



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE
PRODUÇÃO

VINICIUS QUEIROZ DE OLIVEIRA

**ANÁLISE COMPARATIVA ENTRE
ARQUITETURAS DE REDES NEURAIAS
CONVOLUCIONAIS PARA APOIO
DIAGNÓSTICO DE CÂNCER DE PULMÃO A
PARTIR DE IMAGENS DE TOMOGRAFIA
COMPUTADORIZADA**

RECIFE

2024

VINICIUS QUEIROZ DE OLIVEIRA

**ANÁLISE COMPARATIVA ENTRE
ARQUITETURAS DE REDES NEURAIAS
CONVOLUCIONAIS PARA APOIO
DIAGNÓSTICO DE CÂNCER DE PULMÃO A
PARTIR DE IMAGENS DE TOMOGRAFIA
COMPUTADORIZADA**

Trabalho de Conclusão de Curso apresentado a banca avaliadora de Engenharia de Produção, como requisito parcial para obtenção do título de Bacharel em Engenharia de Produção.

Orientadora: Isis Didier Lins, DSc.

RECIFE

2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Oliveira, Vinicius Queiroz de.
Análise Comparativa Entre
Arquiteturas de Redes Neurais
Convolucionais Para Apoio Diagnóstico de
Câncer de Pulmão a Partir de Imagens de
Tomografia Computadorizada / Vinicius
Queiroz de Oliveira. - Recife, 2024. 59 p. :
il., tab.

Orientador(a): Isis Didier Lins
Trabalho de Conclusão de Curso (Graduação) - Universidade
Federal de Pernambuco, Centro de Tecnologia e Geociências,
Engenharia de Produção - Bacharelado, 2024.

1. Redes Neurais Convolucionais. 2. Eficiência Hospitalar.
3. Aprendizado por Transferência. 4. Diagnóstico de Exames
de Imagem. 5. Precisão Diagnóstica. 6. Tomografia
Computadorizada. I. Lins, Isis Didier. (Orientação). II. Título.

**ANÁLISE COMPARATIVA ENTRE
ARQUITETURAS DE REDES NEURAI
CONVOLUCIONAIS PARA APOIO
DIAGNÓSTICO DE CÂNCER DE PULMÃO A
PARTIR DE IMAGENS DE TOMOGRAFIA
COMPUTADORIZADA**

Trabalho de Conclusão de Curso
apresentado à banca avaliadora de
Engenharia de Produção, como
requisito parcial para obtenção do título
de Bacharel em Engenharia de
Produção.

Aprovado em: 14/03/2024

BANCA EXAMINADORA

Profa. Dra. Isis Didier Lins (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. Rodrigo José Pires Ferreira (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dra. Lucia Reis Peixoto Roseli (Examinador Interno)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos a todos que contribuíram direta ou indiretamente para a conclusão desta pesquisa, bem como para minha formação acadêmica e pessoal.

Aos professores e orientadores, meu mais profundo agradecimento por dedicarem seu tempo e cuidado para transmitir conhecimento e por incentivarem minha curiosidade. Aos meus amigos, que estiveram ao meu lado durante toda essa jornada, meu sincero reconhecimento pela companhia e apoio inestimável. E à minha família, por fazer parte integral da minha vida e formação, meu mais profundo agradecimento por seu amor, suporte e compreensão em todos os momentos.

Cada um de vocês desempenhou um papel essencial no meu crescimento acadêmico e pessoal, e sou imensamente grato por todo o apoio e incentivo ao longo do caminho. Suas contribuições foram verdadeiramente valorizadas e deixaram uma marca eterna em minha jornada.

RESUMO

Com o avanço tecnológico e a crescente necessidade de métodos diagnósticos mais precisos, as redes neurais convolucionais (CNNs) emergem como uma ferramenta promissora na medicina diagnóstica, complementando o trabalho dos profissionais de saúde em seus diagnósticos. Este estudo, inserido no contexto de eficiência hospitalar, realiza uma análise comparativa abrangente de diferentes modelos e arquiteturas de CNNs no apoio ao diagnóstico de câncer de pulmão a partir de imagens de tomografia computadorizada (TC), visando aprimorar a agilidade na análise desses exames e melhorar a confiabilidade de seus resultados. São destacadas as diferenças de desempenho entre os modelos, bem como a aplicação de técnicas de aprendizado por transferência para potencializar o seus desempenhos. O objetivo principal é não apenas identificar o modelo mais eficaz para essa tarefa crucial, mas também explorar estratégias inovadoras que possam melhorar significativamente a precisão e a confiabilidade dos diagnósticos de câncer de pulmão, contribuindo tanto para a eficiência em ambientes hospitalares quanto para o conhecimento acerca da eficiência de novos modelos neste contexto específico.

Palavras-chave: redes neurais convolucionais, eficiência hospitalar, aprendizado por transferência, diagnóstico de exames de imagem, precisão diagnóstica, tomografia computadorizada.

ABSTRACT

With the technological advancement and the growing need for more precise diagnostic methods, Convolutional Neural Networks (CNNs) emerge as a promising tool in diagnostic medicine, complementing the work of healthcare professionals in their diagnoses. This study, inserted in the context of hospital efficiency, conducts a comprehensive comparative analysis of different CNN models and architectures in supporting the diagnosis of lung cancer from computed tomography (CT) images, aiming to enhance the agility in analyzing these exams and improve the reliability of their results. Performance differences between models are highlighted, as well as the application of transfer learning techniques to enhance their performances. The main objective is not only to identify the most effective model for this crucial task but also to explore innovative strategies that can significantly improve the accuracy and reliability of lung cancer diagnoses, contributing to both efficiency in hospital environments and advancing knowledge regarding the effectiveness of new models in this specific context.

Keywords: convolutional neural networks, hospital efficiency, transfer learning, diagnostic imaging, diagnostic accuracy, computed tomography.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação gráfica de uma <i>MLP</i>	18
Figura 2 – Arquitetura da rede <i>VGG16</i>	21
Figura 3 – Arquitetura Clássica da rede <i>ResNet50</i>	22
Figura 4 – Processo de classificação com arquitetura <i>EfficientNet</i>	24
Figura 5 – Exemplo de Arquitetura da Rede <i>ConvNext</i>	25
Figura 6 - Amostras da Base de Dados IQ-OTH/NCCD.....	41
Figura 7 - Amostras de imagens após pré-processamento	44
Figura 9 – Matrizes de confusão <i>VGG16</i>	47
Figura 10 – Matrizes de Confusão <i>ResNet50</i>	47
Figura 11 – Matrizes de Confusão <i>EfficientNetV2 B0</i>	48
Figura 12 – Matrizes de Confusão <i>ConvNext Tiny</i>	48
Figura 13 – Gráficos de Acurácia e Perda <i>VGG16</i>	48
Figura 14 – Gráficos de Acurácia e Perda <i>ResNet50</i>	49
Figura 15 – Gráficos de Acurácia e Perda <i>EfficientNetV2 B0</i>	49
Figura 16 – Gráficos de Acurácia e Perda <i>ConvNext Tiny</i>	50
Figura 17 – Imagens de Previsão do Modelo <i>VGG16</i>	51

LISTA DE TABELAS

Tabela 1 – Configurações <i>ImageNet</i>	27
Tabela 2 - Performances <i>ImageNet</i>	27
Tabela 3 – Tipos de Classificação de Erro.....	29
Tabela 4 - Hiperpâmetros dos modelos-base	46
Tabela 5 – Resumo de Desempenho dos Modelos	50

SUMÁRIO

1. INTRODUÇÃO.....	11
1.1 Justificativa e Relevância	12
1.2 Objetivo.....	13
1.2.1 Objetivo Geral	13
1.2.2 Objetivos Específicos.....	13
1.3 Metodologia	13
1.3.1 Categorização da Metodologia.....	14
2. FUNDAMENTAÇÃO TEÓRICA.....	16
2.1 Aprendizado Supervisionado	16
2.1.1 Fatores na Escolha do Algoritmo	16
2.2 <i>Perceptron</i> Multicamadas	18
2.3 Redes Neurais Convolucionais	19
2.4 Modelos de Redes Neurais Convolucionais	21
2.4.1 <i>VGG16</i>	21
2.4.2 <i>ResNet50</i>	22
2.4.3 <i>EfficientNet V2 B0</i>	23
2.4.4 <i>ConvNextT</i>	24
2.4.5 Comparação de Desempenho <i>ImageNet</i>	25
2.5 Métricas Para a Avaliação de Desempenho.....	28
2.5.1 Medição do Erro.....	28
2.5.2 Outras Medidas de Avaliação	28
2.6 Aprendizado por Transferência	30
3. REVISÃO DA LITERATURA.....	32
3.1 Estudos Seleccionados.....	33

4. CONSTRUÇÃO DOS MODELOS.....	39
4.1 Programas e Bibliotecas	39
4.2 Modelagem e Análise.....	40
4.3 Base de Dados Utilizada	40
4.4 Tratamento de Imagens	41
4.5 Balanceamento de Dados	42
4.6 Divisão em Treino, Teste e Validação.....	44
4.7 Treinamento e Fine Tuning dos Modelos	45
5. RESULTADOS	47
5.1 Matrizes de Confusão	47
5.2 Gráficos de Acurácia e Perda ao Longo das Épocas	48
5.3 Resumo de Desempenhos.....	50
5.4 Análise de Resultados	50
6. CONCLUSÃO.....	54
6.1 Limitações e Sugestões Para Trabalhos Futuros	54
REFERÊNCIAS	56

1. INTRODUÇÃO

Em um ambiente hospitalar, diversos são os fatores que podem influenciar na qualidade dos serviços, e cada vez mais ferramentas são utilizadas para a melhoria desses processos para garantir a segurança e satisfação dos clientes e dos profissionais de saúde. Esses sistemas são amplamente utilizados para coletar, armazenar e gerenciar informações essenciais que auxiliam no funcionamento e tomada de decisões médicas em ambientes hospitalares (TAMBLYN, ROBYN, et al, 2003).

Assim, modelos de aprendizado de máquina e inteligência artificial (IA) são integrados a esse fluxo de processos digitais e atuam como uma poderosa ferramenta, com o objetivo primário de entender as relações entre dados clínicos e a situação de pacientes (COIERA, 1997). Além disso, este tipo de tecnologia se demonstra capaz de realizar funções previamente atribuídas exclusivamente a análises humanas, gerando maneiras inovadoras de diagnosticar, tratar e prevenir doenças e enfermidades (MULLAINATHAN, 2022).

Nesse contexto, a área de radiologia com ênfase em diagnóstico e prevenção de câncer em tomografias computadorizadas (*computed tomography* – TC) possui problemas conhecidos e considerados recorrentes, como a demanda por profissionais experientes superar a oferta, ou até mesmo as situações onde os dados disponíveis são muito complexos para serem analisados de uma forma eficiente em diagnósticos manuais (HOSNY, 2018). Essas situações, evidentemente, podem gerar gargalos nítidos em sistemas hospitalares, ao mesmo tempo que possuem características em sua problemáticas com potencial de serem solucionadas pelo uso de modelos de IA com assertividade equiparável a de profissionais experientes, como é o caso de modelos de redes neurais convolucionais (*Convolutional Neural Networks* – CNNs) (PANTANOWITZ, 2020).

Dada a natureza de rápida evolução desse tipo de modelo, o intuito deste trabalho é investigar, avaliar e comparar a performance de arquiteturas de CNNs consolidadas na literatura vigente como confiáveis em situações de diagnósticos de malignidade de nódulos pulmonares em CTs, com a performance de arquiteturas mais modernas que prometem ser mais eficientes e precisas em uma quantidade relativamente limitada de dados.

Na área da oncologia, modelos de inteligência artificial tem sido usados há anos

para auxílio no diagnóstico, sendo utilizados principalmente para prever qual tipo de tratamento o paciente terá maior probabilidade de responder positivamente, baseado em critérios individuais (BHINDER, 2021). Essa eficiência vem sendo mostrada ao longo dos anos como um conceito cada vez mais confiável, com um crescente número de casos onde sua assertividade demonstrou-se superior à de profissionais experientes (PANTANOWITZ, 2020).

Diante desses fatos, é natural um crescente interesse na melhoria desses sistemas e modelos, sendo esse um dos fatores motivadores desta pesquisa.

1.1 Justificativa e Relevância

A crescente incidência e mortalidade associadas ao câncer de pulmão ressaltam a urgência e a relevância de aprimorar as ferramentas de diagnóstico e detecção precoce dessa doença, que se tornou uma das mais frequentes na população mundial e a segunda maior causa de morte por câncer em todo o mundo (SUNG, 2020).

A detecção precoce desempenha um papel crucial na sobrevivência dos pacientes, com estimativas sugerindo um aumento significativo nas taxas de sobrevivência quando o câncer é diagnosticado em estágios iniciais (LENDERLIN, et al., 2013). A complexidade do diagnóstico do câncer de pulmão, aliada à necessidade de identificar e monitorar nódulos pulmonares de forma precisa e eficaz, motiva o contínuo desenvolvimento e investigação de métodos e técnicas capazes de oferecer respostas rápidas e confiáveis.

Além disso, o aperfeiçoamento dessas técnicas computacionais, especialmente aquelas baseadas em processamento de imagens e reconhecimento de padrões em imagens de TC, tem sido impulsionado pelo desejo de oferecer aos profissionais de saúde ferramentas que sejam capazes de auxiliá-los a acelerar análises em situações onde não há oferta suficiente de médicos frente a demanda presente, evitando gargalos e atrasos (HOSNY, 2018).

Os benefícios porém, não se limitam aos resultados de um uso específico. O teste intensivo de novos modelos e algoritmos incentiva sua posterior pesquisa, contribuindo para sua avaliação e potencial melhoria de performance em outros casos de diagnóstico (CURIGLIANO, et al. 2017). Dito isso, a crescente disponibilidade de modelos potencialmente capazes de suprir essas necessidades de formas cada vez mais eficientes, a custos computacionais menores, com alta performance e

necessidade de quantidades cada vez menores de dados, são justificativas para a contínua comparação entre modelos de CNN que se provaram capazes de gerar resultados satisfatórios no apoio ao diagnóstico de doenças usando imagens radiológicas.

1.2 Objetivo

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é realizar uma análise comparativa de desempenho de diferentes modelos e arquiteturas de redes neurais convolucionais (CNNs) consolidados e amplamente testados na literatura, no contexto de apoio ao diagnóstico de câncer de pulmão utilizando imagens de TCs, visando aprimorar a eficiência diagnóstica em ambientes hospitalares.

