Universidade de Brasília – UnB
Campus Gama – FGA
Engenharia Eletrônica

Análise de ausculta pulmonar para ferramentas De Auxílio a Diagnóstico com base em Redes Neurais Convolucionais e Características Cepstrais

Haniel Rodrigues Guimarães de Lima

Orientador: Dr. Cristiano Jacques Miosso

UnB – Universidade de Brasília

FGA – Faculdade Gama

Engenharia Eletrônica

Análise de ausculta pulmonar para ferramentas De Auxílio a Diagnóstico com base em Redes Neurais Convolucionais e Características Cepstrais

Haniel Rodrigues Guimarães de Lima

Orientador: Cristiano Jacques Miosso

Trabalho de Conclusão de Curso
Engenharia Eletrônica

Brasília/DF, Fevereiro de 2025

UnB – Universidade de Brasília

FGA – Faculdade Gama

Engenharia Eletrônica

Análise de ausculta pulmonar para ferramentas De Auxílio a Diagnóstico com base em Redes Neurais Convolucionais e Características Cepstrais

Haniel Rodrigues Guimarães de Lima

Trabalho de Conclusão de Curso submetido à Faculdade UnB Gama da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do grau de bacharel em engenharia eletrônica

Aprovado por:

Cristiano Jacques Miosso

(Orientador)

Prof Dr. Fabiano Araujo Soares

Prof. Dr. Adson Ferreira da Rocha

Brasília/DF, Fevereiro de 2025

Ficha Catalográfica

De Lima, Haniel Rodrigues Guimarães
Análise de ausculta pulmonar para ferramentas
De Auxílio a Diagnóstico com base em Redes
Neurais Convolucionais e Características Cepstrais,
[Distrito Federal], 2025.
49p., 210 × 297 mm (FGA/UnB Gama, Bacharelado em Engenharia Eletrônica, 2022).
Trabalho de Conclusão de Curso, Faculdade UnB Gama, Engenharia Eletrônica

1. Processamento de Sinal
2. MFCC de Auscultas Cardiopulmonares
3. *Compressive Sensing*
4. Processamento de Sinais

I. FGA UnB/UnB.
II. Título (série)

Referência

De Lima, Haniel Rodrigues Guimarães (2025). Análise de ausculta pulmonar para Ferramentas de Auxílio a Diagnóstico com Base em Redes Neurais Convolucionais e Características Cepstrais. Trabalho de Conclusão de Curso, Engenharia Eletrônica, Faculdade UnB Gama, Universidade de Brasília, Brasília, DF, 49p.

Cessão de Direitos

Autor: Haniel Rodrigues Guimarães de Lima

Título: Análise de ausculta pulmonar para Ferramentas de Auxílio a Diagnóstico com Base em Redes Neurais Convolucionais e Características Cepstrais

Grau: Bacharel em Engenharia Eletrônica

Ano: 2025

É concedida à Universidade de Brasília permissão para reproduzir cópias desta monografia de conclusão de curso e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta monografia pode ser reproduzida sem a autorização por escrito do autor.

---

hanielrlima@gmail.com

Brasília, DF – Brasil

Agradecimentos

Dedico este trabalho, primeiramente, a Deus, que me deu resiliência, sabedoria e perseverança para superar cada desafio ao longo desta jornada. Aos meus pais, que sempre me incentivaram a estudar e acreditaram no meu potencial. À minha amada esposa, Lucília, que esteve ao meu lado com amor, apoio e palavras de encorajamento. À minha irmã, seu esposo e minhas sobrinhas, cuja torcida e carinho foram fonte de motivação. À minha família em geral, que foi minha base e minha maior inspiração.
Aos meus amigos, que compartilharam comigo essa caminhada, em especial ao meu amigo Diego M., companheiro de estudos, risadas e de superação nos momentos desafiadores. Aos meus professores e mentores, cuja paciência, conhecimento e dedicação foram fundamentais para o meu crescimento acadêmico e pessoal. Um agradecimento especial ao professor Cristiano Miosso, por seu profissionalismo, cuidado e respeito, qualidades que admiro profundamente, e a sua forma de passar o conhecimento nas aulas. E a todos aqueles que, de alguma forma, contribuíram para que este sonho se tornasse realidade. Essa conquista não é apenas minha, mas de todos que caminharam ao meu lado nessa trajetória.

Resumo

Esta pesquisa apresenta uma alternativa para o processamento de áudio de ausculta cardiopulmonar, com foco em determinadas patologias pulmonares contidas em uma base de dados. O processamento sonoro envolve a separação dos sons das contrações cardíacas e dos sons dos movimentos pulmonares. Como parâmetro de avaliação dessa abordagem, foi utilizada uma rede neural convolucional em duas situações de entrada de dados: uma com a base de dados original, sem modificações, e outra com a mesma base, porém submetida ao processamento de separação. O objetivo é analisar a efetividade da separação sonora e as possíveis vantagens e desvantagens desse procedimento em uma análise computacional. Por meio da rede neural convolucional, foram extraídos dados estatísticos para avaliar o procedimento e definir ações futuras, possibilitando, posteriormente, uma análise clínica. Foi realizada uma análise gráfica e sonora de amostras obtidas da base de dados utilizada para estudos científicos, empregando espectrogramas para comparar os áudios originais com aqueles processados para separação dos sons. Os áudios da base de dados foram processados, convertidos em gráficos de espectrograma de MFCC e submetidos à extração de características MFCC, que foram então utilizadas como entrada na rede neural convolucional. A partir dos dados estatísticos, foi avaliada a performance da rede tanto com os áudios brutos quanto com aqueles processados, que enfatizavam os sons pulmonares. Os resultados mostraram uma melhora de 5,3% na acurácia da identificação da patologia estudada. No entanto, ressalta-se a necessidade de uma base de dados com um maior número de amostras para aprimorar a identificação das características sonoras de outras patologias. Devido ao desbalanceamento das bases de dados utilizadas, apenas uma patologia foi analisada: a Doença Pulmonar Obstrutiva Crônica (COPD). Quanto à separação dos sons pulmonares e cardíacos, recomenda-se que testes futuros sejam conduzidos com especialistas da área da saúde para avaliar se esse processo contribui para a semiologia médica.

Abstract

This research presents an alternative approach to processing cardiopulmonary auscultation audio, focusing on specific pulmonary pathologies contained in a database. The sound processing involves separating the sounds of cardiac contractions from those of pulmonary movements. As an evaluation parameter for this approach, a convolutional neural network was used in two data input scenarios: one with the original, unmodified database and another with the same database but subjected to the separation process. The objective is to analyze the effectiveness of sound separation and the possible advantages and disadvantages of this procedure in a computational analysis. Through the convolutional neural network, statistical data were extracted to assess the procedure and define future actions, ultimately enabling clinical analysis. A graphical and auditory analysis was performed on samples obtained from the database used for scientific studies, employing spectrograms to compare the original audio recordings with those processed for sound separation. The database audio files were processed, converted into MFCC spectrogram graphs, and subjected to MFCC feature extraction, which was then used as input for the convolutional neural network. Based on statistical data, the network's performance was evaluated using both raw audio files and processed ones that emphasized pulmonary sounds. The results showed a 5.3% improvement in the accuracy of identifying the studied pathology. However, it is important to highlight the need for a database with a larger number of samples to enhance the identification of sound characteristics associated with other pathologies. Due to the imbalance in the databases used, only one pathology was analyzed: Chronic Obstructive Pulmonary Disease (COPD). Regarding the separation of pulmonary and cardiac sounds, it is recommended that future tests be conducted with healthcare specialists to assess whether this process contributes to medical semiology.

Sumário

1 Introdução **1**

1.1 Patologias Pulmonares e os Desafios do Diagnóstico 2

1.2 Desenvolvimento do estetoscópio 4

1.3 O Desenvolvimento da Intenet no Brasil e as Possibilidades Geradas para a Telemedicina 5

1.4 Definição do Problema Científico e Proposta de Pesquisa 10

1.5 Objetivos 11

1.5.1 Objetivo Geral 11

1.5.2 Objetivos Específicos 11

2 Fundamentação teórica e Estado da Arte **12**

2.1 Rede Neural Convolucional - CNN 12

2.2 Ausculta Pulmonar 17

3 Materiais e Métodos **19**

3.1 Implementação de Algoritmos para Decomposição e Classificação de Ausculta Pulmonar 19

3.2 Insumos Utilizados 23

3.3 Procedimentos experimentais 31

3.4 Procedimentos de análise 33

4 Resultados e Discussões **35**

5 Conclusão **44**

5.1 Propostas para Estudos Futuros 45

Lista de Tabelas

3.1 Distribuição dos indivíduos na segunda base de dados por condição de saúde, número de participantes, faixa etária e gênero. 27

3.2 Número de amostras por classe, após o acréscimo das amostras modeladas. 30

4.1 Arquitetura da Rede Neural Convolucional(CNN) utilizada no estudo . . 39

4.2 Métricas de desempenho usando a base de dados sem a separação harmônica. 41

4.3 Métricas de desempenho usando a base de dados sem a separação harmônica. 42

Lista de Figuras

2.1 Imagem de Neurônios Cerebrais feita por microscópio eletrônico de varredura, a partir de uma amostra humana do lobo temporal do córtex cerebral de 1mm cúbico, incorporado por resina e laminado em 5 mil fatias e por meio de métodos computacionais foi reslizado a imagem tridimensional, demostrando a complexidade das redes neurais biológicas. 12

2.2 A representação simplificada do neurônio biológico oferece insights sobre o funcionamento da sinapse neuronal, servindo como base inspiradora para o desenvolvimento dos modelos de neurônios artificiais.. 13

2.3 Representação do neurônio artificial, proposta por McCullock e Pitts em 1943. Com uma função de ativação com limite (threshold), que muda a saída de acordo com um limite pré estabelecido na função(f(a)). Que depende do somatório das entradas(x) multiplicadas por seus respectivos pesos(W). 14

2.4 Representação simplificada das multicamadas de uma rede neural. Cada camada é composta por neurônios interconectados, onde cada conexão representa um peso que é ajustado durante o processo de treinamento da rede. Essas multicamadas são organizadas em três tipos principais: camada de entrada, camadas ocultas e camada de saída. 14

2.5 Representação simplificada da CNN. Possui uma arquitetura especializada, projetada para processar dados com estrutura de grade, como imagens e detecção de padrões.Utiliza camadas convolucionais para extrair características importantes das entradas. 17

3.1 Quantidade de pacientes distribuidos pela Faixa Etária, demostrando a heterogeneidade dos arquivos de áudio em relação a idade do paciente. . 24

3.2 Quantidade de pacientes distribuidos pelas patologias diagnosticadas, demostrando o desbalanciamento dos arquivos de áudio em relação as patologias dos pacientes. 26

3.3 Quantidade de pacientes distribuidos pelas patologias diagnosticadas, demostrando o desbalanciamento dos arquivos de áudio em relação as patologias dos pacientes. 28

3.4 Quantidade de pacientes distribuidos pelas patologias diagnosticadas das quatro amostras com maior incidência oriundo da união das duas bases de dados, demostrando o desbalanciamento dos arquivos de áudio em relação as patologias dos pacientes. 29

3.5 Quantidade de amostras distribuidas pelas patologias diagnosticadas das quatro amostras com maior incidência oriundo da união das duas bases de dados e do acréscimo das amostras modeladas, demostrando o balanciamento dos arquivos de áudio em relação as patologias dos pacientes. . . . 30

3.6 Curva da escala Mel em razão da frequência em HZ, demostrando a sua não linearidade com o aumento da frequência, desenvolvida para ser similar a percepção humana em razão da variação da frequência(Hz). 32

4.1 Onda sonora da ausculta - Em uma representação pelas grandezas de Amplitude da onda pelo Tempo. 36

4.2 Forma de onda sonora da parte harmônica, onde se ameniza os sons de grande amplitude, destacando os sons de média e baixa amplitude. Em uma representação pelas grandezas de Amplitude da onda pelo Tempo. . 36

4.3 Forma de onda sonora da parte percursiva, em que se destaca os sons com uma elevada amplitude e minimização dos sons de baixa e média amplitude. A representação do som está pela amplitude da onda no tempo. 37

4.4 Espectrogramas dos três Áudios - Normal, Harmônico e Percussivo. O espectrograma é uma forma de visualização bidimensional da composição de um som, na qual é possível visualizar a amplitude da onda por uma escala de cores, as frequências pela altura dos pontos e a posição temporal pelo deslocamento lateral. 37

4.5 Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados sem a separação. 40

4.6 Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados com a separação. 41

4.7 Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados sem a separação. 41

4.8 Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados com a separação. 42

4.9 Gráfico da Matriz de Confusão da base de dados sem o processo de separação, onde podemos comparar as previsões do modelo com os resultados reais. 43

4.10 Gráfico da Matriz de Confusão da base de dados com o processo de separação, onde podemos comparar as previsões do modelo com os resultados reais. 43

Lista de Símbolos, Nomenclaturas e Abreviações

3DCRT	–	*3D Conformal Radiotherapy*
CREMESP	–	*Conselho Regional de Medicina de São Paulo*
CFM	–	*Conselho Federal de Medicina*
ESPIN	–	*Emergência em Saúde Pública de Importância Nacional*
EUA	–	*Estados Unidos da América*
SUS	–	*Sistema Único de Saúde*
IBGE	–	*Instituto Brasileiro de Geografia e Estatística*
SARS-CoV-2	–	*Severe Acute Respiratory Syndrome Coronavirus 2*
LNCC	–	*Laboratório Nacional de Computação Científica*
RNP	–	*Rede Nacional de Ensino e Pesquisa*
BITNET	–	*Because It's Time Network*
NJE	–	*Network Job Entry*
TCP/IP	–	*Transmission Control Protocol / Internet Protocol*
FAPESP	–	*Fundação de Amparo à Pesquisa do Estado de São Paulo*
Fermilab	–	*Fermi National Accelerator Laboratory*
UFRJ	–	*Universidade Federal do Rio de Janeiro*
ONU	–	*Organização das Nações Unidas*
UUCP	–	*UNIX-to-UNIX Copy Program*
LGPD	–	*Lei Geral de Proteção de Dados Pessoais*
WHO	–	*World Health Organization*

1 Introdução

Nos últimos anos, tem havido avanços significativos no desenvolvimento de sistemas de auxílio ao diagnóstico de patologias pulmonares, empregando análise automatizada de sinais de ausculta pulmonar.

Esses avanços são impulsionados por tecnologias que utilizam algoritmos de aprendizado de máquina, processamento de sinais e sensores avançados, proporcionando uma abordagem mais eficiente e precisa na detecção de condições pulmonares.

Este trabalho tem como objeto de estudo o uso de algoritmos de processamento de áudio e redes neurais convulsionais (CNN), em sons oriundos de bases de dados de auscultas cardiopulmonares adquiridas por estetoscópios digitais, análise dos resultados obtidos após o processamento desses algoritmos, bem como esses algoritmos poderiam auxiliar no desenvolvimento da telessaúde.

O uso de algoritmos de processamento de sinais permite a identificação de padrões acústicos específicos associados a diversas patologias como pneumonia, bronquite e até mesmo doenças mais complexas, como fibrose pulmonar.

A implementação dessas tecnologias tem contribuído significativamente para a melhoria da precisão diagnóstica de modo a permitir uma detecção precoce de patologias pulmonares.

A análise automatizada de sinais de ausculta pulmonar não apenas agiliza o processo de diagnóstico como também reduz a dependência da habilidade individual do profissional de saúde, tornando o diagnóstico mais acessível em diversas configurações de atendimento médico. Não só isso, mas também essa tecnologia pode aperfeiçoar o uso da telemedicina em regiões remotas, permitindo a avaliação de pacientes à distância, facilitar o acompanhamento simultâneo de múltiplos pacientes em ambientes hospitalares e possibilitar a coleta contínua de dados, indo além da ausculta pontual, o que contribui para uma avaliação continua e dinâmica de acordo com a evolução patológica do paciente.

Além disso, a capacidade desses sistemas em processar grandes volumes de dados de forma rápida e eficiente contribui para a geração de *insights* valiosos, auxiliando na compreensão aprofundada de padrões acústicos associados a diferentes condições pulmonares.

1.1 Patologias Pulmonares e os Desafios do Diagnóstico

Na área de patologias pulmonares ainda persistem desafios como a necessidade contínua de melhorar a eficácia dos diagnósticos, a diversidade de manifestações clínicas e a sobreposição de sintomas entre diferentes condições pulmonares.
Em uma consulta convencional em que o profissional da saúde e o paciente estão presentes, o procedimento do exame físico do aparelho respiratório compreende em palpação, percussão, inspeção e ausculta.
Atualmente, em um atendimento virtual, os procedimentos de palpação e percussão ainda não são possíveis, devido à necessidade de contato com o paciente e o conhecimento especializado da técnica do exame.
No entanto, o procedimento de inspeção do tórax, que é separado em dinâmico e estático, permite que a inspeção unicamente do tipo estática seja possível em uma consulta virtual por meio da videoconferência, pois consiste em observar o formato do tórax, simetria torácica, ginecomastia, presença de abaulamentos e retrações, circulação colateral tipo cava superior e lesões de pele.
Já a inspeção do tórax dinâmica trata-se da observação do padrão e ritmo respiratório, a presença de sinais de esforço respiratório, a expansividade e a frequência respiratória. Nessa inspeção só a análise por videoconferência não seria suficiente; no entanto, com o uso do algoritmo proposto neste trabalho em um dispositivo adequado, a referida inspeção poderá apresentar um potencial significativo e melhorias substanciais na obtenção dos parâmetros para execução de tal procedimento.
Palpação, percussão e inspeção não são objetos de estudo deste trabalho, mas apenas o estudo do procedimento de ausculta pulmonar, que é a principal técnica semiológica pulmonar, ou seja, uma parte importante do exame físico realizado por profissionais da saúde para identificar os sons pulmonares produzidos pelo paciente.

