Universidade Federal de Ouro Preto
Escola de Minas
CECAU - Colegiado do Curso de
Engenharia de Controle e Automação

Thallys Augusto Clemente

Inteligência computacional para auxílio no diagnóstico de alterações nas orelhas externa e média

Monografia de Graduação

Ouro Preto, 2024

Thallys Augusto Clemente

Inteligência computacional para auxílio no diagnóstico de alterações nas orelhas externa e média

Trabalho apresentado ao Colegiado do Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheira(o) de Controle e Automação.

Universidade Federal de Ouro Preto

Orientador: Profa. Adrielle de Carvalho Santana, Dr^{a}

Coorientador: Prof. Mateus Coelho, Dr.

Ouro Preto

2024

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

FOLHA DE APROVAÇÃO

Thallys Augusto Clemente

"Inteligência computacional para auxílio no diagnóstico de alterações nas orelhas externa e média"

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

Aprovada em 07 de fevereiro de 2024

Membros da banca

Dra. Adrielle de Carvalho Santana - Orientadora (Universidade Federal de Ouro Preto)
M.Sc. Mateus Coelho Silva - Coorientador (Universidade Federal de Ouro Preto)
Dra. Carla Aparecida de Vasconcelos - Convidada (Superintendência Central de Perícia Médica e Saúde Ocupacional, SEPLAG - MG)
Dr. Pedro Henrique Lopes Silva - Convidado (Universidade Federal de Ouro Preto)

Adrielle de Carvalho Santana, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 09/02/2024

Documento assinado eletronicamente por **Adrielle de Carvalho Santana**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 09/02/2024, às 07:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0664817** e o código CRC **B42BBCF5**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: 3135591533 - www.ufop.br

Agradecimentos

Gostaria de expressar meu mais profundo agradecimento a todos que contribuíram para a realização deste trabalho. Em especial, dedico minha gratidão aos meus orientadores, cujo conhecimento e orientação foram fundamentais para o desenvolvimento deste trabalho.

À minha querida família, expresso minha sincera gratidão pelo amor incondicional e apoio contínuo. Suas palavras de incentivo foram a força motriz por trás de cada passo dado nesta trajetória desafiadora.

Agradeço de coração aos meus amigos, pelo apoio e amizade ao longo dessa jornada acadêmica. Cada momento compartilhado, desde os desafios até as celebrações, enriqueceu significativamente essa experiência. A presença de vocês, com palavras de incentivo nos momentos difíceis e celebrações nas conquistas, iluminou esta trajetória, tornando-a mais leve e memorável.

Aos amigos que se tornaram companheiros de viagem no ônibus rumo à faculdade, meu agradecimento pela amizade constante. Cada risada compartilhada e desafio superado juntos contribuíram para tornar esta jornada acadêmica ainda mais significativa.

A todos que, de alguma forma, estiveram presentes ao longo desta jornada, meu sincero agradecimento. Este trabalho é também fruto do apoio e contribuição de cada um de vocês.

Ao encerrar este capítulo, reflito sobre a inspiradora frase de Gandhi: “Seja a mudança que você quer ver no mundo”. Este trabalho é o resultado de um esforço coletivo, e espero sinceramente que possa contribuir para o avanço e aprimoramento do conhecimento na área.

"Matéria é a parte acidental." (Oliver Lodge)

Resumo

A otoscopia digital é uma ferramenta valiosa na prática clínica, desempenhando um papel importante no auxílio ao diagnóstico rápido e preciso de alterações da orelha externa e interna no aparelho auditivo, contribuindo, assim, para uma melhor qualidade de cuidado e intervenção de saúde. Nesse cenário, a aplicação de Inteligência Artificial tem emergido como uma abordagem promissora para aprimorar a precisão e eficiência do auxílio a diagnóstico. Diante desse contexto, este estudo visa explorar o emprego de Redes Neurais Convolucionais (CNN) na classificação de imagens, com o objetivo de identificar alterações da orelha externa e interna dos pacientes. A pesquisa se propõe a investigar a capacidade das CNNs em auxiliar na triagem, avaliando o desempenho de diferentes arquiteturas em classificar as imagens em grupos distintos. Dessa forma, este projeto não apenas busca identifica as arquiteturas de CNN mais eficazes para estudos futuros, mas também sugere que tais algoritmos têm o potencial de apoiar a avaliação da área da saúde. Ao direcionar a atenção do profissional para áreas de interesse específicas nas imagens da orelha externa e interna, essa abordagem pode reduzir o tempo necessário para diagnósticos assertivos. Destacam-se os resultados notáveis obtidos pelas arquiteturas de CNN, com ênfase especial na *convnext tiny*. Esta arquitetura alcançou um *F1 Score* de 0.9637, especialmente ao lidar com a base de dados específica para otite. O *recall* de 0.9797 destaca a habilidade da *convnext tiny* em capturar quase todas as instâncias positivas de otite, enquanto a precisão de 0.9617 demonstra sua acurácia ao classificar casos positivos. Os resultados obtidos pela *convnext tiny* indicam a eficácia dessa arquitetura na identificação de diversos tipos de otite. No entanto, é crucial reconhecer a complexidade e diversidade das patologias auditivas. O desafio central é a compreensão de que um único modelo pode não ser uniformemente eficiente em diversas doenças do orelha média, sendo interessante investigar a classficação de outras patologias assim como o uso de diferentes bases de dados.

Palavras-chaves: Otoscópio digital; Visão computacional; Inteligência computacional; Otoscopia; Rede neural convulocional; Processamento de imagens médicas; auxílio a diagnóstico na área da saúde; Orelha média; Orelha externa; Machine learning; Classificação de dados;

Abstract

Digital otoscopy is a valuable tool in clinical practice, playing an important role in aiding the rapid and accurate diagnosis of alterations in the external and internal ear within the auditory system, thus contributing to better quality of care and health intervention. In this scenario, the application of Artificial Intelligence has emerged as a promising approach to enhancing the accuracy and efficiency of diagnostic aid. Within this context, this study aims to explore the use of Convolutional Neural Networks (CNNs) in image classification, with the goal of identifying alterations in the external and internal ear of patients. The research aims to investigate the ability of CNNs to assist in screening, evaluating the performance of different architectures in classifying images into distinct groups. Thus, this project not only seeks to identify the most effective CNN architectures for future studies but also suggests that such algorithms have the potential to support health area assessment. By directing the professional's attention to specific areas of interest in external and internal ear images, this approach can reduce the time required for accurate diagnoses. Noteworthy results were achieved by CNN architectures, with special emphasis on the convnext tiny. This architecture achieved an F1 Score of 0.9637, especially when dealing with the specific otitis database. The recall of 0.9797 highlights the ability of convnext tiny to capture almost all positive instances of otitis, while the precision of 0.9617 demonstrates its accuracy in classifying positive cases. The results obtained by convnext tiny indicate the effectiveness of this architecture in identifying various types of otitis. However, it is crucial to recognize the complexity and diversity of auditory pathologies. The central challenge is understanding that a single model may not be uniformly efficient in various middle ear diseases, making it interesting to investigate the classification of other pathologies as well as the use of different databases.

Key-words: Digital Otoscope; Computer Vision; Computational Intelligence; Otoscopy; Convolutional Neural Network; Medical Image Processing; Medical Diagnostic Aid; Middle Ear; Outer Ear; Machine Learning; Data Classification.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

CAE	Canal Auditivo Externo
CNN	*Convolutional Neural Networks*
IA	Inteligência Artificial
MT	Membrana Timpânica
OM	Orelha Média

Sumário

1 Introdução

1.1 Contextualização

A orelha humana é parte importante do sistema auditivo e ela é dividido, como mostra a Figura 1 em: orelha externa, médio e interno. A orelha médio é composta pelo tímpano, uma câmara cheia de ar, que contém uma cadeia de três ossos pequenos (martelo, bigorna, estribo) e a entrada para a tuba auditiva. A orelha média é de suma importância para a realização da transmissão do orelha externa para o orelha interna.

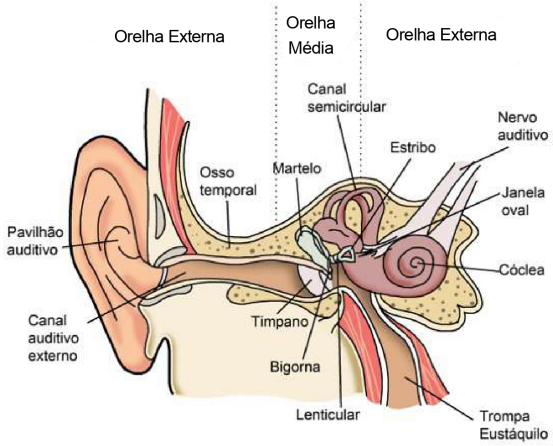

Figura 1 – Anatomia do orelha humana.

Fonte: Anatomia, 2021 [1]

A otoscopia desempenha um papel fundamental na prática da área da saúde, permitindo um acesso direto e rápido ao Meato Acústico Externo(MAE) e à membrana timpânica (MT). É uma técnica de extrema importância no diagnóstico de diversas patologias relacionadas ao orelha média (OM) (SILVA, 2018b). Por meio da otoscopia, é possível identificar condições como otite média serosa, retração da membrana timpânica, otite média crônica com perfuração, timpanosclerose, otite externa aguda difusa, presença de corpos

[1] https://www.anatomiaemfoco.com.br/aparelho-auditivo-audicao/ouvido-humano-anatomia/

estranhos na membrana timpânica, hemotímpano e obstrução do meato acústico externo por rolha de cera.

Vale ressaltar que ao longo do tempo, a técnica da otoscopia evoluiu consideravelmente, culminando na versão moderna que faz uso do otoscópio digital. Esse avanço proporciona uma avaliação mais precisa e detalhada das condições da orelha, melhorando assim a qualidade do diagnóstico e o acompanhamento clínico.

O primeiro otoscópio conhecido foi o de Brunton, como pode ser observado na Figura 2. Este dispositivo consistia em um cilindro de latão com um espéculo auricular em uma extremidade e uma lente de ampliação na outra. Com o passar dos anos e o avanço da tecnologia elétrica, novas melhorias foram incorporadas ao projeto original do otoscópio de Brunton (SILVA, 2018a).

Uma das principais evoluções ocorreu no sistema de iluminação. Os modelos posteriores começaram a incluir lâmpadas em seu interior, proporcionando uma fonte de luz mais eficiente. A grande revolução no otoscópio veio com o desenvolvimento do sistema de iluminação alimentado por pilhas. Os otoscópios modernos utilizam fontes de luz de halogênio, transmitidas por fibras óticas dispostas ao redor de todo o espéculo auricular, como pode ser visto na Figura 3.

Essas inovações permitiram uma visualização mais precisa do canal da orelha média. Além disso, os avanços tecnológicos levaram ao desenvolvimento do otoscópio digital, que oferece ainda mais recursos para a avaliação e diagnóstico das condições da orelha.

Figura 2 – Otoscópio de Brunton

Fonte: historiadelamedicina,2021 [2]

[2] https://www.historiadelamedicina.org

Figura 3 – Otoscópio digital

Fonte: Medicial,2021 [3]

Visão computacional é a ciência responsável pela visão de uma máquina, pela forma como um computador enxerga o meio à sua volta, extraindo informações significativas a partir de imagens capturadas por câmeras de vídeo, sensores, *scanners*, entre outros dispositivos. Estas informações permitem reconhecer, manipular e pensar sobre os objetos que compõem uma imagem (BALLARD; BROWN, 1982). Essa capacidade permite que os computadores reconheçam, analisem e compreendam os objetos presentes em uma imagem.

Nos últimos anos, a tecnologia de visão computacional tem ganhado destaque em diversas áreas, impulsionando estudos e pesquisas em diversas aplicações. Devido à sua versatilidade, ela também pode ser incorporada no campo da área da saúde, com o propósito de auxiliar profissionais como otorrinolaringologistas, fonoaudiólogos, clínicos gerais e pediatras no diagnóstico de alterações relacionadas a orelha.