1.2.2 Objetivos Específicos

- Investigar a performance de modelos de previsão previamente utilizados para propósitos similares ao desta pesquisa através de uma extensa revisão bibliográfica.
- Treinar modelos recentes e consolidados capazes de classificar a malignidade de tumores pulmonares em exames de TC utilizando uma bases de dados de fácil acesso e com amplo reconhecimento na literatura;
- Comparar a performance de diferentes arquiteturas de CNNs em diagnósticos de câncer a partir de diversas métricas pertinentes;
- Avaliar qual modelo deve ser priorizado para auxílio diagnóstico com base em métricas pertinentes.

1.3 Organização do Trabalho

Nesse trabalho, serão introduzidos os principais conceitos e fundamentos teóricos para compreender o desenvolvimento e a análise de modelos de redes neurais convolucionais para a tarefa de categorização das imagens de TC.

No Capítulo 2, será abordada a fundamentação Teórica, explorando conceitos

de Aprendizado Supervisionado, o Perceptron Multicamadas, Redes Neurais Convolucionais e modelos como VGG16, ResNet50, EfficientNet V2 B0 e ConvNextT. Também serão apresentadas métricas de avaliação de desempenho e o conceito de aprendizado por transferência.

No Capítulo 3, será realizada uma revisão da literatura, destacando estudos relevantes que nos darão um melhor contexto do estado da arte desses modelos.

O Capítulo 4 detalhará a Metodologia adotada, incluindo programas e bibliotecas, modelagem, base de dados, tratamento de imagens, balanceamento de dados e treinamento dos modelos.

No Capítulo 5, serão apresentados os resultados obtidos, incluindo análise quantitativa e qualitativa. E por fim, no Capítulo 6, serão abordadas as possíveis limitações e sugestões para pesquisas futuras.

1.3.1 Categorização da Metodologia

As definições de Cauchick (2019) foram utilizadas para classificação da pesquisa, atribuindo descrições pré estabelecidas quanto ao método científico, finalidade, natureza, objetivo e dados.

Quanto ao método científico, a pesquisa adota um método dedutivo, uma vez que parte de teorias e conceitos estabelecidos sobre CNNs para testar e comparar modelos específicos aplicados à categorização de imagens de TC de câncer como malignas, benignas ou normais.

Quanto à finalidade, é classificada como aplicada, pois busca resolver um problema prático, aplicando conhecimentos científicos na prática para melhorar a categorização de imagens médicas.

O objetivo da pesquisa é descritivo, uma vez que se propõe a descrever e comparar a performance de quatro modelos diferentes de CNNs em relação ao problema descrito.

Quanto à natureza da pesquisa, ela é considerada quantitativa, pois envolve a coleta e análise de dados numéricos, como métricas de desempenho dos modelos, taxas de acerto, precisão, *recall*, entre outros.

Em relação aos dados utilizados, a pesquisa se enquadra na categoria de modelagem, uma vez que são construídos e treinados modelos de CNNs utilizando uma base de dados real para representar e compreender o fenômeno complexo da categorização de imagens médicas.

Em resumo, este estudo adota uma abordagem dedutiva e aplicada, com objetivo descritivo, natureza quantitativa e envolvimento de modelagem de dados.

2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, serão abordados temas essenciais para a compreensão posterior do desempenho dos modelos testados na pesquisa, iniciando pelos fundamentos, passando pelo detalhamento de cada modelo e finalizando com as métricas que serão utilizadas.

2.1 Aprendizado Supervisionado

O Aprendizado de Máquina Supervisionado é um ramo da inteligência artificial que se dedica ao desenvolvimento de algoritmos e técnicas capazes de aprender padrões e realizar previsões a partir de dados de entrada, conhecidos como conjunto de treinamento, e suas respectivas saídas desejadas, denominadas rótulos ou classes (CARUANA, 2006).

Nesse contexto, o termo "supervisionado" refere-se à presença de uma supervisão explícita durante o processo de treinamento, onde o algoritmo é apresentado aos dados rotulados e ajusta seus parâmetros de acordo com a comparação entre as previsões geradas e os rótulos fornecidos.

A principal característica do aprendizado supervisionado é a capacidade de generalização, ou seja, a habilidade do modelo de fazer previsões precisas em dados não vistos anteriormente, baseando-se nos padrões aprendidos durante o treinamento. Isso é possível graças à capacidade dos algoritmos supervisionados de capturar relações complexas entre os atributos de entrada e as saídas desejadas, adaptando-se a diferentes cenários e contextos, como problemas de classificação de imagem ou de previsão de séries temporais.

2.1.1 Fatores na Escolha do Algoritmo

Não existe um único algoritmo capaz de resolver todos os problemas de forma eficiente, e por isso há diversos fatores a serem considerados durante a escolha da ferramenta que será utilizada (HASTIE et al., 2009).

Trade-off Viés-Variância

O trade-off viés-variância é um conceito chave no aprendizado de máquina supervisionado, que descreve a relação inversa entre a capacidade de generalização de um modelo e sua complexidade. Em essência, trata-se de equilibrar a simplificação

do modelo (viés) com sua sensibilidade aos dados de treinamento (variância).

Um modelo com alto viés pode falhar em capturar relações complexas nos dados (subajuste), enquanto um modelo com alta variância pode capturar até mesmo o ruído nos dados (sobreajuste). Encontrar o equilíbrio adequado entre viés e variância é crucial para minimizar o erro de generalização do modelo.

Ruídos de Valores de Saída

O ruído nos valores de saída é uma consideração importante no aprendizado de máquina supervisionado, referindo-se à presença de variações aleatórias nos rótulos ou valores de destino associados aos dados de treinamento. Esse ruído pode surgir de várias fontes, como erros de medição, imprecisões nos rótulos fornecidos ou inerente à própria natureza dos dados.

Quando os valores de saída contêm ruído, isso pode afetar adversamente o desempenho do modelo, levando a previsões menos precisas e uma capacidade de generalização comprometida. Modelos sensíveis ao ruído podem tentar ajustar-se excessivamente a essas variações aleatórias, resultando em um desempenho inferior quando aplicados a dados não vistos.

Complexidade do Modelo e Quantidade de Dados

A relação entre a complexidade do modelo e a quantidade de dados de treinamento é crucial no aprendizado de máquina supervisionado. Modelos mais complexos requerem uma quantidade suficiente de dados para evitar o sobreajuste, um problema comum onde modelos não conseguem realizar uma generalização boa o suficiente para fazer previsões em dados nunca vistos.

Encontrar o equilíbrio adequado entre a complexidade do modelo e a quantidade de dados disponíveis é essencial para garantir um desempenho ótimo do modelo e sua capacidade de generalizar para novos dados não vistos. Este tema é amplamente discutido na literatura de aprendizado de máquina e é essencial para o desenvolvimento eficaz de modelos de aprendizado de máquina em aplicações do práticas.

Outros Fatores

Outros fatores a serem considerados ao escolher e aplicar um algoritmo de aprendizado incluem:

Heterogeneidade dos dados: alguns algoritmos são mais fáceis de aplicar em conjuntos de dados com diferentes tipos de características. Muitos exigem que as características sejam numéricas e escalonadas de forma semelhante, enquanto outros, como árvores de decisão, lidam bem com a heterogeneidade.

Redundância nos dados: se as características contiverem informações redundantes, certos algoritmos podem ter desempenho prejudicado devido a instabilidades numéricas, mas isso pode ser resolvido com regularização.

Presença de interações e não linearidades: algoritmos como árvores de decisão e redes neurais são mais adequados para detectar interações complexas entre as características, enquanto métodos lineares exigem especificação manual das interações.

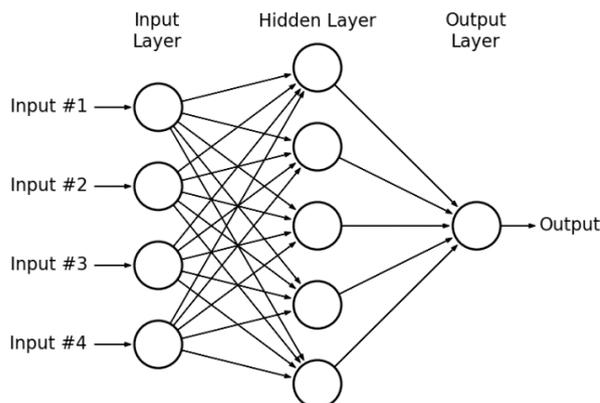
2.2 *Perceptron* Multicamadas

O *Perceptron* Multicamadas, ou *Multilayer Perceptron (MLP)* representa um marco significativo na evolução das redes neurais artificiais, com suas raízes históricas remontando ao conceito original de *Perceptron*, proposto por Frank Rosenblatt na década de 1950 (ROSENBLATT, 1958). O *Perceptron*, embora tenha mostrado promessa como um modelo de aprendizado de máquina, logo foi considerado limitado devido à sua incapacidade de lidar com problemas que não eram linearmente separáveis. Esse obstáculo levou a um declínio no interesse em redes neurais durante as décadas seguintes, à medida que a comunidade científica se voltava para abordagens de aprendizado de máquina mais convencionais.

No entanto, o interesse em redes neurais foi reavivado na década de 1980, com o desenvolvimento do MLP por pesquisadores como Geoffrey Hinton, David Rumelhart e Ronald Williams (RUMELHART et al., 1986). O MLP representou uma mudança de paradigma ao introduzir múltiplas camadas de neurônios interconectados, permitindo a criação de modelos de aprendizado profundo capazes de aprender representações complexas dos dados. Com essa arquitetura mais complexa, o *MLP* superou as limitações do *Perceptron* original e tornou-se uma ferramenta poderosa em uma variedade de aplicações de aprendizado de máquina. Sua estrutura pode ser interpretada na figura 1

:

Figura 1 – Representação gráfica de uma *MLP*



Fonte: MOHAMED, 2015

A fundamentação matemática do MLP é baseada em princípios de álgebra linear e cálculo. Em cada neurônio de uma camada do MLP, ocorre uma combinação linear das entradas ponderadas pelos pesos sinápticos correspondentes, seguida pela aplicação de uma função de ativação não linear. Essa primeira operação pode ser expressa pela equação 1:

$$a_j = f\left(\sum_{i=1}^n w_{ij}x_{ij} + b_j\right) \quad (1)$$

onde a_j representa a saída do neurônio j , f é a função de ativação, w_{ij} são os pesos das conexões entre o neurônio i da camada anterior e o neurônio j da camada atual, x_i são as entradas do neurônio e b_j é o viés do neurônio.

Uma variedade de funções de ativação pode ser utilizada, cada uma introduzindo diferentes características de não-linearidade na rede. Exemplos comuns incluem a função sigmoide logística, a função tangente hiperbólica, a função ReLU (*Rectified Linear Unit*) e outras.

O MLP e suas variantes têm sido amplamente utilizados em uma variedade de aplicações de aprendizado de máquina (GOODFELLOW, 2016). Seu impacto no campo da inteligência artificial tem sido notável, impulsionando avanços significativos na compreensão e modelagem de dados complexos.

2.3 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNNs) são um tipo especializado de rede neural artificial, projetadas especificamente para processar dados estruturados de

forma espacial, como imagens. Elas são compostas por camadas convolucionais que aprendem representações hierárquicas de características nos dados de entrada, tornando-as extremamente eficazes em tarefas de visão computacional, reconhecimento de padrões e processamento de imagem (LECUN et al., 2015).

As CNNs são compostas por diferentes tipos de camadas, incluindo camadas convolucionais, de pooling e totalmente conectadas. Cada camada convolucional consiste em um conjunto de filtros (ou kernels) que são aplicados aos dados de entrada para extrair características específicas. Durante o treinamento, os pesos desses filtros são ajustados através do processo de aprendizado, de modo que as características relevantes sejam detectadas de maneira eficiente (GOODFELLOW et al., 2016).

Na arquitetura de uma CNN, as características são extraídas de forma progressiva à medida que os dados passam de uma camada para outra na rede, e armazenadas em matrizes denominadas mapas de características. Cada neurônio em uma camada K está conectado a um campo receptivo local na camada anterior $K-1$. Essa organização permite que as características sejam combinadas sequencialmente, resultando na extração de características de níveis mais elevados ao longo das camadas da rede.

Dentro da estrutura das Convolutional Neural Networks (CNNs), as camadas de convolução desempenham um papel crucial. Essas camadas são compostas por filtros ajustáveis, os quais são aplicados individualmente a toda a imagem de entrada, resultando na produção de uma variedade de mapas de características (LECUN et al., 2010). Após o processo de treinamento da rede, cada filtro é capaz de se especializar na detecção de uma característica específica, independente de sua localização exata na imagem (HAFEMANN, 2014). Isso significa que a rede aprende a identificar padrões relevantes em diferentes partes da imagem, sem depender de sua posição exata (LECUN et al., 1998). Essa capacidade de generalização torna as CNNs extremamente poderosas em tarefas de reconhecimento visual e processamento de imagens. Pode-se definir o mapa de características pela equação 2:

$$(h_k)_{ij} = (w_k * x)_{ij} + b_k \quad (2)$$

onde k é o índice do neurônio, h_k é o mapa de características de saída k , w_k é o filtro ajustável k , x é o mapa de características de entrada, b_k é o termo de viés k e

"*" representa o processo de convolução espacial bidimensional.

Após cada etapa de convolução, segue-se uma fase de subamostragem (*subsampling*), que consiste na extração de amostras de cada mapa de características. Essas amostras podem ser obtidas de várias formas, como a soma, a média, a seleção do valor máximo (*max pooling*) ou mínimo (*min pooling*) dentro de uma região específica. Essa técnica promove uma síntese ou resumo do mapa de características (LECUN et al., 1998).

Seguindo esses processos, que são responsáveis pela extração de características, os pixels de todos os mapas de características da camada anterior são apresentados à camada completamente conectada. Esta camada é responsável pela classificação dos padrões de entrada, de maneira similar a uma MLP.

2.4 Modelos de Redes Neurais Convolucionais

2.4.1 VGG16

A *VGG16* é uma arquitetura de rede neural convolucional (CNN) proposta por Simonyan e Zisserman em seu artigo "*Very Deep Convolutional Networks for Large-Scale Image Recognition*", apresentado na Conferência de Visão Computacional e Reconhecimento de Padrões (CVPR) de 2014.

