

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE GRADUAÇÃO
CURSO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO**

BEN-HUR DANIEL MORAIS FARINHA

**SISTEMA CLASSIFICADOR DE SONS CARDÍACOS BASEADO EM REDES
NEURAIS ARTIFICIAIS**

**São Leopoldo
2022**

BEN-HUR DANIEL MORAIS FARINHA

**SISTEMA CLASSIFICADOR DE SONS CARDÍACOS BASEADO EM REDES
NEURAIAS ARTIFICIAIS**

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Bacharel em Engenharia de Controle e Automação, pelo Curso de Engenharia de Controle e Automação da Universidade do Vale do Rio dos Sinos (UNISINOS).

Orientador: Prof. Dr. Rodrigo Marques de Figueiredo

São Leopoldo

2022

Dedico este trabalho a minha mãe, que mesmo diante das dificuldades de sua vida, nunca mediu esforços para me tornar um homem de caráter íntegro e honesto. À minha noiva, por sempre me incentivar a seguir em frente e ser meu porto seguro.

AGRADECIMENTOS

A felicidade não se trata de alcançar objetivos e realizações, mas sim dos aprendizados e momentos que vivemos até chegar em cada um destes. E aqui estou eu, prestes a alcançar um dos maiores objetivos da minha vida, a realização do sonho de me tornar engenheiro. Este momento me faz voltar a infância, onde surgiu o meu fascínio pela tecnologia. Lembro-me de desmontar meus brinquedos eletrônicos somente para observar como eram por dentro e como coisas tão pequenas poderiam fazer algo que para mim era mágica, algo sem explicação e simplesmente fantástico. Ali, o destino me mostrou que um dia eu seria um engenheiro e compreenderia os truques daquela mágica. Quis o mesmo destino, que a minha mãe fosse a Virginia, a mulher mais forte que conheci, que, mesmo diante de todas as limitações que a vida lhe impôs, sempre esteve ao meu lado, me incentivando a seguir meus estudos e principalmente sempre me ensinando a ser um homem de caráter íntegro e honesto. À minha mãe, o meu mais profundo agradecimento, sem os seus ensinamentos, eu jamais estaria aqui, escrevendo este texto.

Este trabalho desafiou-me academicamente e psicologicamente, onde tive de tirar forças para sempre seguir em frente diante dos desafios aos quais me deparei. Esta força veio graças ao maior presente que a vida me deu, a mulher da minha vida e meu porto seguro, minha noiva Luana, que esteve sempre ao meu lado, me incentivando e me apoiando nos momentos em que mais precisei. Minha eterna gratidão por ser minha companheira de vida, por me fazer feliz todos os dias e por ter se dedicado estar ao meu lado quando mais precisei, esta conquista também é sua.

Gostaria de agradecer ao meu orientador, Rodrigo, por todos os conhecimentos passados a mim, pela dedicação a enfrentar os desafios deste trabalho ao meu lado, sempre me incentivando a seguir em frente. Deixo aqui meu muito obrigado por aceitar ser o meu orientador e auxiliar na minha pesquisa, que certamente um dia poderá ajudar a salvar mais vidas.

” Alguns homens veem as coisas como são, e dizem ‘Por quê?’ Eu sonho com as coisas que nunca foram e digo ‘Por que não?’”.

(George Bernard Shaw)

RESUMO

A OMS afirma que as doenças cardiovasculares são, atualmente, a maior causa de mortalidade do mundo. Por sua parte, a entidade também observa que a melhora nos processos de prevenção, diagnóstico e tratamento contribuem para a diminuição das taxas de letalidade. Ademais, dentro da ordem de doenças cardíacas, algumas possuem possibilidade de diagnóstico por meio da identificação de sopros cardíacos, que podem ser detectados a partir do exame de ausculta cardíaca – onde o profissional escuta os sons com origem no coração, por meio de um estetoscópio, e realiza a análise da condição do paciente. Isso posto, este trabalho tem por objetivo propor um sistema computacional que realiza a classificação desses sons cardíacos nas categorias normal e anormal. Para tanto, tal sistema utiliza-se da base de dados *Physionet*, em que os sons cardíacos patológicos e não patológicos são utilizados para treinar e validar classificadores baseado em redes neurais artificiais. Assim, são propostos três tipos de classificadores, baseados em MLP, LSTM e CNN, que são submetidos aos áudios da base de dados. A validação destes modelos, por seu turno, será realizada através da avaliação do desempenho dos classificadores quanto à predição dos sons cardíacos, por meio da aplicação das técnicas de validação sugeridas pelas bibliografias. Portanto, ao fim do processo, objetiva-se obter um modelo de classificador robusto e confiável, que seja capaz de realizar predições de sons cardíacos com eficiência. Após aplicação das metodologias junto aos modelos de classificador, verificou-se o melhor desempenho do modelo CNN, obtendo-se uma acurácia de 0,8947, associada a uma sensibilidade de 0,8915 e uma especificidade de 0,8973.

Palavras-chave: sopros cardíacos; classificador; ausculta; redes neurais; MLP, LSTM, CNN.

ABSTRACT

The World Health Organization (WHO) asserts that cardiovascular diseases are, nowadays, the major cause of mortality in the world. For its part, the entity also observes that the improvement in processes of prevention, diagnostic and treatment contribute to the decrease in lethality rates. Furthermore, inside the group of heart diseases, some have the possibility of diagnose through the identification of heart murmurs, that can be detected by means of an exam of cardiac auscultation – where the professional listens to the sounds originating in the heart, through a stethoscope, and performs the analysis of the patient's condition. Therefore, this work focuses on propounding a computer system that executes the classification of these heart murmurs in normal and abnormal. For that reason, the system uses database Physionet, in which the cardiac sounds, both pathological and not, are applied to train and validate classifiers, based in artificial neural networks. In this way, three types of classifiers are proposed, underpinned by MLP, LSTM and CNN, which are submitted to the audios of the database. The validation of these models, for its turn, will be carried out by evaluating the performance of the classifiers in terms of predicting heart sounds, based on the application of validation techniques suggested by the bibliographies. Consequently, by the end of the process, the objective is to obtain a robust and reliable classifier model, which is capable of performing heart sound predictions efficiently. After applying the methodologies to the classifier models, the CNN model performed best, with an accuracy of 0.8947, associated with a sensitivity of 0.8915 and a specificity of 0.8973.

Keywords: heart murmurs; classifier; auscultation; neural networks; MLP, LSTM, CNN.

LISTA DE FIGURAS

Figura 1 – Estrutura do coração	17
Figura 2 – Eventos do ciclo cardíaco associados ao ECG e PCG	18
Figura 3 – Áreas de ausculta para cada valva	19
Figura 4 – PCG de corações normais e anormais.....	20
Figura 5 - Segmentação de áudio	23
Figura 6 – Comportamento dos bancos de filtros Mel	24
Figura 7 – Obtenção das MFCC.....	25
Figura 8 – Espectrograma de frequência	26
Figura 9 – Neurônio Artificial	27
Figura 10 – Arquitetura RNA genérica	27
Figura 11 – Processo de aprendizagem de uma MLP	29
Figura 12 – Funcionamento de uma Rede Recorrente	30
Figura 13 - Célula de memória de uma LSTM.....	31
Figura 14 - Campos receptivos locais	32
Figura 15 – Rede Neural Convolucional.....	33
Figura 16 – Avaliação por amostragem aleatória	34
Figura 17 – Validação Cruzada	35
Figura 18 - Exemplo de matriz de confusão.....	36
Figura 19 – Exemplo de espaço ROC.....	37
Figura 20 – Visão Geral do Sistema.....	60
Figura 21 – Sistema Proposto	61
Figura 22 – Composição da base de dados.....	63
Figura 23 – Etapas de Pré-Processamento	65
Figura 24 – Etapas do sistema classificador	67
Figura 25 – Repartição de Áudios	70
Figura 26 – Formato do vetor MFCC.....	72
Figura 27 – Formato do vetor MFCC normalizado	73
Figura 28 – Formato do vetor MFCC com aplicação da média aritmética.....	74
Figura 29 – Espectrograma de frequência de um som cardíaco	75
Figura 30 – Espectrograma de frequência limitado a 400 Hz.....	76
Figura 31 – Espaço ROC para os resultados da MLP.....	79
Figura 32 – Espaço ROC para os classificadores LSTM	80

Figura 33 – Espaço ROC para os classificadores CNN	82
Figura 34 – Espaço ROC para os dez grupos de validação cruzada	85
Figura 35 – Associação entre áudio e espectrograma de frequência.....	88
Figura 36 – Som Cardíaco Normal x Anormal – Caso 1	90
Figura 37 - Som Cardíaco Normal x Anormal – Caso 2	91

LISTA DE QUADROS

Quadro 1 – Configuração da Base de dados após a retirada de áudios incertos	69
Quadro 2 – Configuração da Base de dados após segmentação dos áudios	71
Quadro 3 – Configuração da Base de dados após equalização.....	71
Quadro 4 - Configuração da Base de dados de treino e validação	77
Quadro 5 - Resultado de desempenho do modelo de MLP	78
Quadro 6 – Resultado de desempenho do modelo de MLP sem aplicação de filtro .	78
Quadro 7 – Resultado de desempenho do modelo de LSTM	80
Quadro 8 – Resultado de desempenho do modelo de CNN.	82
Quadro 9 – Resultado da validação cruzada do classificador	83
Quadro 10 – Matriz de confusão do classificador.....	84
Quadro 11 – Matriz de confusão dos Resultados de Predição.....	86
Quadro 12 – Matriz de confusão do Grupo 1	102
Quadro 13 – Matriz de confusão do Grupo 2	102
Quadro 14 – Matriz de confusão do Grupo 3	102
Quadro 15 – Matriz de confusão do Grupo 4	102
Quadro 16 – Matriz de confusão do Grupo 5	103
Quadro 17 – Matriz de confusão do Grupo 6	103
Quadro 18 – Matriz de confusão do Grupo 7	103
Quadro 19 – Matriz de confusão do Grupo 8	103
Quadro 20 – Matriz de confusão do Grupo 9	104
Quadro 21 – Matriz de confusão do Grupo 10	104

LISTA DE SIGLAS

MLP	<i>Multilayer Perseptron</i>
LSTM	<i>Long Short Term Memory</i>
CNN	<i>Convolutional Neural Network</i>
OMS	Organização Mundial de Saúde
ECG	Eletrocardiograma
PCG	Fonocardiograma
MFCC	<i>Mel Frequency cepstral coefficients</i>
FFT	<i>Fast Fourier Transform</i>
DFT	<i>Discrete Fourier Transform</i>
DCT	<i>Discrete Cosine Transform</i>
STFT	<i>Short Time Fourier Transform</i>
RNA	Redes Neurais Artificiais
RNR	Redes Neurais Recorrentes
LSTM	<i>Long Short-Term Memory</i>
RGB	<i>Red, Green, Blue</i>
ReLU	Função Linear Retificada
ROC	<i>Receiving Operating Characteristics</i>
CARD	<i>Cardiac Auscultatory Recording Database</i>
VSD	<i>Ventricular Septal Defect</i>
TAP - CRNN	<i>Temporal attentive pooling – convolutional recurrent neural network</i>
VP	Verdadeiro positivo
VN	Verdadeiro negativo
FP	Falso Positivo
FN	Falso Negativo
ANOVA	Análise de variância
CV	Coeficiente de variação
LPC	<i>Linear Predictive coding</i>
PLP	<i>Perseptual Linear Predictive</i>
PNCC	<i>Power Normalized Cepstral Coefficients</i>
DAC	Doença Arterial Coronariana

EWT	<i>Empiric Wavelet Transform</i>
KNN	<i>K-Nearest Neighbors</i>
PS	Pressão Sanguínea
RE	Entropia de Renyi
CQT	Transformada Constante-Q
VQT	Transformada Variável-Q
HCQT	Transformada Constante-Q Híbrida

SUMÁRIO

1 INTRODUÇÃO	14
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 Conceitos básicos.....	17
2.1.1 O coração e o sons cardíacos.....	17
2.1.1.1 Ausculta Cardíaca	19
2.1.1.2 Fonocardiograma	20
2.1.2 Pré-Processamento de Dados	20
2.1.2.1 Integração dos conjuntos de dados.....	21
2.1.2.2 Equalização de dados	21
2.1.2.3 Eliminação de dados ruidosos.....	21
2.1.2.4 Transformação de Dados	22
2.1.2.5 Escala Mel.....	23
2.1.2.6 Coeficientes Cepstrais de frequência Mel	23
2.1.2.7 Espectrogramas de frequência.....	25
2.1.3 Redes Neurais Artificiais	26
2.1.3.1 Deep Learning.....	28
2.1.3.2 Arquiteturas de Redes Neurais	28
2.1.3.3 Redes Perseptron multicamadas	28
2.1.3.4 Redes Neurais Recorrentes	29
2.1.3.5 Redes Neurais Convolucionais	31
2.1.4 Avaliação de modelos de predição.....	33
2.1.4.1 Amostragem aleatória	33
2.1.4.2 Validação Cruzada	34
2.1.4.3 Matriz de Confusão	35
2.1.4.4 Análise ROC.....	37
2.2 Trabalho Correlatos	38
3 METODOLOGIA	60
3.1 Sistema Proposto.....	61
3.1 Base de dados	62
3.2 Pré-processamento	63
3.3 Sistema Classificador	66

3.3.1 Redes Neurais Artificiais	66
3.4 Avaliação do sistema	67
4 ANÁLISE DOS RESULTADOS	68
4.1 Resultados da etapa de pré-processamento	68
4.1.1 Tratamento da Base de Dados.....	68
4.1.2 Transformação de Dados para aplicação em classificadores baseados em MLP e LSTM.....	72
4.1.3 Transformação de Dados para aplicação no classificador baseado em CNN..	74
4.2 Resultados de desempenho dos modelos por amostragem aleatória	77
4.3 Resultados da aplicação de validação cruzada.....	83
4.4 Matriz de confusão do classificador.....	84
4.5 Predições de Sons Cardíacos	86
5 DISCUSSÃO	87
6 CONCLUSÃO	93
REFERÊNCIAS.....	96
APÊNDICE A – MATRIZES DE CONFUSÃO DOS GRUPOS DOS TESTES DE VALIDAÇÃO CRUZADA	102

1 INTRODUÇÃO

Segundo dados da Organização Mundial de saúde, as doenças cardiovasculares configuram-se como a principal causa de morte no mundo. Dados de 2019, indicam que as doenças deste grupo for responsáveis por aproximadamente 27 % do total óbitos da população global. A OMS, observa que melhorias nos processos de prevenção, diagnóstico e tratamento, tem contribuído de forma significativa na redução da mortalidade destes tipos de doenças nos últimos anos. Conscientização por hábitos de vida saudáveis, melhoria da infraestrutura de atendimento, bem como desenvolvimento de estudos e tecnologias que permitam diagnósticos mais precisos, constituem-se como fatores que favorecem a redução destas taxas de mortalidade.

Um sopro cardíaco, constitui-se de um som gerado pelo fluxo sanguíneo turbulento dentro do coração. Na grande maioria dos casos, a turbulência do fluxo sanguíneo é normal, produzindo sons chamados de sopros inocentes. Existem também os sopros cardíacos anormais, que podem indicar anormalidades estruturais nas valvas, ou câmaras cardíacas, ou até mesmo conexões anormais entre duas partes do coração. (HARVARD MEDICAL SCHOOL)

Os médicos conseguem escutar estes sons por meio de um estetoscópio, em um exame conhecido como ausculta cardíaca. Existem áreas do tórax nas quais o som de cada valva cardíaca pode ser mais bem auscultado, o cardiologista, durante o exame, distingue os diferentes sons cardíacos valvulares por meio de um processo de eliminação, de forma a movimentar o estetoscópio entre as áreas do tórax, observando a altura do som nos diferentes pontos de forma a diferenciar gradativamente os componentes sonoros de cada valva (GUYTON & HALL, 2011).

Muitos casos de sopros cardíacos são descobertos de maneira inesperada, durante exames físicos de rotina, visto que os sintomas podem ser percebidos pelo paciente somente quando os sopros estão afetando significativamente a função cardíaca (HARVARD MEDICAL SCHOOL). Observa-se que a identificação precisa de sons cardíacos anormais, principalmente em exames físicos de rotina, configura-se como um pré-requisito fundamental para realização de um diagnóstico precoce, auxiliando no eventual tratamento de uma doença e evitando que o paciente procure um cardiologista somente quando sentir sintomas de perda de função cardíaca. Segundo MEYER (2020), a ausculta cardíaca tem reportado uma sensibilidade de

70% e uma especificidade de 98 % para detecção de valvopatias. O autor observa que estes percentuais ainda variam substancialmente a depender da expertise e proficiência do profissional que realiza a ausculta cardíaca.

O objetivo deste trabalho é criar um sistema computacional que detecte de forma precisa, a existência de sons cardíacos anormais em pacientes, a partir da gravação dos sons cardíacos destes indivíduos. Para implementação deste sistema, utilizou-se a literatura, buscando por pesquisas de sistemas que realizem funções semelhantes, ou que apresentem soluções que possam ser aplicadas aos objetivos desta dissertação. Dentre os trabalhos observados, destacam-se o uso de técnicas de *machine learning*, mais especificamente as Redes Neurais Artificiais, as quais vem ganhando notoriedade por sua eficiência em aplicações de reconhecimento de padrões, justificando a utilização deste tipo de técnica neste trabalho.

Ao desenvolver este estudo, busca-se contribuir à comunidade científica com um sistema que possa detectar sons cardíacos anormais, de forma a possibilitar estudos futuros, com relação a sua aplicação em ferramentas que auxiliem os profissionais da saúde, principalmente do atendimento primário, a realizar diagnósticos de sopros cardíacos com maior eficiência e precisão.

Este documento está organizado em 6 capítulos, onde o primeiro constitui-se nesta introdução e após, o capítulo 2 apresenta a fundamentação teórica, onde disserta-se sobre os conceitos básicos para o entendimento dos trabalhos, bem como os trabalhos acadêmicos existentes a respeito da temática desta dissertação, onde focou-se em apresentar artigos relacionados a aplicações de machine learning já implementadas para detecção de sons cardíacos. O capítulo 3 apresenta a metodologia aplicada para desenvolvimento do sistema proposto, bem como um detalhamento das etapas envolvidas na confecção deste sistema. Em seguida, no capítulo 4, apresenta-se os testes e resultados, bem como as metodologias utilizadas nesta etapa, seguido da discussão destes dados obtidos, realizada no capítulo 5. Desenvolve-se a conclusão, por meio do capítulo 6, onde apresenta-se os principais resultados obtidos, bem como sua interpretação e contribuição no cumprimento dos objetivos do sistema proposto. Por fim é apresentada as referências bibliográficas.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo está subdividido em conceitos básico e trabalhos correlatos, a primeira seção disserta sobre os conceitos aplicados, focando a aplicação aos objetivos deste trabalho. A seção de trabalhos correlatos trata-se de uma revisão de estudos já realizados com a aplicações semelhantes às abordadas por este estudo.

O subcapítulo de conceito básicos, inicia abordando os sons cardíacos, suas causas e a forma de diagnosticar patologias cardíacas por meio destes sons, onde observa-se a necessidade de experiência e treinamento do profissional de saúde para realizar um diagnóstico acurado.

A utilização de *deep learning*, associadas redes neurais e as técnicas de análise de sinais vão ao encontro das necessidades do sistema de classificação de sons cardíacos proposto nesta dissertação, visto que o entendimento do funcionamento e aplicação destas técnicas tornam-se essenciais para a utilização no sistema proposto.

As redes neurais ganharam um enfoque neste capítulo, devido ao grande número de aplicações as quais tem sido aplicada, principalmente na área da visão computacional, como ferramenta de identificação de padrões, conceito que também pode ser aplicado aos objetivos desta dissertação.

Nos trabalhos correlatos foi realizada uma busca por trabalhos científicos que abordassem a temática de sons cardíacos, tanto dissertações a respeito da forma como estes sons são percebidos e diagnosticados pelos profissionais da saúde, como a origem dentro do mecanismo de funcionamento do coração humano.

Foram selecionados também trabalhos que abordassem os sons cardíacos e a classificação pela ótica da inteligência artificial, onde foi possível identificar a utilização de diversas técnicas de *machine learning* por parte dos autores, visando a detecção de sons cardíacos patológicos ou outras patologias cardíacas passíveis de serem identificadas por meio dos sons cardíacos.

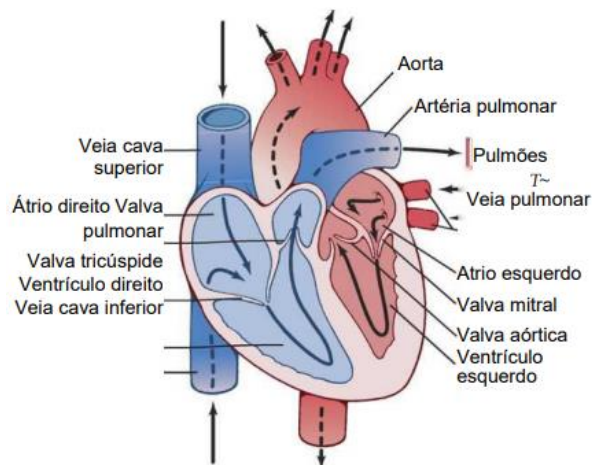
2.1 Conceitos básicos

Esta seção apresenta os conceitos básicos aplicados para o desenvolvimento deste trabalho, onde será realizada uma abordagem a respeito de tópicos de sons cardíacos, aprendizagem profunda (*deep learning*), técnicas de análise de sinais e redes neurais aplicadas a reconhecimento de padrões. Neste subcapítulo, objetiva-se abordar os conceitos com enfoque na aplicação dentro das etapas de desenvolvimento do sistema proposto por esta dissertação.

2.1.1 O coração e o sons cardíacos

Segundo Guyton & Hall (2011), o coração é formado por duas bombas, chamadas de coração direito e esquerdo. A porção direita do coração bombeia sangue para os pulmões, enquanto o esquerdo realiza o bombeamento para os órgãos periféricos. Cada uma destas porções do coração constitui-se de uma bomba pulsátil de duas câmaras, compostas por um átrio e um ventrículo. O átrio comporta-se como bomba de escorva para o ventrículo, de forma a propelir o sangue para seu interior, já os ventrículos fornecem força de bombeamento principal que direciona o sangue pela circulação pulmonar e periférica pelo ventrículo direito e esquerdo, respectivamente. A figura 1 ilustra a estrutura do coração, bem como o fluxo do sangue e as valvas cardíacas.

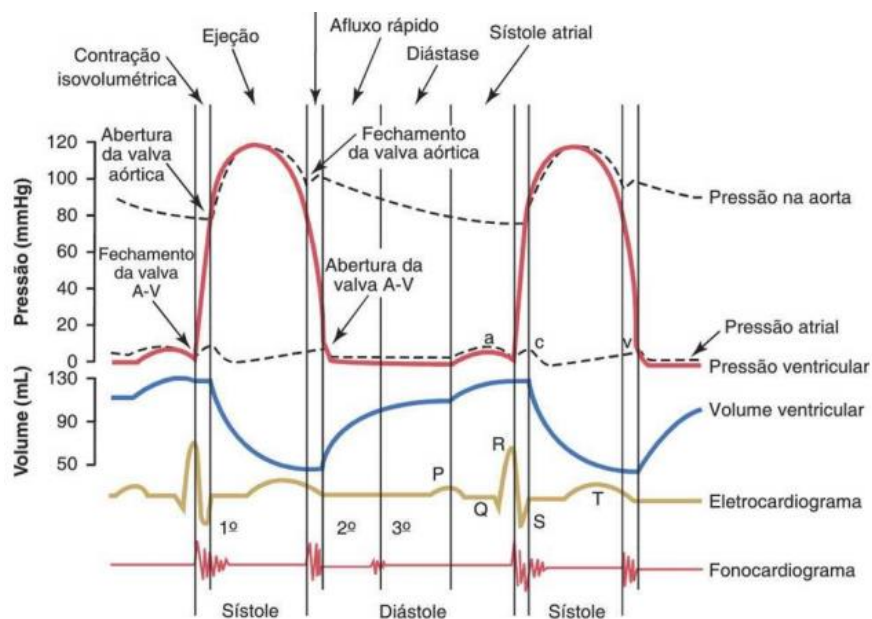
Figura 1 – Estrutura do coração



Fonte: GUYTON & HALL, (2011, p. 107).

O ciclo cardíaco compreende pelo período de diástole seguido de sístole. A diástole trata de um período de relaxamento, onde o coração se enche de sangue, enquanto a sístole trata do período de contração do músculo. Na contração ventricular existe um acúmulo de grande quantidade de sangue nos átrios direito e esquerdo, visto que as valvas atrioventriculares se encontram fechadas. Ao final da sístole, as pressões ventriculares retornam para os baixos valores diastólicos, de forma com que as pressões desenvolvidas nos átrios no processo anterior de sístole forcem a abertura das valvas atrioventriculares. No intervalo de sístole ventricular ocorre um aumento abrupto da pressão ventricular, provocando o fechamento das valvas atrioventriculares, a elevada pressão provoca a abertura das valvas semilunares de forma que o sangue passe a ser lançado para as artérias aorta e pulmonar. Finalizada a sístole, com a diástole ventricular as pressões diminuem, essa diminuição de pressão, associada a alta pressão nas artérias provocam a volta do sangue para os ventrículos de forma a fechar as valvas semilunares. As valvas atrioventriculares, ou tricúspide e mitral possuem função de evitar o refluxo de sangue dos ventrículos para os átrios no processo de sístole, já as valvas semilunares, ou pulmonar e aórtica, tem a atribuição de impedir o refluxo destas valvas para os ventrículos durante a diástole. A figura 2 ilustra os eventos do ciclo cardíaco, bem como sua associação com o eletrocardiograma (ECG) e o fonocardiograma (PCG) (GUYTON & HALL, 2011).

Figura 2 – Eventos do ciclo cardíaco associados ao ECG e PCG



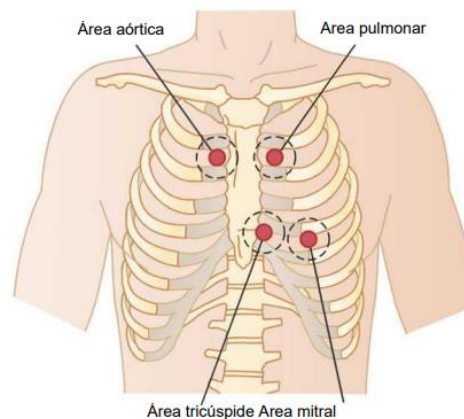
Fonte: GUYTON & HALL, (2011, p. 111).

Os sons cardíacos originam-se do fechamento das valvas, a variação abrupta de pressão causa vibração dos folhetos valvares e os líquidos, originando sons disseminados por toda região do tórax. A contração dos ventrículos e o fechamento das valvas mitral e tricúspide gera o primeiro som, conhecido como primeiro som cardíaco (S1). No final da sístole, o fechamento das valvas pulmonar e aórtica causa o segundo som cardíaco (S2) (GUYTON & HALL, 2011).

2.1.1.1 Ausculta Cardíaca

A ausculta cardíaca consiste em escutar os sons do corpo por meio de um estetoscópio. O cardiologista distingue os sons de diferentes valvas por um processo de eliminação, de forma a movimentar o estetoscópio pelas diferentes áreas de ausculta e observando o padrão do som em cada uma destas áreas e diferenciando gradativamente os componentes de cada valva (GUYTON & HALL, 2011). Abaixo, a figura 3 ilustra regiões do tórax onde cada valva é mais bem auscultada.

Figura 3 – Áreas de ausculta para cada valva



Fonte: GUYTON & HALL, (2011, p. 280).

