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

Luís Felipe Schons Silva

COMPARAÇÃO DE DIFERENTES ARQUITETURAS DE *DEEP LEARNING* PARA A CLASSIFICAÇÃO DE RADIOGRAFIAS DO TÓRAX ENTRE PACIENTES SAUDÁVEIS E DOENTES

Monografia de Graduação

Ouro Preto, 2023

Luís Felipe Schons Silva

COMPARAÇÃO DE DIFERENTES ARQUITETURAS DE *DEEP LEARNING* PARA A CLASSIFICAÇÃO DE RADIOGRAFIAS DO TÓRAX ENTRE PACIENTES SAUDÁVEIS E DOENTES

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

Universidade Federal de Ouro Preto

Orientador: Profª. Adrielle de Carvalho Santana, Dra.

Coorientador: Prof. Mateus Coelho Silva, Me.

Ouro Preto

2023

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

FOLHA DE APROVAÇÃO

Luís Felipe Schons Silva

Comparação de Diferentes Arquiteturas de *Deep Learning* para a Classificação de Radiografias do Tórax Entre Pacientes Saudáveis e Doentes

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

Aprovada em 24 de março de 2023

Membros da banca

Prof. Dra. Adrielle de Carvalho Santana - Orientadora (Universidade Federal de Ouro Preto)
Prof. M.Sc. Mateus Coelho Silva - Coorientador (Universidade Federal de Ouro Preto)
Prof. Dr. Rodrigo Cesar Pedrosa Silva - Convidado (Universidade Federal de Ouro Preto)
Prof. Dr. Eduardo Jose da Silva Luz - Convidado (Universidade Federal de Ouro Preto)

Adrielle de Carvalho Santana, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 24/03/2023

Documento assinado eletronicamente por **Adrielle de Carvalho Santana**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 24/03/2023, às 19:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0496814** e o código CRC **EA812B49**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.003696/2023-18 SEI nº 0496814

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

AGRADECIMENTOS

Primeiramente, agradeço a todos os que acreditaram em mim e se dispuseram a estar comigo durante essa jornada. Agradeço aos meus pais, André e Elenice, por sempre priorizarem minha educação, aos meus irmãos, João e Pedro, pela amizade e apoio. Não posso deixar de agradecer a minha namorada, Elisa, que esteve ao meu lado em toda graduação. Ela foi minha fonte de força e inspiração, acreditando em mim até quando eu mesmo duvidei. Seu amor e companheirismo tornaram essa jornada ainda mais significativa.

Também quero agradecer aos bons professores que conheci durante a graduação, em especial aos meus orientadores, Adrielle e Mateus, por dedicarem seu tempo e expertise para me ajudar na pesquisa e guiar o desenvolvimento deste trabalho. Agradeço à UFOP, pelo ensino de qualidade e pela oportunidade única.

Aos meus amigos, que tornaram minha jornada acadêmica mais divertida e leve, agradeço pelas risadas e momentos compartilhados. Finalmente, agradeço a todos que participaram dessa caminhada, vocês, sem dúvida, fazem a trilha ser mais prazerosa que o destino final, seja ele qual for.

"A inteligência é a capacidade de se adaptar à mudança." (Stephen Hawking)

RESUMO

A radiografia torácica é o exame de imagem mais utilizado na medicina, avaliando a saúde do paciente. No entanto, a interpretação das imagens pode ser subjetiva, complexa e demorada, podendo comprometer a detecção acurada e em tempo hábil de anomalias e doenças. Como resultado, a Inteligência Artificial tem sido cada vez mais utilizada como uma ferramenta promissora para melhorar a precisão e a eficiência do diagnóstico. Neste contexto, o objetivo deste estudo é discutir o uso de redes neurais convolucionais (CNN) para a classificação de radiografias torácicas com o intuito de detectar anormalidades em pacientes. A pesquisa investiga a possibilidade de utilizar as CNNs para auxiliar a tarefa de triagem, avaliando o desempenho de 10 modelos diferentes de arquitetura para classificar as radiografias de tórax em dois grupos: saudáveis e doentes. Os resultados mostram que redes neurais mais rasas, como *VGG16* e *AlexNet*, obtiveram o desempenho superior em relação às mais profundas. Assim, o projeto identifica as melhores arquiteturas para estudos futuros e sugere que o algoritmo pode apoiar a avaliação médica, direcionando a atenção do profissional para a área de interesse e reduzindo o tempo de diagnóstico. Embora o trabalho seja promissor, o uso destas redes em radiografias torácicas apresenta alguns desafios, como a necessidade de aumentar a base de dados e diminuir a presença de ruído nos rótulos.

Palavras-chaves: redes neurais convolucionais, radiografias torácicas, classificação, inteligência artificial, medicina.

ABSTRACT

Chest radiography is the most commonly used imaging examination in medicine, assessing the patient's health. However, image interpretation can be subjective, complex, and time-consuming, potentially compromising the detection of anomalies and diseases. As a result, Artificial Intelligence has been increasingly used as a promising tool to improve diagnostic accuracy and efficiency. In this context, the aim of this study is to discuss the use of convolutional neural networks (CNNs) for the classification of chest radiographs to detect abnormalities in patients. The research investigates the possibility of using CNNs to assist the screening task, evaluating the performance of 10 different architecture models to classify chest radiographs into two groups: healthy and sick. The results show that shallower neural networks, such as VGG16 and AlexNet, achieved superior performance compared to deeper ones. Thus, the project identifies the best architectures for future studies and suggests that the algorithm can support medical evaluation, directing the professional's attention to the area of interest and reducing diagnosis time. Although the work is promising, the use of these networks in chest radiographs presents some challenges, such as the need to increase the database and reduce label noise presence.

Keywords: convolutional neural networks, chest X-rays, classification, artificial intelligence, medicine.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AP	Projeção Antero-Posterior
CNN	Rede Neural Convolucional
DL	*Deep Learning*
HITL	*Human in the loop*
IA	Inteligência Artificial.
MLP	*Multilayer Perceptron*
PA	Projeção Póstero-Anterior
ReLU	Unidade Linear Retificada
RNA	Rede Neural Artificial
UFOP	Universidade Federal de Ouro Preto

SUMÁRIO

1 INTRODUÇÃO

A radiografia é o exame de imagem mais utilizado na medicina para a detecção de doenças (SMITH-BINDMAN; MIGLIORETTI; LARSON, 2008). Ele permite a visualização de estruturas internas do corpo humano, auxiliando médicos e profissionais de saúde na identificação de possíveis anomalias. Entre as modalidades radiográficas, a torácica é uma das mais utilizadas, sendo fundamental no diagnóstico de doenças cardíacas, pulmonares e outras afecções do mediastino (BRESSEM et al., 2020).

A avaliação das imagens de radiografia do tórax é realizada por profissionais da área da saúde, tais como radiologistas e demais médicos. Devido à complexidade das imagens e à subjetividade da avaliação, é comum que ocorram erros humanos ou que certas doenças passem despercebidas pelo avaliador (PINTO; BRUNESE, 2010). Além disso, o processo de inspeção pode ser demorado e trabalhoso, especialmente quando há uma grande quantidade de imagens ou falta de experiência do médico (BUSBY; COURTIER; GLASTONBURY, 2017).

Nesse sentido, a Inteligência Artificial (IA) tem sido utilizada como uma ferramenta importante na classificação de imagens médicas. Em especial, as redes neurais convolucionais (CNN) têm apresentado resultados eficazes, devido à sua capacidade de aprender características importantes presentes nas imagens (GREENSPAN; GINNEKEN; SUMMERS, 2016). Portanto, os algoritmos de *machine learning* têm o potencial de serem aplicados em todos os campos da medicina, desde a descoberta de medicamentos à tomada de decisão clínica, alterando a forma como a medicina é praticada (KER et al., 2017).

Diante disso, neste trabalho, avalia-se 10 diferentes arquiteturas de CNNs para a classificação de radiografias do tórax, utilizando o banco de dados do *NIH Clinical Center* (HOLSTE et al., 2022). Adicionalmente, este estudo apresenta a importância da utilização da IA na área da saúde, especificamente na análise de imagens radiográficas. O trabalho apresenta uma abordagem inovadora, classificando os pacientes em duas possíveis classes: os saudáveis e os doentes, visando assim comparar o desempenho deste método de classificação em relação ao utilizados pela literatura e apresentado na sessão 2.6.

1.1 Objetivos

Nesta pesquisa, pretende-se investigar o uso de 10 modelos de arquiteturas de rede neural convolucional (CNN) para classificar radiografias torácicas e separá-las entre alteradas ou inalteradas, definindo dois grupos de pacientes: os saudáveis e os não saudáveis.

Os objetivos específicos do trabalho são:

- Analisar a especificidade e sensibilidade das redes neurais e a possibilidade de utiliza-las para suporte aos profissionais durante a interpretação do exame.

- Comparar o resultado de diferentes *backbones* utilizados no treinamento da rede.

1.2 Justificativas e Relevância

Com o avanço da tecnologia e a expansão do *big data*, as redes neurais e algoritmos de inteligencia artificial ganharam, recentemente, considerável interesse comercial (GREENSPAN; GINNEKEN; SUMMERS, 2016). Estudos têm ganhado espaço e novas aplicações utilizando algoritmos de *deep learning* (DL) surgem a todo momento. Especialmente na medicina, onde acurácia e tempo de diagnóstico são fatores fundamentais. Nesse sentido, a pandemia de COVID-19 criou uma oportunidade de aplicação e popularização destes algoritmos para análise de imagens médicas, visando alcançar um diagnóstico rápido, eficiente e de baixo custo (SHORTEN; KHOSHGOFTAAR; FURHT, 2021; LIU; SIEGEL; SHEN, 2022; ASLANI; JACOB, 2023).

Dessa forma, o sucesso do *machine learning* em tarefas de visão computacional nos últimos anos vem em um momento oportuno, no qual os registros médicos tendem a se tornar digitais (KER et al., 2017). Somando-se a isto, o uso do diagnóstico por imagem aumentou drasticamente ao longo das últimas décadas, devido a crescente disponibilidade de tecnologias (SMITH-BINDMAN; MIGLIORETTI; LARSON, 2008). Portanto, diante de um aumento significativo de dados *online* e de livre acesso, surge uma grande oportunidade de implementação dos algoritmos de Inteligência Artificial.

Pesquisas envolvendo imagens médicas vêm aumentando significativamente. Pensando nisso, a radiografia do tórax é uma imagem médica frequentemente utilizada em trabalhos de computação visual, devido a sua ampla utilização no meio médico e o nível de padronização da técnica (BRESSEM et al., 2020). Diante disso, espera-se que a construção de uma rede neural convolucional (CNN) para a classificação de raio-x do tórax possa contribuir significativamente para a melhoria do processo de inspeção dessas imagens, além de auxiliar os profissionais da área médica na tomada de decisões clínicas mais precisas. Entretanto, é necessário realizar mais estudos e pesquisar para avançar na construção de um modelo que seja eficiente o bastante para ser implementado para essa tarefa. Nesse contexto, o presente trabalho avança significativamente nesta direção ao identificar as arquiteturas e parâmetros de rede que melhor desempenham na classificação de radiografias torácicas.

1.3 Organização e estrutura

A organização e estrutura do trabalho é descrita da seguinte forma: no capítulo 2 é apresentada a fundamentação teórica, introduzindo as técnicas e os conceitos necessários para o desenvolvimento deste trabalho, abordando, também, trabalhos similares. Em seguida, o capítulo 3 descreve a metodologia utilizada. Posteriormente, encontra-se o capítulo 4, no qual são detalhados os experimentos realizados e os resultados encontrados. Por fim, a conclusão é introduzida no capítulo 5, abordando as conclusões deduzidas da pesquisa, tais quais as sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresenta-se o referencial teórico essencial para a compreensão deste trabalho.

2.1 A Radiografia

Os raios-X são uma técnica de imagem médica que foi desenvolvida no final do século XIX, por Wilhelm Conrad Roentgen, enquanto realizava experiências com um tubo de raios catódicos (MOULD, 1995). Após a descoberta, Roentgen realizou a primeira radiografia humana, representada na figura 1, sendo um raio-x da mão esquerda de sua esposa, Anna Bertha Roentgen. No mesmo ano, o físico entregou um relatório para a Sociedade Físico-Médica de Würzburg, Alemanha, descrevendo todas as suas descobertas (RÖNTGEN, 1896). De acordo com Assmus (1995), após a publicação da pesquisa, o mundo ficou espantado e admirado diante da descoberta e cientistas se apressaram para explorar as novas propriedades da radiação.

Figura 1 – Primeira radiografia humana realizada no mundo: radiografia da mão da Sra. Rontgen, enviada a Ludwig Zehnder em Basiléia - Suíça, um de seus antigos alunos. Fonte: Mould (1995).

Segundo Bushong (2020), o raio-x é produzido a partir de uma fonte de radiação, geralmente um tubo de raio-x, que gera radiação ionizante. Esta radiação atravessa o tecido corporal e é registrada por uma película radiográfica ou por um detector digital. A quantidade de radiação que passa através do tecido depende de sua densidade e espessura. Tecidos mais densos, como o osso, absorvem mais radiação, enquanto tecidos menos densos, como o pulmão, permitem que mais radiação atravesse.

A imagem radiográfica resultante mostra as estruturas corporais com diferentes níveis de opacidade, dependendo da quantidade de radiação que passou através deles. Isso permite aos médicos visualizar estruturas corporais internas e avaliar a presença de anomalias, como fraturas ósseas, acometimentos pulmonares ou cardíacos, dentre outras condições clínicas. Nesse sentido, dentre as aplicabilidades deste exame, destaca-se a radiografia torácica como auxílio e reforço do diagnóstico médico. De acordo com Kelly (2012), a radiografia de tórax é responsável por uma parcela significativa de exames de imagens em todo o mundo, muitas vezes representando o primeiro passo no diagnóstico por imagem, conforme ilustra a figura 2.