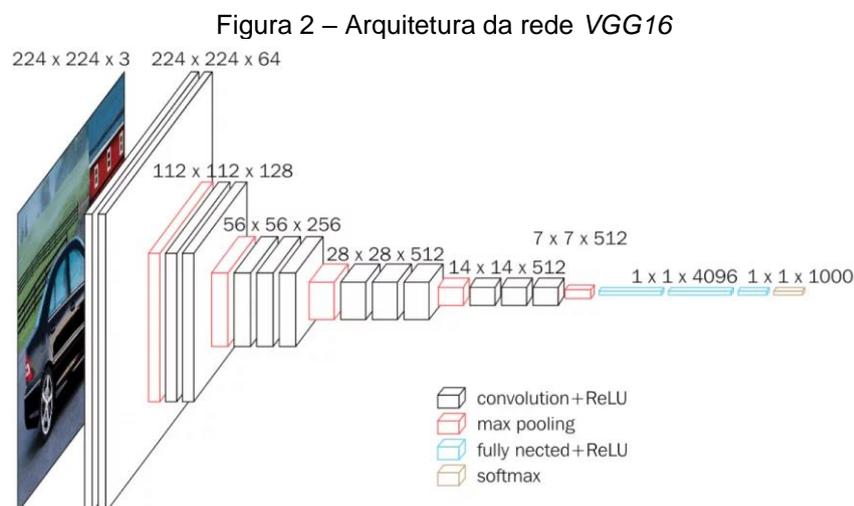
É conhecida por sua simplicidade e profundidade relativa em comparação com outras arquiteturas de CNN contemporâneas. Ela consiste em 16 camadas de convolução e pooling, seguidas por três camadas totalmente conectadas para a classificação final.

Esta arquitetura também é conhecida por sua simplicidade, utilizando convoluções de tamanho fixo (3x3) e pooling (geralmente *max-pooling* de 2x2) em toda a rede, sem a utilização de técnicas mais avançadas, como normalização de lote ou blocos residuais.

Apesar de sua simplicidade, a *VGG16* obteve resultados impressionantes em benchmarks de classificação de imagens, incluindo o conjunto de dados ImageNet. Ela foi uma das primeiras arquiteturas a demonstrar que o aumento da profundidade da rede poderia melhorar significativamente o desempenho em tarefas de visão computacional.

Devido à sua simplicidade e eficácia, a *VGG16* tornou-se uma escolha popular como uma linha de base para várias aplicações de visão computacional, bem como

um ponto de partida comum para experimentos em arquiteturas de redes neurais mais complexas.



Fonte: HINDARTO, 2023

2.4.2 ResNet50

A *ResNet50* é uma arquitetura de rede neural convolucional (CNN) proposta por He et al. (2015). em seu artigo "Deep Residual Learning for Image Recognition", apresentado na Conferência de Visão Computacional e Reconhecimento de Padrões (CVPR) de 2016.

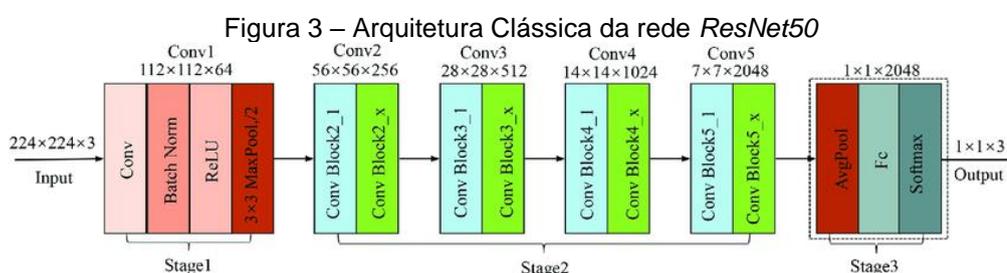
A *ResNet50* é uma versão específica da família de redes ResNet (Redes Residuais) e é notável por sua profundidade relativamente grande (50 camadas) e sua capacidade de resolver o problema de degradação de desempenho que surge com o aumento da profundidade da rede devido a dificuldade de otimização e do problema de desaparecimento de gradiente (HE et. al, 2015).

A principal inovação das redes ResNet, incluindo a *ResNet50*, é o conceito de blocos residuais. Esses blocos residuais permitem que a rede aprenda atalhos ou "resíduos" das conexões de entrada para as saídas dos blocos, o que facilita o treinamento de redes mais profundas. Isso ajuda a prevenir problemas como o desaparecimento do gradiente, que ocorre quando a informação é perdida à medida que passa por muitas camadas na rede, dificultando o treinamento.

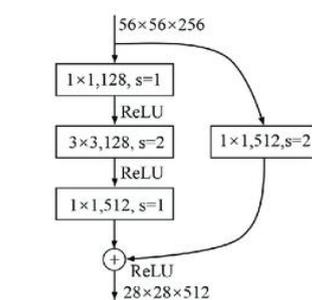
Ela é composta por 50 camadas, incluindo 49 blocos convolucionais e uma camada final de classificação. Cada bloco convolucional consiste em várias camadas convolucionais, normalização de lote e ativações de ReLU, com conexões de atalho

para contornar as camadas convolucionais. Esses blocos residuais são empilhados em cascata para formar sua arquitetura completa.

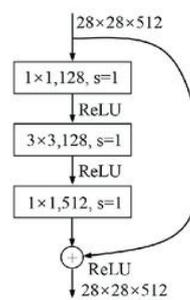
Devido à sua profundidade e à capacidade de aprendizado eficaz de representações complexas, a *ResNet50* tornou-se outra escolha popular em aplicações de visão computacional, como classificação de imagens, detecção de objetos e segmentação semântica. É frequentemente utilizada como uma linha de base para comparar o desempenho de novas arquiteturas de rede neural em várias tarefas. Sua arquitetura é exemplificada na figura 3:



a. Network structure of classical Resnet50 model



b. Residual structure with added scale



c. Residual structures that do not change dimensions

Fonte: DU, 2023

2.4.3 EfficientNet V2 B0

A EfficientNetV2 é uma melhoria da arquitetura original EfficientNet, que foi proposta pela primeira vez em 2019. A ideia por trás da EfficientNet é encontrar um equilíbrio entre eficiência computacional e desempenho em tarefas de visão computacional, como classificação de imagens (TAN; LE, 2021).

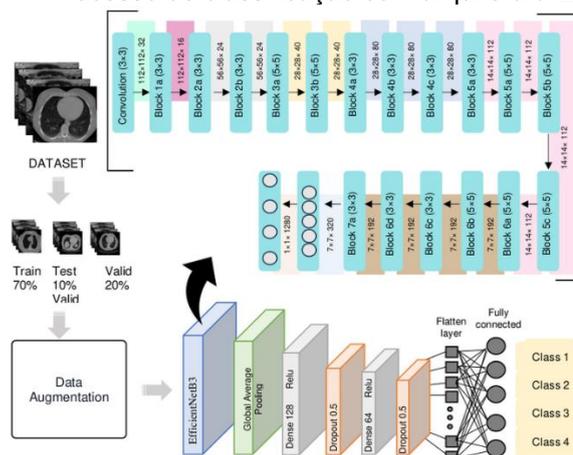
A arquitetura EfficientNet original alcançou um alto desempenho com um número relativamente pequeno de parâmetros, o que a tornou uma escolha popular em aplicações onde recursos computacionais são limitados, como em dispositivos móveis ou sistemas embarcados. Isso foi alcançado usando uma técnica chamada de escalamento composto, que aumenta a profundidade, a largura e a resolução da rede

de maneira balanceada.

A EfficientNetV2, por sua vez, propõe melhorias na arquitetura original, mantendo o mesmo princípio de equilíbrio entre eficiência e desempenho. Uma das principais características da EfficientNetV2 é a adoção de blocos convolucionais chamados "Residual Normalization" (RN), que integram a normalização e o residual connection diretamente no bloco convolucional. Esses blocos RN são eficazes na redução do custo computacional sem comprometer significativamente o desempenho.

Além disso, a EfficientNetV2 também introduz a técnica de "Width Mismatch Correction", que ajusta a largura da rede para lidar melhor com as diferenças de escala entre os estágios da rede. Isso ajuda a melhorar ainda mais a eficiência e o desempenho da rede. Sua arquitetura é exemplificada na figura 4:

Figura 4 – Processo de classificação com arquitetura *EfficientNet*



Fonte: NEFEA, 2023

2.4.4 ConvNextT

ConvNetX (ConvNext) é uma arquitetura de rede neural convolucional proposta em um artigo de pesquisa intitulado "A Convolutional Network for the 2020s", que foi apresentado na Conferência de Visão Computacional e Reconhecimento de Padrões (CVPR) de 2022.

Esta arquitetura foi projetada com o objetivo de superar as limitações das arquiteturas convolucionais tradicionais, como a ResNet e a EfficientNet, e melhorar o desempenho em uma variedade de tarefas de visão computacional (LIU, 2022). Algumas características principais da ConvNetX incluem:

Path Integration Blocks (PIBs): estes blocos são responsáveis por integrar

informações de múltiplos caminhos na rede. Eles ajudam a capturar características complexas em diferentes escalas e níveis de abstração.

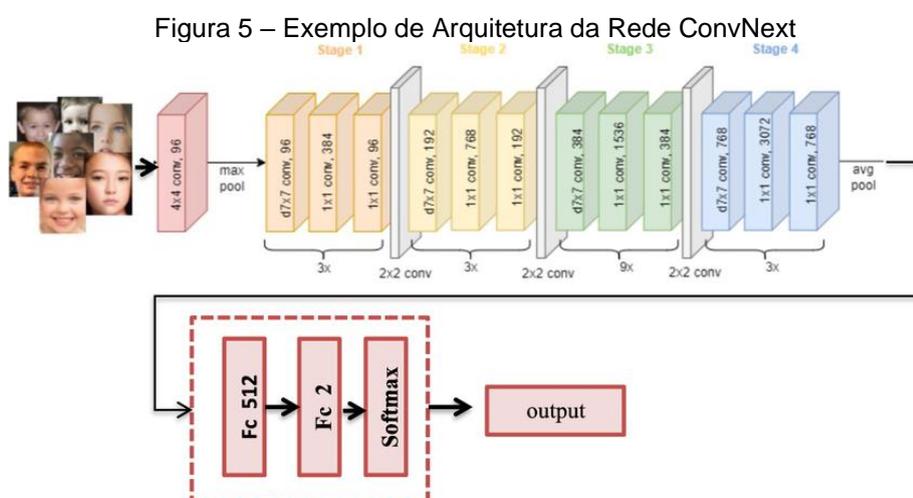
Attention Mechanisms: incorpora mecanismos de atenção em sua arquitetura para destacar regiões importantes nas imagens, melhorando assim a capacidade da rede de se concentrar em características relevantes.

Multi-scale Processing: a rede é projetada para processar informações em várias escalas, o que é essencial para tarefas de visão computacional onde os objetos de interesse podem ter diferentes tamanhos e aspectos.

Efficient Training: a ConvNetX é projetada para ser eficiente em termos de tempo de treinamento e uso de recursos computacionais. Isso é importante para permitir a implantação prática em sistemas com recursos limitados.

State-of-the-art Performance: a arquitetura ConvNetX visa alcançar ou superar o estado-da-arte em uma variedade de conjuntos de dados e tarefas de visão computacional, como classificação de imagens, detecção de objetos e segmentação semântica.

Em resumo, a ConvNetX é uma arquitetura de rede neural convolucional desenvolvida com o objetivo de superar as limitações das arquiteturas convolucionais tradicionais e melhorar o desempenho em uma variedade de tarefas de visão computacional, através do uso de blocos de integração de caminhos, mecanismos de atenção e processamento multi-escala. Sua arquitetura pode ser observada na figura 5:



Fonte: KANWAL, 2023

2.4.5 Comparação de Desempenho *ImageNet*

Como já mencionado, todos os modelos citados foram testados e classificados de acordo com sua performance no conjunto de imagens *ImageNet*. A *ImageNet* é um banco de dados de imagens criado para pesquisa em visão computacional e aprendizado de máquina. Ele contém milhões de imagens de alta resolução organizadas em milhares de categorias diferentes. A *ImageNet* é amplamente utilizada na comunidade de pesquisa em visão computacional como um conjunto de dados de referência para desenvolver e avaliar algoritmos de reconhecimento de objetos e classificação de imagens.

Nesta pesquisa, foram utilizados os pesos dos modelos treinados nos dados da *ImageNet* para ser realizado o aprendizado por transferência, de forma que as características fundamentais deste treinamento pudessem ser reaproveitadas

Um dos marcos mais importantes associados à *ImageNet* é o Desafio de Reconhecimento Visual da *ImageNet* (*ImageNet Large Scale Visual Recognition Challenge*), que ocorreu anualmente de 2010 a 2017 e ajudou a impulsionar avanços significativos na área de visão computacional, especialmente no desenvolvimento de redes neurais convolucionais profundas (CNNs).

Ao serem avaliados os modelos durante esta competição, sete métricas são comumente utilizadas para descrever seu desempenho: *top-1 accuracy*, *top-5 accuracy*, tamanho, parâmetros, profundidade, TPEICPU e TPEIGPU.

Top-1 accuracy: refere-se à porcentagem de vezes em que a classe verdadeira está na posição mais alta (ou seja, a previsão mais provável) feita pelo modelo. Por exemplo, se um modelo classificar uma imagem de um cachorro corretamente como um cachorro, sua *top-1 accuracy* é 100%.

Top-5 accuracy: refere-se à porcentagem de vezes em que a classe verdadeira está entre as cinco previsões mais prováveis feitas pelo modelo. Isso significa que, mesmo que a classe verdadeira não seja a previsão mais alta, contanto que esteja entre as cinco melhores previsões, será considerada correta. Isso é útil porque, em muitos casos, várias classes podem ser razoavelmente confundidas entre si. A *top-5 accuracy* dá uma medida mais ampla da capacidade do modelo de classificar corretamente a imagem.

Parâmetros referem-se ao número de parâmetros ajustáveis em um modelo de aprendizado de máquina, como uma rede neural convolucional (CNN). Mais parâmetros podem indicar um modelo mais complexo, capaz de aprender representações mais detalhadas dos dados, mas também pode levar a um maior risco

de sobreajuste e exigir mais dados para treinar efetivamente.

Profundidade, em redes neurais, se refere ao número de camadas que compõem a arquitetura da rede. Redes mais profundas geralmente têm mais capacidade de aprendizado, mas também podem ser mais difíceis de treinar e podem exigir mais recursos computacionais.

Tempo por etapa de inferência CPU (TPEICPU): Isso se refere ao tempo médio necessário para fazer uma única inferência (ou seja, previsão) usando o modelo, medido em milissegundos (ms) quando executado em uma unidade de processamento central (CPU). É uma medida importante para entender a eficiência do modelo em termos de tempo de execução.

Tempo por etapa de inferência GPU (TPEIGPU): similar ao tempo por etapa de inferência na CPU, mas medido quando o modelo é executado em uma unidade de processamento gráfico (GPU). Em muitos casos, o uso de GPUs pode acelerar significativamente a inferência em comparação com CPUs devido à sua capacidade de processamento paralelo.