As *Convolutional Neural Networks* (CNNs) são amplamente empregadas na visão computacional para tarefas como classificação de imagens, detecção de objetos e segmentação. Essas redes especializadas utilizam camadas de convolução para extrair características fundamentais, seguidas por camadas de *pooling* para redução de dimensionalidade. A principal característica da CNN é a existência de camadas convolucionais que agem como campos receptivos de neurônios e tem como principal aplicação o processamento de informações visuais (PONTI et al., 2017). As CNNs são capazes de aprender padrões hierárquicos complexos em imagens, sendo essenciais para reconhecimento eficaz. O processo envolve preprocessamento de dados, construção da arquitetura, compilação, treinamento, validação, e técnicas avançadas como *transfer learning* e *data augmentation* de dados, garantindo a capacidade do modelo de generalizar e realizar previsões precisas em novos

[3] https://www.medicalcenterbr.com.br

dados (CHUNG et al., 2015)

O objetivo deste estudo é integrar a visão computacional com um otoscópio digital, com o propósito de fornecer suporte no diagnóstico de patologias. O *software* desenvolvido realiza uma análise de uma imagem capturada e, em seguida, identifica automaticamente áreas do orelha média que podem apresentar alguma alteração. Isso proporciona uma valiosa assistência aos profissionais de saúde, agilizando e aprimorando o processo de diagnóstico.

1.2 Justificativas e Relevância

A Inteligência Artificial (IA) tem desempenhado um papel fundamental no auxilio de tarefas rotineiras em clínicas e consultórios da a´rea da saúde, inclusive na capacidade de auxiliar no diagnostico de alterações do indivíduo. Essa tecnologia permite que os sistemas aprendam e melhorem por conta própria, reduzindo a dependência de entrada de dados humanos. Na área da saúde, a IA tem trazido inúmeros benefícios para a qualidade de vida e o bem-estar dos pacientes, revolucionando as práticas médicas e melhorando o atendimento personalizado em clínicas.

A IA desempenha um papel crucial na gestão médica, fornecendo auxilio em diagnósticos mais precisos e decisões clínicas mais assertivas. Ela se tornou um componente essencial na área da saúde, oferecendo um recurso promissor para abordar uma variedade de problemas de saúde (PATEL et al., 2009).

Nesse contexto, a justificativa deste estudo reside na importância fundamental de ampliar o conhecimento por meio da integração da visão computacional com imagens obtidas por um otoscópio digital. Ao explorar essa intersecção, este trabalho busca não apenas expandir o entendimento sobre auxilio em diagnósticos auditivos, mas também oferecer insights cruciais para aprimorar práticas clínicas e contribuir para o avanço da área da saúde.

Os resultados obtidos têm o potencial de fornecer informações valiosas que podem impactar positivamente a tomada de decisões por profissionais de saúde, melhorar a eficácia dos diagnósticos e, consequentemente, a qualidade dos cuidados prestados aos pacientes. Dessa forma, a integração da visão computacional com a análise de imagens de otoscópio digital não apenas representa um avanço técnico, mas também uma significativa contribuição para a prática médica e o conhecimento científico na área auditiva.

1.3 Objetivos

1.3.1 Objetivo Geral

Analisar a eficácia da aplicação de CNNs na classificação de imagens obtidas por um otoscópio digital para identificação de alterações na orelha média. O estudo visa avaliar o desempenho de diferentes arquiteturas de CNN no auxílio a diagnósticos na orelha média, contribuindo para aprimorar a eficiência do processo de avaliação na área da saúde.

1.3.2 Objetivo Especifico

1. Levantamento e organização de uma base de dados para treinamento do sistema computacional inteligente.
2. Teste de diferentes arquiteturas de CNN sem alterações.
3. Obtenção de uma boa arquitetura CNN e realização alteração das camadas de classificação do algortmo para o problema de classificação proposto.
4. Identificação de alterações na orelha externa e média utilizando o modelo proposto.

1.4 Organização e estrutura

No Capítulo 2 concentra-se na revisão bibliográfica, analisando estudos e teorias relacionados. No Capítulo 3, a metodologia adotada é detalhada, explicando métodos e escolhas. O Capítulo 4 apresenta experimentos e resultados, destacando ferramentas e o processo de desenvolvimento. No Capítulo 5, conclui-se o trabalho, discutindo a relevância dos resultados e sugestões para futuras pesquisas.

2 Fundamentação Teórica

Este capítulo tem como finalidade apresentar alguns conceitos, os algoritmos, otóscopios digitais e as tecnologias utilizadas no trabalho, para que seja possivel compreender como funciona.

2.1 Inteligência Artificial aplicada na área da saúde

A Inteligência Artificial está se tornando cada vez mais presente na área da saúde na atualidade, com diversas aplicações que visam aprimorar diagnósticos e tornar tratamentos mais eficazes. Ela tem a capacidade de explorar extensas bases de dados para identificar opções de tratamento mais viáveis com base em casos analisados. Neste trabalho, o foco está na aplicação de algoritmos de classificação na área da saúde. Abaixo, apresento dois exemplos de trabalhos em que algoritmos foram utilizados

Os autores Dutt et al. (2020) do artigo *"Insights into the growing popularity of artificial intelligence in ophthalmology"* comenta que o surgimento de grandes redes neurais, chamadas de aprendizado profundo, é um divisor de águas em muitas aplicações. Na oftalmologia, o aprendizado profundo permitiu que os algoritmos aprendizado de máquina alcançassem precisões aceitáveis para implantação em campo em larga escala.

Em seu trabalho Tacchella et al. (2017), denominado *"Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: a proof-of-principle studym"* a inteligência artificial foi capaz de, diante da combinação de previsões feitas por seres humanos com as de um algoritmo de aprendizado de máquina, demonstrar uma gama maior de informações sobre a progressão da esclerose múltipla. Então é possível observar que a inteligência artificial com a colaboração humana tem sido benéfica também para o auxílio de informações.

Esses exemplos ilustram como a Inteligência Artificial está contribuindo para avanços significativos na área médica, melhorando a precisão dos diagnósticos e a eficácia dos tratamentos. Foram identificados artigos relacionados a pesquisas sobre o aprendizado de máquina aplicado para prever a condição do orelha média. Diversos estudos abordam o uso dessa tecnologia inovadora na antecipação e avaliação das condições da orelha média, contribuindo assim para o avanço e aprimoramento das técnicas de diagnóstico e monitoramento. Essa tendência de pesquisa reflete a crescente importância da integração de abordagens de aprendizado de máquina em contextos da área da saúde, visando aprimorar a precisão e eficiência dos diagnósticos relacionados ao sistema auditivo.

O estudo realizado por Zeng et al. (2021), denominado *" Efficient and accurate*

identification of ear diseases using an ensemble deep learning model", adotou um modelo de transferência de aprendizagem baseado em DensNet-BC169 e DensNet-BC1615, registrando melhorias notáveis, com uma precisão média de 95,59%. As categorias classificadas foram: colesteatoma de orelha média, Otite média supurativa crônica, Sangramento do canal auditivo externo, Cerume impactado, Tímpano normal, Otomicose externa, Otite média secretora e Calcificação da membrana timpânica.

A escolha desses modelos considerou tanto a precisão quanto o tempo de treinamento. A alta precisão foi atribuída à utilização de extensos conjuntos de dados, destacando a diversidade de doenças e a precisão diagnóstica alcançada. O classificador em tempo real foi treinado em dados variados, tornando-o adequado para situações práticas. O conjunto de dados incluiu 41.056 pacientes, com 20.542 imagens (53,55% do total) analisadas, distribuídas por gênero e faixa etária. A divisão em conjuntos de treinamento (80%) e validação (20%), sem repetição, foi consistente em cada modelo treinado, proporcionando uma abordagem robusta. O estudo concluiu que o modelo de aprendizagem profunda é altamente útil na detecção e tratamento precoces de doenças da orelha em ambientes clínicos.

No estudo, realizado por Alhudhaif, Cömert e Polat (2021) denominado *" Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm"*, foi desenvolvido um novo modelo de classificação para otite média, baseado em uma rede neural convolucional (CNN). Para aprimorar a capacidade generalizada do modelo proposto, foram incorporados elementos como a combinação do canal e modelo espacial (CBAM), blocos residuais e a técnica de hipercoluna. Todos os experimentos foram conduzidos utilizando um conjunto de dados de membrana timpânica de acesso aberto, composto por 956 imagens de otoscópios distribuídas em cinco classes. Os resultados demonstraram que o modelo proposto alcançou uma classificação satisfatória, garantindo uma precisão geral de 98,26%, sensibilidade de 97,68% e especificidade de 99,30%.

No trabalho de Zafer (2020) denominado *"Fusing fine-tuned deep features for recognizing different tympanic membranes"*, a atenção foi direcionada para o reconhecimento de condições normais, otite média aguda, otite média supurativa crônica e membrana timpânica com cera de orelha, utilizando recursos profundos fundidos e ajustados provenientes de redes neurais convolucionais profundas (DCNNs) pré-treinadas. Esses recursos foram empregados como entrada em diversas redes, incluindo uma rede neural artificial, o vizinho mais próximo (NN), a árvore de decisão (DT) e a máquina de vetores de suporte (SVM). Adicionalmente, foi introduzido um novo conjunto de dados de membrana timpânica disponível publicamente, composto por um total de 956 imagens de otoscópios.

Os resultados destacaram o desempenho promissor das DCNNs, especialmente evidenciado pelo VGG-16, que alcançou uma precisão de 93,05%. Os recursos profundos

e ajustados contribuíram para a melhoria geral do sucesso na classificação. Em última análise, o modelo proposto demonstrou resultados promissores, atingindo uma precisão de 99,47%, sensibilidade de 99,35% e especificidade de 99,77% ao combinar os recursos profundos fundidos e ajustados com o modelo SVM. Como resultado, este estudo evidencia a utilidade significativa dos recursos profundos fundidos e ajustados no reconhecimento de diferentes membranas timpânicas, destacando sua capacidade de fornecer um modelo totalmente automatizado com alta sensibilidade.

2.2 Visão computacional

Na área da visão computacional, o objetivo é descrever o ambiente percebido por meio de uma ou várias imagens e reconstruir características como forma, iluminação e distribuição de cores. É notável como essa tarefa é realizada de maneira intuitiva por seres humanos e animais, enquanto os computadores necessitam empregar algoritmos e processamento cuidadoso para realizar análises semelhantes. Busca-se capacitar as máquinas a interpretar o mundo visual com a mesma eficácia que os seres vivos, abrindo um vasto campo de aplicações em automação, área da saúde, segurança e muitas outras áreas(SZELISKI, 2010).

As etapas para o processamento de imagem são as seguintes:

1. **Aquisição**: Nesta etapa o objetivo é a aquisição de imagens;
2. **Processamento de imagens**: Esta fase tem como intuito adequar e otimizar os dados visuais adquiridos. Para isso, podem ser aplicadas técnicas como retirada de ruídos, rotação da imagem;
3. **Análise de imagens**: As imagens são tornadas únicas do ponto de vista do computador. Cada imagem é atribuída à uma função única de duas incógnitas independentes, que podem ser visualizadas de forma mais objetiva pelas máquinas;

2.3 Aprendizagem de máquina

Aprendizado de Máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado, bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um *software* que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores. Os diversos sistemas de aprendizado de máquina possuem características particulares e comuns que possibilitam sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado (MONARD; BARANAUSKAS, 2003).

Algoritmos de aprendizagem de máquina se sobressaem em ambientes que variam com o tempo e que requerem readaptação constante. Atualmente o aprendizado de máquina esta presente em diversas áreas e aplicações, dentre elas o reconhecimento de voz, a visão computacional, predições financeiras e auxílio nos diagnósticos da área da saúde (BISHOP; NASRABADI, 2006). O aprendizado de máquina pode ser classificado em supervisionado ou não supervisionado como ilustrado na Figura 4.

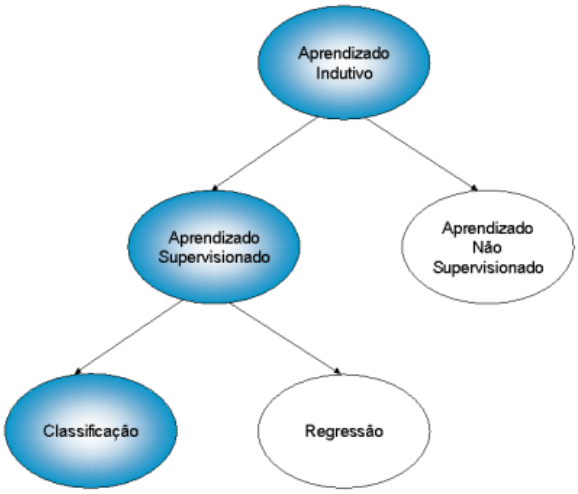

Figura 4 – hierarquia do aprendizado.