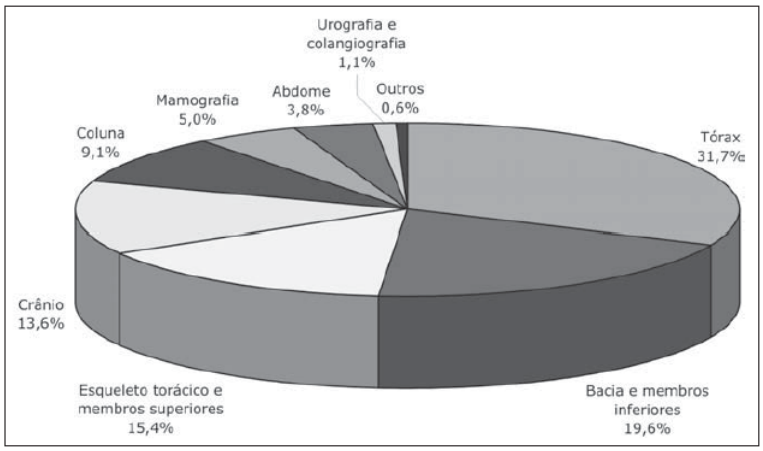


Figura 2 – Distribuição (%) dos exames radiológicos (SIA/SUS), segundo o tipo, realizados em estabelecimentos de saúde do Estado de São Paulo, referente ao período de julho de 2003 a junho de 2004. Fonte: Freitas e Yoshimura (2005).

A tecnologia de imagem radiográfica é utilizada para avaliar uma ampla variedade de condições, incluindo fraturas ósseas, lesões pulmonares, problemas cardíacos, condições gastrointestinais, entre outras. No entanto, é importante destacar que, apesar de ser uma ferramenta valiosa para o diagnóstico, o raio-x não é perfeito. Segundo Busby, Courtier e Glastonbury (2017), os erros humanos, como a interpretação incorreta das imagens ou o uso inadequado da tecnologia, podem levar a resultados falsos, contribuindo com o aumento das taxas de morbidade e mortalidade na área médica e interferindo substancialmente os custos médicos.

2.2 *Machine Learning*

Machine Learning é uma área da inteligência artificial que se concentra em criar modelos matemáticos que possam aprender e se ajustar a novos dados sem a necessidade de programação explícita. Em outras palavras, Goodfellow, Bengio e Courville (2016) afirma em seu livro que um algoritmo de *Machine Learning* é capaz de aprender com dados. Portanto, estes modelos são construídos a partir de grandes conjuntos de informação, com o objetivo de prever resultados futuros a partir destas bases.

Os algoritmos são desenvolvidos para realizar tarefas específicas, divididas entre classificação, regressão e *clustering*. Assim, é possível encontrar aplicações diversas na sociedade, como: processamento de linguagem natural (VASWANI et al., 2017), detecção de objetos (REN et al., 2015), recomendação de produtos (SHI et al., 2019), entre outros. De acordo com Goodfellow, Bengio e Courville (2016), estes modelos podem ser classificados em três categorias principais:

- **Aprendizado supervisionado:** envolve a previsão de saídas a partir de entradas previamente rotuladas. Nesta categoria, o algoritmo de aprendizado é fornecido com uma série de exemplos de treinamento, contendo a entrada e a saída desejada, em seguida é utilizado para fazer previsões sobre novos dados.

- **Aprendizado não-supervisionado:** ao contrário da classificação anterior, o algoritmo não possui rótulos para os dados de treinamento e o objetivo é descobrir padrões ou relações escondidas nos dados. Normalmente, é utilizado com o propósito de entender toda a distribuição de probabilidade que originou um conjunto de dados, seja explicitamente, como a estimativa de densidade, ou implicitamente, para tarefas como a síntese ou remoção de ruídos.

- **Aprendizado por reforço:** é uma categoria de aprendizado em que o algoritmo é treinado por meio de uma série de ações e *feedbacks*, de tal forma que ele aprende a tomar decisões e realizar tarefas com o objetivo de maximizar uma recompensa.

2.3 Redes Neurais Artificiais (RNA)

As Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no funcionamento do sistema nervoso humano. Segundo Haykin (2001), uma rede neural artificial é uma ferramenta de aprendizado de máquina que pode ser usada para modelar sistemas complexos, permitindo que a máquina aprenda a partir de exemplos.

Uma rede neural artificial típica é composta por uma sucessão de camadas de neurônios interconectados, sendo que os neurônios realizam operações matemáticas em seus *inputs* e transmitem sinais para os neurônios nas camadas subsequentes (GOODFELLOW; BENGIO; COURVILLE, 2016). A figura 3 representa uma estrutura típica de RNA.

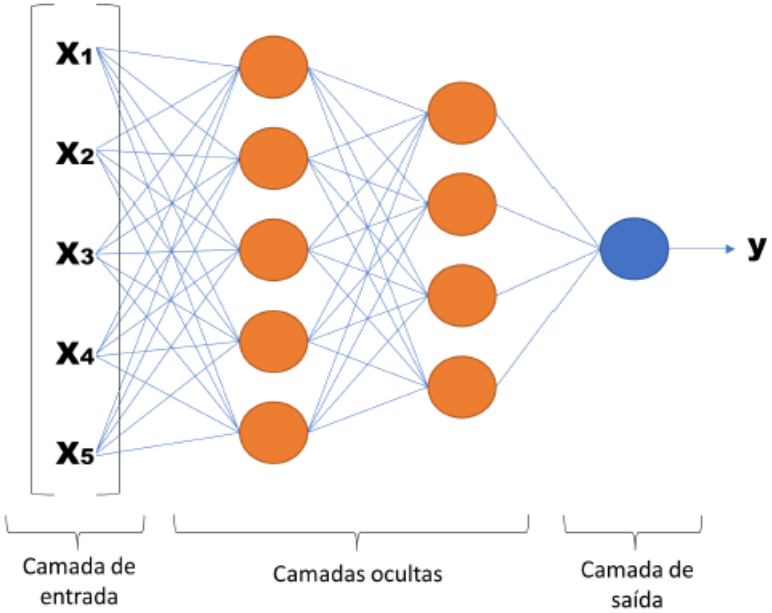


Figura 3 – Estrutura típica de uma rede neural totalmente conectada. Fonte: Mota (2021).

As camadas podem ser classificadas em três tipos: camadas de entrada, camadas ocultas e camadas de saída.

- **Camada de entrada:** É a camada responsável por receber os dados que serão processados pela rede neural. Esses dados podem ser imagens, textos ou valores numéricos. Cada entrada é processada individualmente e é atribuído um peso para a análise.

- **Camada oculta:** É a principal responsável pelo processamento e aprendizado da rede neural. Composta por múltiplos neurônios, estes recebem informações da camada de entrada e passam por uma função de ativação, responsável por determinar o nível de ativação do neurônio. Essa camada é responsável por identificar os padrões nos dados de entrada e criar conexões entre eles.

- **Camada de saída:** É a responsável por produzir a resposta da rede neural. Essa camada é composta por um conjunto de neurônios, que produzem um resultado numérico ou categórico. A mesma é treinada para produzir a resposta desejada a partir dos dados de entrada.

Os neurônios em uma RNA podem ser considerados como a unidade de processamento elemental da rede, sendo interconectados por meio de conexões ponderadas, o que significa que o peso da conexão entre dois neurônios determina a força de influência que um neurônio exerce sobre o outro (HAYKIN, 2001). Em geral, estes pesos são incializados com valores aleatórios e ajustados durante o treinamento da rede, visando otimizar o seu

desempenho. Nesse sentido, o peso associado a cada neurônio é um parâmetro de entrada desta unidade de processamento.

A função de ativação é outra componente importante em uma rede neural artificial. Responsável pela introdução de não-linearidade ao processamento de dados, a função de ativação determina a saída de cada neurônio, transformando a entrada linear em uma saída não linear (CYBENKO, 1989). A escolha da função de ativação adequada é crítica para o desempenho da rede neural, pois ela influencia a capacidade da rede de aprender e generalizar, possibilitando a rede criar modelos mais precisos e eficazes. Entre as funções de ativação mais utilizadas, pode-se citar: a sigmoide, tangente hiperbólica, ReLU (unidade linear retificada), *Softmax*, entre outras, sendo amplamente utilizadas em diferentes aplicações (KRIZHEVSKY; SUTSKEVER; HINTON, 2017).

Matematicamente, pode-se definir as operações de um neurônio k da seguinte forma:

$$f_k = \sum_{j=1}^{m} \omega_{kj} * x_j \tag{2.1}$$

$$z_k = \sigma(f_k + b_k) \tag{2.2}$$

Onde $\omega_{k1}, \omega_{k2}, ..., \omega_{km}$, são os pesos das conexões com o neurônio k; $x_1, x_2, ..., x_m$ são os sinais de entrada; f_k é a saída linear do neurônio; b_k é o bias (fator corretivo); σ é a função de ativação e z_k é o sinal de saída do neurônio k. A figura 4 demonstra o funcionamento do neurônio matemático descrito pelas fórmulas 2.1 e 2.2.

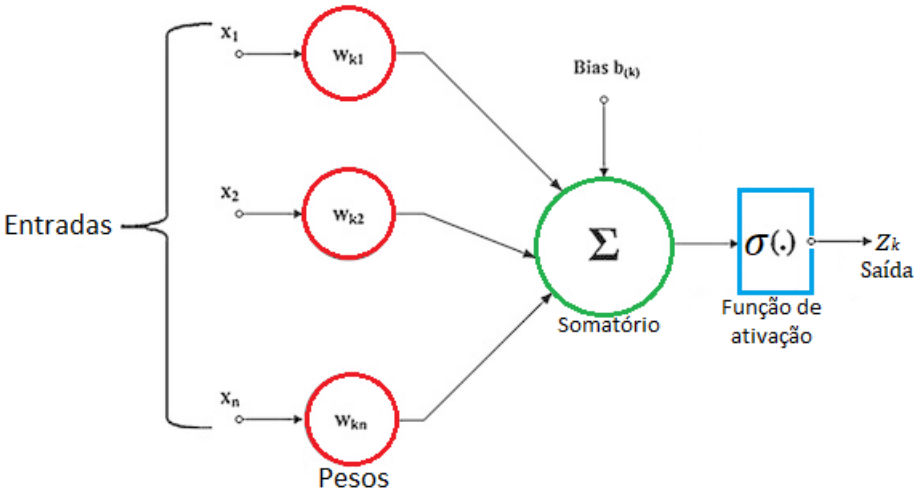


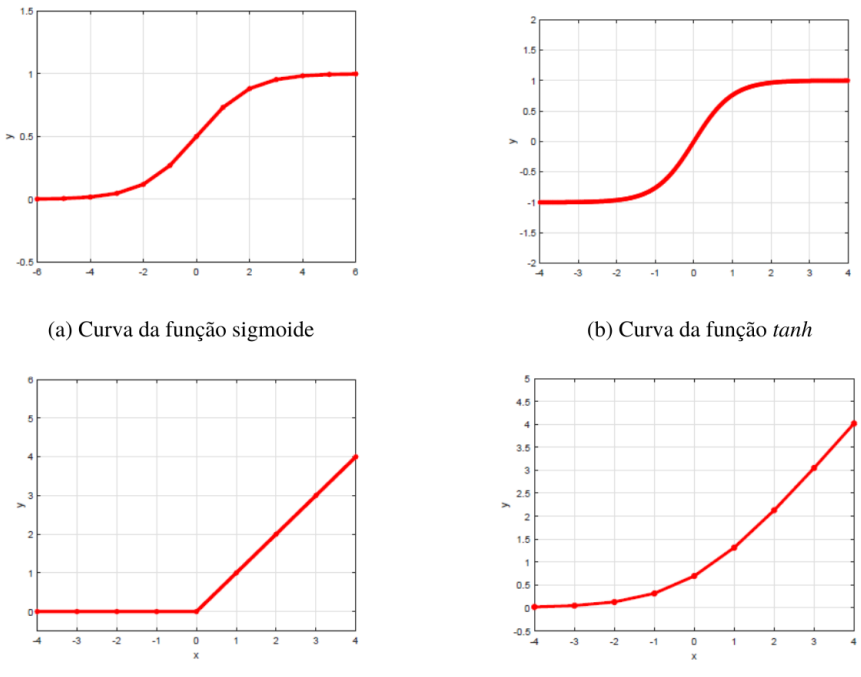
Figura 4 – Representação do nerônio matemático. Fonte: Adaptado de Goodfellow, Bengio e Courville (2016).

As redes neurais podem ser classificadas conforme a sua arquitetura. De acordo com LeCun, Bengio e Hinton (2015), as redes *Multilayer Perceptron* (MLP) são o tipo mais

simples e antigo de rede neural, composta por várias camadas de neurônios, onde cada um destes é conectado a todos os neurônios da camada seguinte. O tipo de arquitetura e os componentes da rede neural utilizada devem ser bem definidos, uma vez que podem afetar o desempenho da rede em diferentes tarefas (HAYKIN, 2001). De acordo com Greenspan, Ginneken e Summers (2016), as redes neurais convolucionais (CNNs) têm sido amplamente utilizadas na classificação de imagens médicas, sendo, portanto, a arquitetura selecionada para este trabalho. Os componentes das redes neurais e as arquiteturas são abordadas nas próximas subsecções.

2.3.1 Função de Ativação

As funções de ativação são usadas para introduzir a não-linearidade em uma rede neural, permitindo que ela capture relações mais complexas entre os dados, aumentando a capacidade de expressão do modelo (WANG, Y. et al., 2020). Segundo Goodfellow, Bengio e Courville (2016), a escolha da função de ativação apropriada é essencial, pois pode afetar a velocidade de treinamento e o desempenho da rede. Nesse sentido, a figura 5 representa a curva de algumas funções de ativação comuns.



(a) Curva da função sigmoide

(b) Curva da função *tanh*

(c) Curva da função ReLu

(d) Curva da função *softplus*

Figura 5 – Curva de quatro funções de ativação comuns. Fonte: Adaptado de Yingying Wang et al. (2020)

A função sigmoide (fig. 5-a) foi a primeira função de ativação comumente utilizada, mapeando o sinal de entrada entre 0 e 1. O seu uso diminuiu nos últimos anos, devido a difusão do gradiente[1] no aprofundamento da rede e por existirem funções computaci-

[1] A difusão do gradiente ocorre quando a informação necessária para ajustar os pesos da rede é perdida

onalmente melhores (GÉRON, 2022). A função *tanh* (fig. 5-b) surge como uma evolução da função anterior, possuindo uma taxa de convergência maior (WANG, Y. et al., 2020). No entanto, o problema da difusão do gradiente ainda está presente nesta função. Nesse sentido, a função ReLu (Unidade Linear Retificada) surge corrigindo este erro, sendo uma função insaturada e computacionalmente mais rápida que as anteriores, o que a tornou popular em modelos atuais, de acordo com Yingying Wang et al. (2020).