Esses tempos são uma média de 30 batches e 10 repetições com as seguintes configurações mostradas na tabela 1:

Tabela 1 – Configurações *ImageNet*

CPU	AMD EPYC Processor (with IBPB) (92 core)
RAM	1.7T
GPU	Tesla A100
BATCH SIZE	32

Fonte: keras.io/api/applications/

Esses parâmetros são essenciais ao comparar modelos e para a análise de quais seriam os que mais teriam potencial de superar modelos antigos em tarefas em geral, já que promovem um aspecto de eficiência do quão bem ele se comporta em relação ao seu custo computacional.

A tabela 2 mostra como cada modelo que foi utilizado neste estudo performou na competição *ImageNet*:

Tabela 2 - Performances *ImageNet*

Modelo	Tamanho (MB)	Top-1 Accuracy	Top-5 Accuracy	Parâmetros (Milhões)	Profundidade	TPEICPU	TPEIGPU
VGG16	528	71.3%	94.5%	22.9	81	109.4	8.1
ResNet50	98	74.9%	92.1%	25.6	107	58.2	4.6
EfficientNet V2 B0	29	78.7%	94.3%	7.2	-	-	-

<i>ConvNext Tiny</i>	109	81.3%	-	28.6	-	-	-
----------------------	-----	-------	---	------	---	---	---

Fonte: <https://keras.io/api/applications/>

Pode-se analisar previamente que os modelos *EfficientNet V2 B0* e *ConvNext Tiny* possuem um grande potencial de performance, vista suas *Top-1 Accuracy* elevadas tendo um número menor ou similar de parâmetros aos modelos mais antigos.

2.5 Métricas Para a Avaliação de Desempenho

Após os treinamento dos modelos, foi feita uma avaliação de seu desempenho. Para isso, nesta seção, serão examinados métodos e métricas de avaliação de algoritmos utilizados.

2.5.1 Medição do Erro

As taxas de acerto e erro de um algoritmo estão relacionadas às porcentagens de vezes em que ele acerta ou erra ao classificar dados no aprendizado de máquina, focando em problemas de classificação discutidos neste trabalho.

A partir dos testes feitos, descobrimos as porcentagens de classificações corretas e incorretas, considerando todos os dados analisados (MONARD; BARANAUSKAS, 2003). A taxa de erro de um classificador k pode ser representada como $erro(k)$, que é calculada comparando o que o classificador diz (k) com o que realmente é (y) (MONARD; BARANAUSKAS, 2003). A acurácia então pode ser descrita conforme a equação 3:

$$acc(h) = 1 - erro(k) \quad (3)$$

2.5.2 Outras Medidas de Avaliação

Para avaliar o desempenho do modelo em diversos contextos, é comum utilizarmos outras métricas de avaliação onde apenas a acurácia sozinha não é capaz de avaliar propriamente o modelo (MONARD; BARANAUSKAS, 2003).

- **Matriz de Confusão**

A matriz de confusão é uma tabela que avalia o desempenho de um modelo de classificação. Ela compara as previsões feitas pelo modelo com os valores reais dos

dados. A matriz é uma representação bidimensional onde as linhas representam as classes reais e as colunas representam as classes previstas pelo modelo. Formalmente, a matriz de confusão pode ser definida como:

Tabela 3 – Tipos de Classificação de Erro

VP	FP
FN	VN

Fonte: O autor

Onde:

VP (Verdadeiros Positivos): Instâncias corretamente classificadas como positivas.

FP (Falsos Positivos): Instâncias incorretamente classificadas como positivas.

FN (Falsos Negativos): Instâncias incorretamente classificadas como negativas.

VN (Verdadeiros Negativos): Instâncias corretamente classificadas como negativas.

- **Acurácia**

A acurácia, como descrita previamente, é uma métrica que mede a precisão global de um modelo de classificação. Ela representa a proporção de predições corretas em relação ao número total de predições feitas pelo modelo. Formalmente, a acurácia pode ser calculada conforme a equação 4:

$$acurácia = \frac{VP + VN}{VP + FP + FN + VN} \quad (4)$$

- **Precisão**

A precisão é uma métrica que mede a proporção de verdadeiros positivos em relação ao total de predições positivas feitas pelo modelo. Formalmente, a precisão é definida como:

$$precisão = \frac{VP}{VP + FP} \quad (6)$$

- **Revocação ou Sensibilidade**

A sensibilidade é uma métrica que mede a proporção de verdadeiros positivos em relação ao total de instâncias que realmente são positivas. Formalmente, a sensibilidade é calculada pela fórmula:

$$revocação = \frac{VP}{VP + FN} \quad (7)$$

- **Pontuação F1**

Esta métrica combina a precisão e recall em uma só, fornecendo uma precisão equilibrada do desempenho do modelo em relação a previsões positivas e negativas. É calculada por:

$$F1 - Score = \frac{2 * precisão * revocação}{precisão + revocação} \quad (8)$$

2.6 Aprendizado por Transferência

O aprendizado por transferência é uma técnica central em aprendizado de máquina que envolve o reaproveitamento de modelos pré-treinados em tarefas relacionadas para melhorar o desempenho em novas tarefas ou conjuntos de dados (THRUN; PRATT, 2012). Essa abordagem é motivada pelo pressuposto de que representações aprendidas em um domínio podem ser úteis em outros domínios, permitindo a transferência de conhecimento para problemas semelhantes.

No contexto do aprendizado profundo, o aprendizado por transferência é comumente realizado utilizando modelos pré-treinados em grandes conjuntos de dados, como o *ImageNet*, que capturam características gerais de alto nível em imagens. Esses modelos pré-treinados são geralmente estruturados em uma arquitetura de rede neural convolucional (CNN) e são capazes de extrair características úteis, como bordas, texturas e formas, que são relevantes para uma variedade de tarefas de visão computacional, e podem ser úteis até mesmo em casos

onde as imagens utilizadas para o treinamento não possuem semelhança aparente com as imagens utilizadas para a previsão final.

Ao aplicar o aprendizado por transferência, os pesos do modelo pré-treinado são inicializados como ponto de partida e, em seguida, ajustados para se adequarem aos novos dados ou tarefa específica. Essa adaptação pode ser feita através de técnicas como ajuste fino (*fine-tuning*), onde os pesos do modelo são atualizados durante o treinamento com os novos dados, ou através da extração de características (*feature extraction*), onde as representações aprendidas pelo modelo pré-treinado são utilizadas como entrada para um novo classificador inserido ao final da arquitetura já treinada no modelo.

De forma geral, pode-se dizer que o aprendizado por transferência pode ser usado como um catalizador para treinamento de novos modelos sem a necessidade de uma grande quantidade de dados de treinamento iniciais, que geralmente são essenciais para um bom desempenho de redes neurais convolucionais treinadas “do zero” (THRUN; PRATT, 2012).

3. REVISÃO DA LITERATURA

A revisão sistemática da literatura é um processo fundamental na condução de pesquisas acadêmicas, permitindo uma análise abrangente e organizada do conhecimento existente sobre um determinado tópico. Este estudo adota diferentes procedimentos para revisar a literatura de forma sistemática (ARAÚJO et al., 2023). Em geral, o processo é composto por três etapas principais: (1) planejamento da revisão, (2) coleta e seleção de artigos, e (3) verificação dos resultados.

No contexto do tema deste trabalho, a revisão da literatura segue um roteiro previamente planejado. Inicialmente, são realizados estudos preliminares para identificar os principais temas e palavras-chave relacionados ao tópico de pesquisa. Essas palavras-chave são combinadas de maneira específica para orientar a busca por artigos relevantes e alinhados ao tema.

As seguintes palavras-chave foram selecionadas, combinadas e traduzidas para o inglês para pesquisa: (i) “Redes Neurais Convolucionais”, (ii) “Tomografia Computadorizada”, (iii) “Nódulo Pulmonar” e (iv) “Aprendizado por Transferência”.

Na segunda etapa, de coleta e seleção de artigos, critérios de inclusão são aplicados para filtrar os artigos encontrados. Apenas artigos em inglês são selecionados, e o intervalo de busca abrange um período específico de tempo para garantir a inclusão dos estudos mais relevantes, definido entre janeiro de 2000 e dezembro de 2023. Para pesquisa, foi utilizada a base de dados Web of Science, que agrega diversas bases de dados e oferece acesso a diversos textos de consulta. Após a aplicação deste filtro, um total de 68 artigos foram encontrados, dando início à parte de seleção de artigos.

Após a seleção dos artigos, uma análise é conduzida para determinar sua relevância e contribuição para a pesquisa. Este processo envolve a leitura completa dos artigos selecionados e a identificação daqueles estritamente relacionados ao problema de pesquisa em questão. O número final de artigos selecionados para análise é apresentado, juntamente com uma explicação das razões para a exclusão de artigos que não atenderam aos critérios de inclusão.

Esta etapa do processo se caracteriza como uma etapa de decisão, pois aqui serão excluídos os artigos que não forem estritamente relacionados ao tema de pesquisa. Primeiramente, foram lidos os títulos e resumos dos artigos encontrados, e para fins deste trabalho, foram filtrados para fora da amostra artigos que focam em

modelos generativos, outros tipos de exame, detecção de outras doenças, detecção de nódulos ao invés de classificação e também artigos que foquem na utilização e comparação de modelos híbridos. Após este filtro, apenas 16 dos 68 artigos foram mantidos pelo critério previamente descrito. Essa amostra foi então lida em sua integridade para análise final, nos restando um total de 12 artigos pertinentes.

3.1 Estudos Selecionados

Huang et al. (2020) introduziu uma abordagem inovadora no diagnóstico de nódulos pulmonares, utilizando uma combinação de Deep Transfer Convolutional Neural Network (DTCNN) e Extreme Learning Machine (ELM) para classificar nódulos benignos e malignos. Pela primeira vez, o DTCNN foi empregado para extrair características de alto nível dos nódulos pulmonares, previamente treinado com o conjunto de dados ImageNet. Em seguida, um classificador ELM foi desenvolvido para a classificação dos nódulos. Os dados para análise foram coletados do Lung Image Database Consortium - Image Database Resource Initiative (LIDC-IDRI) e de um conjunto de dados privado do First Affiliated Hospital of Guangzhou Medical University na China (FAH-GMU). Os resultados para o conjunto de dados LIDC-IDRI demonstraram uma precisão de 94,57%, sensibilidade de 93,69% e especificidade de 95,15%.

Nishio et al. (2018) desenvolveram um método de diagnóstico auxiliado por computador (CADx) para classificação entre nódulo benigno, câncer primário de pulmão e câncer de pulmão metastático e avaliaram: (i) a utilidade da rede neural convolucional profunda (DCNN) para CADx da classificação ternária, comparada com um método convencional (característica de imagem feita à mão mais aprendizado de máquina), (ii) a eficácia da transferência de aprendizado e (iii) o efeito do tamanho da imagem como entrada do DCNN. Entre 1240 pacientes de um banco de dados previamente construído, imagens de tomografia computadorizada e informações clínicas de 1236 pacientes foram incluídas. Para o método convencional, o CADx foi realizado usando padrão binário local de padrão uniforme invariante à rotação em três planos ortogonais com uma máquina de vetor de suporte. Para o método DCNN, o CADx foi avaliado usando a rede neural convolucional VGG-16 com e sem transferência de aprendizado, e a otimização de hiperparâmetros do método DCNN foi realizada por pesquisa aleatória. As melhores precisões de validação médias do CADx foram de 55,9%, 68,0% e 62,4% para o método convencional, o método DCNN

com transferência de aprendizado e o método DCNN sem transferência de aprendizado, respectivamente. Para o tamanho da imagem de 56, 112 e 224, a melhor precisão de validação média para o DCNN com transferência de aprendizado foi de 60,7%, 64,7% e 68,0%, respectivamente. O DCNN foi melhor do que o método convencional para o CADx, e a precisão do DCNN melhorou ao usar a transferência de aprendizado. Além disso, descobriu-se que tamanhos de imagem maiores como entradas para DCNN melhoraram a precisão da classificação de nódulos pulmonares.

Da Nóbrega et al. (2020) investigaram a classificação da malignidade de nódulos pulmonares em imagens de tomografia computadorizada torácica usando transferência de aprendizado e redes neurais convolucionais. O estudo explorou a performance de transferência de aprendizado profundo de imagens não médicas em tarefas de classificação de malignidade de nódulos pulmonares para aprimorar tais sistemas. Inicialmente, 1018 exames de tomografia computadorizada (CT) de tórax e anotações médicas do LIDC/IDRI foram processados. Várias redes neurais convolucionais (*VGG16*, *VGG19*, *MobileNet*, *Xception*, *InceptionV3*, *ResNet50*, *InceptionResNetV2*, *DenseNet169*, *DenseNet201*, *NASNetMobile* e *NASNetLarge*) foram construídas, treinadas no conjunto de dados ImageNet, convertidas em extratores de características e aplicadas nas imagens de nódulos do LIDC/IDRI. Em seguida, cada conjunto de características profundas foi submetido a validações cruzadas de 10 dobras com classificadores de Bayes ingênuos, perceptrons de múltiplas camadas, máquina de vetor de suporte (SVM), vizinhos K-mais próximos (KNN) e floresta aleatória. Finalmente, as métricas de avaliação precisão (ACC), área sob a curva (AUC), taxa verdadeira positiva (TPR), precisão positiva (PPV) e pontuação F1 de cada resultado médio de validação cruzada foram calculadas e comparadas. Os resultados mostraram que o extrator de características profundo baseado na *ResNet50* e o classificador SVM RBF alcançaram uma métrica AUC de 93,1%, uma TPR de 85,38%, uma ACC de 88,41%, uma PPV de 73,48% e uma pontuação F1 de 78,83%. Com base nesses resultados, a transferência de aprendizado profundo mostra ser uma estratégia relevante para extrair características representativas de imagens de TC de nódulos pulmonares.

Ali et al. (2020) propuseram redes neurais convolucionais de textura transferíveis para melhorar o desempenho da classificação de nódulos pulmonares em exames de TC. Uma Camada de Energia (EL) foi incorporada ao esquema, que extrai características de textura da camada de convolução. A inclusão de EL reduz o

número de parâmetros aprendíveis da rede, o que reduz ainda mais os requisitos de memória e a complexidade computacional. O modelo proposto possui apenas três camadas convolucionais e uma EL, em vez de camada de agrupamento. No geral, a arquitetura proposta de CNN compreende nove camadas para extração automática de características e classificação de candidatos a nódulos pulmonares como malignos ou benignos. Além disso, o modelo pré-treinado da CNN proposta também é usado para lidar com o problema de classificação de conjunto de dados menores usando transferência de aprendizado. Este trabalho foi avaliado nos conjuntos de dados publicamente disponíveis LIDC-IDRI e LUNGx Challenge através de diferentes métricas de avaliação, como precisão, especificidade, taxa de erro e AUC. O modelo proposto é treinado por validação cruzada de seis dobras e alcançou uma pontuação de precisão de $96,69\% \pm 0,72\%$ com uma taxa de erro de apenas $3,30\% \pm 0,72\%$. Enquanto isso, a AUC medida e o recall são de $99,11\% \pm 0,45\%$ e $97,19\% \pm 0,57\%$, respectivamente.