Fonte: (MONARD; BARANAUSKAS, 2003) [1]

2.3.1 Aprendizado de máquina não supervisionado

O aprendizado não supervisionado é um paradigma em que não ocorre intervenção humana direta; em vez disso, a máquina desenvolve suas próprias regras de funcionamento com base no reconhecimento de padrões, processo também conduzido por meio de software. Portanto, a análise de dados é realizada de forma automática. Esse método confere maior autonomia à máquina, embora não dependa de *feedback* ou exemplos específicos para operar.

2.3.2 Aprendizado de máquina supervisionado

O aprendizado supervisionado envolve intervenção humana direta. Essencialmente, um ser humano fornece um banco de dados e instrui a máquina a encontrar a resposta desejada de acordo com a necessidade.

Assim, a máquina adquire a capacidade de tomar decisões com base em informações previamente estabelecidas por um operador humano. Portanto, é possível afirmar

[1] https://dcm.ffclrp.usp.br/ augusto/publications/2003-sistemas-inteligentes-cap4.pdf

que o aprendizado supervisionado é competente em rotular os dados de acordo com os padrões definidos em seu sistema.

2.3.3 Redes Neurais Convolucionais (CNN)

As Redes Neurais Convolucionais (Convolutional Neural Networks - CNNs) são uma classe especializada de redes neurais profundas que se destacam em tarefas de processamento de imagem e visão computacional. Elas têm sido a base para inovações significativas em diversas aplicações, desde reconhecimento de objetos e detecção facial até carros autônomos e diagnóstico da área da saúde por imagem.

As vantagens da CNN residem em sua estrutura arquitetônica, que incorpora o emprego de campos receptivos locais, compartilhamento de pesos e, em determinadas situações, subamostragem espacial ou temporal, (LECUN; BOSER et al., 1989). A estrutura fundamental das CNNs é composta por camadas de convolução, camadas de *pooling* e camadas totalmente conectadas. Portanto, conforme o próprio nome indica, as redes neurais convolucionais são essencialmente redes neurais que empregam a operação linear de convolução em pelo menos uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Nesse sentido, pode-se definir a convolução como a integral do produto de duas funções, em que uma delas é deslizada sobre a outra (GÉRON, 2022).

2.3.3.1 Estrutura e Funcionamento das CNNs

A equação é das operações de convolução é dada por:

$$s(t) = (x * w)(t) \tag{2.1}$$

Na terminologia de visão computacional, o caracter x representa a entrada, que é comumente um campo receptivo local, como uma imagem ou um conjunto de características. O conjunto de pesos, representado por w, é conhecido como *kernel*, e a saída, s, é denominada mapa de características, composta pelos resultados das convoluções realizadas entre o kernel e os campos receptivos. Dessa forma, a operação de convolução atua como um filtro, reduzindo ruídos e permitindo que a CNN aprenda a enfocar diferentes regiões da imagem. Isso contribui para minimizar os efeitos de pequenas variações e imperfeições na amostra, como variações na iluminação, posição, escala, orientação, entre outros (LECUN; BENGIO et al., 1995).

As CNNs foram inspiradas na organização da área visual do cérebro humano e são projetadas para lidar com dados de grade, como imagens. Elas se diferenciam das redes neurais tradicionais devido a três características principais: camada convolucional, camada *pooling* (redução) e camada totalmente conectada *(Fully Connect)* como mostra a Figura 5.

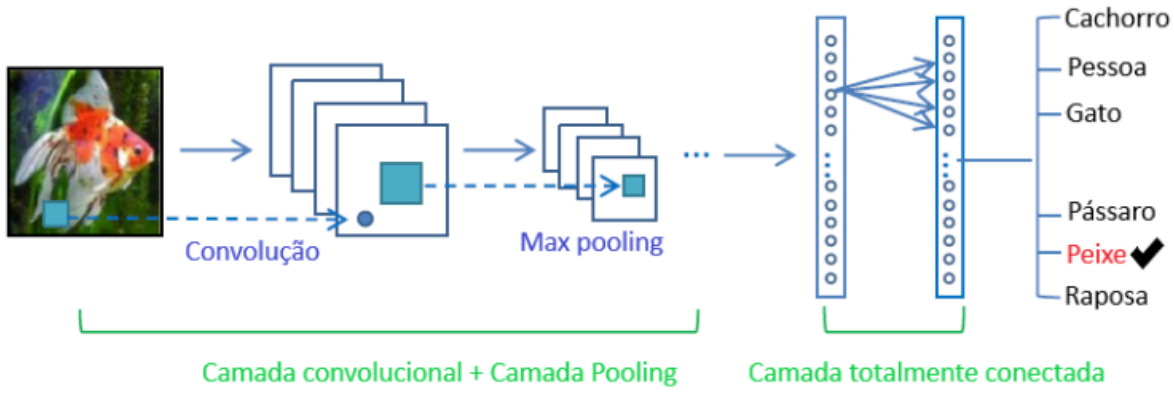

Figura 5 – Estrutura geral de uma CNN. Adaptado de (GUO et al., 2016)

As camadas convolucionais são a espinha dorsal das CNNs. Elas consistem em um conjunto de filtros (*kernels*) que percorrem a imagem em pequenos passos (convolução) e extraem características locais, como bordas, texturas e padrões. Essas características são aprendidas durante o treinamento e formam representações hierárquicas das imagens.

As camadas de *pooling* reduzem a dimensionalidade das características extraídas pelas camadas convolucionais. O *pooling* envolve a combinação de informações em regiões locais, reduzindo o tamanho dos mapas de características e tornando a rede mais eficiente computacionalmente.

Fully Connected é a camada que conecta cada neurônio ou unidade de uma camada à todas as unidades da camada subsequente. Em outras palavras, todos os neurônios de uma camada totalmente conectada estão conectados a todos os neurônios da camada seguinte. Essa camada é comumente utilizada nas partes finais de redes neurais, onde as características extraídas anteriormente são combinadas para realizar tarefas específicas, como classificação ou regressão. A camada totalmente conectada é frequentemente responsável por mapear as características aprendidas para as saídas desejadas.

3 Desenvolvimento

Neste capítulo, é descrito o processo de treinamento das redes e as métricas utilizadas para avaliar o desempenho dos modelos. Foram analisadas seis arquiteturas distintas de redes neurais convolucionais, com o objetivo de realizar a classificação das imagens do canal do orelha média.

3.1 Treinamento e Aprendizado

O *framework TensorFlow* foi fundamental para o projeto. A utilização *TensorFlow* devido à sua versatilidade e às suas poderosas capacidades para treinar modelos de aprendizado profundo. Sua ampla gama de ferramentas e recursos nos permite construir e treinar modelos de forma eficiente e escalável. Desde a construção da arquitetura do modelo até a implementação de algoritmos, o *TensorFlow* oferece uma base sólida e confiável para todas as etapas do nosso processo. Além disso, sua integração com outras bibliotecas e ferramentas de aprendizado de máquina.

Após a revisão de literatura, decidiu-se utilizar apenas CNN para buscar alcançar o objetivo deste trabalho. As CNNs desempenham um papel fundamental em visão computacional, sendo aplicadas em tarefas como reconhecimento de objetos, detecção facial, processamento de imagens médicas, realidade virtual e aumentada, entre outras.

O treinamento de uma CNN é um processo no qual um extenso conjunto de dados de treinamento, contendo imagens rotuladas, é apresentado à rede neural. Durante o treinamento, a CNN ajusta seus parâmetros, representados pelos pesos dos filtros, para minimizar uma função de perda. Essa função de perda mensura a diferença entre as previsões da rede e os rótulos reais das imagens.

3.1.1 Aprendizagem por Transferência

A aprendizagem por transferência é uma técnica crucial para adaptar redes pré-treinadas a novas áreas, proporcionando treinamento consistente para conjuntos de dados menores (ALTUNTAŞ; CÖMERT; KOCAMAZ, 2019). Redes como AlexNet, MobileNetV2, VGG19, EfficientNetV2_S, ConvNetX_Small, ConvNetX_Tiny e ConvNetX_Tiny3 passam por um estágio de pré-treinamento na base de dados ImageNet. Isso permite que as redes adquiram representações de características úteis, transferíveis para tarefas específicas por meio de ajuste fino ou transferência de aprendizado. A referência à base de dados ImageNet é comum ao descrever o processo de treinamento inicial (BRESSEM et al., 2020).

3.1.2 Função de Ativação

Uma função de ativação em redes neurais é uma função matemática aplicada a cada unidade (ou neurônio) em uma camada da rede neural. Ela determina se um neurônio deve ser ativado ou não com base em sua entrada ponderada. A ativação é a saída gerada pela aplicação dessa função, e ela introduz não-linearidade na rede, permitindo que a rede neural aprenda relações complexas nos dados (GOODFELLOW; BENGIO; COURVILLE, 2016). Algumas das funções de ativação mais comuns incluem a função sigmoide, a tangente hiperbólica (tanh), a função de retificação linear (ReLU) e suas variantes, como Leaky ReLU.

A função Softmax é frequentemente usada como função de ativação na camada final de modelos, especialmente em processos de regressão e classificação. Para tarefas de classificação, o vetor de valores D-dimensionais é enviado a partir da camada anterior. A função Softmax normaliza esses valores para o intervalo [0,1], gerando um vetor D-dimensional reconstruído. Essa função é essencial para garantir que a saída da rede seja interpretável de maneira probabilística. Em tarefas de classificação, a entrada é associada à classe com a probabilidade mais alta (MAHARJAN et al., 2020).

A função Softmax é definida matematicamente por:

$$\text{Softmax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{3.1}$$

onde z é o vetor de entrada com K elementos, e $\text{Softmax}(z)_i$ é o i-ésimo elemento da saída da função Softmax. A função exponencial e^{z_i} destaca o valor mais alto em z, enquanto a normalização pelo somatório das exponenciais garante que a saída seja uma distribuição de probabilidades, somando-se a 1.

3.1.3 Otimizador

Em redes neurais, um otimizador é um algoritmo utilizado para ajustar os pesos e os viéses durante o treinamento, com o objetivo de minimizar a função de perda. A escolha do otimizador pode impactar significativamente a eficácia e a eficiência do treinamento da rede neural (GOODFELLOW; BENGIO; COURVILLE, 2016). Os mais comuns são: gradiente descendente, gradiente descendente estocástico ,*root mean square propagation*, *adaptive moment estimation(Adam)*.

O otimizador Adam, que significa Estimação Adaptativa de Momento, é amplamente utilizado no treinamento de redes neurais (KINGMA; BA, 2014). Ele combina as características do RMSprop e do Momentum, proporcionando taxas de aprendizado adaptativas e eficiência na convergência. A taxa de aprendizado é ajustada individualmente para cada parâmetro, utilizando uma média móvel de gradientes e suas magnitudes (KINGMA;

BA, 2014). A incorporação de momentos suaviza oscilações nos gradientes, acelerando a convergência ao aumentar o movimento em direção aos mínimos locais (RUDER, 2016).

O Adam também incorpora o RMSprop, normalizando os gradientes pela raiz quadrada da média móvel exponencial dos quadrados dos gradientes. Essa normalização é eficaz para mitigar problemas de convergência em espaços de parâmetros com escalas discrepantes (KINGMA; BA, 2014). Apesar de sua eficiência, é importante ajustar hiperparâmetros, como a taxa de decaimento e os parâmetros de momento, para otimizar o desempenho (KINGMA; BA, 2014).

O otimizador utilizado foi o Adam, com os parâmetros apresentados na Tabela 1. Os valores definidos consideram a disponibilidade de RAM-GPU e foram ajustados para otimizar o desempenho dentro das limitações de hardware.

Tabela 1 – Parâmetros do Otimizador Adam

Parâmetro	**Valor**
amsgrad	False
betas	(0.9, 0.999)
capturable	False
differentiable	False
eps	1e-08
foreach	None
fused	None
lr	0.001
maximize	False
weight_decay	0

3.1.4 Critérios de avaliação de algoritmo

Após concluir o treinamento da rede, registra-se a sua duração e, em seguida, procede-se ao teste do modelo. Durante esse teste, é viável avaliar o desempenho da rede por meio de critérios como erro medio, precisão, *recall* e *f1-score*, os quais são calculados a partir da matriz de confusão exibida ao final do teste. Essa matriz é uma tabela que apresenta a frequência com que cada classe foi corretamente ou incorretamente classificada pelo modelo. Com base nos valores dessa matriz, é possível calcular diversas métricas de desempenho. A matriz 2x2 em questão possui quatro resultados potenciais.