A função Relu (fig. 5-c) retorna zero na saída para todas as entradas menores ou iguais a zero, e retorna o próprio valor da entrada para valores maiores que zero. Apesar das vantagens desta função, o fato de zerar as entradas negativas pode causar a morte de alguns neurônios, impactando o resultado do modelo (GÉRON, 2022). Dessa forma, a função *softplus* (fig. 5-d) diminui a possibilidade de morte dos neurônios, porém traz um custo computacional maior em comparação a função ReLu (WANG, Y. et al., 2020).

2.3.2 *Backbones*

As redes neurais convolucionais permitiram treinar redes com muitas dimensões e dados, encontrando diversas aplicações (GOODFELLOW; BENGIO; COURVILLE, 2016). Nesse sentido, a competição *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) de 2012 foi um importante marco para a popularização das CNN, quando a sua utilização se tornou a abordagem dominante na competição (GÉRON, 2022). Desde então, numerosas variantes deste modelo foram propostas, sendo as mais famosas: *AlexNet*, *VGGs*, *ResNets*, *MobileNet*, entre outras.

As variantes que surgem são frequentemente pré-treinadas em grandes conjuntos de dados, como o *ImageNet*, e posteriormente ajustadas para tarefas específicas, usando técnicas de transferência de aprendizado. Nesse sentido, um *backbone* é uma rede conhecida e treinada em muitas outras tarefas, tendo previamente demonstrado a sua eficácia (ELHARROUSS et al., 2022). Os *backbones*, também chamados como modelos de base ou modelos de arquitetura, são uma parte fundamental das redes neurais convolucionais usadas em tarefas de visão computacional.

2.4 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNNs) foram popularizadas por Yann LeCun, enquanto estudante de doutorado na Universidade de Toronto, em 1989. LeCun, juntamente com seus colegas Yoshua Bengio e Geoffrey Hinton, publicou um artigo intitulado "*Backpropagation Applied to Handwritten Zip Code Recognition*"[2], descrevendo o uso de uma CNN, associada ao *backpropagation*, para reconhecer códigos postais manuscritos.

à medida que se propaga pelas camadas, dificultando o treinamento de redes muito profundas.

[2] Ver mais: LeCun, Boser et al. (1989)

Segundo LeCun, Bengio et al. (1995), as CNN possuiriam algumas vantagens em relação a arquitetura de Rede Totalmente Conectada (*Fully Connected*). Primeiramente, as redes convolucionais requerem menos parâmetros para treinar em comparação com redes totalmente conectadas, evitando problemas como limitações de *hardware* e o sobreajuste do modelo aos dados de treinamento, conhecido como *overfitting*. Além disso, a principal desvantagem das redes *Fully Connected* está no pré-processamento dos dados, que pode resultar na perda de informação espacial e temporal da amostra de entrada.

As vantagens da CNN estão na sua arquitetura, trazendo o uso de campos receptivos locais, pesos compartilhados e, em alguns casos, a subamostragem espacial ou temporal (LECUN; BENGIO et al., 1995). A arquitetura básica das CNNs consiste em camadas de convolução, camadas de *pooling* e camadas totalmente conectadas. Assim, como o próprio nome já sugere, as redes neurais convolucionais são simplesmente redes neurais que utilizam a operação linear de convolução em pelo menos uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Nesse sentido, pode-se definir a convolução como a integral do produto de duas funções, em que uma delas é deslizada sobre a outra (GÉRON, 2022).

As operações de convolução são denotadas da seguinte forma:

$$s(t) = (x * \omega)(t) \tag{2.3}$$

Onde, na terminologia de visão computacional, x é o ***input***, ou campo receptivo local, sendo, normalmente, uma imagem ou um conjunto de características. ω é o conjunto de filtro de pesos, denominado ***kernel***, e a saída, s, é chamada de mapa de características, ou ***feature map***, composta pelo resultado das convoluções realizadas entre o kernel e os campos receptivos. Assim, a convolução funciona como um filtro, reduzindo os ruídos e permitindo que a CNN aprenda a focar em diferentes partes da imagem, minimizando os efeitos de pequenas variações e defeitos na amostra, como a iluminação, posição, escala, orientação, entre outras (LECUN; BENGIO et al., 1995).

A figura 6 apresenta um exemplo da operação de convolução. Na imagem, a figura (a) é representada em forma matricial na figura (b). Esta matriz passará pela operação de convolução e terá suas dimensões reduzidas, o resultado é representado na figura (c). A operação inicia-se definindo um filtro, no caso da figura 6 este parâmetro é uma submatriz, respectiva ao "olho" à esquerda. Em seguida, a matriz imagem (b) é varrida da esquerda para a direita, coluna a coluna até a margem à direita, após esse processo a varredura volta para a margem à esquerda e desloca-se uma linha abaixo. A cada passo da varredura, ocorre a soma dos produtos efetuados entre (b) e o filtro, utilizando regras especiais (d). Ao final deste processo, tem-se uma matriz denominada mapa de características (c), representando a semelhança de cada região da imagem com o filtro utilizado (NEVES,

2021).

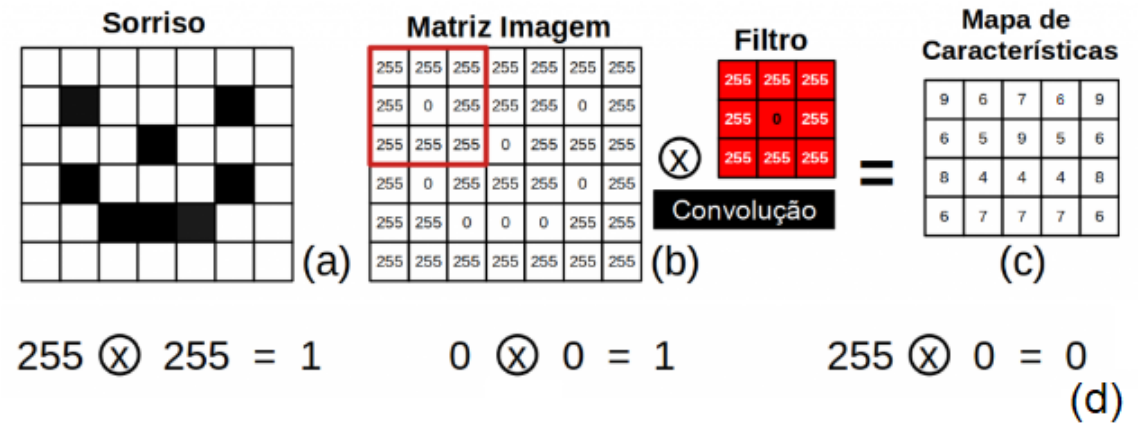


Figura 6 – Operação de convolução realizada em uma matriz (b) correspondente a imagem ilustrada (a), utilizando as regras especiais (d). O resultado da operação está ilustrado em (c). Fonte: Adaptado de Neves (2021)

Em seguida, após a camada de convolução, uma camada de *pooling* é aplicada, com o objetivo de reduzir as dimensões do mapa de recursos gerado na camada anterior. A função de *pooling* visa reduzir o tamanho espacial da representação da imagem, buscando preservar a informação mais importante e diminuir a complexidade computacional, a fim de evitar o *overfitting* e melhorar o desempenho da rede (GOODFELLOW; BENGIO; COURVILLE, 2016). As funções mais comuns são o *max pooling* e o *average pooling*. A figura 7 representa a operação de *pooling*, utilizando a função *max pooling*. A operação inicia-se com a varredura do mapa de características, resultante de uma convolução. Assim, a varredura passa por cada região não sobreposta da imagem, retendo o valor máximo destas e descartando as demais informações.

Figura 7 – Operação de *pooling* realizada no mapa de características resultante de uma convolução, utilizando a função *max pooling*. Fonte: Adaptado de Neves (2021)

2.5 Treinamento de redes profundas

O treinamento de redes profundas é uma das áreas mais importantes de *Machine Learning*, permitindo que modelos complexos sejam criados e ajustados para solucionar

problemas em diversas áreas, sendo um dos principais desafios em *deep learning*. Conforme mencionado em capítulos anteriores, a escolha do modelo e dos parâmetros de treinamento, influenciam diretamente os resultados da rede. Nesse sentido, a complexidade pode aumentar com o número de camadas do sistema, gerando problemas como o desaparecimento ou a explosão do gradiente, o *overfitting* aos dados de treinamento ou simplesmente demandar um tempo de treinamento inviável (GÉRON, 2022).

Nesse sentido, visando otimizar o treinamento e reduzir o impacto dos problemas mencionados, existem técnicas que podem ser implementadas. A seguir, serão abordadas as técnicas importantes para a compreensão deste trabalho, são elas: a normalização dos dados e os otimizadores.

2.5.1 Normalização

A normalização é uma técnica usada para tornar os dados de entrada mais padronizados e facilitar o treinamento da rede. Nesse sentido, a normalização mais comum em redes neurais é a normalização por lotes (*Batch Normalization*), proposta em 2015 por Ioffe e Szegedy (2015), com o objetivo de evitar os problemas com o gradiente. A técnica consiste na adição de uma camada responsável por ajustar a distribuição do sinal, antes da função de ativação. Dessa forma, as entradas de um lote são centralizadas em zero e normalizadas, permitindo que o modelo aprenda a escala ótima e a média das entradas de cada camada (GÉRON, 2022). Além da técnica de *Batch Normalization*, outras técnicas de normalização comuns são: normalização por instância[3] e normalização por camada[4].

2.5.2 Otimizadores

Os algoritmos de otimização são responsáveis por ajustar os pesos e *bias* da rede para minimizar uma função de perda, que representa a diferença entre a saída real da rede e a saída desejada. Os otimizadores são uma parte fundamental do treinamento de redes neurais, pois podem influenciar significativamente a convergência e o desempenho da rede. Existem vários algoritmos de otimização disponíveis, sendo que os mais comuns são: Gradiente acelerado de Nesterov (NAG, do inglês *Nesterov Accelerated Gradient*), *Stochastic Gradient Descent* (SGD), AdaGrad, RMSProp e Adam. De acordo com Géron (2022), o otimizador Adam é o mais utilizado e indicado para a maioria dos modelos.

2.6 Uso da IA em exames de imagem

A aplicação da Inteligência Artificial (IA) na área da saúde tem sido cada vez mais utilizada em diversas tecnologias, incluindo o processamento de exames de imagem. De

[3] Proposta por: Ulyanov, Vedaldi e Lempitsky (2016)
[4] Proposta por: Ba, Kiros e Hinton (2016)

acordo com Aung, Wong e Ting (2021), a IA tem o potencial de revolucionar a medicina, ao reduzir o tempo e os custos associados aos exames, permitindo um diagnóstico mais rápido e preciso. No entanto, a baixa disponibilidade de dados públicos e a dificuldade de coletar amostras, tanto em termos de custo quanto temporalmente, são fatores limitantes no avanço da IA no domínio médico (GREENSPAN; GINNEKEN; SUMMERS, 2016). Entretanto, nos últimos anos alguns pesquisadores direcionaram seus esforços em construir *datasets* públicos e de livre acesso, com o objetivo de permitir que a comunidade científica estude modelos de predição para tais dados. Portanto, esta seção do estudo explora trabalhos relacionados, abordando o uso de IA em exames de imagem do tórax, subdividindo-se em subseções englobando os *datasets* criados pela comunidade.

2.6.1 Banco de dados *ChestX-ray14*

O projeto de Xiaosong Wang et al. (2017) é um importante marco inicial para o desenvolvimento de CNNs para classificação de raio-x de tórax. O trabalho apresenta o *ChestX-ray8*[5], um grande banco de dados hospitalar composto por mais de 100.000 radiografias torácicas, com o objetivo de padronizar e compartilhar dados para a aplicação de métodos de *Machine Learning* em radiologia. A pesquisa buscou classificar as imagens em 8 categorias de doenças, utilizando uma rede neural convolucional profunda baseada nos *backbones*: AlexNet, GoogLeNet, VGGNet-16 e ResNet-50. Os resultados foram promissores, demonstrando uma superioridade da arquitetura ResNet-50 em relação as demais redes. Além disso, o trabalho fomentou novos estudos ao disponibilizar o *dataset*.

Assim, o trabalho de Yao et al. (2017) também utiliza o banco de dados *ChestX-ray8*, apresentando um método inovador de *Machine Learning* para diagnóstico médico, que explora as informações de dependência entre diferentes doenças. Os autores utilizaram uma rede neural convolucional densamente conectada, similar a *DenseNet* de Huang et al. (2017), para codificar as imagens de entrada. Em seguida, o vetor resultante da codificação é utilizado como entrada em uma rede neural recorrente, permitindo que o modelo explore dependências estatísticas entre as 14 classes de doenças, visando melhorar a precisão de suas previsões. As arquiteturas utilizadas foram treinadas do zero, sem o uso de *backbones*, buscando garantir que as características específicas dos dados fossem capturadas pelos modelos. Os resultados da pesquisa foram comparados aos de Xiaosong Wang et al. (2017) e apresentaram uma precisão superior.

O estudo de Rajpurkar et al. (2018) também utilizou o *dataset ChestX-ray14*, porém o objetivo foi comparar os resultados da rede neural convolucional desenvolvida com a atuação de 9 radiologistas. O método proposto foi a CNN denominada *CheXNeXt*, construída utilizando a arquitetura *DenseNet* com 121 camadas. O algoritmo foi confrontado

[5] Banco de dados diponibilizado pelo NIH e utilizado neste trabalho. Posteriormente renomeado como *ChestX-ray14*, devido ao número de patologias presentes.

com 6 radiologistas certificados (média de 12 anos de experiência) e 3 residentes sêniores, alcançando o mesmo desempenho dos radiologistas em 11 patologias, perdendo em apenas 3. Entretanto, o algoritmo levou cerca de 1.5 minutos para avaliar as 420 imagens, ao passo que os profissionais de saúde demoraram 240 minutos.