Apostolopoulos et al. (2021) focaram na classificação automática e confiável de SPNs em exames de TC extraídos de um sistema PET/CT. Para alcançar a tarefa mencionada, foi aplicado Aprendizado Profundo com Redes Neurais Convolucionais (CNN). Foram comparadas e avaliadas as estratégias de treinar arquiteturas específicas de CNN do zero e a estratégia de transferência de aprendizado, utilizando CNNs pré-treinadas de ponta. Para aumentar os conjuntos de treinamento, foi realizada a ampliação de dados. O banco de dados disponível publicamente de exames de TC, denominado Lung Image Database Consortium e Image Database Resource Initiative (LIDC-IDRI), também foi utilizado para expandir ainda mais o conjunto de treinamento e foi adicionado ao conjunto de dados PET/CT. Os resultados destacam a eficácia da transferência de aprendizado e da ampliação de dados para a tarefa de classificação de pequenos conjuntos de dados. A melhor precisão obtida no conjunto de dados PET/CT atingiu 94%, utilizando uma proposta de modificação de uma CNN de ponta, chamada VGG16, e aprimorando o conjunto de treinamento com o conjunto de dados LIDC-IDRI. Além disso, a proposta de modificação supera em termos de sensibilidade vários estudos semelhantes, que exploram os benefícios da transferência de aprendizado.

Zhang et al. (2019) implementaram recentemente uma rede neural convolucional (CNN) no diagnóstico auxiliado por computador (CAD) para a classificação de nódulos pulmonares de imagens de TC torácica. O maior desafio na

classificação de imagens médicas com a ajuda da CNN é a dificuldade de adquirir amostras suficientes, e o overfitting é um problema comum quando não há imagens suficientes para treinamento. A transferência de aprendizado tem se mostrado razoável para lidar com esses problemas com um valor de perda aceitável. O modelo clássico LeNet-5 é usado para classificar nódulos pulmonares de imagens de TC torácica, incluindo nódulos pulmonares benignos e malignos e diferentes malignidades dos nódulos malignos. As imagens de TC são obtidas do Lung Image Database Consortium e Image Database Resource Initiative (LIDC-IDRI), onde tanto a varredura de nódulos pulmonares quanto as anotações de nódulos são disponibilizadas. O modelo é implementado na validação cruzada de 10 pastas (CV) para testar a robustez do modelo de classificação treinado. O resultado demonstra que a transferência de aprendizado do LeNet-5 é boa para classificar nódulos pulmonares de imagens de TC torácica, e os valores médios de precisão do Top-1 são de 97,041% e 96,685%, respectivamente.

Li & Liu (2021) propuseram um método eficaz de classificação de nódulos pulmonares solitários malignos com base no Faster R-CNN melhorado e na estratégia de transferência de aprendizado. No estudo, foi proposta uma rede neural convolucional profunda para classificar as imagens de TC de câncer de pulmão em quatro categorias: adenocarcinoma pulmonar, carcinoma de células escamosas do pulmão, câncer de pulmão metastático e tipos normais de câncer de pulmão. Alguns tipos de imagens de TC de pulmão de alta resolução têm caracteres desnecessários, como um grande número de características de continuidade de alta densidade, alvos de nódulos pulmonares de tamanho pequeno, complexidade de fundo da imagem de TC, e assim por diante. A estratégia de pré-processamento de sub-bloco de imagem de TC foi usada para extrair características de nódulos para aprimoramento e aliviar os problemas mencionados. Os resultados experimentais mostraram que o sistema proposto foi eficaz em resolver problemas como alta taxa de falso-positivo e longo tempo de custo de classificação com base no método de detecção Faster R-CNN original. Além disso, a estratégia de transferência de aprendizado foi usada para melhorar a eficiência da classificação para evitar o problema de overfitting causado por alguns poucos exemplos rotulados de conjuntos de dados de câncer de pulmão. Os resultados da classificação das imagens de TC de pulmão mostraram que o método proposto teve uma precisão de detecção média de 89,7% e reduziu a taxa de erro para atender às necessidades clínicas.

Zheng et al. (2020) propuseram uma rede neural convolucional profunda (CNN) com módulo de transferência de escala (STM) e operação de fusão de múltiplas características, chamada STM-Net, para a classificação do risco de nódulos pulmonares em adenocarcinoma. O STM-Net foi treinado e testado para a previsão do estágio de risco. A precisão e a área média sob a curva ROC alcançadas pelo método foram de 95,455% e 0,987, respectivamente, para o conjunto de dados ZSDB. Os resultados experimentais mostram que o STM-Net aumenta significativamente a precisão da classificação sobre abordagens semelhantes do estado-da-arte. O método proposto será um auxiliar eficaz para ajudar os médicos no diagnóstico da classificação de risco de nódulos pulmonares em adenocarcinoma em estágio inicial.

Atiya et al. (2023) desenvolveram um método de classificação baseado em CNN duplo-estado utilizando aprendizado de transferência para a detecção e classificação de tipos de câncer de pulmão, como adenocarcinoma, carcinoma de células escamosas e carcinoma de pequenas células, em imagens de TC torácica. Durante o treinamento, o modelo *ResNet50* atingiu uma precisão de 94% usando aprendizado de transferência de duplo estado, enquanto durante a validação e teste, alcançou 92,57% e 96,12% de precisão, respectivamente. O modelo DSTL baseado em CNNs profundas também superou modelos do estado-da-arte em tarefas de classificação. Em resumo, a precisão e a eficácia na triagem e detecção de câncer de pulmão podem ser melhoradas utilizando métodos de aprendizado de transferência de duplo estado e abordagens baseadas em CNN profundas.

Yadlapalli et al. (2022) propuseram um método de classificação inteligente de malignidades pulmonares utilizando técnicas de aprendizado profundo. Eles utilizaram redes neurais convolucionais (CNNs) para a classificação de nódulos pulmonares em imagens de TC. O modelo *VGG16* usando aprendizado de transferência maximizou a precisão da classificação em comparação com *VGG19*, *MobileNet* e *DenseNet169*. A precisão do *VGG16* foi de 91,28%, seguido por *VGG19* com 89,39%, *MobileNet* com 85,60% e *DenseNet169* com 83,71%.

Sakshiwala & Singh (2023) propuseram uma nova arquitetura de rede neural convolucional em várias escalas para a classificação de nódulos pulmonares benignos e malignos em imagens de TC. Eles utilizaram a estratégia de aprendizado de transferência para inicializar os pesos da arquitetura em várias escalas. O método proposto alcançou uma precisão de 93,88%, sensibilidade de 93,36% e especificidade de 93,26% na classificação da malignidade do nódulo.

Shi et al. (2019) propuseram um método de redução de falso positivo para a detecção de nódulos pulmonares em fatias de TC usando uma abordagem de aprendizado de transferência baseada em CNNs profundas. O método proposto alcançou uma sensibilidade geral de 87,2% com 0,39 FPs por varredura, o que é maior do que 85,4% com 4 FPs por varredura obtidos por outro método do estado-da-arte.

A aplicação de técnicas avançadas de aprendizado de máquina, especificamente o uso de redes neurais convolucionais (CNNs) e transferência de aprendizado, tem mostrado um grande potencial no desenvolvimento de sistemas de diagnóstico auxiliado por computador (CAD) para a classificação de nódulos pulmonares em imagens de TC.

Esta tecnologia, porém, está em constante desenvolvimento, e cada vez mais modelos surgem com capacidade de realizar previsões mais assertivas com uma quantidade cada vez menor de dados, como é o caso da *EfficientNetV2*.

Pode-se perceber, por exemplo, que nos estudos selecionados, modelos como *VGG16* e *ResNet50* são mencionados com frequência, o que reforça sua confiabilidade de performance através da soma de novos estudos que as utilizam, mas também deixam uma lacuna quanto a comparação dessas arquiteturas com novas arquiteturas de fácil acesso e implementação.

Neste trabalho serão investigados alguns desses novos modelos que não possuem tanta abrangência de estudos que avaliam seu desempenho no cenário de previsão em tomografias computadorizadas, como os modelos *EfficientNet* e *ConvNext*, contribuindo assim para o desenvolvimento deste campo.

4. CONSTRUÇÃO DOS MODELOS

Nesta seção, serão detalhados os recursos e processos utilizados para a modelagem e realização das comparações entre arquiteturas de convolução.

4.1 Programas e Bibliotecas

Para a programação, a linguagem escolhida foi a linguagem Python devido à sua flexibilidade, facilidade de uso e vasta comunidade de desenvolvedores. O Python possui uma ampla gama de bibliotecas disponíveis para aprendizado de máquina, o que facilita a implementação e experimentação de diferentes algoritmos e arquiteturas de redes neurais. Sua sintaxe simples e legibilidade também tornam o Python uma escolha popular entre os cientistas de dados e pesquisadores em inteligência artificial.

Na etapa de modelagem, Foi feito o uso de 4 bibliotecas no total, sendo elas: *TensorFlow*, *Scikit-Learn*, *Matplotlib* e *Keras*.

A biblioteca TensorFlow é amplamente reconhecida por sua capacidade de construir e treinar modelos de redes neurais de forma eficiente, sendo especialmente útil para tarefas de aprendizado profundo. Ela oferece uma variedade de ferramentas e APIs flexíveis para a implementação de redes neurais em uma variedade de cenários, desde classificação de imagens até processamento de linguagem natural. Além disso, o TensorFlow é desenvolvido e mantido pelo Google, o que garante suporte contínuo e atualizações frequentes.

A biblioteca Scikit-Learn é uma ferramenta fundamental para tarefas de aprendizado de máquina clássico, oferecendo uma variedade de algoritmos de classificação, regressão, clustering e pré-processamento de dados. Sua interface consistente e fácil de usar torna a construção e avaliação de modelos de aprendizado de máquina mais acessível para os desenvolvedores. Neste contexto específico, foi utilizada para análise das métricas e resultados dos modelos.

Matplotlib é uma biblioteca poderosa para visualização de dados em Python, fornecendo uma grande variedade de gráficos e ferramentas de plotagem para análise exploratória de dados, apresentação de resultados e comunicação de informações de forma clara e eficaz.

Keras é uma API de alto nível para construção e treinamento de redes neurais, projetada para ser simples e intuitiva, enquanto mantém a flexibilidade e o desempenho necessários para tarefas complexas de aprendizado profundo. Keras é

integrado ao TensorFlow e fornece uma interface amigável para criar e experimentar com diferentes arquiteturas de rede neural, além de possibilitar o uso de técnicas avançadas para melhora do desempenho de modelo através de ferramentas integradas de tuning de hiperparâmetros.

4.2 Modelagem e Análise

Para investigar os resultados relacionados à questão abordada, conduziu-se uma análise que combinou aspectos qualitativos e quantitativos da execução de quatro modelos de aprendizado de máquina na análise de imagens, utilizando as arquiteturas *VGG16*, *ResNet50*, *EfficientNetV2B0* e *ConvNext Tiny*. Os parâmetros considerados para avaliação foram: acurácia, precisão, sensibilidade e pontuação F1.

Para a modelagem, optou-se por empregar o ambiente interativo *Jupyter*, uma plataforma flexível e amplamente utilizada que permite a criação, execução e documentação de código de forma interativa, facilitando a exploração dos dados e o desenvolvimento dos modelos. A escolha do Jupyter proporcionou um ambiente ágil, possibilitando a experimentação rápida e a visualização imediata dos resultados obtidos.

Além disso, para otimizar o desempenho computacional durante o treinamento dos modelos, recorreu-se à tecnologia CUDA, uma plataforma de computação paralela desenvolvida pela Nvidia. A utilização do CUDA permitiu a execução de operações intensivas em computação de forma eficiente na GPU, acelerando significativamente o processo de treinamento dos modelos e reduzindo o tempo necessário para alcançar resultados satisfatórios.

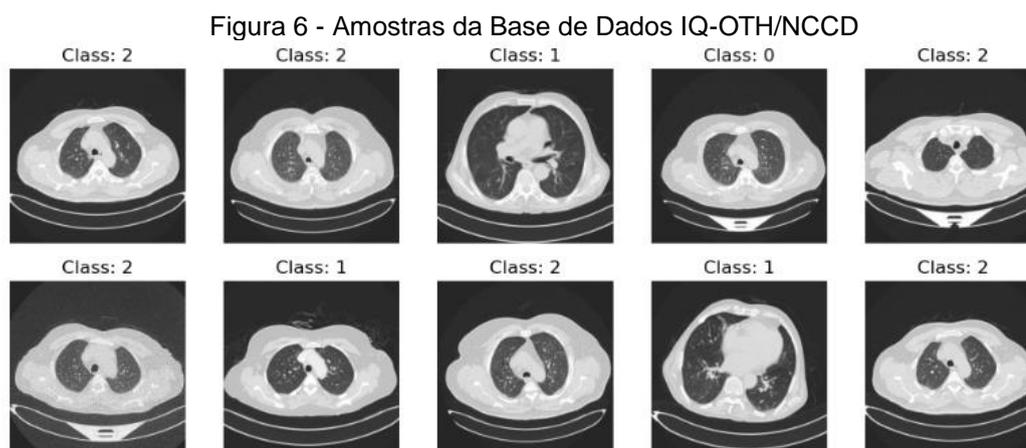
4.3 Base de Dados Utilizada

Para o treinamento dos modelos, foi utilizada a base de dados do Hospital de Ensino de Oncologia do Iraque/Centro Nacional de Doenças Cancerígenas (IQ-OTH/NCCD), que pode ser acessada através da plataforma online Kaggle. Essa base foi coletada pelos hospitais mencionados durante três meses, em 2019, e possui um total de 1190 imagens de cortes de tomografias computadorizadas, de um total de 110 casos, agrupados nas categorias de “benigno”, com 15 casos, “maligno”, com 40 casos, e “normal”, com 55 casos.

Cada TC contém vários cortes, variando de 80 a 200 cortes, cada um

representando uma imagem do tórax humano com diferentes lados e ângulos. Os 110 casos variam em gênero, idade, escolaridade, área de residência e estado civil.

