropout* igual a 0,5; número máximo de épocas igual a 6.000, em que cada época consiste de um ciclo de treinamento completo. Para a taxa de aprendizagem, foi empregado um esquema de diminuição progressiva. A taxa de aprendizagem inicial utilizada foi de 0,01, com um período de queda igual a 200 épocas e um fator de queda igual a 0,75. Esses parâmetros foram utilizados no intuito de melhorar o desempenho da rede durante a fase do treinamento, diminuindo as oscilações e permitindo uma convergência mais rápida.

Os trabalhos de simulação dessa dissertação foram realizados utilizando-se o software Matlab<sup>®</sup> versão 2018b. A máquina utilizada possuía um processador Intel i7-8700 com uma velocidade de 3.2 GHz, com 16 GB de RAM e GPU de 8 GB GeForce GTX 1070.



## 5 RESULTADOS

Inicialmente, na Figura 28 é mostrada a curva de convergência da acurácia da rede CNN-2D, que apresentou melhor desempenho das redes analisadas neste trabalho.

No início da fase do treinamento, em virtude da alto valor da taxa de aprendizagem, ocorreram muitas oscilações. No entanto, como a taxa de aprendizagem diminui com o aumento das épocas, essas oscilações foram diminuindo.

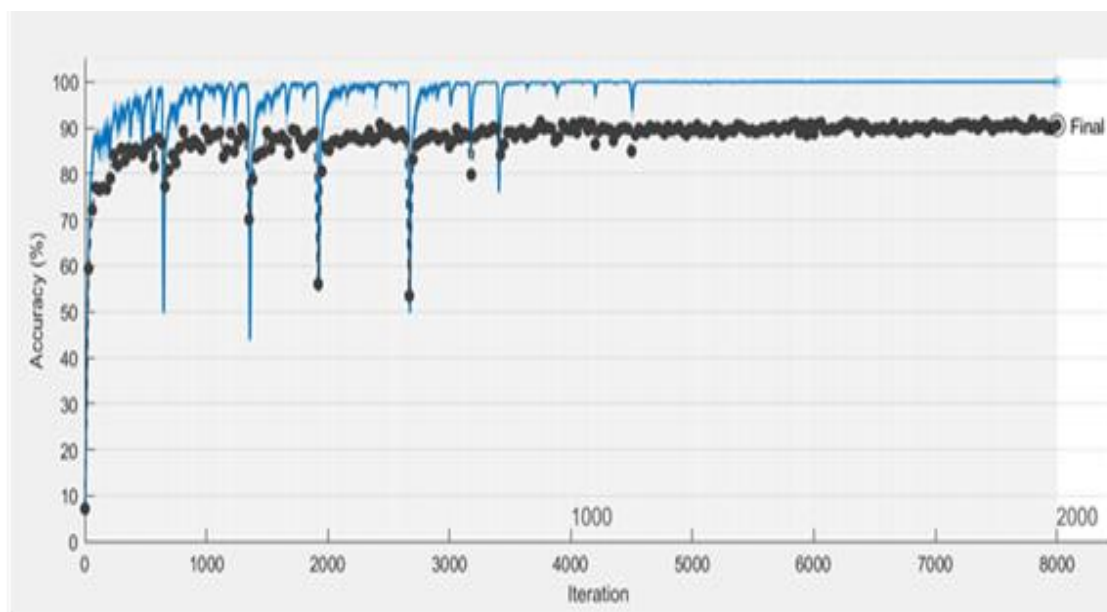


Figura 28-Desempenho da Arquitetura da Rede CNN

Fonte: Autor

Na Tabela 1 mostra-se o resultado do desempenho dos quatro tipos de redes neurais profundas propostas para a classificação de arritmias neste trabalho CNN-1D, CNN-2D, LSTM e GRU. Conforme mostrado, a rede que apresentou os melhores resultados para todas as métricas foi a rede CNN-2D, com uma acurácia de 99,13% e uma precisão de 92,31%. A rede com pior desempenho foi a rede CNN-1D, com acurácia de 94,29% e precisão de 52,27%. As redes recorrentes, LSTM e GRU, tiveram um desempenho intermediário entre as duas redes anteriormente citadas.

Na Figura 29 mostramos uma matriz de confusão, com os valores de precisão para as 17 classes de arritmias, obtidos com a rede que apresentou o melhor desempenho, a rede CNN-2D. Todas essas arritmias estão elencadas no Apêndice I deste trabalho.

Tabela 1- Resultado das 4 arquiteturas de redes profundas avaliadas: CNN-1D, CNN-2D, LSTM e GRU

Arquitetura	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	Precisão (%)	F1-Score (%)
1D CNN	94,29± 2,14	47,64± 18,05	97,00 ± 1,92	52,27±27,28	49,85±22,37
2D CNN	99,13±0,65	93,07±6,10	99,52±0,42	92,31±6,87	92,68±5,64
LSTM	96,21±1,65	64,65±15,15	98,09±1,48	68,13±24,72	66,34±18,53
GRU	95,71±1,85	62,99±16,12	97,70±1,55	63,51±24,19	63,24±18,35

		Confusion Matrix																	
Output Class	AFIB	26 5.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	32.9% 7.1%	
	AFL	0 0.0%	22 4.3%	1 0.2%	1 0.2%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	34.6% 15.4%	
	APB	2 0.4%	0 0.0%	29 5.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	35.3% 14.7%	
	B	0 0.0%	6 1.2%	0 0.0%	28 5.5%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	30.0% 20.0%	
	F	0 0.0%	1 0.2%	0 0.0%	0 0.0%	29 5.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	36.7% 3.3%	
	IVR	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	27 5.3%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	33.1% 6.9%	
	LB BBB	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	24 4.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	2 0.4%	0 0.0%	0 0.0%	0 0.0%	32.8% 17.2%	
	NSR	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	29 5.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	33.5% 6.5%	
	PR	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	1 0.2%	0 0.0%	29 5.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	33.5% 6.5%	
	PVC	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	30 5.9%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	33.8% 6.3%	
	RBBB	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	29 5.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	36.7% 3.3%	
	SDHB	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	26 5.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	36.3% 3.7%	
	SVTA	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	27 5.3%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%	
	T	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	29 5.7%	0 0.0%	0 0.0%	100% 0.0%	
	VFL	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	30 5.9%	0 0.0%	36.8% 3.2%	
	VT	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	30 5.9%	100% 0.0%	
	WPW	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	37.1% 12.9%	
		36.7%	73.3%	96.7%	93.3%	96.7%	90.0%	90.0%	96.7%	100%	96.7%	96.7%	90.0%	96.7%	100%	100%	90.0%	92.4%	
		13.3%	26.7%	3.3%	6.7%	3.3%	10.0%	20.0%	3.3%	3.3%	0.0%	3.3%	13.3%	10.0%	3.3%	0.0%	10.0%	7.6%	
		AFIB	AFL	APB	B	F	IVR	LB BBB	NSR	PR	PVC	RBBB	SDHB	SVTA	T	VFL	VT	WPW	
		Target Class																	

Figura 29- Matriz de Confusão mostrando os valores de precisão para as 17 arritmias, obtidos com a rede CNN-2D

Fonte: Autor

A Tabela 2 mostra uma comparação do desempenho da rede CNN-2D com resultados obtidos em trabalhos previamente publicados na literatura, que também realizaram a classificação das 17 arritmias classificadas no trabalho, ora apresentado e que também utilizaram o banco de dados utilizado nesse trabalho, o banco de dados “MIT – BIH database”. Ressalta-se que esses trabalhos foram revisados no capítulo de revisão da literatura. Através dos resultados apresentados na Tabela 2, verifica-se que a técnica de classificação proposta nesse trabalho, utilizando redes CNN-2D encontra-se no estado da arte.

Tabela 2- Comparação dos resultados obtidos com a rede CNN-2D com outros resultados publicados

Referência	Ano	Precisão (%)
YILDIRIM et al.	2018	91,33
PLAWIAK	2017	91,40
Método Proposto (2D CNN)	2021	92,31

Fonte: Autor

## 6 CONCLUSÕES

Neste trabalho, propusemos uma nova metodologia para detecção de arritmias em sinais de ECG por meio de arquiteturas profundas CNN-2D. Os resultados obtidos foram comparados com outras arquiteturas profundas de redes CNN-1D, LSTM e GRU. As principais características do método proposto foram as seguintes: avaliação de apenas um registro do ecocardiograma, ou seja, de apenas uma pulsação, com 225 amostras; conversão desse registro em uma imagem de 15x15 pixels. Os resultados obtidos com a rede CNN-2D foram muito superiores àqueles obtidos com as arquiteturas CNN-1D, LSTM e GRU. A rede CNN-2D obteve uma acurácia de 99,13 % e uma precisão de 92,31%. Esses valores colocam o método proposto no estado da arte, quando comparado a outros resultados previamente publicados na literatura. Em trabalhos futuros, para melhorar os desempenhos obtidos, pretendemos utilizar um conjunto de dados maior, usando uma ou mais técnicas de aumento de dados, nos casos em que não se consiga mais dados. Para isso, pretendemos utilizar *flipping*, que consiste no espelhamento das imagens na vertical ou na horizontal; para girá-las, a técnica de rotação; para deslocá-las, a de translação; para incliná-las, a de cisalhamento (*shear*); para aumentá-las ou diminuí-las, a de escala (*zoom*); entre outras técnicas. Por fim, implementar o método proposto em dispositivos móveis, que acessarão a aplicação via web.

Registramos, por fim, que tanto o objetivo principal quanto todos os objetivos específicos propostos foram cumpridos na íntegra.

## REFERÊNCIAS

ABDULLAH, L. A.; AL-ANI, M. S. CNN-LSTM based model for ECG arrhythmias and myocardial infaction classification. v. 5, p. 601-606, 2020.

ALVES, G. **Entendendo redes neurais convolucionais (CNN)**. 8 out. 2018. Disponível em <https://medium.com/neuronio-br/entendendo-redes-convolucionais-cnns-d10359f21184> . Acesso em 16 fev. 2021.

AZEVEDO, L. P. **Aplicação de redes neurais artificiais no processo de orquídeas do gênero cattleya**. Disponível em: <https://www.ifmg.edu.br/sabara/biblioteca/trabalhos-de-conclusao-de-curso/tcc-documentos/TCCLucasAzevedo.pdf>. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, Sabará, p. 49, 2016.

BOSCARIOLI, C.; DA SILVA, L.A.; PERES, S. M. **Introdução à mineração de dados: com aplicações em R**. Elsevier. 2016.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B.. **Redes Neurais Artificiais: Teoria e Aplicações**, 2 ed., Rio de Janeiro – RJ: Editora LTC, 2007.

CHAGAS, E. T. D. O. **Avaliação de imóveis utilizando redes neurais artificiais** – revisão sistemática de literatura. Revista Científica Multidisciplinar Núcleo do Conhecimento, n. 4, p. 55-75, abr. 2019. Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/redes-neurais-artificiais>. Acesso em 16 ago. 2021.

COLAH, C. **Understanding LSTM networks**. 27 ago. 2015. Disponível em: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. 22 mar. 2021.

COSTA, E. M. M.; LUDWIG JÚNIOR, O. **Redes Neurais: Fundamentos e aplicações com programas em C**. Rio de Janeiro. Ciência Moderna. 2007.