Através da ausculta é possível ter um direcionamento na formulação da hipótese do diagnóstico. Os sons respiratórios habituais que podemos escutar nesta fase da avaliação são gerados quando o ar flui de forma turbulenta e entra em contato com as paredes e protuberâncias nas divisões da árvore brônquica. É importante avaliar esses sons com relação à sua frequência, qualidade e predominância durante o ciclo de respiração.
Os sons respiratórios anormais são conhecidos como ruídos adventícios e, ainda que os ruídos anormais possam surgir ocasionalmente em pulmões de pessoas saudáveis, sua presença constante em respirações consecutivas indica uma anormalidade subjacente.
Esses ruídos anormais podem ser classificados em sons descontínuos e contínuos. Os sons respiratórios descontínuos são breves, com duração inferior a 20 milissegundos, não apresentam uma característica musical e são conhecidos como crepitações, as quais podem

ter dois tipos: finas, que são curtas, agudas e menos intensas, e grossas, que têm uma duração maior, sendo graves e mais intensas.
Os sons respiratórios contínuos não implicam que o som seja constante durante todo o ciclo respiratório, mas sim que sua duração sonora seja de 250 milissegundos ou mais dentro desse ciclo. E essa diferenciação de duração é feita verificando-se o registro do som em relação ao tempo, por exemplo, pela análise de gráficos. Esses sons têm características musicais, conforme a seguir. Os sons contínuos incluem os roncos e os sibilos [1] e os sons de patologias pulmonares são nomeados com cornagem ou estridor, sibilo, ronco, estertor, crepitante ou fino, estertor subcrepitante ou bolhoso, grasnido, atrito pleural e sopro tubário.
De acordo com pesquisas da *World Health Organization* (WHO), referentes aos anos de 2000 a 2019, a doença pulmonar obstrutiva crônica e as infecções respiratórias inferiores são a terceira e quarta causa de morte no mundo, chegando a 5,8 milhões de mortes no planeta só no ano de 2019 [2], o que destaca a importância em ter um olhar direcionado para doenças pulmonares e o estudo dessa área de interesse.
Entre os tipos mais comuns de patologias pulmonares, podemos destacar algumas como: asma, pneumonia, bronquite, bronquiolite, tuberculose, apneia do sono, doença pulmonar obstrutiva crônica (DPOC), enfisema pulmonar, doenças intersticiais, fibrose cística, embolia pulmonar, hipertensão pulmonar, displasia broncopulmonar, câncer de pulmão. E cada uma dessas patologias vai apresentar sintomas e sinais diferentes, ao passo que o fator que provoca essas condições também pode ser totalmente diferente. A tuberculose, por exemplo, é uma patologia causada por uma bactéria de nome *Mycobacterium tuberculis*, e a pneumonia pode ser causada por vírus, bactéria, inalação excessiva de produtos químicos, aspiração de vômito e fungos e em uma embolia pulmonar pode ser desenvolvida em razão de efeitos colaterais de medicamentos, ao passo que a asma ocorre por causa de uma alteração na sensibilidade dos brônquios.

Uma patologia que é importante é a Doença Pulmonar Obstrutiva Crônica (DPOC) este um termo engloba condições pulmonares progressivas, como enfisema e bronquite crônica, que causam dificuldades respiratórias [3]. O enfisema envolve danos aos sacos de ar nos pulmões, enquanto a bronquite crônica refere-se à inflamação prolongada das vias aéreas [3]. O principal fator de risco para DPOC é o tabagismo, responsável por mais de 70% dos casos em países de alta renda [4]. Outros fatores incluem exposição prolongada a poluentes ambientais, como fumaça de combustíveis sólidos, e fatores genéticos, como a deficiência de alfa-1 antitripsina [4]. // A DPOC causa sintomas como falta de ar, tosse persistente com produção de muco e chiado [4]. Esses sintomas podem limitar atividades diárias e levar a exacerbações graves. Globalmente, a DPOC foi responsável por 3,5 milhões de mortes em 2021, representando aproximadamente 5% de todas as

mortes [4]. Embora os danos pulmonares causados pela DPOC sejam permanentes, tratamentos podem aliviar os sintomas e melhorar a qualidade de vida [5] A prevenção da DPOC é possível principalmente evitando o tabagismo e reduzindo a exposição a poluentes ambientais. Diagnóstico precoce e tratamento adequado são essenciais para melhorar o prognóstico e a qualidade de vida dos pacientes [5].//

O processo de ausculta por meio do estetoscópio é um elemento que direciona o especialista para chegar a um diagnóstico, principalmente se for uma patologia pulmonar. Com o recurso da ausculta estando disponível para o uso na telessaúde, o atendimento on-line dará um passo para uma melhora nos serviços prestados ao paciente, podendo trazer uma contribuição na qualidade e nas informações coletadas pelo especialista, com a possibilidade de levar isso para regiões isoladas.
No período da pandemia de SARS-CoV-2 - *Severe Acute Respiratory Syndrome Coronavirus 2* a necessidade de tecnologias como essa ficou evidente, pois, se a cultura de telemedicina fosse mais difundida no Brasil, muitas pessoas poderiam ser atendidas por especialistas utilizando a telessaúde. Nesse sentido, é preciso aprimorar os estudos de tecnologias e equipamentos que podem compor uma consulta de telessaúde com o objetivo de abranger o acesso à saúde da população e melhorar a qualidade da teleconsulta.
Além disso, questões éticas e de privacidade devem ser cuidadosamente consideradas no desenvolvimento e implementação desses sistemas, garantindo a proteção dos dados do paciente e o respeito às normas regulatórias.
O sistema de auxílio no diagnóstico representa uma promissora fronteira na medicina, oferecendo benefícios significativos para profissionais de saúde e pacientes, ao mesmo tempo que demanda atenção cuidadosa a desafios técnicos e éticos.

1.2 Desenvolvimento do estetoscópio

O estetoscópio é um instrumento médico essencial utilizado para auscultar os sons internos do corpo humano, especialmente os sons do coração e dos pulmões. A história do estetoscópio remonta ao início do século XIX. Os parágrafos seguintes apresentarão uma breve visão geral da sua evolução:
Invenção por René Laennec (1816): O estetoscópio foi inventado por René Théophile Hyacinthe Laennec, um médico francês, em 1816. Na época, a prática comum para auscultar os sons do corpo envolvia colocar o ouvido diretamente no peito do paciente. Laennec desenvolveu o estetoscópio para superar a dificuldade de examinar pacientes obesos ou de gênero oposto [6]. Seu primeiro estetoscópio era um simples tubo de papel dobrado.
Ao longo dos anos, o estetoscópio passou por várias modificações e melhorias. A versão

original de Laennec foi aprimorada por outros médicos e inventores, substituindo o tubo de papel por materiais mais resistentes e leves. Em meados do século XIX, George Cammann, um médico americano, introduziu melhorias significativas, incluindo um estetoscópio de duas extremidades [7].
Estetoscópio Biauricular (1900s): No início do século XX, o estetoscópio biauricular, com duas extremidades conectadas a tubos que se ligam aos ouvidos, tornou-se padrão. Este design é mais semelhante aos estetoscópios modernos.
Materiais e Tecnologia Modernos: Avanços contínuos na tecnologia e materiais levaram a melhorias adicionais no design e na funcionalidade dos estetoscópios. Materiais como o aço inoxidável se tornam comuns e os estetoscópios agora vêm em diferentes modelos e estilos para atender às necessidades específicas de diferentes especialidades médicas.
Mais recentemente surgiram estetoscópios eletrônicos e digitais. Esses dispositivos podem amplificar os sons corporais, permitindo uma análise mais detalhada. Além disso, eles podem ter recursos de gravação para documentação e compartilhamento de resultados.
Hoje em dia o estetoscópio continua sendo uma ferramenta fundamental para médicos e profissionais de saúde, desempenhando um papel crucial no diagnóstico e monitoramento de condições médicas.

1.3 O Desenvolvimento da Intenet no Brasil e as Possibilidades Geradas para a Telemedicina

O primórdio da Internet no Brasil deu-se em 1992, quando foi implementada a primeira rede de Internet brasileira. A rede tinha fins acadêmicos e conectava dez estados e o Distrito Federal, com a capacidade na transferência de dados de 64Kb/s. Mas desde 1988 o Laboratório Nacional de Computação Científica (LNCC), localizado em Petrópolis – RJ, já tinha estabelecido uma conexão com a Universidade de Maryland, por meio da rede BITNET - *(Because It's Time Network)* que era uma rede que permitia a troca de mensagens e transferência de arquivos entre os seus usuários. O protocolo de comunicação utilizado era o NJE - *(Network Job Entry)* e não o protoloco TCP/IP – *(Transmission Control Protocol / Internet Protocol)* usado hoje em dia, mas esse protocolo também permitia trocar informações via Internet [8].
Outras instituições brasileiras também realizaram conexões, como a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que se conectou com o *Fermi National Accelerator Laboratory* (Fermilab) um laboratório localizado em Chicago, Illinois, EUA, e a Universidade Federal do Rio de Janeiro (UFRJ) que se conectou a rede BITNET, sendo essa a terceira instituição brasileira a ter acesso a essa tecnologia. Mas a rede

brasileira só começou a ser criada depois da realização da Rio-92, que foi uma das principais conferências ambientais organizada pela Organização das Nações Unidas (ONU). Na Conferência das Nações Unidas para o Meio Ambiente e Desenvolvimento, realizada no Rio de Janeiro-RJ, que tinha como principal objetivo o debate sobre questões do meio ambiental e o conceito de desenvolvimento sustentável.
Nesse cenário, era do interesse dos organizadores que, durante a Rio-92, o acesso à Internet fosse garantido. Com isso, foi feito um grande investimento em equipamentos e infraestrutura para que essa exigência fosse alcançada. Ficando de legado desse evento toda essa estrutura de rede de Internet que já estava operando com o protocolo UUCP - *(UNIX-to-UNIX Copy Program)*, conhecido como Internet discada. A partir desse investimento colaborativo, possibilitou-se o desenvolvimento da primeira rede de Internet no Brasil, a qual conectou inicialmente dez estados e o Distrito Federal. Essa rede de Internet era somente para fins acadêmicos, com conexão entre pesquisadores brasileiros e internacionais [9].
Só em maio de 1995 ocorreu a abertura da Internet comercial para a população brasileira com a expansão de infraestrutura da rede de internet, realizando, assim, conexões com outras capitais brasileiras e cabeamentos para os EUA , saindo tanto de Brasília e de São Paulo e atingindo uma velocidade de transferência de dados de 2Mb/s. Em 2000 foi inaugurado o backbone da Rede Nacional de Ensino e Pesquisa (RNP) 2, dando início a uma infraestrutura de rede avançada capaz de fornecer conexões de Banda Larga, melhorando significantemente a transferência de dados, permitindo pela primeira vez uma transmissão de vídeo em modo online, ou seja, a taxa de transmissão e recepção era igual o maior que a taxa de visualização do vídeo [10].

Nessa época e com o rápido crescimento da internet o Conselho Regional de Medicina de São Paulo (CREMESP) percebeu a necessidade de definir algumas recomendações aos médicos e esclarecer os direitos dos usuários da internet no tocante ao uso dela com fins de busca de informações, produtos, serviços e saúde on-line. A entidade elaborou a primeira resolução sobre o tema, a Resolução nº 97 de 2001 da regional CREMESP, que proibia o uso da Internet como instrumento de consultas médicas, diagnósticos clínicos, prescrição de medicamentos ou tratamento do paciente. Somente era permitido usar a internet de modo informativo e para orientações de saúde que fossem de utilidade pública visando cuidados pessoais, bem-estar, higiene, nutrição e para soluções de problemas de saúde coletiva, como dúvidas sobre vacina e doenças transmissíveis. Tudo isso se justificava em razão das limitações da internet e a preocupação para não se ferissem os princípios éticos da medicina [11].
No ano seguinte, em 2002, o Conselho Federal de Medicina (CFM), com base no instrumento jurídico da Resolução nº 1643/2002 do CFM, define pela primeira vez o que é

a telemedicina, a saber: é um exercício da medicina por meio da utilização de recursos de comunicação de dados e audiovisuais, tendo o objetivo de educação, pesquisa e assistência. Além disso, definiu-se também como deve ser a segurança dos dados do paciente e a garantia do sigilo profissional, respeitando a Lei Geral de Proteção de Dados Pessoais (LGPD) que afirma ser necessário proteger os direitos fundamentais de liberdade e de privacidade de cada indivíduo. No entanto, a resolução só trata da definição de telemedicina e esclarece as diretrizes de como se deve proceder o atendimento do paciente. Nessa resolução, não é permitida a utilização da telemedicina no território brasileiro [12].

Com o desenvolvimento da internet, o emprego de novas facilidades de comunicação, as interações interpessoais, o acesso à informação e o crescimento das infraestruturas de rede, o acesso à internet em 2018, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), havia alcançado o índice de 79,1% dos domicílios no Brasil. Observando a constante inovação e desenvolvimento de novas tecnologias de comunicação e informação e constatando que essas tecnologias da época já eram suficientes para um intercâmbio entre os próprios médicos e pacientes, o CFM regulamentou essa matéria por meio da Resolução CFM nº 2.227/2018. A partir disso, a telemedicina tornou-se um modo de prestação de serviços médicos mediados por tecnologia, bem como definidas as disciplinas que devem ser observadas e como devem ser realizadas.

Por conseguinte, foi permitido o seu exercício em território nacional, descrevendo vários ramos da telemedicina, a saber: teleconsulta, teleatendimento, teleinterconsulta, telediagnóstico, telecirurgia, teleconferência de ato cirúrgico, teletriagem, telemonitoramento, teleorientação e teleconsultoria. Todas essas categorias foram descritas a fim de esclarecer o que são, em que circunstâncias podem ser usadas e as responsabilidades dos envolvidos [13]. Essa resolução tinha o potencial de modernizar o atendimento ao público e auxiliar no acesso universal à saúde da população.

No entanto, essa resolução foi recebida pela maioria da classe médica como prejudicial ao código de ética da medicina, ferindo, em especial, o princípio da não maleficência, ou seja, até em que ponto a telemedicina não traria malefícios ao paciente. Com isso, no ano seguinte, por meio da Resolução 2228/2019, o CFM revogou a resolução de 2018 e restabeleceu como vigentes as antigas diretrizes da resolução de 2002 [14].

No entanto, no final de 2019 com o coronavírus *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) que provocou uma pandemia, o governo brasileiro, através do Ministério da Saúde, iniciou o período de Emergência em Segura Pública de Importância Nacional (ESPIN), mediante a Portaria nº188 de 03 de fevereiro de 2020 [15], com o intuito de prover o atendimento de pessoas e os serviços de saúde. Em seguida, foi publicada a Portaria nº 467 de 20 de março de 2020 que, em caráter excepcional e temporário, regulamentava e operacionalizava a telemedicina, bem como aprovada pelo Congresso Nacional e sancionada pelo presidente da República a Lei 13.989 de 15 de abril de 2020. Esses dispositivos legais permitiam o uso da telemedicina em caráter emergencial durante

o período de ESPIN ocasionada pelo coronavírus (SARS-CoV-2), mas só enquanto durasse a crise [16, 17].
A telemedicina foi uma solução para as pessoas terem uma assistência médica, mesmo em condições de isolamento domiciliar; Em 2021 a internet já chegava a 90,0% dos domicílios do país, alcançando 65.620 mil domicílios com acesso a internet[18]. Nesse período, a telemedicina no Brasil avançou bastante nesse tipo de atendimento médico pelo emprego de recursos tecnológicos de transmissão de dados. De acordo com o artigo 6º da Lei 13.989/2020, após o ESPIN o CFM teria a competência na regulamentação da telemedicina. O fim da ESPIN foi anunciado em abril de 2022, acabando oficialmente em maio de 2022 [19].

Atualmente a telemedicina está regulamentada no Brasil por meio da Resolução do CFM nº2.314/2022, publicada dia 5 de maio de 2022, que define a telemedicina como o exercício da medicina mediado por tecnologias digitais de informação e comunicação. Com o intuito de promover a saúde pública em regiões de difícil acesso e em comunidades que têm escassez de atendimento especializado de saúde, a Portaria 1.348 de 02 de junho de 2022, publicada pelo Ministério da Saúde, veio para regulamentar a telemedicina no Sistema Único de Saúde (SUS), com o objetivo de, por meio dessa ferramenta, proporcionar o direito à universalidade da saúde pública [20, 21].