A pesquisa desenvolvida por Baltruschat et al. (2019) adotou uma abordagem diferente das anteriores, considerando parâmetros não-imagem, como a idade dos pacientes, genêro e o tipo de aquisição (AP ou PA). O trabalho utilizou a arquitetura ResNet-50, experimentando as variações ResNet-38 e ResNet-101. Além disso os autores propuseram o desenvolvimento de uma rede a partir do zero, sem a utilização de *backbones*. Os resultados demonstram que as características não-imagem não melhoraram significativamente os resultados, trazendo a hipótese de que os recursos das amostras já possuem essas informações codificadas. Os dados encontrados também demonstram uma superioridade da arquitetura ResNet-38 em relação as demais, especialmente quando treinadas do zero.

2.6.2 Banco de dados *CheXpert*

O trabalho de Irvin et al. (2019) foi o responsável pelo desenvolvimento da base de dados *CheXpert*. O *dataset* desenvolvido possui 224.316 radiografias do tórax de 65.240 pacientes, classificadas entre 14 classes. A rotulagem dos dados foi realizada a partir de um rotulador que pode extrair observações de relatórios de radiologistas, contando com um rótulo de incerteza. A partir do banco de dados rotulado, os autores testaram várias arquiteturas de redes neurais convolucionais, como: ResNet-152, DenseNet121, Inception-v4 e SEResNeXt101, concluindo que a arquitetura DenseNet121 produziu os melhores resultados. Em seguida, a pesquisa avaliou o modelo desenvolvido comparando-o com 3 radiologistas certificados. Os resultados demonstraram que o modelo conseguiu superar pelo menos 2 dos 3 radiologistas na detecção de 4 patologias relevantes.

A base de dados *CheXpert* não só permitiu que novos estudos fossem realizados, ela também se tornou uma competição para verificar quais são os melhores modelos. Assim, a pesquisa de Pham et al. (2019) conseguiu desenvolver uma rede que, na época da pesquisa, se tornou a mais eficiente. Os autores trabalharam com uma CNN baseada na arquitetura DenseNet-121, introduzindo um novo procedimento de treinamento em que as dependências entre doenças e os rótulos de incerteza são explorados e integrados no treinamento das CNNs. Além disso, o trabalho constatou que as arquiteturas podem variar a precisão de classificação para cada doença, sendo que algumas apresentam melhores resultados para determinadas patologias. Nesse sentido, os autores estudaram a precisão das arquiteturas DenseNet-121, DenseNet-169, DenseNet-201, Inception-ResNet-v2, Xception e NASNetLarge, propondo um método que explora o melhor resultado de cada um destes *backbones*.

Os bancos de dados de raio-x de tórax difundiram o uso de diferentes *backbones*

para classificar as imagens. Nesse sentido, Bressem et al. (2020) estudaram e compararam 15 modelos de CNN utilizando 5 arquiteturas diferentes, ResNet, DenseNet, VGG, SqueezeNet, Inception v4 e AlexNet. Além disso, os autores variaram o *Batch Size*, testando 16 e 32 imagens por iteração. Considerando a literatura, geralmente, redes mais profundas permitiam alcançar resultados melhores neste *dataset*. Entretanto, os autores concluíram ao final da pesquisa que o aumento da complexidade e profundidade das redes neurais artificiais para a classificação das radiografias de tórax nem sempre é necessária para alcançar bons resultados. Nesse sentido, o trabalho alcançou resultados similares ao de outros trabalhos utilizando redes mais rasas, contendo poucas camadas, como a AlexNet usando oito camadas. A vantagem de se utilizar estes modelos é o ganho computacional, uma vez que podem ser treinadas mais rapidamente.

O uso de diferentes arquiteturas pre-existentes se tornou comum na classificação de raio-x do tórax, muitas vezes utilizando os pesos da *ImageNet*. Tendo isto em vista, o trabalho de Raghu et al. (2019) busca avaliar a transferência de aprendizado destas redes utilizando pesos definidos, como os da *ImageNet*. Os autores avaliaram as arquiteturas ResNet50 e Inception-v3, além de criar uma rede neural convolucional simples, com poucas camadas. Os resultados do estudo demonstram que a aprendizagem por transferência oferece ganhos de desempenho limitados, enquanto arquiteturas muito menores podem executar a mesma tarefa de forma similar aos modelos *ImageNet*. Além de imagens de raio-x do tórax, os autores também avaliaram imagens de retina dos olhos, visando classificar os dados em doenças oculares.

3 METODOLOGIA

Nesta seção, apresenta-se o treinamento das redes e as métricas analisadas para avaliar o desempenho dos modelos. Em resumo, foram avaliadas 10 diferentes arquiteturas de redes neurais convolucionais, com o intuito de classificar as imagens radiográficas dada a presença de anormalidades ou não, classificando-as entre pacientes saudáveis e doentes.

3.1 Treinamento da rede neural

A tarefa de classificação das imagens foi realizadas utilizando-se uma rede neural convolucional. Os modelos avaliados foram as arquiteturas: VGG16, MobileNet, MobileNetV2, MobileNetV3Small, DenseNet121, DenseNet169, ResNet50, ResNet101 e EfficientNetV2B0. Os testes foram realizados utilizando os pesos da *ImageNet* e inicializando as redes com pesos aleatórios, a fim de comparar os resultados. O treinamento das redes ocorreu aplicando-se a linguagem de programação *Python* em conjunto as bibliotecas Keras, pandas, *scikit-learn* e TensorFlow, em execução no Jupyter Notebook. Os códigos utilizados foram disponibilizados no GitHub com licença de código aberto[1]. A pesquisa foi realizada em uma estação de trabalho física, rodando em Windows 11 e com uma placa gráfica NVIDIA GeForce RTX 2060 (6 GB de RAM GDDR6).

Os dados foram rotulados binariamente entre pacientes saudáveis e pacientes doentes, representados por 0 e 1, respectivamente. Os parâmetros de treino definidos foram o *batch size* de tamanho 32 e uma resolução das imagens de 256x256 pixels para todos os modelos. Outras resoluções também foram validadas em algumas arquiteturas, a título de comparação, sendo elas de 128x128, 320x320 e 512x512 pixels, neste último caso com um *batch size* de tamanho 16. Os valores definidos são limitados pela disponibilidade de RAM-GPU, sendo assim, buscou-se explorar o melhor desempenho possível considerando as limitações de *hardware*.

O otimizador escolhido foi o ADAM, sendo o modelo comumente utilizado pela literatura[2], com a taxa de aprendizado definida como $lr = 0.001$. Durante o treinamento, a taxa de aprendizado é reduzida por um fator de 10, quando a perda de validação não melhora por 5 épocas. A fim de prevenir o sobreajuste (*overfitting*) das redes, adotou-se um critério de parada antecipada, quando a perda de validação não diminui por 15 épocas consecutivas. Após cada época, a rede é salva e armazenada, permitindo restaurar a rede com o melhor desempenho, assim como proposto por Rajpurkar et al. (2018).

[1] Disponível em: https://github.com/LuisSchons/Multiple-CNNs-for-chest-xray

[2] (YAO et al., 2017; RAJPURKAR et al., 2018; BALTRUSCHAT et al., 2019; RAGHU et al., 2019; IRVIN et al., 2019)

3.2 Métricas de desempenho

Ao final do treinamento da rede, é computada a sua duração e, em seguida, realiza-se o teste do modelo. Neste, é possível avaliar a rede desenvolvida pelos critérios de precisão, *recall*, ou sensibilidade e *f1-score*, calculadas a partir da matriz de confusão, impressa no final do teste. A matriz é uma tabela que demonstra a frequência com a qual cada classe foi classificada corretamente e incorretamente pelo modelo. Utilizando como base os valores da matriz de confusão, é possível calcular várias métricas de desempenho. A matriz 2x2 possui quatro possíveis resultados:

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

A partir dos valores da matriz de confusão é possível calcular as demais métricas avaliadas durante o teste do modelo. A precisão mede a quantidade de vezes que o modelo acerta em relação ao total de vezes que ele tenta acertar, calculado por: $VP/(VP+FP)$, seu *range* vai de 0 a 1, sendo que valores mais próximos de 1 indicam uma precisão melhor. Já o *recall* mede a quantidade de vezes que o modelo acerta em relação ao total de vezes que ele deveria ter acertado, calculado por: $VP/(VP+FN)$, seu *range* é o mesmo da precisão. Por fim, o *f1-score* é uma métrica que combina precisão e *recall* de maneira equilibrada, seu valor é medido realizando a média harmônica destas duas métricas, podendo variar de 0 a 1, sendo que valores mais próximos de 1 indicam melhores resultados. As métricas geradas são importantes para ajudar a ajustar os parâmetros da rede neural e melhorar o seu desempenho em problemas de classificação. Dessa forma, é possível avaliar os modelos em relação a confiabilidade dos resultados e eficiência do treino.

4 EXPERIMENTOS E RESULTADOS

Este capítulo apresenta a base de dados utilizada e discute os protocolos, os experimentos e os resultados obtidos.

4.1 Base de dados

A base de dados de radiografias torácicas, denominada *ChestX-ray14* e disponibilizada pelo *NIH Clinical Center*[1], é uma das maiores coleções de radiografias do tórax, contendo 112.120 imagens de 30.805 pacientes (RAJPURKAR et al., 2018). Os dados são rotulados binariamente entre 14 patologias e uma classe denominada *"No Finding"*, indicando a ausência de doenças, obtidos a partir de métodos de extração automática de relatórios dos profissionais de saúde. As imagens seguem um tipo de projeção padronizado, variando a posição de amostragem entre póstero-anterior (PA) e ântero-posterior (AP). Dessa forma, o banco de dados demonstra ser ideal para a pesquisa, eliminando os desafios de rotulagem dos dados, conforme citado por Greenspan, Ginneken e Summers (2016).

Além disso, as imagens possuem as informações de gênero e idade, permitindo realizar análises estatísticas acerca da representatividade dos dados. Nesse sentido, a figura 8 ilustra a distribuição das amostras, classificadas por gênero e idade.

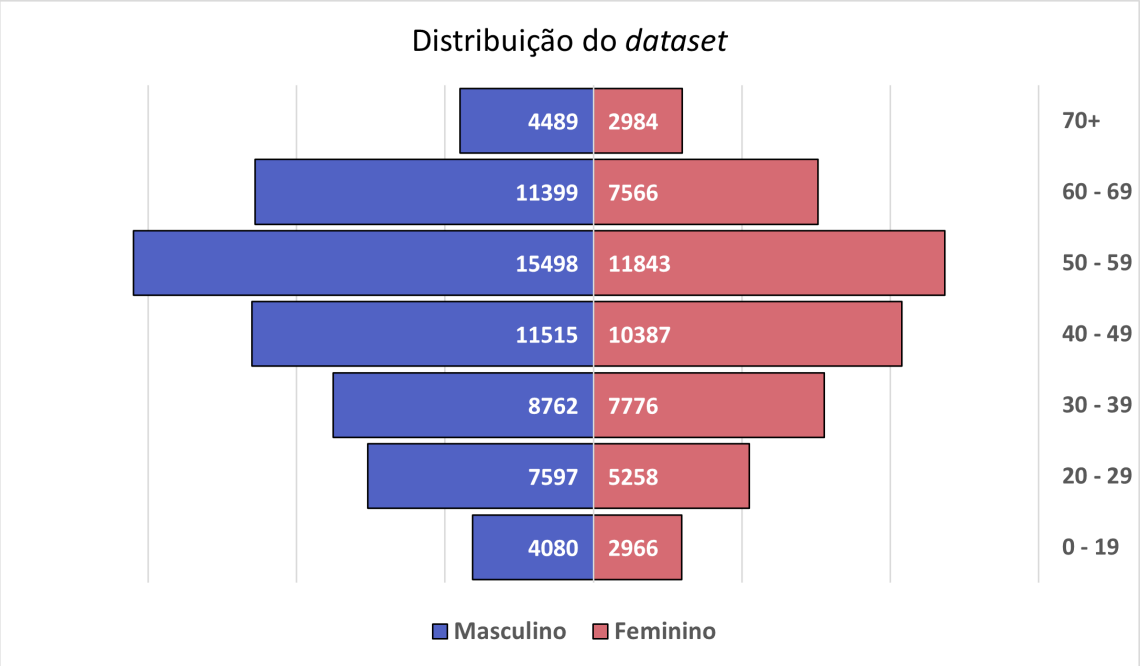


Figura 8 – Distribuição do *dataset* classificado por gênero e idade. Fonte: autoria própria.

[1] Acesso em: https://nihcc.app.box.com/v/ChestXray-NIHCC/

A partir do gráfico ilustrado na figura 8, é possível perceber que os dados são bem distribuídos, possuindo uma pequena predominância de dados masculinos sobre os femininos. As idades com mais registros de imagens radiológicas estão entre 40 e 59 anos, sendo a média de idade 46,87 anos com desvio padrão de 16,60 anos (BALTRUSCHAT et al., 2019). O *dataset* foi divido, aleatoriamente, em 70% para treinamento, 10% para validação e 20% para teste, seguindo a divisão utilizada pelos trabalhos de Xiaosong Wang et al. (2017), Yao et al. (2017) e Baltruschat et al. (2019). Inicialmente, a pesquisa de Xiaosong Wang et al. (2017) demonstra variações insignificantes nos resultados ao variar arbitrariamente as divisões dos dados, posteriormente, reafirmada pelo trabalho de Yao et al. (2017). Nesse sentido, para permitir uma melhor comparação entre diferentes arquiteturas, utilizou-se uma única divisão da base de dados.

4.2 Análise dos resultados obtidos com pesos pré-treinados do modelo *ImageNet*

O uso de pesos pré-treinados em modelos de redes neurais convolucionais (CNNs) tem se tornado uma prática comum em tarefas de visão computacional. Segundo Bressem et al. (2020), essas redes neurais costumam ser treinadas em grandes conjuntos de dados publicamente disponíveis, como o *ImageNet* e, portanto, já são capazes de reconhecer vários recursos de imagem, podendo ser aplicadas em outras tarefas. O modelo *ImageNet*, em particular, foi treinado em um conjunto de dados com mais de 1 milhão de imagens e 1000 classes distintas, tendo sido amplamente utilizado para inicializar os pesos de CNNs na tarefa de classificação de radiografias do tórax (WANG, X. et al., 2017).

Nesta seção do trabalho, analisa-se os resultados obtidos ao utilizar pesos pré-treinados do modelo *ImageNet* na tarefa de classificação das imagens do *NIH Clinical Center*. Assim, a tabela 1 ilustra as métricas do conjunto de teste, organizadas por arquitetura, utilizando uma resolução de 256x256 pixels e um *Batch Size* de tamanho 32.