A partir da figura acima, observa-se as áreas em que os sons cardíacos das valvas cardíacas se tornam mais nítidos de ser escutados por meio do estetoscópio. Além de escutar, é possível visualizar a forma de onda dos sons cardíacos por meio de um exame chamado de fonocardiograma.

2.1.2.1 Integração dos conjuntos de dados

A integração de conjuntos de dados é aplicada quando os dados disponíveis são originários de fontes diferentes. Antes de submeter esses dados aos modelos, é necessário promover a integração em um conjunto único, formando uma base de dados centralizada. Deve-se garantir que todos os dados movidos para a base de dados central representem o mesmo atributo (FACELI, *et. al.*, 2021).

2.1.2.2 Equalização de dados

A equalização de dados é aplicada principalmente em problemas de classificadores. Em alguns conjuntos de dados reais, não é incomum que exista uma prevalência maior de dados de uma classe em relação as demais, tornando este conjunto de dados desbalanceados. Esta característica pode comprometer o desempenho de modelos classificadores, visto que podem tender a classificar os novos dados à classe majoritária do conjunto. Uma das soluções propostas pelas bibliografias trata-se de redefinir o conjunto de dados, onde é possível retirar alguns dados da classe majoritária ou acrescentar novos dados a classe minoritária. Deve se atentar que a redução dos dados e o aumento dos dados podem causar os fenômenos de *underfitting* e *overfitting*, no primeiro, o modelo não é capaz ajustar-se aos dados, devido a serem escassos e no segundo, com o aumento da base de dados pode ocorrer um ajuste exagerado ao conjunto de dados (FACELI, *et. al.*, 2021).

2.1.2.3 Eliminação de dados ruidosos

Segundo FACELI, *et. al.* (2021), os ruídos são compreendidos como variações ou erros aleatórios na geração ou na medida de um dado. A presença de ruídos, em alguns casos pode ser identificada por meio de *outliers*, que consistem em dados como valores discrepantes quando comparados aos demais valores da base de dados. Submeter um modelo a um conjunto de dados pode levar a um *overfit*, de forma que o modelo tenda a identificar características específicas destes ruídos e não as de interesse. Por outro lado, em alguns casos as características desejadas estão presentes justamente nestes dados, nestes casos, a retirada não é indicada (FACELI, *et. al.*, 2021).

2.1.2.4 Transformação de Dados

Em muitas oportunidades, os registros das bases de dados, encontram-se em escala diferentes, ou mesmo estando em mesma escala, podem variar dentro de faixas de valores diferentes em comparação a outras amostras do conjunto. Esta variação de magnitude de valores, muitas vezes não são desejadas, alguns algoritmos de modelagem são altamente sensíveis a dados com grande variação. Para resolver este tipo de problema, implementa-se a normalização dos dados. Esta técnica é indicada quando existe uma grande variação de magnitude entre atributos diferentes dos dados. A bibliografia indica duas principais formas de implementação da normalização: por reescala e por padronização. A primeira, converte os dados para uma escala que, geralmente varia entre 0 e 1 e a segunda técnica define um valor central e um valor de espalhamento comuns para todos os dados de cada amostra (FACELI, et. al., 2021). Abaixo, demonstra-se os métodos de cálculo dos dois tipos de normalização, a equação 1 descreve o cálculo da normalização por reescala, enquanto a equação 2 denota o cálculo da normalização por padronização, onde μ denota a média aritmética e σ denota o desvio padrão do conjunto de dados em análise.

$$V_{novo} = \min + \frac{V_{atual} - menor}{maior - menor} (max - \min) \quad (1)$$

Na equação 1, é possível notar que os valores normalizados, são calculados com base na proporção entre o valor dos dados perante os valores máximos e mínimos do conjunto, de forma que todos os dados que passem pelo processo de normalização variem dentro da mesma escala.

$$V_{novo} = \min + \frac{V_{atual} - \mu}{\sigma} \quad (2)$$

A equação 2 permite observar que, no processo de normalização por padronização, subtrai-se os dados do ponto central do conjunto, representado pela média aritmética μ e divide-se pelo desvio padrão σ , obtendo-se uma métrica que representa a proporção entre o desvio do dado em relação ao seu ponto central e a variação dos dados deste conjunto.

A etapa de pré-processamento, por vezes exige algumas abordagens complementares as técnicas demonstradas neste capítulo. Em aplicações de áudio, existem técnicas que podem ser aplicadas para converter os dados em formatos que evidenciem as características de interesse dentro do conjunto de registros. O próximo

tópico deste capítulo destina-se a ilustrar algumas técnicas que são aplicadas ao pré-processamento e extração de características em áudios.

2.1.2.5 Escala Mel

A escala mel consiste em uma escala de frequência, que denota a percepção de um tom pelo ouvido humano, esta escala tem a característica de possuir um comportamento aproximadamente linear até 1 kHz e logarítmico acima deste valor (SANTOS, 1997). A equação 3 é utilizada para relacionar a frequência linear e a mesma métrica em escala Mel:

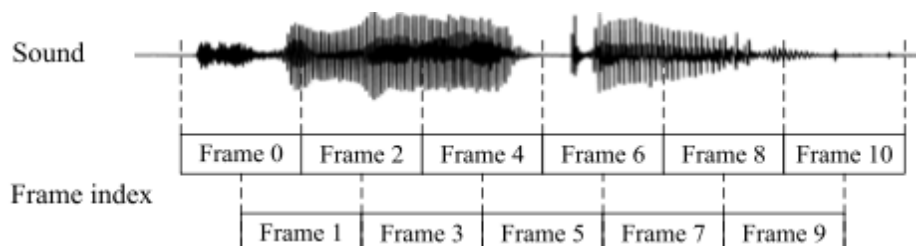
$$Mel(f) = 2595 * \log_{10} \left(1 + \frac{f}{700} \right) \quad (3)$$

A utilização de escala mel, é utilizada em diversas aplicações de reconhecimento de voz, as quais buscam por características específicas dentro do sinal sonoro. Uma das técnicas mais utilizadas para identificação de características, constitui-se nos coeficientes cepstrais de frequência mel, que serão descritos nos próximos tópicos desta dissertação.

2.1.2.6 Coeficientes Cepstrais de frequência Mel

Os coeficientes Cepstrais de frequência Mel, ou MFCC, constitui-se como uma técnica de extração de características de sons, amplamente utilizada por sistemas de reconhecimento de voz. A MFCC implementa um algoritmo, onde, inicialmente o áudio é segmentado em pedaços ou *frames*, e um vetor de características sonoras é retirada de cada uma destas parcelas de áudio. A figura 7 ilustra o processo de segmentação do sinal. (J. Jo, H. Yoo and I. -C, 2016)

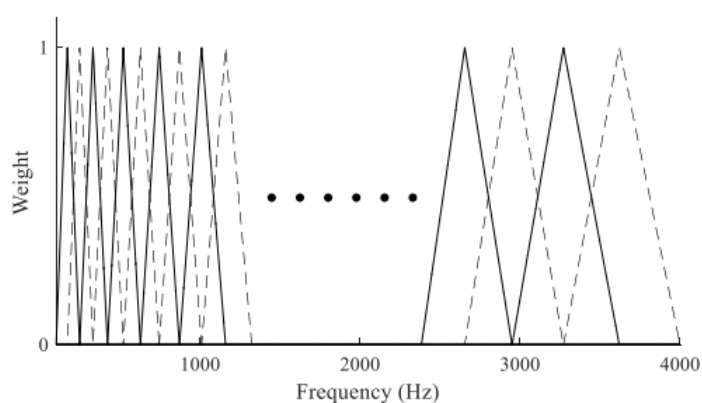
Figura 5 - Segmentação de áudio



Fonte: J. Jo, H. Yoo a I. -C, (2016, p.754).

Após o processo de segmentação, o sinal de áudio passa por uma etapa de pré-ênfase, onde os componentes de alta frequência são evidenciados por meio de um filtro passa altas. Seguido da evidenciação das altas frequências, realiza-se um processo de janelamento. Como observado na figura 8, existe uma sobreposição de *frames* resultante da etapa de segmentação do sinal, executa-se o janelamento *Hamming*. (J. Jo, H. Yoo and I. -C, 2016). Realizado janelamento, extrai-se o espectro de distribuição de energia de energia dos frames, por meio da transformada rápida de Fourier (*Fast Fourier Transform*) ou FFT, convertendo o sinal, originalmente no domínio de tempo para o domínio da frequência. Em seguida, eleva-se os módulos da FFT ao quadrado, de forma a converter o espectro de amplitudes para um espectro de energia do sinal (Q. Li et al., 2020). Aplica-se então os chamados bancos de filtro mel, que objetivam simular a percepção de som do ouvido humano, que possui a característica de ser mais sensível a baixas frequências. Aplica-se a distribuição de filtros que aumentam sua largura de banda, a medida em que o espectro de frequência se aproxima das frequências maiores, de forma que, para baixas frequências, os filtros sejam mais seletivos em comparação com as altas frequências (Q. Li et al., 2020). A figura 8 ilustra o comportamento dos bancos de filtro Mel.

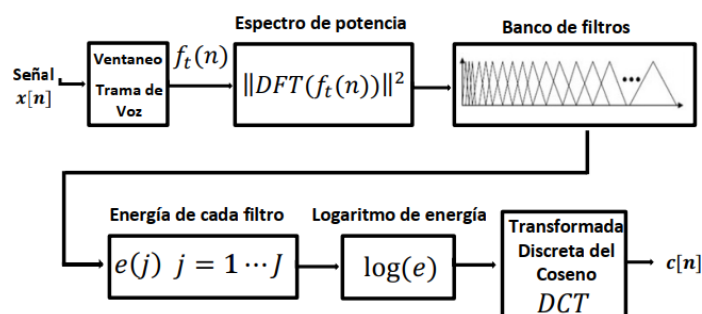
Figura 6 – Comportamento dos bancos de filtros Mel



Fonte: Q. Li et al (2020, p.48722).

Após passar pelos bancos de filtro, o sinal obtido passa por uma operação, onde aplica-se a transformada discreta de cosseno (DCT) e multiplica-se pelo logaritmo da energia do sinal obtida da FFT. A partir desta operação, é possível obter-se os coeficientes mel cepstrais para cada frame do sinal (Q. Li et al., 2020). A figura 9 resume os processos envolvidos para obtenção das MFCC.

Figura 7 – Obtenção das MFCC



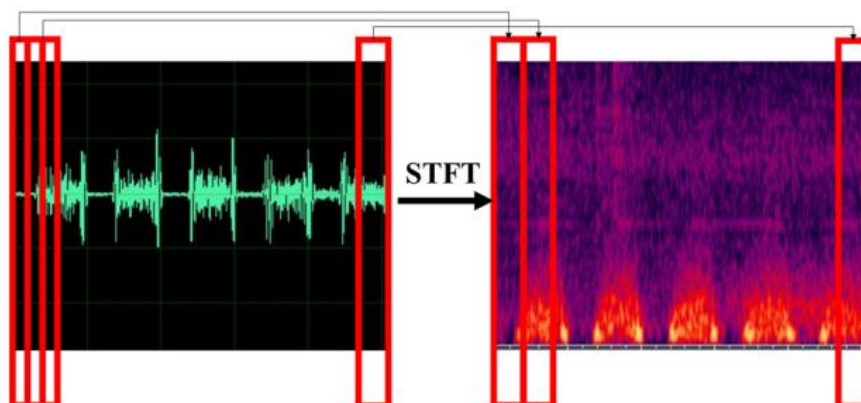
Fonte: T. Villa-Cañas *et al.*, (2012).

Os MFCCs não se constituem como única forma de identificação de características em sinais de áudio. Uma das formas alternativas de obtenção destas características pode ser por meio de imagens que representem a dinâmica dos sinais. Estas imagens podem ser expressas por meio de espectrogramas de frequência, permitindo a visualização do comportamento dos componentes de frequência ao longo do tempo.

2.1.2.7 Espectrogramas de frequência

Os espectrogramas de frequência são representações que permitem a visualização de um sinal tanto no domínio do tempo quanto na frequência. Para esta visualização ser possível, torna-se necessário aplicar a FFT em frações do sinal. Esta operação também é conhecida como *Short Time Fourier Transform* (STFT). Aplica-se a FFT em janelas de tempo fixos do sinal, de maneira a obter-se a informação do comportamento das frequências ao longo destes intervalos (Z. M. Amin, 2002). A figura 10 ilustra um exemplo de espectrograma de frequência de um sinal de áudio.

Figura 8 – Espectrograma de frequência



Fonte: WANG *et al.* (2020, p.3).

A partir da execução da etapa de pré-processamento, os dados estão devidamente tratados para serem submetidos ao treinamento e validação com auxílio de algoritmos de *machine learning*, dentre os quais se destacam as Redes Neurais Artificiais, devido a sua grande gama de aplicabilidades.

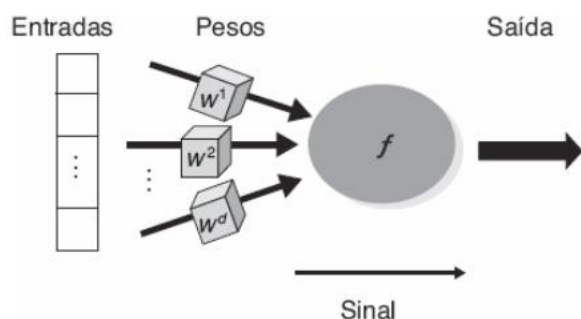
2.1.3 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas computacionais distribuídos, constituídos por diversas unidades interconectadas, que possuem a incumbência de realizar processamentos simples, cada uma destas unidades pode ser chamada de neurônio. Os neurônios podem ser distribuídos em diversas camadas e interligados por conexões. Em muitas arquiteturas, a conexões possuem pesos associados, que ponderam a entrada recebida por cada neurônio da rede, os pesos assumem valores positivos e negativos, a depender do comportamento da conexão ser excitatório o inibitório. As ponderações fazem parte do processo de aprendizado, onde tem seus valores ajustados e codificam o conhecimento adquirido pela rede. (FACELI, *et. al.*, 2021). Segundo FACELI, *et. al.* (2021), uma RNA possui dois expoentes, a arquitetura e o aprendizado, a primeira está relacionada com o tipo, número de neurônios, e forma de conexão destes e o aprendizado relaciona as regras utilizadas para ajuste de pesos e as informações que são utilizadas por esta regra.

Dentro da arquitetura da RNA, o neurônio constitui-se como a unidade de processamento fundamental, estas unidades recebem valores de entrada e de saída, o valor de entrada recebido passa por uma atribuição de pesos, que são combinados

por meio de uma função de ativação, que disponibiliza seu resultado para a saída do neurônio. A figura 11 ilustra o funcionamento básico de um neurônio artificial.

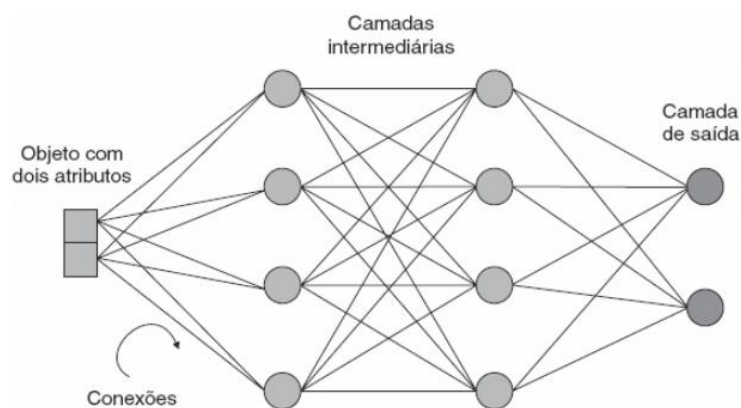
Figura 9 – Neurônio Artificial



Fonte: FACELI, *et. al.* (2021, p.103).

Dentro das funções de ativação, destacam-se duas que são amplamente utilizadas, a função sigmoideal (Sigmoid) e linear retificada (ReLU), na função Sigmoid, existe um valor de limiar que define quando a saída é ativada ou não. No momento em que a soma das entradas do neurônio ultrapassa os valores de limiar, o neurônio é ativado, neste tipo de função, a saída varia em um intervalo de 0 a 1, sendo 1 o estado ativo. A função ReLU, retorna o valor 0, caso receba valores negativos e para valores positivo, retorna o próprio valor recebido. Em uma arquitetura RNA genérica, ao utilizar-se duas ou mais camadas, os neurônios podem receber em suas entradas, sinais de saídas de neurônio de camadas anteriores, bem como enviar seus sinais de saída para a entrada de neurônios de camadas adiante (FACELI, *et. al.*, 2021). A figura 12 ilustra uma arquitetura RNA genérica.

Figura 10 – Arquitetura RNA genérica



Fonte: FACELI, *et. al.* (2021, p.105).

A medida em que se objetiva a construção de modelos mais robustos, muitas vezes torna-se necessário aumentar a arquitetura das redes neurais, seja o número de neurônios, camadas escondidas ou ambos. As redes neurais constituídas de muitas camadas entram dentro do conceito de *deep learning*, que será discutido no próximo item deste capítulo.

2.1.3.1 Deep Learning

O *Deep Learning* (aprendizado profundo), consiste em uma subárea do *Machine Learning* (Aprendizado de máquina). A *Deep Learning* é associada às redes neurais artificiais profundas, que, de maneira sucinta, usam camadas de neurônios matemáticos, de forma à saída de cada camada comportar-se com uma entrada para a camada seguinte gerando um fluxo de informações passadas entre as camadas da rede. Como arquitetura geral, uma *Deep Learning* possui uma camada de entrada, seguida pelas camadas ocultas e pela camada de saída. O *Deep learning* constitui-se como uma das principais ferramentas de visão computacional, onde destacam-se o reconhecimento de áudio e de imagens. (DATA SCIENCE ACADEMY, 2022).

2.1.3.2 Arquiteturas de Redes Neurais

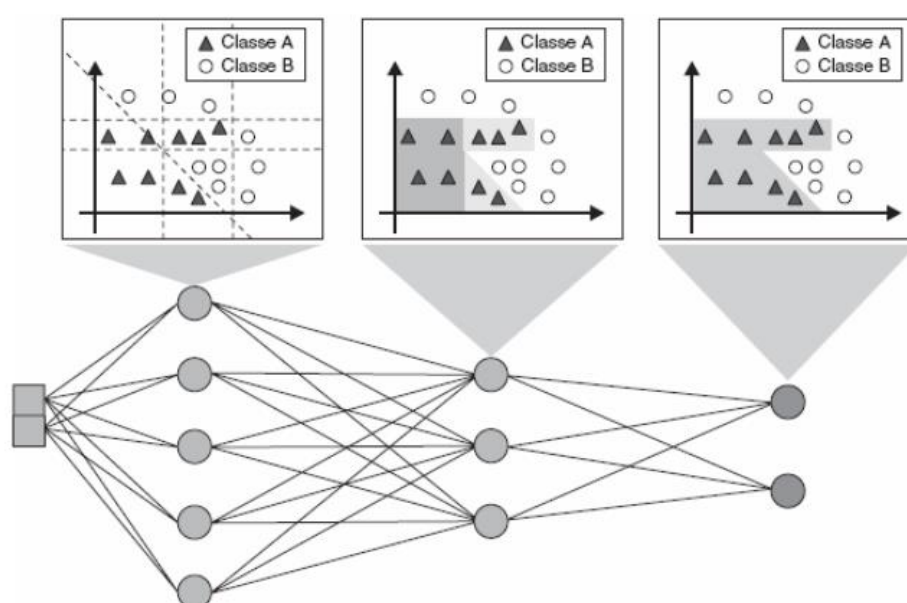
As redes neurais apresentam algumas variações em suas arquiteturas. A depender da aplicação, algumas destas arquiteturas são mais indicadas, em comparação as demais. Este subcapítulo destina-se a apresentar três das principais arquiteturas, bem como seus princípios de funcionamento e aplicabilidade.

2.1.3.3 Redes Perceptron multicamadas

Segundo FACELI, *et. al.* (2021), a rede perceptron multicamadas, ou *multilayer perceptron* (MLP), possuem uma ou mais camadas intermediárias e uma camada de saída. Neste tipo de arquitetura, cada neurônio desempenha uma função específica, de forma que, a função de cada neurônio seja a combinação das funções executadas pelos neurônios da camada anterior. A medida em que existem mais camadas na rede, maior é a complexidade do processamento realizado e das funções executadas pelos neurônios. As saídas da MLP são atreladas a classes dos dados de entrada. O

treinamento de uma MLP é realizado pela comparação entre a resposta obtida e a resposta desejada, onde, a diferença entre estes valores indica o erro de previsão para a entrada apresentada, este erro é utilizado nas próximas iterações de treinamento para ajustar os pesos aplicados às entradas dos neurônios das camadas intermediárias (FACELI, *et. al.*, 2021). A figura 13 ilustra o processo de aprendizagem de uma MLP que realizada a classificação de exemplos recebidos em duas classes distintas.

Figura 11 – Processo de aprendizagem de uma MLP



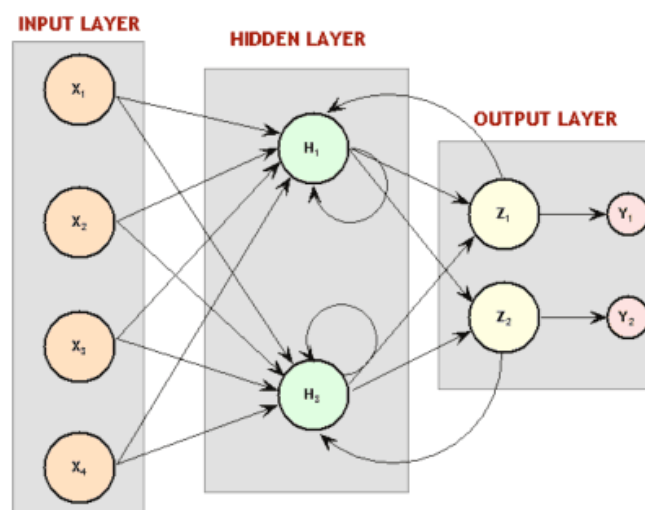
Fonte: FACELI, *et. al.* (2021, p.109).

2.1.3.4 Redes Neurais Recorrentes

As Redes Neurais Recorrentes (RNR), tem como principal característica, o sentido em que os processamentos são realizados em sua rede. Enquanto em redes como MLP, as conexões entre os neurônios fluem da camada de entrada até a camada de saída, em uma rede recorrente, estas conexões fluem tanto de camadas anteriores para camadas à frente, quanto no sentido contrário. Segundo *Data Science Academy* (2022), os exemplos apresentados a uma RNR em etapas anteriores podem afetar os resultados de novos exemplos que são submetidos a rede, de forma que exemplos antigos podem auxiliar na previsão de novos exemplos. As RNR tornaram-se muito úteis, especialmente em aplicações com dados sequenciais e temporais,

como áudio, processamento de linguagem natural e séries temporais (DATA SCIENCE ACADEMY, 2022). A figura 14 ilustra o funcionamento básico de uma RNR.

Figura 12 – Funcionamento de uma Rede Recorrente

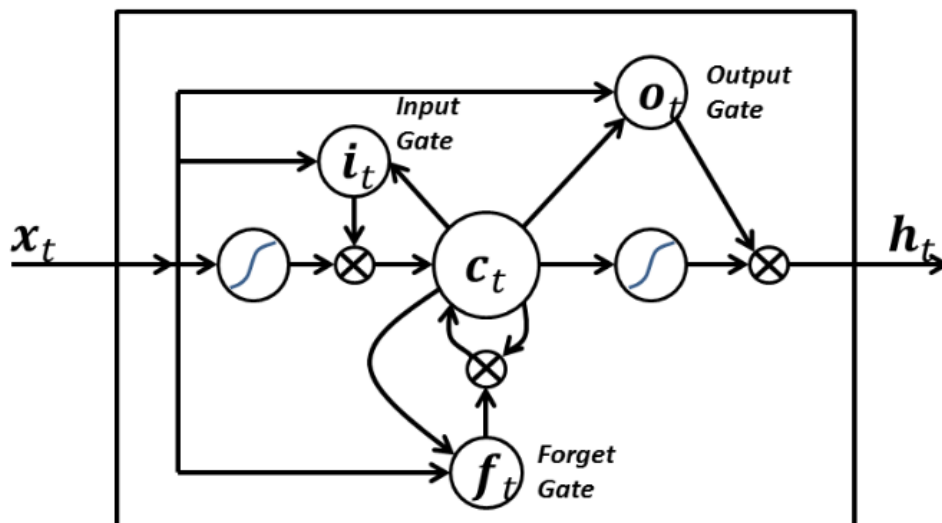


Fonte: *Data Science Academy* (2022).

Dentro das redes recorrentes, destaca-se uma variação em especial, as redes *Long Short-Term Memory* (LSTM). Ao aplicar este tipo de variação, possibilita-se que a rede recorrente continue seu processo de aprendizagem por um número maior de amostras de tempo, de forma relacionar de maneira mais eficaz as causas e efeitos de amostras de tempos mais antigas às amostras que vão sendo mais recentemente apresentadas a rede. (DATA SCIENCE ACADEMY, 2022).

A LSTM possui uma célula, chamada célula de memória, que se situa fora do fluxo de dados da rede recorrente. Esta célula possui a incumbência de decidir os dados armazenados, escritos e lidos, por meio das portas de entrada e saída também chamados de *Gate Input* e *Gate Output*. Quando a informação é submetida a estas portas, aplica-se um filtro, que, baseado em pesos atribuídos, decidem se esta informação é repassa a frente ou não. Estes pesos são calculados e calibrados ao longo do processo de aprendizagem da rede, de forma que estas células sejam capazes de controlar informações que entram e que saem da unidade, excluindo informações que não são consideradas relevantes para o aprendizado (DATA SCIENCE ACADEMY, 2022). A figura 15 ilustra como a célula de memória realizada o controle de informações que são repassadas adiante.

Figura 13 - Célula de memória de uma LSTM



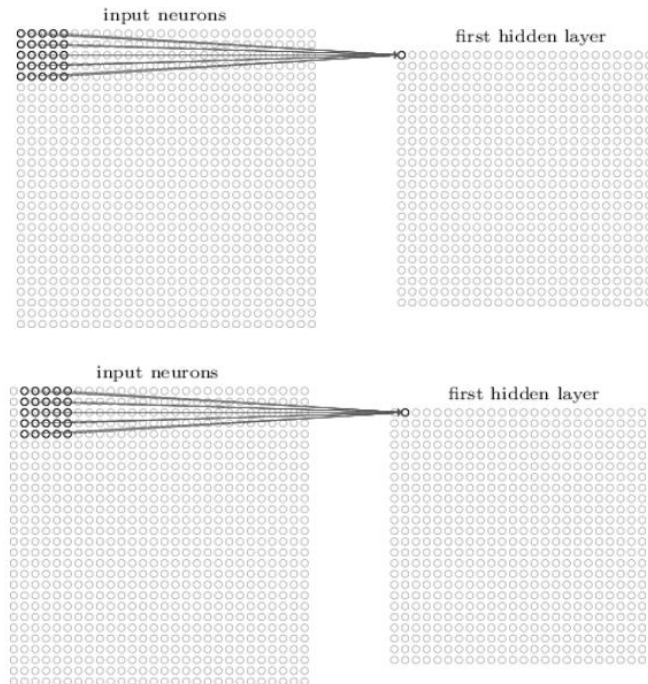
Fonte: *Data Science Academy* (2022).