- Verdadeiro Positivo (VP): O modelo previu corretamente a classe positiva.
- Falso Positivo (FP): O modelo previu erroneamente a classe positiva quando a classe real era negativa.
- Falso Negativo (FN): O modelo previu erroneamente a classe negativa quando a classe real era positiva.

- Verdadeiro Negativo (VN): O modelo previu corretamente a classe negativa.

A partir dos valores da matriz de confusão, é possível calcular as demais métricas avaliadas durante o teste do modelo. A precisão mensura a quantidade de acertos do modelo em relação ao total de tentativas de acerto, sendo calculada por: $\frac{VP}{VP+FP}$. Sua escala varia de 0 a 1, onde valores mais próximos de 1 indicam uma precisão mais elevada.

Já o recall, por sua vez, avalia a quantidade de acertos do modelo em relação ao total de vezes que ele deveria ter acertado, sendo calculado por: $\frac{VP}{VP+FN}$. Seu intervalo também varia de 0 a 1, sendo equiparado à escala da precisão.

Por fim, o f1-score é uma métrica que combina precisão e recall de forma equilibrada, sendo obtido pela média harmônica dessas duas métricas. Seu valor pode variar de 0 a 1, sendo que resultados mais próximos de 1 indicam um desempenho mais favorável. Essas métricas são cruciais para ajustar os parâmetros da rede neural e aprimorar sua eficácia em problemas de classificação. Assim, é possível avaliar os modelos considerando a confiabilidade dos resultados e a eficiência do treinamento.

4 Experimentos e Resultados

Este capítulo aborda a base de dados utilizada, discutindo os protocolos experimentais e apresentando os resultados obtidos.

4.1 Base de dados

Os parâmetros de treinamento foram estabelecidos com um tamanho de lote (*batch size*) de 8 e uma resolução de 100x100 pixels para todos os modelos. Essas escolhas foram feitas levando em consideração a disponibilidade de RAM-GPU, buscando atingir o melhor desempenho possível dentro das limitações de *hardware*. Neste âmbito, um conjunto de dados de membranas timpânicas de acesso aberto foi disponibilizado pelo grupo *CTG*. O conjunto de dados consiste em um total de 956 imagens brutas de otoscópio coletadas de pacientes voluntários internados no Hospital Özel Van Akdamar entre 10/2018 e 06/2019 (ZAFER, 2020).

O número de amostras normais de membranas timpânicas é 535, enquanto o número de amostras de *acute otitis media*, *chronic suppurative otitis media* e *Earwax* é 119, 63 e 140, respectivamente. A propósito, as amostras pertencentes a *otitis externa* (41), *ear ventilation tube* (16), *foreign bodies in the ear* (3), *pseudo-membranes* (11) e *tympanosclerosis* (28).

Cada amostra de otoscópio do banco de dados foi avaliada por três otorrinolaringologistas. As amostras pertencentes a diferentes classes foram armazenadas nas pastas especificadas nomeadas considerando os tipos de *OM*. As imagens de baixa qualidade por falta de luz, etc. também foram isoladas do banco de dados. Um resumo do conjunto de dados *Membrana timpanica* é dada na Tabela 2.

Tabela 2 – Quantidade de amostras por categoria.

Categoria	Quantidade
Normal Tympanic membrane	535
Acute Otitis Media (AOM)	119
Chronic suppurative Otitis Media	63
Earwax	140
Otitis externa	41
Ear ventilation tube	16
Foreign bodies in the ear	3
Pseudo membranes	11
Tympanosclerosis	28
Total	956

A Figura6 mostra um exmeplo de cada alteração na orelha externa e média presente na base de dados:

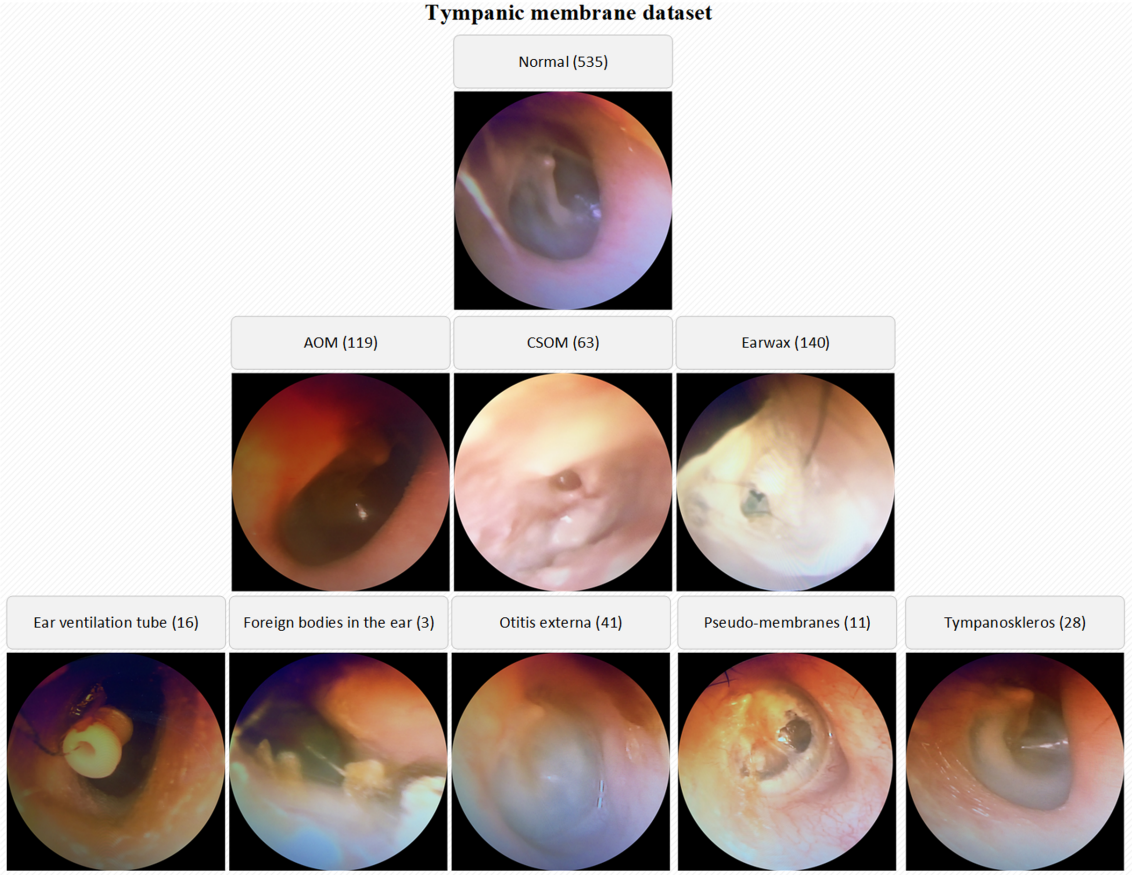

Figura 6 – Amostra da base de dados

Fonte: Adi Alhudhaif Zafer Cömert (2021)[1]

Os dados utilizados foram divididos para análise em três conjuntos distintos: primeiro, uma base denominada *"AuAnomaliaOuNao"*, classificando as imagens como normais (Tímpano Normal) ou anormais (os demais itens da tabela); segundo, uma base denominada *"AuMulticlasse"*, que classifica cada classe presente na tabela de forma independente; terceiro, uma base chamada *"AuSoOtite"*, cujo objetivo é classificar se o canal do orelha média apresenta Otitis Media Aguda (OMA), Otitis Media Supurativa Crônica, Otitis Externa ou não (Tímpano Normal).

Devido ao reduzido número de imagens disponíveis para algumas doenças, decidiu-se excluir as condições de *Foreign bodies in the ear*, *Pseudo membranes* e *Ear ventilation tube* do processo de avaliação. Essa decisão visa garantir uma análise mais robusta e confiável, concentrando-se nas condições para as quais há uma quantidade maior de dados disponíveis.

[1] https://peerj.com/articles/cs-405/

A distribuição de imagens na base intitulada *"AuSoOtite"*, foi feita como mostra a Tabela 3:

Tabela 3 – Base de dados *"AuSoOtite"*.

Categoria	Quantidade
Normal Tympanic membrane	535
Acute Otitis Media (AOM)	119
Chronic suppurative Otitis Media	63
Otitis externa	41
Total	758

A distribuição de imagens na base intitulada, *"AuMulticlasse"*, foi feita como mostra a Tabela 4:

Tabela 4 – Base de dados *"AuMulticlasse e AuAnomaliaOuNao"*.

Categoria	Quantidade
Normal Tympanic membrane	535
Acute Otitis Media (AOM)	119
Chronic suppurative Otitis Media	63
Earwax	140
Otitis externa	41
Tympanosclerosis	28
Total	926

A distribuição de imagens na base intitulada,*"AuAnomaliaOuNao"*, foi feita como mostra a Tabela 4. A diferença é como são separadas as bases. Na *"AuMulticlasse"* é classificado classe por classe enquanto na *"AuAnomaliaOuNao"* todas as alterações na orelha média são agrupadas em uma base e compara-se tem alterações na orelha média ou não na mambrana timpânica.

A estratificação ao dividir conjuntos de dados visa assegurar que cada partição mantenha uma distribuição aproximada das classes da base original. Esse método é especialmente importante em problemas onde há múltiplas instâncias relacionadas entre si, como imagens de uma mesma pessoa em um problema de classificação de doença. A estratificação busca evitar a dispersão de dados relacionados entre diferentes partições, garantindo que, por exemplo, imagens de uma mesma pessoa permaneçam juntas. Essa abordagem considera tanto as classes quanto os grupos na distribuição dos dados. (CHOO et al., 2020).

Em todas as três situações, a divisão da base foi realizada com 80% para treinamento, 10% para validação e 10% para teste.

Os momentos invariantes têm sido extensivamente utilizados no reconhecimento de padrões em imagens em diversas aplicações, graças às suas propriedades de invariância em relação a translações, escalonamentos e rotações da imagem (HUANG; LENG, 2010).Para

expansão da base de dados, nas pastas de treino e validação, foram aplicadas três rotações em todas as imagens após a separação, com rotações de 90º, 180º e 270º em relação à imagem original.

Antes da realização da rotação das imagens, observou-se que a acurácia dos modelos estava consistentemente baixa, variando em torno de 65% a 73%. Essa performance aquém do esperado sugeria uma limitação na capacidade dos modelos em reconhecer e generalizar padrões presentes nas imagens, especialmente em diferentes orientações.

A rotação das imagens foi realizada, considerando que é uma das transformações mais comuns que podem ocorrer na realidade, proporcionando uma ampla gama de orientações. Isso enriqueceu significativamente a variedade de perspectivas presentes no conjunto de dados. A distribuição resultante reflete a diversidade introduzida pelo processo de rotação, destacando a abrangência do conjunto de treinamento em termos de diferentes ângulos de visualização. Essa análise da distribuição pós-rotação é crucial para compreender como a expansão da base de dados influencia a preparação do modelo para reconhecer padrões em diversas orientações, contribuindo assim para um treinamento mais robusto e eficiente.

Base de dados após as realização das rotações:

A nova distribuição de imagens na base intitulada *"AuAnomaliaOuNao"* está resumida na Tabela 5:

Tabela 5 – Nova base de dados após rotação e retirada das classes com poucas imagens *"AuAnomaliaOuNao"*.

Categoria	Quantidade treino	Quantidade validação	Quantidade teste
Anormal	1211	164	44
Normal	1659	212	55

A nova distribuição de imagens na base intitulada *"AuMulticlasse"*, ficou como mostra a Tabela 6:

Tabela 6 – Nova base de dados após rotação e retirada das classes com poucas imagens *"AuMulticlasse"*.

Categoria	Quantidade treino	Quantidade validação	Quantidade teste
Normal Tympanic membrane	1659	212	53
Acute Otitis Media (AOM)	384	44	12
Chronic suppurative Otitis Media	204	24	6
Earwax	448	52	14
Otitis externa	132	16	4
Tympanosclerosis	88	12	3

A nova distribuição de imagens na base intitulada *"SoOtite"*, ficou como mostra a Tabela 7:

Tabela 7 – Nova base de dados após rotação e retirada das classes com poucas imagens *"SoOtite"*.