Tabela 1 – Métricas do conjunto de teste, organizadas por arquitetura, utilizando uma resolução de 256x256 pixels e um *Batch Size* de tamanho 32.

Redes	**Classe**	**Precisão**	***Recall***	***f1-score***
VGG16	0	**0,71**	0,67	**0,69**
	1	**0,64**	**0,68**	**0,66**
MobileNet	0	0,65	0,65	0,65
	1	0,59	0,59	0,59

Continua na próxima página

Tabela 1 – *Continuação da tabela*

Redes	**Classe**	**Precisão**	***Recall***	***f1-score***
MobileNetV2	0	0,65	0,62	0,63
	1	0,58	0,63	0,60
DenseNet121	0	0,66	0,64	0,65
	1	0,60	0,63	0,61
DenseNet169	0	0,64	0,64	0,64
	1	0,58	0,58	0,58
ResNet50	0	0,69	0,68	**0,69**
	1	**0,64**	0,65	0,64
ResNet101	0	0,68	0,70	**0,69**
	1	**0,64**	0,62	0,63
EfficientNetV2B0	0	0,68	0,67	0,67
	1	0,62	0,64	0,63
Xception	0	0,66	0,66	0,66
	1	0,61	0,61	0,61
AlexNet	0	0,64	**0,71**	0,67
	1	0,62	0,54	0,58

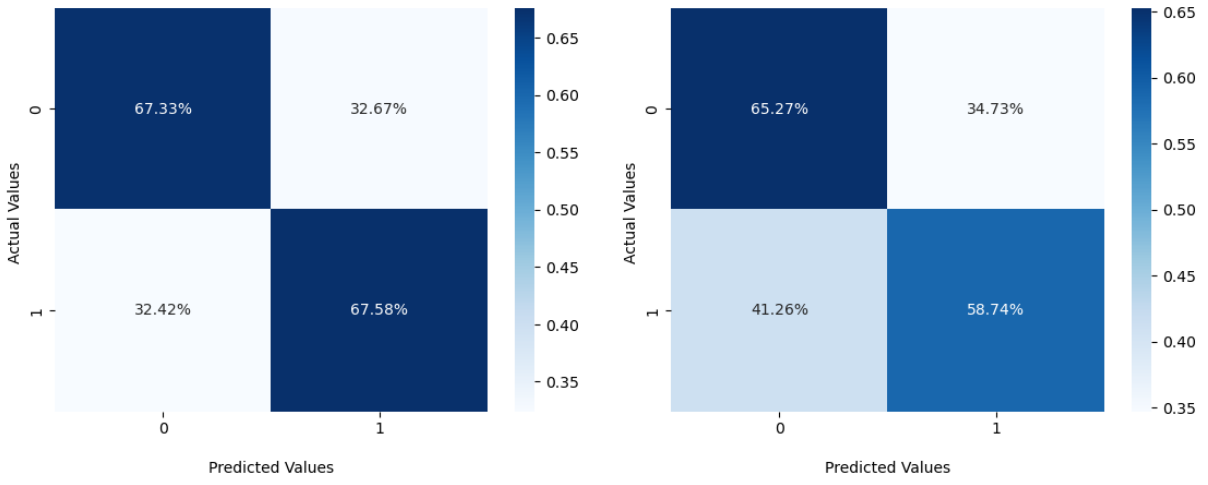
Fim da tabela

Na tabela 1, a coluna de **Classe** representa as classificações realizadas pelo modelo, sendo que os valores: 0 indicam pacientes saudáveis e 1 indicam pacientes doentes. Os melhores resultados de cada métrica para cada classe foram destacados em negrito. Analisando os dados obtidos, observa-se pouca variação de **Precisão**, ***Recall*** e ***f1-score*** entre os modelos de arquitetura utilizados. Apesar disso, é possível afirmar que os melhores resultados predominaram na arquitetura VGG16, o que motivou a realização de mais treinamentos utilizando esta arquitetura e modificando os parâmetros de treino.

As arquiteturas que mais se destacaram foram a VGG16, ResNet e AlexNet, demonstrando que redes neurais mais profundas não necessariamente desempenham melhor que redes mais rasas para a classificação proposta. As redes mencionadas são compostas pelas camadas:

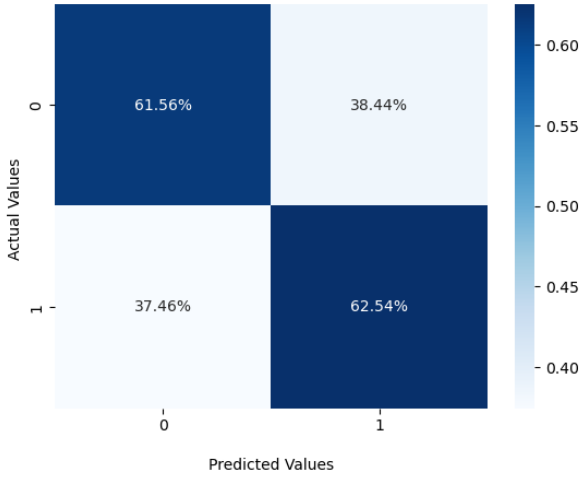
- **AlexNet:** 5 camadas convolucionais e 3 camadas totalmente conectadas;
- **VGG16:** 13 camadas convolucionais e 3 camadas totalmente conectadas;
- **ResNet50:** 49 camadas convolucionais e 1 camada totalmente conectada;
- **ResNet101:** 100 camadas convolucionais e 1 camada totalmente conectada;

De forma a sustentar os resultados apresentados, outra conclusão semelhante foi proposta pelo trabalho de Bressem et al. (2020), que após analisar diversos modelos de CNN, encontrou o melhor desempenho de classificação nas arquiteturas *AlexNet*, *ResNet-34* e *VGG16*. As figuras 9a-9j representam as matrizes de confusão do conjunto de teste para cada rede convolucional avaliada, que permitem calcular as métricas dispostas na tabela 1.

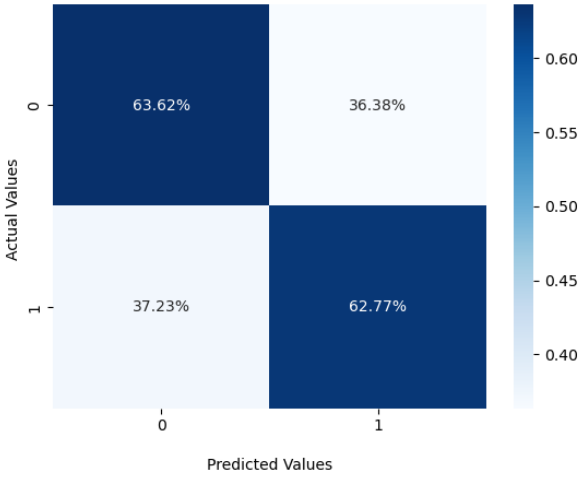


(a) VGG16

(b) MobileNet

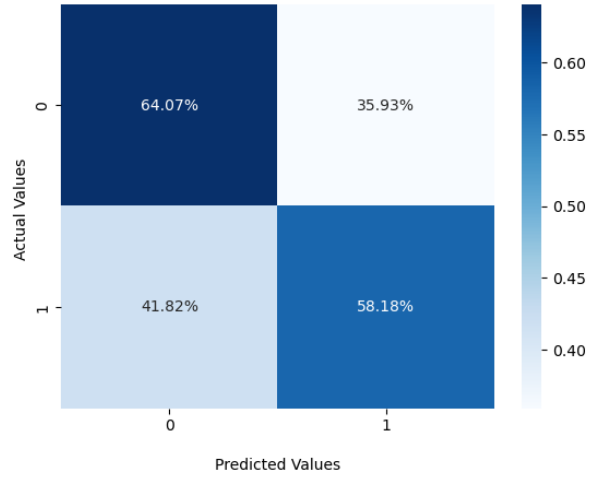


(c) MobileNetV2

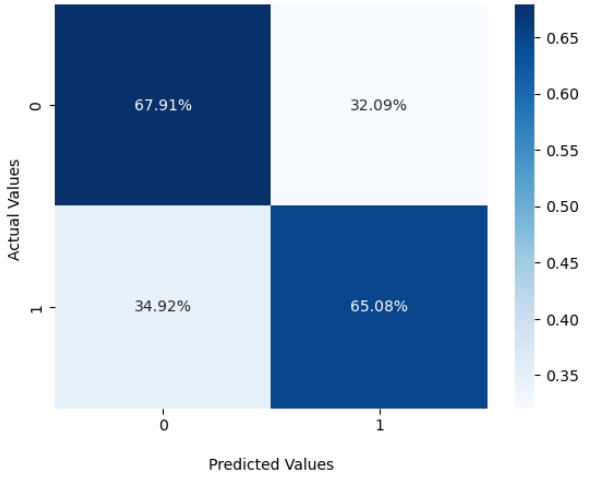


(d) DenseNet121

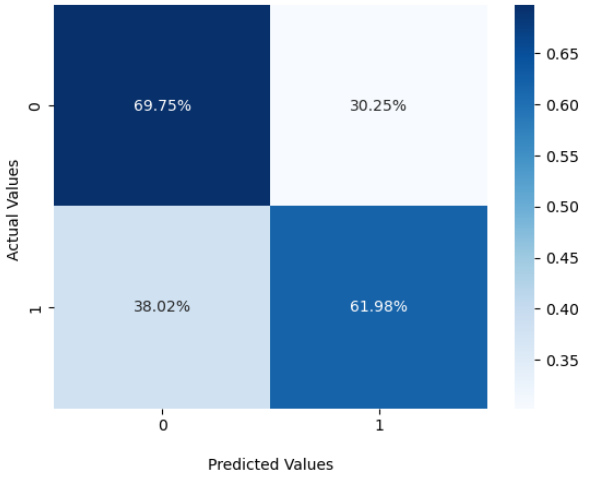
Figura 9 – Matrizes de confusão das arquiteturas utilizadas e avaliadas. Fonte: autoria própria.



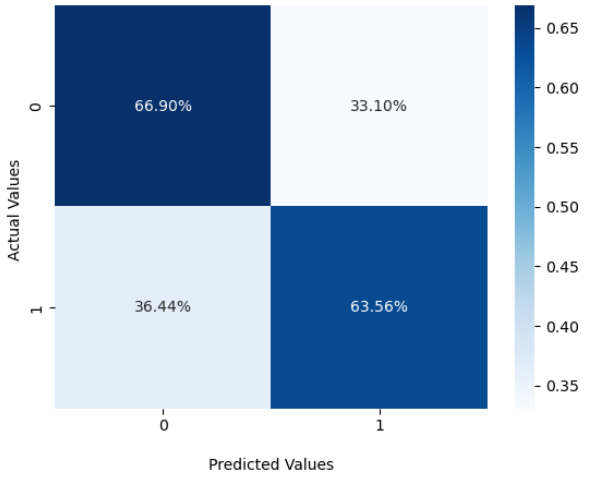
(e) DenseNet169



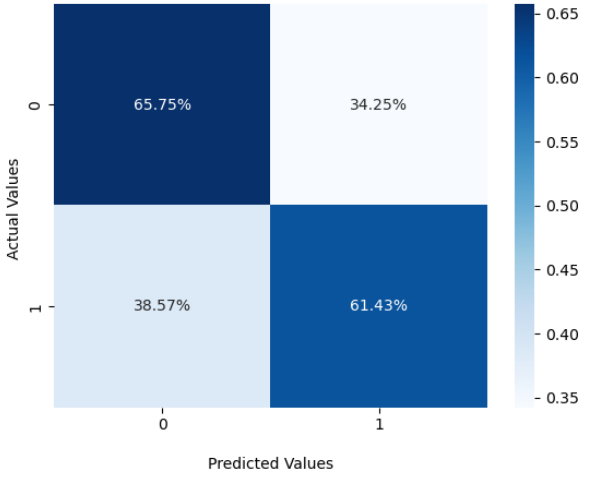
(f) ResNet50



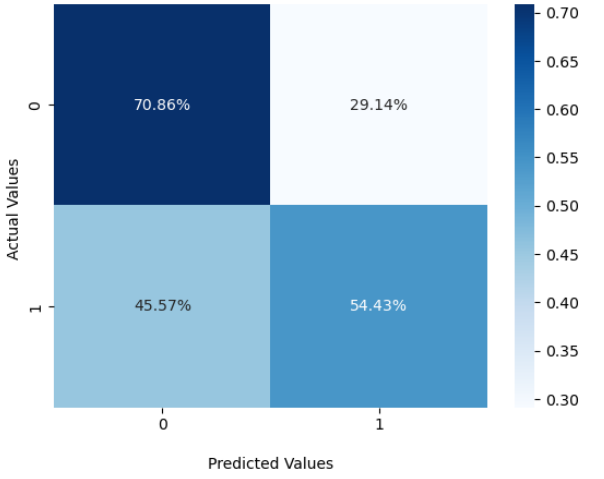
(g) ResNet101



(h) EfficientNetV2B0



(i) Xception



(j) AlexNet

Figura 9 – Matrizes de confusão das arquiteturas utilizadas e avaliadas. Fonte: autoria própria.

A interpretação dos valores da matriz de confusão e das métricas de desempenho depende do contexto do problema e dos objetivos do modelo. Por exemplo, em problemas

de detecção de fraudes financeiras, é mais importante maximizar a sensibilidade do que a precisão, ou seja, o objetivo é detectar o máximo possível de fraudes verdadeiras, mesmo que em alguns casos falsas fraudes sejam consideradas como verdadeiras. Ao contrário disso, em problemas de detecção de spam em *emails*, pode ser mais importante maximizar a especificidade, buscando evitar classificar mensagens legítimas como spam.

No contexto da radiologia, os autores Oliveira et al. (2010) definem a sensibilidade como a probabilidade de um teste detectar uma anomalia na presença da doença, representado pela taxa de Verdadeiros Positivos. Já a especificidade é a probabilidade de um teste não detectar anomalias quando o paciente está saudável, isto é, a taxa de Verdadeiros Negativos. A escolha entre estes conceitos é uma questão crucial na radiologia. Se o objetivo for detectar qualquer anomalia possível, a sensibilidade deve ser maximizada. No entanto, isso pode resultar em muitos resultados falso-positivos, ocasionando tratamentos desnecessários e o uso excessivo dos recursos disponíveis. Por outro lado, se o objetivo é garantir que o diagnóstico seja correto, a especificidade deve ser maximizada. Essa medida tende a reduzir os resultados falso-positivos e diminuir a ansiedade e o estresse desnecessários para o paciente, porém há a possibilidade de detectar-se menos anomalias, aumentando o risco de complicações (OLIVEIRA et al., 2010). O teste perfeito seria aquele sem resultados falso positivos ou falso negativos, entretanto estes exames normalmente não existem.