2.1.3.5 Redes Neurais Convolucionais

As redes neurais convolucionais, ou CNN são modelos frequentemente utilizados em aplicações de visão computacional, visto que apresentam uma grande eficácia no reconhecimento de imagens, podendo ser aplicadas para resolução de problemas em diversas áreas, como: carros autônomos, robótica, segurança e diagnósticos médicos. Também é possível aplicar este tipo de rede a arquivos de áudio, quando este é representado em imagem por um diagrama de espectro. As CNNs tratam as imagens como objetos tridimensionais, representando a decomposição da imagem nos espectros vermelho, verde e azul (RGB). Esta rede recebe na verdade três conjuntos de matrizes, cujas dimensões denotam o tamanho em pixels da imagem. A medida em que estas imagens percorrem a CNN, estas dimensões sofrem alterações, de forma a gerar em sua camada de saída uma série de probabilidades para cada classe definida. A classe que obtiver a maior probabilidade torna-se então a classificação da imagem de entrada. A partir de seu princípio de funcionamento, a rede identifica relações espaciais e temporais, aplicando filtros de relevância à imagem a qual é submetida, estas imagens são reduzidas pela CNN, de forma a otimizar o processamento sem perda de recurso. Existem três pilares principais se tratando de redes convolucionais: *pooling*, compartilhamento de pesos e campos receptivos locais. Os campos receptivos locais trata-se de um agrupamento de pixels, na qual cada neurônio da primeira camada escondida da rede estará

conectado, de forma que todos os neurônios desta camada serão conectados a agrupamentos de pixels de mesma dimensão, deslocando estes grupos ao longo da imagem a qual a rede é submetida (*DATA SCIENCE ACADEMY, 2022*). A figura 16 sintetiza os agrupamentos de pixel realizados e sua associação com cada neurônio oculto, bem como seu deslocamento ao longo da imagem.

Figura 14 - Campos receptivos locais

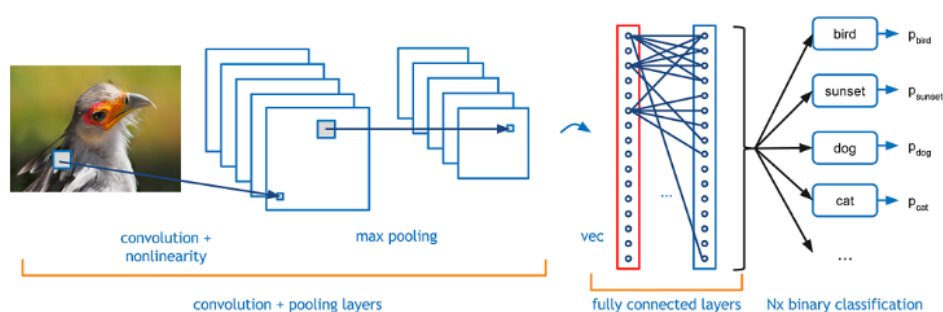


Fonte: *Data Science Academy (2022)*

O compartilhamento de pesos em uma CNN, na verdade trata-se de que, para cada neurônio presente na camada oculta são utilizados os mesmos pesos e características, de forma que, o mesmo recurso possa ser mapeado em diferentes locais da imagem, formando o que se chama de mapa de recursos. Em uma rede convolucional, utiliza-se de vários mapas de recurso, cada um sendo capaz de identificar algum tipo de característica. (*DATA SCIENCE ACADEMY, 2022*).

A camada de pooling realiza o agrupamento dos mapas de recursos, de forma a sintetizar estes recursos, obtendo-se um número de parâmetros menor, que é repassado as próximas camadas da rede. Geralmente, camadas de pooling são utilizadas após as camadas convolucionais. (*DATA SCIENCE ACADEMY, 2022*). A figura 17 demonstra os componentes de uma rede neural convolucional.

Figura 15 – Rede Neural Convolucional



Fonte: *Data Science Academy* (2022)

Após escolher uma arquitetura de RNN que melhor se adapte ao modelo que deseja-se desenvolver, realizar o treinamento e a validação, necessita-se avaliar o desempenho deste sistema, garantindo a confiabilidade. Os próximos tópicos deste capítulo destinam-se a explicar algumas técnicas de avaliação que podem ser aplicadas a modelos de predição.

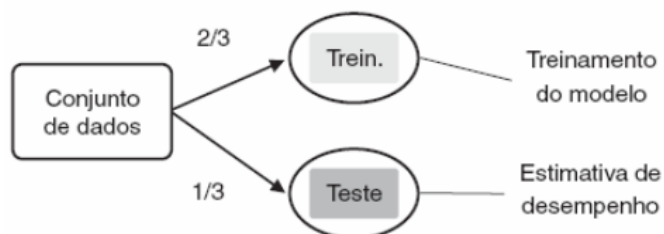
2.1.4 Avaliação de modelos de predição

Após elaborar um modelo de classificador que seja capaz de realizar predições, torna-se necessário avaliar o desempenho e a robustez, este subcapítulo visa abordar algumas das principais técnicas que podem ser implementadas na avaliação destes modelos.

2.1.4.1 Amostragem aleatória

Esta técnica consiste em dividir os dados de treinamento e validação em uma proporção, geralmente de 2/3, sendo esta primeira fração os dados de treinamento e 1/3 para validação do modelo. Deve ser garantido a condição e exclusividade mútua dos dados de treinamento e validação, ou seja, dados que estado no grupo de treino não podem figurar no de testes, a recíproca também é verdadeira. A partir da divisão dos dados, calcula-se a acurácia do modelo, quando submetido aos dados de teste (FACELI, *et. al.*, 2021). A figura 18 ilustra as etapas da avaliação por amostragem aleatória.

Figura 16 – Avaliação por amostragem aleatória



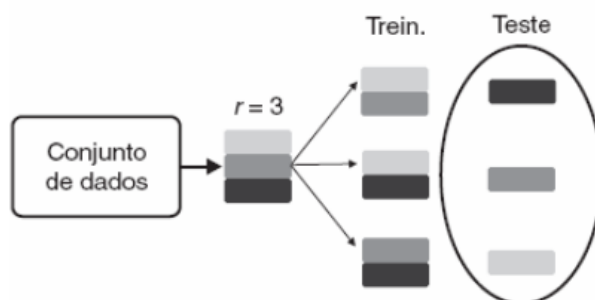
Fonte: FACELI, *et. al.* (2021, p. 151).

A abordagem por amostragem aleatória por vezes pode apresentar resultados que não são confiáveis, visto que, aleatoriamente são escolhidos os conjuntos de treino e validação, pode ser obtida uma combinação de conjuntos, de forma que os resultados sejam muito bons, porém, não se tem a garantia de que, se fossem gerados novos conjuntos de treino e teste se obteria um resultado de igual qualidade. A partir deste problema, necessita-se treinar e validar um modelo com diferentes combinações entre os conjuntos, em uma técnica conhecida como validação cruzada (FACELI, *et. al.*, 2021).

2.1.4.2 Validação Cruzada

Na técnica de validação cruzada, os dados são divididos em n de grupos. Um destes grupos é retirado dos demais para constituir os dados de validação, os restantes são utilizados para treino do modelo. Este processo é repetido n vezes, e em cada uma das n iterações, utiliza-se um dos grupos para teste. O desempenho final do modelo é determinado pela média do desempenho do modelo nas n iterações. Em comparação a avaliação por amostragem aleatória, a validação cruzada garante uma avaliação mais precisa da robustez do modelo, visto que este é treinado e validado em com diferentes combinações do conjunto de dados, algo que não ocorre na técnica ilustrada anteriormente (FACELI, *et. al.*, 2021). A figura 19 ilustra a implementação da validação cruzada utilizando 3 grupos.

Figura 17 – Validação Cruzada



Fonte: FACELI, *et. al.* (2021, p. 151).

A partir da validação do modelo, pode-se calcular algumas métricas que especifiquem mais detalhadamente o desempenho do sistema em requisitos mais específicos. No caso de um classificador, pode-se ter interesse em obter o índice de acerto e de erro das previsões para cada classe de forma individual. Estes índices podem ser visualizados em uma matriz, chamada de matriz de confusão.

2.1.4.3 Matriz de Confusão

A matriz de confusão, consiste em uma maneira de visualização do desempenho de classificadores, por meio da construção de uma matriz que relaciona previsões corretas e incorretas para cada classe (FACELI, *et. al.*, 2021). Esta matriz tem dimensões 2x2 apresentando as seguintes variáveis:

- Verdadeiro Positivo (VP): Número de exemplos positivos classificados corretamente.
- Verdadeiro Negativo (VN): Número de exemplos negativo classificados corretamente.
- Falso Positivo (FP): Número de exemplos negativo classificados erroneamente como positivos.
- Falso Negativo (FN): Número de exemplo positivos classificados erroneamente como negativos.

A figura abaixo exemplifica a formatação de uma matriz de confusão:

Figura 18 - Exemplo de matriz de confusão

		Classe predita	
		+	-
Classe verdadeira	+	VP	FN
	-	FP	VN

Fonte: FACELI, *et. al.* (2021, p. 152).

A partir dos dados da matriz de confusão, é possível extrair informações a respeito do desempenho do modelo, sendo esta descritas e obtidas pelas equações abaixo:

- Taxa de Falso Negativo:

$$TFN = \frac{FN}{VP + FN} \quad (4)$$

- Taxa de Falso Positivo:

$$TFP = \frac{FP}{FP + VN} \quad (5)$$

- Taxa de Erro Total:

$$TErr = \frac{FP + FN}{N} \quad (6)$$

- Taxa de acerto, ou acurácia:

$$Acc = \frac{VP + VN}{n} \quad (8)$$

$$n = VP + VN + FP + FN \quad (9)$$

- Precisão: Taxa de exemplos positivos classificados corretamente em relação ao total de total de casos preditos como positivos:

$$Prec = \frac{VP}{VP + FP} \quad (10)$$

- Sensibilidade: Taxa de acerto de predições positivas:

$$Sens = \frac{VP}{VP + FN} \quad (11)$$

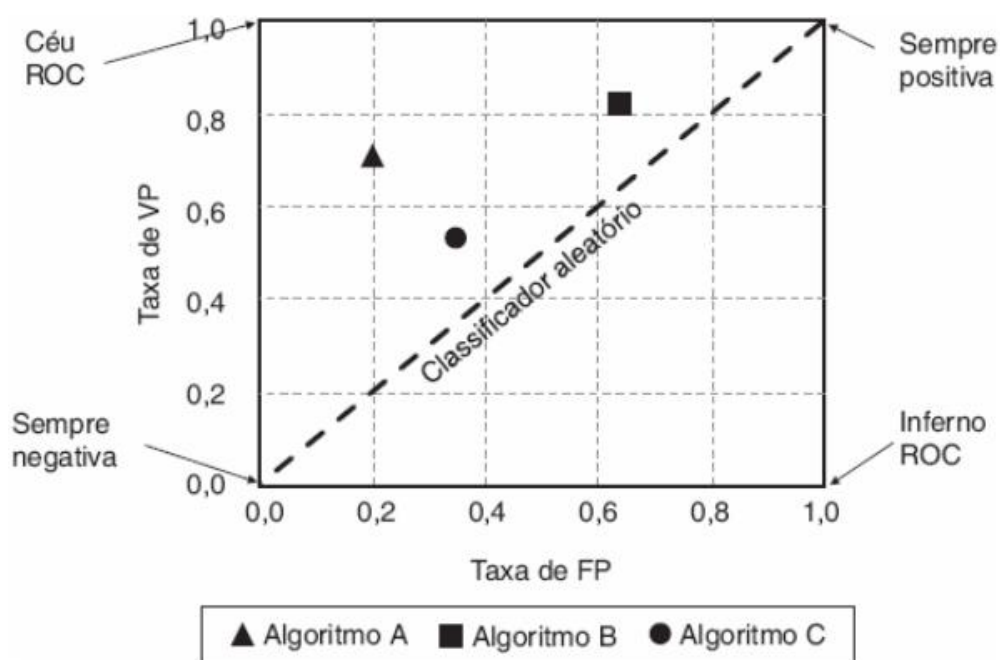
- Especificidade: Taxa de acerto de predições negativas:

$$Esp = \frac{VN}{VN + FP} \quad (12)$$

2.1.4.4 Análise ROC

Segundo FACELI, et. al. (2021), a análise ROC (*Receiving Operating Characteristics*) constitui-se como uma alternativa de avaliação de classificadores binários. Este método utiliza-se da visualização de um gráfico formado pelas taxas de falsos positivos e a taxa de verdadeiros positivos, representando os eixos x e y, respectivamente. A figura 21 ilustra o espaço ROC para comparação de três algoritmos.

Figura 19 – Exemplo de espaço ROC



Fonte: FACELI, et. al. (2021, p. 155).

Na figura acima, a linha pontilhada representa o limiar para um classificador ser considerado aleatório, de forma que, modelos presentes abaixo desta linha possuem comportamentos de predição insatisfatórios. Neste tipo de comparação, objetiva-se considerar o melhor modelo aquele que possui a maior taxa de verdadeiro positivos e a menor taxa de falsos positivos, buscando-se pelo algoritmo alinhado mais acima e a esquerda do gráfico, em relação aos outros modelos (FACELI, et. al., 2021).

2.2 Trabalho Correlatos

Esta seção tem como objetivo reunir informações a respeito de trabalhos já realizados com abordagem semelhante a este estudo. Iniciou-se a pesquisa por trabalhos correlatos primeiramente por dissertações da área da saúde, onde os autores abordam desde a geração do som cardíaco, tipos de sons cardíacos até a identificação de patologias por meio destes, a partir da ausculta cardíaca, a qual exige uma grande experiência e treinamento do profissional para aplicação adequada de detecção de possíveis anomalias nos sons cardíacos.

Em seguida buscou-se relatos de autores que abordassem o tema à luz da inteligência artificial. As dissertações encontradas na bibliografia trazem técnicas de *machine learning* para classificação de sons cardíacos e outras patologias cardíacas, bem como relatos a respeito de métodos de pré-processamento, que incluem tratamento de sinais, segmentação e identificação de características de acústicas de sinais de fonocardiograma.

O relato de Meyer (2020) revisa tópicos relacionados a ausculta cardíaca em adultos, bem como técnicas de diferenciação dos tipos de sopros cardíacos. A ausculta cardíaca é uma das técnicas mais utilizadas para detectar alterações anatômicas e fisiológicas cardiovasculares, constituindo-se muitas vezes como o primeiro diagnóstico indicativo de problemas relacionados a válvulas cardíacas. Esta técnica possui uma sensibilidade de 70 % na detecção de valvopatias, a depender da expertise e proficiência do examinador. Os sopros cardíacos resultam da turbulência causado pelo fluxo de sangue no coração, a intensidade destes sons depende de fatores como tamanho dos vasos por onde flui o sangue, volume e diferença de pressão, quando flui por estreitamentos. A ferramenta utilizada para a ausculta se constitui no estetoscópio, este possui o modelo padrão, com princípio de funcionamento acústico, possuindo um sino para sons de baixas frequências e um diafragma para alta frequência. O estetoscópio eletrônico utiliza de técnicas de processamento de sinal, possibilitando o especialista a escolha de utilização do sino, diafragma e faixas de frequência para ausculta do paciente, além de permitir a gravação de sons para posterior análise. Segundo o autor, indica-se que o clínico realize a ausculta em quatro áreas do tórax: tricúspide, pulmonar, aórtica e mitral onde é possível uma melhor distinção dos sons cardíacos. O relato cita que os sopros

podem ser sistólicos ou diastólicos, tendo suas características classificadas por intensidade, frequência, configuração, tempo, qualidade, localização e radiação.

Em sua dissertação, Thompson *et al.* (2018) realiza a avaliação quantitativa e qualitativa de algoritmos de inteligência artificial voltados a detecção de sopros cardíacos. Segundo o estudo, mesmo com avanços recentes na área de processamento de sinais e surgimento algoritmos de classificação de sopros cardíacos, a aceitação deste tipo de tecnologia enfrenta grande resistência da área clínica, muito em função da ausência de dados objetivos relacionados ao desempenho e qualidade deste tipo de tecnologia. O estudo, visando suprir uma demanda por dados de desempenho e qualidade, desenvolve um algoritmo de avaliação objetiva para sistemas inteligência artificial de detecção de sopros cardíacos, caracterizando-se, segundo o conhecimento dos autores, por ser a primeira avaliação abrangente e objetiva deste tipo de algoritmo já relatada na bibliografia. Utilizou-se gravações de sons cardíacos de pacientes atendidos no centro ambulatorial Johns Hopkins obtidos do banco de dados Johns Hopkins *Cardiac Auscultatory Recording Database* (CARD), os dados foram de 1200 pacientes com patologias cardíacas e sem. As gravações tiveram duração de 20 segundos, sendo obtidas com auxílio de um estetoscópio digital em diversas localidades do tórax dos pacientes. Simultaneamente com as gravações realizou-se um eletrocardiograma (ECG). A base de dados também disponibilizou diagnósticos de ecocardiograma, seguido de uma descrição dos sopros cardíacos e de suas características, realizadas por um cardiologista. O algoritmo de detecção de sopros cardíacos avaliado pelo estudo foi desenvolvido pela *CSD Labs*, Áustria, este, constitui-se de um servidor em nuvem, onde envia-se a gravação do sopro e este realiza a análise, os autores esclarecem que este algoritmo não utilizou a base de dados do Johns Hopkins para treinamento do sistema. O algoritmo para mensuração de performance analisou todas as gravações do banco de dados do centro clínico Hopkins em uma única sessão de processamento em lote. As características clínicas e os diagnósticos de ECG associados a cada paciente foram então comparados aos resultados das análises do algoritmo da *CSD Labs*, que recebeu as gravações deste mesmo banco de dados. Os sinais de ECG foram analisados pelo autor visando obter uma frequência cardíaca médio padrão ouro, tendo como objetivo constituir uma métrica de comparação entre esta estimativa e as análises do algoritmo da *CSD Labs*. Analisou-se primeiramente a sensibilidade e especificidade do algoritmo para classificar um sopro como patológico ou normal. Como análise secundária, o autor

incluiu os efeitos da sensibilidade e especificidade da certeza das medidas do algoritmo em função do local de tórax onde foram feitas as gravações, idade do paciente, frequência cardíaca, intensidade do sopro, e diagnóstico patológico primário. Comparou-se a estimativa de frequência cardíaca média padrão ouro com as estimativas realizada pelo algoritmo. O algoritmo de medição de performance conseguiu analisar 89 % das gravações, houve concordância entre o algoritmo em teste e a frequência cardíaca médio padrão ouro obtida pelo ECG em 95 % dos casos. A análise primária do algoritmo apresentou sensibilidade de 93 % e especificidade de 81 % para detecção de casos patológicos, com uma acurácia de 88 %. Quando analisados os locais de ausculta do tórax, foi observada uma precisão maior na borda esternal inferior e superior esquerdas. Ao analisando o desempenho do algoritmo para grupos de idade, o grupo de idade menor que 1 ano apresentou uma alta sensibilidade associada com uma baixa especificidade (98 % e 53% respectivamente com 87% de acurácia), grupos de idade de 1 a 12 anos apresentaram sensibilidade de 95% e 76 %, com 88 % de acurácia respectivamente, enquanto grupos com idade superior a 12 anos contam com 87 % de sensibilidade, 91 % de especificidade e 89 % de acurácia. Analisando a sensibilidade para detecção de sopros patológicos, os autores observaram que os sopros associados a estenose aórtica e defeito no septo ventricular (patologias mais comuns observadas na base de dados) foram detectadas com uma sensibilidade superior a 90 %.

No trabalho de WANG *et al.* (2020), aborda-se o reconhecimento de sopros cardíacos em pacientes com defeito no septo ventricular (VSDs). O defeito no septo ventricular consiste em um tipo de cardiopatia congênita causado por problemas no desenvolvimento do septo interventricular, caracterizando-se por ser o tipo mais comum de má-formação do coração, sendo detectado já no momento do nascimento. Este tipo de cardiopatia ocorre em aproximadamente 2 a 6 vezes a cada 1000 nascimentos, sendo responsável por 30 % dos casos de cardiopatias em crianças e adolescentes. Os sopros cardíacos podem ser auscultados em pacientes com VSD, sendo conhecidos por serem comumente holos sistólicos e tendo origem na turbulência do sangue entre os ventrículos esquerdo e direito. A ausculta cardíaca é convencionalmente utilizada para o diagnóstico de VSD, onde observa-se que a acurácia depende fortemente da experiência do clínico, podendo ser um desafio para clínicos mais novos e, portanto, menos experientes, dada esta necessidade, os autores identificam que uma ferramenta que reconheça de maneira automática os

sopros cardíacos poderia auxiliar no diagnóstico de cardiopatias. O estudo ressalta o aumento gradativo do uso de inteligência artificial em diagnósticos assistidos por computador, exemplificando pelo grande uso de algoritmos de reconhecimento e classificação de imagens médicas que vem sendo desenvolvidas utilizando aprendizagem profunda (deep learning). Houve avanços significativos no uso de Redes Neurais artificiais e Redes neurais artificiais profundas na detecção e classificação de sopros cardíacos, redes neurais convolucionais também já foram utilizadas para identificação de sopros cardíacos. O objetivo do estudo consiste em desenvolver um algoritmo de reconhecimento automático de sopros cardíacos de origem sistólica de pacientes com defeito no septo ventricular, utilizando uma rede neural recorrente convolucional de pooling de tempo temporal (TAP-CRNN). A base de dados utilizada no estudo consiste em gravações de sopros cardíacos de pacientes do Hospital Universitário Nacional do Taiwan, realizadas com auxílio de um estetoscópio eletrônico. Sons de pacientes com diagnóstico de VSD foram categorizados no grupo VSD, enquanto sopros de paciente sem má-formação cardíaca significativa foram categorizados no grupo normal. A ausculta cardíaca nos pacientes foi realizada por um cardiologista com trinta anos de experiência, visando confirmar a ocorrência de um sopro sistólico patológico em pacientes com VSD, uma ecocardiografia também foi realizada nos pacientes para confirmar o diagnóstico da doença. Para cada paciente, as gravações dos sopros cardíacos foram realizadas por enfermeiras treinadas para o estudo, sob supervisão de um cardiologista experiente. O procedimento foi repetido duas vezes, com as gravações realizadas nas áreas aórtica, pulmonar, aórtica secundária (ponto Erb), tricúspide e mitral. Os registros coletados foram divididos em dois grupos, treinamento e teste. O Algoritmo utilizou-se da transformada de Fourier rápida de curta duração para converter os sinais de fonocardiograma (PCG) para representação de tempo-frequência, constituindo diagramas de espectro, o que permite observar a característica de sinais temporais variando sobre o tempo. O estudo utilizou uma TAP-CRNN para classificação, onde a camada convolucional foi utilizada para extrair padrões de deslocamento de frequências invariantes, as camadas recorrentes exploram o padrão temporal global das representações, a camada TAP gera os resultados de classificação, extraíndo os padrões de tempo temporal e atribuindo peso as características do espectro. O estudo contabilizou sons cardíacos de 76 pacientes, 51 com VSD e 25 sem má-formação cardíacas significantes, resultando num total de 525 gravações de pacientes com VSD

e 251 de pacientes do grupo normal. Analisou-se o desempenho do algoritmo de classificação executando procedimento de validação por divisão treinamento-teste e validação cruzada K-fold, modelos de algoritmos de Redes Neurais Convolucionais (CNN) e Redes Neurais Convolucionais Recorrentes CRNN foram validados da mesma forma para comparação com a TAP-CRNN desenvolvida no estudo. Na validação por divisão treinamento-teste, a base de dados foi dividida em 70 % (191 registros normais e 351 registros com sopro sistólico) para treinamento e 30% (60 registros normais e 178 registros de sopro sistólico) para teste da performance. A métrica demonstrou que a TAP-CRNN apresentou uma acurácia superior aos outros modelos, apresentando resultados de sensibilidade e especificidade de 97% e 98 %, respectivamente, contra 88% e 85 % obtidos da CNN, 92% e 93% obtido da CRNN. A validação cruzada K-fold foi utilizada para verificar a confiabilidade da performance do sistema, a base de dados de PCG foi dividida em quatro grupos (K = 4), com proporções iguais entre o grupo normal e o grupo com VSD, os dados de três grupos foram utilizados para treinamento da TAP-CRNN e o quarto grupo utilizado para teste do algoritmo. A sensibilidade e especificidade média do K-fold foi de 97,18 % e 91,98 %, respectivamente, confirmando que a TAP-CRNN tem capacidade para produzir resultados satisfatórios e confiáveis. O estudo avaliou por último a capacidade de reconhecimento dos sopros da TAP-CRNN nas diferentes regiões de ausculta, obtendo 100 % de especificidade e sensibilidade para as regiões da segunda área aórtica e tricúspide, com a sensibilidade decrescendo nas áreas aórtica (95,5 %), pulmonar (94,1 %) e mitral (94,1 %).

O artigo de Tagashira & Nakagawa (2020) visa o estudo a detecção de sons cardíacos anormais e coleta de dados biométricos. A autenticação biométrica, utiliza dados de imagens ou sinais biomédicos como informações biológicas que indicam características individuais. Imagens de digitais dos dedos, palma da mão, forma dos vasos sanguíneos, íris ou face são utilizados em diversas aplicações de autenticação. Estes tipos de sistemas possuem o ônus de eventualmente realizar uma falsa identificação, muitas vezes causadas por dispositivos que imitam a parte do corpo humano em questão. Sinais biológicos, como pulsação, eletrocardiograma (ECG) ou sopros cardíacos proporcionam maior robustez em comparação as métricas mais comuns, visto que são praticamente impossíveis de serem imitados. Estes sistemas mais robustos exigem a aquisição de muitos dados biométricos, juntamente com a constante atualização da base de dados, para fornecer uma identificação individual

confiável e estável. A partir da necessidade de compor uma base de dados considerável, alguns estudos relacionados a autenticação por ECG e pulsação, utilizaram-se dos serviços de saúde para compor e atualizar suas bases de dados. O estudo de (Tagashira & Nakagawa, 2020) identifica que não existem estudos a respeito de sopros cardíacos que façam a relação entre aplicações de gerenciamento de saúde e dados biométricos, o autor enfatiza que esta base de dados de sons cardíacos pode ser eficientemente acumulada ao fazê-la juntamente com o diagnóstico dos sopros cardíacos. Sistemas de gerenciamento de cuidados de saúde são importantes para prevenir doenças relacionadas ao estilo de vida, hábitos alimentares, ambiente de trabalho e falta de exercícios físicos são alguns exemplos de estilos de vida que podem vir a causar doenças, dentre as quais estão as doenças cardíacas. Como doenças causadas por estilo de vida podem permanecer um longo período sem se manifestarem, problemas cardíacos podem progredir silenciosamente. Estes fatores levam o autor a enfatizar que a necessidade de se fazer um monitoramento constante do estado de saúde, de forma que sopros cardíacos, por exemplo, diagnosticados em estágio inicial, podem ser tratados de forma a prevenir o desenvolvimento de um estado de saúde mais severo. Desta forma, o estudo tem como objetivo propor um sistema biométrico de fácil utilização que trabalhe juntamente com sistemas de cuidado de saúde baseados na coleta de dados de sopros cardíacos. Os dados de referência precisam ser periodicamente atualizados para a autenticação biométrica, de forma a propiciar a utilização dos mesmos tanto para autenticação biométrica quanto para detecção de sopros cardíacos anormais. Tagashira & Nakagawa (2020), propõe o cálculo dos padrões dos sons cardíacos a partir da potência temporal do som, a qual é obtida por meio da soma dos componentes do espectro de potência (PS), proveniente da análise de tempo-frequência dos sopros cardíacos. A diferença entre os dados de referência e sopros cardíacos mensurados são avaliadas pela distância de Mahalanobis (MD), pelo método de Mahalanobis-Taguchi (MT). Os dois sons predominantes em um coração saudável, são conhecidos como primeiro e segundo som cardíaco, S1 e S2, respectivamente, onde S1 representa o fechamento das valvas mitral e tricúspide, podendo ser auscultado de forma mais clara na região do Apex, S2 manifesta-se no fechamento das valvas aórtica e pulmonar. No estudo, a ausculta cardíaca foi realizada na região do Apex, onde o sopro pode ser auscultado com mais clareza. O estudo realizou o cálculo de oito padrões (K1 a K8), onde seis deles (K2 a K7) são

utilizados para detectar sons cardíacos anormais e sete (K2 a K8) para realizar a autenticação pessoal. Os padrões de sopros cardíacos foram calculados partindo-se da medida do mesmo em pacientes, esta forma de onda do som obtido é então modificada de forma identificar a repetição periódica do sinal, devendo esta estar situada entre os dois pontos S1, gerando o padrão K1. O padrão K2 constitui-se do intervalo entre S1 e S2. As posições de S1 e S2 foram calculadas pela detecção de picos em uma curva spline cúbica gerada por 100 amostras de dados. Embora alguns pontos de pico tenham sido calculados, o estudo analisou estes pontos partindo do pressuposto que a frequência cardíaca normalmente está na faixa de 50 a 120 batimentos por minuto, logo, assume-se que K1 varia seu intervalo entre 0,5 s e 1,25 s, K2 foi determinada de maneira empírica entre 0,125 s e metade do intervalo de K1. O espectro de Potência é calculado via transformada rápida de Fourier, utilizando janela retangular como função de janela. O estudo faz referência a faixa de frequência onde sopros cardíacos anormais podem ser observados, estando estes na faixa de 20 a 400 Hz, desta forma, ajustou-se a largura da função janela para uma resolução de frequência em torno de 400 Hz. Utilizou-se o método MT para classificar os sopros cardíacos e o método MD para constituir uma escala de medida, um conjunto de características é padronizado e um grupo normal é criado e utilizado como referência para escala de medição do método MT. O estudo primeiramente detectou sopros anormais com dados experimentais e então avaliou-se a autenticação por biométrica, realizou-se a ausculta de dez participantes saudáveis por meio de um estetoscópio digital, 19 casos de tipos de doenças, incluindo sopros cardíacos, foram coletados de um CD, que consistia em gravações de sopros cardíacos anormais, coletados por uma equipe de médicos, como parte de um livro de treinamento. Calculou-se um valor MD onde relacionou-se as gravações de cada paciente com os 19 tipos de casos de doenças. Em seguida, investigou-se o limiar para identificação de sons cardíacos anormais, por meio da relação entre taxa de ocorrência e limiar dos valores de MD. O grau de anormalidade do sopro foi estimado a partir das alterações de MD ao longo do tempo, de forma que monitorando os valores de MD periodicamente, será possível identificar sopros cardíacos com antecedência. A performance de autenticação biométrica foi avaliada com base na relação entre a taxa de falsa rejeição e a taxa de falsa aceitação. O estudo apresentou uma performance de autenticação variando entre uma taxa de 90% e 100 % no reconhecimento correto para os dez participantes das pesquisas, e 100 % para o reconhecimento de sopros cardíacos.