Com essas novas regulamentações, abriram-se a possibilidade de outras profissões também usarem esse método de atendimento, não estando mais restrito à medicina somente, compondo assim, um ecossistema de telessaúde.

Contudo, ainda é difícil realizar alguns procedimentos de forma remota. O protocolo usado atualmente em um atendimento clínico começa com a observação do relato do paciente sobre o histórico dos sintomas que provocaram o desvio do estado de saúde. Com as informações obtidas, o especialista questiona o paciente visando complementar as informações recebidas. Essa parte do atendimento tem por nome anamnese, que, além da História da moléstia atual, compreende também a identificação do paciente, interrogatório sobre diversos aparelhos, queixa e duração, condições socioeconômicas, antecedentes pessoais, familiares e hábitos de vida. Com essas primeiras informações em um contexto de telemedicina, o especialista pode constatar as anormalidades sentidas pelo paciente no exame físico,mas enfrenta grandes limitações para formular um diagnóstico completo, pela própria natureza do tipo de atendimento [22].
Um dos processos empregados pelos especialistas durante um exame físico e a ausculta pulmonar, uma das principais técnicas semiológica pulmonares.

Devido às limitações para realizar exames físicos em uma consulta on-line, observa-se a necessidade de criação de novas tecnologias que venham suprir essas carências, a fim de que as consultas a distância tenham uma eficiência similar às consultas presenciais.
Este trabalho tem foco no estudo de uma proposta de tecnologia para auscultas pulmonares e suporte no diagnóstico por meio de processamento dos sinais obtidos e para auxiliar o especialista na telessaúde.

1.4 Definição do Problema Científico e Proposta de Pesquisa

Os pontos em que essa proposta de trabalho poderá auxiliar usando a diferenciação de sons serão:

- Em treinamento de ausculta pulmonar para discentes que devem desenvolver esta técnica e habilidade de diferenciação sonora.
- Em ambientes com vários ruido externos, para este o incremento de uma supressão de ruido irá contribuir substancialmente.
- Para uma ausculta direcionada aos sons produzidos pelos pulmões ou pelo coração.

E usando a tecnologia proposta por completo, poderá auxiliar nos seguintes pontos:

- Na obtenção de dados fisiológicos sonoros em consultas de telemedicina, corroborando para um melhor atendimento de pacientes remotos.
- Auxiliar os especialistas na caracterização do som pulmonar, auxiliando os diagnósticos feito por eles.
- Na ausculta remota de pacientes neonatais diminuindo as manipulações dos exames físicos, em casos em que a criança fica internada onde é necessário realizar exames físicos a cada 4 ou 6 horas, dependendo da gravidade do caso. Por causa desse curto intervalo de tempo entre os exames, o sono dessas crianças fica prejudicado.
- Na monitoração de vários pacientes em um hospital. O especialista pode monitorar seus pacientes remotamente e em tempo continuo.
- Comparações do exame de ausculta feita pelo especialista em dias diferentes.
- Acompanhamento da evolução dos sintomas, por meio de aferições feitas pelo paciente.
- Monitoramento particular da evolução dos sintomas, assim com o oxímetro, termômetro e medidores de pressão são dispositivos disponíveis para o público, este dispositivo seguiria a mesma linha de disponibilidade.

Assim a proposta de pesquisa desse trabalho é:
A utilização de CNN e de processamento de sinais em áudios de estetoscópios digitais, pode auxiliar na análise da ausculta pulmonar?

1.5 Objetivos

1.5.1 Objetivo Geral

Estudo do uso de processamento de áudio, em sons cardiopulmonares obtidos por estetoscópios digitais, analisando se o processamento provê uma separação de componentes patológicas pulmonares. Com o intuito de separar os sons gerados pelo coração dos gerados pelo pulmão, e o estudo do impacto na classificação utilizando o áudio original em comparação com o áudio pré-processado, e a avaliação será feita com base na verificação do impacto na classificação dos áudios.

1.5.2 Objetivos Específicos

Tendo como foco o objetivo geral, determinados objetivos específicos foram pensados para que isso seja alcançado, conforme a seguir:

- Implementar um sistema de processamento de áudio para a ausculta pulmonar, com base em técnicas de processamento de áudio que são usadas na indústria de áudio, para efeito de separação de componentes de baixa frequência.
- Levantamento de bases publicamente disponíveis de sinais de ausculta cardiorrespiratória, sobretudo associadas a patologias pulmonares.
- Implementar um classificador baseado no estado da arte, no caso a CNN para a classificação das patologias pulmonares.
- Fazer a avaliação de desempenho sistemática do classificador, com base em métricas objetivas de desempenho como: acurácia, precisão e revocação.
- Avaliar comparativamente o desempenho dos classificadores testados e verificar a possibilidade de melhoras com base no pré-processamento inicial.
- Realizar o meio de processamento de sinais a separação dos sons referentes ao coração e ao pulmão
- Fazer uma análise de comparação entre os áudios em estado bruto e os que foram especificados sendo majoritariamente pulmonares, no desempenho da CNN.
- Fazer uma análise estatística para verificar a confiabilidade dos resultados de classificação obtidas e reportar as métricas de desempenho com base nesse testes estatísticos.

2 Fundamentação teórica e Estado da Arte

2.1 Rede Neural Convolucional - CNN

As redes neurais artificiais são concebidas com base nas estruturas complexas das redes neurais biológicas presentes no cérebro humano conforme a figura 2.1. Ao examinar os neurônios cerebrais, observamos uma miríade de conexões neurais, que desempenham um papel fundamental na transmissão de informações e na execução de diversas funções cognitivas e motoras. Essas redes neurais biológicas compartilham características que são importantes para a compreensão e modelagem das redes neurais artificiais.

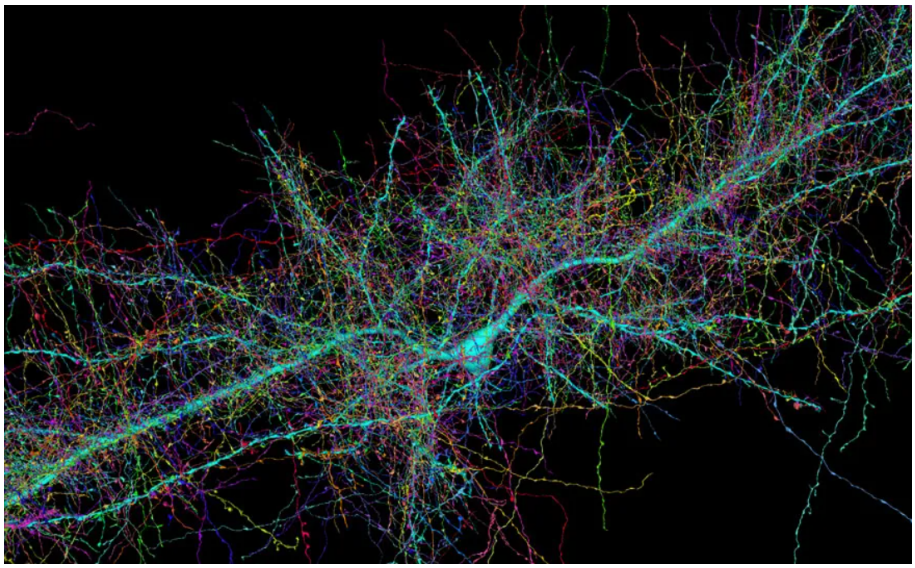


Figura 2.1. Imagem de Neurônios Cerebrais feita por microscópio eletrônico de varredura, a partir de uma amostra humana do lobo temporal do córtex cerebral de 1mm cúbico, incorporado por resina e laminado em 5 mil fatias e por meio de métodos computacionais foi reslizado a imagem tridimensional, demostrando a complexidade das redes neurais biológicas.
Fonte[23]: https://h01-release.storage.googleapis.com/gallery.html. Data: 09/04/2024

Primeiramente, destacamos a vasta quantidade de neurônios presentes no cérebro, que se interconectam formando uma rede complexa e intrincada. Essa interconexão entre neurônios é responsável pela complexidade e pelas habilidades diversas do cérebro humano, tais como percepção visual, linguagem, locomoção, entre outras.

Além disso, é crucial compreender que a comunicação entre os neurônios ocorre por meio dessas conexões, permitindo a troca eficiente de informações. Quando nos expressamos verbalmente, por exemplo, as conexões neurais estão estruturadas de maneira a viabilizar a produção linguística. Da mesma forma, ao aprender um novo idioma, novas conexões neurais são estabelecidas para facilitar a comunicação naquela língua específica.
Portanto, é a troca dinâmica e contínua de informações entre os neurônios que possibilita o processo de aprendizagem e a aquisição de novas habilidades ao longo da vida. Essa capacidade adaptativa das redes neurais, tanto biológicas quanto artificiais, é importante para a o desempenho e desenvolvimento de sistemas inteligentes e de aprendizado de máquina [24]

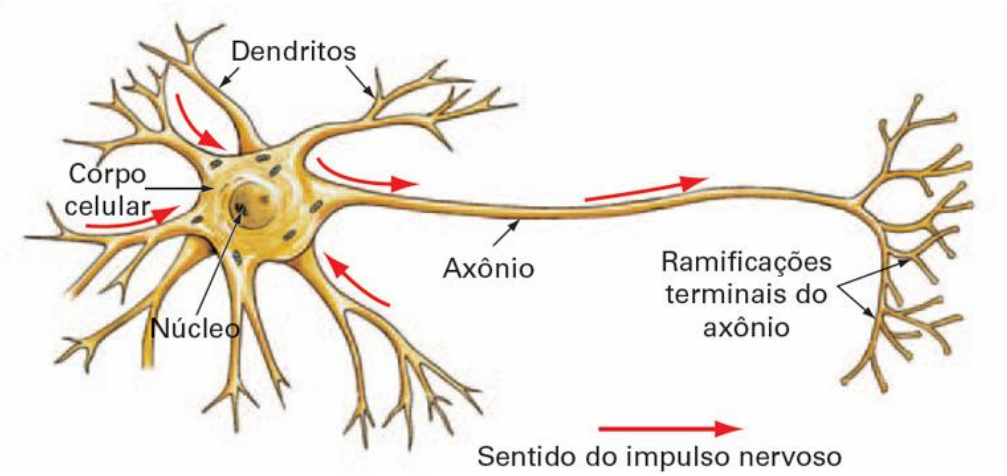


Figura 2.2. A representação simplificada do neurônio biológico oferece insights sobre o funcionamento da sinapse neuronal, servindo como base inspiradora para o desenvolvimento dos modelos de neurônios artificiais..
Fonte [24]: https://www.deeplearningbook.com.br. Data: 09/04/2024

Ao examinarmos um neurônio em destaque na figura 2.2, identificamos vários seguimentos essenciais. Primeiramente, destacam-se os dendritos, que são as estruturas receptoras do neurônio, responsáveis por receber informações de outros neurônios. Em seguida, temos o corpo celular, onde ocorre o processamento dessas informações recebidas. Posteriormente, encontramos o axônio, responsável por transmitir a informação processada até as extremidades do neurônio, conhecidas como terminais do axônio, onde a informação é repassada para outros neurônios.
Essa transferência de informações entre os dendritos e os terminais dos axônios é denominada sinapse nervosa. As sinapses nervosas podem ocorrer de duas maneiras distintas: química ou elétrica.
O neurônio artificial compartilha semelhanças com a estrutura dos neurônios biológicos. Ao observar a representação gráfica na figura 2.3, destacam-se as entradas, cada uma com seu respectivo peso, o processamento do sinal realizado pela função soma e a função de

ativação. O resultado dessas operações é a saída, que será encaminhada para os neurônios subsequentes.

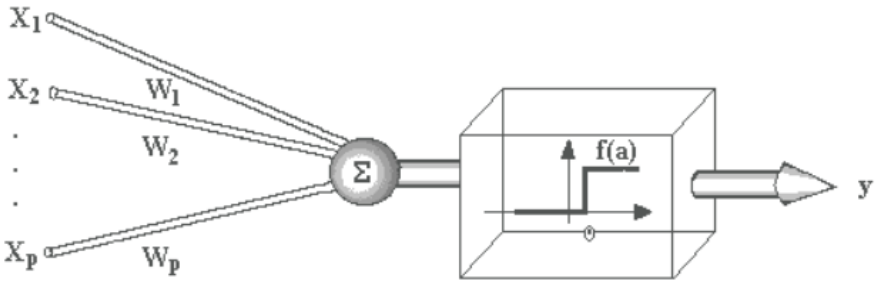


Figura 2.3. Representação do neurônio artificial, proposta por McCullock e Pitts em 1943. Com uma função de ativação com limite (threshold), que muda a saída de acordo com um limite pré estabelecido na função(f(a)). Que depende do somatório das entradas(x) multiplicadas por seus respectivos pesos(W).

Fonte[25]: https://sites.icmc.usp.br/andre/research/neural/. Data: 09/04/2024

Para resolver problemas que não são linearmente separáveis, é necessária a configuração de redes de multicamadas. Nesse contexto, os neurônios são organizados em camadas, sendo a camada intermediária entre a entrada e a saída denominada camada oculta, como se pode observar na figura 2.4 [25].

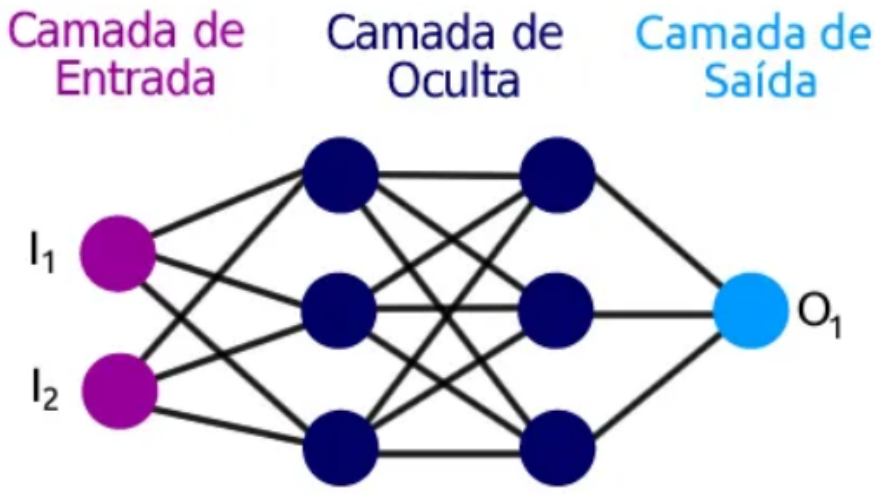


Figura 2.4. Representação simplificada das multicamadas de uma rede neural. Cada camada é composta por neurônios interconectados, onde cada conexão representa um peso que é ajustado durante o processo de treinamento da rede. Essas multicamadas são organizadas em três tipos principais: camada de entrada, camadas ocultas e camada de saída.

Fonte[26]: https: medium.com/brasil-ai. Data: 09/04/2024

As redes neurais convolucionais são amplamente utilizadas na tarefa de visão computacional, uma subárea da inteligência artificial voltada para o processamento de vídeos e

imagens. Esse campo visa capacitar programas computacionais com habilidades de visão semelhantes às dos seres humanos, permitindo a identificação de objetos, pessoas e outros elementos visuais.

Em uma Rede Neural Convolucional (CNN), durante o processamento de uma imagem, não são utilizados todos os pixels como entrada devido à enorme quantidade de dados envolvida, o que tornaria o processamento excessivamente complexo. Em vez disso, realiza-se um pré-processamento da imagem para identificar suas características mais relevantes para cada classe, reduzindo assim o número de entradas dos neurônios e priorizando os pixels mais importantes. Para isso, são realizadas três etapas antes da fase da rede neural densa, conhecidas como operação de convolução, *pooling* e achatamento (*flattening*).

A operação de convolução consiste em adicionar cada elemento da imagem aos seus vizinhos, ponderados por um kernel. Ao processar a imagem, ela é visualizada como uma matriz, enquanto o kernel é uma matriz menor que é multiplicada pela matriz da imagem, resultando em mudanças na imagem. Na equação 2.1 temos a convolução discreta entre um sinal $f[i,j]$ de domínio bidimensional discreto e uma resposta impulsional $g[i,j]$, obtendo uma resposta discreta $h[i,j]$ que é usada na estrutura da CNN.

$$h[i,j] = f[i,j] * g[i,j] \tag{2.1}$$

$$h[i,j] = \sum_{a=0}^{R-1} \sum_{b=0}^{C-1} f[a,b] \cdot g[i-a, j-b] \tag{2.2}$$

Em que:

- $i = 0, 1, ..., R-1$;
- $j = 0, 1, ..., C-1$;
- $f = imagemoriginal$

Isso pode incluir realce de bordas, efeitos de suavização, entre outros. Dentre os diversos tipos de kernels disponíveis, os mais comumente utilizados nessa etapa são os de detecção de bordas, os quais realçam as bordas dos elementos presentes na imagem[27].

Os kernels, também conhecidos como detectores de características, desempenham um papel crucial no processo de convolução. Após a aplicação da convolução, obtém-se uma

matriz menor em relação à matriz original da imagem, o que simplifica o processamento. Essa matriz resultante é denominada mapa de características (*feature map*). Embora possa haver perda de algumas informações da imagem original, o objetivo é detectar as partes mais relevantes, preservando suas características distintivas.