De acordo com Oliveira et al. (2010), entende-se que é importante, portanto, considerar o contexto clínico ao avaliar a qualidade de uma radiografia. Nesse sentido, a rede *VGG16* (Fig. 9a) apresentou a melhor sensibilidade, com um valor de 67,58%, enquanto a *AlexNet* (Fig. 9j) a melhor especificidade, com um valor de 70,86%. Novamente, os resultados demonstram uma superioridade de desempenho das arquiteturas com menos camadas.

Uma vantagem em utilizar estas arquiteturas, consideradas rasas em relação as demais, está na redução dos requisitos de *hardware* necessários, permitindo um treinamento mais curto em comparação às redes mais profundas (BRESSEM et al., 2020). Essa afirmação pode ser confirmada pela análise da Tabela 2, que apresenta os tempos de processamento de cada arquitetura com base nos conjuntos de treinamento e validação utilizados.

Tabela 2 – Duração do treinamento dos modelos avaliados, utilizando uma resolução de 256x256 pixels e um *Batch Size* de tamanho 32.

Redes	**Duração**		
	Época	**Média**	**Total**
VGG16	103s - 156s	148.24s	2h 33m 11s

Continua na próxima página

Tabela 2 – *Continuação da tabela*

Redes	**Duração**		
	Época	**Média**	**Total**
MobileNet	100s - 151s	144.79s	5h 16m 07s
MobileNetV2	103s - 192s	149.95s	4h 14m 55s
DenseNet121	104s - 240s	156.52s	2h 39m 08s
DenseNet169	103s - 207s	154.77s	2h 57m 59s
ResNet50	104s - 204s	155.07s	2h 29m 54s
ResNet101	105s - 160s	148.12s	2h 25m 39s
EfficientNetV2B0	100s - 156s	147.79s	2h 45m 02s
Xception	103s - 160s	147.74s	4h 43m 10s
AlexNet	102s - 189s	152.13s	1h 18m 36s

Fim da tabela

Os valores da tabela foram medidos durante a etapa de treinamento da rede, sendo que a coluna de **Época** representa a menor e a maior duração de cada época. Em seguida, calcula-se a média destes tempos, disposta na coluna **Média**. Por fim, a soma de todos os tempos de treinamento é exposta na última coluna, denominada **Total**.

Analisando a Tabela 2, pode-se observar um tempo de treinamento significativamente menor do modelo *AlexNet*, seguido pelas arquiteturas *ResNet* e *VGG16*. Uma vantagem de tempos de treinamento mais curtos é a possibilidade de integrar métodos de melhoria no treino das redes, como técnicas de pré-processamento dos dados ou o treinamento com a presença humana, denominado *Human in the loop* (HITL). Estes métodos podem ser utilizadas para contornar o ruído dos rótulos, presente no banco de dados do *NIH Clinical Center* (BALTRUSCHAT et al., 2019), como erros ou ambiguidades na classificação das imagens, o que pode levar a resultados imprecisos do modelo. Nesse sentido, a técnica de HITL pode ser utilizada para identificar e corrigir esses erros. A presença de um profissional permite intervir e corrigir a rede em passos críticos, refinando o conjunto de dados de entrada, além de permitir a avaliação e interpretação dos resultados produzidos pelo modelo, garantindo que eles sejam relevantes e úteis para a tarefa em questão.

Reduzir a duração da etapa de treinamento da rede possibilita realizar mais testes no algoritmo, utilizando outros parâmetros com o objetivo de melhorar o desempenho da CNN. Além disso, com a redução dos requisitos de *hardware* necessários, é possível

também, utilizar resoluções de imagem maiores, permitindo que mais informações sejam processadas pela rede. Assim, a tabela 3 ilustra os resultados obtidos no conjunto de teste, após treinar os modelos *VGG16*, uma vez que apresentou os melhores resultados na Tabela 1 e Figura 9a; e o modelo *DenseNet121*, por ser uma arquitetura amplamente utilizada na literatura. As redes selecionadas foram treinadas com hiperparâmetros de treinamento e resoluções de imagens de entrada diferentes dos utilizados anteriormente.

Tabela 3 – Métricas do conjunto de teste, organizadas por arquitetura, utilizando hiperparâmetros de resolução e *Batch Size* diferentes.

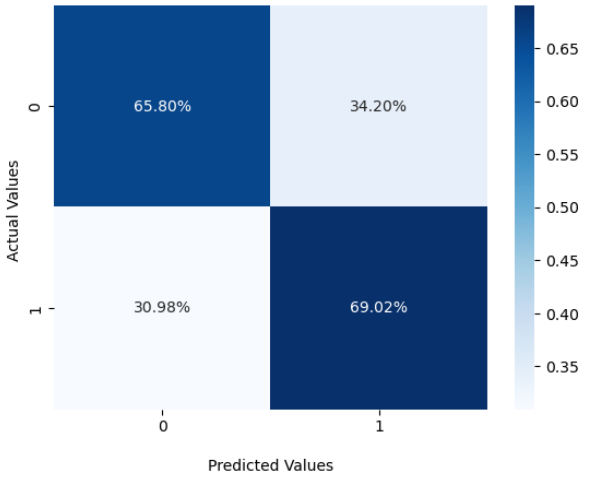
Redes	**Resolução**	***Batch Size***	**Classe**	**Precisão**	***Recall***	***f1-score***
VGG16	128	32	0	**0,71**	**0,67**	**0,69**
			1	**0,64**	**0,68**	**0,66**
VGG16	320	32	0	0,69	0,66	0,68
			1	0,63	0,66	0,65
VGG16	512	16	0	0,68	0,66	0,67
			1	0,62	0,65	0,63
DenseNet121	128	32	0	0,67	0,65	0,66
			1	0,61	0,63	0,62
DenseNet121	320	32	0	0,65	0,63	0,64
			1	0,59	0,61	0,60
DenseNet121	512	16	0	0,65	0,59	0,62
			1	0,57	0,63	0,60

Fim da tabela

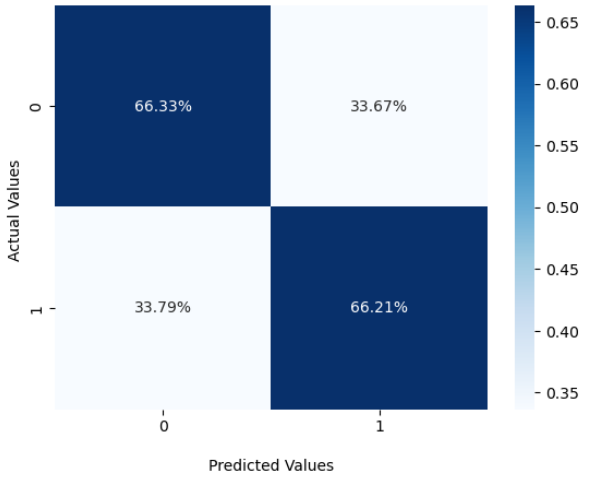
A resolução da imagem tem grande correlação com os resultados do treinamento de uma rede neural convolucional, afetando diretamente a capacidade da CNN de aprender recursos discriminativos. Quanto maior for a resolução da imagem, maior será a quantidade de informações disponíveis para a CNN aprender, o que pode levar a resultados mais precisos e confiáveis. Nesse sentido, Baltruschat et al. (2019) destaca que imagens com maiores resoluções podem melhorar o desempenho da rede na detecção de pequenas estruturas, que poderiam indicar certas patologias, sinalizadas pela presença de nódulos e massas de diferentes dimensões. Outra afirmação similar foi proposta por Bressem et al.

(2020), em relação a pequenos pneumotórax[2], que podem ser escondidos com a redução da escala das imagens.

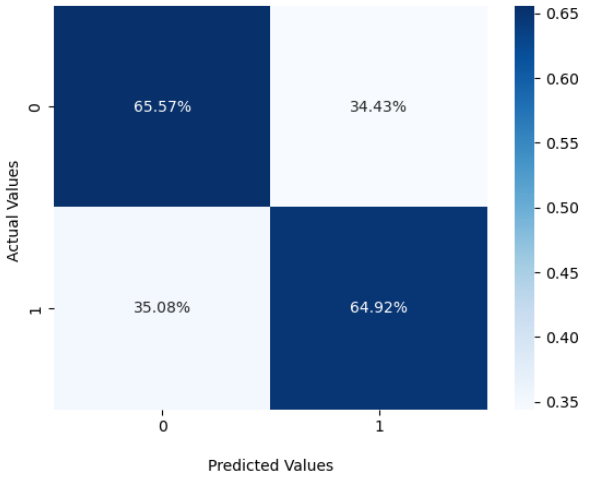
A resolução é um importante hiperparâmetro durante o treinamento da rede, porém nem sempre o aumento deste hiperparâmetro melhora os resultados. Resoluções muito altas de imagens podem trazer informações irrelevantes para o modelo, o que pode tornar o treinamento mais lento e complexo, afetando negativamente a capacidade da CNN de generalizar e fazer previsões precisas em novos conjuntos de dados. Isso pode ser observado ao analisar a tabela 3, onde o aumento da resolução das imagens piorou o desempenho da rede. As figuras 10a -10f representam as matrizes de confusão do conjunto de teste para as rede da tabela 3, que permitem calcular as métricas dispostas na mesma tabela.



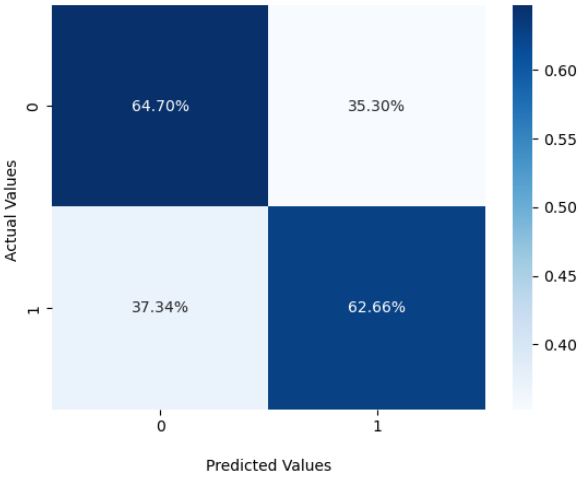
(a) VGG16
128px & 32*Batch Size*



(b) VGG16
320px & 32*Batch Size*



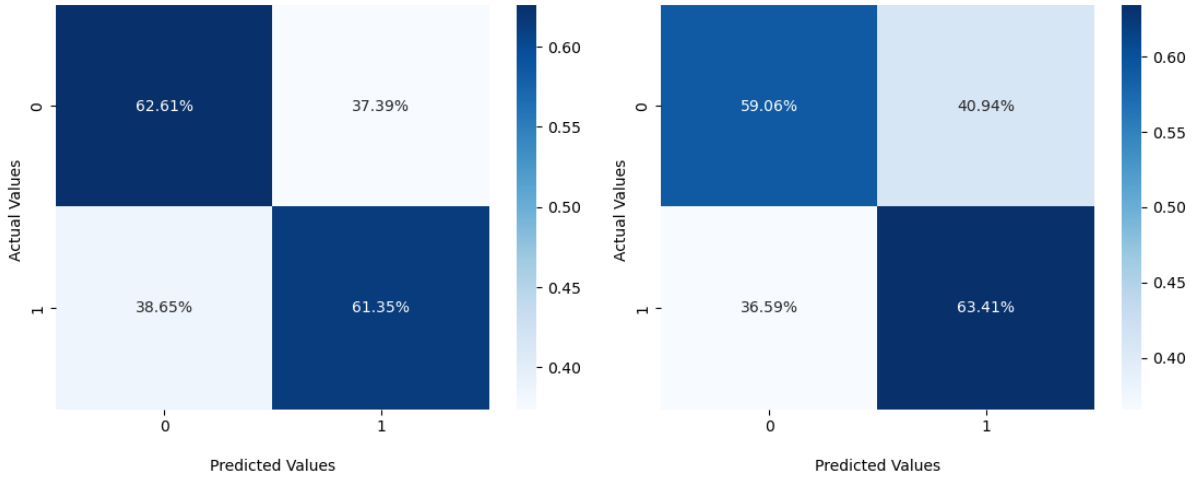
(c) VGG16
512px & 16*Batch Size*



(d) DenseNet121
128px & 32*Batch Size*

Figura 10 – Matrizes de confusão das arquiteturas treinadas utilizando hiperparâmetros de treinamento e um *Batch Size* diferentes dos anteriores. Fonte: autoria própria.

[2] Quando há a presença de ar no espaço pleural, isto é, na membrana interna do tórax.



(e) DenseNet121 320px & 32*Batch Size*

(f) DenseNet121 512px & 16*Batch Size*

Figura 10 – Matrizes de confusão das arquiteturas treinadas utilizando hiperparâmetros de treinamento e um *Batch Size* diferentes dos anteriores. Fonte: autoria própria.

Os resultados da tabela 3 e das figuras 10a - 10f demonstram ganhos irrelevantes nas métricas analisadas ao aumentar a resolução da imagem de entrada, apontando o melhor desempenho na rede *VGG16*, com uma resolução de 128px.

Ao analisar os resultados expostos pela tabela 3 e pelas figuras 10a - 10c, é possível perceber uma grande similariedade nos dados. Isso ocorre devido a limitação de tamanho presente na rede VGG16 e identificada posteriormente. A rede VGG16 foi originalmente proposta para classificar imagens com uma resolução de 224 pixels. Dessa forma, ao utilizar imagens com resoluções maiores que 224 pixels, a rede redimensiona as imagens novamente para a resolução de entrada da arquitetura. Assim, é possível determinar a causa dos resultados para a arquitetura VGG16 terem sido tão próximos utilizando outros hiperparâmetros de resolução.