Em Boulares *et al.* (2021), realiza-se a análise e classificação de sopros cardíacos por meio de sinais de fonocardiograma (PCG). Doenças cardiovasculares são uma das maiores causas de óbito do mundo, sendo responsável por 17,3 milhões de mortes por ano, com uma estimativa de crescimento desta taxa para 23,6 milhões até 2030. Consideradas emergências médicas, os problemas cardiovasculares precisam ser detectados o mais rápido possível, de forma que este diagnóstico precoce pode diminuir consideravelmente a condição de insuficiência cardíaca. Diagnósticos de problemas cardiovasculares podem ser realizados utilizando as técnicas tradicionais de ausculta cardíaca baseadas em estetoscópio, fonocardiograma ou ecocardiograma. Usualmente, cardiologistas usam o estetoscópio para auscultar os sons cardíacos do paciente e buscam encontrar sopros cardíacos anormais. Sons cardíacos normais e anormais se diferenciam principalmente pelos seus sinais de PCG, estas diferenças são evidenciadas nas características de tempo, amplitude, intensidade, homogeneidade e conteúdo espectral. De acordo com pesquisas, profissionais com menos experiência e estudantes de medicina apresentam uma acurácia em torno de 20 % a 40 % na identificação de sopros cardíacos, profissionais experientes apresentam uma acurácia em torno de 80% para a mesma métrica, desta forma, os autores identificam que o desenvolvimento de uma solução precisa, acessível e de fácil utilização pode auxiliar significativamente os pacientes no diagnóstico precoce de sopros cardíacos, bem como estabilizar ou curar doenças cardiovasculares já existentes. O estudo objetiva desenvolver um método automático de análise e classificação de sopros cardíacos, sendo utilizado para detectar patologias cardíacas em aplicações clínicas. As principais contribuições do estudo concentram-se na proposta de um modelo de pré-processamento baseado em um filtro de resposta ao impulso infinito (IIR) para eliminação automática de ruído, um método de segmentação automática de ciclo cardíaco (HCS) baseado em detecção de envelope, utilizando decomposição de onda de Daubechie's e uma nova abordagem de seleção de segmentos ciclos cardíacos (HCS), baseada em clusters de características do PCG retransmitidos no modelo de mistura gaussiana (GMM). No estudo, os autores desenvolveram um método de combinação de aprendizagem supervisionada e não supervisionada. O modelo implementado permite o reconhecimento de sons cardíacos normais e anormais. A etapa de pré-processamento foi dividida em quatro passos: tratamento de ruído, segmentação automática do ciclo cardíaco, espectro de imagens de frequência-Mel e

seleção de segmentos por clusters. Na etapa de tratamento de ruído, um filtro IRR foi utilizado para separar o ruído do sinal, na segmentação do ciclo cardíaco primeiramente reduziu-se a frequência de amostragem do sinal para 2 kHz, visto que se tinha uma amostragem inicial de 44.1 kHz e que a maioria dos sinais de sons cardíacos estão na faixa de 25 Hz a 120 Hz, após a redução de amostragem, o sinal foi então normalizado. Realizou-se a detecção de envelope a partir da decomposição de onda de Daubechie's, os sinais de baixa frequência foram obtidos calculando-se o limiar adaptativo, por meio da decomposição de onda em coeficientes. Após, foi executado uma reconstrução de onda para extrair as baixas frequências. Por último, calculou-se a entropia de Shannon média, levando-se em conta os segmentos de sons de baixa frequência e o número de amostras de sinal por segmento. A identificação dos segmentos de sons cardíacos foi realizada a partir do cálculo do ciclo cardíaco baseado na função de correlação, após definido o ciclo cardíaco os componentes do sinal podem ser identificados e segmentados. A técnica de Espectro de imagens com frequência-Mel (MFCC) foi utilizada para extrair características do espectro do PCG a armazenar estes padrões em imagens de formato PNG, para este procedimento ser possível, os autores executaram os seguintes passos: A partir da janela de Hamming, o sinal de PCG foi dividido em pedaços acústicos, a saída deste procedimento é um vetor que representa as características cepstais de cada pedaço, após, aplicou-se a transformada discreta de Fourier (DFT) em cada pedaço. Para cada pedaço de DFT é mantida apenas o logaritmo do espectro da amplitude, a fim de conservar a propriedade de volume do sinal. Obteve-se as características essenciais de frequência por meio de MFCC associada a suavização de espectro e à saída deste procedimento, aplicou-se a transformada discreta de cosseno, de forma a obter as características MFCC do sinal PCG. Na etapa de seleção de segmentos por clusters (grupos), buscou-se eliminar segmentos indesejados que poderiam influenciar no resultado de reconhecimento dos sons. Os autores utilizaram Redes neurais convolucionais (CNN) para classificação dos sopros cardíacos, devido à falta de uma base de dados de treinamento de sinais PCG considerável, a técnica de aprendizagem da (CNN) foi adaptada, por meio do ajuste de dos modelos CNN pré-treinados existentes, cujos já foram treinados pela base de dados ImageNet. As camadas convolucionais pré-treinadas foram conservadas e utilizadas para extração de características, para realizar o ajuste da técnica de aprendizagem foram adicionadas quatro camadas. O estudo contou com as bases de dados Pascal e *Physionet*, a base de dados *Physionet*

contém amostras de 665 e 2575 sopros normais e anormais, respectivamente, estas gravações contêm uma faixa de duração de 8s a 40 s. Com relação a base de dados pascal, foram utilizados dados de dois grupos de treinamento (A e B), cada grupo contendo classes de sopros normais, anormais e extrassístole, com um total de, 231, 129 e 65 amostras, respectivamente para cada classe. Após passar pelas etapas de pré-processamento, segmentação e seleção de segmentos, os sinais PCG são transformados em imagens PNG e treinados pelo modelo de CNN. Após executar os procedimentos de pré-processamento, o estudo testou 17 tipos de modelos de CNN do Keras com e sem o uso do procedimento de seleção de segmentos, utilizando a base de dados Pascal, contendo as três classes (normal, anormal e extrassístole), CNNs que não utilizaram seleção de segmentos apresentaram como melhor acurácia de validação média obtida um valor de 0,81, que ocorreu para dois modelos de CNN e a melhor taxa de verdadeiros positivos de 0,73 que ocorreu em somente um modelo de CNN, as mesmas métricas foram feitas com CNNs dotadas do procedimento de seleção de segmentos, obtendo os melhores valores de 0,87 de acurácia e 0,83 de taxa de verdadeiros positivos (TPR) e 0,81 de precisão, ambos valores ocorreram somente para um modelo. O estudo também experimentou a base de dados *Physionet*, contendo duas classes (Normal e anormal), também foram testados 17 modelos de CNNs, apresentando a melhor acurácia de 0,97, TPR de 0,944, precisão de 0,944 e especificidade de 0,946.

Em seu estudo, Begic *et al.*, (2021) aborda a classificação de sopros cardíacos originados por cardiopatias congênitas, por meio de técnicas de *machine learning*. Estima-se que 15 milhões de bebês e crianças de países de baixa e média renda, anualmente morrem ou experimentam algum tipo de complicação originada por doenças cardíacas que teriam possibilidade de prevenção ou tratamento, cardiopatias congênitas (CHD) constituem-se como a maior causa de mortes entre crianças e bebês. A sobrevivência de pacientes com cardiopatia congênita depende da severidade do caso, mas também do diagnóstico e tratamento precoces. Falta de infraestrutura de sistemas de saúde, número reduzido de especialistas e a habilidade do profissional para realizar o diagnóstico são alguns dos fatores que justificam um alto índice de mortalidade em países de baixa e média renda. Sopros cardíacos podem ser auscultado com estetoscópio em crianças e adolescentes 50 % dos casos, estes sopros podem ser associados a CHD, porém na maioria dos casos representam sopros normais para a faixa etária. Em muitos casos, médicos do atendimento de

saúde primário ou pediatras encontram dificuldades para diferenciar sopros normais e anormais, acabando por encaminhar o paciente para o atendimento especializado. Estima-se que CHDs possuem uma incidência 4/1000 a 50/1000 casos em nascidos vivos, o que justifica a importância de exames clínicos adequados e a importância desta atividade no atendimento de saúde primário. Sopros cardíacos interpretados de maneira incorreta podem vir a serem fatais para o paciente, o que justifica uma grande demanda por sistemas que sejam capazes de classificar e reconhecer sopros relacionadas a CHDs. A utilização destes sistemas pode vir a aumentar a acurácia de diagnósticos no atendimento primário e por consequência diminuir a demanda pelo atendimento de um especialista, de forma a aliviar a demanda do sistema de saúde como um todo. O objetivo do estudo trata-se de determinar quais parâmetros dos exames físicos dos pacientes do atendimento primário tem maior impacto para a classificação de sopros cardíacos. O artigo apresenta resultados do desenvolvimento de um sistema de classificação baseado em cinco algoritmos de *machine learning*. A pesquisa baseou-se em dados coletados da clínica pediátrica do centro clínico universitário de Saravejo, um total de 116 crianças com idade entre 1 e 15 anos foram envolvidas no estudo, a base de dados dividiu-se em 56,9 % de pacientes do sexo masculino e 43,1 % do sexo feminino, a idade média do grupo foi de 7 anos. As informações incluídas no estudo são provenientes de exames previamente realizados por especialistas experientes, onde os pacientes foram submetidos as seguintes etapas: obtenção de informação do paciente, histórico clínico, informação a respeito dos hábitos da mãe durante a gravidez, realização de exames no paciente utilizando os métodos clínicos: exame físico, ausculta, fonocardiografia, exames laboratoriais, raio-x e eletrocardiograma (ECG). Todos os pacientes examinados foram diagnosticados pelos especialistas quanto a presença de sopros cardíacos inocentes ou sopros relacionados a CHDs, destes, 67,2% tiveram sopros inocentes e 32,8 % sopros que sugeriam presença de CHDs. O desenvolvimento do classificador por machine learning consistiu inicialmente de 68 entradas (informações obtidas dos pacientes) e um parâmetro de saída, que pode assumir os valores de sopro inocente ou anormal. O desenvolvimento do classificador foi dividido em duas etapas: análise dos dados de exames médicos e classificação, na primeira etapa foram aplicados métodos de machine learning para extração de características, os seguintes algoritmos de aplicação de pesos foram implementados: InfoGain, GainRatio, Reliefe e Correlação. Na segunda etapa, seguindo a prática de desenvolvimento de modelos

preditivos em biomedicina, aplicou-se diferentes técnicas de machine learning para classificação dos sopros cardíacos. O estudo montou cinco grupos, um grupo contando com os dados originais e os outros quatro resultantes da aplicação da seleção de características em cada um dos modelos de machine learning, estes cinco grupos constituem como entradas para o classificador de sopros e são divididos em 70 % dados de treinamento de 30 % dados de validação. O classificador foi proposto a partir da utilização dos seguintes algoritmos de machine learning: Naive Bayes (NB), Regressão Logística (LR), Árvore de decisão (AD), Random Forest (RF) e máquina de vetores de suporte (SVM). A performance do sistema foi avaliada a partir da acurácia, calculada com base nos valores de verdadeiros positivos, verdadeiros negativos, falsos positivos e falso negativos. O classificador de sopros cardíacos, apresentou sua performance de validação mais baixa quando se utilizou o algoritmo de Naive Bayes, representando uma acurácia máxima de 40,1 %, a melhor performance de validação foi obtida pela máquina de vetores de suporte, alcançando 97,1 % de acurácia, porém mostrou uma habilidade maior para reconhecer sopros inocentes do que sopros anormais, principalmente por conta do tamanho reduzido dos dados. O estudo selecionou os dados mais relevantes dentro de cada um dos grupos de: exame físico, ausculta, histórico clínico, eletrocardiograma e raio-x para formar um sistema especialista com 17 entradas formadas por estes dados, obteve-se uma acurácia de 94,28 % para o classificador de sopros cardíacos nesta abordagem.

Em Li *et.al* (2020), discute-se a doença arterial coronariana (DAC), esta é a principal doença causadora de óbitos em seres humanos. A maioria dos métodos de detecção da doença se constituem por métodos invasivos, de alto custo ou métodos que não possibilitam o diagnóstico precoce. Alterações nos sons cardíacos e em sua forma geralmente são os primeiros indícios de doenças cardíacas, alguns estudos provaram que os sopros turbulentos diastólicos são características da doença coronária, porém existem doenças relacionadas a válvulas cardíacas que também podem ser identificadas pelos sopros diastólicos, desta forma, os autores identificam a necessidade de se realizar a diferenciação entre sopros cardíacos originados por valvopatias e DAC. O método proposto para distinção entre sopros valvulares e DAC, inclui as etapas de: amostragem, normalização, retirada de ruído, segmentação, decomposição do som cardíaco por meio de transformada wavelet (EWT) empírica, extração de características e classificação do sopro, utilizou-se gravações de sopros cardíacos coletados da base de dados de Michigan. Os sinais de sopro cardíaco foram

decompostos em faixas de frequência, onde, analisando o espectro de cada faixa, os autores puderam identificar que os sinais de DAC estavam concentrados em sua maioria em frequências de 300 Hz a 400 Hz, enquanto sopros relacionados a valvopatias, concentravam-se em frequências entre 200 Hz e 250 Hz. O estudo obteve uma sensibilidade de 94,7 %, especificidade de 93,4 % e acurácia de 88,7 % para identificação da DAC, quando o objetivo foi de identificar cardiopatias, obteve-se uma sensibilidade de 93,3 %, especificidade de 94,6 %, e acurácia de 88,6 %.

Em sua pesquisa, Ahmad et al. (2019) disserta sobre as doenças cardíacas, estas são uma das líderes globais no quesito mortalidade, tendo uma tendência de aumento de mortalidade até o ano de 2030. A detecção precoce de problemas cardíacos torna-se vital e é realizada por meio de ecocardiograma ou pela ausculta de sopros cardíacos realizadas por médicos que necessitam de uma extensiva experiência e treinamento para identificar sons cardíacos anormais. Verificando esta problemática, o estudo tem como objetivo propor um sistema de *machine learning* baseado no reconhecimento dos sopros cardíacos e classificação. A base de dados do estudo é constituída de amostras de sopros cardíacos de pacientes do departamento de cardiologia do Hospital Escola Ayub, 283 sons foram coletados com um estetoscópio eletrônico, sendo 175 sopros normais e 108 anormais, onde contabilizou-se quatro tipos de sopros anormais diferentes. Todas as amostras de sopros foram diagnósticas por um especialista e validadas por um exame de ecocardiograma. O sinal do som cardíaco passou pelos processos de pré-processamento, onde o sinal foi passado por um filtro de remoção de ruídos, seguido da segmentação do sinal em componentes, por meio de envelopamento e utilizando as características no domínio do tempo do sinal de fonocardiograma (PCG), para este processo, utilizou-se a energia de Shannon média normalizada. Após a etapa de segmentação, realizou-se a extração de características do sinal, 50 coeficientes de frequência Mel Cepstrais (MFCC) baseados em características, foram extraídos utilizando os intervalos de sístole e diástole. Um vetor de característica do MFCC foi utilizado para treinamento de sete tipos de máquinas de vetores de suporte (SVM) e *K-Nearest Neighbors* (KNN) que atuaram na classificação e detecção de doenças cardíacas. Para treinamento e validação dos modelos de classificador utilizou-se os métodos de validação cruzada quádrupla e validação *holdout*, no primeiro, as amostras de dados foram divididas aleatoriamente em cinco grupos, o classificador é treinado por quatro conjuntos e validado pelo quinto conjunto, no segundo método, dividiu-se

os dados de maneira aleatória com 80% para treinamento e 20 % para validação. O melhor resultado de acurácia do classificador foi de 92,6 %, utilizando uma SVM de média gaussiana, para utilização de algoritmos KNN, obteve-se 88,3 % como melhor resultado, utilizando uma KNN de aplicação de pesos.

O estudo de Ho *et.al.* (2018) tem como abordagem principal doenças cardíacas originadas por sopros e sua detecção. A gravação de sons de válvulas cardíacas ou por fonocardiograma, realizadas por profissionais de pouca experiência e pouco treinamento, associada a uma baixa qualidade dos dispositivos utilizados para este fim, frequentemente traz consigo problemas relacionados a baixa sensibilidade do equipamento e posicionamento incorreto do equipamento por parte do profissional, podendo resultar em diagnósticos equivocados. Os Smartphones vem sendo utilizados em larga escala para diversas aplicações de monitoração de saúde pessoal, os autores atentam-se para importância de se realizar um monitoramento cardiovascular frequente e não invasivo e para a possibilidade da realização deste tipo de atividade com auxílio de smartphones. Verificando esta problemática, o estudo tem como objetivo desenvolver um algoritmo efetivo para detecção de anomalias cardíacas fisiológicas, a partir da aquisição de sons cardíacos por meio de smartphones. Realizou-se a coleta de sinais de fonocardiograma (PCG) em diversos pontos do tórax dos pacientes, por meio de um *headfone* e com fones de ouvido conectados ao smartfone, para comparação dos dois métodos. As características dos sons cardíacos foram extraídas por meio da análise dos componentes de alta frequência na escala de tempo por meio da transformada *wavelet*. Os experimentos do classificador foram realizados utilizando a base de dados *Physionet*, consistindo em 3153 gravações de sopros cardíacos de pacientes, com 2488 sopros normais e 665 sopros anormais, os ruídos proeminentes foram suprimidos por meio de um algoritmo de limiar, foi realizado a extração de características de todas as gravações da base de dados, sendo os resultados destes processos armazenados em uma base de dados para treinamento e validação. O estudo realizou a extração de quatro características do sinal PCG para cada paciente, sendo estas: pressão sanguínea (PS), relacionada com os picos de amplitude do primeiro e segundo som de ausculta (S1 e S2), valores de entropia dos picos de amplitude periódicos para os componentes de baixa frequência, soma dos picos de amplitude e frequência dominante do acúmulo dos componentes de alta frequência. Antes de realizar o teste do classificador, estimou-se a distribuição de probabilidades a priori bayesiana de cada característica. O algoritmo de AdaBoost foi

utilizado para treinamento e validação da base de dados, atingindo uma acurácia próxima a 100 % para o diagnóstico de doenças cardíacas.

O artigo de Daoud *et al.* (2017) tem como objeto de estudo sinais de fonocardiograma (PCG) e a detecção e classificação de patologias cardíacas com base nestes sinais. Um PCG é composto principalmente por dois sons característicos, principais, S1 e S2. Observa-se a presença de sons cardíacos anormais principalmente por influência de disfunções nas valvas cardíacas, desta forma, os autores observam para a utilidade da análise dos sinais PCG como complemento a ausculta cardíaca, de forma a proporcionar uma melhora significativa no diagnóstico de anomalias cardíacas, visto que a maior problemática da análise de sinais PCG está na detecção dos sons S1 e S2 e em ruídos no sinal, o que muitas vezes compromete a análise. O estudo utilizou a transformada discreta *wavelet* para realizar a decomposição do sinal de PCG em diferentes faixas de frequências, para discriminar altas e baixas frequências, estimou-se a entropia de Renyi (RE) para os coeficientes de detalhe, um limiar destes coeficientes foi aplicado a fim de distinguir os componentes S1 e S2 do sinal PCG e o sopro cardíaco, permitindo a detenção de sons anormais. A partir da caracterização dos resultados de detecção no domínio frequência-tempo, foi possível extrair características dos sopros anormais, como faixa de frequência, frequência de picos e tempo de duração destes sons anormais. Validou-se o método a partir da utilização de sinais de PCG das patologias cardíacas: estenose aórtica, regurgitação mitral e aórtica, estalo de abertura e estenose pulmonar, onde verificou-se faixas os faixas de frequência e duração do sopro característicos de cada doença, onde os autores verificaram que os valores obtidos para estas métricas iam ao encontro de outros trabalhos referenciados no estudo, considerando que a técnica apresentada apresenta bons resultados.

Em Wang (2020) discute-se o diagnóstico de doenças cardíacas congênitas em crianças, a partir dos sinais de fonocardiograma (PCG). A ausculta cardíaca constitui-se com uma importante ferramenta no diagnóstico precoce de doenças cardíacas. Estes sons cardíacos podem ser gravados por um estetoscópio eletrônico, sendo possível a geração do fonocardiograma, onde é possível visualizar as ondas acústicas do som auscultado. Existem dois sons principais (S1 e S2), gerados pelo fechamento das valvas cardíacas, de forma que sons considerados normais não apresentam sopros entre S1 e S2, ao passo que sons patológicos apresentam sopros, na forma de ruído entre S1 e S2. Os autores observam a escassez de estudos relacionado a

interpretação de PCG para crianças, bem como atentam-se ao fato de uma expressiva taxa de incidência de doenças cardíacas congênitas (CHD) presente para esta faixa etária. Dada esta problemática, os autores objetivam o desenvolvimento de um método de diagnóstico inteligente para sopros cardíacos em crianças com doenças cardíacas congênitas. Utilizou-se um estetoscópio eletrônico para realizar as gravações de sons cardíacos de 86 pacientes pediátricos com idade de 4 meses a 16 anos, sendo 24 pacientes com sopros normais e 64 com sopros originados por CHD, estes diagnósticos foram confirmados por um cardiologista por meio de ecocardiografia. Em seguida os sinais de PCG das gravações passaram pela etapa de normalização e retirada de ruídos do sinal, reduziu-se a frequência de amostragem do sinal com posterior normalização, a retirada de ruído foi realizada a partir da transformada discreta wavelet. Na etapa de segmentação do sinal PCG, a partir do sinal foi decomposto em coeficientes de detalhe aproximados a partir da transformada wavelet discreta e de uma função Daubechies 6. Associou-se esta decomposição o produto Hadamard para obter-se o sinal recombinação para baixas frequências, onde encontram-se os sons S1 e S2, a energia de Shannon média normalizada foi utilizada para o envelope do sinal recombinação, os picos deste envelope foram detectados, baseados na alteração de magnitude em pontos de amostras consecutivos do envelope, o instante de ocorrência destes picos tornam-se candidatos a serem pontos de S1 e S2, o intervalo de duração da sístole e diástole também foi utilizado para determinação destes pontos. Extraíu-se do PCG um total de 10 características relacionadas a CHD, estas características foram utilizadas como entradas para 86 redes neurais artificiais com uma estrutura de rede de três camadas, a saída de cada rede neural tem como resultado a classificação dos sopros como normal ou anormal. O estudo obteve como resultado para o diagnóstico de sopro cardíaco uma acurácia de 93 %, associada a uma sensibilidade de 93,5 % e uma especificidade de 91,7 %.

O relato de Tiwari *et al.* (2021) trata da análise e classificação de sons cardíacos por meio de sinais de fonocardiograma (PCG). Estima-se que doenças cardíacas são responsáveis por 31% do total de mortes em um ano, o que torna as doenças cardíacas como líderes no quesito mortalidade, sendo que a maioria destas mortes ocorrem em países de baixa e média renda, onde muitas vezes tem-se uma assistência médica de menor qualidade e de elevado custo. Nota-se que a única maneira de promover a diminuição de letalidade de doenças cardíacas está no diagnóstico em estágio inicial. Os autores propuseram espectrogramas baseados em

Coeficientes de frequência Mel cepstral, transformada constante-Q (CQT), transformada variável-Q (VQT) e transformada constante-Q híbrida (HCQT) para realizar a classificação dos sinais PCG. O resultado do classificador é utilizado para treinamento de um modelo de rede neural convolucional. A base de dados utilizada no estudo consistiu em dados originados do desafio PASCAL de classificação de sons cardíacos, consistindo em gravações de sons cardíacos em ambiente clínico a partir da utilização de um estetoscópio eletrônico. Os dados são divididos em dois grupos (A e B), no grupo A constam quatro classes de sons, nomeadas de normal, som cardíaco extra, sopro cardíaco e artefato, o grupo B conta com as classes normal, extrassístole e sopro. A base de dados unificada em uma única base de dados com cinco classes, Estes registros foram divididos em 80% para treinamento e 20% para validação. O estudo realizou a classificação de diversas classes de sons cardíacos obtendo uma acurácia de 96% para o sistema proposto.