DALVI, R. F. **Suporte a detecção e classificação de câncer a partir de mamografias digitalizadas e redes neurais convolucionais**. Dissertação (Mestrado em Engenharia Elétrica) - Centro Tecnológico, Universidade Federal do Espírito Santo. Vitória, p. 106. 2018.

DEXTRO, R. B. **Anatomia do coração humano**. Disponível em: [www.infoescola.com/anatomia-humana/coracao/](http://www.infoescola.com/anatomia-humana/coracao/). Acesso em 19 jan.2021.

DOS SANTOS, V. S. **"Coração humano"**; Brasil escola. Disponível em: <https://brasilecola.uol.com.br/biologia/coracao-humano.htm>. Acesso em 17 dez. 2020.

FERREIRA JÚNIOR, J. R. **Redes neurais recorrentes**. 11 jun. 2019. Disponível em: <https://medium.com/@web2ajax/redes-neurais-recorrentes-lstm-b90b720dc3f6>. Acesso em 21 fev. 2021.

FRANÇA, D. **O que um estudante de medicina precisa saber sobre doenças cerebrovasculares?** 01 out. 2018. Disponível em: <https://www.bestmed.com.br/blog/doencas-cerebrovasculares/>. Acesso em 01 nov. 2021.

GAO, J.; et al. **An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset.** Hindawi Journal of healthcare engineering, v.2019, p. 10, 2019.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press. 2016.

GUYTON, A. C. Tratado de Fisiologia Médica. 9 ed. Guanabara Koogan.

HAYKIN, S. **Redes neurais: Princípios e práticas.** 2 ed. New Jersey: Pearson Education, 2008.

HOCHREITER, S.; SCHIMIDHUBER, J. **Long short term memory.** Neural computation, MIT Press, v. 9, n. 8, p. 1735-1780, 1997.

YILDIRIM, Ö. et al. **Arrhythmia detection using deep convolutional neural network with long duration ECG signals.** Computers in Biology and Medicine. v. 102, p. 411-420, 2018.

LANHELLAS, R. **Redes neurais artificiais: Algoritmo backpropagation.** Disponível em: <https://devmedia.com.br/redes-neurais-artificiais-algoritmo-backpropagation/28559>. 2013. Acesso em 29 mar. 2021.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. **Convolutional Networks and Applications in Vision.** 2010.

MAGALHÃES, L. **Sistema circulatório.** 12 nov. 2020, Disponível em: <https://www.todamateria.com.br/sistema-circulatorio/#:~:text=No%20cora%C3%A7%C3%A3o%2C%20o%20sangue%20arterial,os%20diversos%20tecidos%20do%20corpo>. Acesso em 17 dez. 2020.

MARIANO, D. Métricas de avaliação em machine learning. 25 abr. 2021. Disponível em: <https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>. Acesso em 13 dez. 2021.

MINSKY, M.; PAPERT, S. **Perceptrons: an introduction to computational geometry.** Mit Press. 1969.

MITCHELL, L. B. **Fibrilação atrial. Manual MSD.** jul. 2019. Disponível em: <https://www.msdmanuals.com/pt-br/profissional/doen%C3%A7as-cardiovasculares/arritmias-e-doen%C3%A7as-de-condu%C3%A7%C3%A3o/fibrila%C3%A7%C3%A3o-atrila-fa>. Acesso em 26 ago. 2021.

MITCHELL, L. B. **Batimentos Atriais Prematuros.** Manual MSD. fev. 2021. Disponível em: <https://www.msdmanuals.com/pt-br/casa/dist%C3%BArbiosdocora%C3%A7%C3%A3o-e-dos-vasos>

sangu%C3%ADneos/arritmias-card%C3%ADacas/batimentos-atriais-prematurados. Acesso em 26 ago. 2021.

MOODY, G. B. **Glossary**. Disponível em: [http://ecg.mit.edu/dbpg/dbu\\_105.htm](http://ecg.mit.edu/dbpg/dbu_105.htm). Acesso em 09 dez. 2021.

MUSTAQUEEM, A.; Anwar, S. M.; Majid, M. **A modular cluster based collaborative recommender system for cardiac patients**, Artif. Intell. Med., vol. 102, nov. 2019, p. 101761, 2020.

OLIVEIRA, A. T. **Classificação automática de arritmias cardíacas usando uma combinação de redes neurais**. Tese (Doutorado em Engenharia Mecânica com ênfase em Mecatrônica) – Faculdade de Engenharia Mecânica, Campinas, p. 81. 2019.

PAEZ, R. **Doenças e sintomas**: Entenda o que são as doenças isquêmicas do coração. 09 abr. 2020. Disponível em: <https://www.rodrigopaez.com.br/publicacoes/entenda-o-que-sao-as-doencas-isquemicas-do-coracao/>. Acesso em 01 nov. 2021.

PECHARKI, C. H.S.; STEMPI, E. E.; WITT, F. A. **Sistema portátil para monitoramento contínuo da atividade elétrica cardíaca**. Trabalho de Conclusão de Curso (Graduação em Engenharia Industrial Elétrica com Ênfase em Eletrônica) - Universidade Tecnológica Federal do Paraná, Curitiba, p. 108. 2013.

PETMEZAS, G. et al. **Automated atrial fibrillation detection using a hybrid CNN-LSTM Network on Imbalanced ECG Datasets**. Biomedical signal processing and control. v. 63, p. 102194, 2020.

PHYSIONET. **MIT-BIH Arrhythmia database**. Disponível em: <https://www.physionet.org/content/mitdb/1.0.0/>. Acesso em 28 fev. 2021.

PLAWIAK, P. et al. **Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system**. Expert Systems with Applications. v. 92, p. 334-349, 2017.

RODRIGUES, A. **Direto ao ponto**: Síndrome Wolf-Parkinson-white. 2018. Disponível em: <https://blog.jaleko.com.br/direto-ao-ponto-sindrome-de-wolff-parkinson-white/>. Acesso em 26 ago. 2021.

RODRIGUES, V. B. **Métricas de avaliação**: acurácia, precisão, recall ... quais as diferenças? 12 abr. 2019. Disponível em: <https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>. Acesso em 13 dez. 2021.

ROMANTRI, R.; SURANTHA, N. **Arrhythmia classification using 2D convolutional neural network**. International journal of advanced computer science and applications. v. 11, n. 4, 2020.

SBC. SOCIEDADE BRASILEIRA DE CARDIOLOGIA. **Estatística cardiovascular Brasil**: 2020 traz informações sobre as doenças cardiovasculares e seus impactos.



Disponível em: <https://www.portal.cardiol.br/post/sociedade-brasileira-de-cardiologia-apresenta-nova-plataforma-de-dados>. Acesso em 28 dez. 2020.

SOBRAC. SOCIEDADE BRASILEIRA DE ARRITMIAS CARDÍACAS.

Disponível em <https://sobrac.org/home/>. Acesso em 28 fev. 2021.

SRIVASTAVA, N. et al. **Dropout**: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, v. 15, p. 1929–1958, 2014.

SWAPNA, G.; SOMAN, K.P.; VINAYAKUMAR, R. **Automated detection of cardiac arrhythmia using deep learning techniques**. *Procedia Computer Science*. v. 132, p. 1192-1201, 2018.

SWEIS, R. N.; JIVAN, A. **Visão geral da doença coronariana**. Disponível em: <https://www.msmanuals.com/pt-br/profissional/doen%C3%A7as-cardiovasculares/doen%C3%A7a-coronariana/vis%C3%A3o-geral-da-doen%C3%A7a-coronariana>. Acesso em 01 nov. 2021.

THOMAS, L.; KHETRAPAL, A. **Doença cardíaca reumático**: Sintomas. Disponível em: [https://www.news-medical.net/health/Rheumatic-heart-disease-symptoms-\(Portuguese\).aspx](https://www.news-medical.net/health/Rheumatic-heart-disease-symptoms-(Portuguese).aspx). Acesso em 01 nov. 2021.

ULLAH, A. et al. **Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation**. *Remote sensing*. v. 12, p. 1685, 2020.

WEBSTER, J. G. **Medical Instrumentation**: Application and Design 4th. John Wiley and Sons. 2009.

WEN, L. et al. **A New Data-Driven Intelligent Fault Diagnosis by Using Convolutional Neural Network**. v. 10-13-Dec, 2017.

WERBOS, P.J. **Backpropagation through time**: What it does and how to do it. *Proceedings of the IEEE*, IEEE, v. 78, n. 10, p. 1550-1560, 1990.

WHO. WORLD HEALTH ORGANIZATION. Disponível em: [https://www.who.int/healthtopics/cardiovascular diseases/#tab=tab\\_1](https://www.who.int/healthtopics/cardiovascular diseases/#tab=tab_1). Acesso em 28 dez. 2020.

WU, M. **A study on arrhythmia via ECG signal classification using the convolutional neural network**. *Frontiers in computational neuroscience*, v. 14, 2021.



## APÊNDICE I

Relação das Arritmias do MIT-BIH Arrhythmia Database:

1. Ritmo sinusal normal (*Normal Sinus Rhythm* - NSR);
2. Batimento atrial prematuro (*Atrial Premature Beat* - APB);
3. Vibração atrial (*Atrial Flutter* - AFL);
4. Fibrilação atrial (*Atrial Fibrillation* - AFIB);
5. Taquiarritmias supraventricular (*Supraventricular Tachyarrhythmia* - SVTA);
6. Pré-excitação (*Pre-excitation* - WPW);
7. Contração ventricular prematura (*Premature Ventricular Contraction* - PVC);
8. Bigeminia ventricular (*Ventricular Bigeminy* - Bigeminy);
9. Trigemismo ventricular (*Ventricular Trigeminy* - Trigeminy);
10. Taquicardia ventricular (*Ventricular Tachycardia* - VT);
11. Ritmo idioventricular (*Idioventricular Rhythm* - IVR);
12. Vibração ventricular (*Ventricular Flutter* - VFL);
13. Fusão de batimento ventricular e normal (*Fusion of Ventricular and Normal Beat* - Fusion);
14. Batida do ramo esquerdo (*Left Bundle Branch Block Beat* - LBBBB);
15. Batida do ramo direito (*Right Bundle Branch Block Beat* - RBBBB);
16. Bloqueio cardíaco de segundo grau (*Second Degree Heart Block* - SDHB);e
17. Ritmo de marcapasso (*Pacemaker Rhythm* - PR).

Essa base de dados está disponível para *download* no site do Physionet (Physionet, 2021), no endereço eletrônico:

<https://www.physionet.org/content/mitdb/1.0.0/>. Esse site foi acessado pelo autor em 28 de fevereiro de 2021.