Em seguida, entra em cena a função de ativação ReLU (*Rectified Linear Unit*), que, na etapa da convolução, é aplicada aos mapas de características. Essa função zera todos os valores menores que zero e mantém inalterados os valores maiores que zero, promovendo a não linearidade na rede neural.

Durante a camada de convolução, ocorre a aplicação dos mapas de características, permitindo que a CNN identifique quais desses mapas resultam em melhores desempenhos na detecção de padrões.

A segunda etapa é o *pooling*, que desempenha um papel fundamental na seleção das características mais relevantes dos mapas obtidos na etapa de convolução. Isso é essencial para reduzir o *overfitting* e eliminar ruídos desnecessários, promovendo uma representação mais compacta e significativa dos dados.

O *pooling* pode ser realizado de várias formas, sendo a mais comum o *max pooling*, que seleciona o valor máximo de uma região específica do mapa de características. Além disso, existem outras opções, como o *min pooling*, que seleciona o valor mínimo, e o *average pooling*, que calcula a média dos valores dentro da região definida. Essas diferentes abordagens oferecem flexibilidade na seleção das características mais importantes, permitindo uma adaptação mais eficaz às necessidades específicas de cada problema.

E a terceira etapa é o *flattening*, que desempenha um papel crucial na transição da matriz resultante do *pooling*, denominada mapa de características agrupadas (*pooled feature map*), para um formato vetorial. Esse processo consiste essencialmente em converter a estrutura matricial em um vetor unidimensional, que será então utilizado como entrada na rede neural densa. Essa transformação é fundamental para permitir que as informações extraídas pelas camadas convolucionais sejam processadas de maneira eficiente pela rede neural densa, contribuindo assim para o aprendizado e a tomada de decisões do modelo. Na figura 2.5 podemos observar todas as camadas da rede neural convolucional, no entanto ela não se restringe só nessa sequencia de camadas, podendo ter diferentes

sequencias de camadas.

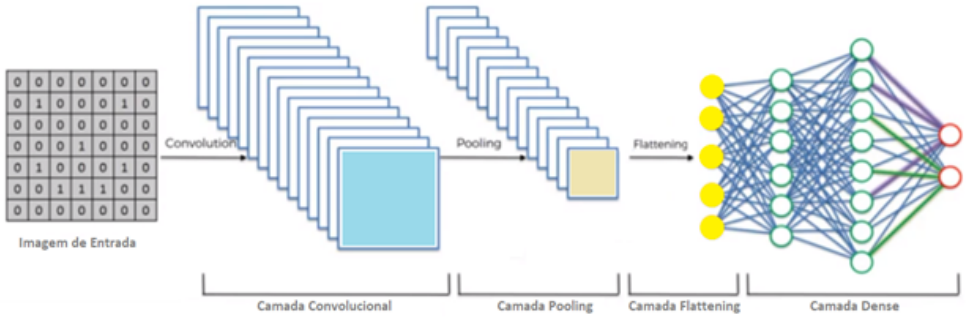


Figura 2.5. Representação simplificada da CNN. Possui uma arquitetura especializada, projetada para processar dados com estrutura de grade, como imagens e detecção de padrões.Utiliza camadas convolucionais para extrair características importantes das entradas.

Fonte[28]: https://medium.com. Data: 09/04/2024

2.2 Ausculta Pulmonar

Os procedimentos de ausculta pulmonar em um exame físico normalmente tem as seguintes diretrizes. A ausculta pulmonar deve ser realizada com o paciente sentado ou em pé e que respire profundamente com a boca aberta. Isso permite uma maior entrada de ar e elimina bastante os ruídos produzidos pela passagem do ar pelas vias aéreas superiores, melhorando assim a percepção dos sons. Durante o exame da parte posterior do tórax, o paciente deve cruzar os braços, enquanto durante o exame da parede lateral do tórax, o paciente deve colocar as mãos na cabeça para melhor exposição dessas áreas. Os sons respiratórios normais ouvidos durante essa parte do exame são criados pela turbulência do ar ao colidir com as paredes e saliências das bifurcações da árvore brônquica. Esses sons devem ser avaliados quanto à frequência, qualidade e predominância durante o ciclo respiratório. O murmúrio vesicular é o som respiratório normal predominante encontrado na maior parte do tórax. Esse som tem um componente inspiratório mais longo e mais intenso em comparação com o componente expiratório.

Nas áreas da parede anterossuperior do tórax e nas regiões supraescapulares, o som tende a ser mais volumoso. Isso é especialmente notável na parede anterossuperior direita devido às características do brônquio fonte direito, que é mais largo e fica mais próximo da parede torácica em comparação com o lado esquerdo. O som traqueal, encontrado ao auscultar a traqueia, é extremamente áspero, com um componente expiratório mais prolongado e intenso do que o inspiratório. Encontrar esse som em outras áreas do tórax é anormal e pode indicar consolidação pulmonar [29].

Os sons respiratórios anormais são conhecidos como ruídos adventícios e incluem:

• Cornagem ou estridor: um ruído áspero, principalmente durante a inspiração, causado pela passagem turbulenta do ar por áreas obstruídas nas vias aéreas superiores fora do tórax, como a laringe ou traqueia. Às vezes, pode ser ouvido a certa distância do paciente. Pode ocorrer em condições como estenose da traqueia, tumores na laringe, edema da glote, laringite aguda e difteria.
• Sibilo: um som agudo, principalmente durante a expiração, mas também pode ocorrer durante a inspiração. Semelhante a um assobio, chiado ou miado de gato, é causado pelo estreitamento dos brônquios de pequeno calibre devido à constrição da musculatura lisa, edema da mucosa e/ou acúmulo de secreção intraluminal. Pode ocorrer em condições como asma, doença pulmonar obstrutiva crônica (DPOC) e infecções brônquicas. Quando localizado em uma região específica, sugere obstrução parcial devido a neoplasia ou corpo estranho.
• Ronco: um ruído grave que ocorre tanto durante a inspiração quanto durante a expiração, semelhante ao ronco de uma pessoa dormindo. Os roncos desaparecem ou mudam de localização com a tosse e são causados pela presença de secreção nas vias aéreas dos brônquios de grande e médio calibre. Podem estar presentes em infecções brônquicas, asma e DPOC.
• Estertor crepitante ou fino: um som agudo que ocorre no final da inspiração, estendendo-se até o início da expiração. Semelhante ao som de cabelos roçando ou ao desprendimento de duas superfícies de velcro, pode ser produzido pela abertura sequencial de alvéolos previamente colapsados devido à presença de exsudato em seu interior, como na pneumonia, ou por alterações no interstício pulmonar, como em doenças intersticiais e congestão pulmonar.
• Estertor subcrepitante ou bolhoso: um ruído grave presente desde o início da inspiração até o final da expiração, que pode ser aliviado com a tosse. Ocorre devido ao fechamento e abertura de bronquíolos com secreção em sua luz e é comum em condições como bronquiectasias, bronquite aguda, bronquite crônica e edema agudo de pulmão, onde toda a via aérea está preenchida por líquido.
• Grasnido: um som agudo, principalmente durante a inspiração, semelhante ao piado de uma gaivota, que ocorre em pacientes com doenças que afetam os bronquíolos.
• Atrito pleural: um ruído áspero e áspero, audível tanto na expiração quanto na inspiração, com maior intensidade nesta última. É produzido por doenças que inflamam as pleuras visceral e parietal, como infecções, neoplasias e doenças autoimunes com envolvimento pleural. Sua semelhança ao som do atrito entre duas superfícies de couro torna-o mais audível na face lateral inferior do tórax, devido à maior expansibilidade nessa área.
• Sopro tubário: um ruído grave, mais intenso durante a expiração e semelhante ao som de um tubo sendo soprado. Este sopro é característico de áreas de consolidação pulmonar, quando um brônquio permeável se estende até o local da condensação pulmonar [29].

3 Materiais e Métodos

3.1 Implementação de Algoritmos para Decomposição e Classificação de Ausculta Pulmonar

Estrutura da rede neural:

Uma Rede Neural Convolucional (CNN), também conhecida como ConvNet, é um tipo especializado de arquitetura de rede neural profunda projetada para processar dados que têm uma grade de topologia, como imagens. As CNNs foram especialmente desenvolvidas para resolver problemas de visão computacional, embora também tenham se mostrado eficazes em tarefas envolvendo séries temporais e dados de grade semelhantes. A estrutura da Rede Neural no projeto é a seguinte: O input é uma imagem ou imagem com características MFCCs, em seguida temos uma camada de entrada, segue uma camada de convolução com uma função de ativação ReLU; em seguida temos uma camada de dropout e uma camada de Maxpooling, acrescida por mais uma camada convolucional, aplicação de ativação ReLU e outra sequência de Dropout e MaxPooling, seguida por uma sequência de somente dois blocos de camadas convolucionais e MaxPooling. Na sequência, temos o *Flatten* que irá fazer a transformação da matriz em vetor para ser enviado às camadas densas da rede neural, primeira camada densa, aplicação de dropout, mais uma camada densa, aplicação da função de ativação softmax que irá gerar uma probabilidade para cada uma das classes, e no final as previsões dos classificadores. Apresentando 91.658 parâmetros ou pesos que serão ajustados durante o treinamento. Os parâmetros são compostos por Camadas Convolucionais: A característica fundamental das CNNs é a presença de camadas convolucionais. Essas camadas aplicam operações de convolução para extrair características específicas da entrada. Cada filtro na convolução age como um detector de características, aprendendo padrões como bordas, texturas e formas em diferentes regiões da imagem.

Camadas de Pooling: Após as camadas convolucionais, as camadas de pooling são frequentemente utilizadas para reduzir a dimensionalidade espacial da representação e diminuir a quantidade de parâmetros na rede. A operação de pooling reduz a resolução da imagem, mantendo as características mais importantes. Em cada quadrante é reduzido

para o dado de maior valor.
Camadas Totalmente Conectadas: No final da arquitetura, geralmente existem camadas totalmente conectadas que recebem as características extraídas e realizam a classificação final. Essas camadas são semelhantes às encontradas em redes neurais tradicionais.
Funções de Ativação: Funções de ativação como ReLU (Rectified Linear Unit) são comumente aplicadas após as operações convulsionais. Essas funções introduzem não-linearidades na rede, permitindo que a rede aprenda representações mais complexas.
Regularização e Normalização: Técnicas como dropout e normalização em lotes (batch normalization) são frequentemente empregadas para prevenir overfitting e acelerar o treinamento. A porcentagen de dropout usada nesse projeto foi de 40% dos neurônios desligados.
Transfer Learning: Devido ao grande número de parâmetros e à necessidade de grandes conjuntos de dados para treinamento eficaz, as CNNs muitas vezes se beneficiam de técnicas de transfer learning. Isso envolve usar uma rede pré-treinada em uma tarefa semelhante e adaptá-la para a tarefa específica.

Algumas linguagens de programação foram sondadas para a implementação do projeto, como C++/C, MATLAB e Python. A linguagem de programação escolhida para a realização do projeto foi o Python, por causa da familiaridade com a linguagem adquirida por meio da utilização dessa linguagem em projetos anteriores, bem como sua eficiência para processamento de dados.
O Python é uma linguagem de programação de alto nível, interpretada e tem um propósito geral, sendo usada em diferentes áreas. Essa linguagem é conhecida por sua sintaxe clara e legível. Foi desenvolvida por Guido van Rossum e lançada pela primeira vez em 1991. Desde então, tornou-se uma das linguagens mais populares no cenário da programação. Sua versatilidade e vasta biblioteca de módulos contribuem para a ampla adoção em diversos campos, incluindo ciência de dados e aprendizado de máquina, que é uma área de atuação desse projeto.
Python se destaca como uma excelente escolha para trabalhar com redes neurais por várias razões. Primeiramente sua sintaxe clara e expressiva facilita o desenvolvimento de código, tornando-o mais legível e compreensível. Isso é particularmente importante em projetos complexos como o desenvolvimento de redes neurais, onde a clareza do código é crucial para entender e ajustar os modelos. Além do mais, a maioria das linhas do código possui comentários para uma melhor compreensão e documentação do programa. Além disso, Python oferece uma ampla gama de bibliotecas especializadas em aprendizado de máquina, como TensorFlow, PyTorch e Keras, que simplificam o processo de implementação e treinamento de redes neurais. Essas bibliotecas são construídas sobre a eficiência de implementações em linguagens de baixo nível, mas fornecem interfaces

simples e abstratas, facilitando a experimentação e o desenvolvimento rápido de modelos. Outro fator determinante é a comunidade ativa de desenvolvedores em torno da linguagem Python. Isso resulta em uma grande quantidade de recursos, tutoriais e suporte disponíveis on-line, o que é crucial para quem está começando ou enfrentando desafios específicos ao trabalhar com redes neurais, o que torna a explicação de uma dúvida ou o entendimento de uma aplicação mais acessível e com várias opções de fontes.

Logo, a linguagem Python destaca-se como uma boa escolha para o desenvolvimento de redes neurais devido à sua sintaxe clara, bibliotecas especializadas, comunidade ativa e facilidade de uso, proporcionando um ambiente propício para a inovação e avanços no campo do aprendizado de máquina.

O Google Colab, ou Colaboratory, é uma plataforma gratuita que roda em nuvem e oferece ambientes de execução de notebooks Jupyter para desenvolvimento de código em Python. Essa ferramenta tem ganhado popularidade, especialmente na comunidade de ciência de dados e aprendizado de máquina, devido a várias vantagens como:

O Colab é gratuito e executa na infraestrutura de nuvem do Google. Isso elimina a necessidade de configurar e manter ambientes locais, permitindo que os códigos executem em grandes servidores sem custos adicionais.

Também a plataforma dispõe de uma opção de uso de *hardware*, fornecendo acesso gratuito a GPUs (Unidades de Processamento Gráfico) e TPUs (Unidades de Processamento Tensorial) atualizados.

Isso é crucial para acelerar o treinamento de modelos de aprendizado de máquina e processamento de dados complexos, tornando o Colab uma escolha atrativa para projetos intensivos em computação. Embora este projeto não exija um grande poder de processamento, o ambiente do colab permite fácil escalonamento para o uso pesado de processamento.

Outra vantagem é a integração com Google Drive. Os notebooks podem ser facilmente salvos no Google Drive, proporcionando uma maneira conveniente de compartilhar e colaborar em projetos. Isso também permite o acesso dos notebooks de qualquer dispositivo conectado à internet, sem a preocupação de perder o projeto, caso o computador local tenha algum problema.

A plataforma dispõe de várias bibliotecas e frameworks pré-instalados. No Colab há muitas bibliotecas populares de Python e frameworks de aprendizado de máquina pré-instalados. Isso simplifica a configuração e economiza tempo, pois podemos começar a trabalhar imediatamente sem a preocupação com instalações de bibliotecas e atualizações.

Apresenta também uma facilidade de colaboração, embora não seja o caso desse projeto. Mas no ambiente do Colab é possível a colaboração em tempo real semelhante aos notebooks Jupyter tradicionais, podendo compartilhar facilmente os notebooks com outras pessoas, facilitando assim o trabalho em equipe e a revisão de códigos.

A biblioteca principal usada na análise de áudio desse projeto foi a Librosa. Trata-se

de uma biblioteca em Python desenvolvida para análise de música e áudio. Projetada para lidar com tarefas relacionadas ao processamento de sinais de áudio, a Librosa oferece uma ampla gama de ferramentas e funcionalidades que são valiosas para músicos, pesquisadores em áudio e cientistas de dados. Algumas das principais características e funcionalidades da Librosa incluem:

A extração de características de áudio. a Librosa é útil para extrair características relevantes de sinais de áudio, como espectrogramas, cromagramas, tonnetz, entre outros. Essas representações são essenciais para tarefas como classificação de gênero musical, reconhecimento de acordes e análise de padrões sonoros.

A biblioteca oferece ferramentas para carregar e manipular sinais de áudio de forma eficiente. Além disso, ela simplifica tarefas comuns como resampling, alteração de pitch e tempo, análise rítmica, detecção de batidas e estimativa de tempo musical. Essas ferramentas são essenciais para entender a estrutura temporal da música.

Integração com visualização gráfica. A biblioteca possui recursos de visualização que facilitam a interpretação e análise de características extraídas. Isso inclui gráficos para representação espectral, cromagramas, entre outros.

Possibilita suporte a diversos formatos de áudio. A Librosa é capaz de lidar com uma variedade de formatos de áudio comuns, permitindo a análise de arquivos WAV, MP3, FLAC, aumentando a gama de bancos de áudios possíveis de serrem trabalhados no projeto.

Por ser uma biblioteca voltada para Python, a Librosa tem uma boa compatibilidade com outras bibliotecas como NumPy e SciPy, proporcionando uma integração eficiente em fluxos de trabalho mais amplos de análise de dados.

Também apresenta uma documentação abundante e bem documentada, o que facilita o aprendizado e a utilização por parte dos desenvolvedores. Há exemplos práticos e tutoriais disponíveis para orientar os usuários em várias tarefas comuns.