Categoria	Quantidade treino	Quantidade validação	Quantidade teste
Anormal	540	84	22
Normal	1713	212	53

4.2 Utilizando redes pré-treinadas sem modificações

Diante da vasta gama de arquiteturas de redes neurais pré-treinadas disponíveis, foi essencial estabelecer critérios de pré-seleção para escolher aquelas mais adequadas ao escopo do estudo. Para esse fim, foram consideradas duas métricas fundamentais: acurácia e erro médio. A acurácia, que representa a capacidade da rede neural de fazer previsões corretas, e o erro médio, que mensura a magnitude média dos desvios entre as previsões do modelo e os rótulos reais, foram escolhidos como indicadores essenciais para avaliar o desempenho das arquiteturas.

Esses critérios foram aplicados a cada rede neural pré-treinada, possibilitando uma avaliação comparativa de seu desempenho antes de sua implementação no estudo. A escolha dessas métricas se justifica pela necessidade de assegurar que as redes neurais selecionadas fossem capazes de proporcionar resultados confiáveis e precisos, aspectos fundamentais para a validade dos experimentos conduzidos. Essa abordagem de pré-seleção contribuiu para orientar a escolha das arquiteturas mais apropriadas, considerando as demandas específicas do estudo em questão.

Os métodos empregados neste estudo aproveitam o banco de dados de treinamento da ImageNet, uma vasta coleção de imagens em milhares de categorias. Esses dados são essenciais para treinar modelos de aprendizado profundo, como as Redes Neurais Convolucionais (CNNs), usadas neste trabalho. Ao se basear na ImageNet, os modelos podem aprender a reconhecer padrões complexos e extrair características relevantes das imagens otoscópicas, melhorando assim sua capacidade de diagnosticar alterações na orelha média e externa.

4.2.1 Estrutura das Redes Utilizadas

4.2.1.1 AlexNet

O texto a seguir é um resumo de informações do artigo de referência Krizhevsky, Sutskever e Hinton (2012). A AlexNet é uma arquitetura de rede neural convolucional que ganhou destaque ao vencer a competição ImageNet em 2012. Desenvolvida por Alex Krizhevsky, a rede é composta por 8 camadas convolucionais no bloco *"features"*, seguidas por uma

camada de *pooling* médio adaptativo, e 5 camadas totalmente conectadas no bloco *"classifier"*. A arquitetura introduziu inovações, como o uso de funções de ativação não lineares, como ReLU, e a aplicação de técnicas como *dropout* para combater o *overfitting*. A *AlexNet* contribuiu significativamente para o avanço do campo de visão computacional e inspirou o desenvolvimento de arquiteturas subsequentes.

4.2.1.2 MobileNetV2

O texto a seguir é um resumo de informações do artigo de referência Sandler et al. (2018).A *MobileNetV2* é uma arquitetura de rede neural convolucional projetada para eficiência computacional, especialmente em dispositivos móveis. Composta por 19 blocos no bloco "features"(blocos 0 a 18), a rede utiliza um design inovador chamado *blocos invertidos de convolução de gargalo móvel* (MBConv). Cada bloco MBConv consiste em camadas convolucionais, normalização por lote (*BatchNorm*), ativação linear por unidade sigmoidal (*SiLU*), e módulos de *Squeeze-and-Excitation* (SE). A *MobileNetV2* termina com uma camada totalmente conectada no bloco "classifier"(camada 1). Essa arquitetura foi projetada para atingir um equilíbrio entre desempenho e eficiência, sendo particularmente adequada para dispositivos móveis e aplicações com recursos computacionais limitados,

4.2.1.3 VGG19

O texto a seguir é um resumo de informações do artigo de referência Simonyan e Zisserman (2014). A *VGG19* é uma arquitetura de rede neural convolucional que faz parte da família *Visual Geometry Group* (VGG). Composta por 13 camadas convolucionais no bloco *"features"* (camadas 0 a 36), a rede utiliza convoluções com filtros pequenos de 3x3, seguidas por camadas de *BatchNorm* e funções de ativação ReLU. Após o bloco *"features"*, a *VGG19* inclui uma camada de *pooling* adaptativo médio (*avgpool*) e finaliza com 3 camadas totalmente conectadas no bloco *"classifier"* (camadas 0 a 6). Essa arquitetura é conhecida por sua simplicidade e eficácia, contribuindo para o entendimento da importância de camadas profundas em redes neurais convolucionais.

4.2.1.4 EfficientNetV2

O texto a seguir é um resumo de informações do artigo de referência Tan e Le (2021). A distribuição da rede é dada por *Features (Camadas 0 a 6):* A rede possui camadas convolucionais organizadas em blocos MBConv (*Mobile Inverted Bottleneck Convolution*). Cada bloco consiste em uma sequência de operações, incluindo convoluções, normalização por lote (*BatchNorm*), ativação SiLU (uma variação da função de ativação sigmoidal) e operações de Squeeze-and-Excitation (SE). A probabilidade de eliminação estocástica (*stochastic depth*) também é aplicada.

Classifier (Camadas 7 a 10): A rede termina com uma camada de convolução convencional seguida por uma camada de *BatchNorm* e ativação SiLU. Em seguida, há uma camada de *pooling* adaptativo médio (*avgpool*) e um bloco *classifier* que consiste em uma camada de *dropout* e uma camada totalmente conectada (*linear*) que produz a saída final.

Parâmetros Específicos:

Stochastic Depth (Probabilidade de eliminação estocástica): A probabilidade de eliminação estocástica varia de 0.1 a 0.2 nos blocos MBConv.

Outras Observações:

A utilização de ativações SiLU e operações Squeeze-and-Excitation pode contribuir para melhorar a representação e a eficiência do modelo.

4.2.1.5 ConvNeXt Small

O texto a seguir é um resumo de informações do artigo de referência Liu et al. (2022). A arquitetura é organizada em três principais seções: *features*, *avgpool* e *classifier*.

Bloco *features*:

Começa com uma camada convolucional inicial *Conv2d* com 96 filtros de tamanho 4x4 e passo 4x4, seguida por normalização (*LayerNorm2d*). Em seguida, há uma sequência de 24 blocos *CNBlock*, cada um contendo:

- Uma camada convolucional com 96 grupos (camadas *Conv2d* e *LayerNorm*).
- Processamento de permutação (*Permute*).
- Duas camadas lineares, uma de entrada 96 e saída 384, e outra de entrada 384 e saída 96.
- Processamento de permutação adicional.

O bloco possui uma camada de ativação *GELU* e utiliza a técnica de *dropout estocástico* (*StochasticDepth*). Entre os blocos *CNBlock*, há camadas adicionais de normalização (*LayerNorm2d*) seguidas por camadas convolucionais (*Conv2d*).

Camada *avgpool*:

É uma camada de *pooling* adaptativo médio (*AdaptiveAvgPool2d*) com tamanho de saída 1x1.

Bloco *classifier*:

Começa com uma camada de normalização (*LayerNorm2d*). Segue para uma camada de achatamento (*Flatten*). Termina com uma camada totalmente conectada (*Linear*) de entrada 768 e saída 1000, que geralmente representa as classes de saída.

4.3 Utilizando redes pré-treinadas sem modificações

A arquitetura utilizada aqui é uma versão simplificada de um modelo de rede neural convolucional. Cada bloco *CNBlock* possui uma estrutura padrão de convolução, normalização, permutação, linear, ativação e *dropout estocástico*.

Neste trabalho, foram conduzidas 100 épocas em cada modelo de rede neural para avaliar o desempenho em diferentes métricas. Os detalhes dos resultados para cada modelo são apresentados na Tabela 8, destacando a acurácia nos conjuntos AuAnomaliaOuNao, AuMulticlasse e AuSoOtite.

Ao analisar os resultados, observamos que o modelo *convnext_tiny* apresentou consistentemente uma porcentagem média de acurácia superior em comparação com os demais modelos. Especificamente, obteve taxas de acerto de 85,64% em *AuAnomaliaOuNao*, 83,64% em *AuMulticlasse* e 87,84% em *AuSoOtite*. Esses valores destacam a robustez e eficácia do modelo em diversas situações.

O modelo *"convnext_tiny"* também demonstrou um desempenho notável, com percentuais de acertos de 84,30%, 79,24% e 88,85% nos conjuntos *"AuAnomaliaOuNao"*, *"AuMulticlasse"* e *"AuSoOtite"*, respectivamente. Esses resultados evidenciam a capacidade do modelo em lidar com múltiplas classes e anomalias.

Por fim, o modelo *MobileNetV2* mostrou consistência, com acurácia de 82,18%, 77,73% e 87,50% nos conjuntos *AuAnomaliaOuNao*, *AuMulticlasse* e *AuSoOtite*, respectivamente. Sua eficácia e eficiência o posicionam como uma opção robusta para diversas tarefas de classificação.

Esses resultados proporcionam uma visão abrangente do desempenho dos modelos, permitindo uma seleção informada do modelo mais adequado para diferentes contextos de aplicação. .

Tabela 8 – Resultados acurácia.

Modelo	AuAnomaliaOuNao	AuMulticlasse	AuSoOtite
alexnet	77,3936	69,5455	84,7973
MobileNetV2	82,1809	77,7273	87,5
VGG19	75,266	60	81,7568
efficientnet_v2_s	74,4681	58,4848	83,7838
convnext_small	84,3005	79,2424	88,8514
convnext_tiny	85,6383	83,6364	87,8378

Os resultados apresentam os erros médios para cada método nos conjuntos na tebela 9 *AuAnomaliaOuNao*, *AuMulticlasse* e *AuSoOtite*. Cada métrica foi avaliada individualmente:

- **AuAnomaliaOuNao:**

  - Observa-se que o método *convnext_tiny* apresentou o menor erro médio (0,4006), indicando uma boa capacidade de generalização e acerto.
  - Os métodos *MobileNetV2* e *convnext_small* também demonstraram desempenhos notáveis, com baixos valores de erro (0,4366 e 0,4187, respectivamente).

- **AuMulticlasse:**
  - *convnext_tiny* mais uma vez liderou com o menor erro médio (0,6456), seguido por *MobileNetV2* (0,9071) e *convnext_small* (0,6484).

- **AuSoOtite:**
  - *convnext_tiny* novamente apresentou o menor erro médio (0,3005), indicando boa capacidade de classificação em casos mais desafiadores.
  - *convnext_small* e *MobileNetV2* também mostraram bom desempenho, com erros de 0,3157 e 0,3643, respectivamente.

Esses resultados sugerem que, em geral, o método *convnext_tiny* tende a ter um desempenho superior, enquanto *convnext_small* e *MobileNetV2* também mostram resultados promissores. Essas conclusões podem orientar a escolha do modelo mais adequado para diferentes cenários de aplicação.

Tabela 9 – Erros médios para cada método.

Método	AuAnomaliaOuNao	AuMulticlasse	AuSoOtite
alexnet	0,5588	1,5072	0,4015
MobileNetV2	0,4366	0,9071	0,3643
VGG19	0,5782	1,3786	0,4672
efficientnet_v2_s	0,5412	1,1787	0,4206
convnext_small	0,4187	0,6484	0,3157
convnext_tiny	0,4006	0,6456	0,3005

4.3.1 Gráfico de erros

4.3.1.1 Alexnet

Os gráficos (Figuras 7, 8 e 9) mostram os erros médios de treino e validação ao longo das épocas durante o processo de treinamento do modelo, usando a Alexnet. Eles são cruciais para avaliar o desempenho do modelo, identificar sobreajuste ou subajuste e guiar ajustes na arquitetura ou nos dados de treinamento.

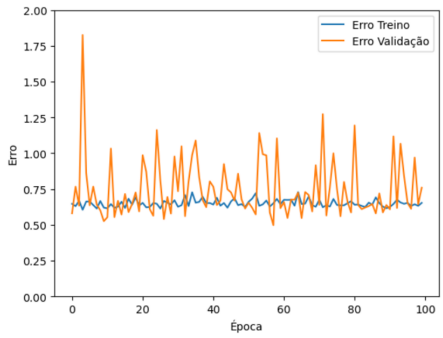

Figura 7 – Gráfico do erro médio e validação Alexnet base: AuanomaliaOunao

Fonte: Autoria própria.

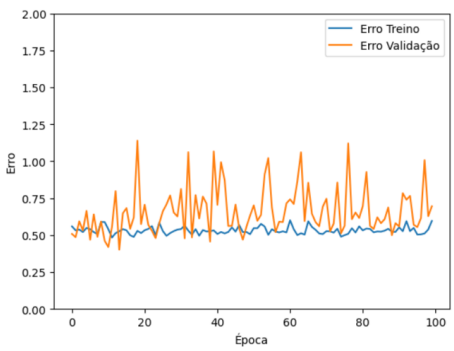

Figura 8 – Gráfico do erro médio e validação Alexnet base: ausootite

Fonte: Autoria própria.