## APÊNDICE II

Registros extraídos dos sinais de ECG de cada arritmia utilizada neste trabalho:

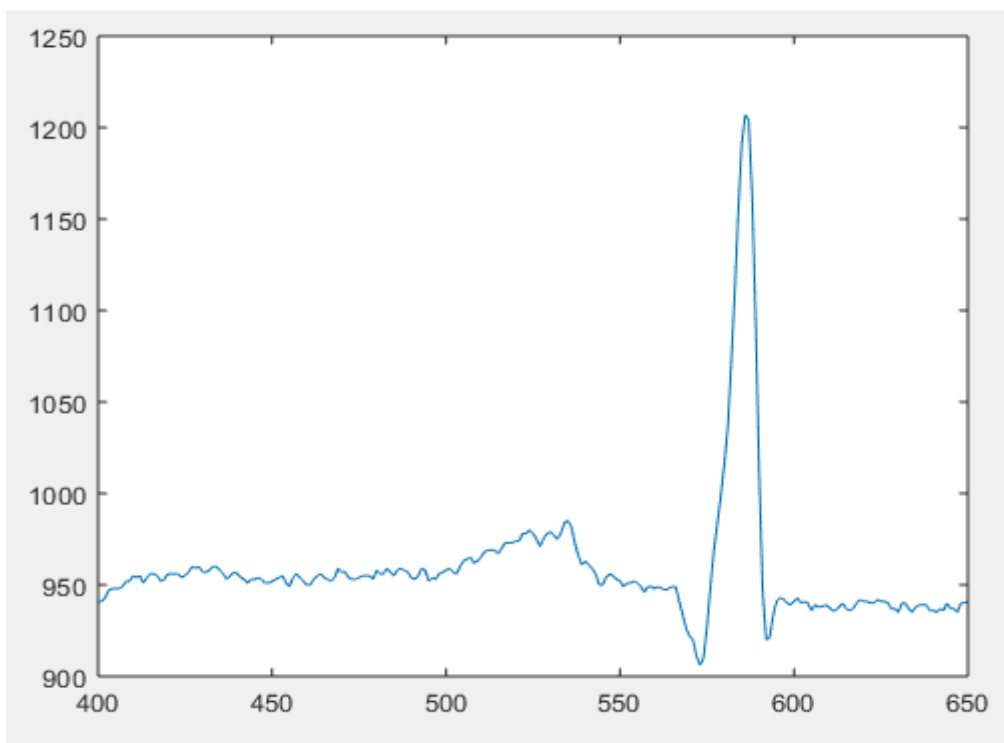


Figura 30- Registro de um trecho do sinal de ECG do sinal NSR

Fonte: Autor

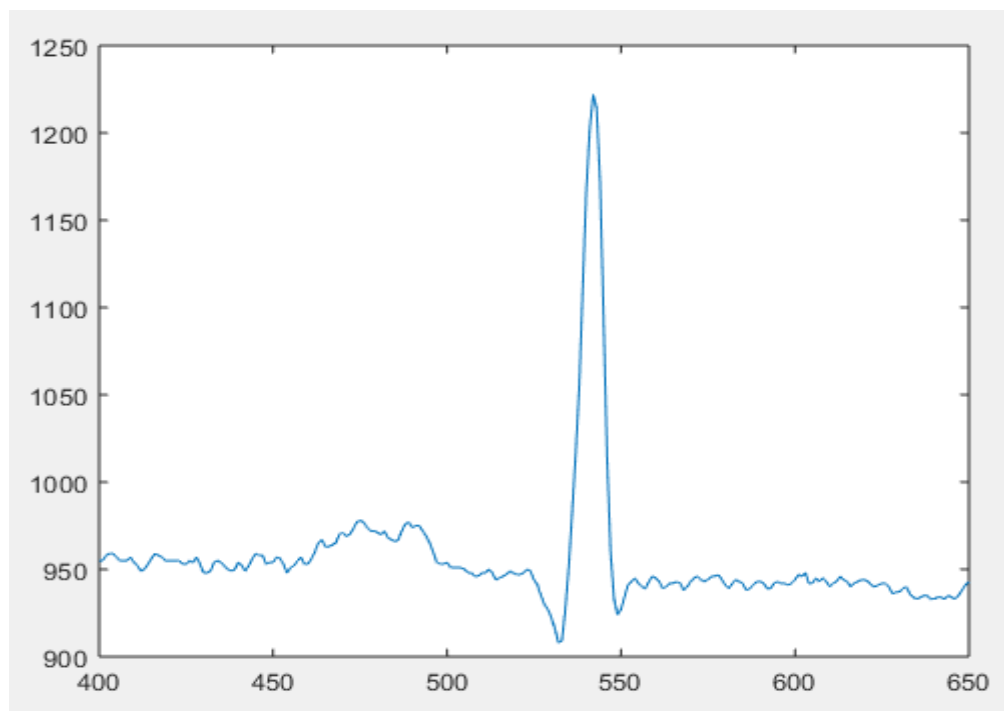


Figura 31- Registro de um trecho do sinal de ECG da arritmia APB

Fonte: Autor

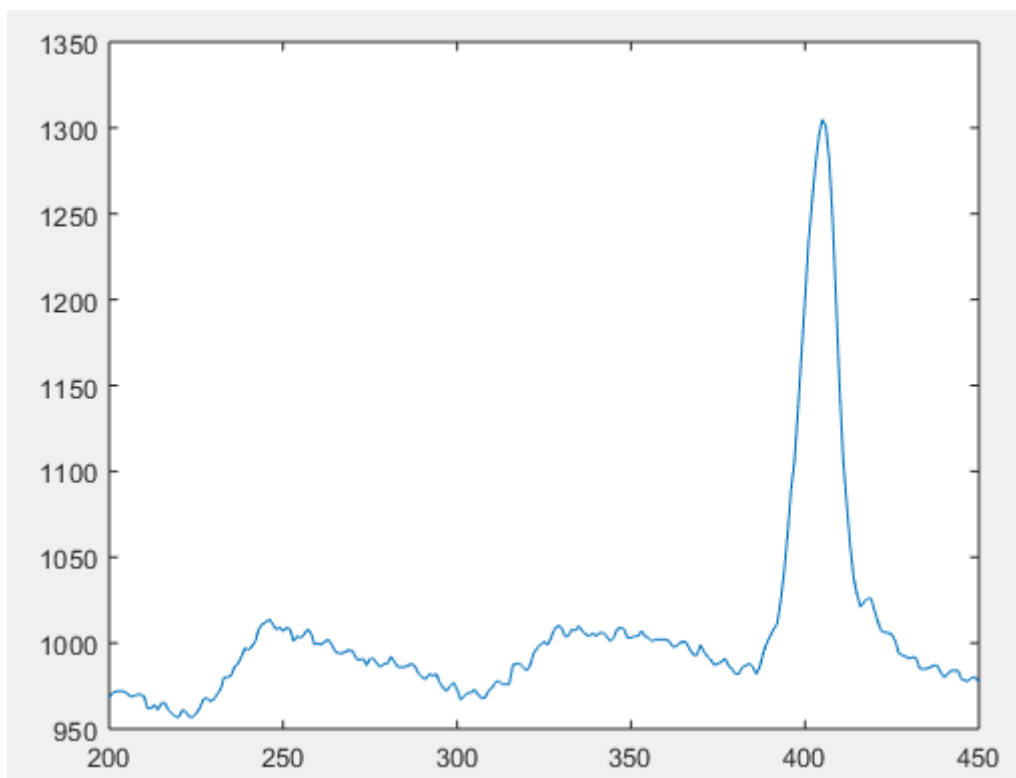


Figura 32- Registro de um trecho do sinal de ECG da arritmia AFL

Fonte: Autor

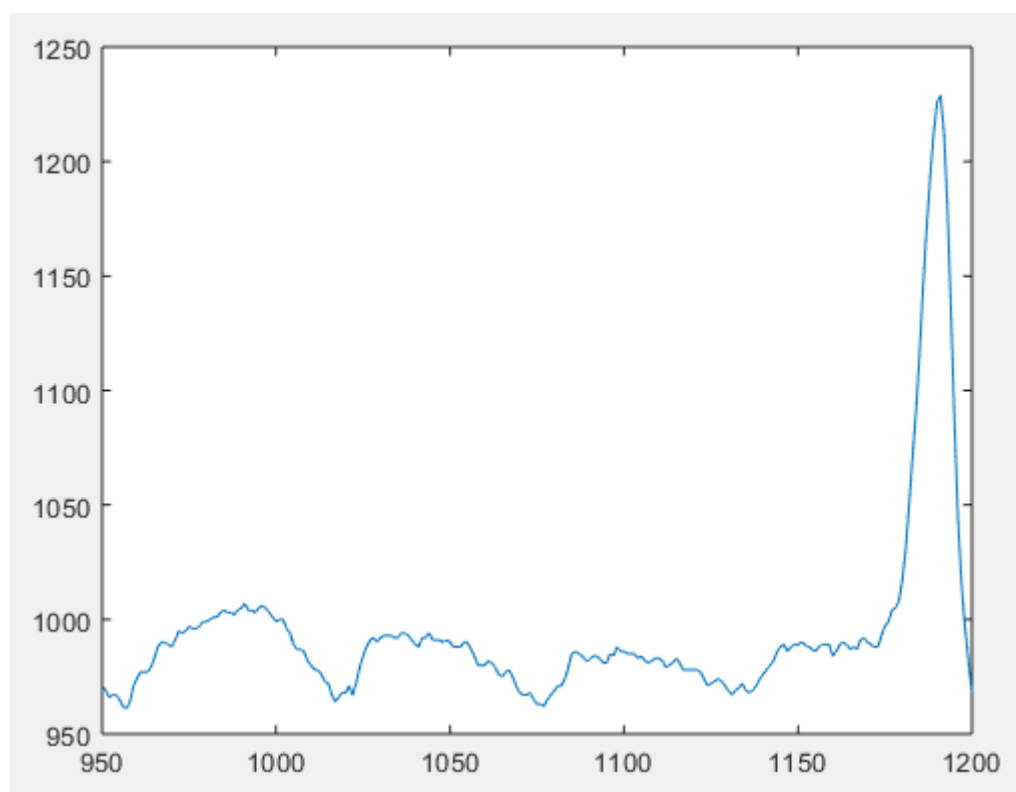


Figura 33- Registro de um trecho do sinal de ECG da arritmia AFIB

Fonte: Autor

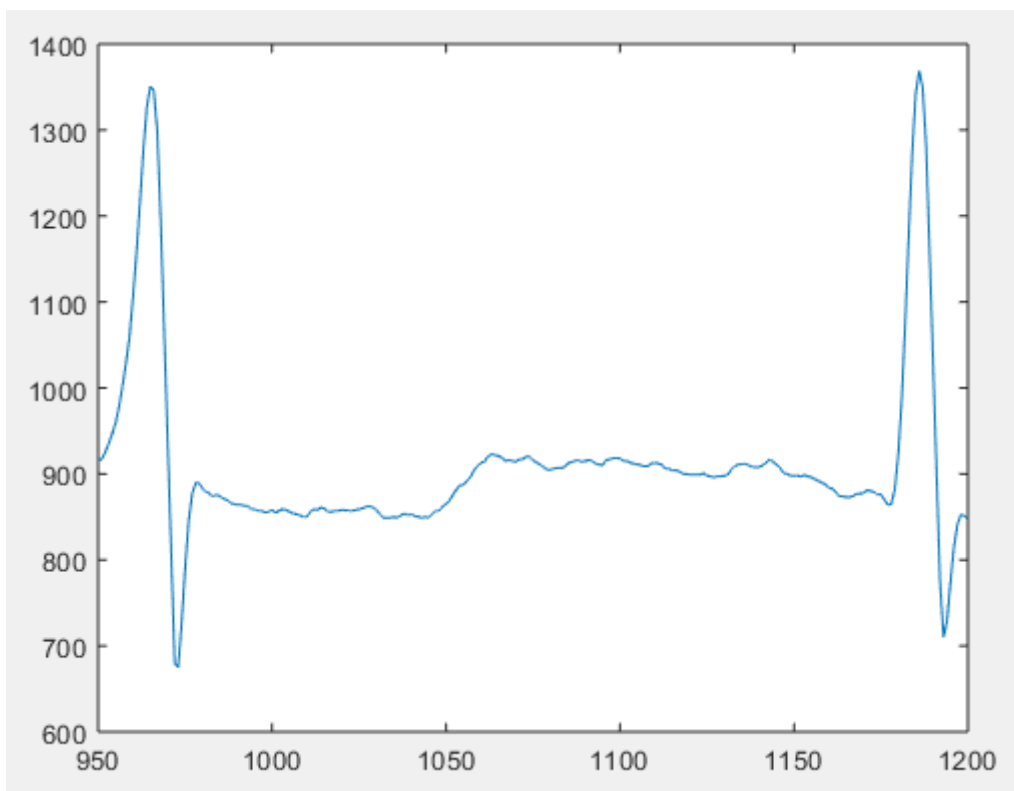


Figura 34- Registro de um trecho do sinal de ECG da arritmia SVTA

Fonte: Autor

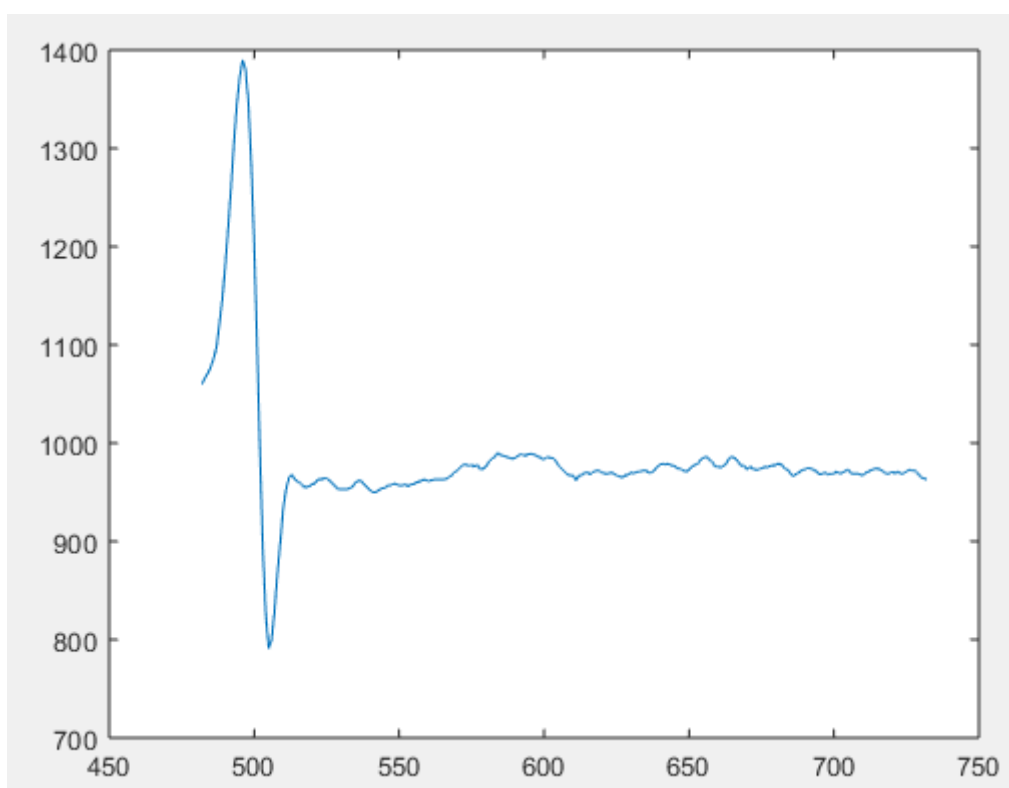


Figura 35- Registro de um trecho do sinal de ECG da arritmia WPW

Fonte: Autor