O *TensorFlow* é uma biblioteca de código aberto desenvolvida pela Google para realizar tarefas de aprendizado de máquina e deep learning. Lançado em 2015, o TensorFlow tornou-se uma das ferramentas mais populares e amplamente utilizadas no campo da inteligência artificial. Ela oferece uma estrutura flexível e abrangente que abrange desde o treinamento de modelos simples até arquiteturas complexas de deep learning. Essa biblioteca foi escolhida para a programação da rede neural convolucional.

A principal abstração no TensorFlow é o conceito de tensores, que são arranjos multidimensionais de dados. Esses tensores fluem através de um gráfico computacional, onde as operações matemáticas são aplicadas para realizar o processamento. O TensorFlow utiliza uma representação em grafo para otimizar o desempenho e facilitar a execução distribuída em hardware especializado como GPUs e TPUs.

Principais características e aspectos do TensorFlow são:

Flexibilidade e escalabilidade. No TensorFlow é oferecido suporte a uma ampla variedade

de tarefas de aprendizado de máquina, desde modelos simples até arquiteturas complexas de deep learning. Ele é projetado para ser escalável, podendo ser utilizado em diferentes plataformas desde dispositivos móveis até clusters de servidores.
O TensorFlow incorpora o Keras como sua API de alto nível, o que simplifica a construção e o treinamento de modelos, tornando o código mais legível e acessível para iniciantes. A combinação do TensorFlow com o Keras oferece uma abordagem poderosa e intuitiva para o desenvolvimento de modelos de aprendizado de máquina.
Ele possui uma comunidade vasta e ativa, com desenvolvedores contribuindo para a biblioteca e compartilhando recursos, tutoriais e modelos pré-treinados. Isso cria um ecossistema dinâmico que impulsiona a inovação e a colaboração.
Também o TensorFlow inclui o TensorBoard, que é uma ferramenta de visualização que ajuda a entender e depurar modelos. Ele fornece gráficos interativos, métricas de treinamento e visualização de tensores, facilitando a análise e melhoria de modelos.
Suporta implementações em diferentes plataformas, incluindo desktops, servidores, dispositivos móveis e nuvem. Isso permite a implementação de modelos em uma variedade de cenários e ambientes, podendo estender sua atuação para atender às suas necessidades específicas na criação de novas operações, funções de otimização e até mesmo integrando com hardware especializado.
O TensorFlow é amplamente utilizado em diversas aplicações, como visão computacional, processamento de linguagem natural, reconhecimento de voz, entre outros, destacando sua versatilidade para uma ampla gama de tarefas de aprendizado de máquina.

3.2 Insumos Utilizados

Foram utilizadas duas bases de dados distintas para compor o conjunto de sons pulmonares empregados nesta pesquisa. A partir dessa seleção dos dados obtidos nessas fontes, foram identificadas e selecionadas as quatro ocorrências sonoras de maior prevalência, que servirão como foco para a continuidade do estudo. Essa seleção visa reduzir a necessidade de manipulações excessivas, promovendo maior equidade entre os conjuntos de amostras e minimizando o desbalanceamento inerente às bases de dados utilizadas. Espera-se que essa abordagem resulte em uma maior precisão dos resultados, garantindo uma análise mais consistente.

Uma das bases de dados dos sons respiratórios utilizada no projeto resulta da colaboração entre duas equipes de investigação, uma em Portugal e outra na Grécia. Ela compreende um total de 920 gravações anotadas, com durações variadas de 10 segun-

dos a 90 segundos. Essas gravações foram obtidas a partir de 126 pacientes, totalizando 5,5 horas de registros que englobam 6.898 ciclos respiratórios. Dentre esses ciclos, 1.864 apresentam estertores, 886 contêm sibilos e 506 incluem tanto estertores quanto sibilos. Os dados abrangem uma diversidade de faixas etárias, contemplando crianças, adultos e idosos conforme figura 3.1.

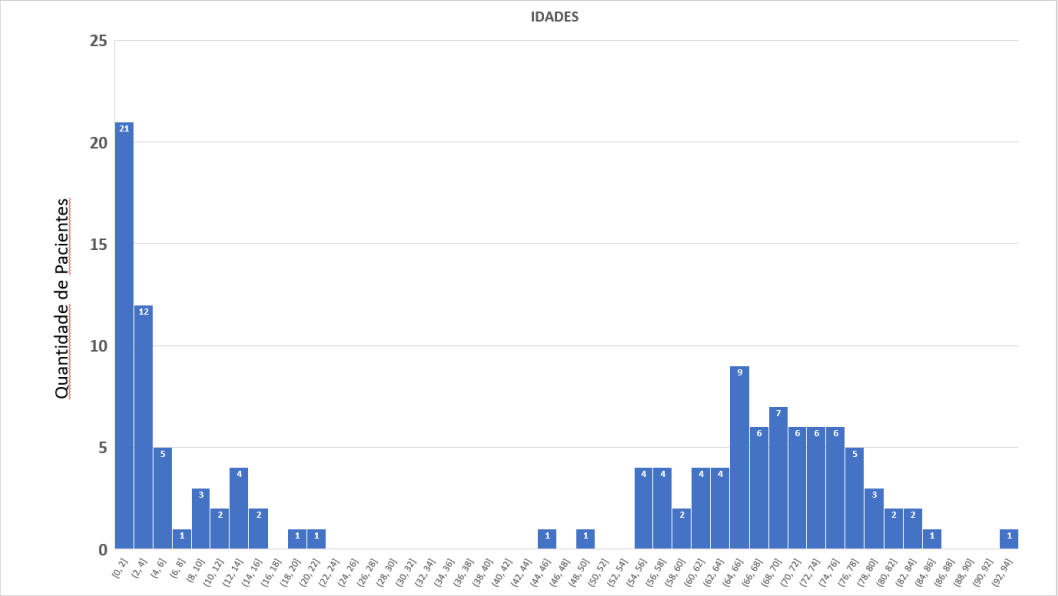


Figura 3.1. Quantidade de pacientes distribuidos pela Faixa Etária, demostrando a heterogeneidade dos arquivos de áudio em relação a idade do paciente.
Fonte: Elaborada pelo autor, a partir do software Excel. 2023.

A base de dados é composta por 45 pacientes do sexo feminino, 79 pacientes do sexo masculino e 1 paciente que não teve os dados informados. Além disso, a base inclui tanto sons respiratórios limpos como gravações ruidosas, simulando condições realistas da vida cotidiana.
No nome de cada arquivo de áudio constam as informações complementares, além de um arquivo .txt com o restante das informações de cada paciente. Os arquivos de áudio dispõem das seguintes informações no próprio nome do arquivo.

O nome tem o seguinte formato: [fsID]-[índice de gravação]-[localização da captura do áudio]- [Modo de aquisição]-[classID]-[Equipamento usado].wav, onde:

- [fsID] = é o ID do paciente que foi obtido o som;
- [índice de gravação] (1b1, 2b2, 2b1 ...);
- [localização da captura do áudio] sendo dividido assim:

  - a. Traqueia (Tc);
  - b. Anterior Esquerdo (Al);
  - c. Anterior Direito (Ar);
  - d. Posterior Esquerdo (Pl);
  - e. Posterior Direito (Pr);
  - f. Lateral Esquerdo (Ll);
  - g. Lateral Doreito (Lr);
- [Modo de aquisição] dividido em dois tipos:
  - Canal mono (sc);
  - Canal stereo (mc);
- [classID] é uma identificação numérica da classe de som das patologias com as seguites associações:
  - 0 = Asthma – Asma;
  - 1 = Bronchiectasis – Bronquiectasia;
  - 2 = COPD - Doença Pulmonar Obstrutiva Crônica;
  - 3 = Healthy – Normal;
  - 4 = LRTI - Infecção do Trato Respiratório Inferior;
  - 5 = Pneumonia;
  - 6 = URTI - Infecção do Trato Respiratório Superior;
  - 7 = Broquiolite;
- [Equipamento usado] é a última identificação e referente ao modelo de estetoscópio digital usado , onde:
  - a. AKG C417L Microphone (AKGC417L);
  - b. 3M Littmann Classic II SE Stethoscope (LittC2SE);
  - c. 3M Litmmann 3200 Electronic Stethoscope (Litt3200);
  - d. WelchAllyn Meditron Master Elite Electronic Stethoscope (Meditron).

Um exemplo prático da nomeclatura do arquivo **104_1b1_Ar_sc_Litt3200.wav**. Esses classificadores são os dados presentes no nome de cada arquivo de áudio. Já no arquivo em .txt estão presentes seis dados, onde o primeiro é o número identificador do

paciente, o segundo refere-se à idade do paciente, o terceiro diz respeito ao sexo do paciente; já o quarto dado contém as informações do IMC dos pacientes adultos, ao passo que o quinto e sexto dados são referentes, respectivamente, ao peso e à altura das crianças que participaram dessa coleta de dados.
Dois exemplo de desse dados: de criança: **102 0.75 F NA 9.8 73** e de um adulto: **111 63 M 28.4 NA NA**.

Essa base de dados apresenta um desbalanceamento uma distribuição desigual das classes de sons pulmonares, onde algumas ocorrências sonoras estão representadas em maior quantidade do que outras como podemos observar na figura 3.2 . Esse desequilíbrio pode comprometer o desempenho de modelos preditivos, uma vez que algoritmos de aprendizado de máquina tendem a favorecer as classes mais frequentes, resultando em vieses e reduzindo a capacidade de generalização do modelo. Para mitigar esse problema, é essencial adotar estratégias como balanceamento de amostras, técnicas de reamostragem ou ajustes nos pesos das classes, garantindo uma análise mais equitativa e uma maior robustez na interpretação dos resultados.

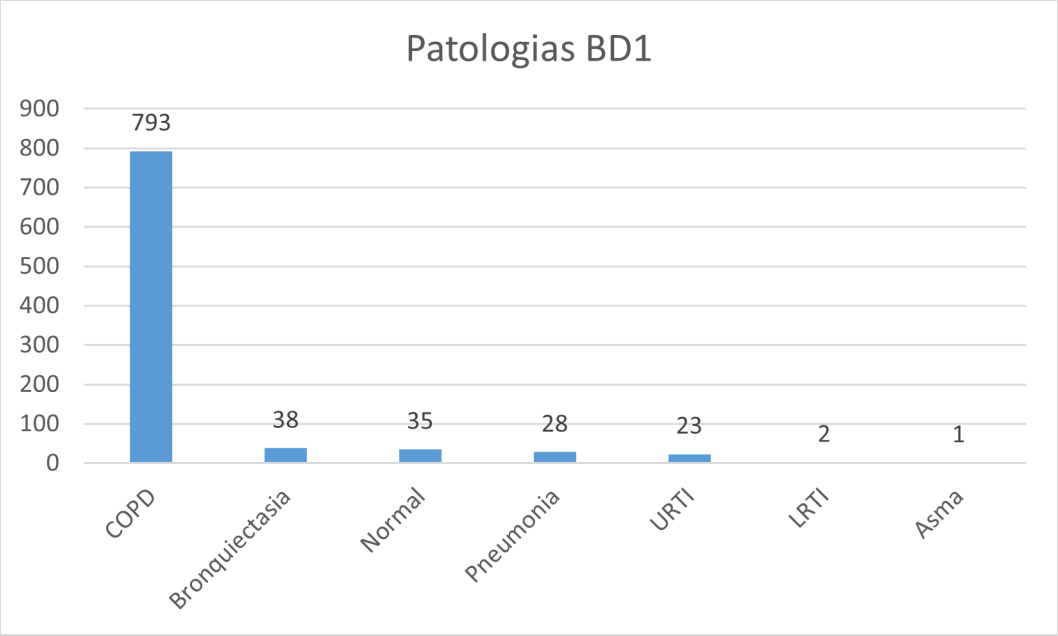


Figura 3.2. Quantidade de pacientes distribuidos pelas patologias diagnosticadas, demostrando o desbalanciamento dos arquivos de áudio em relação as patologias dos pacientes.

Fonte: Elaborada pelo autor, a partir do software Excel. 2024.

Para complementar a análise e ampliar as amostras do estudo, foi utilizada uma segunda base de dados contendo gravações de sons respiratórios de 112 indivíduos do Oriente Médio que enfrentam diversas condições de saúde pulmonar. Dentre esses participantes, 35 eram saudáveis e 77 apresentavam alguma patologia respiratória de acordo

com a tabela 3.1 . A idade dos indivíduos variava entre 21 e 90 anos, com uma média de 50,5 anos e um desvio padrão de 19,4. A amostra era composta por 43 mulheres e 69 homens.

Tabela 3.1. Distribuição dos indivíduos na segunda base de dados por condição de saúde, número de participantes, faixa etária e gênero.

Condição de Saúde	**Nº de Indivíduos**	**Faixa Etária**	**Gênero**
Normal	35	18–81	11 mulheres, 24 homens
Asma	32	12–72	17 mulheres, 15 homens
Pneumonia	5	36–70	2 mulheres, 3 homens
DPOC	9	42–76	1 mulher, 8 homens
BRON	3	20–68	1 mulher, 2 homens
Insuficiência Cardíaca	21	20–83	9 mulheres, 12 homens
Fibrose Pulmonar	5	44–90	2 mulheres, 3 homens
Derrame Pleural	2	70–81	0 mulheres, 2 homens

Essa base de dados contém apenas uma gravação por indivíduo. A duração das gravações varia entre 5 e 30 segundos, tempo suficiente para abranger pelo menos um ciclo respiratório. O tempo máximo de gravação é determinado pelas limitações do estetoscópio eletrônico utilizado, e não houve restrições quanto à duração mínima impostas ao médico responsável pelo exame dos pacientes. Os arquivos de dados são nomeados seguindo um padrão específico, começando com um código de filtragem representado pelas letras B, D ou E. Em seguida, a letra P precede um número sequencial exclusivo para cada paciente, seguido por um sublinhado. O nome do arquivo também contém informações sobre o diagnóstico, o tipo de som gravado, a localização da medição no tórax, a idade e o gênero do paciente. Foram aplicados três tipos distintos de filtros nas gravações. O modo Bell, indicado pela letra B, amplifica sons na faixa de frequência de [20-1000] Hz, com ênfase nos sons de baixa frequência entre [20-200] Hz. O modo Diafragma, identificado pela letra D, amplia a faixa de frequência para [20-2000] Hz, destacando sons entre [100-500] Hz. Já o modo Estendido, representado pela letra E, opera dentro da faixa de [20-1000] Hz, enfatizando frequências entre [50-500] Hz. Dessa forma resultou-se em uma distribuição de acordo como a figura 3.3.

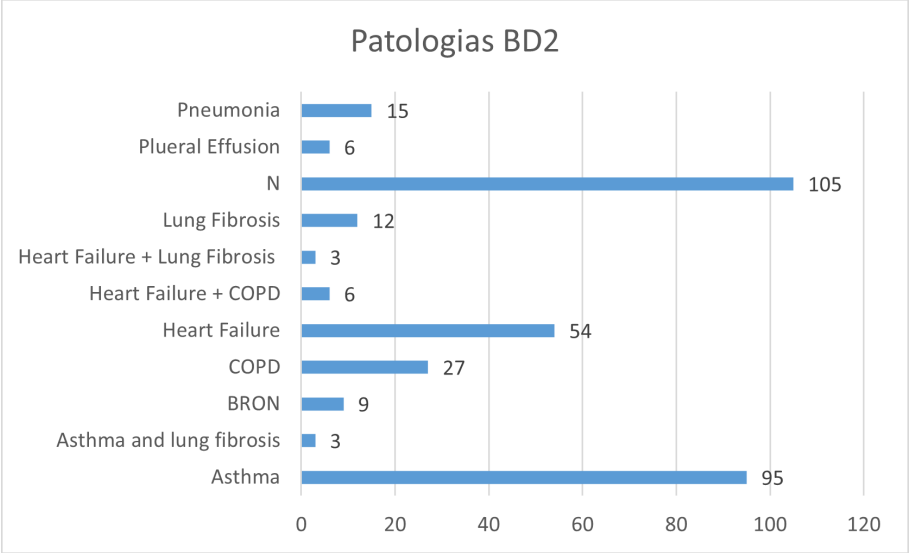


Figura 3.3. Quantidade de pacientes distribuidos pelas patologias diagnosticadas, demostrando o desbalanciamento dos arquivos de áudio em relação as patologias dos pacientes.

Fonte: Elaborada pelo autor, a partir do software Excel. 2024.

Uma das principais dificuldades encontradas no desenvolvimento deste projeto foi a obtenção de uma base de dados balanceada, contendo uma quantidade significativa de amostras suficientes para que a rede neural pudesse ser treinada de maneira eficaz. Para superar essa limitação, a primeira estratégia adotada foi a seleção das quatro condições de saúde com maior incidência, utilizando-se das duas bases de dados apresentadas. Assim, foram escolhidos os seguintes conjuntos de amostras: Doença Pulmonar Obstrutiva Crônica (COPD), normal, asma e pneumonia. A base de dados resultante, conforme ilustrado na Figura 3.4, será utilizada como fonte primária.