Chowdhury *et al.* (2019) relata sobre os sons cardíacos bem como o diagnóstico destes com base no fonocardiograma (PCG). Doenças cardiovasculares são umas das principais causas de morte no mundo, sendo que em países menos desenvolvidos, uma maior parte da população acaba desenvolvendo este tipo de doença. A ausculta cardíaca constitui-se de um dos exames básicos mais utilizados para a detecção de patologias cardíacas, porém a realização exige uma grande sensibilidade do equipamento à ruídos e a habilidade e experiência do médico em identificar possível anomalias cardíacas. Observando a crescente necessidade por sistemas de detecção precoce de doenças cardíacas, os autores propõem um modelo de protótipo de um estetoscópio inteligente, a fim de monitorar os sons cardíacos do paciente e diagnosticar em tempo real a presença de alguma anormalidade. O sistema é composto de dois subsistemas, um estetoscópio digital portátil e um sistema de tomada de decisão alocado em um computador. Estes dois subsistemas comunicam-se utilizando Bluetooth Low Energy. O estetoscópio utilizado trata-se de um modelo tradicional, onde foi adaptado um microfone. O estetoscópio tem a função de fazer a aquisição do sinal acústico, que é amplificado e filtrado por meio de um circuito eletrônico seguido de digitalização e por último transmissão destes dados para o sistema de tomada de decisão. Foram utilizados dados de sons cardíacos do desafio *Physionet* 2016, onde constam sons cardíacos classificados como normais e anormais, estes dados foram utilizados para treinamento e validação de algoritmos de aprendizagem da máquina, onde tinha-se como objetivo identificar o algoritmo com

melhor performance e os parâmetros ótimos para garantir o melhor desempenho do sistema. Os sons cardíacos da base de dados passaram por uma etapa de pré-processamento, onde filtrou-se ruídos e picos do sinal das gravações, realizou-se a segmentação dos dados, os autores observaram que o som cardíaco S1 está localizado nas regiões de maiores picos do PCG, desta forma, o ciclo cardíaco foi obtido como o intervalo entre pontos de S1, segmentando o sinal a estes intervalos. Extraíu-se característica do sinal por meio de domínio de tempo, domínio da frequência e coeficientes de frequência Mel cepstral, resultado num total de 27 características extraídas do PCG. A base de dados foi dividida em 80 % para treinamento e 20 % para validação, foram treinados diversos algoritmos de *machine learning* que tiveram sua performance comparada. Utilizou-se a análise de componentes *Neighborhood* para determinar a características mais determinantes a serem consideradas pelo classificador, de um total de 27 características, 15 foram selecionadas com maior contribuição na classificação. Os algoritmos de *machine learning* foram treinados novamente considerando o número menor de características do sinal a ser considerado. Os autores obtiveram uma acurácia de 97 % para classificação de sopros anormais e 88 % de acurácia para classificação de sopros normais.

Em Takahashi *et.al* (2021), estuda-se as doenças cardíacas congênicas e a detecção dos mesmo por meio dos sons cardíacos. As doenças cardíacas congênicas (CHD) constituem-se como umas das anomalias mais comuns, ocorrendo em aproximadamente 1% dos nascimentos. Frequentemente, estas anomalias existem no período neonatal sem manifestar-se por meio de sopros cardíacos ou outros sinais clínicos, aproximadamente, um quarto das crianças com CHD são diagnosticadas em exames de rotina. A ausculta cardíaca constitui-se como um dos exames cardíacos mais tradicionais, porém possui o ônus de depender das habilidades e experiência do médico clínico. Os autores observam a existência de uma demanda por uma ferramenta simples que possa ser utilizada por clínicos para detectar sopros cardíacos sem a necessidade de manejo da posição do paciente, visto que em bebês, este manuseio pode vir a causar mudanças abruptas na pressão arterial do paciente. O estudo tem como objetivo verificar a capacidade de detecção de sopros cardíacos em pacientes com CHD, por meio de sensores piezoelétricos. O estudo utilizou nove crianças no período neonatal com CHD sistólicos e nove crianças da mesma faixa etária sem patologias cardíacas. A ausculta foi realizada por um médico clínico

experiente, seguido de um exame de ecocardiograma, utilizado para confirmar os diagnósticos realizados pelos clínicos. O sensor piezoelétrico (PS) foi desenvolvido para detectar as vibrações mecânica causada pelos batimentos cardíacos do paciente, foram realizados registros dos sons cardíacos utilizando PS, quanto com um estetoscópio eletrônico para futura comparação. O sinal adquirido pelo OS passou por digitalização e filtragem por meio de um filtro passa altas, de maneira a possibilitar a visualização dos sons cardíacos S1 e S2. Prosseguiu-se com a segmentação do sinal, a partir da utilização de algoritmos de extração de envelope, o estudo utilizou a transformada Hilbert para esta função. A amplitude dos sons cardíacos sistólicos foi avaliada utilizando signal-to-noise, foram comparados os pacientes saudáveis com paciente com CHD. Por meio do método de correlação de Spearman, determinou-se a correlação entre as amplitudes dos sons cardíacos registradas pelo PS e pelo estetoscópio eletrônico. O estudo verificou que a amplitude dos sons cardíacos detectados pelo PS nos casos de CHD foi significativamente superior a pacientes sem patologia. Os sons cardíacos sistólicos registrados pelo sensor demonstraram uma alta correlação com o estetoscópio eletrônico gerando como resultado uma correlação de 0,899, com uma significância menor que 0,01.

O artigo de Gersh *et al.* (2021), tem como objeto de estudo o diagnóstico clínico diferencial entre sons cardíacos normais e patológicos. A ausculta cardíaca consiste em uma das ferramentas mais utilizadas para detectar alterações na anatomia e fisiologia cardiovascular, o correto manuseio do exame possui grande influência para obter-se um diagnóstico mais acurado. O estudo tem como objetivo realizar uma revisão sobre as manobras mais utilizadas durante a ausculta cardíaca. O exame padrão para indicar a causa de um sopro cardíaco consistente no ecocardiograma, porém este geralmente é indicado em caso de diagnóstico e avaliação de paciente com suspeita ou diagnóstico confirmado de algum tipo de valvopatia. Observa-se que durante a inspiração ocorrem alterações no volume dos ventrículos, causados pelo retorno venoso no mesmo, essas alterações resultam em mudanças na intensidade dos sons cardíacos. O posicionamento do paciente da posição supinada para ortostática para resulta em diminuição do retorno venoso e conseqüente redução de volume dos ventrículos, além da possibilidade de redução da pressão arterial e da frequência cardíaca, estas mudanças hemodinâmicas podem ser associadas valvopatias como: estenose pulmonar e aórtica, regurgitação mitral e tricúspide. Já o agachamento causa aumento no retorno venoso e possível aumento da pressão

arterial. Este tipo de manobra pode apresentar indicativos de patologias como: regurgitação mitral, prolapso da valva mitral e tetrologia de Fallot. A Manobra de Valsalva frequentemente utilizada para realizar o diagnóstico diferencial, induz uma diminuição na intensidade de sopros de estenose aórtica e um aumento na intensidade de sopros originados de cardiopatia hipertrófica. A manobra de aperto de mão é utilizada para diagnóstico diferencial entre os sopros sistólicos originados por estenose aórtica ou sopros originados por regurgitação da valva mitral.

Em seu trabalho, Meyer (2021) tem como proposta a revisão de tópicos da ausculta de sons cardíacos. A ausculta cardíaca tem como objetivo caracterizar sons cardíacos e sopros cardíacos, desta forma o estudo objetiva revisar a classificação dos tipos de sons cardíacos possíveis de serem auscultados. Os sons cardíacos são classificados em duas categorias maiores: sons de alta e baixa frequência. Os sons de alta frequência originam-se da abertura ou fechamento das valvas cardíacas, onde inclui-se os sons de fechamentos das valvas mitral e tricúspide e sons de fechamento da valva aórtica e pulmonar. As baixas frequências, incluem os sons cardíacos S3 e S4. Verifica-se que a intensidade do primeiro som cardíaco (S1), auxilia na avaliação da função e da hemodinâmica do ventrículo esquerdo. Uma divisão ampla constante do segundo som cardíaco (S2) constitui-se como um indicativo de defeito no septo atrial. Sons de galope em S3, associados a ausência de regurgitação mitral, normalmente indicam uma elevada pressão diastólica final no ventrículo esquerdo. O som S4 pode ser frequentemente observado em pacientes com patologias como: isquemia miocárdica aguda e hipertrofia do ventrículo esquerdo. O relato discute sobre a utilização de estetoscópios para realização da ausculta cardíaca, onde verifica-se a aderência dos clínicos a utilização de estetoscópio digitais, que possuem recursos como processamento de sinais e atenuação de ruídos que auxiliam em uma qualidade melhor de ausculta. Observe-se que este tipo de estetoscópio varia em sua capacidade de capturar sons cardíacos relevantes e em sua capacidade de rejeitar ruídos de ambiente, assim sendo necessário que os clínicos levem estas informações em conta no momento de escolher o modelo do equipamento.

O estudo de Oliveira *et al.* (2021) tem como escopo a ausculta cardíaca pediátrica. A doenças cardíacas constituem-se como principal causa de morte no mundo, sendo responsáveis por 31% das mortes globais. A ausculta cardíaca constitui-se como uma das técnicas mais utilizadas para identificação de patologias cardíacas no mundo, aplicações de decisões assistidas por computador podem auxiliar

os clínicos em suas decisões de diagnósticos. Os autores identificam que as aplicações existentes para este propósito têm a limitação de fornecer uma resposta binária (normal ou anormal) para o som cardíaco, observado a partir do fonocardiograma (PCG). Desta forma, identificando uma demanda por um banco de dados que disponibilize de forma pública descrições mais detalhadas sobre sons cardíacos anormais, o artigo propõe a confecção de um banco de dados de sons cardíacos pediátricos. Realizou-se a coleta de 5282 gravações de sons cardíacos de um total de 1568 paciente. Um especialista realizou a descrição manual das gravações com relação a tempo, forma, tom, classificação e qualidade, todos os dados foram coletados com auxílio de um estetoscópio digital. Realizou-se a segmentação dos áudios coletados de forma a identificar os sons fundamentais S1 e S2. Dois especialistas realizaram a inspeção dos resultados de segmentação do áudio de forma a realizar a correção dos mesmo em caso de segmentação incorreta dos sons. Os áudios com as corretas segmentações foram então salvos na base de dados. O estudo espera que a base de dados seja uma ferramenta de grande utilidade para futuros trabalhos que necessitem de uma base de dados mais ampla e detalhada.

A dissertação de Chowdhury *et al.* (2020) aborda o fonocardiograma (PCG) e suas aplicações no diagnóstico de sons cardíacos. O exame de PCG fornece informações importantes a respeito dos sons cardíacos, tais como tempo de duração, frequência cardíaca, entre outros parâmetros relevantes para determinar a função e a condição de funcionamento das valvas cardíacas. O estudo observa que a identificação de sons cardíacos patológicos somente pela ausculta cardíaca torna-se uma tarefa de maior nível de dificuldade, visto que exige habilidade e experiência do clínico para identificar eventuais sopros cardíacos. Desta forma os autores atentam-se para uma carência por ferramentas de análise dos sinais de PCG que auxiliem na análise e identificação de sons cardíacos anormais. Os autores objetivam combinar técnicas de processamento de sinal e deep learning para realizar o tratamento, compressão, segmentação e classificação de sinais PCG de forma eficiente. Para o tratamento dos sinais, realizou-se a remoção de ruído seguido de compressão dos dados por meio de técnicas baseadas na transformada wavelet discreta. Um algoritmo de segmentação de dados foi aplicado, utilizando o envelope de energia de Shannon e detecção de passagem por zero, de forma a segmentar o sinal em: S1, intervalo da sístole, S2 e intervalo da diástole. Uma etapa de retirada de informações do sinal foi realizada por meio de um diagrama de espectro de potência em escala Mel e

coeficientes de frequência Mel cepstral. A etapa de classificação dos sons foi realizada com base em deep learning, onde um modelo de rede neural profunda feed-forward de 5 camadas foi utilizada. O estudo obteve uma acurácia global de 97,1 %.

A partir da revisão da bibliografia a respeito dos sons cardíacos, pela abordagem da área saúde, e por meio da compreensão da origem, características e forma de diagnóstico do sons cardíacos, verifica-se a necessidade da uma identificação de patologias cardíacas mais acurada, visto que a detecção de anomalias por meio ausculta cardíaca exige treinamento e experiência do profissional, recurso este que muitas vezes não está presente no atendimento primário de saúde, impedindo muitas vezes um diagnóstico precoce.

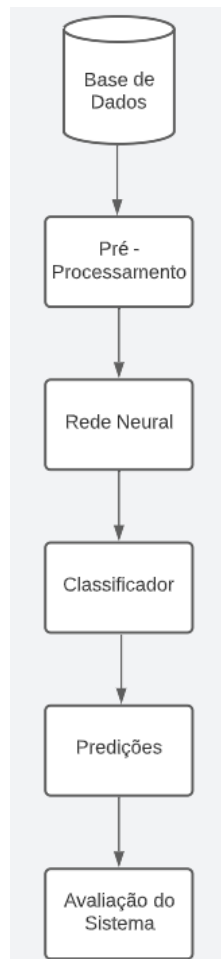
Pela análise das dissertações da área computacional, verificou-se uma grande quantidade de modelos de *machine learning* implementados tanto para classificação de sons cardíacos quanto a identificação de outras patologias cardíacas por meio destes sons. As ferramentas de pré-processamento relatadas nas dissertações, como retirada de ruídos, segmentação, identificação de características de sinal de áudio, constituem-se como referências importantes para este estudo, visto que estas ferramentas podem ser utilizadas no sistema proposto.

As redes neurais foram amplamente citadas nos estudos, para modelar sistemas de reconhecimentos dos padrões dos sons cardíacos, nestas propostas também foi possível atentar-se para a abordagem dos autores quanto ao treinamento e validação das redes neurais e divisão da base de dados para desempenho destas tarefas. Os relatos também garantem uma grande contribuição no que diz respeito a base de dados, diversos autores citaram as fontes destas bases, o que é de grande utilidade para este estudo, visto que estes dados também podem ser utilizados para o desenvolvimento do sistema proposto por esta dissertação.

3 METODOLOGIA

Os tópicos a seguir descrevem a metodologia aplicada no desenvolvimento deste trabalho, bem como os detalhes de cada etapa do processo. Inicialmente apresenta-se o processo de obtenção do banco de dados, com gravações reais de sopros cardíacos. A partir desta etapa, constrói-se o algoritmo computacional responsável por sugerir diagnósticos de sopros cardíacos, iniciando pela etapa de pré-processamento dos dados, seguido de aplicação ao sistema classificador de sons cardíacos. De posse do modelo de classificador, inicia-se a etapa de predições de novo sons cardíacos, seguido pela avaliação do desempenho do sistema quanto a classificação de pacientes com sons cardíacos normais e anormais. O fluxograma da figura 22 ilustra as etapas básicas da metodologia utilizada.

Figura 20 – Visão Geral do Sistema



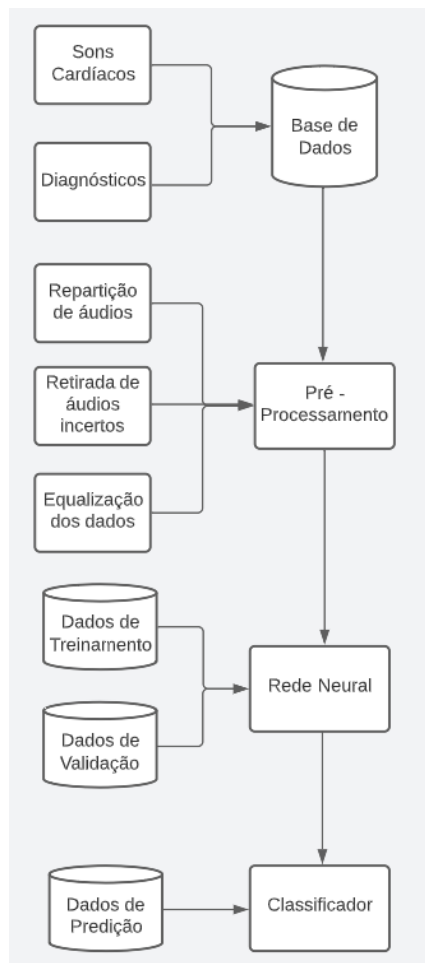
Fonte: Elaborado pelo autor.

A partir do diagrama da figura 22 é possível identificar, em linhas gerais as etapas a serem seguidas para confecção do sistema classificador. As etapas de modelagem do sistema classificador e pré-processamento, demandam maior poder computacional, devido a utilização de algoritmos que implementam as técnicas necessária para realização de cada etapa. Na avaliação do sistema, novamente exige-se poder computacional para implementação de algoritmos que sejam capazes de implementar técnicas para cálculo da eficácia do sistema quanto a classificação dos sons cardíacos utilizados para validação e predição.

3.1 Sistema Proposto

A figura 23 sintetiza o sistema classificador de sons cardíacos proposto.

Figura 21 – Sistema Proposto



Fonte: Elaborado pelo autor.

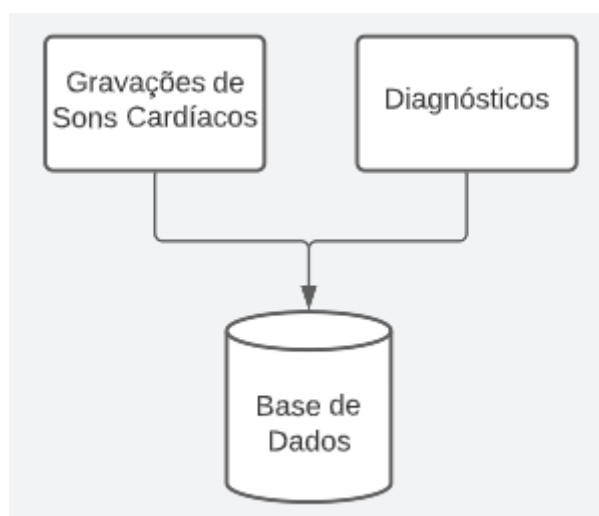
O sistema proposto é constituído pelo classificador, sendo este treinado e validado a partir dos sons cardíacos da base de dados. O conjunto de dados utilizado possui gravações de sons cardíacos, classificadas como normais, anormais e incertos, sendo este último representado por áudios que possuem grande quantidade de interferências, não sendo possível considerar como uma gravação confiável. Visando garantir confiabilidade ao sistema, os áudios classificados como incertos foram descartados. O conjunto de dados possui gravações com duração desde 5 segundos até alguns minutos, a fim de padronizar a duração dos áudios utilizados para treinamento do sistema, bem como promover um aumento no número de gravações disponíveis para treinamento, todas as gravações da base de dados original foram repartidas em novos áudios de 5 segundos. Após esta etapa, verificou-se que, dentro das bases de dados disponibilizadas, algumas possuíam diferentes proporções em relação ao número de áudios classificados como normais e anormais. De forma a tornar a base de dados de treinamento e validação equalizada, descartou-se alguns destes áudios. De posse da base de dados devidamente tratada, três tipos de RNN foram utilizadas para treinamento do classificador, posteriormente, fez-se comparação dos resultados de desempenho de classificação para os diferentes tipos de RNN. Por fim, os áudios descartados na etapa de equalização da base de dados foram submetidos ao classificador, a fim de realizar previsões de novos sons cardíacos. Os detalhes da metodologia desenvolvida em cada etapa do processo serão discutidos nos próximos itens deste capítulo.

3.1 Base de dados

Os dados utilizados pelo sistema são disponibilizados pelo evento *PhysioNet/Computing in Cardiology (CinC) Challenge 2016* e documentados no artigo de (Liu et al., 2016) e (Cliford et al., 2016). O evento supracitado, consistiu em um desafio internacional, onde os participantes foram convidados a implementar algoritmos que realizem a classificação de sons cardíacos nas categorias normal, anormal e incerto, quando se faz necessário refazer a gravação do som cardíaco. Segundo Liu et al. (2016), a base de dados é composta por gravações de sons cardíacos, coletados em uma grande variedade de ambientes, sejam estes em clínicas ou mesmo visitas domiciliares a pacientes. Cliford et al. (2016) afirma que a base de dados possui um total de 3163 gravações coletadas de 764 pacientes, sendo

estes saudáveis ou portadores de condições cardíacas consideradas patológicas. Esta base de dados é formada por seis grupos, onde cada grupo caracteriza uma base de dados proveniente de institutos que forneceram gravações de sons cardíacos ao desafio. Cliford et al. (2016), observa que, para o desafio, todos áudios da base de dados foram reamostrados para a frequência de 2000 Hz. A figura 24 ilustra a forma de composição da base de dados.

Figura 22 – Composição da base de dados



Fonte: Elaborado pelo autor.

Após os dados estarem devidamente organizados nos devidos grupos de classificação, foram utilizados para treinamento e validação do sistema classificador.

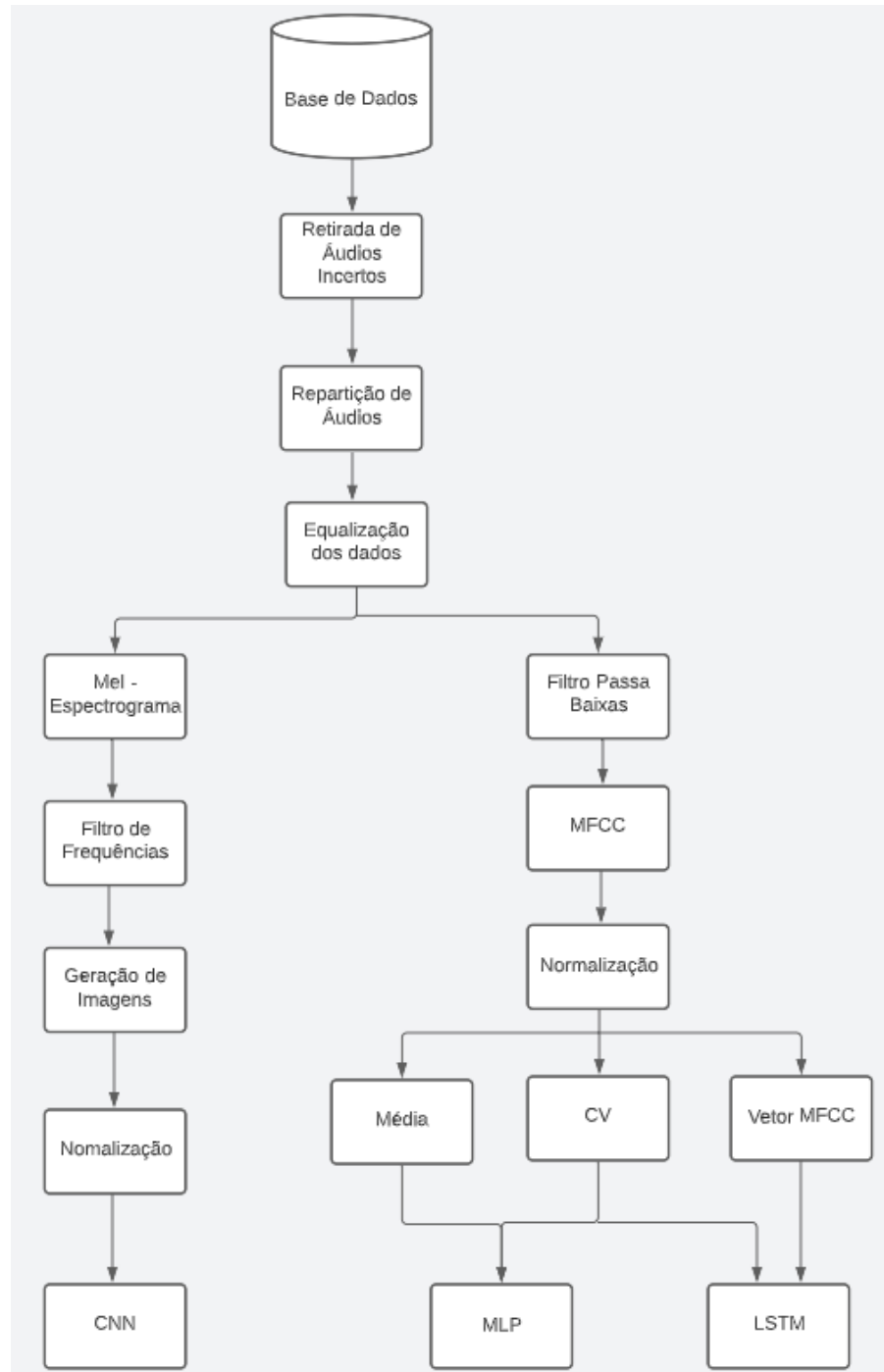
3.2 Pré-processamento

Na etapa de pré-processamento objetiva-se realizar o tratamento das amostras disponibilizados pela base de dados. Primeiramente organizou-se a base de dados a ser utilizada para o treinamento do classificador, por meio da retirada de áudios classificados como incertos, seguido da segmentação das gravações em áudios de 5 segundos de duração e finalização com o balanceamento dos dados, de forma que o número de áudios classificados como normais seja igual aos anormais.

As técnicas de tratamento de dados aplicadas variaram de acordo com o tipo rede neural utilizada para treinar o classificador, porém seguem o mesmo padrão: filtro do sinal, extração de características, seguido de normalização dos dados. Para o classificador baseado em CNN, as características são extraídas por meio de imagens

do espectrograma de frequências em escala mel, de cada um dos áudios da base de dados. Visando extrair uma imagem que demonstre somente as áreas de interesse do som cardíaco, utilizou-se a informação do estudo de Tagashira & Nakagawa (2020), onde atenta-se para a presença de sons cardíacos anormais nas faixas de frequências de 20 a 400 Hz, desta forma, limitou-se a representação do espectro de frequência a esta faixa. Após a implementação do filtro de frequências, as imagens são submetidas ao processo de normalização. Para os classificadores baseados em MLP e LSTM, um filtro digital passa baixas é utilizado para filtrar frequências até 400 Hz, seguindo a mesma linha da limitação do espectrograma de frequências aplicados as imagens. A partir do sinal filtrado, 13 características são obtidas por meio de MFCC, para cada áudio. Aplicando está técnica de extração de características, o áudio é fracionado em períodos de aproximadamente 23 ms, onde são retiradas 13 características por período de forma a obter-se um vetor de dimensões 216 x 13, para os áudios de 5s de duração. De posse do vetor de características, foram exploradas três técnicas para tratar este conjunto: média aritmética, coeficiente de variação e utilização do vetor de características íntegro. Ao calcular a média aritmética buscou-se o valor médio de cada um dos 13 coeficientes do MFCC ao longo das 216 amostras. O coeficiente de variação (CV) buscou extrair a informação de dispersão do valor de cada coeficiente ao longo das amostras de áudio. Na utilização do vetor íntegro, foi utilizado vetor de MFCC orginalmente obtido após o processo de normalização, este vetor foi submetido somente a LSTM, devido a arquitetura MLP não suportar entradas multidimensionais. A figura 23 demonstra as etapas do pré-processamento.

Figura 23 – Etapas de Pré-Processamento



Fonte: Elaborado pelo autor.

Ao final desta etapa, os dados, devidamente tratados são submetidos ao sistema classificador, servindo de dados de entrada para as respectivas redes neurais, cuja incumbência será classificar os sons cardíacos.

3.3 Sistema Classificador

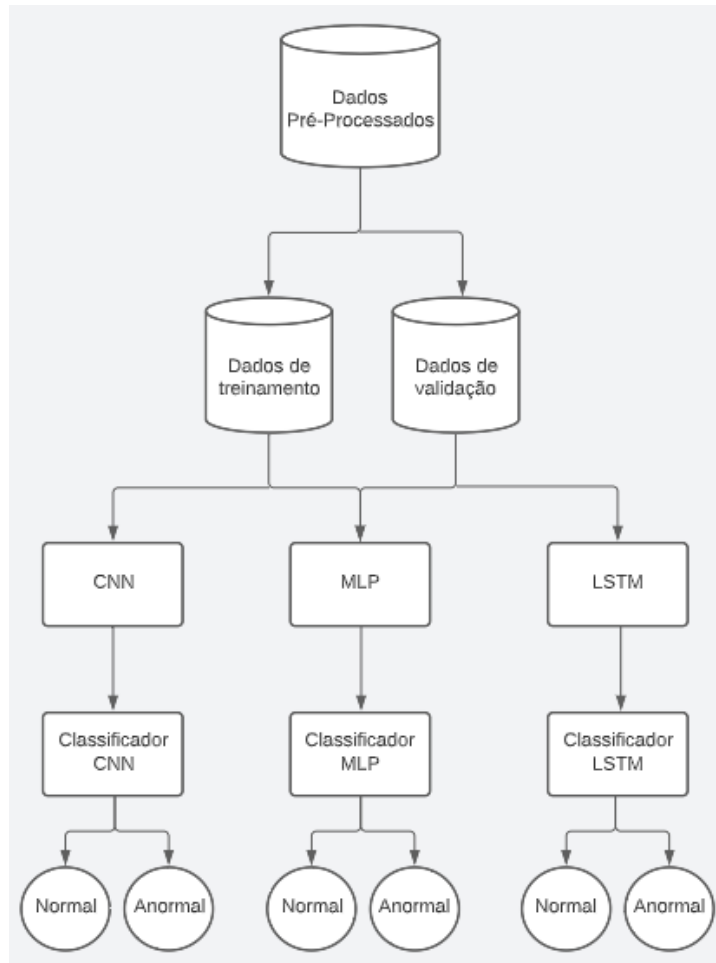
Um sistema de classificação tem por objetivo atribuir uma ou mais classes aos dados ao qual é submetido. Em muitas aplicações, os classificadores buscam separar os dados em função das características. Essas características tornam-se então as classes nas quais o classificador possuía a incumbência de atribuir a cada dado que é submetido a sua entrada. O sistema classificador aplicado nesta dissertação é do tipo binário, atribuindo aos dados as classes normal e anormal, que denotam sons cardíacos patológicos e não patológicos. A partir do uso de RNA encontram-se as características predominantes nos dados que influenciam na tomada de decisão que atribui aos dados uma das classes.