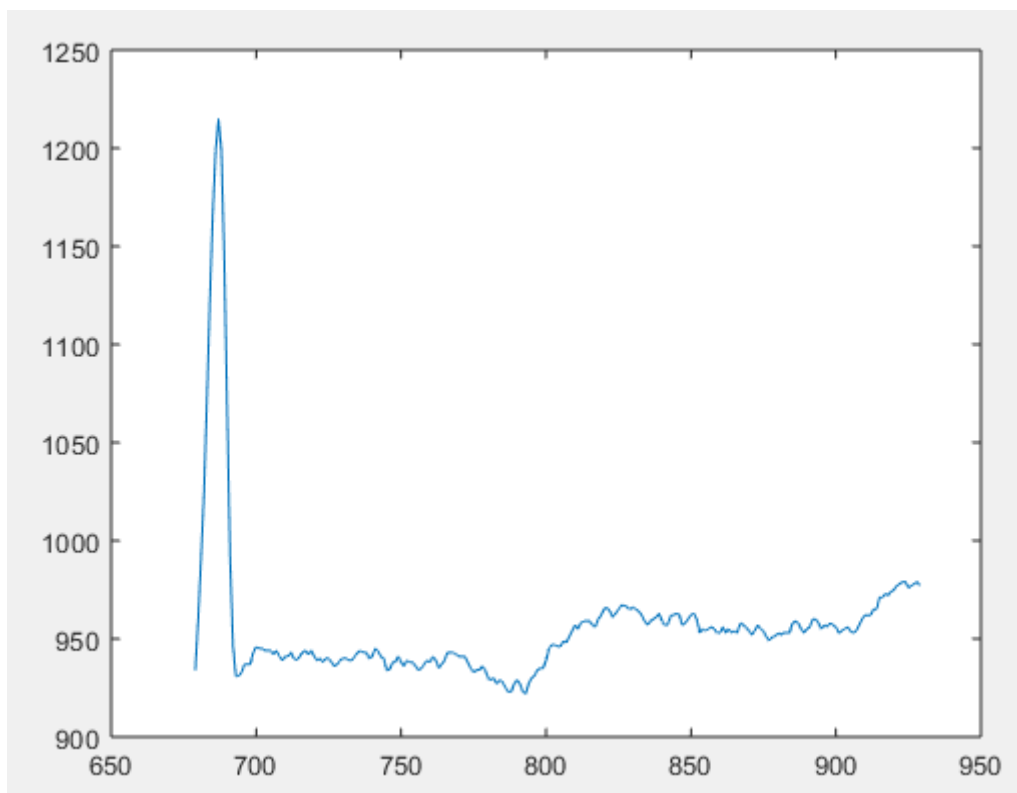


Figura 36- Registro de um trecho do sinal de ECG da arritmia PVC

Fonte: Autor

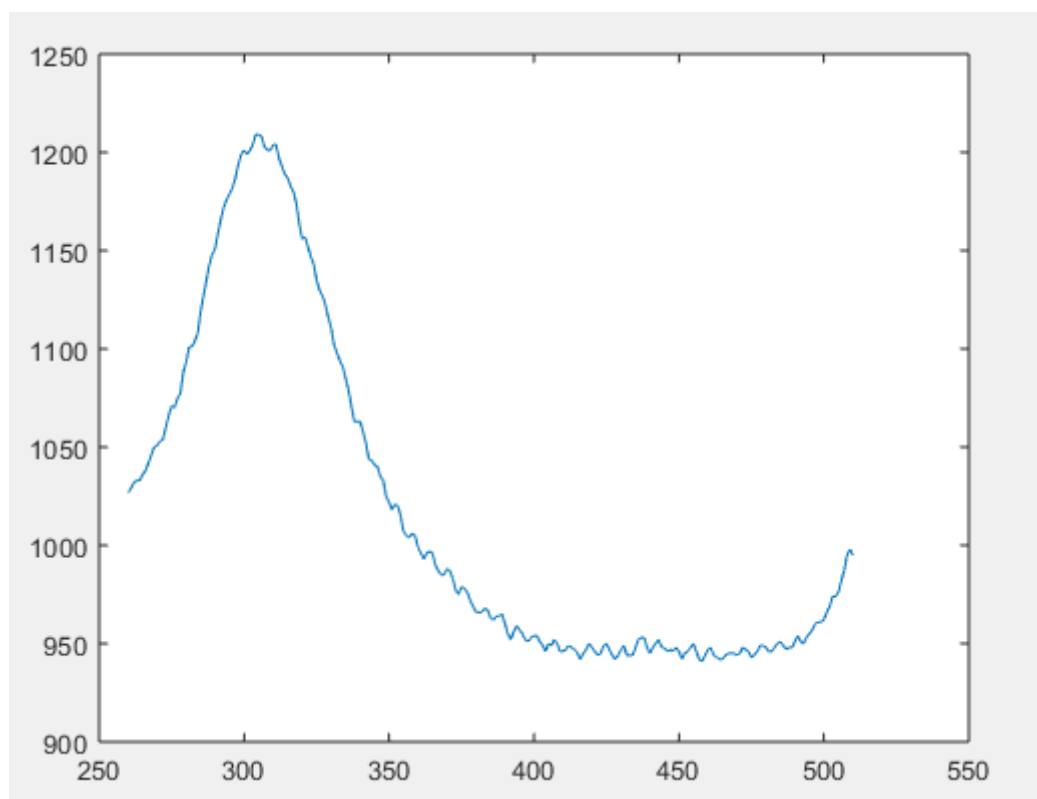


Figura 37- Registro de um trecho do sinal de ECG da arritmia Bigeminy

Fonte: Autor

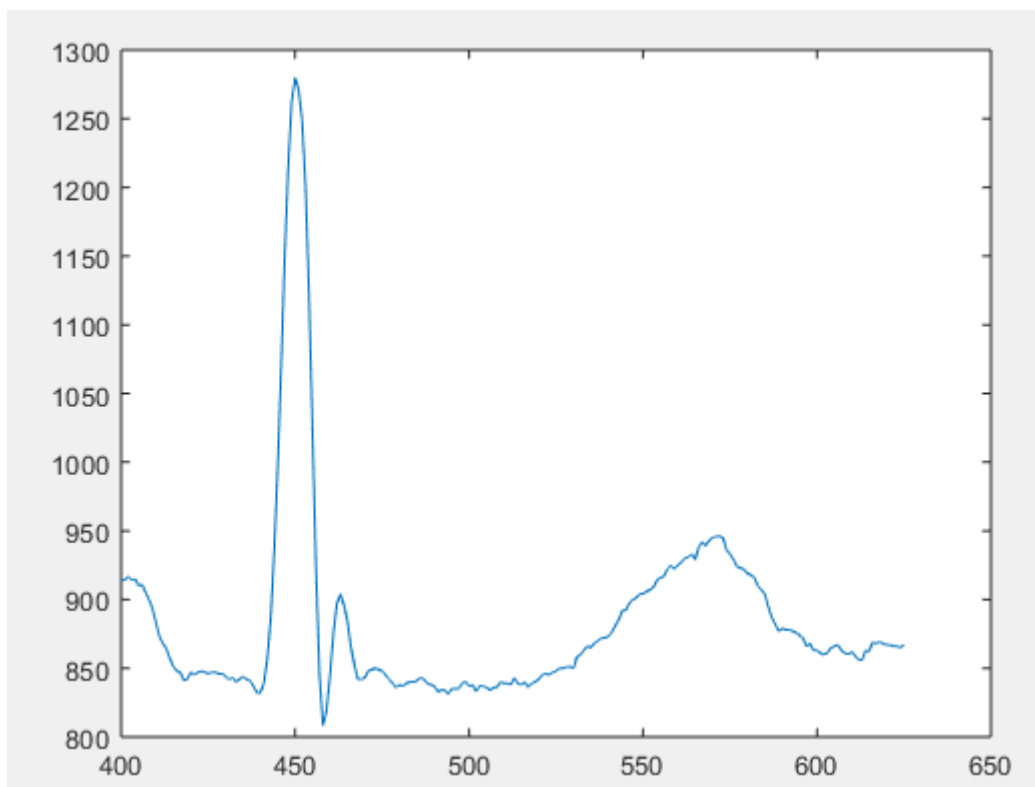


Figura 38- Registro de um trecho do sinal de ECG da arritmia Trigeminy

Fonte: Autor

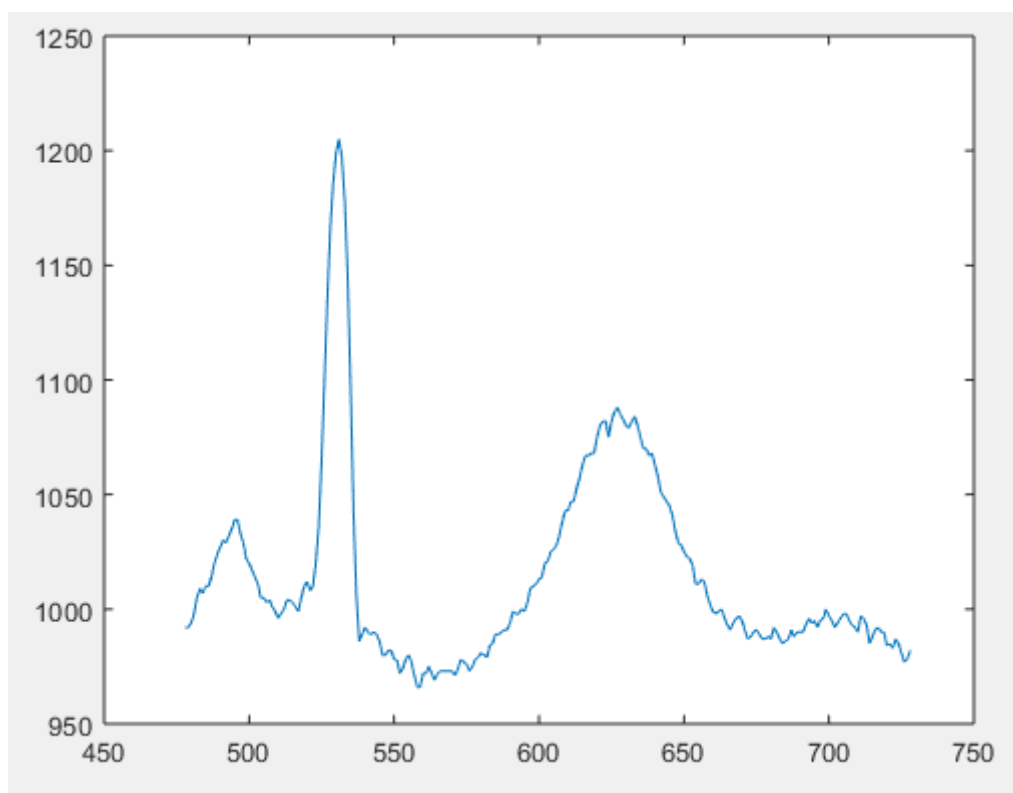


Figura 39- Registro de um trecho do sinal de ECG da arritmia VT

Fonte: Autor

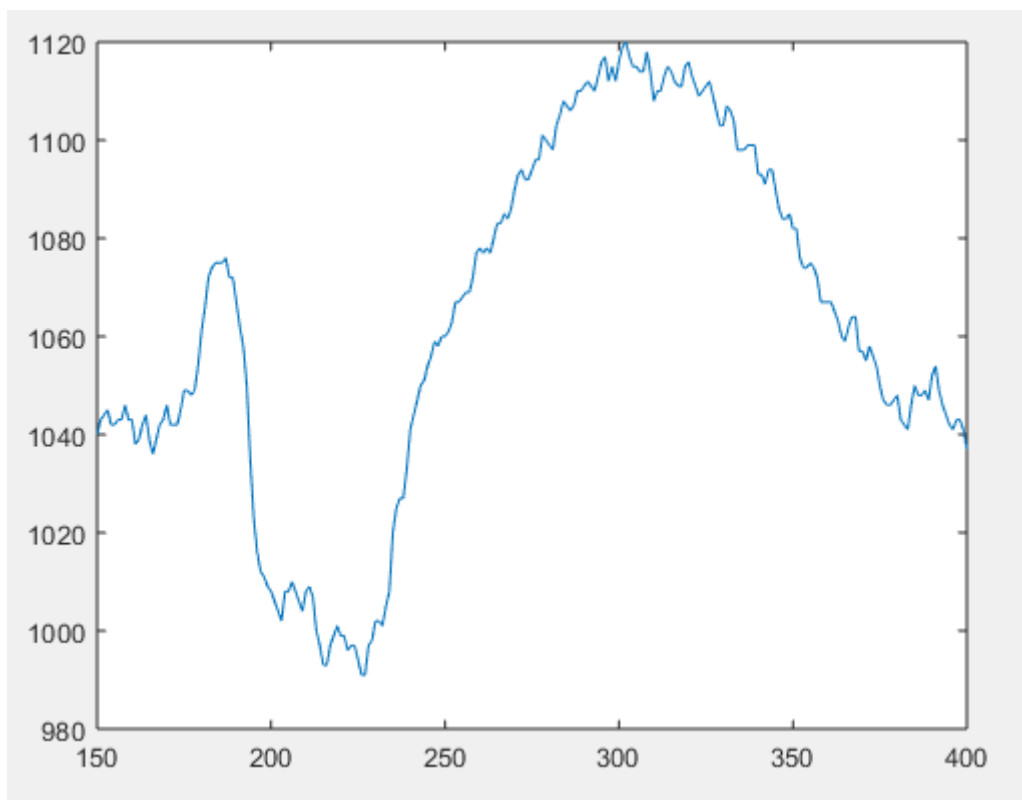


Figura 40- Registro de um trecho do sinal de ECG da arritmia IVR

Fonte: Autor

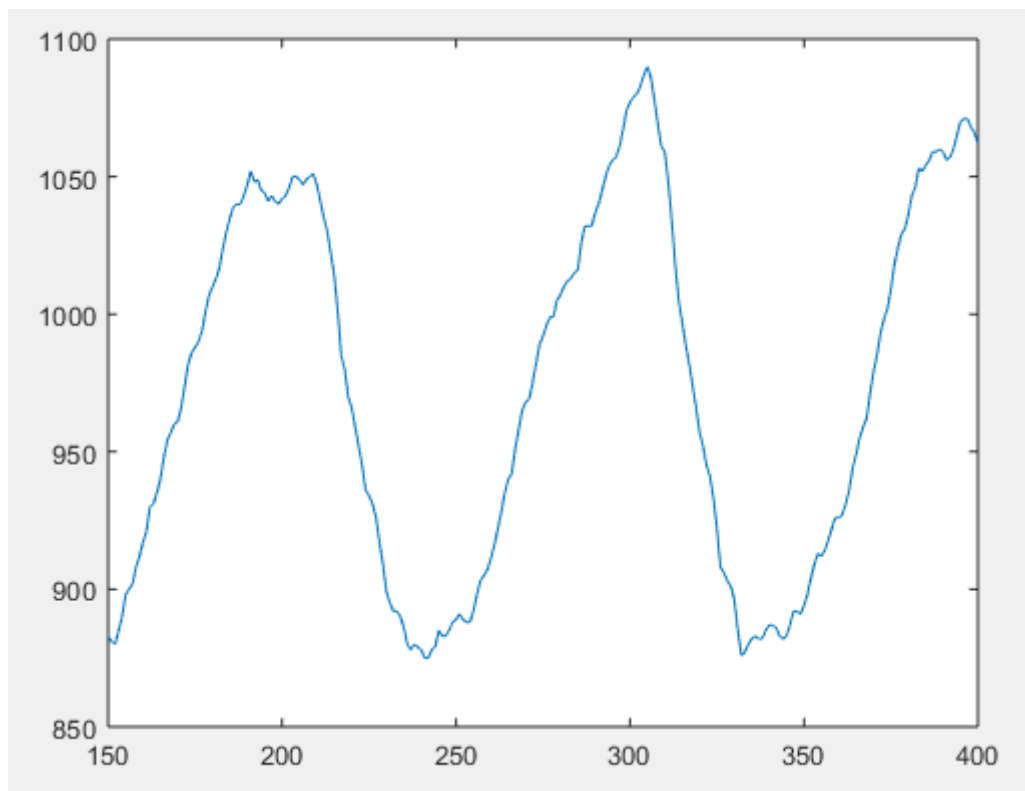


Figura 41- Registro de um trecho do sinal de ECG da arritmia VFL

Fonte: Autor

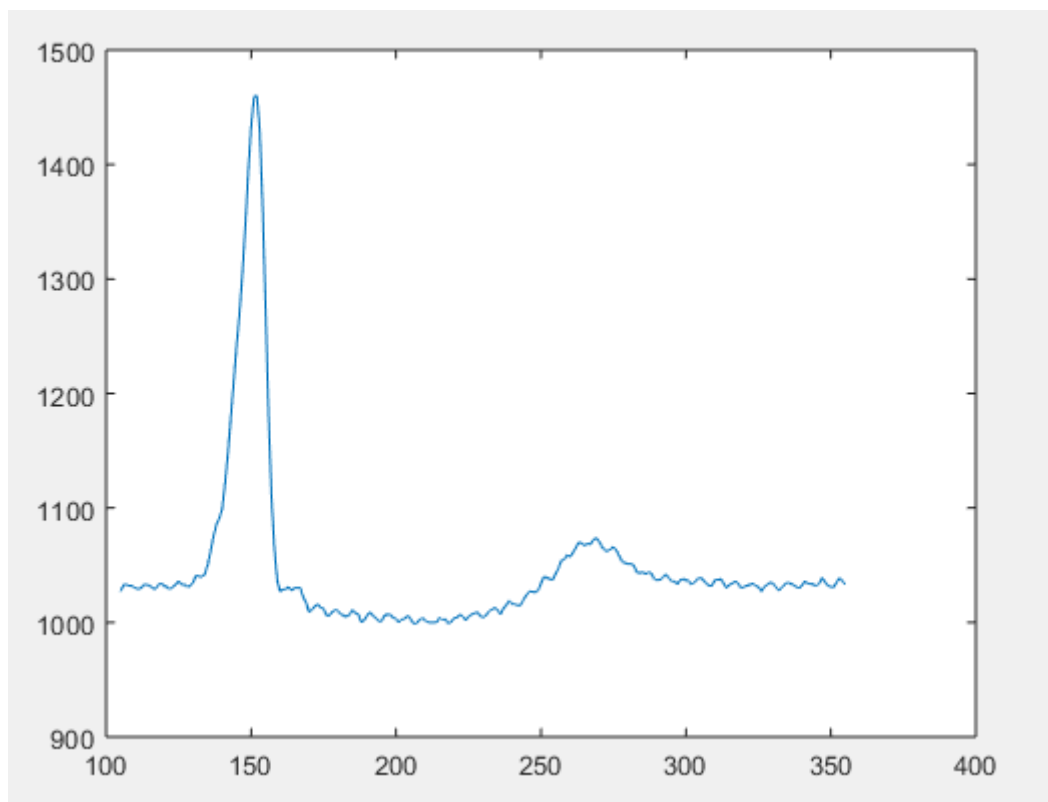


Figura 42- Registro de um trecho do sinal de ECG da arritmia Fusion