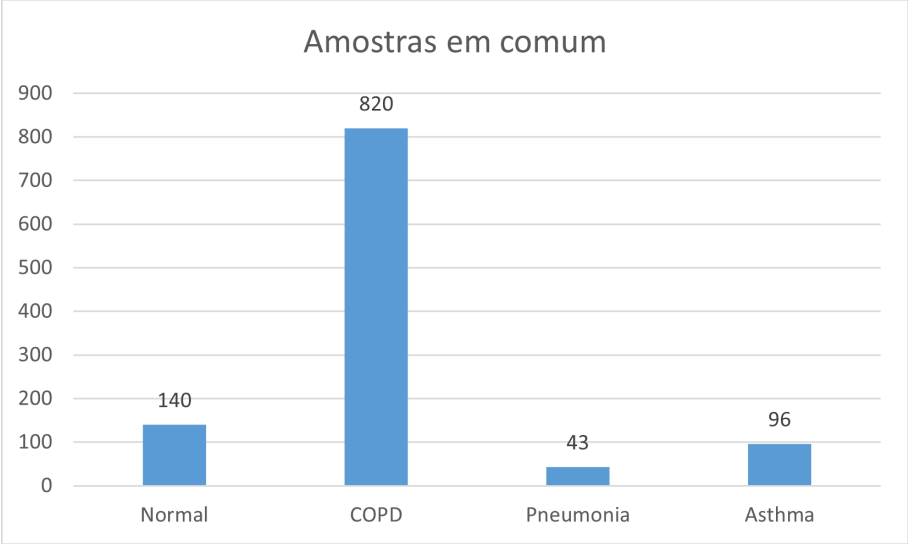


Figura 3.4. Quantidade de pacientes distribuidos pelas patologias diagnosticadas das quatro amostras com maior incidência oriundo da união das duas bases de dados, demostrando o desbalanciamento dos arquivos de áudio em relação as patologias dos pacientes.

Fonte: Elaborada pelo autor, a partir do software Excel. 2024.

No entanto, apenas a patologia Doença Pulmonar Obstrutiva Crônica (COPD) apresentou um número significativo de amostras, totalizando 820, após a combinação das duas bases de dados, conforme ilustrado na Figura 3.4.

Diante do desbalanceamento entre as amostras obtidas, a segunda estratégia adotada foi a criação de amostras modeladas. Para isso, foram realizadas pequenas modificações nas amostras originais dos conjuntos normal, pneumonia e asma, com o objetivo de aumentar a quantidade dessas amostras e melhorar o desempenho da rede neural, tornando a base de dados balanceada. Essas modificações consistiram em variações no fator multiplicativo de velocidade, como ajustes de 1,05x ou 0,95x da velocidade original. Também foi realizada a alteração do tom sonoro das amostras por meio de convolução, tornando-as levemente mais agudas ou mais graves, utilizando passos de semi-tom a partir da nota Dó da quarta oitava(C4), que tem uma frequência de aproximadamente 261,626 Hz. A frequência do quarto Dó pode ser calculada pela seguinte fórmula:

$$f = f_{\text{ref}} \times 2^{\frac{n}{12}} \tag{3.1}$$

Onde:

- f_{ref} é a frequência de referência (440 Hz para A4),
- n é o número de semitons entre a nota de referência (A4) e a nota desejada (C4),

- f é a frequência da nota desejada. Sabemos que C4 está 9 semitons abaixo de A4. Aplicando a fórmula:

$$f_{C4} = 440 \times 2^{\frac{-9}{12}} \approx 261,626\,\text{Hz} \tag{3.2}$$

Portanto, a frequência do Dó da quarta oitava (C4), partindo do Lá 440 Hz, é aproximadamente 261,626 Hz.

A aplicação dessas técnicas de balanceamento possibilitou a obtenção de uma base de dados balanceada, conforme apresentado na Tabela 3.2 e na Figura 3.5. Isso pode contribuir significativamente para o treinamento da rede neural, promovendo uma distribuição mais equilibrada das classes e, consequentemente, reduzindo viéses.

Tabela 3.2. Número de amostras por classe, após o acréscimo das amostras modeladas.

Classe	**Número de Amostras**
Asma	873
Normal	840
Pneumonia	832
COPD	820

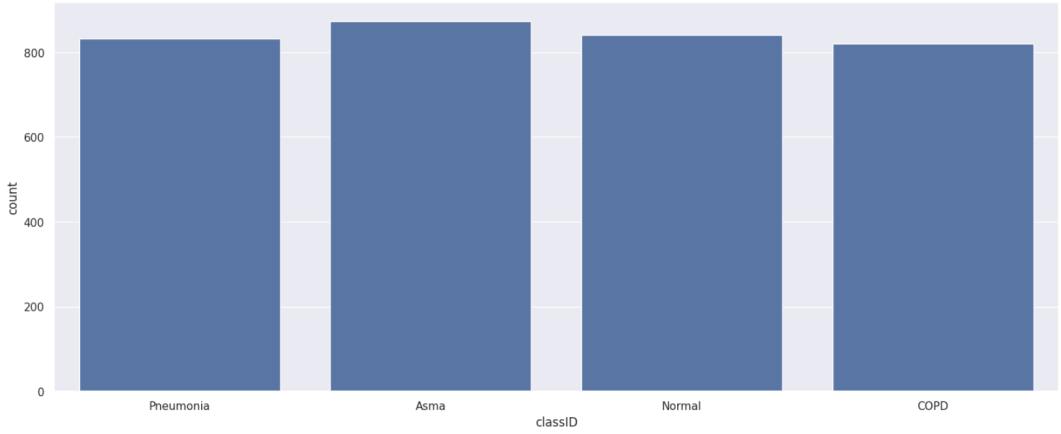


Figura 3.5. Quantidade de amostras distribuidas pelas patologias diagnosticadas das quatro amostras com maior incidência oriundo da união das duas bases de dados e do acréscimo das amostras modeladas, demostrando o balanciamento dos arquivos de áudio em relação as patologias dos pacientes.

Fonte: Elaborada pelo autor, a partir do software Pyton. 2025.

Esse processo também melhora a capacidade de generalização do modelo. No entanto, é importante ressaltar que, embora as amostras das três classes minoritárias tenham sido significativamente aumentadas por meio de modelagem, os resultados obtidos serão

analisados exclusivamente para a patologia COPD. Isso ocorre porque a classe COPD não passou por modelagem para aumento de amostras, mantendo as amostras extraídas diretamente de suas fontes originais, sem modificações.

3.3 Procedimentos experimentais

Estudando a biblioteca Librosa e explorando suas capacidades no processamento de áudio, identificou-se uma aplicação potencial para a análise de sons provenientes da ausculta cardiopulmonar. Na biblioteca Librosa, há uma função denominada **'.decompose.hpss'**, cujo propósito é separar dois componentes musicais distintos: a parte percussiva e a parte harmônica.
A componente percussiva em uma composição musical é responsável pelo ritmo e compasso, frequentemente associada a sons de bateria, tambores e pratos, geralmente ligados a instrumentos percussivos. Por outro lado, a componente harmônica desempenha o papel na melodia e nos arranjos musicais, sendo comumente associada a instrumentos que emitem várias notas musicais e conseguem manter a emissão sonora, geralmente esses instrumentos estão associados a instrumentos de cordas e sopro.
Ao aplicar esse entendimento musical ao áudio de uma ausculta cardiopulmonar, percebeu-se que o som cardíaco se assemelha a uma componente percussiva provocado pelo rápido batimento do coração, enquanto o som pulmonar assemelha-se a uma componente harmônica, já que o som produzido pelo pulmão é produzido pelo fluxo de ar no sistema respiratório. Essa observação levou à decomposição do áudio cardiopulmonar, com a interação de separar o som produzido pelo coração como uma componente percussiva do som do pulmão, como uma componente harmônica. Os resultados dessa análise podem ser apreciados nos áudios disponíveis nos links e nas figuras 4.1, 4.2 e 4.3 com a representação da amplitude no tempo e nos espectrogramas de cada forma de áudio. Assim, implementou-se essa função em um áudio de ausculta cardiopulmonar para testar o raciocínio proposto.
Os procedimentos experimentais para treinamento da CNN seguiram algumas etapas. Os passos gerais envolvidos no treinamento da CNN foram primeiramente a escolha da base de dados, optando-se por uma base de dados representativa que tem potencial para contribuir no estudo desse trabalho. A base de dados foi dividida em três subconjuntos, conforme a seguir. O primeiro conjunto foi o de treinamento, composto por 80% dos dados, totalizando em valores absolutos 2692 áudios de ausculta cardiopulmonar. Os outros dois subconjuntos, cada um com 336 e 337 áudios, representando 10% dos áudios da base de dados cada um. Esses dois subconjuntos são o de validação e o de teste. O conjunto de validação é usado em cada época do treinamento, obtendo-se, assim, os parâmetros de cada época, verificando-se se existe uma melhora nos percentuais após cada interação. Já o conjunto de testes só é usado no final de todas as épocas. E serve para analisar a

evolução da CNN, com áudios novos e diferentes que a CNN não empregou no treinamento nem nas validações.

Na realização do pré-processamento dos dados, os arquivos de áudios são processados para um espectrograma de Coeficiente Cepstral de Frequência Mel (MFCC). Essa escala de Mel, que podemos observar na figura 3.6, e é fundamental no uso de áudio em aplicações de *machine learning*, pois ela imita a percepção sonora humana.

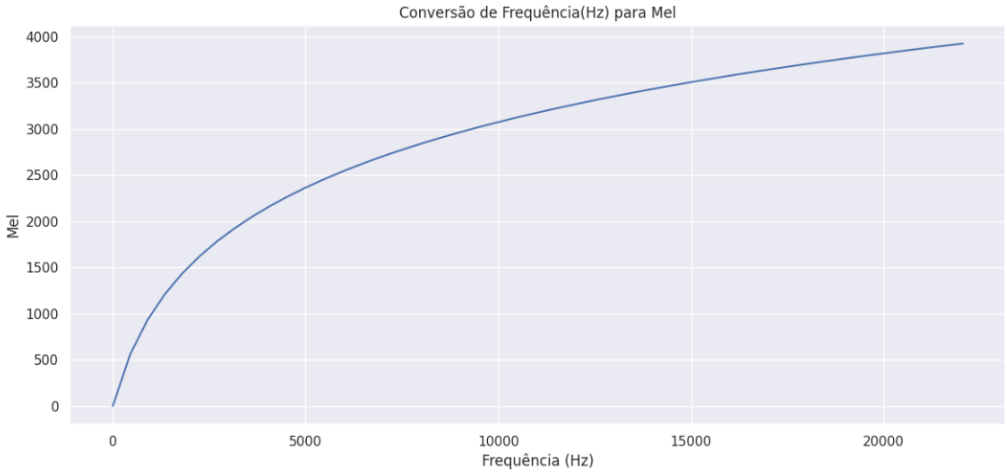


Figura 3.6. Curva da escala Mel em razão da frequência em HZ, demostrando a sua não linearidade com o aumento da frequência, desenvolvida para ser similar a percepção humana em razão da variação da frequência(Hz).

Fonte: Elaborada pelo autor, a partir do software Python. 2025.

Com isso, aplica-se uma normalização média cepstral para formar a entrada dos dados na CNN. Um outro pré-processamento que foi implementado, foi o uso da separação harmônica das amostras.

Foi projetada a arquitetura da CNN, conforme já foi mencionado, sendo divididas as classes de classificação em quatro, de acordo com as patologias da base de dados. Foi utilizado o otimizador **Adam** e a métrica de avaliação foi a *accuracy*. Também foi usada um *dropout* de quarenta por cento dos neurônios da CNN, para regularizar e diminuir o *overfitting*.

Essa porcentagem foi definida após realizações de testes com variações nas porcentagens de *dropout* nas camadas da CNN. Com variação de 10% a 40% na primeira camada de *dropout*, variação de 20% a 40% na segunda camada de *dropout* da CNN e variação de 20% a 50% na camada densa da CNN, última camada de *dropout*. A menor porcentagem de acerto da patologia COPD foi de 75,65% de acerto com a configuração das três camadas de *dropout* sendo 30%, 20% e 20%, respectivamente, e o melhor resultado com 83,7% foi colocar as três camadas com *dropout* de 40%, valores que foram usados para este estudo.

Cada época é composta de 32 arquivos de áudio para treinamento e, após cada treina-

mento, é feita uma validação com os áudios separados para isso. As variáveis como quantidade de época, número de interações e quantidade de camadas da CNN não serão estudadas neste estudo. Este estudo tem como foco a diferença entre a performance de uma mesma CNN, mas com base de dados diferentes, em relação se teve o pré-processamento da separação harmônica ou sem esse pré-processamento, onde por meio de comparações entre os resultados estatísticos dessa interações, será observado se este pré-processamento de separação pode progredir para testes com profissionais de saúde, afim de melhorar a percepção dos sons característicos de cada patologia e/ou melhorar as CNN ou outras redes neurais nas classificação de patologias pulmonares.

3.4 Procedimentos de análise

Este trabalho incluir um processamento de sinal que destaca os sons produzidos pelos pulmões e um modelo de rede neural convolucional, logo para avaliar a performance deste pré-processamento, será usada uma CNN para fazer a validação os dois tipos de áudios.

Comparando as saídas dos classificadores da CNN utilizando dois tipos de dados de áudio, onde em uma análise será dado como entrada os áudios cardiopulmonares em sua forma bruta sem a extração de qualquer característica dos arquivos e na outra análise será usado os mesmos arquivos de áudios só que para diferenciação de análise será feito um pré-processamento que deixará os sons pulmonares em destaque. A comparação desses resultados será importante para este estudo já que poderemos analisar se o pré-processamento imposto aos arquivos de áudios melhorou o desempenho da CNN, foi irrelevante para os resultados ou se piorou os resultados da CNN, já que uma possibilidade e que o pré-processamento possa suprimir características importantes quando a CNN discrimina os arquivos nos classificadores existentes.

Uma forma de análise são os parâmetros da própria CNN, que será avaliada com base nos modelos testados. Durante essa avaliação, serão exibidas métricas de precisão e perda para cada base de dados diferente. Os gráficos serão utilizados para visualizar o desempenho da CNN em diferentes épocas de avaliação, facilitando a seleção da melhor forma de tratar a base de dados.

Após a análise do desempenho da CNN, serão estabelecidos parâmetros estatísticos específicos para cada classe. No entanto, apenas os dados da classe COPD serão consi-

derados, uma vez que essa não passou por modelagem ou aumento artificial de amostras, preservando a integridade dos dados originais. Os parâmetros estatísticos a serem analisados incluem a precisão da classificação (*accuracy*), a taxa de revocação (*recall*), que mede a proporção de amostras corretamente identificadas dentro da classe, e o *F1-score*, métrica que equilibra precisão e revocação para avaliar a eficácia do modelo de maneira mais robusta.

4 Resultados e Discussões

A primeira pesquisa realizada, testando essa função de separação harmônica e percussiva foi com a utilização de apenas um áudio proveniente de uma ausculta cardiopulmonar, tendo como parâmetro de avaliação da função de separação o próprio sistema auditivo humano. Observando a eficiência na separação das diferentes componentes sonoras, percebe-se que há uma atenuação da outra componente; no entanto, foram mesclados alguns fragmentos sonoros. Mas estes resultados preliminares dão esperança para progredir no estudo de possíveis aplicação deste processamento no áudio em uma ausculta cardiopulmonar. Para uma melhor análise do áudio é recomendado o uso de fones de ouvido e, se possível, está um ambiente silencioso. Os links em azul apontam para arquivos armazenados no google Drive, logo é recomendado usar o navegador do google Chrome e estar logado em conta própria. No caso de acesso em condições diferentes e a reprodução do áudio no navegador não ser possível, existe a possibilidade de fazer o download do arquivo de áudio na mesma página. Os arquivos de áudios estão na extensão .mp3.

Este é o link do áudio normal, a importância de sua análise é observar o conteúdo bruto, sem a separação das componentes musicais. É o mesmo áudio descrito na figura 4.1.

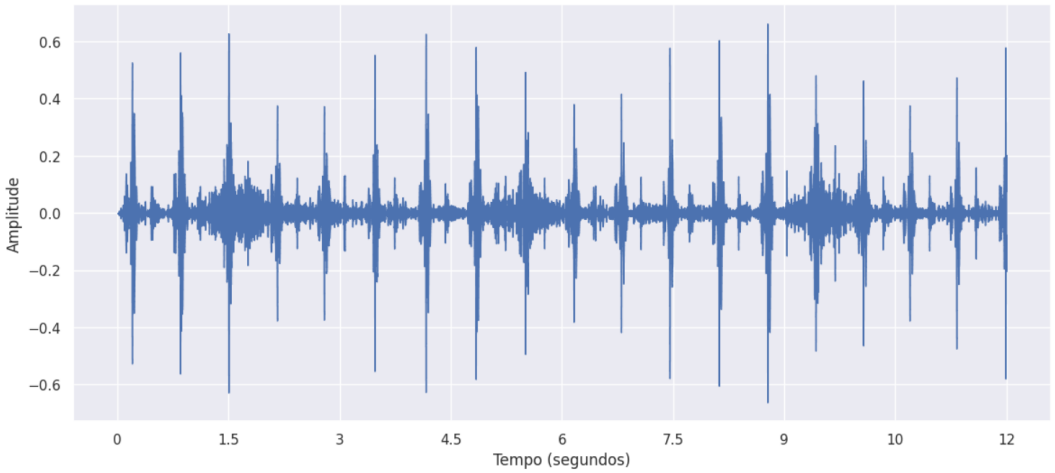


Figura 4.1. Onda sonora da ausculta - Em uma representação pelas grandezas de Amplitude da onda pelo Tempo.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

Este é o link do áudio da parte harmônica, que em sua maioria é composto dos sons gerados pelo fluxo de ar no sistema respiratório. Na figura 4.2 podemos observar se as características sonoras predominantes estão nos sons de baixa amplitude, pois as partes harmônicas tendem a mudar de frequência, sem mudar tanto a intensidade sonora.