Abaixo são mostradas alguns exemplos de imagens que serão usadas, categorizadas nas classes 0, 1 e 2 para representar os casos benignos, que são casos com a presença de tumores que não apresentam riscos à saúde; malignos, que apresentam risco e normais, que não possuem nódulos:



Fonte: Base de Dados IQ-OTH/NCCD

4.4 Tratamento de Imagens

No âmbito do desenvolvimento de modelos de Deep Learning, é fundamental compreender e aplicar técnicas adequadas de pré-processamento de imagens para maximizar o desempenho. O pré-processamento de imagens desempenha um papel crucial na preparação dos dados de entrada para os modelos de Deep Learning, influenciando diretamente sua capacidade de aprendizado e generalização.

Para garantir uma abordagem consistente e eficaz, é necessário realizar uma série de etapas de pré-processamento em todas as imagens de entrada. Uma prática comum é redimensionar as imagens para uma dimensão específica, neste caso, 224x224 pixels, usado no treinamento com a base *ImageNet*, para garantir que todas as entradas tenham o mesmo tamanho e formato. Isso facilita o processamento e reduz a carga computacional durante o treinamento e inferência do modelo.

Além disso, dependendo do modelo utilizado, podem ser aplicadas outras transformações às imagens de entrada. Por exemplo, para os modelos *VGG16*, as

imagens são convertidas do padrão RGB para BGR (*Blue-Green-Red*), e em seguida, cada canal de cor é centrado em zero em relação aos valores médios do conjunto de dados *ImageNet*, sem escala adicional. Essa etapa de pré-processamento é crucial para alinhar as características das imagens de entrada com as expectativas do modelo, garantindo resultados mais precisos e consistentes.

Para o modelo *EfficientNetV2*, espera-se que as entradas sejam tensores de pixels em ponto flutuante com valores na faixa de 0 a 255. Isso significa que as imagens de entrada devem ser representadas como tensores de pixels, onde cada pixel é um valor de ponto flutuante na faixa de 0 a 255. Essa faixa representa os níveis de intensidade de cada pixel na imagem.

Para o modelo *ConvNetX*, as transformações específicas de pré-processamento podem variar dependendo da implementação e das especificações do modelo.

Para facilitar esses procedimentos de pré-processamento, a biblioteca Keras oferece uma função nativa chamada *ImageDataGenerator*. Essa função automatiza o processamento de imagens, permitindo que os desenvolvedores especifiquem facilmente uma variedade de transformações, como redimensionamento, rotação, espelhamento, normalização e muito mais. Isso simplifica o fluxo de trabalho de desenvolvimento e permite uma implementação coesa e eficiente de modelos de Deep Learning para tarefas de visão computacional.

4.5 Balanceamento de Dados

Além das técnicas de pré-processamento de imagens, outro aspecto importante a considerar no desenvolvimento de modelos de *Deep Learning* para visão computacional é o balanceamento de dados. O desequilíbrio de classes, onde uma ou mais classes têm um número significativamente menor de exemplos do que outras, pode levar a um viés no modelo e impactar negativamente sua capacidade de generalização.

Ao analisar a distribuição das classes dentro do conjunto de dados, observamos uma disparidade marcante no número de imagens entre as categorias de exames. Especificamente, a categoria de nódulos benignos possui apenas 120 imagens, enquanto a categoria de nódulos malignos conta com 561 imagens e a categoria de imagens normais possui 416 exemplos. Essa disparidade na distribuição das classes pode ter um impacto adverso no treinamento do modelo, levando a uma

tendência de classificar erroneamente as classes majoritárias em detrimento das minoritárias.

Uma abordagem comum para lidar com o desequilíbrio de classes é o uso de técnicas de balanceamento de dados, como a sobre-amostragem (*over-sampling*) ou sub-amostragem (*under-sampling*). No contexto da sobre-amostragem, que foi utilizada nesta pesquisa, uma das técnicas mais populares é o *SMOTE* (*Synthetic Minority Over-sampling Technique*), que gera exemplos sintéticos para as classes minoritárias, aumentando assim sua representação no conjunto de dados.

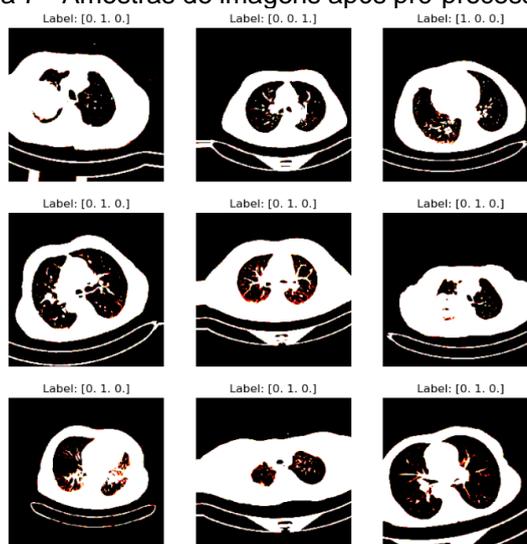
O *SMOTE* é particularmente útil quando as classes minoritárias são insuficientes para fornecer informações significativas ao modelo. Ele funciona criando exemplos sintéticos interpolando os recursos de exemplos semelhantes das classes minoritárias, ajudando a equilibrar a distribuição das classes no conjunto de dados.

Para implementar o *SMOTE* em *Python*, utiliza-se a biblioteca *Keras* com a função *ImageDataGenerator*, que fornece uma implementação facilitada da técnica. Com o auxílio desta biblioteca, podemos aplicar o *SMOTE* aos nossos dados de treinamento antes de alimentá-los ao modelo de *Deep Learning*. Isso ajuda a melhorar a capacidade do modelo de aprender com todas as classes de maneira equilibrada e, assim, melhorar seu desempenho na tarefa de classificação de imagens.

Esta função é responsável por gerar em tempo real variações das imagens originais da base de dados em *batches*, mudando suas características originais de acordo com parâmetros selecionados, sendo eles: rotação da imagem em até 10 graus, deslocamento horizontal e vertical em até 10% da largura da imagem, zoom de até 10% e cisalhamento em até 10%.

Abaixo são mostrados exemplos das imagens utilizadas no treinamento após a etapa de pré-processamento e aumento de dados:

Figura 7 - Amostras de imagens após pré-processamento



Fonte: O autor

4.6 Divisão em Treino, Teste e Validação

O processo de divisão dos dados em conjuntos de treinamento, validação e teste desempenha um papel fundamental no desenvolvimento e avaliação de modelos de aprendizado de máquina. Cada conjunto de dados tem uma função específica e contribui de maneira única para o processo de treinamento e avaliação do modelo.

O conjunto de dados de treinamento é utilizado para ajustar os parâmetros do modelo durante o processo de treinamento. Os exemplos contidos neste conjunto são apresentados ao modelo repetidamente, permitindo que ele aprenda a mapear os dados de entrada para as saídas desejadas.

O conjunto de dados de validação é usado para ajustar os hiper parâmetros do modelo e monitorar o desempenho durante o treinamento. Isso envolve a seleção de valores ótimos para parâmetros como taxa de aprendizado, tamanho do lote e arquitetura da rede neural. Ao avaliar o desempenho do modelo em um conjunto de dados separado do conjunto de treinamento, pode-se detectar e corrigir problemas de sobreajuste (*overfitting*) ou sub-ajuste (*underfitting*) antes que o modelo seja aplicado aos dados de teste.

Por fim, o conjunto de dados de teste é usado para avaliar o desempenho final do modelo após o treinamento e validação. Este conjunto de dados é mantido completamente separado dos conjuntos de treinamento e validação e é usado apenas uma vez, após o treinamento do modelo estar completo. Ao avaliar o desempenho do modelo em dados "nunca vistos", é possível obter uma estimativa precisa de como ele

se sairá em situações do mundo real.

Nesta pesquisa, foi utilizada a proporção de 80%, 20% para treino e teste, respectivamente, mantendo a proporção para cada categoria de dados. Durante o treino, o parâmetro de validação foi definido com 0.2, resultando em um conjunto aleatório de imagens selecionadas do conjunto de treino em cada época, servindo então como conjunto de validação durante o treinamento.

4.7 Treinamento e Fine Tuning dos Modelos

Para cada modelo, inicializa-se a instância utilizando a biblioteca *Keras* e especificamos a utilização dos pesos previamente treinados na competição *ImageNet*, visando a alavancagem do aprendizado prévio no contexto do nosso modelo atual. Em seguida, congelam-se todas as camadas do modelo carregado para preservar o conhecimento já adquirido. No entanto, é importante destacar que o modelo possui uma camada final configurada para classificar as 1000 categorias originais do conjunto de dados *ImageNet*, o que nos conduz à próxima etapa do processo. Nessa etapa, remove-se esta última camada de classificação para 1000 categorias e adicionamos duas camadas finais de treinamento: uma responsável pela adaptação aos novos dados de tomografia computadorizada, densamente conectada com 128 neurônios, e desativação randômica de 20% como técnica de regularização, e outra camada de saída com ativação *softmax*, destinada a classificar cada imagem em uma das três categorias propostas.

Posteriormente, o modelo é compilado utilizando o otimizador Adam com uma taxa de aprendizado de 0.0001, a função de perda de entropia cruzada categórica, tamanho do lote de imagens de 32 e a métrica de acurácia como objetivo. O ajuste do modelo é realizado por até 50 épocas de treinamento, com uma estratégia de interrupção precoce implementada para evitar o sobreajuste. Essa estratégia monitora a função de perda dos dados de validação, diminuindo o valor da taxa de aprendizado por um fator de 0.3 quando não há melhora na perda em até 5 épocas, além disso, o modelo foi configurado para ser interrompido caso não houver melhoria de perda nos dados de validação após dez épocas consecutivas.

A tabela abaixo sintetiza a configuração dos hiperparâmetros para os modelos iniciais com todas as camadas de base congeladas:

Tabela 4 - Hiperpâmetros dos modelos-base

	Valor
Formato do input	(224,224,3)
Taxa de aprendizado	0.0001
Função de perda	Entropia cruzada categórica
Métrica Objetivo	Acurácia
Épocas	50
Tamanho do lote	32
Paciência de interrupção	10
Paciência de diminuição de aprendizado	5
Fator de diminuição	0.3
Divisão de validação	0.2

Fonte: O autor

Após a conclusão do treinamento inicial, o modelo treinado é submetido a um processo adicional chamado *fine-tuning*, no qual camadas adicionais do modelo pré-treinado são ativadas para um novo treinamento, visando aprimorar os resultados obtidos anteriormente. No entanto, é necessário realizar essa etapa com cautela para evitar o sobreajuste. Para mitigar esse risco, reduzimos significativamente a taxa de aprendizado por um fator de 10, assumindo um novo valor de 0.00001 e limitando a capacidade dessa etapa de alterar drasticamente os pesos já treinados.

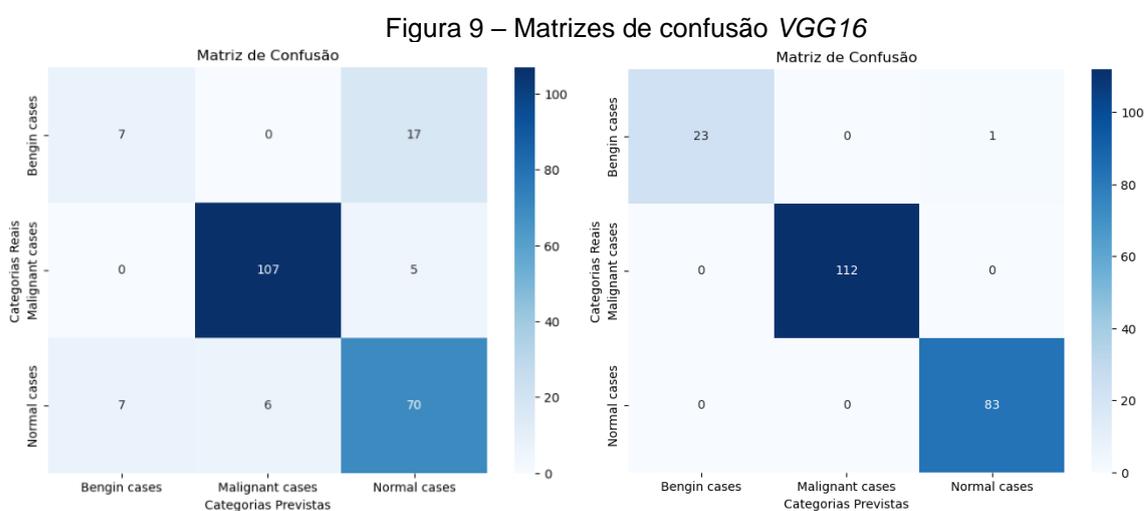
Este processo é realizado para cada um dos modelos propostos, e sua performance é analisada com as métricas previamente mencionadas na seção 2.3.

5. RESULTADOS

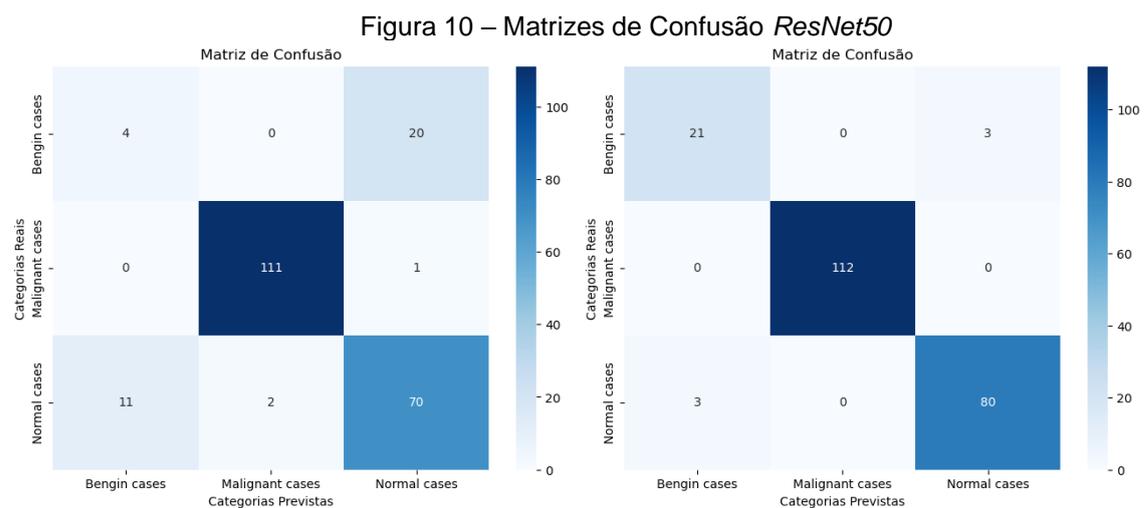
Na seção de resultados abaixo, são comparadas as etapas de treinamento inicial apenas com as camadas extras inseridas no modelo final, com os resultados da etapa posterior de *fine-tuning*.