Fonte: Autor

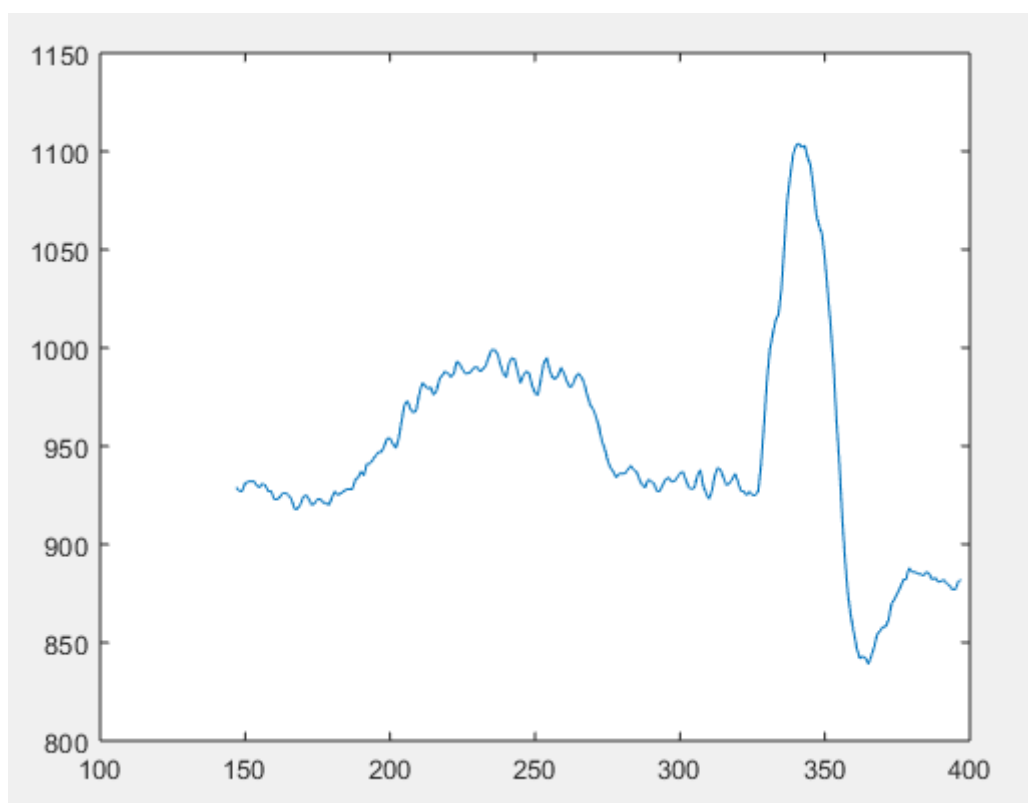


Figura 43- Registro de um trecho do sinal de ECG da arritmia LBBB

Fonte: Autor



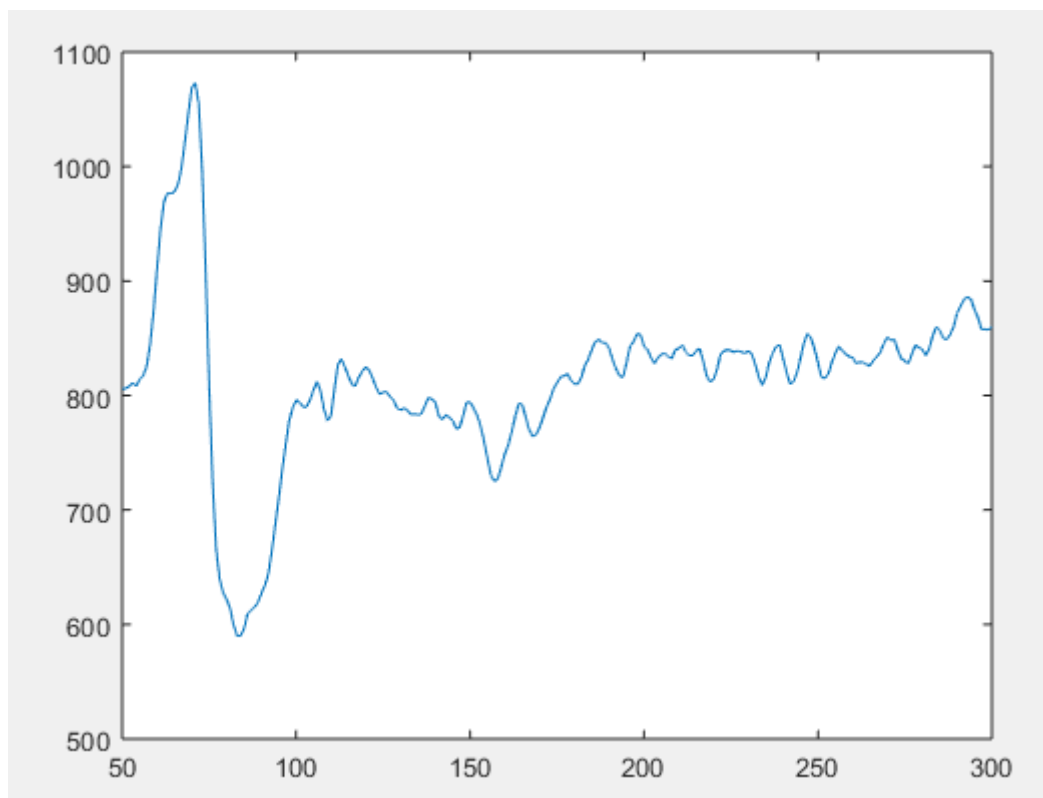


Figura 44- Registro de um trecho do sinal de ECG da arritmia RBBB

Fonte: Autor

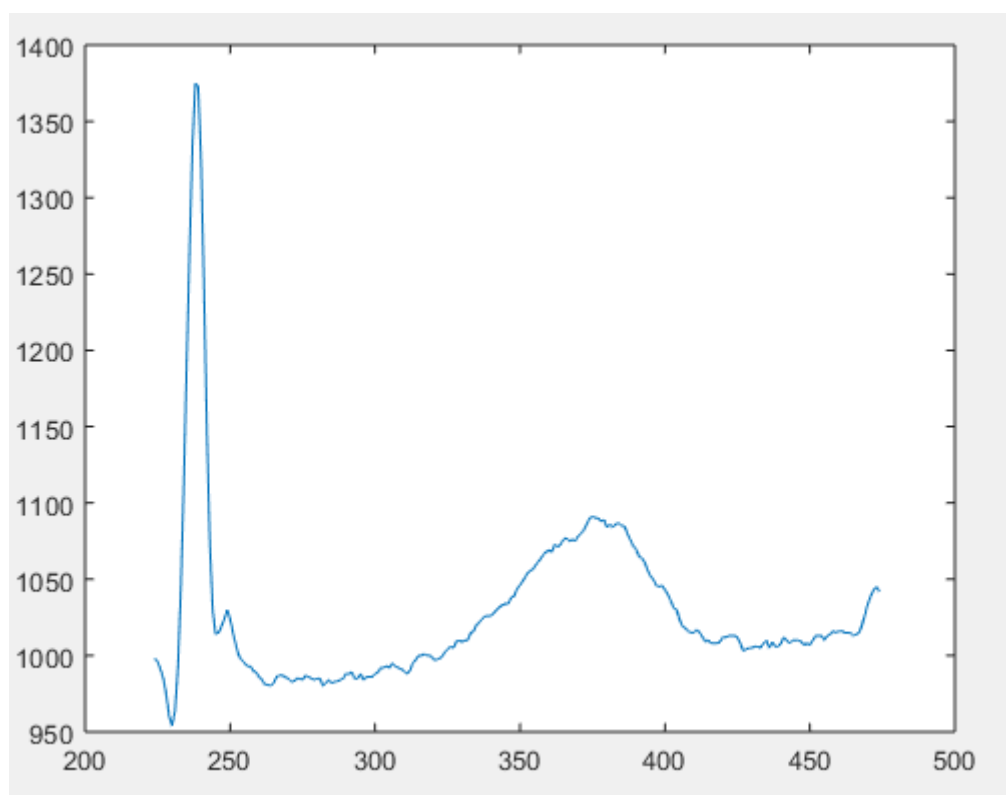


Figura 45- Registro de um trecho do sinal de ECG da arritmia SDHB

Fonte: Autor

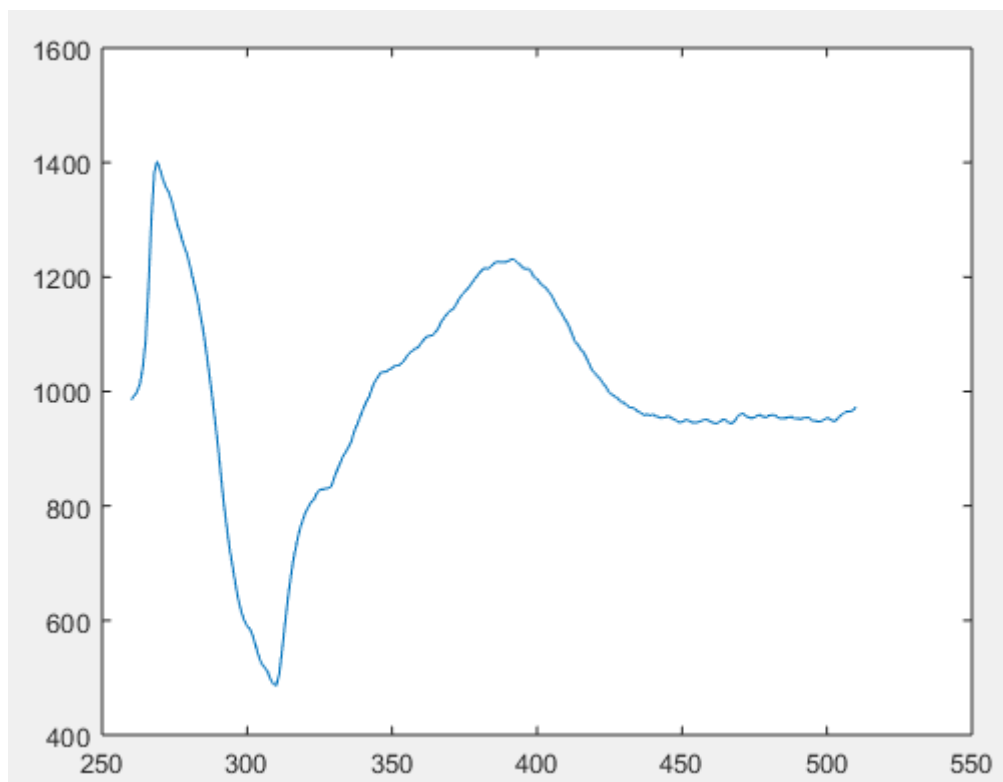


Figura 46- Registro de um trecho do sinal de ECG da arritmia PR

Fonte: Autor

### APÊNDICE III

TRABALHO ACEITO PARA PUBLICAÇÃO NA:

43RD ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE

ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 01/11/2021 A

05/11/2021 – ONLINE

# A New Approach to Classify Cardiac Arrhythmias Using 2D Convolutional Neural Networks \*

J. R. G. Santana, M. G. F. Costa, *Member, IEEE* and C. F. F. Costa Filho, *Member, IEEE*

**Abstract**— Cardiovascular diseases are the number one cause of death worldwide. Detecting cardiovascular diseases in its early stages could effectively reduce the mortality rate by providing timely treatment. In this study, we propose a new methodology to detect arrhythmias, using 2D Convolutional Neural Networks. The main characteristic of the proposed methodology is the use of 15 x 15 pixels gray-level images, containing the values of a heartbeat of the ECG signal. This work aims to detect 17 arrhythmias. To validate and test the proposed methodology, MIT-BIH database, the main benchmark database available in literature, was used. When compared to other results previously published, the obtained precision, 92.31%, is in the state-of-the-art.