3.3.1 Redes Neurais Artificiais

O sistema classificador consistirá em um algoritmo de redes neurais artificiais (RNA) que receberá os dados da etapa de pré-processamento. Estes dados serão divididos em uma taxa de 75% para treinamento e 25% validação do modelo, sendo cada amostra de cada grupo escolhida de forma aleatória. As redes neurais necessitam primeiro ser parametrizadas e após são treinadas. O processo de treinamento é realizado junto da validação, sendo assim, para cada época de treinamento o modelo é validado, fornecendo um resultado de acurácia quanto aos dados de teste. Ao fim do treinamento, os parâmetros da época que obteve a maior acurácia serão considerados para constituir o modelo do sistema classificador.

O sistema deverá realizar a classificação dos sons cardíacos entre duas categorias, patológicos e não patológicos. Três tipos de redes neurais serão utilizados para treinamento do sistema classificador: MLP, LSTM e CNN, objetivando comparar o desempenho de cada tipo de algoritmo e por fim escolhendo-se aquele que obtiver o melhor desempenho. O diagrama da figura 26 ilustra a metodologia que será utilizada para implementação do sistema classificador.

Figura 24 – Etapas do sistema classificador



Fonte: Elaborado pelo autor.

Os três modelos de classificadores serão avaliados em relação a sua acurácia, quando submetido aos dados de validação, aquele que obtiver o melhor resultado nesta métrica será considerado o melhor modelo e posteriormente será submetido a etapa de avaliação do sistema, que será descrita no próximo item deste capítulo.

3.4 Avaliação do sistema

A validação do sistema será realizada por meio de validação cruzada e acuracidade, na etapa de validação cruzada, os dados serão divididos em 10 grupo e o modelo será treinado 10 vezes, alternando o grupo que será utilizado para treinamento, de forma a garantir a exclusão mútua entre os grupos de treino e validação. A avaliação do desempenho do sistema, se dará por meio da visualização da matriz de confusão obtida dos resultados de validação.

4 ANÁLISE DOS RESULTADOS

Este capítulo destina-se a apresentar os testes realizados com diferentes algoritmos de redes neurais, MLP, LSTM e CNN, seguido da avaliação. Organizou-se esta seção em cinco tópicos, no primeiro serão apresentados os resultados das operações de pré-processamento dos dados, bem como os alguns detalhes que foram observados nesta etapa. No próximo tópico, serão elucidados os resultados comparativos de desempenho de cada um dos algoritmos, os modelos serão, em primeiro momento avaliados por amostragem aleatória, onde será obtida a acurácia de cada modelo. Em seguida, o modelo que obteve melhor resultado na avaliação comparativa do tópico anterior será submetido a análise de robustez, por validação cruzada de dez grupos. No quarto tópico do capítulo, implementa-se a matriz de confusão, onde é avaliada o índice de acerto do classificador para cada uma das duas classes. Por fim, no último tópico os resultados obtidos serão discutidos.

4.1 Resultados da etapa de pré-processamento

O pré-processamento dos dados seguiu as etapas da metodologia. A primeira etapa a ser cumprida trata-se do tratamento da base, onde objetiva-se verificar a integridade dos dados, bem como manusear estes dados de forma a garantir que estejam nos formatos adequados. Seguido do tratamento dos dados, realiza-se a transformação dos registros de acordo com os modelos de RNA ao qual serão aplicados, por meio da extração de MFCC dos áudios e geração de imagens do espectrograma de frequências. Os próximos subcapítulos visam descrever detalhadamente estes processos.

4.1.1 Tratamento da Base de Dados

O pré-processamento dos dados seguiu a metodologia descrita no capítulo 3 desta dissertação. Implementou-se primeiro a retirada de áudios incertos, a base de dados possui um documento que descreve o nome o arquivo de áudio, associado ao diagnóstico do som cardíaco, seguido da informação relativa à certeza sobre o diagnóstico proferido. Desta forma, visando evitar que o sistema classificador seja treinado com dados duvidosos, eles foram retirados da base de dados. O quadro

abaixo sumariza a composição da base de dados antes e depois do processo de retirada de áudios incertos.

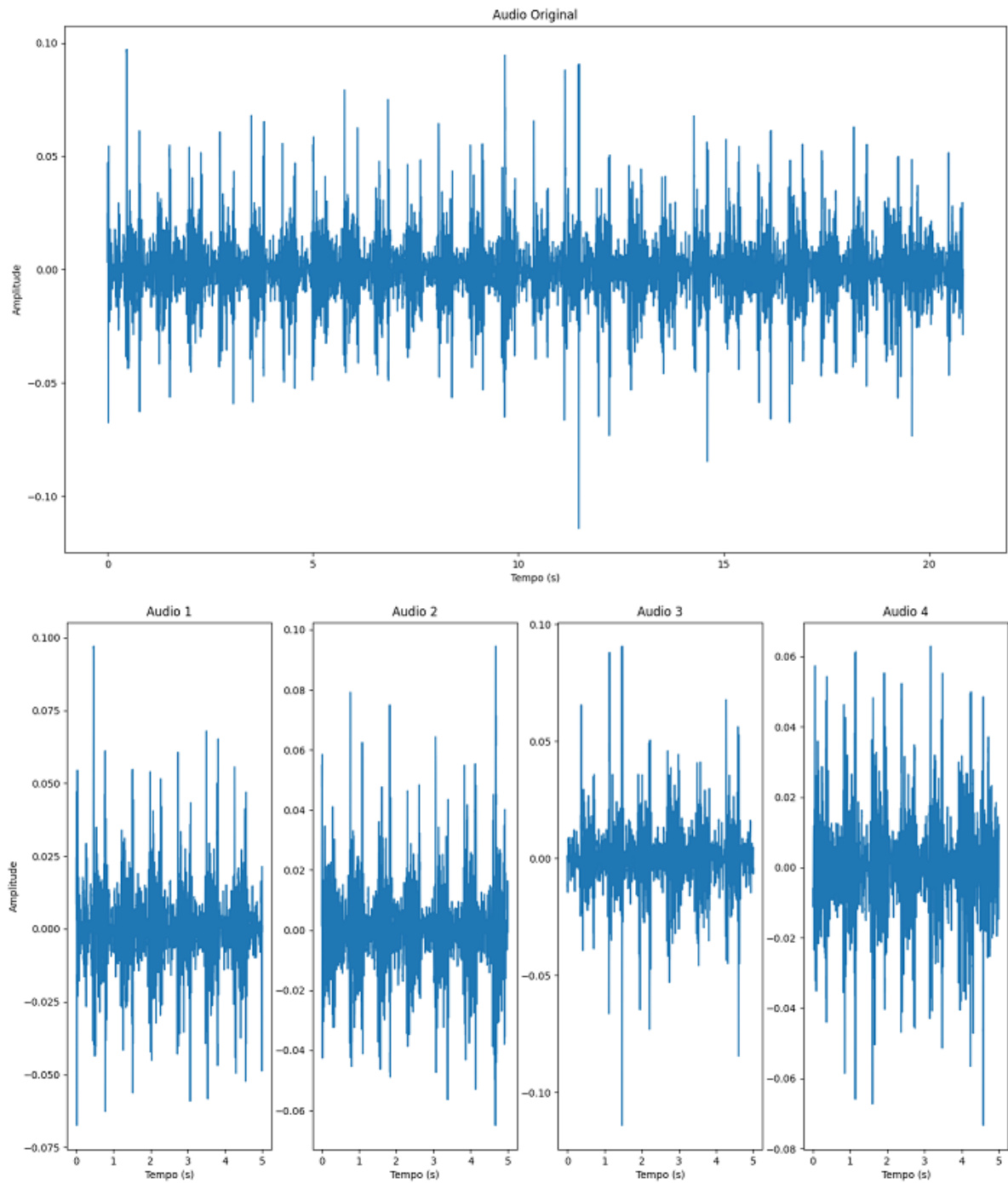
Quadro 1 – Configuração da Base de dados após a retirada de áudios incertos

Base de Dados Original				Base de Dados Filtrada		
Áudios Normais	Áudios Anormais	Áudios Incertos	Total de Áudios	Áudios Normais	Áudios Anormais	Total de Áudios
2488	665	279	3153	2302	572	2874

Fonte: Elaborado pelo autor.

Após filtrar os dados duvidosos da base de dados, os registros foram repartidos em áudios de 5 segundos de duração. Esta etapa visou padronizar o tamanho dos áudios, bem como promover o aumento da base de dados disponível, visto que alguns possuíam até 120 segundos de duração. A figura 27 exemplifica o processo de segmentação dos áudios, realizado em uma amostra da base de dados.

Figura 25 – Repartição de Áudios



Fonte: Elaborado pelo autor.

A partir desta segmentação, a base de dados foi aumentada, obtendo uma nova configuração que é demonstrada no quadro 2.

Quadro 2 – Configuração da Base de dados após segmentação dos áudios

Base de Dados com Áudios Segmentados		
Áudios Normais	Áudios Anormais	Total de Áudios
1691	1635	3326

Fonte: Elaborado pelo autor.

Após o processo de segmentação, notou-se um desequilíbrio na proporção entre áudios normais e anormais, desta forma, foi necessária a retirada de alguns destes registros para tornar o conjunto de dados equilibrado, este processo buscou equilibrar individualmente cada grupo da base de dados e ao final transferir os áudios dos subgrupos para a base de dados total. Estes áudios retirados foram posteriormente utilizados para realizar previsões com o sistema classificador. O quadro 3 ilustra a configuração final da base de dados que será submetida ao sistema classificador e a base de dados de previsões.

Quadro 3 – Configuração da Base de dados após equalização

Base de Dados do Classificador			Base de Dados de Predição		
Áudios Normais	Áudios Anormais	Total de Áudios	Áudios Normais	Áudios Anormais	Total de Áudios
1586	1586	3172	105	49	154

Fonte: Elaborado pelo autor.

A partir da execução da equalização dos dados, a base de dados está pronta para ser submetida aos processos de transformação de dados, onde são aplicadas duas abordagens distintas: a utilização de imagens para o modelo baseado em CNN e a extração de vetores de MFCC para os modelos de MLP e LSTM, que serão descritos nos tópicos seguintes.

4.1.2 Transformação de Dados para aplicação em classificadores baseados em MLP e LSTM

Nesta etapa, são realizados os pré-processamentos necessários para submeter os dados aos classificadores que utilizam as arquiteturas MLP e LSTM. A primeira ação a ser executada, consiste em filtrar as frequências dos sinais de áudio para valores de até 400 Hz, por meio de a aplicação de um filtro digital passa baixas do tipo Butterworth de quarta ordem, seguido da extração dos vetores de MFCC. A partir da utilização da biblioteca Librosa, escrita em linguagem Python, foi possível utilizar uma função que implementa o cálculo das MFCCs. Ao utilizar esta função torna-se necessário fornecer o número de características desejadas, que neste caso foram 13. Os intervalos de segmentação de áudio, por padrão da biblioteca estão configurados como 512 *frames*, utilizando um intervalo de amostragem de 22050 *frames* por segundo, obtém-se então segmentos de áudio de 23,22 ms. Como explicado no capítulo 2, a MFCC opera de maneira que são extraídas *n* características por segmento de áudio. Como estamos lidando com áudios de 5 segundos de duração e 13 características, a aplicação da função MFCC a uma amostra de áudio, retornará um vetor de dimensões 216 x 13. A figura 28 ilustra o formato do vetor MFCC de uma amostra de dados da base.

Figura 26 – Formato do vetor MFCC

Frame	C0	C1	C2	---	C10	C11	C12
0	-586.442	118.804	73.4625	---	17.2233	7.53845	1.02638
1	-589.369	117.886	79.1044	---	21.5623	7.9809	-1.81293
2	-590.202	117.568	80.8345	---	23.0701	9.56327	-1.31495
3	-591.151	116.609	80.8676	---	20.6032	7.74132	-1.93343
4	-571.629	140.877	96.4292	---	17.871	6.10631	-2.21469
5	-565.939	147.912	100.802	---	17.4973	5.871	-2.65158
...
210	-589.301	118.189	79.8694	---	24.0227	10.5849	0.306695
211	-590.098	117.073	78.8068	---	21.7998	7.67574	-3.04784
212	-589.076	117.493	76.622	---	19.0434	4.95801	-5.74512
213	-588.967	117.702	76.9883	---	19.0631	5.55889	-4.23935
214	-592.069	114.657	77.317	---	19.9619	6.61931	-3.1118
215	-584.183	111.64	86.7259	---	21.275	16.4375	-1.17178

Fonte: Elaborado pelo autor.

A partir da figura 28, é possível verificar que existem diferenças entre as magnitudes dos valores para cada coeficiente. Desta forma, indica-se a normalização destes valores, visando converter os valores de todos dos coeficientes para a mesma escala. A normalização por reescala foi aplicada aos vetores MFCC, de forma aos valores de cada coeficiente assumirem valores dentro do intervalo de 0 a 1. A figura 29 exemplifica o formato do vetor MFCC após a normalização por reescala.

Figura 27 – Formato do vetor MFCC normalizado

Frame	C0	C1	C2	C10	C11	C12
0	0.330831	0.256054	0.00318511	0.306113	0.460394	0.658093
1	0.254539	0.237198	0.143104	0.525436	0.484691	0.506056
2	0.232845	0.230657	0.186008	0.601649	0.571585	0.532721
3	0.208105	0.210944	0.186829	0.476956	0.471535	0.499604
4	0.716898	0.709747	0.572754	0.338852	0.38175	0.484543
5	0.865182	0.854334	0.68119	0.319965	0.368828	0.461149
...
210	0.256332	0.243424	0.162076	0.649801	0.627688	0.619556
211	0.235543	0.220479	0.135721	0.537443	0.467933	0.43993
212	0.262178	0.229113	0.0815392	0.398113	0.318692	0.295498
213	0.265024	0.233413	0.0906227	0.399108	0.351689	0.376128
214	0.184176	0.170835	0.0987764	0.44454	0.409921	0.436505
215	0.389713	0.108814	0.332114	0.510913	0.949078	0.540388

Fonte: Elaborado pelo autor.

Após a etapa de normalização do vetor MFCC, verificou-se a possibilidade obter, para cada coeficiente, um único valor que represente a variação das magnitudes deste coeficiente dentro do intervalo de 216 frames, desta forma, optou-se por obter estes valores por meio de medidas de tendência central, a partir da média aritmética de cada coeficiente e de variabilidade relativa, por meio do coeficiente de variação de cada coeficiente da MFCC, de forma que, cada áudio agora possui um vetor de características de dimensões 1 x 13 associado. A figura 30 ilustra um exemplo do formato dos dados após a aplicação da média aritmética aos coeficientes do vetor MFCC.

Figura 28 – Formato do vetor MFCC com aplicação da média aritmética

Audio	C0	C1	C2	----	C10	C11	C12
0	0.603811	0.637972	0.774825	----	0.307373	0.322533	0.280227
1	0.682961	0.731336	0.506563	----	0.310824	0.31098	0.35453
2	0.673322	0.706787	0.568405	----	0.230487	0.257412	0.260677
3	0.668913	0.673401	0.705169	----	0.202415	0.376853	0.37639
4	0.774002	0.819066	0.952065	----	0.322675	0.399469	0.423404
5	0.765025	0.546375	0.868176	----	0.274457	0.446096	0.408367

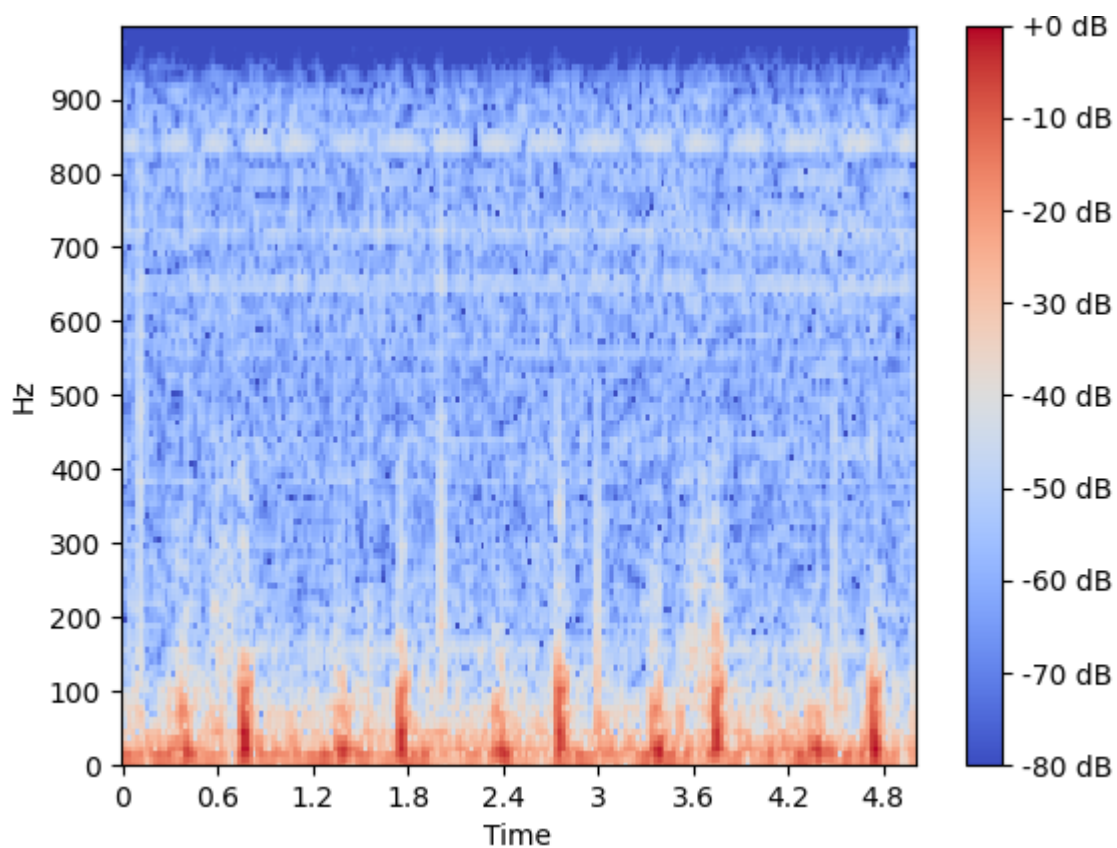
Fonte: Elaborado pelo autor.

A partir dos três formatos de vetores MFCC: vetor íntegro 216 x 13, vetor de média aritmética 1 x 13 e vetor de coeficientes de variação 1 x 13, foi construído um arquivo de formato .csv contendo as informações de cada um dos áudios: identificação da amostra, vetor MFCC associado e diagnóstico. A partir da leitura deste arquivo, extraem-se as informações necessárias para o treinamento e validação dos modelos de MLP e LSTM.

4.1.3 Transformação de Dados para aplicação no classificador baseado em CNN

A etapa de transformação dos dados para o classificador baseado em CNN, consiste em converter os sinais de áudio em uma representação em espectrograma de frequências, na forma de imagem. Nesta etapa, novamente utilizou-se a biblioteca Librosa para geração destes espectrogramas. Optou-se por gerar as representações em escala mel, devido a sua associação com a sensibilidade do ouvido humano, o que pode auxiliar o sistema a compreender a relação entre os áudios da base de dados e os diagnósticos fornecidos pelos profissionais de saúde, baseados no exame de ausculta cardíaca. Para utilização dos espectrogramas, necessitou-se configurar as janelas de tempos em que são aplicadas FFT, este valor foi configurado para 512 *frames*, obtendo-se espectros em janelas de aproximadamente 23 ms. Uma escala de cores também pode ser atribuída ao espectrograma, de forma que as regiões de maior concentração de energia possam ser destacadas por meio de cores, são representadas em escala decibel. Abaixo, demonstra-se o espectro de frequência de uma amostra da base de dados.

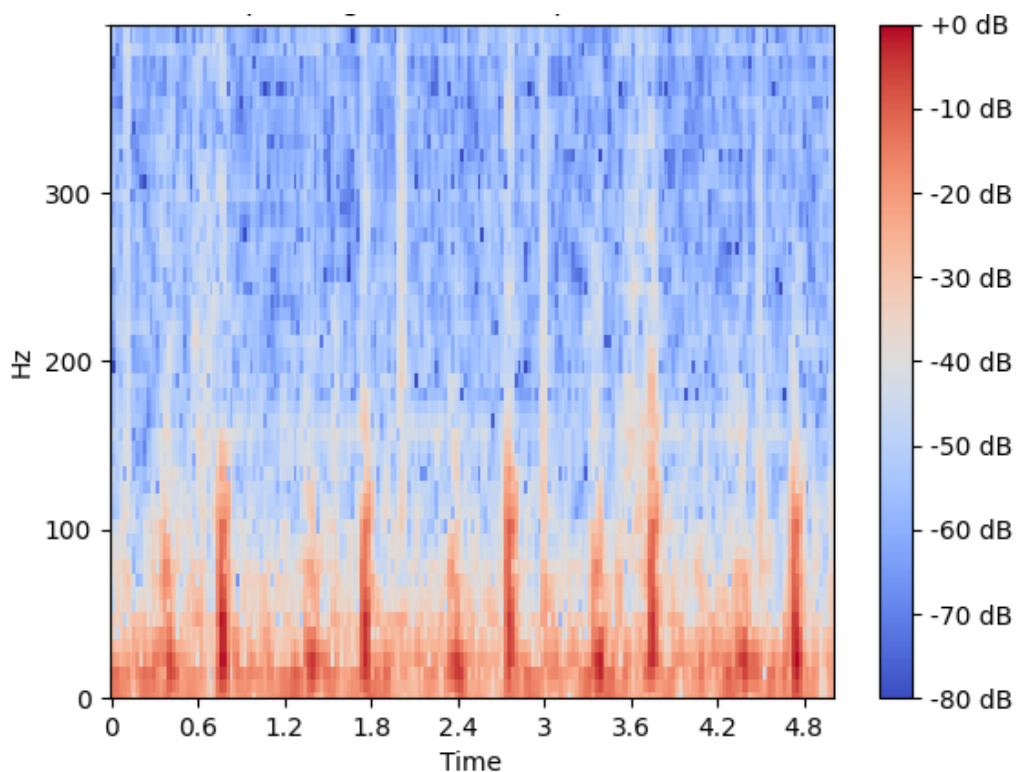
Figura 29 – Espectrograma de frequência de um som cardíaco



Fonte: Elaborado pelo autor.

A partir dos espectrogramas de frequência, é possível limitar os valores de frequência, a regiões onde estão presentes os sons cardíacos. Como sugere Tagashira e Nakagawa (2020), limitou-se as representações para frequências até 400 Hz, como demonstra a figura 32.

Figura 30 – Espectrograma de frequência limitado a 400 Hz



Fonte: Elaborado pelo autor.

Os espectrogramas de frequência de cada um dos áudios foram convertidos para imagens em formato .png. De posse das imagens, com auxílio da biblioteca python chamada numpy, as imagens foram convertidas em vetores, onde cada pixel é expresso por um vetor de 3 dimensões, correspondentes aos componentes de cores RGB (vermelho, verde e azul) da imagem, que variam em uma escala de 0 a 255. De forma a padronizar as imagens, necessita-se normalizar as mesmas, por meio da divisão destes vetores por 255, de forma a obter-se vetores que variam dentro dos valores 0 e 1.

A partir dos processos de pré-processamento acima citados, os dados estão prontos para serem submetidos aos algoritmos de RNA, onde o processo de treinamento e validação será realizado, obtendo-se os modelos de classificadores.

4.2 Resultados de desempenho dos modelos por amostragem aleatória

Neste subcapítulo, objetiva-se apresentar os resultados do processo de treinamento e validação dos modelos de MLP, LSTM e CNN. Os dados foram avaliados por meio de amostragem aleatória, de forma que, aleatoriamente, separou-se 75% dos dados para treinamento e 25 % dos dados para validação, garantindo a exclusividade mútua para estes grupos. O mesmo conjunto de dados de treino e teste foi aplicado aos três modelos de classificadores. Foi garantido também o equilíbrio dos grupos, de forma que ambos possuem aproximadamente o mesmo número de amostras de sons normais e anormais. O quadro 4 ilustra a distribuição dos grupos de treino e teste:

Quadro 4 - Configuração da Base de dados de treino e validação

Base de Dados de Treinamento			Base de Dados de Validação		
Áudios Normais	Áudios Anormais	Total de Áudios	Áudios Normais	Áudios Anormais	Total de Áudios
1187	1192	2379	399	394	793

Fonte: Elaborado pelo autor.

A partir da separação dos grupos de treinamento e validação, aplicou-se aos modelos de classificadores, onde o desempenho de cada uma destes foi medido por meio da acurácia. Os próximos tópicos descrevem os resultados obtidos para cada tipo de classificador.

4.2.1 Desempenho do classificador baseado em MLP

De posse dos dados de treino e teste, realizou-se o treinamento do modelo neural baseado em MLP. Primeiro foi necessário parametrizar a rede neural, fornecendo os dados de número de camadas, número de neurônios por camada, função de ativação, função de inicialização de pesos, otimizador e função de perda. Estes parâmetros foram variados, de forma a encontrar a arquitetura que apresenta o melhor resultado de acurácia. Realizada a parametrização do modelo, configura-se o número de épocas e tamanho de lote na qual o modelo dever ser treinado. Para o

modelo de MLP, utilizou-se a função de ativação ReLu associada a inicialização de pesos uniforme para as camadas de entrada e intermediária e a função softmax, que se constitui em uma variação da sigmoid foi utilizada na camada de saída do modelo. O otimizador Adam, associado a função de perda de entropia cruzada categórica esparsa apresentou os melhores resultados e foi aplicado ao modelo. Realizou-se o treinamento por 200 épocas, com tamanho de lote de 32. O quadro 5 demonstra a acurácia obtida pelo classificador para cada as métricas aplicadas, bem como a parametrização da arquitetura do modelo.

Quadro 5 - Resultado de desempenho do modelo de MLP

Métrica	Camadas	Neurônios por Camada	Acurácia	Sensibilidade	Especificidade
Média Aritmética	5	(128, 64, 32, 8, 2)	0,6608	0,6345	0,6867
CV	5	(128, 64, 32, 8, 2)	0,6683	0,5381	0,7970

Fonte: Elaborado pelo autor.

A partir dos resultados de acurácia do modelo, atentou-se a possibilidade de perda de informação em função da aplicação do filtro passa baixa em 400 Hz, de forma que poderia haver a possibilidade de existirem informações importantes nas frequências mais altas que não estavam sendo consideradas. Desta forma, retirou-se o filtro e processou-se novamente os dados, computando de forma íntegra. Obtendo os resultados demonstrados pelo quadro 6.

