

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Explorando Características Radiômicas Profundas em Modelos de Aprendizado Profundo por meio de Imagens Médicas de COVID-19

Márcus Vinícius Lobo Costa

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Márcus Vinícius Lobo Costa

Explorando Características Radiômicas Profundas em Modelos de Aprendizado Profundo por meio de Imagens Médicas de COVID-19

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Agma Juci Machado Traina

USP – São Carlos
Março de 2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

C837e Costa, Márcus Vinícius Lobo
Explorando Características Radiômicas Profundas em
Modelos de Aprendizado Profundo por meio de Imagens
Médicas de COVID-19 / Márcus Vinícius Lobo Costa;
orientadora Agma Juci Machado Traina. -- São
Carlos, 2024.
98 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2024.

1. Aprendizado Profundo. 2. Radiômica. 3.
Aprendizado Profundo baseado em Radiômica. 4.
Características Profundas. I. Traina, Agma Juci
Machado , orient. II. Título.

Márcus Vinícius Lobo Costa

**Exploiting Deep Radiomics Features in Deep Learning
Models using COVID-19 Medical Images**

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Agma Juci Machado Traina

**USP – São Carlos
March 2024**

*Para todos os nortistas que um dia sonharam em ter a oportunidade de estudar na melhor
Universidade do Brasil, Universidade de São Paulo.*

AGRADECIMENTOS

Gostaria primeiramente de agradecer à minha família. Em especial minha mãe Klelnyr, minha vó Walnir, meu tio Kleber e minha irmã Ana Livia, por todo apoio e incentivo incondicional em todos os momentos e decisões. Sem vocês nada seria possível. Amo vocês!

À minha orientadora Profa. Dra. Aγμα Juci Machado Traina, pela oportunidade de tê-la como orientadora deste trabalho. Agradeço pela confiança, conselhos, paciência, e principalmente por sua contribuição em minha formação acadêmica e profissional ao longo deste período, que juntamente com o Prof. Dr. Caetano Traina Junior me acolheram muito bem.

Agradeço ao grupo de pesquisa de Bases de Dados e Imagens (GBDI), por toda a estrutura fornecida, onde sempre foi cultivado um ambiente pró-discussão de diversos assuntos pertinentes à pesquisa, além de sempre ajudarem durante este projeto de Mestrado. Em especial agradeço: Lucas Santiago, Erikson Aguiar, Igor Alberte, Junior Clementino, Mirela Cazzolato, André Moreira, João Pedro, Afonso Matheus, Mariana e Willian Dener, tornando a caminhada mais leve. Meu muito obrigado! Agradeço também aos amigos externos que incentivaram e me apoiaram nesta trajetória: Uthant Vicentin, Gustavo Balbinot, Wan Song, Fátima Eduarda, Sherlon Almeida, Guilherme Montroni, Luan Souza, Angelo Mendes e Marcos Gôlo.

Por fim, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)¹, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)², e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)³ pelo apoio financeiro para a execução deste trabalho.

A todos que, de alguma forma, contribuíram com a realização deste trabalho.

Obrigado!

¹ Processo: 152760/2021-0.

² Código de financiamento 001.

³ Processos: 016/17078-0 e 2020/07200-9.

*Tu já se perguntou o preço pra entrar na História
Ou reparou como realizar nos revigora?
Lembrou que a cada noite alguma nova mãe que chora?
Só quem não tentou voar não sabe como dói ir embora
(Atravessei São Paulo – Duzz)*

RESUMO

COSTA, M. V. L. **Explorando Características Radiômicas Profundas em Modelos de Aprendizado Profundo por meio de Imagens Médicas de COVID-19**. 2024. 98 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