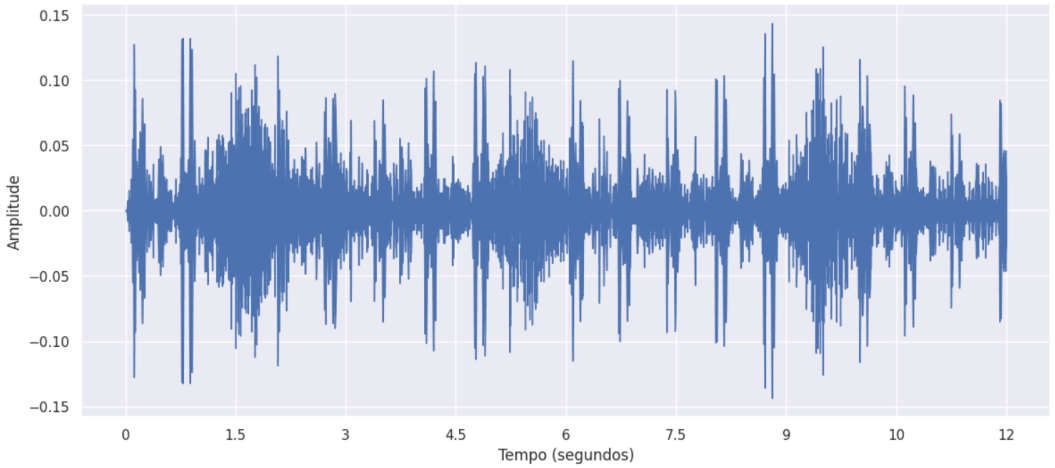


Figura 4.2. Forma de onda sonora da parte harmônica, onde se ameniza os sons de grande amplitude, destacando os sons de média e baixa amplitude. Em uma representação pelas grandezas de Amplitude da onda pelo Tempo.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

Este é o link do áudio da parte percussiva. Um som percussivo se assemelha ao sinal de um impulso unitário, no qual há uma grande amplitude. Observando a figura 4.3, o áudio expresso é composto predominantemente das amplitudes com mais intensidades, caracterizando o som reproduzido do coração.

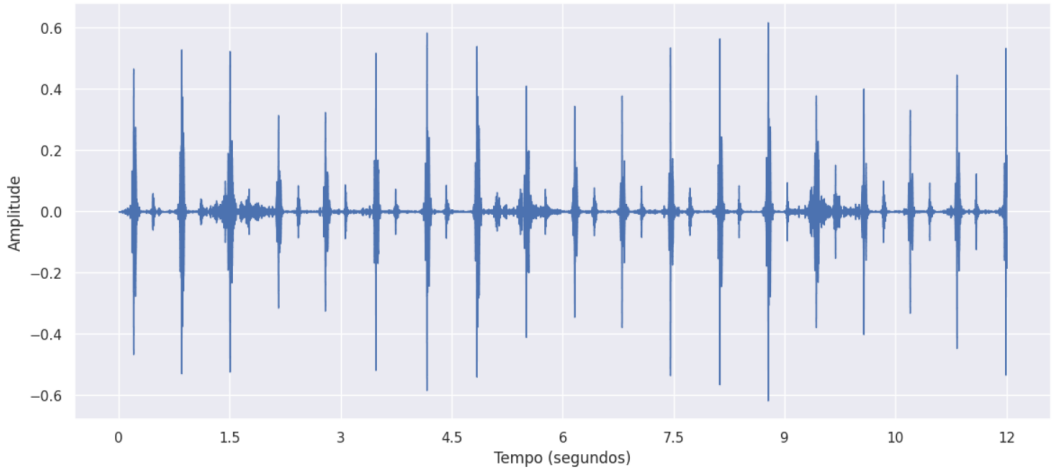


Figura 4.3. Forma de onda sonora da parte percursiva, em que se destaca os sons com uma elevada amplitude e minimização dos sons de baixa e média amplitude. A representação do som está pela amplitude da onda no tempo.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

A representação por espectrograma como na figura 4.4 fornece uma representação tridimensional do sinal de áudio, onde o eixo horizontal representa o tempo, o eixo vertical representa a frequência e a intensidade da cor indica a amplitude do sinal em diferentes frequências ao longo do tempo. Isso é essencial para analisar como as características sonoras evoluem ao longo do tempo, permitindo identificar componentes específicas de um áudio, como tons, harmônicos, ruídos, transientes e outros eventos sonoros. Nessa figura podemos observar as diferenças entre os sinais que foram separados e comparar as porções que foram extraídas do som original.

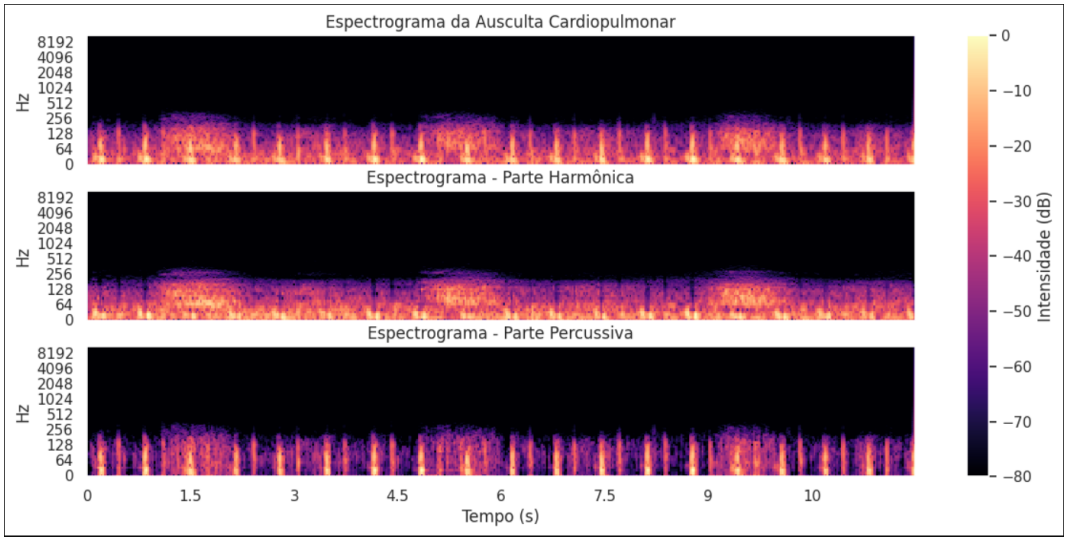


Figura 4.4. Espectrogramas dos três Áudios - Normal, Harmônico e Percussivo. O espectrograma é uma forma de visualização bidimensional da composição de um som, na qual é possível visualizar a amplitude da onda por uma escala de cores, as frequências pela altura dos pontos e a posição temporal pelo deslocamento lateral.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

Neste estudo também foi realizado uma análise usando uma CNN com o intuito de observar se a performance dessas separações por meio estatísticos e quais os resultados apresentados na análise, quando comparados.

Neste estudo, também foi realizada uma análise utilizando uma Rede Neural Convolucional (CNN) com o objetivo de avaliar a eficácia do processo de separação dos sons cardiopulmonares por meio de métricas estatísticas, analisando e comparando os resultados obtidos. A avaliação do desempenho da CNN permitirá verificar se o pré-processamento dos sinais sonoros contribui para uma melhor classificação das amostras ou se, eventualmente, impacta negativamente na capacidade de generalização do modelo.

A análise foi conduzida a partir de uma base de dados balanceada, composta por quatro classes distintas. A primeira classe corresponde a indivíduos sem doenças pulmonares, denominada "Normal", enquanto as demais representam patologias específicas: Asma, Pneumonia e Doença Pulmonar Obstrutiva Crônica (COPD). A uniformização das amostras entre essas classes foi essencial para garantir que a CNN pudesse aprender padrões de forma equilibrada, sem viés excessivo para uma única classe.

Para a extração de características dos áudios, foram utilizados os coeficientes cepstrais de frequência mel (MFCCs), uma técnica amplamente aplicada no processamento de sinais acústicos. Com base nesses coeficientes, a base de dados foi segmentada em conjuntos de treinamento, teste e validação, da divisão dos atributos das classes e os atributos previsores. Essa abordagem permitiu que a CNN fosse treinada a partir de características relevantes dos sinais de áudio, aumentando a robustez do modelo na diferenciação entre as classes.

Os espectrogramas extraídos dos arquivos de áudio podem ser interpretados como imagens bidimensionais (2D), o que possibilita a aplicação de técnicas avançadas de classificação de imagens para a análise desses dados. Especificamente, foram empregadas Redes Neurais Convolucionais (CNNs), que demonstram alto desempenho em tarefas de reconhecimento de padrões visuais e têm sido amplamente utilizadas para a classificação de sinais acústicos.

Tabela 4.1. Arquitetura da Rede Neural Convolucional(CNN) utilizada no estudo

Camada (Tipo)	**Formato de Saída**	**Parâmetros**
conv1d (Conv1D)	(None, 31, 64)	704
dropout (Dropout)	(None, 31, 64)	0
max_pooling1d (MaxPooling1D)	(None, 7, 64)	0
conv1d_1 (Conv1D)	(None, 7, 128)	82,048
activation (Activation)	(None, 7, 128)	0
dropout_1 (Dropout)	(None, 7, 128)	0
max_pooling1d_1 (MaxPooling1D)	(None, 1, 128)	0
flatten (Flatten)	(None, 128)	0
dense (Dense)	(None, 64)	8,256
dropout_2 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 4)	260
activation_1 (Activation)	(None, 4)	0
	Total de Parâmetros	**91,268 (356.52 KB)**

A arquitetura da rede neural foi definida após uma série de testes e ajustes experimentais. A estrutura da CNN adotada segue as camadas representadas na Figura 4.1, totalizando 91.268 parâmetros treináveis. Para o treinamento do modelo, foram realizadas 84 épocas, utilizando lotes (*batches*) de 32 arquivos de áudio por iteração. Esse número de épocas foi determinado com base na convergência dos erros e no comportamento das curvas de aprendizado, evitando tanto o sobreajuste (*overfitting*) quanto o subajuste (*underfitting*).

Os resultados obtidos foram avaliados por meio das métricas de precisão (*accuracy*) e função de perda (*loss*). As figuras 4.5 e 4.6 apresentam as curvas de *accuracy*, permitindo a observação do desempenho do modelo ao longo do treinamento e da validação. Já as figuras 4.7 e 4.8 ilustram as métricas de *loss*, fornecendo informações sobre a estabilidade do aprendizado e a eficiência do processo de otimização. A análise dessas métricas é essencial para compreender o impacto do pré-processamento na capacidade de classificação da CNN e, consequentemente, na distinção entre as diferentes classes de amostras analisadas.

Nas tabelas 4.2 e 4.3 podemos observar os resultados estatísticos. Em uma primeira analise observamos um baixo desempenho da rede neural que teve a separação harmônica mas no entanto se compararmos só o COPD, classe essa, que não sofreu modificação apresentou um resultado melhor de 5,3% em relação aos parâmetros COPD que não teve

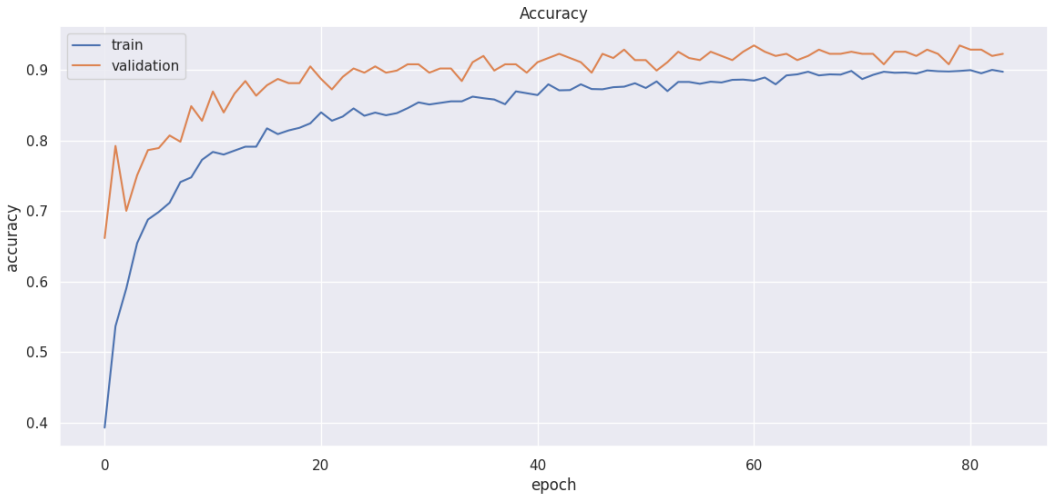


Figura 4.5. Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados sem a separação.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

modificações no áudio. As principais métricas utilizadas foram:

A Precisão que avalia a proporção de acertos entre as previsões positivas feitas pelo modelo, ou seja, a porcentagem de previsões positivas que realmente pertencem à classe correta. Um alto valor de precisão indica que há poucos falsos positivos.

E o recall que mede a capacidade do modelo de identificar corretamente os casos positivos dentro do conjunto de dados, ou seja, a proporção de instâncias relevantes que foram corretamente classificadas. Um recall elevado indica que poucos casos positivos foram erroneamente classificados como negativos.

Além dessas métricas, também foi analisado o f1-score, que representa a média harmônica entre a precisão e o recall, equilibrando ambas as métricas para oferecer uma visão mais robusta do desempenho do modelo. A acurácia geral foi utilizada para avaliar a proporção total de classificações corretas dentro do conjunto de testes. Essas métricas permitem uma avaliação abrangente da eficácia do modelo, identificando suas forças e possíveis limitações na discriminação entre as diferentes classes analisadas. Podemos também observar as matriz de confusão nas figuras 4.9 e 4.10 que mostram o desempenho das CNNs na fase de teste com novas amostras. E quais patologias são confundidas com outra classes.

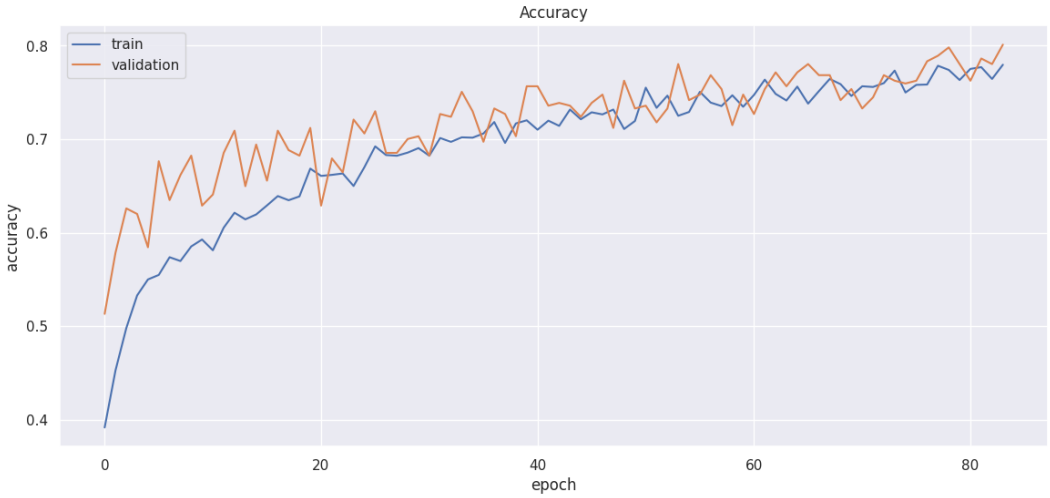


Figura 4.6. Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados com a separação.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

Tabela 4.2. Métricas de desempenho usando a base de dados sem a separação harmônica.