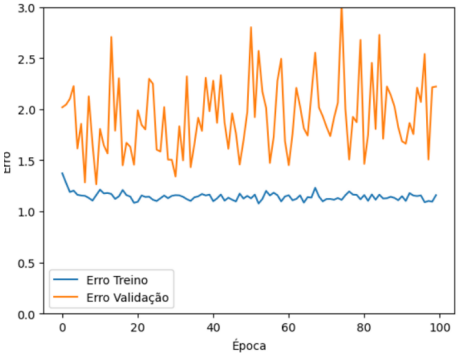

Figura 9 – Gráfico do erro médio e validação Alexnet base: Multiclasse

Fonte: Autoria própria.

4.3.1.2 MobileNetV2

Os gráficos (Figuras 10,11 e 12) mostram os erros médios de treino e validação ao longo das épocas durante o processo de treinamento do modelo, usando a *MobileNetV2*. Eles são cruciais para avaliar o desempenho do modelo, identificar sobreajuste ou subajuste e guiar ajustes na arquitetura ou nos dados de treinamento.

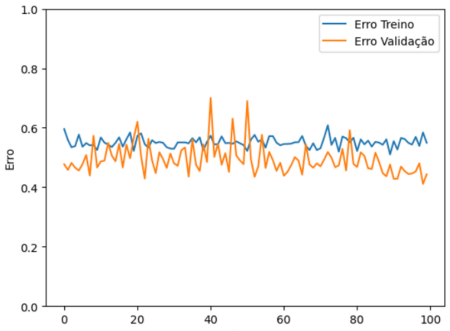

Figura 10 – Gráfico do erro médio e validação MobileNetV2 base: AuanomaliaOunao

Fonte: Autoria própria.

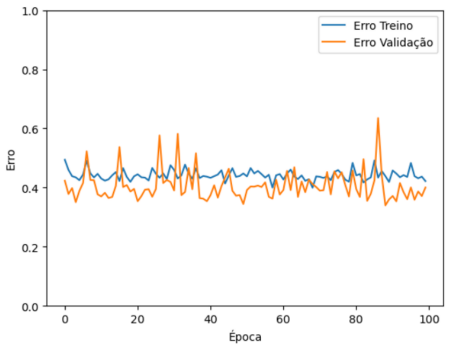

Figura 11 – Gráfico do erro médio e validação MobileNetV2 base: ausootite

Fonte: Autoria própria.

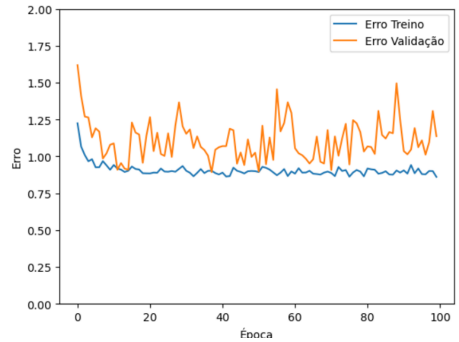

Figura 12 – Gráfico do erro médio e validação MobileNetV2 base: Multiclasse

Fonte: Autoria própria.

4.3.1.3 VGG19

Os gráficos (Figuras 13,14 e 15) mostram os erros médios de treino e validação ao longo das épocas durante o processo de treinamento do modelo, usando a *VGG19*. Eles são cruciais para avaliar o desempenho do modelo, identificar sobreajuste ou subajuste e guiar ajustes na arquitetura ou nos dados de treinamento.

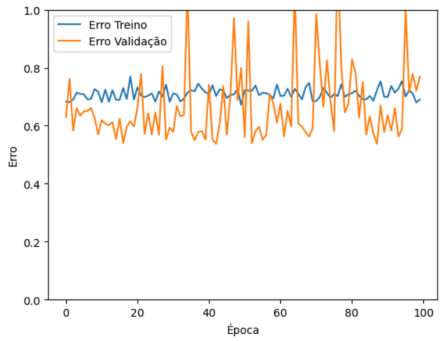

Figura 13 – Gráfico do erro médio e validação VGG19 base: AuanomaliaOunao

Fonte: Autoria própria.

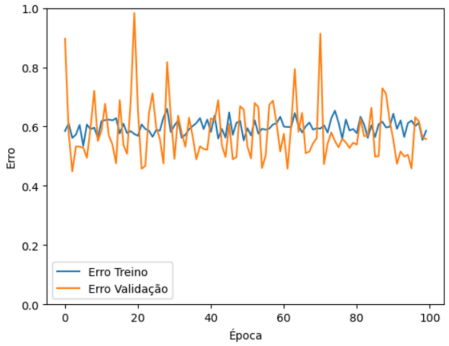

Figura 14 – Gráfico do erro médio e validação VGG19 base: ausootite

Fonte: Autoria própria.

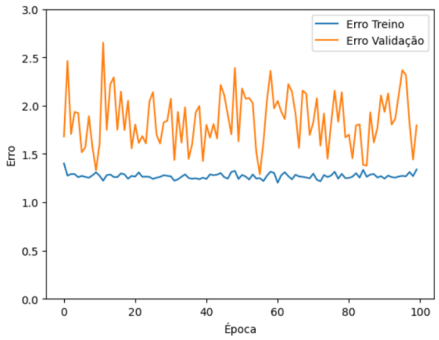

Figura 15 – Gráfico do erro médio e validação VGG19 base: Multiclasse

Fonte: Autoria própria.

4.3.1.4 Efficientnet v2 s

Os gráficos (Figuras 16,17 e 18) mostram os erros médios de treino e validação ao longo das épocas durante o processo de treinamento do modelo, usando a *efficientnet v2 s*. Eles são cruciais para avaliar o desempenho do modelo, identificar sobreajuste ou subajuste e guiar ajustes na arquitetura ou nos dados de treinamento.

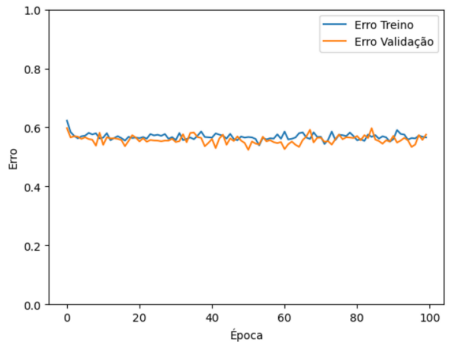

Figura 16 – Gráfico do erro médio e validação efficientnet v2 s base: AuanomaliaOunao

Fonte: Autoria própria.

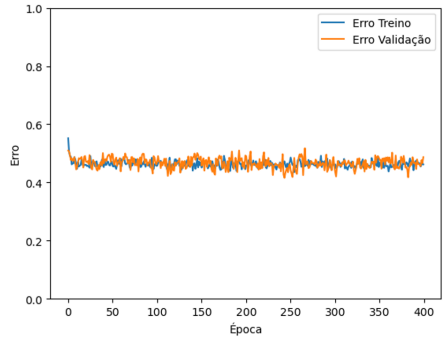

Figura 17 – Gráfico do erro médio e validação efficientnet v2 s base: ausootite

Fonte: Autoria própria.

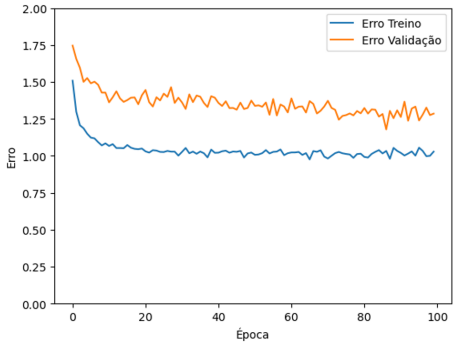

Figura 18 – Gráfico do erro médio e validação efficientnet v2 s base: Multiclasse

Fonte: Autoria própria.

4.3.1.5 Convnext small

Os gráficos (Figuras 19,20 e 21) mostram os erros médios de treino e validação ao longo das épocas durante o processo de treinamento do modelo, usando a *convnext small*. Eles são cruciais para avaliar o desempenho do modelo, identificar sobreajuste ou subajuste e guiar ajustes na arquitetura ou nos dados de treinamento.

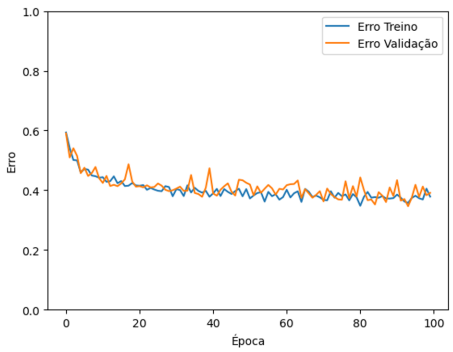

Figura 19 – Gráfico do erro médio e validação convnext small base: AuanomaliaOunao

Fonte: Autoria própria.

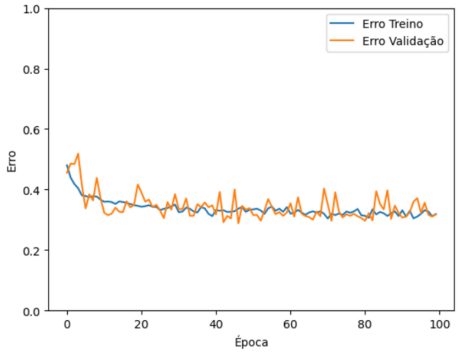

Figura 20 – Gráfico do erro médio e validação convnext small base: ausootite

Fonte: Autoria própria.

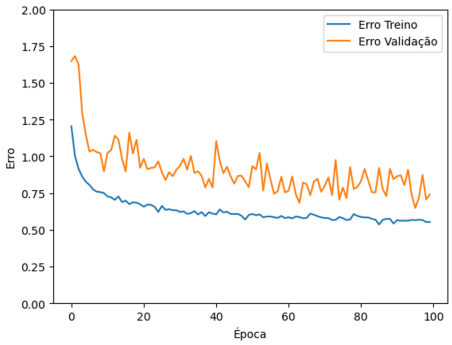

Figura 21 – Gráfico do erro médio e validação convnext small base: Multiclasse

Fonte: Autoria própria.

4.3.1.6 Convnext tiny

Os gráficos (Figuras 22,23 e 24) mostram os erros médios de treino e validação ao longo das épocas durante o processo de treinamento do modelo, usando a *convnext tiny*. Eles são cruciais para avaliar o desempenho do modelo, identificar sobreajuste ou subajuste e guiar ajustes na arquitetura ou nos dados de treinamento.

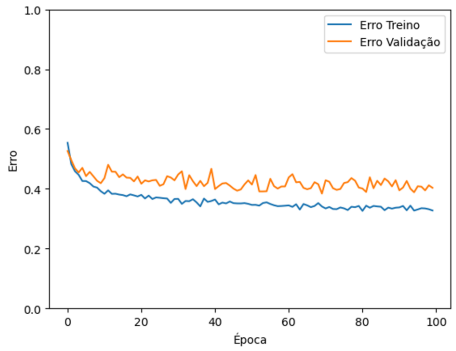

Figura 22 – Gráfico do erro médio e validação convnext tiny base: AuanomaliaOunao

Fonte: Autoria própria.

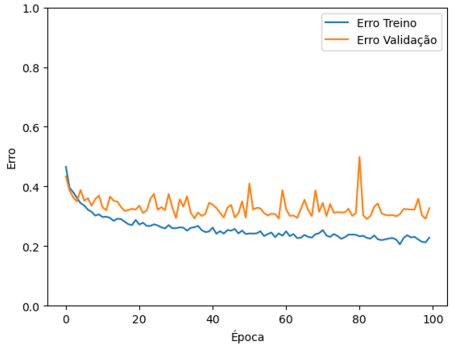

Figura 23 – Gráfico do erro médio e validação convnext tiny base: ausootite

Fonte: Autoria própria.

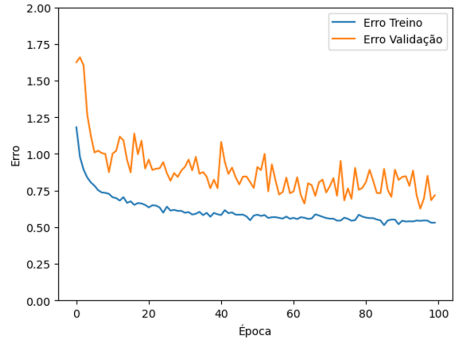

Figura 24 – Gráfico do erro médio e validação convnext tiny base: Multiclasse

Fonte: Autoria própria.