5.1 Matrizes de Confusão

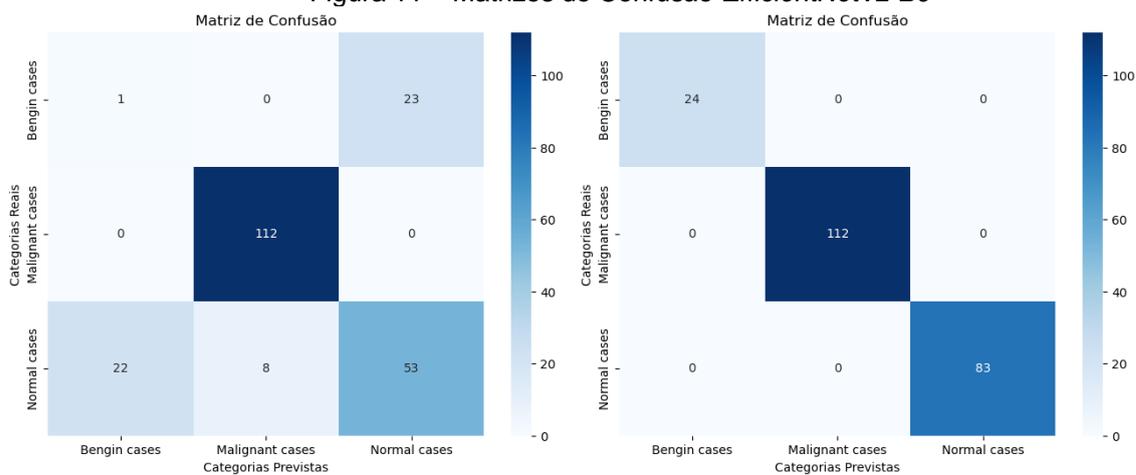
Abaixo estão dispostas as matrizes de confusão resultantes dos treinamentos antes do *fine-tuning* (à esquerda) e após *fine-tuning* (à direita):



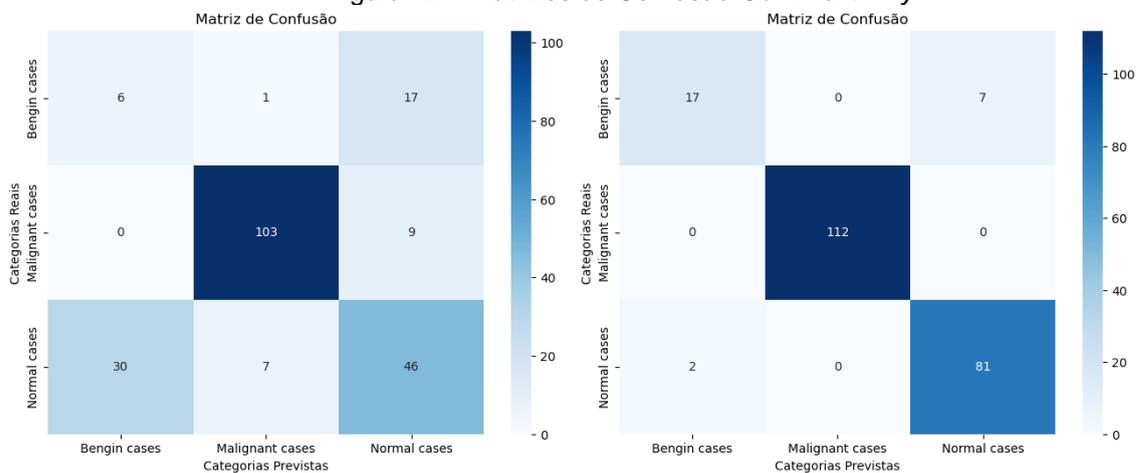
Fonte: O autor



Fonte: O autor

Figura 11 – Matrizes de Confusão *EfficientNetV2 B0*

Fonte: O autor

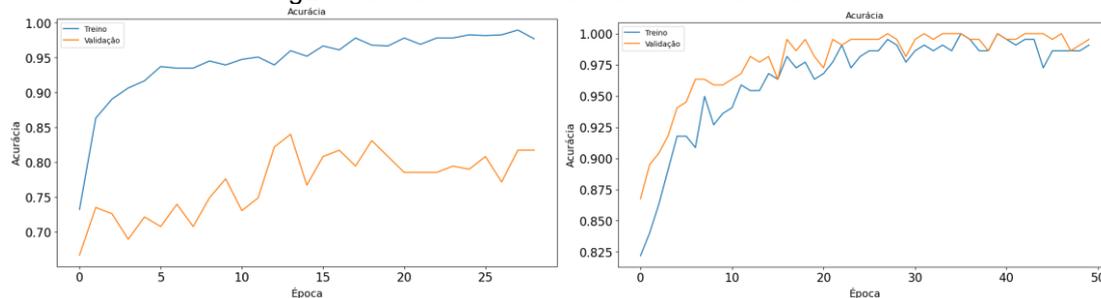
Figura 12 – Matrizes de Confusão *ConvNext Tiny*

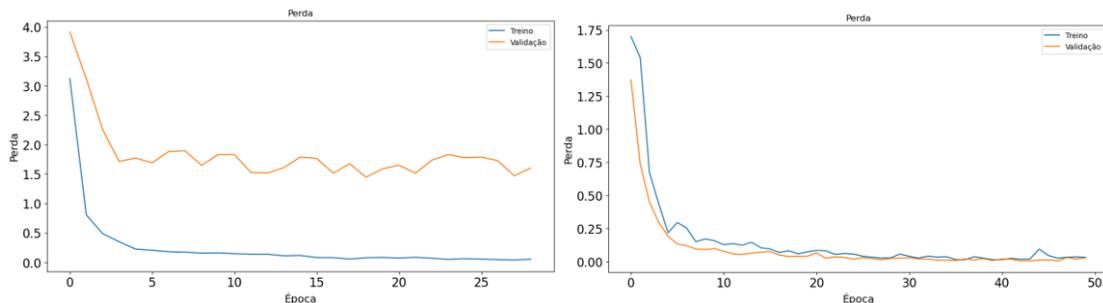
Fonte: O autor

5.2 Gráficos de Acurácia e Perda ao Longo das Épocas

- **VGG16**

Figura 13 – Gráficos de Acurácia e Perda VGG16

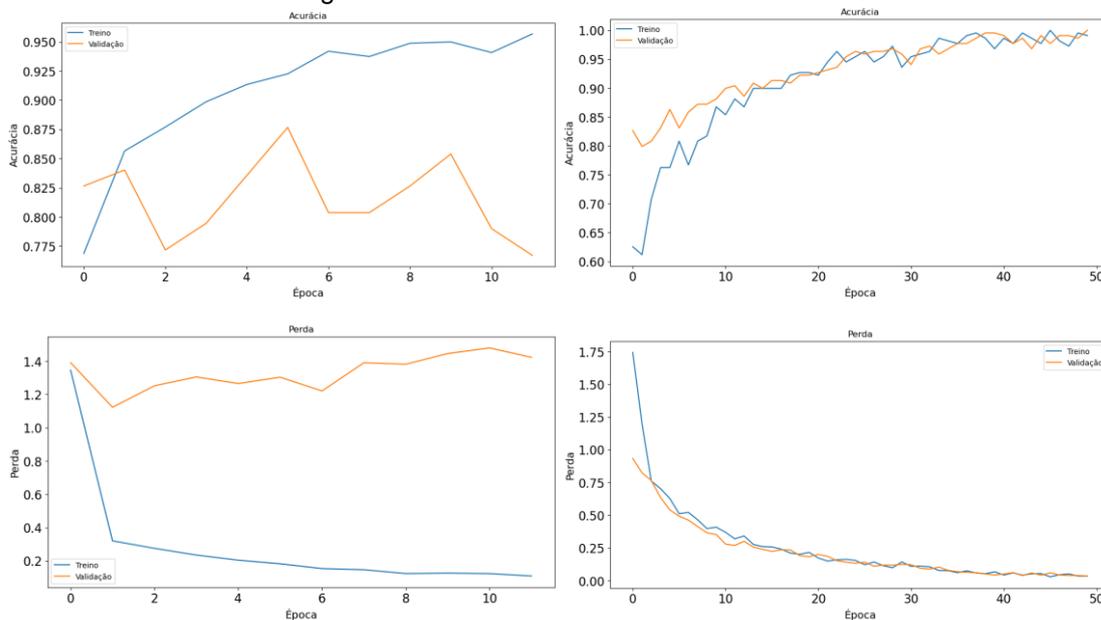




Fonte: O autor

- ResNet50**

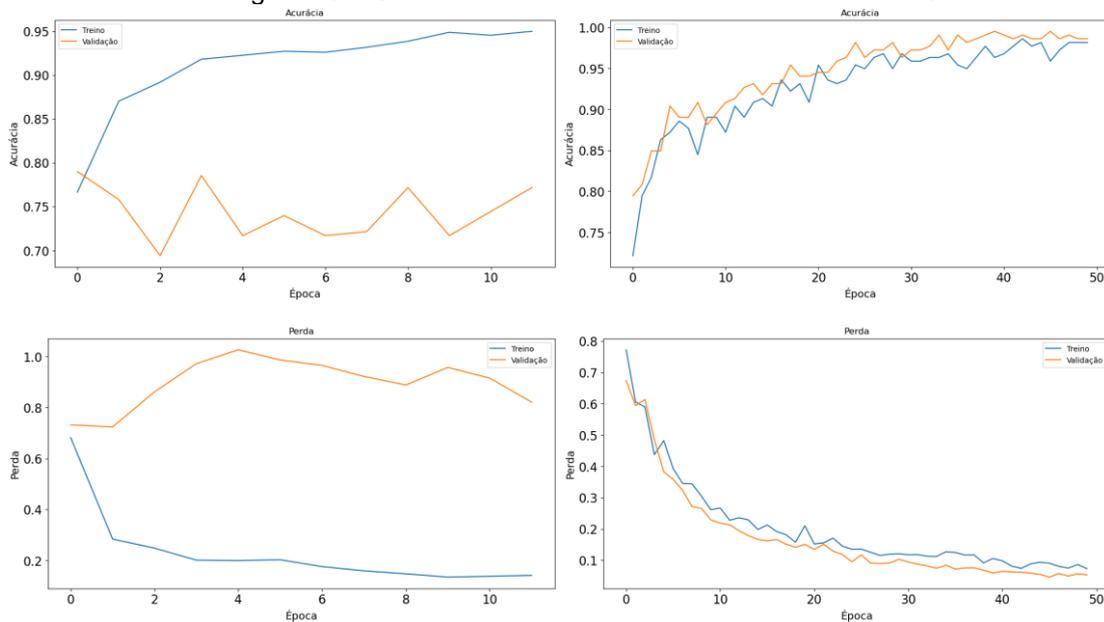
Figura 14 – Gráficos de Acurácia e Perda *ResNet50*



Fonte: O autor

- EfficientNetV2 B0**

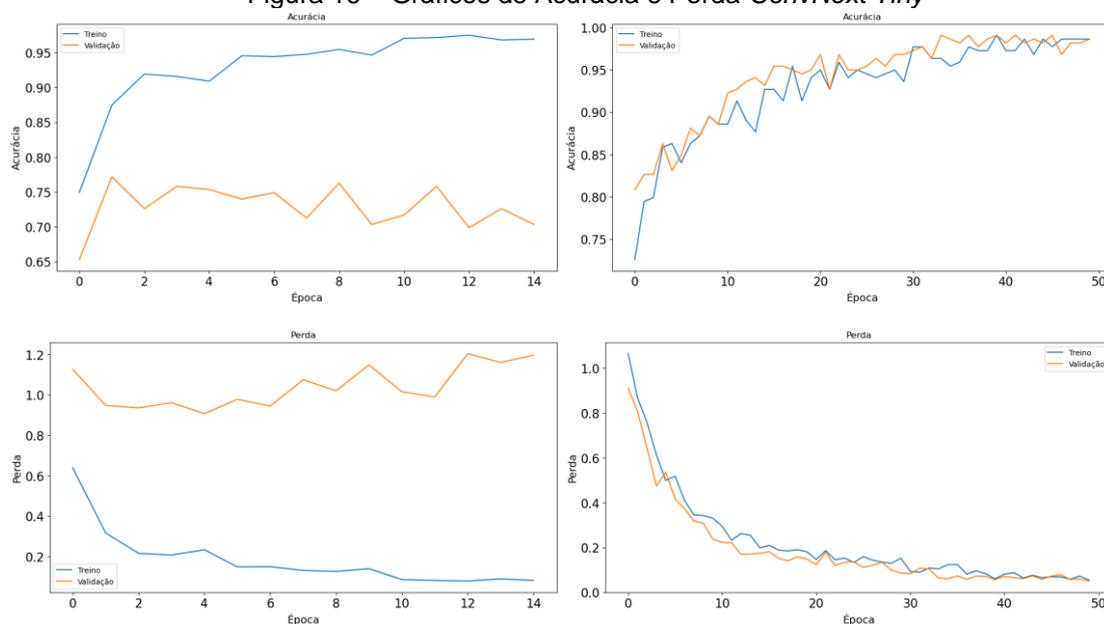
Figura 15 – Gráficos de Acurácia e Perda *EfficientNetV2 B0*



Fonte: O autor

- **ConvNext Tiny**

Figura 16 – Gráficos de Acurácia e Perda *ConvNext Tiny*



Fonte: O autor

5.3 Resumo de Desempenhos

Na tabela abaixo é possível averiguar o desempenho de cada modelo, antes e depois da etapa de *fine-tuning*.