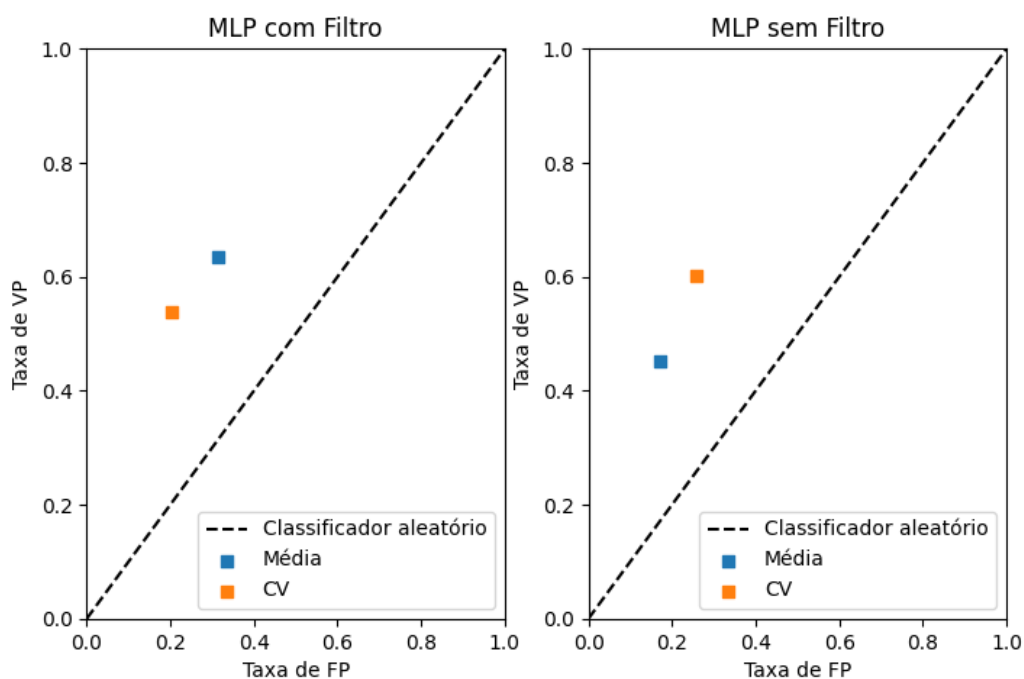
Quadro 6 – Resultado de desempenho do modelo de MLP sem aplicação de filtro

Métrica	Camadas	Neurônios por Camada	Acurácia	Sensibilidade	Especificidade
Média Aritmética	5	(128, 64, 32, 8, 2)	0,6406	0,4518	0,8271
CV	5	(128, 64, 32, 8, 2)	0,6721	0,6015	0,7419

Fonte: Elaborado pelo autor.

A partir da obtenção dos resultados do classificador baseado em MLP, verificou-se o desempenho do modelo a partir da variação da medida descritiva e pela retirada no filtro de frequências. A figura 33 ilustra o espaço ROC para os modelos que implementaram o filtro e os que não o fizeram.

Figura 31 – Espaço ROC para os resultados da MLP



Fonte: Elaborado pelo autor.

Verificou-se que a retirada do filtro não resultou em melhoras significativas no desempenho do sistema, porém, optou-se por implementar este teste nas demais arquiteturas de RNA, visto que o algoritmo do modelo também pode influenciar no desempenho do classificador.

4.2.2 Desempenho do classificador baseado em LSTM

O treinamento do classificador baseado em LSTM seguiu a mesma metodologia do modelo MLP, com exceção ao fato de que, para a LSTM são utilizados dois tipos de dados de entrada, os vetores MFCC constituídos pela média e pelo CV dos vetores originais e os próprios vetores MFCC íntegros. Para cada tipo de entrada a arquitetura apresentou algumas diferenças com relação a parametrização de número de camadas e de neurônios. Foram utilizadas as funções de ativação ReLu e Softmax, nas camadas intermediárias e de saída, respectivamente, associadas a inicialização de pesos uniforme. Os parâmetros de otimizador e função de perda foram os mesmos da MLP: Adam e entropia cruzada categórica esparsa. As redes LSTM as quais foram submetidos os vetores MFCC reduzidos foram treinadas por 200 épocas e tamanho de lote 32, enquanto o modelo LSTM que foi submetido ao vetor íntegro foi

treinado por 100 épocas e tamanho de lote 16. Em ambos os modelos, foram atribuídos os parâmetros em que se obteve melhor desempenho do modelo quanto a acurácia relativa à previsão dos dados de teste. O mesmo processo de retirada do filtro de sinais, realizado com a MLP foi aplicado aos modelos LSTM. O quadro abaixo ilustra os resultados obtidos para cada um dos modelos de LSTM.

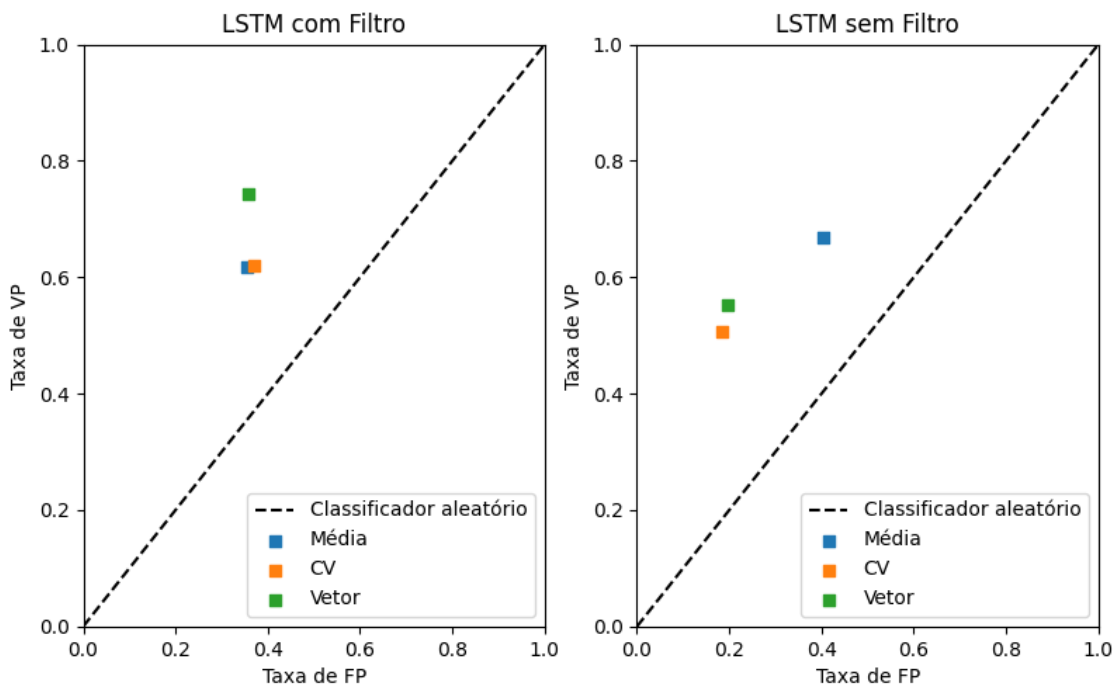
Quadro 7 – Resultado de desempenho do modelo de LSTM

Métrica	Camadas	Neurônios por Camada	Acurácia		Sensibilidade		Especificidade	
			Com Filtro	Sem Filtro	Com Filtro	Sem Filtro	Com Filtro	Sem Filtro
Média Aritmética	5	(128, 64, 32, 8, 2)	0,6318	0,6343	0.6179	0,6691	0.6462	0.5974
CV	5	(128, 64, 32, 8, 2)	0,6494	0,6620	0.6691	0.5076	0.6289	0.8145
Vetor	3	(64, 32, 2)	0,6910	0,6810	0.7428	0.5521	0.6432	0.8020

Fonte: Elaborado pelo autor.

A partir dos resultados obtidos, verificou-se o desempenho dos sistemas para os diferentes formatos de dados e filtros de frequência aplicados. A figura abaixo ilustra o espaço ROC onde podem ser observadas a influência de cada uma destes fatores.

Figura 32 – Espaço ROC para os classificadores LSTM



Fonte Elaborado pelo autor.

A partir dos resultados obtidos, é possível observar que a retirada do filtro impactou no desempenho de alguns classificadores, principalmente quanto a especificidade, indicando a possibilidade de existir informações relevantes nas frequências mais altas.

4.2.3 Desempenho do classificador baseado em CNN

O modelo de classificador baseado em CNN, seguiu a mesma linha das demais arquiteturas de RNA, porém a CNN exige a configuração de alguns parâmetros a mais: dimensão do filtro da camada convolucional e da camada de pooling, as quais foram parametrizadas com as dimensões, 5x5 e 2x2, respectivamente. A arquitetura do modelo conta com 6 camadas convolucionais, 6 camadas de pooling e uma camada densa, sendo a última antes da camada de saída. As camadas de pooling sempre são adicionadas após a camada convolucional. As funções de ativação mantiveram-se as mesmas dos modelos anteriores, ReLu e Softmax, porém, para inicialização de pesos, a função normal apresentou melhor desempenho. A função de perda, manteve-se entropia cruzada categórica esparsa e o otimizador foi alterado para Adamax, visto que, associado a inicialização de pesos normal, apresentou melhor desempenho. Este modelo foi treinado por 70 épocas, utilizando o tamanho de lote de 64. Assim como nos demais modelos, implementou-se o a retirada do filtro de sinais, como a etapa de pré-processamento das imagens exige menos poder computacional em comparação a MFCC, explorou-se a questão do filtro de sinais com mais profundidade, promovendo a análise do desempenho dos sistemas nas seguintes situações: Filtro de 0 a 400 Hz, Filtro de 0 a 250 Hz, Filtro 400 Hz a 1000 Hz e 0 a 1000 Hz, O quadro 8 ilustra os resultados obtidos para o modelo CNN.

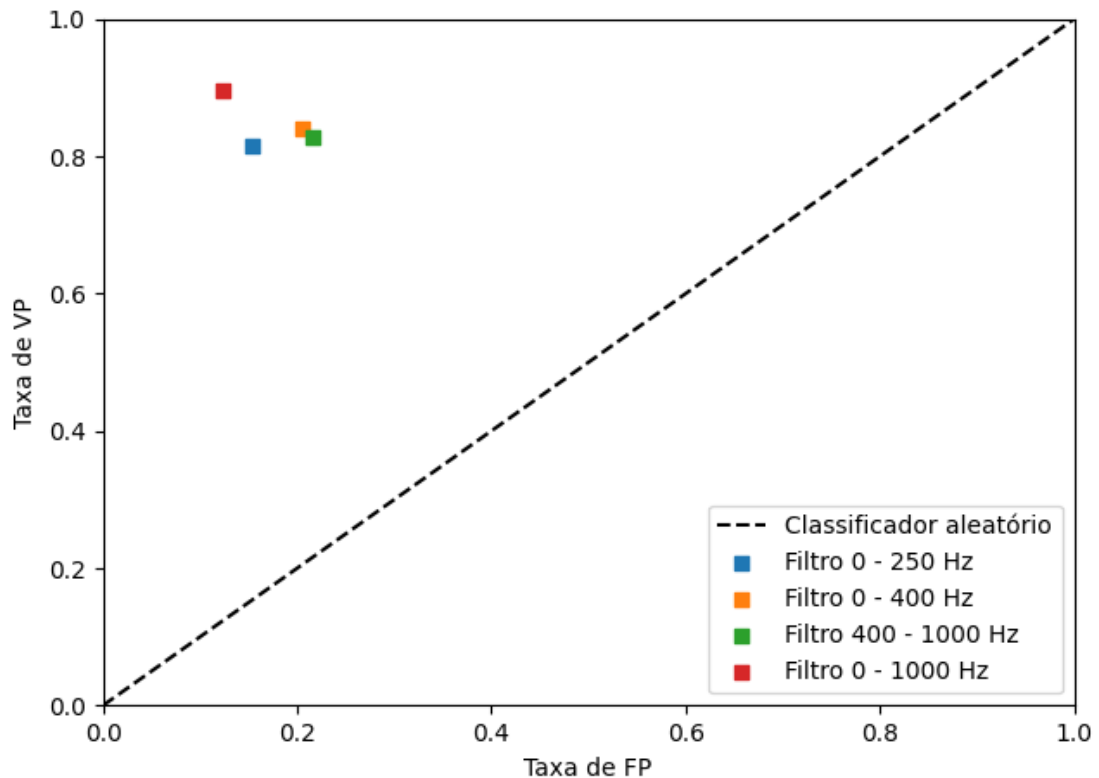
Quadro 8 – Resultado de desempenho do modelo de CNN.

Faixa de frequência do filtro	Neurônios por Camada	Acurácia	Sensibilidade	Especificidade
0 – 250 Hz	(24, 36, 48, 96, 192, 384, 1024, 2)	0,8310	0,8147	0,8471
0 – 400 Hz	(24, 36, 48, 96, 192, 384, 1024, 2)	0,8171	0,8401	0,7944
400 Hz – 1000 Hz	(24, 36, 48, 96, 192, 384, 1024, 2)	0,8058	0,8274	0,7844
0 – 1000 Hz	(24, 36, 48, 96, 192, 384, 1024, 2)	0,8865	0,8959	0,8771

Fonte: Elaborado pelo autor.

De posse destes resultados, é possível verificar o desempenho do sistema para as diferentes faixas de frequências. A figura abaixo ilustra o espaço ROC que fornece informações importantes a respeito do desempenho do modelo para diferentes faixas de frequências.

Figura 33 – Espaço ROC para os classificadores CNN



Fonte: Elaborado pelo autor

A partir dos resultados é possível observar que a aplicação da CNN com auxílio das imagens de espectrogramas de frequência apresentou resultados bem superiores aos modelos LSTM e MLP. Também é possível observar que a retirada do filtro (faixa 0 – 1000 Hz) impactou positivamente nos resultados de todas as métricas do classificador. Conforme observado na metodologia deste trabalho, o modelo que apresentasse o melhor desempenho na avaliação por amostragem aleatória seria considerado o mais exitoso, de forma a ser submetido às próximas métricas de avaliação de modelos que serão descritas nos próximos tópicos.

4.3 Resultados da aplicação de validação cruzada

A partir do modelo de CNN obtido, submeteu-se o mesmo ao processo de validação cruzada de 10 grupos. Para cada uma das dez combinações de grupos de treino validação, obtém-se a acurácia do classificador quanto à predição dos dados de teste. Ao fim deste processo, retira-se a média aritmética das acurácias obtidas, constituindo-se este valor como a acurácia do sistema. A validação cruzada permite verificar a robustez do classificador, visto que é exposto a diferentes conjuntos de dados. Espera-se que os valores de acurácia se mantenham semelhantes nas diferentes combinações de grupos. O quadro 9 apresenta os valores de acurácia obtidos para cada grupo, seguido da acurácia média obtida, caracterizando a acurácia definitiva do classificador.

Quadro 9 – Resultado da validação cruzada do classificador

Grupo	Acurácia
1	0,9119
2	0,8994
3	0,8644
4	0,8738
5	0,8770
6	0,8896
7	0,9022
8	0,9243
9	0,8927
10	0,9117
Média	0,8947

Fonte: Elaborado pelo autor.

A partir do quadro 9, observa-se que os valores de acurácia do classificador mantiveram-se relativamente próximos dentre todos os grupos, provando que, mesmo exposto a diferentes combinações de treino e teste, é capaz de obter bons resultados, obtendo uma acurácia média de 0,8947. A partir dos resultados obtidos na validação cruzada, avalia-se o sistema quanto a classificação específica das duas classes, estas métricas são descritas pela matriz de confusão, que será abordada no próximo tópico.

4.4 Matriz de confusão do classificador

De posse dos resultados do classificador quando a validação cruzada, deseja-se obter um detalhamento a respeito do desempenho quanto a classificação de sons normais e anormais de forma separada. Estas informações podem ser obtidas por meio da matriz de confusão, que fornece as informações de acerto e erro de previsões para cada classe. Como este modelo foi validado em 10 combinações de conjuntos diferentes, serão elaboradas o mesmo número de matrizes. A partir destas, retira-se a média aritmética de cada um de seus elementos, a fim de constituir a matriz de confusão final, que sintetiza os resultados do classificador. Optou-se por expressar os valores da matriz em taxas, de forma que a matriz possua as informações: taxa de verdadeiros positivos, taxa de falso positivos, taxa de verdadeiros negativos e taxa de falsos negativos. O quadro 10 mostra as matrizes de confusão final do classificador, a matrizes de todos os grupos podem ser visualizadas e maior detalhe no apêndice A deste trabalho.

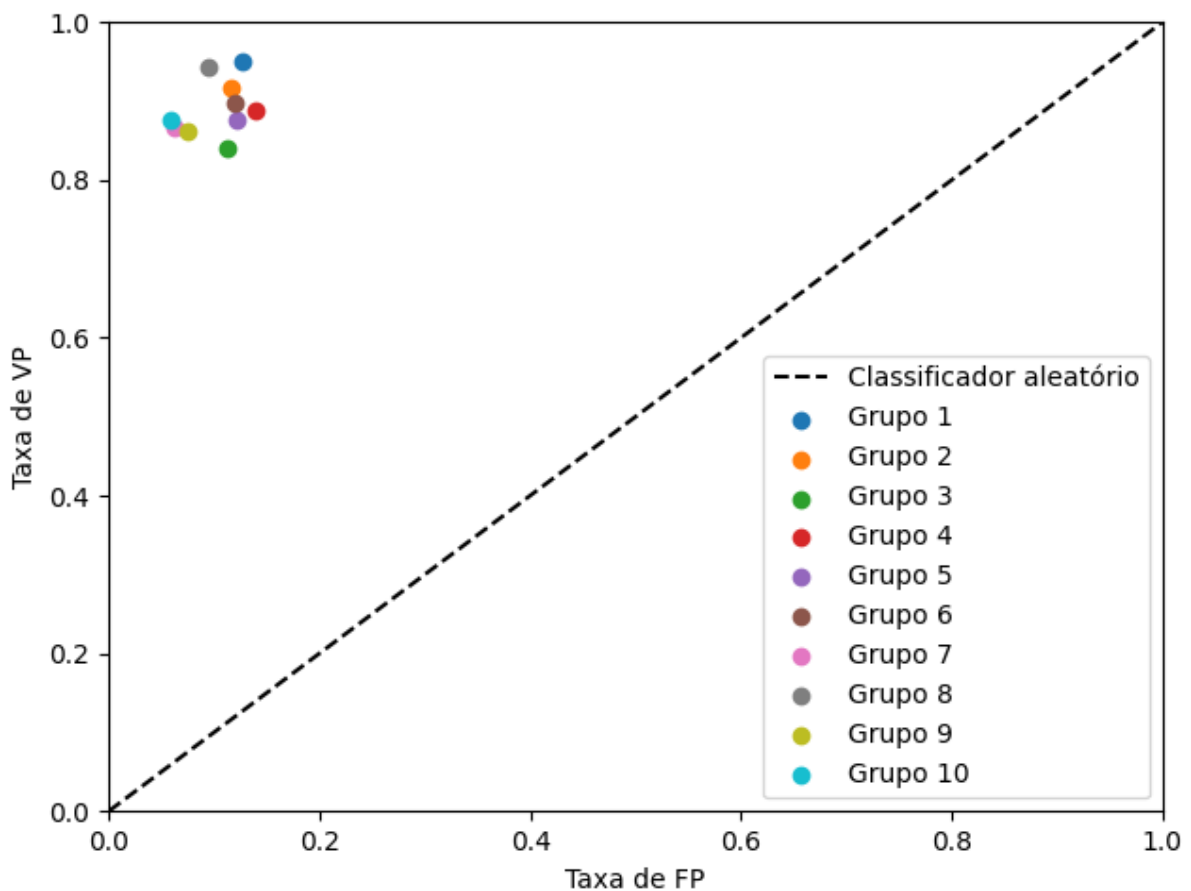
Quadro 10 – Matriz de confusão do classificador

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8915	0,1085
Normal	0,1027	0,8973
Precisão	0,8926	
Acurácia	0,8947	

Fonte: Elaborado pelo autor.

A partir da matriz de confusão sintetizada do classificador, destacam-se os valores de sensibilidade (taxa de verdadeiros positivos) e especificidade (taxa de verdadeiros negativos), assumindo os valores 0,8915 e 0,8973, respectivamente. A partir deste resultado, observa-se que o classificador obtém um desempenho satisfatório quando a classificação individual de ambas as classes. A figura 36, sintetiza, os dados da matriz de confusão, obtidos para cada um dos dez grupos da validação cruzada, por meio do espaço ROC.

Figura 34 – Espaço ROC para os dez grupos de validação cruzada



Fonte: Elaborado pelo autor.

De posse das métricas de avaliação de desempenho, realiza-se a etapa de predição de novos sons cardíacos por parte do modelo. Os resultados das predições são discutidos no próximo subcapítulo.

4.5 Predições de Sons Cardíacos

A etapa de predição de sons cardíacos, objetiva aproveitar as gravações de sons cardíacos que foram descartadas na etapa de equalização da base de dados para submeter as mesmas ao modelo final do classificador. Deve-se ressaltar que, como estes áudios foram retirados da base de dados, não participaram dos processos de treinamento e validação do sistema, constituindo-se como dados novos para o classificador. Como apresentado no quadro 3, a base de dados de predição conta com um total de 154 gravações de sons cardíacos, sendo 105 normais e 49 anormais. Como esta base de dados está desequilibrada, a matriz de confusão torna-se essencial para avaliação dos valores preditos pelo classificador. Os dados foram submetidos ao classificador que obteve o melhor desempenho durante a validação cruzada, desta forma, o melhor modelo torna-se aquele que foi treinado e validado pelo conjunto de dados do grupo 8, cujo obteve acurácia de 0,9243, conforme mostrou o quadro 9. O quadro a seguir mostra a matriz de confusão e a acurácia obtida da predição destes sons cardíacos.

Quadro 11 – Matriz de confusão dos Resultados de Predição

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8367	0,1633
Normal	0,0667	0,9333
Precisão	0,8542	
Acurácia	0,9026	

Fonte: Elaborado pelo autor.

A partir destes resultados, observa-se que o modelo apresentou um bom resultado de acurácia, seguido de uma especificidade elevada de 0,9333 e uma sensibilidade de 0,8367, menor que a métrica anterior, porém alcançando valores satisfatórios.

Finalizada a etapa de avaliação dos modelos de classificadores, existem subsídios suficientes para realizar-se a interpretação e a discussão dos resultados obtidos, o que será feito no próximo capítulo desta dissertação.

5 DISCUSSÃO

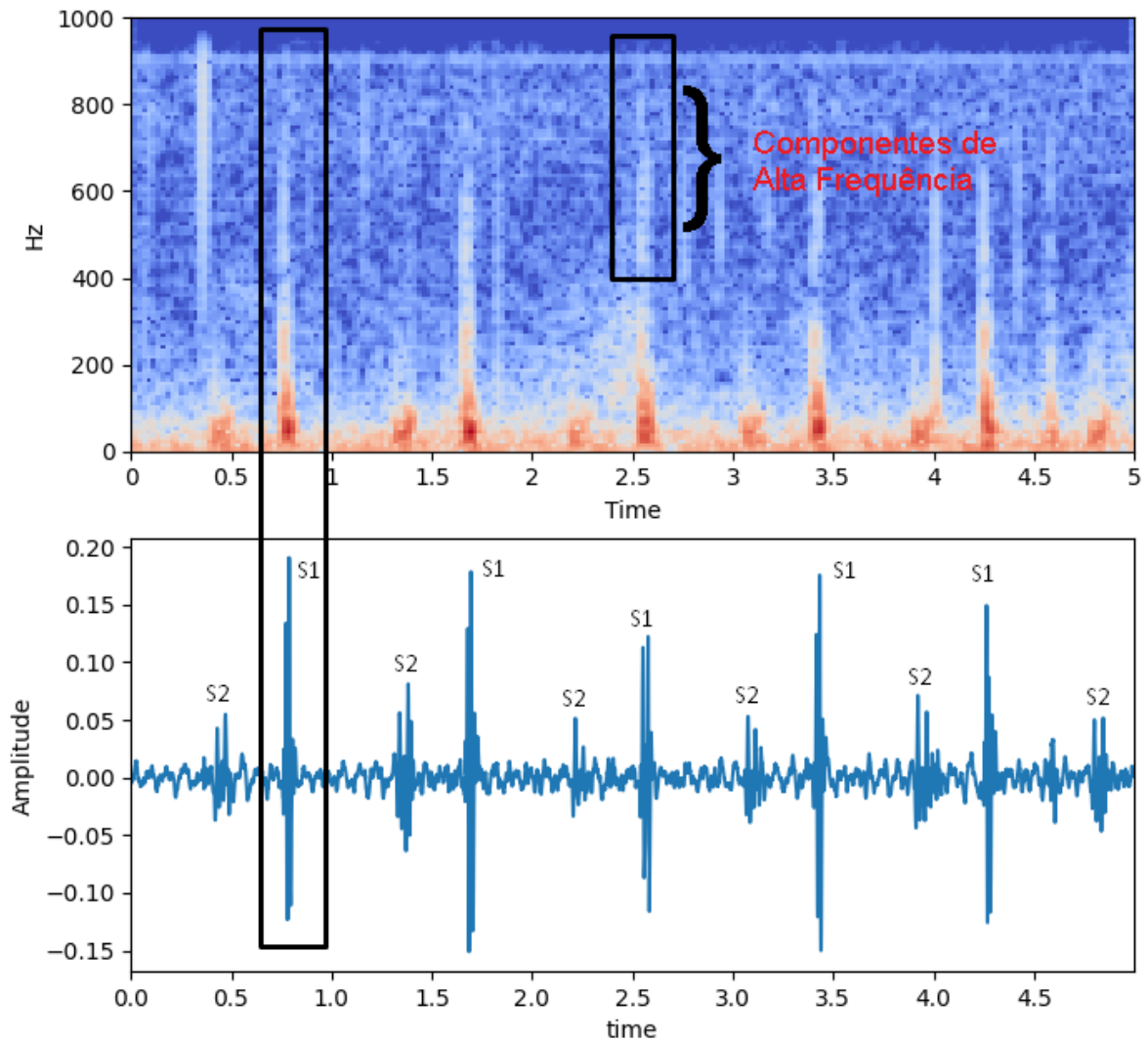
O capítulo anterior destinou-se a apresentar os resultados obtidos das métricas de análise de desempenho aplicadas ao classificador, por meio das técnicas de amostragem aleatória, validação cruzada e matriz de confusão. Estes resultados permitiram avaliar o sistema quanto a sua acuracidade e robustez.

A aplicação das redes neurais MLP e LSTM como abordagem inicial à solução do problema do classificador, associada a extração de características por meio da MFCC para constituir os dados de entrada das destas redes neurais, foi avaliada pelo método de amostragem aleatória, apresentando resultados de acurácia, sensibilidade e especificidade, para diferentes formatos dos dados de entrada e com a aplicação do filtro passa baixa, configurado em 400 Hz e sem a aplicação deste. Quanto ao formato dos dados de entrada, a aplicação do coeficiente de variação aos dados dos vetores MFCC apresentou-se a técnica mais eficaz de medida descritiva dos coeficientes MFCC, indicando que, a medida do grau de dispersão do valor de cada coeficiente, ao longo dos frames do áudio trouxe informações mais relevantes do que a média aritmética destes coeficientes. A utilização do vetor MFCC íntegro, aliado à rede LSTM, forneceu o melhor resultado dentre os modelos que utilizaram esta técnica de extração de características, tendo sua acurácia ligeiramente maior que as demais técnicas. Assim, observa-se que as técnicas de medidas descritivas aplicadas resultaram em perda de informações, resultando em um desempenho inferior. Quanto a análise de aplicação dos filtros de frequência, as alterações não impactaram de maneiras significativa para os modelos MLP, de forma que, mudanças mais expressivas ocorreram nos classificadores LSTM, onde, mesmo observando-se acurácias semelhantes entre as duas variações, as taxas de falsos positivos diminuíram nos classificadores que adotaram a retirada do filtro. Mesmo que, a diminuição da taxa de falsos não signifique necessariamente uma melhora no desempenho do classificador, é possível observar indícios de que realmente existem informações relevantes nas altas frequências dos áudios, que podem contribuir para a performance do classificador, deste que a técnica de extração de características seja a ideal.