4.4 Utilizando redes pré-treinadas com modificações

Nessa etapa, foram incorporadas duas camadas lineares (*Linear*) e duas camadas de ativação ReLU (*ReLU*) nos modelos que se destacaram na etapa anterior: *convnext tiny*, *convnext small* e *MobileNetV2*. Essas adições visaram aprimorar a capacidade desses modelos em capturar relações mais complexas e não-lineares nos dados de entrada.

As camadas lineares foram introduzidas para realizar transformações lineares nos recursos extraídos, ampliando assim a capacidade do modelo de aprender representações mais sofisticadas. Juntamente com as camadas de ativação ReLU, que introduzem não-linearidades, essas modificações têm o potencial de melhorar significativamente o desempenho na identificação de padrões complexos no conjunto de dados.

Além disso, a última camada adicionada, *LogSoftmax*, sugere a aplicação de uma função de softmax logarítmica, aprimorando a capacidade de interpretação das saídas do modelo. Essas adaptações específicas foram implementadas nos modelos previamente destacados, buscando otimizar seu desempenho para cenários mais desafiadores e tarefas mais complexas.

Após as modificações os valores de acurácia foram os apresentados na Tabela 10. Analisando os resultados, percebe-se que o modelo convnextTiny obteve os melhores resultados em todas as métricas e cenários em comparação com os outros modelos (MobileNetV2 e convnextSmall). Isso sugere que, mais uma vez, a convnextTiny demonstrou um desempenho superior em relação às outras arquiteturas avaliadas.

Tabela 10 – Resultados nova acurácia.

Método	AuAnomaliaOuNao	AuMulticlasse	AuSoOtite
MobileNetV2	87,2340	82,8788	86,4865
convnext_small	86,1702	82,6573	89,2504
convnext_tiny	88,9043	89,1156	90,8784

Os erros foram os apresentados na Tabela 11:

Tabela 11 – Erros médios após alteração .

Método	AuAnomaliaOuNao	AuMulticlasse	AuSoOtite
MobileNetV2	0,3827	0,7103	0,3609
convnext_small	0,3989	0,6987	0,3598
convnext_tiny	0,2384	0,5905	0.3103

Após a conclusão do treinamento e validação, o modelo foi desafiado com imagens nunca antes encontradas.Alguns dos resultados das classificações realizadas são apresentados abaixo. Nas Figuras 25, 26 e 27, destacam-se as classificações de alterações na orelha média na membrana timpânica. Por outro lado, nas Figuras 29, 28 e 30, são exibidas

imagens da membrana timpânica em estado normal. As classificações foram corretas e foram realizadas na *convnext tiny* na base de dados *AuSoOtite*.

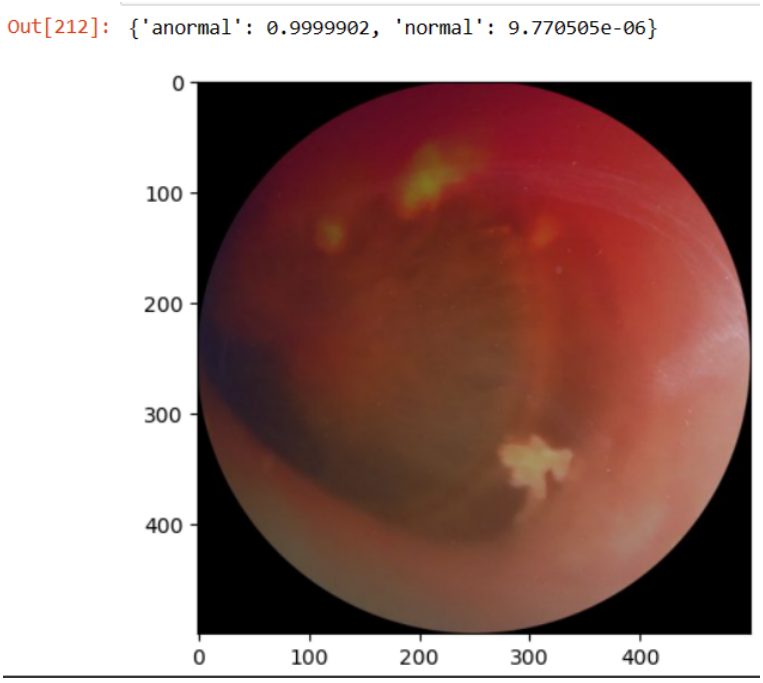

Figura 25 – Classificação 1

Fonte: Autoria própria.

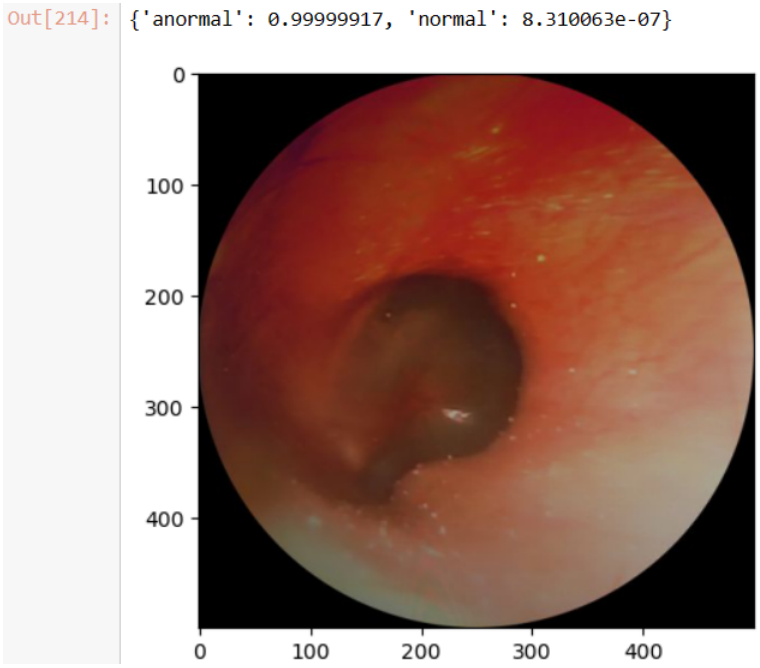

Figura 26 – Classificação 2

Fonte: Autoria própria.

```
Out[215]: {'anormal': 0.9947076, 'normal': 0.0052923583}
```

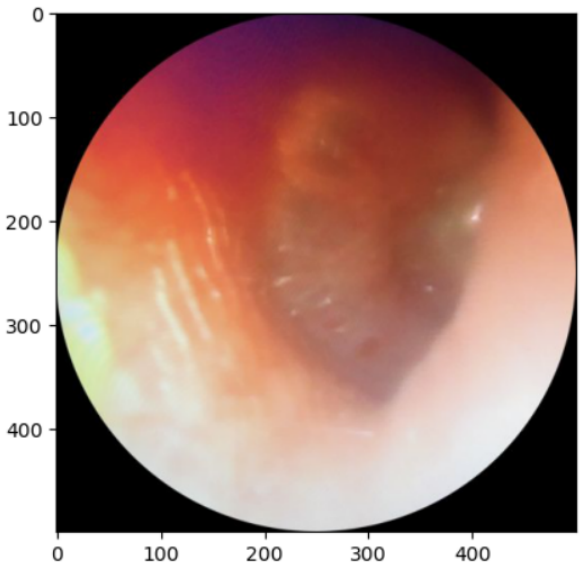

Figura 27 – Classificação 3

Fonte: Autoria própria.

```
Out[275]: {'normal': 0.9998374, 'anormal': 0.00016259906}
```

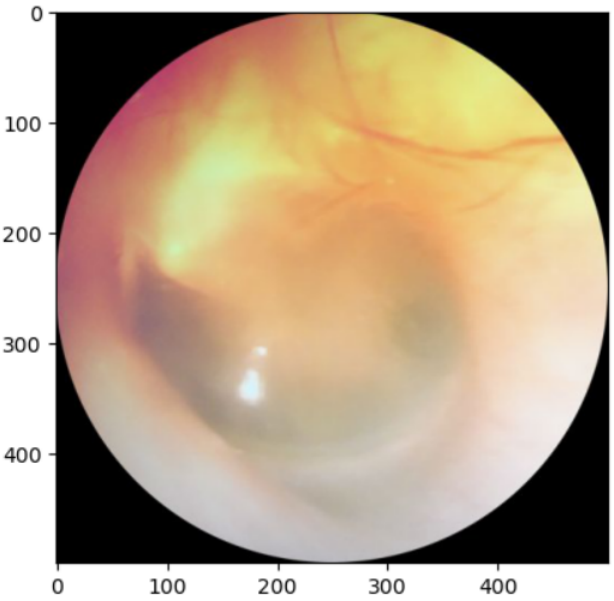

Figura 28 – Classificação 4

Fonte: Autoria própria.

```
Out[277]: {'normal': 0.98762465, 'anormal': 0.012375354}
```

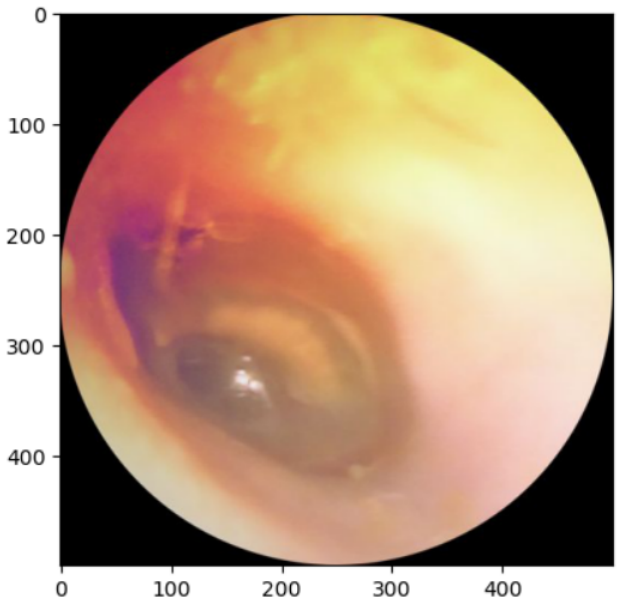

Figura 29 – Classificação 5

Fonte: Autoria própria.

```
Out[278]: {'normal': 0.999793, 'anormal': 0.00020706137}
```

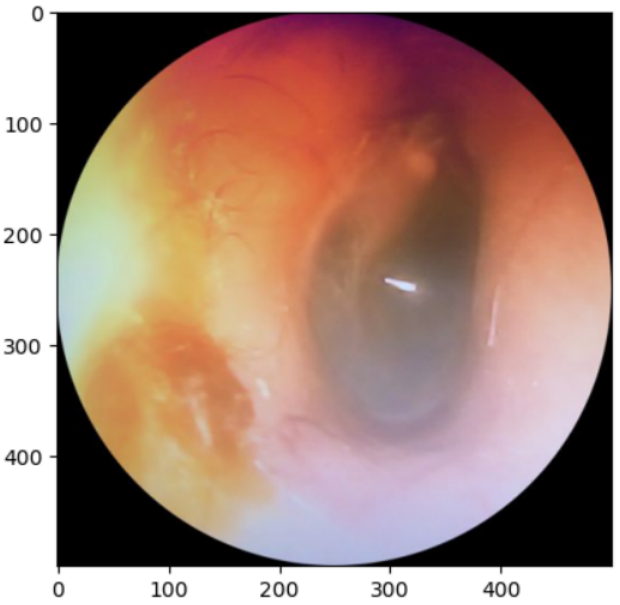

Figura 30 – Classificação 6

Fonte: Autoria própria.

4.5 Análise das métricas de desempenho Convnext Tiny

Após realizar ajustes nas camadas finais das CNNs, elas foram incorporadas a fim de buscar o melhor valor das métricas, *F1 Score*,*Recall* e *Precision*. Essa adaptação busca otimizar o desempenho global das CNNs, proporcionando uma avaliação mais detalhada de sua eficácia.

A inclusão dessas métricas adicionais no código, no modelo *Convnext Tiny* proporciona uma análise mais abrangente e refinada do desempenho dos modelos. O F1 Score, que combina precisão e *recall*, oferece uma visão equilibrada da capacidade do modelo em lidar com falsos positivos e falsos negativos. O *Recall*, também conhecido como taxa de verdadeiros positivos, destaca a habilidade do modelo em identificar corretamente todas as instâncias positivas. Por fim, a Precision mensura a proporção de instâncias positivas identificadas corretamente em relação ao total de instâncias identificadas como positivas.