**Clinical Relevance**— The presented work provides an automatic method to detect arrhythmias in ECG signals by a new methodology.

## INTRODUCTION

According to the World Health Organization [1], cardiovascular diseases (CVDs) are the number one cause of death worldwide, claiming about 17.9 million lives each year. CVDs are a group of diseases of the heart and blood vessels that include rheumatic heart disease, cerebrovascular disease, coronary heart disease and other conditions. Four out of 5 deaths from CVDs are due to heart attacks and strokes, and one third of these deaths occur prematurely in people under 70 years of age.

According to the Cardiovascular Statistics Brazil 2020 report [2], of the Brazilian Society of Cardiology, CVDs were the main cause of death in Brazil in the year 2017. Among CVDs, ischemic heart diseases were the cause number 1.

Detecting CVDs in its early stages could effectively reduce the mortality rate by providing timely treatment [3]. One of the common sources of CVDs is cardiac arrhythmia, which is characterized by the fact that heartbeats deviate from your normal pattern. A normal heartbeat varies with age, body size, activity and emotions. In cases where the heartbeat seems fast or slow,

the condition is known as a palpitation. An arrhythmia does not necessarily mean that the heart is beating too fast or too slowly, it indicates that the heart is following an irregular beating pattern.

Conventionally, for the diagnosis of arrhythmias, cardiologists visually inspect 12-lead ECG waveforms in a digital image format [4]. It is common that ECG signals lasting many hours and even days, like Holter monitoring, need to be analysed. This is a very time-consuming and exhausting procedure, which significantly limits the impartiality of the diagnosis. This limitation can be eliminated with the use of computational techniques for the automatic detection of arrhythmia and ECG classification.

In the literature, we have identified two groups of work developed for the automatic detection of arrhythmias. The first group comprises older works that used classic machine learning tools, while the second group comprises more recent works that use deep learning techniques, such as convolutional and recurrent neural networks.

Among others, in the first group of works we identified the use of the following classic machine learning techniques: wavelet transform [5], support vector machine [6], multilinear single value decomposition [7], hidden Markov models [8], etc.

In the second group of works, we identified two trends. The first one comprises the works that use only Convolutional Neural Networks (CNNs), the most frequent being one-dimensional networks [9][10]. Other studies use two-dimensional CNNs [11][12], converting the one-dimensional ECG tracings into a 2D array with dimensions of 128x128 or 64x64, for example. The second trend

includes works that use hybrid architectures, a serial composition of CNN and Recurrent Neural Networks (RNNs) [13][14][15]. In this hybrid architecture, CNNs are used to feature extraction. The features extracted, after a dimensionality reduction step, through maxpooling layers, feed the input of a RNN, responsible for the ECG classification, according to normal or altered states (arrhythmias).

In this work, we also used 2D CNNs for the classification of ECG, however, 2D ECG data was composed in a different way from that had been used in [11][12]. To obtain the input signal of the 2D CNN, 225 samples of a heartbeat of one-dimensional ECG signal are converted into a 15 x 15 pixels gray-level image. From a computational point of view, working with smaller images puts less stress on low memory devices when, for example, the implementation is embedded in mobile devices.

The advantage of working with a 2D signal instead of a 1D signal, is that the convolution operation with 2D kernel explores new neighborhood relations (neighborhoods above and below) in addition to the neighborhood relation already explored with 1D convolution operation (neighborhoods on the left and on the right).

To validate and test the proposed methodology we use the most used database to benchmark algorithms developed for detecting arrhythmias, known as the MIT-BIH database [16]. Most published academic papers, that use this database, classified 5 or 8 15 cardiac disfunctions. Few studies have been carried out aiming to classify all 17 classes present in MIT-BIH database (normal ECG, pacemaker signal, and 15 cardiac disfunctions) [10][6].

We also intend to present a performance comparison between 2D CNN, 1D CNN and RNNs (Long-Short-Time-Memory -LSTM and Gated-Recurrent-Units - GRU), in the classification of the 17 classes present in the MIT-BIH database, a gap little explored in the literature.

## METHODOLOGY

### Materials

In this work, the ECG arrhythmia database of MIT-BIH was used. The ECG signals, obtained

from 47 patients, were sampled at a rate of 360 Hz. The base consists of 48 annotated records, with an approximate duration of 30 minutes of heartbeat [9]. 17 classes are present in the database: Normal Sinus Rhythm (NSR), pacemaker rhythm (PR) and the following 15 cardiac disfunctions: Atrial Premature Beat (APB), Atrial Flutter (AFL), Atrial Fibrillation (AFIB), Supraventricular Tachyarrhythmia (SVTA), Pre-excitation (WPW), Premature Ventricular Contraction (PVC), Ventricular Bigeminy (B), Ventricular Trigeminy (T), Ventricular Tachycardia (VT), Idioventricular Rhythm (IVR), Ventricular Flutter (VFL), Fusion of Ventricular and Normal Beat (F), Left Bundle Branch Block Beat (LBBBB), Right Bundle Branch Block Beat (RBBBB), Second-Degree Heart Block (SDHB).

For each heartbeat, a one-dimensional record with 225 samples around an R peak, not necessarily with it centralized, was extracted. 150 records of each class were randomly extracted, totaling 2550 records.

### Pre-Processing

From each one-dimensional record was generated one intensity image. Figure 1 illustrates the steps employed in this generation.



Figure 1. Steps to obtain a 15 x 15 pixels gray-level image of a record with 225 samples extracted around an R-wave peak: record extraction, record normalization, 1D-2D conversion, gray-level image generation.

The normalization step, shown in equation (1), aims obtaining sample values in the range [0 255].

$$s_n = \frac{s - s_{min}}{s_{max} - s_{min}} \cdot 255 \quad (1)$$

where:

- $s_n$ : normalized sample value
- $s$ : non - normalized sample value
- $s_{min}$ : minimum sample value
- $s_{max}$ : maximum sample value

In the gray-level image, the pixels intensities correspond to the values of the normalized heartbeat ECG signal. Figure 2 shows corresponding examples of the process steps presented in Figure 1.

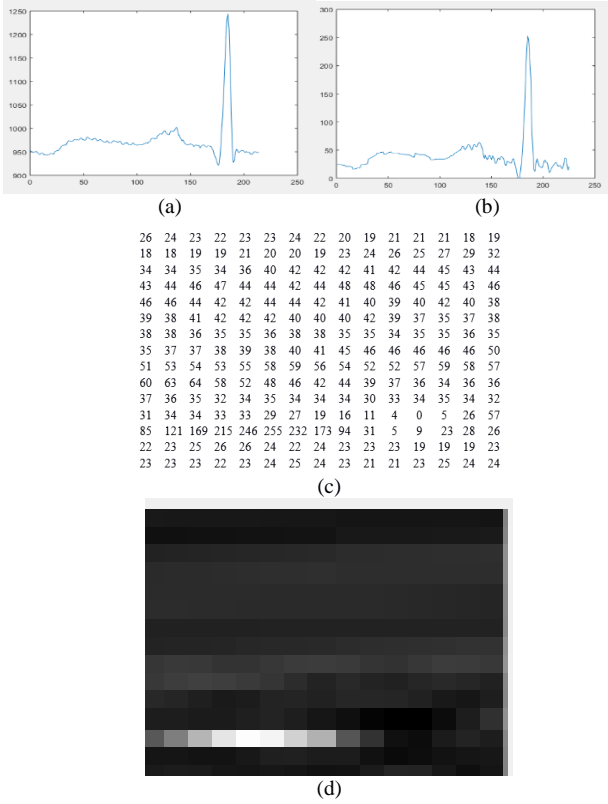


Figure 2. Corresponding examples of process steps shown in Figure 1: (a) Extracted record; (b) Normalized record; (c) 1D-2D conversion and (d) gray-level image obtained.

### Deep Neural Networks and Training Parameters

Figure 3 shows the architectures of neural networks implemented. The 1D and 2D CNN architectures are shown in Figure 3(a), while the LSTM and GRU architectures are shown in Figure 3(b).



Figure 3: (a) 1D or 2D CNN architecture and (b) LSTM or GRU architecture

The differences between the 1D CNN and 2D CNN architectures are the kernel sizes and the input layer. In 1D CNN, the kernel size is  $[1 \ n]$ , while in 2D CNN the kernel size is  $[n \ n]$ . Both convolutional kernels have  $n = 2$ . The size of input layer is  $[1 \ 225]$  in 1D CNN, corresponding to a vector with 255 values of a heartbeat. The size of input layer is  $15 \times 15$  in 2D CNN, and corresponds to a gray-level image, as shown in Figure 2(d). In the CNN architecture there are seven features extraction blocks, consisting of the following layers: Conv→Batch→Relu. The number of filters of the convolutional layers is equal to 96. A stride of 1 is used. In the sequence, there are the following layers: fully connected layer (200 neurons)→ dropout layer → fully connected layer (17 neurons) →softmax layer → classification layer.

The recurrent neural networks, LSTM or GRU, are formed by the following layers: a sequence input layer (input size = 225) → a LSTM or GRU layer (4096 neurons) → a fully connected layer (512 neurons), a dropout layer → a fully connected layer (17 neurons) → a softmax layer and a classification layer.

The methodology used for training and testing was the 5-fold-cross-validation. In each fold 80% of data is used for training and 20%, for validation. In the training, after an exhaustive search, the following parameters were chosen *bath size* = 512, *L<sub>2</sub> regularization factor* = 0.1, *dropout factor* = 0.5, *maximum number of epochs* = 2000. For learning rate, a piecewise schedule was employed, with *initial learning rate* = 0.01, a *drop period* = 200 epochs and a *drop factor* = 0.75. Figure 4 shows the precision during a training section of a 2D CNN. At the beginning of the training, due to the high values of the learning rate, several oscillations occurred. However, as the learning rate decreased with the increase in the number of epochs, these oscillations also decreased. Three optimization methods were evaluated: Stochastic Gradient Descent with Momentum - SGDM, Adaptive Moment Estimation Optimizer - ADAM and Root Mean Square Propagation - RMSProp. The best performances were obtained with the SGDM

optimizer. We notice that the best performance of the training, evaluated in the validation set, does not necessarily occur at the end of the training. That is why we use the MATLAB checkpoint feature, recording the network with the best performance during the training.