A partir da verificação dos resultados dos modelos MLP e LSTM, abordou-se a extração de características dos sons cardíacos pelo processamento das imagens dos espectrogramas de frequência, como forma alternativa as MFCCs. A partir da

utilização de CNN, o modelo foi treinado e validado, passando posteriormente pela avaliação de desempenho do classificador. Na etapa de avaliação por amostragem aleatória, verificou-se o desempenho do sistema para diferentes faixas de frequências dos áudios. Os resultados comparativos dos classificadores CNN para as diferentes frequências, apontou o classificador associado a faixa de frequência de 0 a 1000 Hz, que se constitui dos áudios não filtrados, como o mais exitoso entre todos, confirmando os indícios visto nos resultados de modelos anteriores, de que existem informações relevantes nas altas frequências. A figura 37, ilustra o comportamento do espectrograma nas altas frequências de um dos áudios da base de dados.

Figura 35 – Associação entre áudio e espectrograma de frequência



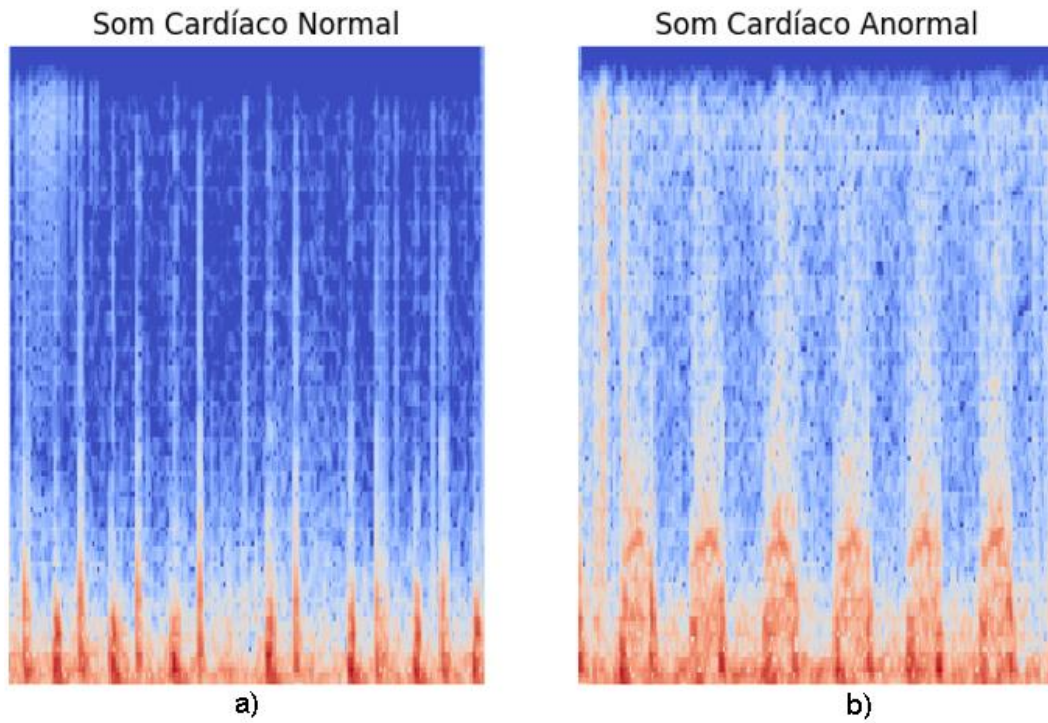
Fonte: Elaborado pelo autor.

A partir da figura 37, é possível elaborar a hipótese para o incremento de desempenho do classificador a partir da retirada do filtro de frequências. Observa-se no sinal de áudio, que a ocorrência dos sons cardíacos S1 e S2 está associada a regiões de concentração de energia no espectrograma. Especialmente na ocorrência do primeiro som cardíaco, nota-se a presença de feixes brancos verticais que se deslocam para as altas frequências do espectro. A presença destas frequências elevadas durante a ocorrência de S1 e S2, pode ser explicada por variações abruptas do sinal quando a ocorrência destes eventos, resultado em informações relevantes que podem ser extraídas pela CNN.

O modelo foi submetido a validação cruzada por dez grupos, realizada junto a matriz de confusão, onde foram obtidos excelentes resultados, demonstrando a precisão e robustez do sistema classificador, este desempenho pode ser confirmado visualmente pelo espaço ROC da figura 36, onde observa-se que as taxas de verdadeiros positivos se apresentaram elevadas em todas as combinações de grupos, esta informação, associada as baixas taxas de falsos positivos e a uma acurácia média 0,8947 e 0,8926, permitem concluir que classificador apresentou resultados acima do esperado.

Torna-se importante salientar que, o classificador obteve um excelente desempenho, classificando sons cardíacos que muitas vezes são difíceis de serem diferenciados e classificados por um humano. Mesmo quando se converteu os sons cardíacos para espectrogramas, após a inspeção das imagens a olho nu, muitas vezes as marcantes características das imagens permitiam verificar claramente a presença de um som cardíaco patológico, o que, por outras vezes não é tão claro. A figura 38 ilustra casos em que os diagnósticos são claros e casos em que esta identificação não é trivial.

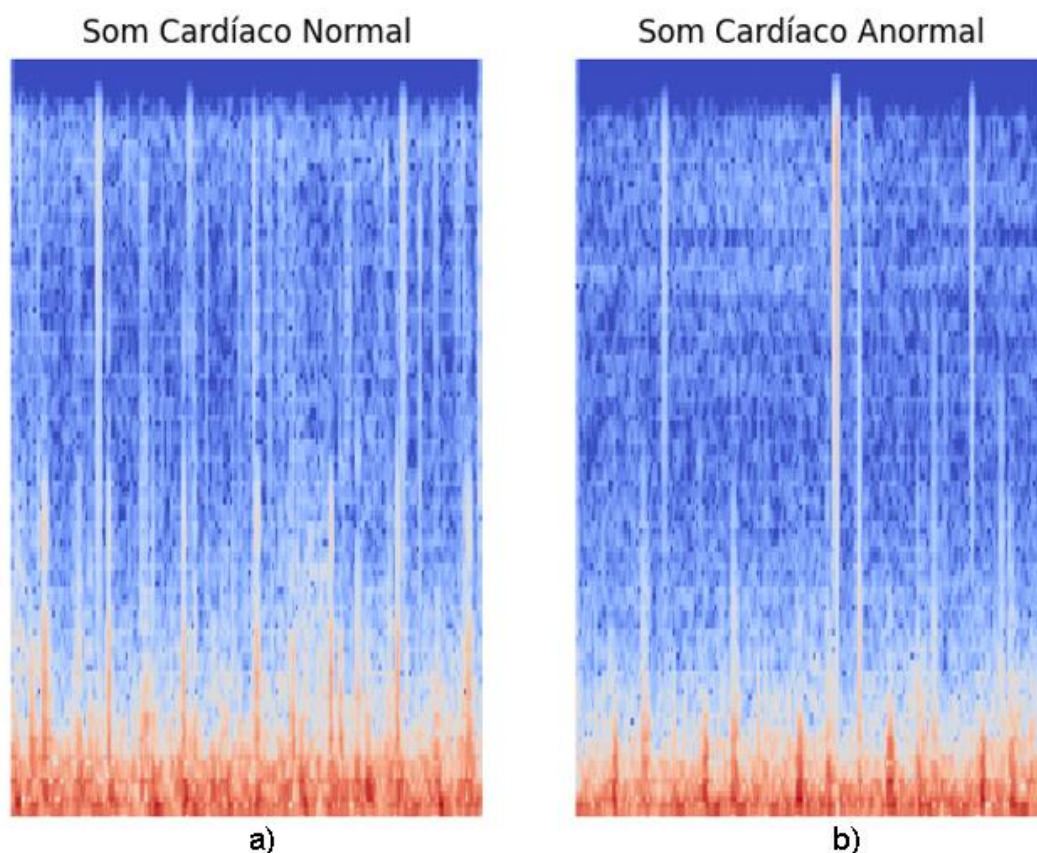
Figura 36 – Som Cardíaco Normal x Anormal – Caso 1



Fonte: Elaborado pelo autor.

A figura 38 ilustra um caso em que existem características marcantes em uma das imagens que permitem, verificar que o som cardíaco da imagem b, possui atributos visuais bem distintos da imagem a, sugerindo que se trata de um som cardíacos patológico. Porém, como mencionado anteriormente, nem todas as imagens possuem características tão claras. As imagens abaixo, retiradas da base de dados ilustram um destes casos.

Figura 37 - Som Cardíaco Normal x Anormal – Caso 2



Fonte: Elaborado pelo autor

Na imagem acima, verifica-se uma grande similaridade visual entre as imagens a e b, sendo muito difícil para um humano ser capaz de identificar qual das imagens representa o sopro cardíaco patológico. Em caso assim, verifica-se os bons resultados do classificador. Mesmo com imagens aparentemente parecidas, o sistema é capaz de encontrar características na imagem que sugerem o diagnóstico correto.

A partir da realização desta discussão de resultados, acredita-se veementemente no ineditismo deste trabalho, os argumentos para esta crença tratam-se primeiramente pela capacidade do classificador de identificar os sopros cardíacos com eficiência, visto que este foi exposto a uma base de dados de validação de tamanho considerável. Acredita-se que esta eficiência possa estar justificada justamente pela associação dos componentes de baixa frequência aos de alta frequência do espectro, o que foi possível devido aos estudos relativos à eficiência dos classificadores para diferentes faixas de frequência do espectro, algo que também acredita-se ter sido implementado neste trabalho de forma inédita. Este trabalho também realizou comparações de performances de diferentes algoritmos RNA, variando as técnicas de

extração de características, bem como o formato nas qual estas são tratadas e submetidas as entradas das redes neurais, não foi encontrado nas referências bibliográficas nenhum estudo que combinasse estas abordagens. Por fim, espera-se que as abordagens inovadoras apontadas neste estudo possam constituir-se como referência para futuros estudos, tanto na área de sons cardíacos, como em outras aplicações que possam utilizar-se de tópicos abordados neste estudo.

6 CONCLUSÃO

Este estudo, tinha como objetivo elaborar um sistema de classificação de sons cardíacos robusto e confiável, de forma a apresentar uma solução de tecnologia que fosse capaz de ser submetida ao áudio de um exame de ausculta cardíaca, buscar características nos mesmo e sugerir um diagnóstico com base em suas análises. Para atingir estes objetivos vislumbrou-se a inteligência artificial, por meio das técnicas de *machine learning*, com uma ferramenta que poderia tornar este sistema possível de ser implementado.

A partir da ideia inicial, buscou-se por trabalhos relacionados ao assunto que pudessem nortear a pesquisa. A bibliografia a respeito de sistemas classificadores mostrou-se escassa em com algumas lacunas que poderiam ser exploradas. Dos trabalhos correlatos, extraiu-se a ideia de extração de características por MFCC, bem como os uso de espectrogramas de frequência associados a CNN, destas mesmas obras, identificou-se oportunidades de explorar pontos que não foram investigados pelos autores. Dentro dos trabalhos, não foi identificada a utilização de MFCCs aplicadas aos algoritmos MLP e LSTM, dedicados a problemática dos sons cardíacos. A utilização da CNN foi abordada por alguns autores, porém não foi explorada a problemática das frequências dos sons cardíacos e sua relação com a eficácia de classificação dos modelos. Por fim, muitas obras contaram com bases de dados menores, o que pode prejudicar o desempenho do classificador. Os trabalhos que utilizaram, a base de dados *Physionet*, abordada por esta dissertação, contaram com os mesmos dados aqui utilizados, porém, observou-se que esta possui áudios com diferentes tempos de duração. Neste ponto atentou-se a repartir estes áudios em intervalos fixos, de forma a padronizar suas durações, bem como promover o aumento dos números de áudios disponíveis para treinar o classificador.

O sistema classificador foi treinado e validado com três diferentes modelos neurais, sendo posteriormente avaliado de maneira geral quanto aos seus desempenhos e por último, o algoritmo que obteve o melhor desempenho, passou por uma avaliação detalhada, aplicando métricas de avaliação de modelos sugeridos pelas bibliografias.

Os resultados observados após a avaliação e comparação dos modelos mostram que o classificador baseado em CNN cumpriu os objetivos deste trabalho,

provando-se robusto e confiável, sendo capaz de classificar com eficácia tanto os dados de validação, como os dados destinados a realizar novas predições.

Com base em todos os tópicos abordados neste estudo, bem como os resultados obtidos, ficam alguns pontos que poderiam constituir-se como melhorias ou sugestões para trabalhos futuros na área de sons cardíacos ou em aplicações que possam utilizar-se das mesmas ferramentas:

- Explorar outras técnicas de extração de características de áudio, como *Linear Predictive coding* (LPC), *Perceptual Linear Predictive* (PLP) e *Power Normalized Cepstral Coefficients* (PNCC), que se constituem de técnicas aplicadas a reconhecimento de fala.
- Utilizar-se das informações deste trabalho a respeito das componentes de alta frequência para promover uma pesquisa que se aprofunde neste tema e possa identificar de forma mais precisa as frequências onde os sons cardíacos estão localizados.
- Propor uma nova base de dados de sons cardíacos com um número de áudios expressivo, que abranja diversos ambientes hospitalares, bem como executar as gravações de sons cardíacos em estetoscópios de diferentes fabricantes e características.
- Implementar um sistema classificador que seja capaz de identificar sons cardíacos patológicos juntamente com o tipo de sopro do paciente.
- Propor um produto, constituído por um estetoscópio que possua um sistema embarcado, onde executa-se um algoritmo de classificação de sons cardíacos, proporcionando um diagnóstico em tempo real.

A partir das sugestões de trabalhos futuros, bem como os próprios resultados deste trabalho, vislumbram-se grandes oportunidades de desenvolvimento de aplicações voltadas a saúde, onde a cada dia mais utiliza-se a tecnologia a favor de salvar vidas. O sistema classificador proposto neste trabalho pode servir como base para o desenvolvimento de produtos que auxiliem os profissionais de saúde no atendimento primário, de forma a sugerir um diagnóstico que em caso de classificação positiva, encaminhe-se o paciente a um cardiologista que poderia realizar uma investigação mais acurada da situação. Desta forma, o sistema colaborará para o diagnóstico precoce de patologias cardíacas, que atualmente são a maior causa de

mortalidade do mundo e carecem por tecnologias que auxiliem os profissionais da saúde a combatê-las.

REFERÊNCIAS

WORLD HEALTH ASSOCIATION. **The top 10 causes of death**. 2020. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Acesso em: 16 jun. 2022.

HARVARD MEDICAL SCHOOL. **Heart Murmur**. 2022. Disponível em: https://www.health.harvard.edu/a_to_z/heartmurmuratoz#:~:text=About%2010%25%20of%20adults%20and,produced%20by%20a%20normal%20heart.. Acesso em: 16 jun. 2022.

HALL, John E.; GUYTON, Arthur C. **Tratado de fisiologia médica**. 12. ed. Rio de Janeiro: Elsevier, 2011. xxi, 1151 p. ISBN 9788535237351.

MEYER, Theo e. **Auscultation of cardiac murmurs in adults**. 2020. Disponível em: https://www.uptodate.com/contents/auscultation-of-cardiac-murmurs-inadults?search=Auscultation%20of%20cardiac%20murmurs%20in%20adults&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1. Acesso em: 25 mar. 2022.

THOMPSON, W. Reid; REINISCH, Andreas J.; UNTERBERGER, Michael J.; SCHRIEFL, Andreas J.. Artificial Intelligence-Assisted Auscultation of Heart Murmurs: validation by virtual clinical trial. **Pediatric Cardiology**, [S.L.], v. 40, n. 3, p. 623-629, 12 dez. 2018. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s00246-018-2036-z>.

WANG, Jou-Kou; CHANG, Yun-Fan; TSAI, Kun-Hsi; WANG, Wei-Chien; TSAI, Chang-Yen; CHENG, Chui-Hsuan; TSAO, Yu. **Automatic recognition of murmurs of ventricular septal defect using convolutional recurrent neural networks with temporal attentive pooling**. Scientific Reports, [S.L.], v. 10, n. 1, p. 1-10, dez. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41598-020-77994-z>.

TAGASHIRA, Mizuho; NAKAGAWA, Takafumi. **Biometric Authentication Based on Auscultated Heart Sounds in Healthcare**. *International Journal Of Computer Science*, [s. l.], v. 47, n. 3, p. 343-349, set. 2020.

BOULARES, Mehrez; ALOTAIBI, Reem; ALMANSOUR, Amal; BARNAWI, Ahmed. **Cardiovascular Disease Recognition Based on Heartbeat Segmentation and Selection Process**. International Journal Of Environmental Research And Public Health, [S.L.], v. 18, n. 20, p. 10952, 18 out. 2021. MDPI AG. <http://dx.doi.org/10.3390/ijerph182010952>.

BEGIC, Edin; POKVIC, Lejla Gurbeta; BEGIC, Zijo; BEGIC, Nedim; DEDIC, Mirza; MRSIC, Denis; JAMAKOVIC, Mesud; VILA, Naim; BADNJEVIC, Almir. **From heart murmur to echocardiography ± congenital heart defects diagnostics using machinelearning algorithms**. Medicina Academica Mostariensia, Zagreb, v. 9, n. 2, p. 236-246, dez. 2021.

LI, Haixia; REN, Guofeng; YU, Xiaopu; WANG, Dongying; WU, Shizhe. **Discrimination of the Diastolic Murmurs in Coronary Heart Disease and in Valvular Disease**. IEEE Access, [S.L.], v. 8, p. 160407-160413, 2020. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/access.2020.3021093>.

AHMAD, M. Sheraz; MIR, Junaid; ULLAH, Muhammad Obaid; SHAHID, Muhammad Laiq Ur Rahman; SYED, Muhammad Adnan. **An efficient heart murmur recognition and cardiovascular disorders classification system**. Australasian Physical & Engineering Sciences In Medicine, [S.L.], v. 42, n. 3, p. 733-743, 16 jul. 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s13246-019-00778-x>.

HO, Wen-Hsien; CHEN, Yenming J.; ZHANG, Yuzhen; TAO, Yanyun; KUO, Hsin-Wen. **Heart diseases detection from noisy recordings of Smartphone devices**. Journal Of Mechanics In Medicine And Biology, [S.L.], v. 18, n. 04, p. 1850039, jun. 2018. WorldScientific Pub Co Pte Lt. <http://dx.doi.org/10.1142/s0219519418500392>.

DAOUD, Boutana; NAYAD, Kouras; BRAHAM, Barkat; MESSAOUD, Benidir. **Heart murmurs detection and characterization using wavelet Analysis with renyi entropy.** *Journal Of Mechanics In Medicine And Biology*, [S.L.], v. 17, n. 06, p. 1750093, set. 2017. World Scientific Pub Co Pte Lt. <http://dx.doi.org/10.1142/s0219519417500932>.

WANG, Jiaming; YOU, Tao; YI, Kang; GONG, Yaqin; XIE, Qilian; QU, Fei; WANG, Bangzhou; HE, Zhaoming. **Intelligent Diagnosis of Heart Murmurs in Children with Congenital Heart Disease.** *Journal Of Healthcare Engineering*, [S.L.], v. 2020, p. 1-9, 11 maio 2020. Hindawi Limited. <http://dx.doi.org/10.1155/2020/9640821>.

TIWARI, Shamik; JAIN, Anurag; SHARMA, Akhilesh Kumar; ALMUSTAFA, Khaled Mohamad. **Phonocardiogram Signal Based Multi-Class Cardiac Diagnostic Decision Support System.** *IEEE Access*, [S.L.], v. 9, p. 110710-110722, 2021. Institut of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/access.2021.3103316>.

CHOWDHURY, Muhammad E.H.; KHANDAKAR, Amith; ALZOUBI, Khawla; MANSOOR, Samar; TAHIR, Anas M.; REAZ, Mamun Bin Ibne; AL-EMADI, Nasser. **Real-Time Smart-Digital Stethoscope System for Heart Diseases Monitoring.** *Sensors*, [S.L.], v. 19, n. 12, p. 2781, 20 jun. 2019. MDPI AG. <http://dx.doi.org/10.3390/s19122781>.

TAKAHASHI, Kiichi; ONO, Kyoichi; ARAI, Hirokazu; ADACHI, Hiroyuki; ITO, Masato; KATO, Akie; TAKAHASHI, Tsutomu. **Detection of Pathologic Heart Murmurs Using a Piezoelectric Sensor.** *Sensors*, [S.L.], v. 21, n. 4, p. 1376, 16 fev. 2021. MDPI AG. <http://dx.doi.org/10.3390/s21041376>.

OLIVEIRA, Jorge; RENNA, Francesco; COSTA, Paulo Dias; NOGUEIRA, Marcelo; OLIVEIRA, Cristina; FERREIRA, Carlos; JORGE, Alipio; MATTOS, Sandra; HATEM, Thamine; TAVARES, Thiago. **The CirCor DigiScope Dataset: from murmur detection to murmur classification.** *IEEE Journal Of Biomedical And Health Informatics*, [S.L.], v. 26, n. 6, p. 2524-2535, jun. 2022. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/jbhi.2021.3137048>

CHOWDHURY, Tanzil Hoque; POUDEL, Khem Narayan; HU, Yating. Time-Frequency Analysis, Denoising, Compression, Segmentation, and Classification of PCG Signals. **Ieee Access**, [S.L.], v. 8, p. 160882-160890, 2020. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/access.2020.3020806>.

GERSH, Bernard J. **Physiologic and pharmacologic maneuvers in the differential diagnosis of heart murmurs and sounds**. 2021. Disponível em: https://www.uptodate.com/contents/physiologic-and-pharmacologic-maneuvers-in-the-differential-diagnosis-of-heart-murmurs-and-sounds?search=Physiologic%20and%20pharmacologic%20maneuvers%20in%20the%20differential%20diagnosis%20of%20heart%20murmurs%20and%20sounds&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1. Acesso em: 01 abr. 2022.

MEYER, Theo e. **Auscultation of heart sounds**. 2021. Disponível em: https://www.uptodate.com/contents/auscultation-of-heart-sounds?search=Auscultation%20of%20heart%20sounds&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1. Acesso em: 05 abr. 2022

DATA SCIENCE ACADEMY. **Deep Learning Book Brasil**. Disponível em: https://www.deeplearningbook.com.br/?gclid=Cj0KCQjwpv2TBhDoARIsALBnVnIDB2vxDsIODEyWf6FnmMAhSM5OgJqIFiX9CKQ4NZ4MQas2b3sqtoaAtigE Aw_wcB. Acesso em: 02 maio 2022.

HAYKIN, Simon S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2007. 1 recurso online. ISBN 9788577800865.

FACELI, Katti et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011. xvi, 378 p. ISBN 9788521621461.

J. Jo, H. Yoo and I. -C. Park, "**Energy-Efficient Floating-Point MFCC Extraction Architecture for Speech Recognition Systems**," in **IEEE Transactions on Very Large Scale Integration (VLSI) Systems**, vol. 24, no. 2, pp. 754-758, Feb. 2016, doi: 10.1109/TVLSI.2015.2413454.Q.

LI, Qin; YANG, Yuze; LAN, Tianxiang; ZHU, Huifeng; WEI, Qi; QIAO, Fei; LIU, Xinjun; YANG, Huazhong. **MSP-MFCC: energy-efficient mfcc feature extraction method with mixed-signal processing architecture for wearable speech recognition applications**. *Ieee Access*, [S.L.], v. 8, p. 48720-48730, 2020. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/access.2020.2979799>.

T. Villa-Cañas et al., **Automatic detection of laryngeal pathologies using cepstral analysis in Mel and Bark scales**, 2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA), 2012, pp. 116-121, doi: 10.1109/STSIVA.2012.6340567.

Z. M. Amin, **Detection of migraine in EEG by joint time-frequency representation**, Student Conference on Research and Development, 2002, pp. 494-497, doi: 10.1109/SCORED.2002.1033165.

SANTOS, S.C.B.; ALCAIM, A. **Um Sistema de Reconhecimento de Voz Contínua Dependente da Tarefa em Língua Portuguesa.** Journal Of Communication And Information Systems, [S.L.], v. 17, n. 2, p. 135-147, 30 dez. 2002. Sociedad Brasileira de Telecomunicacoes. <http://dx.doi.org/10.14209/jcis.2002.6>.

E. HALL, John; E. HALL, Michael. **Text Book of Medical Physiology.** 14. ed. Philadelphia: Elsevier, 2021. 1083 p.

G. D. Clifford et al., "**Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016,**" 2016 Computing in Cardiology Conference (CinC), 2016, pp. 609-612

SHAW, George Bernard. **Frases célebres para monografias, dissertações e teses.** Disponível em: <https://posgraduando.com/frases-celebres-para-monografias-dissertacoes-e-teses/>. Acesso em: 03 nov. 2022.

APÊNDICE A – MATRIZES DE CONFUSÃO DOS GRUPOS DOS TESTES DE VALIDAÇÃO CRUZADA

Quadro 12 – Matriz de confusão do Grupo 1

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,9500	0,0500
Normal	0,1266	0,8734
Precisão	0,8837	
Acurácia	0,9119	

Fonte: Elaborado pelo autor.

Quadro 13 – Matriz de confusão do Grupo 2

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,9156	0,0844
Normal	0,1159	0,8841
Precisão	0,8812	
Acurácia	0,8994	

Fonte: Elaborado pelo autor.

Quadro 14 – Matriz de confusão do Grupo 3

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8408	0,1592
Normal	0,1125	0,8875
Precisão	0,8800	
Acurácia	0,8644	

Fonte: Elaborado pelo autor.

Quadro 15 – Matriz de confusão do Grupo 4

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8875	0,1125
Normal	0,1401	0,8599
Precisão	0,8659	
Acurácia	0,8738	

Fonte: Elaborado pelo autor.

Quadro 16 – Matriz de confusão do Grupo 5

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8750	0,1250
Normal	0,1206	0,8794
Precisão	0,9006	
Acurácia	0,8770	

Fonte: Elaborado pelo autor.

Quadro 17 – Matriz de confusão do Grupo 6

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8987	0,1013
Normal	0,1195	0,8805
Precisão	0,8837	
Acurácia	0,8896	

Fonte: Elaborado pelo autor.

Quadro 18 – Matriz de confusão do Grupo 7

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,8662	0,1338
Normal	0,0625	0,9375
Precisão	0,9315	
Acurácia	0,9022	

Fonte: Elaborado pelo autor.

Quadro 19 – Matriz de confusão do Grupo 8

Classe Verdadeira	Classe Predita	
	Anormal	Normal
Anormal	0,9437	0,0563
Normal	0,0955	0,9045
Precisão	0,9096	
Acurácia	0,9243	

Fonte: Elaborado pelo autor.

Quadro 20 – Matriz de confusão do Grupo 9

Classe Verdadeira	Classe Preditada	
	Anormal	Normal
Anormal	0,8608	0,1392
Normal	0,0755	0,9245
Precisão	0,9189	
Acurácia	0,8927	

Fonte: Elaborado pelo autor.

Quadro 21 – Matriz de confusão do Grupo 10

Classe Verdadeira	Classe Preditada	
	Anormal	Normal
Anormal	0,8767	0,1233
Normal	0,0585	0,9415
Precisão	0,9275	
Acurácia	0,9117	

Fonte: Elaborado pelo autor.