O aumento na quantidade de parâmetros a serem aprendidos pela rede aumenta a necessidade de RAM-GPU, ocasionando a diminuição do *Batch Size* de 32 para 16 amostras. Além disso, o tempo de treinamento das redes tende a aumentar, o que diminui a possibilidade de testar diferentes hiperparâmetros, a fim de encontrar melhores resultados de desempenho das redes. Nesse sentido, a tabela 4 apresenta os tempos de processamento de cada arquitetura, com base nos conjuntos de treinamento e validação utilizados para os modelos propostos.

Tabela 4 – Duração do treinamento dos modelos avaliados, utilizando hiperparâmetros de resolução e um *Batch Size* diferentes.

Redes	**Resolução**	***Batch Size***	**Duração**		
			Época	**Média**	**Total**
VGG16	128	32	103s - 171s	146.70s	3h 22m 56s
VGG16	320	32	110s - 176s	160.76s	3h 35m 24s
VGG16	512	16	166s - 211s	196.54s	3h 55m 51s
DenseNet121	128	32	102s - 167s	143.29s	3h 03m 53s
DenseNet121	320	32	107s - 165s	153.72s	3h 12m 09s
DenseNet121	512	16	137s - 187s	167.56s	3h 21m 04s

Fim da tabela

A tabela 4 demonstra um aumento no tempo médio de cada época proporcional ao aumento da resolução, como esperado. O tempo total de treinamento pode variar entre cada modelo, uma vez em que este é impactado pelo número de épocas, a depender da convergência da rede e dos critérios de parada definidos.

A partir disso, surge a necessidade de uma comparação entre os resultados alcançados pelo trabalho e demais pesquisas similares. Entretanto, essa comparação é dificultosa, uma vez que o presente trabalho buscou explorar os algoritmos de *Deep Learning* como uma ferramenta de auxilio ao radiologista, podendo ser praticada na triagem médica ao classificar os dados entre duas classes possíveis: pacientes doentes ou saudáveis. Já os estudos citados na seção 2.6, os autores buscaram classificar os dados para laudar a condição específica de cada paciente. Em situações nas quais o objetivo é triar muitos pacientes e identificar rapidamente sua condição, isso não é útil, uma vez que seriam necessárias mais avaliações e acompanhamentos médicos. A triagem possibilitaria esta investigação mais direcionada a partir dos resultados da IA, além de auxiliar o profissional em determinar a necessidade de exames mais invasivos.

Contudo, realizando uma comparação direta com outros estudos, pode-se observar uma certa proximidade dos resultados encontrados neste trabalho com os da literatura. As métricas variam entre as diversas patologias classificadas. O trabalho de Xiaosong Wang et al. (2017), em seu melhor modelo de rede, encontrou o AUC[3] mais alto de 0,835 para Edema e o mais baixo de 0,609 para Infiltração torácica, possuindo um AUC médio

[3] Área Sob a Curva: Métrica utilizada para avaliar o desempenho de modelos de classificação binária.

de 0,738. Já o estudo de Yao et al. (2017) alcançou, na melhor rede desenvolvida, os valores de AUC de 0,914 para Hernia e 0,695 para Infiltração torácica, apresentando um AUC médio de 0,798; no entanto esse valor cai para 0,762 na ausência de patologias. A pesquisa de Baltruschat et al. (2019) obteve os valores de 0,937 para Hernia e 0,694 para Infiltração, com uma média de 0,806 caindo para 0,727 na ausência de doenças. Neste trabalho, utilizando os pesos da *ImageNet*, o melhor AUC encontrado foi para a rede *VGG16* (Fig. 9a), com o valor de 0,675. Apesar do resultado não superar as outras pesquisas citadas, existem alguns fatores que influenciam esse resultado.

O primeiro fator a ser observado é que, nesta pesquisa, os parâmetros e configurações dos algoritmos provavelmente não foram selecionados de forma tão precisa quanto em outros trabalhos. O foco deste estudo é a comparação de diferentes arquiteturas, em contrapartida à otimização de uma única rede específica. Manter os parâmetros constantes nos modelos também pode ter afetado certas arquiteturas mais do que outras, diminuindo a comparabilidade entre as redes avaliadas. Apesar deste ponto negativo, observa-se a mesma limitação em demais projetos de comparação, como em Baltruschat et al. (2019).

O banco de imagens utilizado, *ChestX-ray14*, foi classificado utilizando um algoritmo de rotulagem dos dados. Dessa forma, é possível assumir um erro nos rótulos do conjunto, identificado por Baltruschat et al. (2019) de 10%, sendo relativamente alto. Além disso, as 112.120 radiografias podem ser insuficientes para o modelo ser capaz de generalizar os dados. Nesse sentido, os resultados e o desempenho das redes podem melhorar com o emprego de técnicas de pré-processamento dos dados, visando eliminar o erro dos rótulos, e metodologias de aumento dos dados, como *data augmentation*[4], para fornecer mais experiência para o modelo.

A base de dados da *ImageNet* é um conjunto de imagens bastante diverso, com 1000 classes possíveis para suas amostras e mais de um milhão de dados. Esta ampla variedade de objetos e cenários, difere significativamente das imagens radiográficas de tórax em diversas características, como por exemplo o padrão de cor rgb presente nos dados *ImageNet* em contraposição a escala de cinza nas imagens radiológicas (BRESSEM et al., 2020; BALTRUSCHAT et al., 2019). Assim, as redes pré-treinadas com pesos da *ImageNet* podem ter dificuldade em extrair as informações relevantes para a classificação binária dos pacientes de acordo com a presença ou ausência de doenças no tórax. Nesse sentido, a seção 4.3 aborda a adoção de estratégias específicas para ajustar os pesos da rede de forma mais eficiente.

[4] Técnica de processamento de imagens que consiste em criar novas amostras de treinamento a partir das existentes, por meio da aplicação de transformações como rotações, *zooms* e reflexões.

4.3 Análise dos resultados obtidos com pesos iniciados aleatoriamente

A inicialização de redes neurais com pesos aleatórios é uma prática comum na área de *Machine Learning*. A abordagem atribui valores aleatórios aos pesos da rede que são ajustados durante o treinamento. A vantagem deste método é a possibilidade de encontrar pesos que melhor se adéquem ao objetivo do modelo. Dessa forma, esta seção do estudo analisa os resultados obtidos ao inicializar a CNN com valores aleatórios.

Os testes foram realizados configurando a rede para inicializar com pesos aleatórios, que são ajustados durante o treinamento. Além disso, o modelo foi definido para a escala de cinza, o que não era possível ao utilizar os pesos da *ImageNet*. Dessa forma, é possível diminuir a complexidade da CNN, reduzindo o tempo necessário de treinamento. Os resultados das métricas utilizadas são apresentadas na Tabela 5, referente ao conjunto de teste.

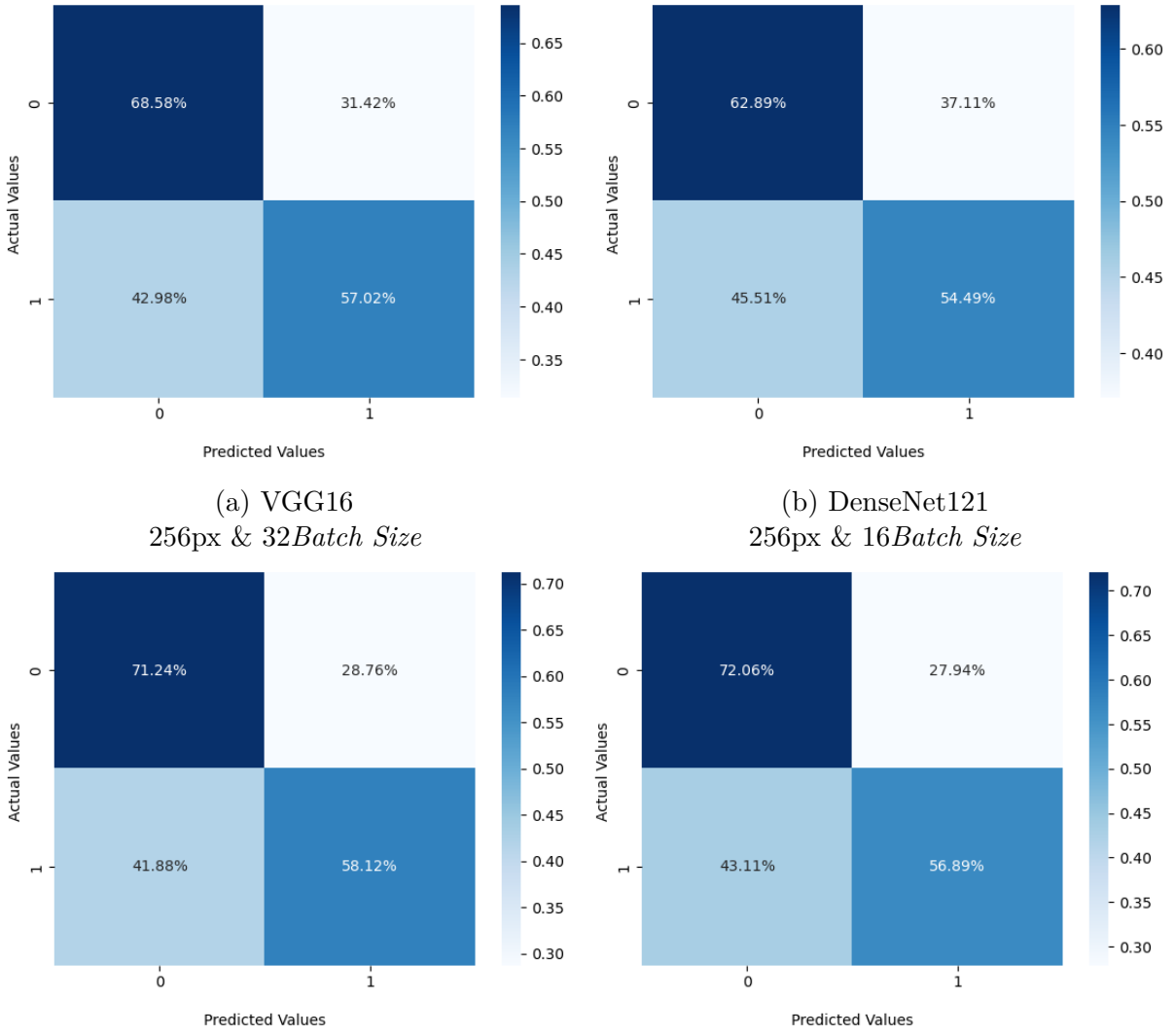
Tabela 5 – Métricas do conjunto de teste, organizadas por arquitetura e inicializadas com pesos aleatórios.

Redes	**Resolução**	***Batch Size***	**Classe**	**Precisão**	***Recall***	***f1-score***
VGG16	256	32	0	0,65	0,69	0,67
			1	0,61	0,57	0,59
DenseNet121	256	16	0	0,61	**0,72**	0,66
			1	0,59	0,47	0,52
AlexNet	256	32	0	**0,66**	0,71	**0,69**
			1	**0,64**	**0,58**	**0,61**
AlexNet	320	32	0	**0,66**	**0,72**	**0,69**
			1	**0,64**	0,57	0,60

Fim da tabela

Analisando os resultados da tabela, é possível observar uma melhora de ***Recall*** na identificação de pacientes saudáveis, representados pelo valor 0 na coluna de ***Classe***, em relação aos resultados da sessão 4.2. Na tabela, os melhores resultados foram destacados em negrito, havendo uma predominância na arquitetura *AlexNet*. Esse cenário indica, novamente, que redes mais rasas são mais eficientes para a tarefa proposta.

As figuras 11a - 11d representam as matrizes de confusão do conjunto de teste para as rede da tabela 5. As matrizes permitem calcular as métricas dispostas na mesma tabela.



(a) VGG16
256px & 32*Batch Size*

(b) DenseNet121
256px & 16*Batch Size*

(c) AlexNet
256px & 32*Batch Size*

(d) AlexNet
320px & 32*Batch Size*

Figura 11 – Matrizes de confusão das arquiteturas treinadas, inicializando a rede com pesos aleatórios. Fonte: autoria própria.

A partir da análise das matrizes, é possível constatar uma melhora na especificidade do modelo ao treinar a rede com pesos aleatórios. No entanto, esse ganho custou significativamente a sensibilidade, diminuindo a taxa de Verdadeiros Positivos. Dessa forma, é possível dizer que ao inicializar a rede com pesos aleatórios há um ganho de especificidade, porém ao utilizar os pesos da *ImageNet* o ganho é de sensibilidade. Isso ocorre pois os pesos adquiridos pelos modelos ao treinar a classificação de relações complexas, como a do banco de dados da *ImageNet*, podem auxiliar a rede a identificar anomalias presentes nas imagens.

A inicialização da rede com pesos aleatórios traz a necessidade de ajustar os pesos para encontrar os melhores resultados. Esse fato tende a aumentar o tempo de treinamento necessário para a convergência das redes. No entanto, ao definir o modelo para a escala de cinza, as dimensões das redes diminuem, tendendo a reduzir o tempo de treinamento. Nesse sentido, a tabela 6 apresenta os tempos de processamento de cada arquitetura, com base nos conjuntos de treinamento e validação utilizados para as arquiteturas propostas.

Tabela 6 – Duração de treinamento dos modelos avaliados, inicializando a rede com pesos aleatórios.

Redes	**Resolução**	***Batch Size***	**Duração**		
			Época	**Média**	**Total**
VGG16	256	32	154s - 166s	156,12s	2h 32m 23s
DenseNet121	256	16	129s - 156s	134,23s	1h 47m 58s
AlexNet	256	32	123s - 148s	129,28s	1h 40m 37s
AlexNet	320	32	127s - 150s	129,59s	1h 43m 14s

Fim da tabela

A partir da tabela, é possível observar que o tempo de treinamento da rede *AlexNet* é inferior às demais redes. Isso permite testar mais hiperparâmetros e aumentar a resolução das imagens de entrada, possibilitando avaliar a arquitetura *AlexNet* com uma resolução de 320px. Assim, foi possível encontrar o melhor resultado de especificidade, de 72,06% (fig. 11d), porém com resultados baixos de sensibilidade.

O trabalho de Pham et al. (2019) constatou que obter uma boa pontuação de AUC para todas as doenças usando uma única CNN é particularmente difícil. Isso se deve ao fato do desempenho da classificação variar de acordo com as arquiteturas de rede utilizadas. Nesse contexto, os autores propuseram a utilização de um conjunto de modelos de *Deep Learning*, explorando o melhor desempenho de cada arquitetura. Nesse contexto, uma solução similar pode ser proposta para este trabalho, explorando os melhores resultados de cada rede avaliada, a fim de encontrar um algoritmo mais eficiente.