Tabela 5 – Resumo de Desempenho dos Modelos

	Pré Fine-Tuning				Pós Fine-Tuning				Tempo Total de Treino
	Acurácia	Precisão	Revocação	F1-score	Acurácia	Precisão	Revocação	F1-score	
<i>VGG16</i>	0.84	0.83	0.84	0.83	1.00	1.00	1.00	1.00	769s
<i>ResNet50</i>	0.84	0.82	0.84	0.83	0.97	0.97	0.97	0.97	585s
<i>EfficientNetV2 B0</i>	0.76	0.75	0.76	0.75	1.00	1.00	1.00	1.00	521s
<i>ConvNext Tiny</i>	0.71	0.73	0.71	0.72	0.96	0.96	0.96	0.96	510s

Fonte: O autor

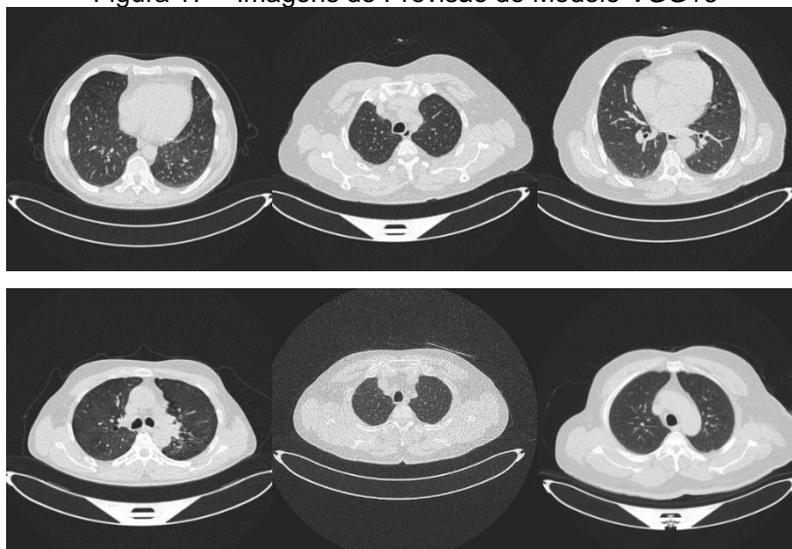
5.4 Análise de Resultados

A análise dos resultados obtidos ao treinar e avaliar os modelos *VGG16*,

ResNet50, *EfficientNetV2 B0* e *ConvNext Tiny* para a tarefa específica de classificação de nódulos pulmonares em imagens de tomografia computadorizada revela uma série de *insights* valiosos para a aplicação prática desses modelos em cenários médicos.

Começando pela análise das métricas de desempenho, observamos que todos os modelos alcançaram resultados satisfatórios após o processo de *fine-tuning*, com altas pontuações de acurácia, precisão, revocação e F1-score. No entanto, é importante notar que, antes do *fine-tuning*, os modelos apresentaram dificuldades em classificar corretamente as amostras da classe benigna, frequentemente rotulando-as erroneamente como "normal". Essa tendência pode ser atribuída à complexidade da tarefa de distinção entre esse tipo de tumor e imagens onde o paciente se encontra normal, sendo então necessário a etapa de *fine-tuning* para detectar características mais sutis desses dados. Na primeira linha da figura 17 estão alguns exemplos de amostras da categoria "Normal" que foram classificadas como nódulos benignos em comparação a reais benignos, e na segunda linha, casos corretamente classificados como "Normais":

Figura 17 – Imagens de Previsão do Modelo VGG16



Fonte: Base de Dados IQ-OTH/NCCD

Em primeira vista, não é possível detectar quais características específicas estão sendo detectadas pelo modelo para sua categorização errônea antes da etapa de *fine-tuning*. Porém, após passarem por ela, todos os modelos indicaram melhoria significativa no desempenho, mostrando o quão essencial é sua realização para detecção e aprendizado de características minuciosas que podem não ser percebidas

com facilidade.

Ambos os modelos mais antigos, *VGG16* e *ResNet50*, demonstraram desempenho semelhante após o *fine-tuning*, alcançando altas pontuações em todas as métricas avaliadas. No entanto, o *ResNet50* apresenta uma arquitetura mais profunda e uma menor quantidade de parâmetros em comparação com o *VGG16*, o que sugere uma maior eficiência computacional.

O modelo *EfficientNetV2 B0*, embora tenha apresentado um desempenho ligeiramente inferior antes do *fine-tuning*, alcançou resultados excelentes após o processo, com pontuações perfeitas em todas as métricas. Além disso, o *EfficientNetV2 B0* se destaca pela sua arquitetura eficiente, com um número significativamente menor de parâmetros em comparação com os modelos mais antigos, apresentando também os melhores resultados em relação ao tempo total que levou para ser treinado. Isso o torna uma escolha atraente para aplicações onde recursos computacionais são limitados.

O modelo *ConvNext Tiny* apresentou desempenho ligeiramente inferior em comparação com os outros modelos, tanto antes quanto após o *fine-tuning*, tendo também um tamanho em parâmetros, comparável ao modelo *VGG16*. Isso pode indicar que, para esta tarefa específica, os modelos anteriores parecem mais atraentes para implementações posteriores.

Além disso, os modelos em geral apresentaram desempenhos equivalentes ou até superiores aos estudados durante a revisão de literatura, também se alinhando a métricas de observações empíricas em outras instâncias não acadêmicas, o que reforça a confiabilidade de seus desempenhos.

Considerando os resultados obtidos, os sistemas hospitalares podem se beneficiar significativamente ao utilizar modelos mais recentes e eficientes, como o *EfficientNetV2 B0*, para esta tarefa específica de classificação. Esse modelo apresentou uma combinação única de desempenho superior e eficiência, permitindo diagnósticos mais precisos com sob as mesmas condições de dados necessários para o treino, o que pode levar a uma melhoria significativa na qualidade do atendimento ao paciente e na eficiência operacional.

Em suma, embora todos os modelos tenham demonstrado capacidade de aprendizado e desempenho satisfatório após o *fine-tuning*, o modelo *EfficientNetV2 B0*, mostra-se particularmente promissor devido à sua eficiência computacional e desempenho superior. Essas características o torna uma opção ideal para

implementação em sistemas hospitalares, onde a precisão e a eficiência são fundamentais para o seu sucesso.

6. CONCLUSÃO

Neste trabalho, destaca-se a importância de avaliar novos modelos de aprendizado de forma contínua para situações específicas, e demonstra-se que essa prática é capaz de contribuir de forma significativa, tanto em questão de avaliação de performance desses modelos, quanto em sua contribuição como referência para trabalhos futuros.

Aqui foi ilustrado claramente como modelos mais recentes, como o *EfficientNetV2 B0*, podem oferecer benefícios significativos em termos de desempenho e eficiência em comparação com modelos mais antigos e consolidados, como o *VGG16* e o *ResNet50*.

Ao realizar estudos comparativos como este, não apenas avançamos no desenvolvimento de métodos mais eficazes para diagnóstico e tratamento médico, mas também são fornecidos insights valiosos para pesquisas futuras. A compreensão das vantagens e desvantagens de diferentes arquiteturas de modelo nos permite refinar abordagens e direcionar esforços para áreas onde há maior potencial de impacto.

Para profissionais de saúde, os resultados desta pesquisa têm implicações diretas na prática clínica. Ao adotar modelos mais eficientes e precisos para auxiliar no diagnóstico e na tomada de decisões, os profissionais podem oferecer um atendimento mais rápido e preciso aos pacientes. Isso não apenas melhora a qualidade dos cuidados de saúde, mas também pode levar a diagnósticos mais precoces e tratamentos mais eficazes, resultando em melhores resultados para os pacientes.

Além disso, a implementação de modelos mais eficientes, como o *EfficientNetV2 B0*, pode resultar em ganhos significativos em termos de custos de treinamento de modelos. Com uma arquitetura mais leve e eficiente, esses modelos exigem menos recursos computacionais para treinamento e inferência, o que pode reduzir os custos operacionais e permitir uma adoção mais ampla em ambientes clínicos.

6.1 Limitações e Sugestões Para Trabalhos Futuros

Como principais limitações deste trabalho, destaca-se a falta de um método

definido de priorização de escolha de imagens para treinamento dentro na própria base de dados. Isto porque, como discutido, há uma gama de 80 a 200 imagens por exame de TC por paciente, e nem todas possuem a mesma relevância para a previsão de categorias.

Há neste contexto uma grande oportunidade de melhoria, que não foi abordada nesta pesquisa e pode vir a afetar a generalização e eficácia dos modelos em ambientes clínicos.

Além disso, o trabalho apresentado se caracteriza por apenas identificar se a imagem de TC apresenta ou não nódulos malignos, benignos ou não apresenta nódulos. Isto abre espaço para uma melhoria onde, além de categorizar, o modelo seja capaz de segmentar e destacar o formato e localização específica no nódulo, facilitando sua visualização e análise para profissionais de saúde.

Dito isto, abaixo encontram-se sugestões para trabalhos futuros:

- Uma pesquisa extensa e sistemática acerca da seleção das imagens mais relevantes em exames de TC para o contexto de câncer pulmonar, sendo esta seleção através de métodos automatizados ou não automatizados a fim de melhorar a generalização e eficiência dos modelos.
- Pesquisa e aplicação de técnicas de pré-processamento específicas para imagens de TC que auxiliem no desempenho dos modelos em identificar e categorizar os nódulos pulmonares.
- Implementação de uma função de mapeamento e segmentação dos nódulos pulmonares nas imagens, facilitando sua visualização e análise por profissionais de saúde.

REFERÊNCIAS

- ALI, Imdad et al. Efficient lung nodule classification using transferable texture convolutional neural network. *Ieee Access*, v. 8, p. 175859-175870, 2020.
- APOSTOLOPOULOS, Ioannis D. et al. Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques. *Medical & Biological Engineering & Computing*, v. 59, n. 6, p. 1299-1310, 2021.
- ARAÚJO, Lavínia Maria Mendes et al. Technology selection and ranking: Literature review and current applications in oil & gas industry. *Geoenergy Science and Engineering*, p. 211771, 2023.
- ATIYA, Shaik Ummay; RAMESH, N. V. K.; REDDY, B. Naresh Kumar. Classification of non-small cell lung cancers using deep convolutional neural networks. *Multimedia Tools and Applications*, p. 1-30, 2023.
- BHINDER, Bhavneet et al. Artificial intelligence in cancer research and precision medicine. *Cancer discovery*, v. 11, n. 4, p. 900-915, 2021.
- CAUCHICK-MIGUEL, Paulo Augusto. *Metodologia científica para engenharia*. Elsevier Brasil, 2019.
- CARUANA, Rich; NICULESCU-MIZIL, Alexandru. An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd international conference on Machine learning*. 2006. p. 161-168.
- COIERA, Enrico. *Guide to medical informatics, the internet and telemedicine*. Chapman & Hall, Ltd., 1997.
- CURIGLIANO, Giuseppe et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Annals of Oncology*, v. 28, n. 8, p. 1700-1712, 2017.
- DA NOBREGA, Raul Victor M. et al. Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Computing and Applications*, v. 32, p. 11065-11082, 2020.
- DU, Xinwu et al. A method for detecting the quality of cotton seeds based on an improved *ResNet50* model. *Plos one*, v. 18, n. 2, p. e0273057, 2023.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep feedforward networks. *Deep learning*, n. 1, 2016.
- HAFEMANN, Luiz G.; OLIVEIRA, Luiz S.; CAVALIN, Paulo. Forest species recognition

- using deep convolutional neural networks. In: 2014 22Nd international conference on pattern recognition. IEEE, 2014. p. 1103-1107.
- HASTIE, Trevor et al. The elements of statistical learning: data mining, inference, and prediction. New York: springer, 2009.
- HINDARTO, Djarot. Comparative Analysis VGG16 Vs MobileNet Performance for Fish Identification. International Journal Software Engineering and Computer Science (IJSECS), v. 3, n. 3, p. 270-280, 2023.
- HOSNY, Ahmed et al. Artificial intelligence in radiology. Nature Reviews Cancer, v. 18, n. 8, p. 500-510, 2018.
- HUANG, Xufeng et al. Deep transfer convolutional neural network and extreme learning machine for lung nodule diagnosis on CT images. Knowledge-Based Systems, v. 204, p. 106230, 2020.
- KANWAL, Ayesha et al. A hybrid framework for detection of autism using ConvNeXt-T and embedding clusters. The Journal of Supercomputing, p. 1-23, 2023.
- LE, Quoc. Efficientnetv2: Smaller models and faster training. In: International conference on machine learning. PMLR, 2021. p. 10096-10106.
- LEDERLIN, M. et al. Management strategy of pulmonary nodule in 2013. Diagnostic and interventional imaging, v. 94, n. 11, p. 1081-1094, 2013.
- LIU, Zhuang et al. A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 11976-11986.
- MOHAMED, Hassan et al. Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes: Case study el burullus lake. In: International water technology conference. 2015. p. 12-14.
- MULLAINATHAN, Sendhil; OBERMEYER, Ziad. Solving medicine's data bottleneck: Nightingale Open Science. Nature Medicine, v. 28, n. 5, p. 897-899, 2022.
- NAFEA, Ahmed Adil et al. A Deep Learning Algorithm for Lung Cancer Detection Using EfficientNet-B3. Wasit Journal of Computer and Mathematics Science, v. 2, n. 4, p. 68-76, 2023.
- NISHIO, Mizuho et al. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PloS one, v. 13, n. 7, p. e0200721, 2018.
- PANTANOWITZ, Liron et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation

- and deployment study. *The Lancet Digital Health*, v. 2, n. 8, p. e407-e416, 2020.
- ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain [J]. *Psychological review*, v. 65, n. 6, p. 386-408, 1958.
- RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. *nature*, v. 323, n. 6088, p. 533-536, 1986.
- SAKSHIWALA; SINGH, Maheshwari Prasad. A new framework for multi-scale CNN-based malignancy classification of pulmonary lung nodules. *Journal of Ambient Intelligence and Humanized Computing*, v. 14, n. 5, p. 4675-4683, 2023.
- SAZEWICZ, Stanislaw. Oncological diagnostics using an ensemble of convolutional neural networks. *Journal of Digital Imaging*, v. 34, n. 6, p. 1327-1337, 2021.
- SHI, Zhenghao et al. A deep CNN based transfer learning method for false positive reduction. *Multimedia Tools and Applications*, v. 78, p. 1017-1033, 2019.
- SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- SUNG, Hyuna et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, v. 71, n. 3, p. 209-249, 2021.
- TAMBLYN, Robyn et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. *Cmaj*, v. 169, n. 6, p. 549-556, 2003.
- TAN, Mingxing; LE, Quoc. Efficientnetv2: Smaller models and faster training. In: *International conference on machine learning*. PMLR, 2021. p. 10096-10106.
- THRUN, Sebastian; PRATT, Lorien (Ed.). *Learning to learn*. Springer Science & Business Media, 2012.
- YADLAPALLI, Priyanka; BHAVANA, D.; GUNNAM, Suryanarayana. Intelligent classification of lung malignancies using deep learning techniques. *International Journal of Intelligent Computing and Cybernetics*, v. 15, n. 3, p. 345-362, 2022.
- ZHANG, Chao et al. Toward an expert level of lung cancer detection and classification using a deep convolutional neural network. *The oncologist*, v. 24, n. 9, p. 1159-1165, 2019.
- ZHENG, Jie et al. Pulmonary nodule risk classification in adenocarcinoma from CT images using deep CNN with scale transfer module. *IET Image Processing*, v. 14, n. 8, p. 1481-1489, 2020.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. nature, v. 521, n. 7553, p. 436-444, 2015.