Classe	**Precisão**	**Recall**	**F1-score**	**Suporte**
Asma	0.83	0.98	0.90	93
COPD	0.93	0.90	0.91	87
Normal	0.97	0.91	0.94	76
Pneumonia	0.97	0.86	0.91	80
Acurácia		0.91 (336 amostras)		
Média Macro	0.92	0.91	0.92	336
Média Ponderada	0.92	0.91	0.91	336

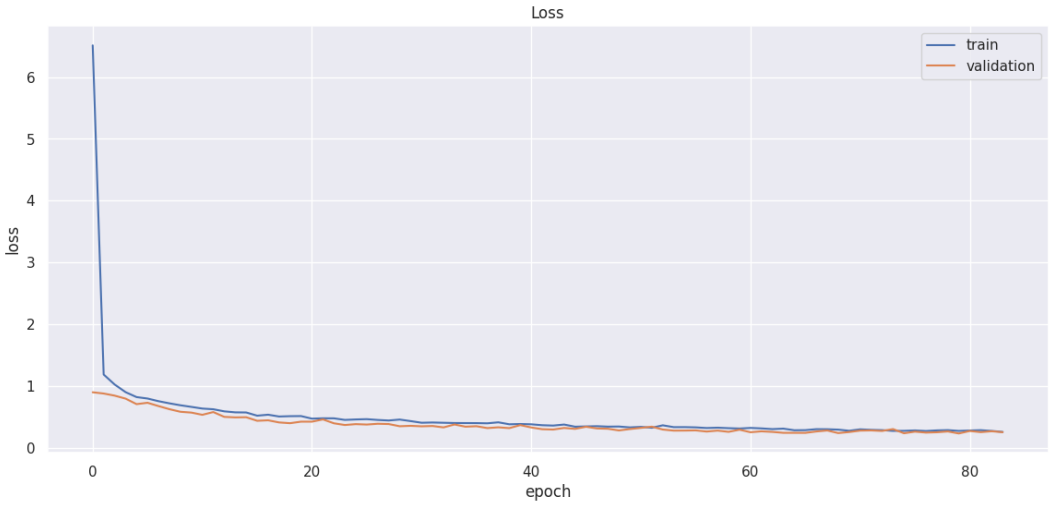


Figura 4.7. Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados sem a separação.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

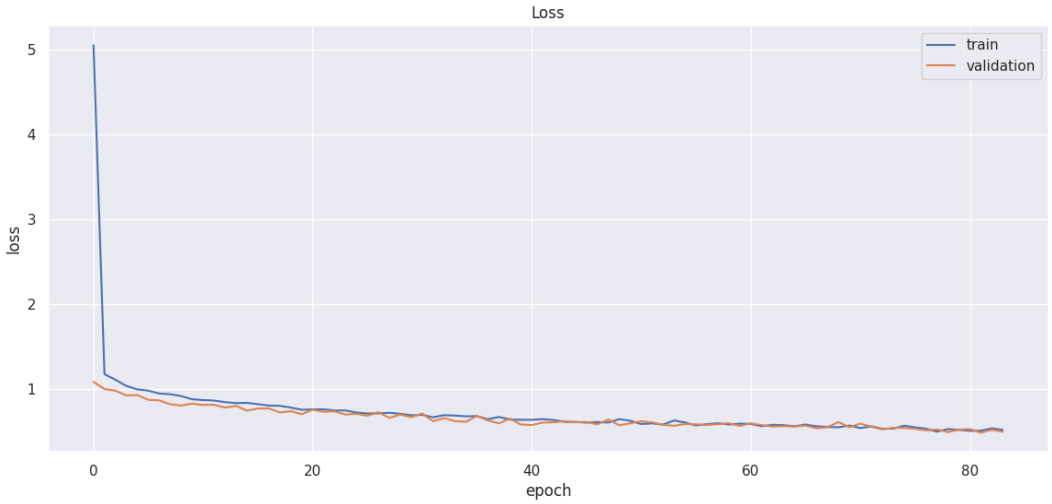


Figura 4.8. Gráfico da acurácia durante o treinamento da rede neural, usando a base de dados com a separação.

Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

Tabela 4.3. Métricas de desempenho usando a base de dados sem a separação harmônica.

Classe	**Precisão**	**Recall**	**F1-score**	**Suporte**
Asma	0.68	0.71	0.69	93
COPD	1.00	0.89	0.94	87
Normal	0.68	0.66	0.67	76
Pneumonia	0.82	0.91	0.86	80
Acurácia		0.79 (336 amostras)		
Média Macro	0.80	0.79	0.79	336
Média Ponderada	0.80	0.79	0.79	336

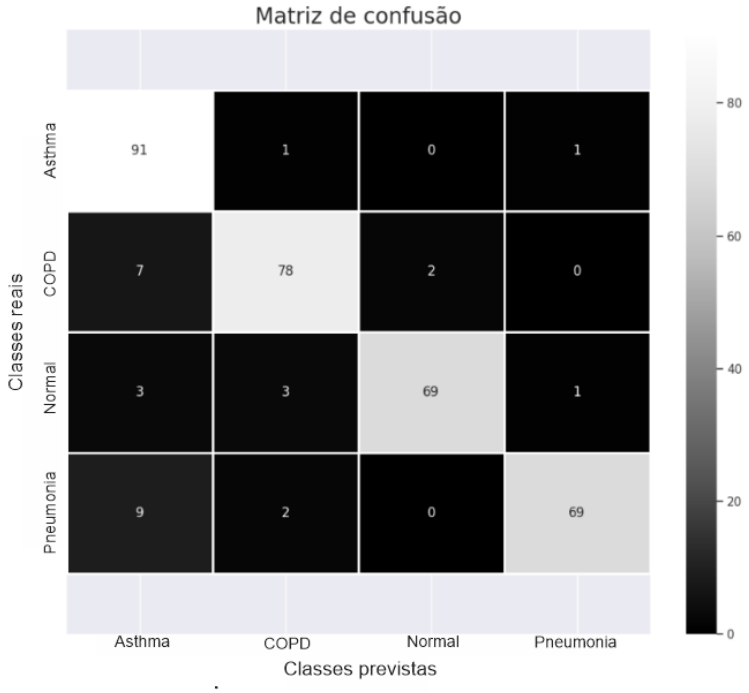


Figura 4.9. Gráfico da Matriz de Confusão da base de dados sem o processo de separação, onde podemos comparar as previsões do modelo com os resultados reais.
Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

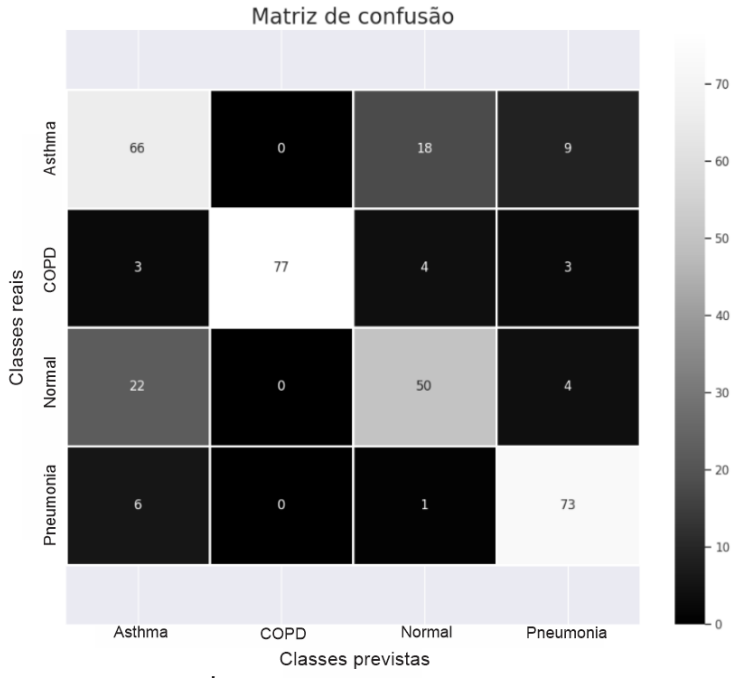


Figura 4.10. Gráfico da Matriz de Confusão da base de dados com o processo de separação, onde podemos comparar as previsões do modelo com os resultados reais.
Fonte: Elaborada pelo autor, a partir da linguagem python. 2025.

5 Conclusão

Este estudo abordou diversas etapas fundamentais para a análise e classificação de patologias pulmonares a partir de sinais de auscultas cardiopulmonares. Inicialmente, foi realizado um pré-processamento dos dados visando a remoção seletiva dos sons predominantemente gerados pelos pulmões e pelos batimentos cardíacos. Esse procedimento permitiu a obtenção de sinais acústicos mais distintos, facilitando a separação e a análise das características relevantes para cada tipo de som.

Além disso, um modelo de rede neural convolucional (CNN) foi desenvolvido com o propósito de avaliar a eficácia do pré-processamento na melhoria do desempenho da classificação sem a necessidade de uma avaliação feita do profissionais da área da saúde. Os resultados obtidos demonstraram que a utilização de técnicas de destaque dos sons pulmonares pode impactar a precisão da CNN, contribuindo para uma melhor distinção entre diferentes condições patológicas. As métricas de *precision*, *recall* e *f1-score* foram empregadas para mensurar o desempenho do modelo, evidenciando que a abordagem adotada tem potencial para auxiliar no diagnóstico de doenças respiratórias.

Embora o foco deste estudo tenha sido a análise de patologias pulmonares, a metodologia aplicada pode ser expandida para investigações futuras envolvendo a classificação de doenças cardíacas, uma vez que a extração e a separação dos sinais cardíacos também foram consideradas ao longo do processo.

Dessa forma, este trabalho não apenas contribui para o desenvolvimento de modelos mais robustos na identificação de doenças respiratórias, como também abre caminho para estudos futuros que explorem a combinação de sinais pulmonares e cardíacos na detecção de diferentes condições clínicas. Com a expansão e refinamento da base de dados, bem como a aplicação de técnicas avançadas de aprendizado profundo, espera-se alcançar modelos cada vez mais precisos e eficazes para uso na prática médica.

5.1 Propostas para Estudos Futuros

Para aprimorar as pesquisas e expandir os resultados obtidos neste estudo, algumas direções futuras são sugeridas:

- **Desenvolvimento de uma base de dados**: A criação de uma base de dados mais equilibrada e abrangente é essencial para garantir um treinamento mais robusto e representativo das redes neurais, melhorando a generalização dos modelos aplicados à análise de sinais cardiopulmonares.
- **Estudo com profissionais e estudantes da área da saúde**: Realizar avaliações com especialistas e acadêmicos da área médica permitirá validar tecnicamente os métodos utilizados, proporcionando um respaldo fundamentado na prática clínica e no contexto diário da semiologia médica.
- **Aprimoramento dos algoritmos de redes neurais**: Após a construção de uma base de dados adequada, será possível explorar novos algoritmos de redes neurais e otimizar os parâmetros já estabelecidos neste estudo, visando melhorias na acurácia e eficiência dos modelos.
- **Realizar testes para viabilizar uso na telemedicina**: Após os resultados obtidos dos algoritmos, testar o uso desse conceito com paciente voluntários em umcontexto de consulta de telemedicina assistida por um profissional da saúde.

Referências Bibliográficas

[1] CAMPANA, Álvaro O. Exame Clínico - Sintomas e Sinais em Clínica Médica. Grupo GEN, 2010. E-book. ISBN 978-85-277-1955-1. Disponível em: https://integrada.minhabiblioteca.com.br//books/978-85-277-1955-1/. Acesso em: 28 set. 2023.

[2] A visualized overview of global and regional trends in the leading causes of death and disability 2000-2019. World Health Organization - (WHO): https://www.who.int/data/stories/leading-causes-of-death-and-disability-2000-2019-a-visual-summary. Acessado em 01/11/2023.

[3] What is COPD?. Disponível em: https://www.copdfoundation.org/What-is-COPD/Understanding-COPD/What-is-COPD.aspx. Acesso em: 03 mar. 2025.

[4] Chronic obstructive pulmonary disease (COPD) fact sheet. Overview about chronic obstructive pulmonary disease (COPD). Disponível em: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd). Acesso em: 03 mar. 2025.

[5] Chronic obstructive pulmonary disease (COPD). National Health Service (NHS). Disponível em: https:https://www.nhs.uk/conditions/chronic-obstructive-pulmonary-disease-copd/. Acesso em: 03 mar. 2025.

[6] De L'Auscultation Médiate; ou Traité du Diagnostic des Maladies des Poumons et du Cœur, fondé principalement sur ce Nouveau Moyen d'Exploration. *Edinburgh Medical and Surgical Journal*, 18(72):447, July 1822. Publisher: Biomedical Journal Digitization Project.

[7] Richard A Reinhart. George phillip cammann (1804–1863): A physician's contribution to the modern stethoscope and auscultatory percussion. *Journal of Medical Biography*, 31(4):234–245, 2023. PMID: 34279155.

[8] Evolução da internet no Brasil. Portal da Rede Nacional de Ensino e Pesquisa: https://www.rnp.br/noticias/evolucao-da-internet-no-brasil. Acessado em 28/10/2023.

[9] Eco-92, 30 anos de uma experiência pioneira de Internet. Revista Eletrônica de Jornalismo Científico: https://www.comciencia.br/eco-92-30-anos-de-uma-experiencia-pioneira-de-internet. Acessado em 30/10/2023.

[10] Evolução da internet no Brasil. Rede Nacional de Ensino e Pesquisa – (RNP): https://www.rnp.br/noticias/evolucao-da-internet-no-brasil. Acessado em 25/10/2023.

[11] Resolução da CREMESP nº97, de 20 de fevereiro de 2001. Portal do Conselho Nacional de Medicina: https://sistemas.cfm.org.br/normas/visualizar/resolucoes/SP/2001/97. Acessado em 25/10/2023.

[12] Resolução da CFM – nº 1.643,de 26 de agosto de 2002. Portal do Conselho Nacional de Medicina: https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2002/1643. Acessado em 25/10/2023.

[13] Resolução da CFM – nº 2.227/2018. Portal do Conselho Nacional de Medicina: https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2018/2227. Acessado em 25/10/2023.

[14] Resolução da CFM – nº 2.228 de 06 de março de 2019. Portal do Conselho Nacional de Medicina: https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2019/2228. Acessado em 25/10/2023.

[15] Imprensa Nacional. BRASIL. PORTARIA Nº 188, DE 3 DE FEVEREIRO DE 2020. Dispõe sobre a declaração de Emergência em Saúde Pública de Importância Nacional (ESPIN) em decorrência da Infecção Humana pelo novo Coronavírus do Ministério da Saúde. Diário Oficial da República do Brasil. Brasília,DF, Edição: 24-A, Seção: 1 - Extra, p.1, fev,2020.

[16] Imprensa Nacional. PORTARIA Nº 467, DE 20 DE MARÇO DE 2020 – DOU. Portal da Imprensa Nacional do Brasil: https://www.in.gov.br/web/dou. Acessado em 30/10/2023.

[17] Imprensa Nacional. LEI Nº 13.989, DE 15 DE ABRIL DE 2020 - DOU. Portal da Imprensa Nacional do Brasil: https://www.in.gov.br/web/dou. Acessado em 30/10/2023.

[18] Agência IBGE. Indicadores IBGE : pesquisa nacional por amostra de domicílios contínua: https://painel.ibge.gov.br/pnadc/. Acessado em 30/10/2023.

[19] Ministério da Saúde declara fim da Emergência em Saúde Pública de Importância Nacional pela Covid-19. Portal do Ministério da Saúde: https://www.gov.br/saude/pt-br/assuntos/noticias/2022/abril. Acessado em 01/11/2023.

[20] Resolução do CFM nº 2.314, de 05 de maio de 2022. Portal do Conselho Nacional de Medicina: https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2022/2314. Acessado em 30/10/2023.

[21] Ministério da Saúde. PORTARIA GM/MS Nº 1.348, DE 2 DE JUNHO DE 2022. Dispõe sobre as ações e serviços de Telessaúde no âmbito do Sistema Único de Saúde (SUS): https://bvsms.saude.gov.br/bvs/saudelegis/gm/2022/prt1348_03_06_2022.html Acessado em 30/10/2023.

[22] CAMPANA, Álvaro O. Exame Clínico - Sintomas e Sinais em Clínica Médica: Grupo GEN, 2010. E-book. ISBN 978-85-277-1955-1. Disponível em: https://integrada.minhabiblioteca.com.br//books/978-85-277-1955-1/. Acesso em: 03 nov. 2023.

[23] Gallery | H01 Release.Lichtman laboratory at Harvard University and Connectomics at Google team . Neurônios do lobo temporal do cortex cerebral. Disponível em:https://h01-release.storage.googleapis.com/gallery.html Acesso em: 10 abr. 2024.

[24] Equipe DSA. Capítulo 4 - O Neurônio, Biológico e Matemático. Deep Learning Book., December 2022. Disponível em: https://www.deeplearningbook.com.br/o-neuronio-biologico-e-matematico. Acesso em: 10 abr. 2024.

[25] Redes Neurais Artificiais. Disponível em: https://sites.icmc.usp.br/andre/research/neural. Acesso em: 10 abr. 2024.

[26] Murillo Grübler. Entendendo o funcionamento de uma Rede Neural Artificial, June 2018. aibrasil. Disponível em:https://medium.com/brasil-ai/entendendo-o-funcionamento-de-uma-rede-neural-artificial-4463fcf44dd0, Acesso em: 9 abr. 2024.

[27] Image Kernels explained visually. Explained Visually. Disponível em: https://setosa.io/ev/image-kernels. Acesso em: 10 abr. 2024.

[28] Balaji C. Convolutional Neural Network(CNN), January 2023. C onvolutional Neural Network(CNN) are type of feed-forward neural network whose connectivity structure. Medium. Disponível em:https://medium.com/@balajicena1995/convolutional-neural-network-cnn-7acf8ede9d49, Acesso em: 9 abr. 2024.

[29] MASSON, Antonio A.; MUXFELDT, Elizabeth; SUBILHAGA, Janice; et al. Semiologia Essencial na Prática Médica: O Que Todo Clínico Deve Saber. Thieme Brazil, 2022. E-book. ISBN 9786555721089. Disponível em: https://integrada.minhabiblioteca.com.br//books/9786555721089/. Acesso em: 28 set. 2023.