The experiments were performed using MATLAB<sup>®</sup> version 2018b, a 3.2 GHz Intel i7-8700 processor computer with 16 GB RAM and 8 GB GeForce GTX 1070 GPU.

#### Evaluation Metrics

The following metrics were used for evaluation: accuracy (Acc), sensitivity (Sens) or recall, specificity (Spec) and precision (P) and F1-score. For each fold, these metrics are calculated according to equations (2), (3), (4), (5) and (6), respectively [10]. Accuracy can be a misleading metric for imbalanced data sets. To make up for this deficiency, we also calculated specificity and sensitivity. Precision talks about how precise/accurate the model is out of those predicted positive, how many of them are actual

positive. F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall and there is an uneven class distribution (large number of Actual Negatives).

$$Acc = \frac{1}{17} \sum_{c=1}^{17} \frac{T_p^c + T_n^c}{T_p^c + T_n^c + F_p^c + F_n^c} \quad (2)$$

$$Sens = \frac{1}{17} \sum_{c=1}^{17} \frac{T_p^c}{T_p^c + F_n^c} \quad (3)$$

$$Spec = \frac{1}{17} \sum_{c=1}^{17} \frac{T_n^c}{T_n^c + F_p^c} \quad (4)$$

$$P = \frac{1}{17} \sum_{c=1}^{17} \frac{T_p^c}{T_p^c + F_p^c} \quad (5)$$

$$F1\text{-score} = 2 * \frac{(P * Sens)}{(P + Sens)} \quad (6)$$

where  $T_p^c$  denotes the true positives: all  $c$  instances that are classified as  $c$ ;  $T_n^c$  denotes true negatives: all non- $c$  instances that are not classified as  $c$ ;  $F_p^c$  denotes the false positives: all non- $c$  instances that are classified as  $c$ ;  $F_n^c$  denotes false negatives: all  $c$  instances that are not classified as  $c$ .

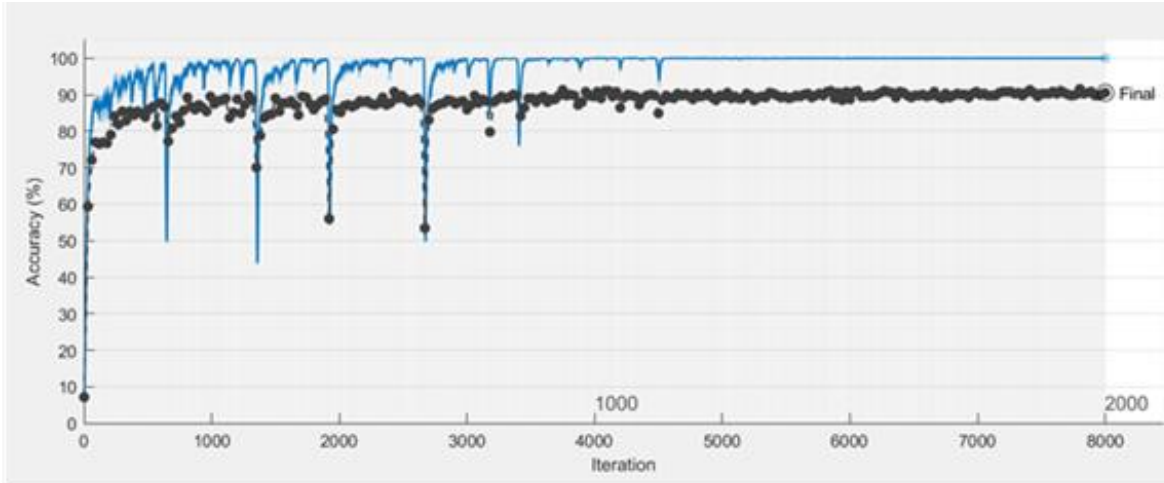


Figure 4. Precision during the training section of a 2D CNN. blue curve: training precision. black-curve – validation precision.

#### Results and discussions

Table I shows the performance of all deep neural networks evaluated in this work: 1D CNN, 2D CNN, LSTM and GRU, with 5-fold-cross validation. As shown, the best performance was obtained with the 2D CNN network, with a precision of 92.31%. Figure 5 shows a confusion table obtained with a 2D CNN.

Table II shows a comparison of the performance of the 2D CNN with other results previously published in the literature for classification of 17 classes of MIT-BIH database. As shown, the results obtained in this paper are in the state-of-the-art.

As mentioned, some previous arrhythmias detection works use gray-level images [11] [14] [12]. In these works, the authors used, as input



signal, a gray-level image of the ECG tracing (like an ECG signal photo) with different sizes: 256 x 256, 192 x 128, and 64 x 64 pixels, respectively. In this work, we also use a gray-level image. However, the intensity values of the pixels correspond to the values of the normalized heartbeat ECG signal. Therefore, all pixels of the gray-level image carry relevant information about ECG signal.

Also considering that, this image is significantly smaller than those used in aforementioned papers, the proposed method is more suitable to be implemented on platforms with reduced memory capacity, such as mobile devices.

Although the most recent studies for arrhythmias detection use hybrid architectures, CNN-LSTM, in this work we managed to obtain cutting-edge performance, using only a CNN architecture.

We credit this better performance of the 2D-CNN to the fact that new neighborhood relationships are explored through a proposed gray-level image that all pixels carry information about a heartbeat of the ECG signal.

In future works, we intend to 1) explore data augmentation, creating new images through small gray level changes in the pixels of the gray-level image; 2) evaluate other arrhythmia classifications in 2, 5 and 8 classes; 3) implement the proposed method in mobile devices.

TABLE I RESULTS OF THE 4 ARCHITECTURES EVALUATED IN THIS STUDY (MEAN±SD)

Architecture	Accuracy	Sensitivity	Specificity	Precision	F1-Score
1D CNN	94.29±2.14	47.64±18.05	97.00±1.92	52.27±27.28	49.85±22.37
2D CNN	99.13±0.65	93.07±6.10	99.52±0.42	92.31±6.87	92.68±5.64
LSTM	96.21±1.65	64.65±15.15	98.09±1.48	68.13±24.72	66.34±18.53
GRU	95.71±1.82	62.99±16.12	97.70±1.55	63.51±24.19	63.24±18.35

		Confusion Matrix																			
Output Class	AFIB	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	AFIB	5.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	AFL	0	22	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
	AFL	0.0%	4.3%	0.2%	0.2%	0.0%	0.0%	0.2%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	APB	2	0	29	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
	APB	0.4%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	14.7%
	B	0	6	0	28	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	B	0.0%	1.2%	0.0%	5.5%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	F	0	1	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0.0%	0.2%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	IVR	0	0	0	1	0	27	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	IVR	0.0%	0.0%	0.0%	0.2%	0.0%	5.3%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	LBBBB	1	0	0	0	0	0	24	0	0	0	0	1	2	0	0	0	0	0	1	0
	LBBBB	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	4.7%	0.0%	0.0%	0.0%	0.0%	0.2%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%
	NSR	0	1	0	0	0	0	0	29	0	0	0	0	1	0	0	0	0	0	0	0
	NSR	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	PR	0	0	0	0	0	1	1	0	29	0	0	0	0	0	0	0	0	0	0	0
	PR	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	PVC	0	0	0	0	1	0	0	0	0	0	30	0	0	0	1	0	0	0	0	0
	PVC	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	5.9%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	RBBBB	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	1	0
	RBBBB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%
	SDHB	0	0	0	0	0	0	1	0	0	0	0	26	0	0	0	0	0	0	0	0
	SDHB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	5.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	SVTA	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0
	SVTA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0
	T	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	VFL	0	0	0	0	0	0	1	0	0	0	0	0	0	0	30	0	0	0	0	0
	VFL	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.9%	0.0%	0.0%	0.0%	0.0%	0.0%
	VT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0
	VT	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.9%	0.0%	0.0%	0.0%	0.0%
	WPW	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	27	0
	WPW	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	5.3%	0.0%
		86.7%	73.3%	86.7%	83.3%	86.7%	80.0%	80.0%	86.7%	86.7%	100%	86.7%	86.7%	80.0%	86.7%	100%	100%	80.0%	82.4%	86.7%	7.6%
		13.3%	26.7%	13.3%	16.7%	13.3%	10.0%	20.0%	13.3%	13.3%	0.0%	13.3%	13.3%	10.0%	13.3%	0.0%	0.0%	10.0%	17.6%	13.3%	0.0%
		AFIB	AFL	APB	B	F	IVR	LBBBB	NSR	PR	PVC	RBBBB	SDHB	SVTA	T	VFL	VT	WPW			

Figure 5. Confusion Matrix obtained with 2D CNN showing precision values.



TABLE II COMPARISON OF THE 2D CNN WITH OTHER RESULTS PREVIOUSLY PUBLISHED IN LITERATURE

Work	Year	Precision
Yildirim [10]	2018	91.33
Plawiak [6]	2017	91.40
Proposed Method (2D CNN)	2021	92.31

## Conclusion

In this paper we proposed a new methodology for arrhythmias detection in ECG signals, using 2D CNNs. The main characteristic of the proposed method was the use of 15x15 pixels gray-level images of the ECG signal, with the pixel's intensities corresponding to the ECG values. The obtained results, with a precision of 92.31% is in the state-of-the-art. In future works we intend to explore data augmentation to improve the results. We also intend to implement the proposed method in mobile devices.

## References

- [1] WHO, "WHO," 2021. [Online]. Available: [https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1).
- [2] G. M. M. Oliveira *et al.*, "Cardiovascular Statistics – Brazil 2020," *Arq. Bras. Cardiol.*, vol. 115, no. 3, pp. 308–439, 2020.
- [3] A. Mustaqeem, S. M. Anwar, and M. Majid, "A modular cluster based collaborative recommender system for cardiac patients," *Artif. Intell. Med.*, vol. 102, no. November 2019, p. 101761, 2020.
- [4] G. A. Tadesse *et al.*, "Cardiovascular disease diagnosis using cross-domain transfer learning," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 4262–4265, 2019.
- [5] Y. Jung and W. J. Tompkins, "Detecting and Classifying Life-threatening ECG Ventricular Arrhythmias using Wavelet Decomposition," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 3, pp. 2390–2393, 2003.
- [6] P. Plawiak, "Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals," *Swarm Evol. Comput.*, vol. 39, no. April, pp. 192–208, 2018.
- [7] M. Bousse, G. Goovaerts, N. Vervliet, O. Debals, S. Van Huffel, and L. De Lathauwer, "Irregular heartbeat classification using Kronecker Product Equations," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 438–441, 2017.
- [8] D. Novák, D. Cuesta-Frau, T. Al Ani, M. Aboy, P. Mico, and L. Lhotská, "Speech recognition methods applied to biomedical signals processing," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 26 I, pp. 118–121, 2004.
- [9] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
- [10] A. Ullah, S. Muhammad Anwar, M. Bilal, and R. Majid Mehmood, "Classification of arrhythmia by using deep learning with 2-d ecg spectral image representation," *arXiv*, vol. 2, pp. 1–14, 2020.