A Tabela 12 apresenta métricas de desempenho para três cenários distintos: *AuAnomaliaOuNao*, *AuMulticlasse* e *AuSoOtite*.

Tabela 12 – Metricas de desempenho.

Método	f1score	Recall	Precision
AuAnomaliaOuNao	0.6418	0.7128	0.6902
AuMulticlasse	0.2019	0.2139	0.3454
AuSootite	0.9637	0.9797	0.9617

No cenário *AuAnomaliaOuNao*, observou-se um desempenho razoável na detecção de alterações na orelha média ou não, com um *F1 Score* de 0.6418. O equilíbrio entre precisão (0.6902) e *recall* (0.7128) indica a capacidade do modelo de identificar a maioria das alterações na orelha média reais, minimizando falsos positivos e falsos negativos.

Em relação ao cenário *AuMulticlasse*, o modelo enfrenta desafios, evidenciados pelo baixo *F1 Score* de 0.2019. O *recall* (0.2139) sugere dificuldades na identificação efetiva de todas as classes presentes, enquanto a precisão (0.3454) revela falta de consistência ao prever instâncias positivas. Provavelemente uma base de dados com mais amostras e mais equilibrada, resultaria em melhores resultados.

No cenário *AuSoOtite*, destaca-se o notável desempenho do modelo, com um F1 Score de 0.9637. O *recall* elevado (0.9797) destaca a capacidade do modelo de capturar quase todas as instâncias positivas de otite, enquanto a precisão (0.9617) indica um otímo desempenho ao classificar casos positivos.

Para o cenário *AuSoOtite*, foi realizado a analise da matriz de confusão. A análise da matriz de confusão revela-se como uma ferramenta para avaliar o desempenho do modelo. Ela proporciona uma representação organizada da relação entre as previsões feitas pelo modelo e os rótulos verdadeiros dos dados. Essa abordagem mais abrangente na

avaliação do desempenho dos modelos contribui para uma compreensão mais aprofundada de sua eficácia.

A matriz de confusão é uma ferramenta usada na avaliação do desempenho de modelos de classificação. Ela fornece uma visão detalhada das previsões feitas pelo modelo em relação aos resultados reais. A matriz de confusão na Figura 31 e seus resultados são explicados a seguir:

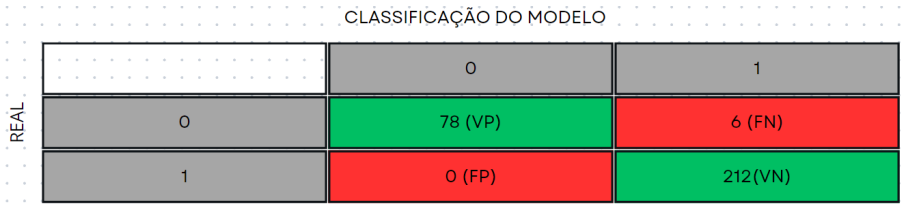

Figura 31 – Matriz de confusão

Fonte: Autoria própria.

- **Verdadeiros Positivos (VP):** 78 casos foram corretamente previstos como positivos pelo modelo. Isso significa que o modelo identificou corretamente 78 instâncias como pertencentes à classe positiva.
- **Falsos Negativos (FN):** 6 casos foram incorretamente previstos como negativos pelo modelo, quando na verdade eram positivos. Isso indica que o modelo falhou em reconhecer 6 instâncias que deveriam ter sido classificadas como pertencentes à classe positiva.
- **Falsos Positivos (FP):** Não houve casos incorretamente previstos como positivos pelo modelo quando na verdade eram negativos. Isso significa que o modelo não cometeu erros ao classificar instâncias negativas como positivas.
- **Verdadeiros Negativos (VN):** 212 casos foram corretamente previstos como negativos pelo modelo. Isso indica que o modelo identificou corretamente 212 instâncias como pertencentes à classe negativa.

4.6 Disponibilidade do código

Os *scripts* criados no desenvolvimento deste trabalho estão disponiveis em: código usado.

5 Conclusão

O presente trabalho teve como objetivo a comparação de diversos modelos de Redes Neurais Convolucionais (CNN) para a classificação de alterações do canal da orelha média.

A análise inicial, fundamentada em métricas de desempenho, como acurácia e erro médio, revelou que os modelos *MobileNetV2*, *convnext_small* e *convnext_tiny* , quando avaliados sem modificações, destacaram-se como os mais eficazes. Essa fase inicial de classificação, sem alterações nos modelos, desempenhou um papel crucial como um filtro na seleção dos modelos mais promissores. Esses critérios iniciais foram determinantes para a identificação dos modelos que se sobressaíram, proporcionando, posteriormente, uma análise mais aprofundada de seus desempenhos específicos em diversas condições

As modificações implementadas nos modelos foram fundamentais para conduzir um estudo focado na melhoria do desempenho. Além da avaliação do *F1 Score*,*recall* e precisão. Essas alterações foram determinantes para alcançar resultados superiores em comparação com as configurações iniciais.

O desempenho do modelo,*convnext_tiny* especialmente ao classificar a base *soOtite*, com um *F1 Score* de 0.9637. O *recall* de (0.9797) destaca a capacidade do modelo de capturar quase todas as instâncias positivas de otite, enquanto a precisão (0.9617) indica o desempenho ao classificar casos positivos. Os resultados promissores da *convnext_tiny* sugerem, assim, a eficácia na identificação de diferentes tipos de otite.

No entanto, o modelo não atingiu a espectativa na classificação de múltiplas patologias. É importante reconhecer a complexidade e diversidade das patologias auditivas. Especula-se que um único modelo pode não ser igualmente eficiente em diversas doenças. Outro possível motivo, pode ser a pequena quantidade de amostras para as diferentes patologias classificadas e o desbalanceamento da base de dados nesse caso.

A análise dos dados revelou que, para algumas condições específicas, o classificador apresentou desempenho inferior. Como perspectiva para futuras pesquisas, recomenda-se aprofundar as avaliações nos modelos que não obtiveram resultados satisfatórios. A realização de novos testes e ajustes nesses modelos pode fornecer *insights* valiosos para a escolha do melhor modelo em diferentes contextos clínicos. Além disso, a implementação em tempo real dos modelosutilizando imagens capturadas por um otoscópio digital é uma direção promissora, pois poderia contribuir significativamente para o auxílio ao diagnóstico rápido e eficaz de doenças no canal do orelha média. Essa abordagem facilitaria a integração das ferramentas desenvolvidas neste trabalho no contexto prático da área da saúde, proporcionando suporte aos profissionais de saúde no processo de identificação e tratamento precoce de doenças no canal do orelha média. Essas considerações abrem

caminho para avanços na aplicação prática da pesquisa em um cenário clínico mais amplo.

Referências

ADI ALHUDHAIF ZAFER CÖMERT, Kemal Polat. Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm. *PeerJ Computer Science*, v. 1, n. 1, p. 37, 2021. Citado 0 vez na página 27.

ALHUDHAIF, Adi; CÖMERT, Zafer; POLAT, Kemal. Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm. *PeerJ Computer Science*, PeerJ Inc., v. 7, e405, 2021. Citado 1 vez na página 17.

ALTUNTAŞ, Yahya; CÖMERT, Zafer; KOCAMAZ, Adnan Fatih. Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. *Computers and Electronics in Agriculture*, Elsevier, v. 163, p. 104874, 2019. Citado 1 vez na página 22.

BALLARD; BROWN. *Computer-Vision*. Prentice Hall, 1982. Citado 1 vez na página 13.

BISHOP, Christopher M; NASRABADI, Nasser M. *Pattern recognition and machine learning*. Springer, 2006. v. 4. Citado 1 vez na página 19.

BRESSEM, Keno K et al. Comparing different deep learning architectures for classification of chest radiographs. *Scientific reports*, Nature Publishing Group UK London, v. 10, n. 1, p. 13590, 2020. Citado 1 vez na página 22.

CHOO, Hyunwoo et al. Influenza screening via deep learning using a combination of epidemiological and patient-generated health data: development and validation study. *Journal of Medical Internet Research*, JMIR Publications Toronto, Canada, v. 22, n. 10, e21369, 2020. Citado 1 vez na página 28.

CHUNG, Junyoung et al. Gated feedback recurrent neural networks. In: PMLR. INTERNATIONAL conference on machine learning. 2015. P. 2067–2075. Citado 1 vez na página 14.

DUTT, Sreetama et al. Insights into the growing popularity of artificial intelligence in ophthalmology. *Indian Journal of Ophthalmology*, Wolters Kluwer–Medknow Publications, v. 68, n. 7, p. 1339, 2020. Citado 1 vez na página 16.

GÉRON, Aurélien. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc.", 2022. Citado 1 vez na página 20.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. MIT press, 2016. Citado 3 vezes nas páginas 20, 23.

GUO, Yanming et al. Deep learning for visual understanding: A review. *Neurocomputing*, Elsevier, v. 187, p. 27–48, 2016. Citado 0 vez na página 21.

HUANG, Zhihu; LENG, Jinsong. Analysis of Hu's moment invariants on image scaling and rotation. In: IEEE. 2010 2nd international conference on computer engineering and technology. 2010. v. 7, p. v7–476. Citado 1 vez na página 28.

KINGMA, Diederik P; BA, Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. Citado 4 vezes nas páginas 23, 24.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. v. 25. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Citado 1 vez na página 30.

LECUN, Y.; BOSER, B. et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 1989. DOI: 10.1162/neco.1989.1.4.541. Citado 1 vez na página 20.

LECUN, Yann; BENGIO, Yoshua et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, Cambridge, MA USA, v. 3361, n. 10, p. 1995, 1995. Citado 1 vez na página 20.

LIU, Zhuang et al. A ConvNet for the 2020s. *CoRR*, abs/2201.03545, 2022. arXiv: 2201.03545. Disponível em: https://arxiv.org/abs/2201.03545. Citado 1 vez na página 33.

MAHARJAN, Sunil et al. A novel enhanced softmax loss function for brain tumour detection using deep learning. *Journal of neuroscience methods*, Elsevier, v. 330, p. 108520, 2020. Citado 1 vez na página 23.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, Manole Ltda, v. 1, n. 1, p. 32, 2003. Citado 1 vez nas páginas 18, 19.

PATEL, Vimla L et al. The coming of age of artificial intelligence in medicine. *Artificial intelligence in medicine*, Elsevier, v. 46, n. 1, p. 5–17, 2009. Citado 1 vez na página 14.

PONTI, Moacir Antonelli et al. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: IEEE. 2017 30th SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T). 2017. P. 17–41. Citado 1 vez na página 13.

RUDER, Sebastian. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016. Citado 1 vez na página 24.

SANDLER, Mark et al. Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation. *CoRR*, abs/1801.04381, 2018. arXiv: 1801.04381. Disponível em: http://arxiv.org/abs/1801.04381. Citado 1 vez na página 31.

SILVA, Catarina Lopes. *História e Evolução da Otoscopia*. 2018. Diss. (Mestrado) – Universidade de Lisboa, Lisboa. Citado 1 vez na página 12.

SILVA, Catarina Lopes. *História e evolução da otoscopia*. 2018. Tese (Doutorado). Citado 1 vez na página 11.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado 1 vez na página 31.

SZELISKI, Richard. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010. Citado 1 vez na página 18.

TACCHELLA, Andrea et al. Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: a proof-of-principle study. *F1000Research*, Faculty of 1000 Ltd, v. 6, 2017. Citado 1 vez na página 16.

TAN, Mingxing; LE, Quoc V. EfficientNetV2: Smaller Models and Faster Training. *CoRR*, abs/2104.00298, 2021. arXiv: 2104.00298. Disponível em: https://arxiv.org/abs/2104.00298. Citado 1 vez na página 31.

ZAFER, Cömert. Fusing fine-tuned deep features for recognizing different tympanic membranes. *Biocybernetics and Biomedical Engineering*, v. 40, n. 1, p. 40–51, 2020. ISSN 0208-5216. DOI: https://doi.org/10.1016/j.bbe.2019.11.001. Disponível em: https://www.sciencedirect.com/science/article/pii/S0208521619304681. Citado 2 vezes nas páginas 17, 26.

ZENG, Xinyu et al. Efficient and accurate identification of ear diseases using an ensemble deep learning model. *Scientific Reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 10839, 2021. Citado 1 vez na página 16.