5 CONCLUSÃO

O trabalho apresentado neste estudo se propôs a comparar diferentes arquiteturas de redes neurais convolucionais (CNN) para a tarefa de classificação de radiografias de tórax entre pacientes doentes e saudáveis. Embora dos resultados obtidos não tenham sido superiores aos da literatura, foi possível verificar que redes mais rasas, como a *VGG16* e a *AlexNet*, apresentam melhores desempenhos para essa tarefa.

Os resultados permitem comparar diferentes *backbones* e identificar as melhores arquiteturas para futuros estudos direcionados. Os melhores resultados de sensibilidade e especificidade encontrados, de 69,02% (fig. 10a) e 72,06% (fig. 11d) respectivamente, podem facilitar o processo de triagem, no qual o objetivo é identificar os pacientes que necessitam de atenção e avaliações médicas mais detalhadas. No entanto, faz-se necessário realizar mais estudos a partir dos resultados apresentados, visando confirmar quais redes realmente apresentaram os melhores resultados, como por exemplo a realização de validação cruzada e o aumento do número de casas decimais, em trabalhos futuros.

O objetivo da pesquisa foi comparar os diferentes modelos de arquitetura de redes neurais convolucionais. Portanto, os parâmetros e configurações dos algoritmos foram mantidos constantes entre cada modelo, indicando que provavelmente não foram selecionados de forma tão precisa quanto em outros trabalhos. Assim, o estudo não focou em otimizar uma única rede específica, visando alcançar os melhores resultados de classificação. Contudo, a pesquisa direciona estes estudos ao identificar as características presentes nas arquiteturas com os melhores desempenhos, fomentando o desenvolvimento de um algoritmo robusto e eficiente, inovando a forma como as imagens radiográficas são analisadas.

Entretanto, algumas limitações foram identificados no decorrer do estudo, como o ruído presente nos rótulos da base de dados utilizada, o que pode ter afetado a qualidade dos resultados. A utilização de técnicas, como o *"human in the loop"* podem garantir mais precisão e confiabilidade nas classificações. Além disso, a quantidade limitada de dados de treinamento pode ter comprometido a generalização do modelo, tornando necessário o aumento da quantidade de dados disponíveis, incorporando outros *Datasets* como o *CheXpert*. Destaca-se, ainda, a implementação de técnicas para a criação de novas amostras, como o *Data Augmentation*, em estudos futuros. Adicionalmente a isso, em estudos futuros é necessário utilizar métodos de extração de características, com o objetivo de entender como o modelo está processando as informações e identificando padrões nos dados, ajudando a interpretação e explicação dos resultados do modelo.

REFERÊNCIAS

ASLANI, S.; JACOB, J. Utilisation of deep learning for COVID-19 diagnosis. *Clinical Radiology*, v. 78, n. 2, p. 150–157, 2023. ISSN 0009-9260. DOI: https://doi.org/10.1016/j.crad.2022.11.006. Disponível em: https://www.sciencedirect.com/science/article/pii/S0009926022007188. Citado 1 vez na página 20.

ASSMUS, Alexi. Early history of X rays. *Beam Line*, v. 25, n. 2, p. 10–24, 1995. Disponível em: https://www.slac.stanford.edu/pubs/beamline/25/2/25-2-assmus.pdf. Citado 1 vez na página 22.

AUNG, Yuri YM; WONG, David; TING, Daniel SW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *British medical bulletin*, v. 139, n. 1, 2021. Disponível em: https://www.researchgate.net/profile/David-Wong-49/publication/354006928_The_promise_of_artificial_intelligence_A_review_of_the_opportunities_and_challenges_of_artificial_intelligence_in_healthcare/links/616bf4d4b90c51266254fe8c/The-promise-of-artificial-intelligence-A-review-of-the-opportunities-and-challenges-of-artificial-intelligence-in-healthcare.pdf. Citado 1 vez na página 32.

BA, Jimmy Lei; KIROS, Jamie Ryan; HINTON, Geoffrey E. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016. Disponível em: https://arxiv.org/abs/1607.06450. Citado 1 vez na página 31.

BALTRUSCHAT, Ivo M et al. Comparison of deep learning approaches for multi-label chest X-ray classification. *Scientific reports*, Springer, v. 9, n. 1, p. 1–10, 2019. Citado 10 vezes nas páginas 33, 35, 38, 43, 44, 48.

BRESSEM, Keno K. et al. Comparing different deep learning architectures for classification of chest radiographs. *Nature*, v. 1, n. 1, p. 1–16, 2020. DOI: 10.1038/s41598-020-70479-z. Disponível em: https://www.nature.com/articles/s41598-020-70479-z. Citado 8 vezes nas páginas 19, 20, 34, 38, 40, 42, 44, 48.

BUSBY, Lindsay P.; COURTIER, Jesse L.; GLASTONBURY, Christine M. Bias in Radiology: The How and Why of Misses and Misinterpretations. *Radiological Society of North America*, dez. 2017. DOI: 10.1148/rg.2018170107. Disponível em: https://doi.org/10.1148/rg.2018170107. Citado 2 vezes nas páginas 19, 23.

BUSHONG, Stewart C. *Radiologic science for technologists e-book: physics, biology, and protection*. Elsevier Health Sciences, 2020. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=fV4MEAAAQBAJ&oi=fnd&pg=PP1&dq=Radiologic+Science+for+Technologists:+Physics,+Biology,+and+Protection&ots=Evfp25f9Sn&sig=66uJ2KTpjj-EbcZdps8ogQS7rC4. Citado 1 vez na página 22.

CYBENKO, George. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals, and systems (MCSS)*, Springer, v. 2, n. 4, p. 303–314, 1989. Citado 1 vez na página 26.

ELHARROUSS, Omar et al. Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv:2206.08016*, 2022. Disponível em: https://arxiv.org/abs/2206.08016. Citado 1 vez na página 28.

FREITAS, Marcelo Baptista de; YOSHIMURA, Elisabeth Mateus. Levantamento da distribuição de equipamentos de diagnóstico por imagem e freqüência de exames radiológicos no Estado de São Paulo. *Radiologia Brasileira*, SciELO Brasil, v. 38, p. 347–354, 2005. Disponível em: http://www.rb.org.br/detalhe_artigo.asp?id=1447&idioma=Portugues. Citado 0 vez na página 23.

GÉRON, Aurélien. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc.", 2022. Citado 7 vezes nas páginas 28, 29, 31.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. MIT Press, 2016. Citado 7 vezes nas páginas 23, 24, 26–30.

GREENSPAN, Hayit; GINNEKEN, Bram van; SUMMERS, Ronald M. Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE*, 2016. DOI: 10.1109/TMI.2016.2553401. Disponível em: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7463094. Citado 5 vezes nas páginas 19, 20, 27, 32, 37.

HAYKIN, Simon. *Redes neurais: princípios e prática*. Bookman Editora, 2001. Citado 3 vezes nas páginas 24, 25, 27.

HOLSTE, Gregory et al. Long-Tailed Classification of Thorax Diseases on Chest X-Ray: A New Benchmark Study, p. 22–32, ago. 2022. DOI: 10.1007/978-3-031-17027-0_3. Disponível em: https://link.springer.com/chapter/10.1007/978-3-031-17027-0_3. Citado 1 vez na página 19.

HUANG, Gao et al. Densely connected convolutional networks. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. 2017. P. 4700–4708. Disponível em: http://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html. Citado 1 vez na página 32.

IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: PMLR. INTERNATIONAL conference on machine learning. 2015. P. 448–456. Disponível em: http://proceedings.mlr.press/v37/ioffe15.html. Citado 1 vez na página 31.

IRVIN, Jeremy et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: 01. PROCEEDINGS of the AAAI conference on artificial intelligence. 2019. v. 33, p. 590–597. Disponível em: https://ojs.aaai.org/index.php/AAAI/article/view/3834. Citado 2 vezes nas páginas 33, 35.

KELLY, Barry. The chest radiograph. *The Ulster medical journal*, Ulster Medical Society, v. 81, n. 3, p. 143, 2012. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3632825/. Citado 1 vez na página 23.

KER, Justin et al. Deep learning applications in medical image analysis. *Ieee Access*, IEEE, v. 6, p. 9375–9389, 2017. DOI: 10.1109/ACCESS.2017.2788044. Disponível em: https://ieeexplore.ieee.org/abstract/document/8241753. Citado 2 vezes nas páginas 19, 20.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. Citado 1 vez na página 26.

LECUN, Yann; BENGIO, Yoshua et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, Citeseer, v. 3361, n. 10, p. 1995, 1995. Disponível em: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e26cc4a1c717653f323715d751c8dea7461aa105. Citado 3 vezes na página 29.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado 1 vez na página 26.

LECUN, Yann; BOSER, Bernhard et al. Backpropagation applied to handwritten zip code recognition. *Neural computation*, MIT Press, v. 1, n. 4, p. 541–551, 1989. Disponível em: https://ieeexplore.ieee.org/abstract/document/6795724/. Citado 1 vez na página 28.

LIU, Tianming; SIEGEL, Eliot; SHEN, Dinggang. Deep Learning and Medical Image Analysis for COVID-19 Diagnosis and Prediction. *Annual Review of Biomedical Engineering*, v. 24, p. 179–201, jun. 2022. DOI: 10.1146/annurev-bioeng-110220-012203. Disponível em: https://doi.org/10.1146/annurev-bioeng-110220-012203. Citado 1 vez na página 20.

MOTA, Mariana Regina Ferreira. *REDE CONVOLUCIONAL PARA VERIFICAÇÃO BIOMÉTRICA BASEADA EM EEG EM AMBIENTE MULTITAREFA*. 2021. Graduação – UNIVERSIDADE FEDERAL DE OURO PRETO, Ouro Preto, Brasil. Citado 0 vez na página 25.

MOULD, RF. Röntgen and the discovery of X-rays. *The British journal of radiology*, The British Institute of Radiology, v. 68, n. 815, p. 1145–1176, 1995. Disponível em: https://www.birpublications.org/doi/abs/10.1259/0007-1285-68-815-1145. Citado 1 vez na página 22.

NEVES, Davi. *Deep Learning e Industria 4.0: Introdução Às Redes Neurais Convolucionais*. DECOM - UFOP, out. 2021. http://www2.decom.ufop.br/imobilis/deep-learning-e-a-industria-4-0-introducao-as-redes-neurais-convolucionais/. Acessado: 2023-02-16. Citado 1 vez nas páginas 29, 30.

OLIVEIRA, Glória Maria de et al. Revisão sistemática da acurácia dos testes diagnósticos: uma revisão narrativa. *Revista do Colégio Brasileiro de Cirurgiões*, SciELO Brasil, v. 37, p. 153–156, 2010. Disponível em: https://www.scielo.br/j/rcbc/a/XTbJWZPVvCYTgG6tDNBbpKn/abstract/?lang=pt. Citado 3 vezes na página 42.

PHAM, Hieu H. et al. *Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels*. arXiv, 2019. DOI: 10.48550/ARXIV.1911.06475. Disponível em: https://arxiv.org/abs/1911.06475. Citado 2 vezes nas páginas 33, 51.

PINTO, Antonio; BRUNESE, Luca. Spectrum of diagnostic errors in radiology. *World Journal of Radiology*, v. 2, n. 10, p. 377–383, out. 2010. DOI: 10.4329/wjr.v2.i10.377. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2999012/pdf/WJR-2-377.pdf. Citado 1 vez na página 19.

RAGHU, Maithra et al. Transfusion: Understanding Transfer Learning with Applications to Medical Imaging. *CoRR*, abs/1902.07208, 2019. arXiv: 1902.07208. Disponível em: http://arxiv.org/abs/1902.07208. Citado 2 vezes nas páginas 34, 35.

RAJPURKAR, Pranav et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS medicine*, Public Library of Science San Francisco, CA USA, v. 15, n. 11, e1002686, 2018. Disponível em: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002686. Citado 4 vezes nas páginas 32, 35, 37.

REN, Shaoqing et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: CORTES, C. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2015. v. 28. Disponível em: https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf. Citado 1 vez na página 24.

RÖNTGEN, Wilhelm Conrad. On a new kind of rays. *Science*, American Association for the Advancement of Science, v. 3, n. 59, p. 227–231, 1896. Disponível em: https://www.science.org/doi/10.1126/science.3.59.227. Citado 1 vez na página 22.

SHI, Xingjian et al. Deep Matrix Factorization Models for Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 31, n. 9, p. 1745–1757, 2019. Disponível em: https://www.ijcai.org/Proceedings/2017/0447.pdf. Citado 1 vez na página 24.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M.; FURHT, Borko. Deep Learning applications for COVID-19. *Journal of Big Data*, jan. 2021. DOI: 10.1186/s40537-020-00392-9. Disponível em: https://doi.org/10.1186/s40537-020-00392-9. Citado 1 vez na página 20.

SMITH-BINDMAN, Rebecca; MIGLIORETTI, Diana L.; LARSON, Eric B. Rising use of diagnostic medical imaging in a large integrated health system. *Health Affairs*, v. 27, n. 6, p. 1491–1502, 2008. DOI: 10.1377/hlthaff.27.6.1491. Disponível em: https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.27.6.1491. Citado 2 vezes nas páginas 19, 20.

ULYANOV, Dmitry; VEDALDI, Andrea; LEMPITSKY, Victor. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. Disponível em: https://arxiv.org/abs/1607.08022. Citado 1 vez na página 31.

VASWANI, Ashish et al. Attention is All you Need. In: GUYON, I. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Citado 1 vez na página 24.

WANG, Xiaosong et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. 2017. P. 2097–2106. Disponível em: http://openaccess.thecvf.com/content_cvpr_2017/html/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.html. Citado 6 vezes nas páginas 32, 38, 47.

WANG, Yingying et al. The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences*, MDPI, v. 10, n. 5, p. 1897, 2020. Disponível em: https://www.mdpi.com/661772. Citado 4 vezes nas páginas 27, 28.

YAO, Li et al. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501*, 2017. Disponível em: https://arxiv.org/abs/1710.10501. Citado 5 vezes nas páginas 32, 35, 38, 48.