A análise de imagens médicas desempenha um papel importante na tomada de decisão de médicos especialistas. Especificamente na detecção da COVID-19, as abordagens de Aprendizado Profundo (AP) e a abordagem Radiômica obtiveram resultados promissores no decorrer dos últimos anos. No entanto, os resultados por meio de aprendizado profundo são difíceis de interpretar/visualizar e a abordagem radiômica elenca etapas sucessivas, como: a aquisição de imagens, o processamento de imagens, a segmentação, a extração de características e a análise. Neste contexto, este trabalho de Mestrado propõe a abordagem DEELE-Rad (*Deep Learning-based Radiomics*), que utiliza a integração de aprendizado profundo e a abordagem radiômica, com intuito de auxiliar na detecção de COVID-19. O DEELE-Rad utiliza modelos de aprendizado profundo para extrair 100, 128, 200, e 300 características radiômicas profundas relevantes para avaliar a COVID-19. Utilizando várias fontes de imagens, com 392 exames representativos de radiografia de tórax (raio-X). Desta maneira, evitam-se etapas sucessivas da radiômica utilizando aprendizado profundo e aplicando ajuste-fino nas redes da VGG-16, ResNet50V2 e a DenseNet201. O DEELE-Rad considera um conjunto de algoritmos de Aprendizado de Máquina (AM) para validar os resultados, fornecendo um comitê (*ensemble*) de classificadores para classificação de COVID-19. Com resultados experimentais do DEELE-Rad, foi possível observar que o DEELE-Rad obteve um melhor desempenho quando utilizou-se 300 características radiômicas profundas provenientes da DenseNet201, tornando-o mais preciso em até 8,8% comparada aos modelos de aprendizado profundo de ponta a ponta. O DEELE-Rad têm como proposta a análise e avaliação visual de características radiômicas profundas, a fim de apresentar interpretabilidade e explicabilidade tornando-o mais explicável. De maneira geral, o DEELE-Rad pode aumentar o desempenho da classificação binária em um cenário real. Por fim, destacamos que o DEELE-Rad pode ser adaptado para criar outras ferramentas de radiômica baseadas em aprendizado profundo e pode fornecer perspectivas de interpretabilidade e explicabilidade de radiômicas profundas por meio dos modelos de aprendizado profundo.

Palavras-chave: Aprendizado Profundo, Radiômica Profunda, Características Profundas, Radiômica, Imagem Médica.

ABSTRACT

COSTA, M. V. L. **Exploiting Deep Radiomics Features in Deep Learning Models using COVID-19 Medical Images**. 2024. 98 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Medical image analysis plays an essential role in aiding physicians in decision-making. Specifically in detecting COVID-19, deep learning (DL) and radiomics approaches have achieved promising results in recent years. However, deep learning results are hard to interpret/visualize, and the radiomic approach encompasses successive steps, such as image acquisition, image processing, segmentation, feature extraction, and analysis. In this context, this Master's thesis proposes the DEELE-Rad (*Deep Learning-based Radiomics*) approach, which integrates deep learning and the radiomic approaches, aiding in detecting COVID-19. DEELE-Rad uses deep learning models to extract 100, 128, 200, and 300 deep radiomic features relevant to assessing COVID-19. Multiple image sources, with 392 representative chest X-ray examinations. This way, successive steps of radiomics are avoided using deep learning and fine-tuning on the VGG-16, ResNet50V2, and DenseNet201 networks. The DEELE-Rad considers a set of Machine Learning (ML) algorithms to further validate our results, providing an ensemble learning model to detect COVID-19. With experimental results from DEELE-Rad, it was possible to observe that DEELE-Rad performed better when using 300 deep radiomic features from DenseNet201, making it more accurate by up to 8.8% compared to end-to-end deep learning models. DEELE-Rad aims to visually analyze and evaluate deep radiomic features to make them more interpretable and explainable. Overall, DEELE-Rad can increase binary classification performance in a real scenario. Finally, we highlight that DEELE-Rad can be adapted to create other deep learning-based radiomics tools and can provide insights into the interpretability and explainability of deep radiomics through deep learning models.

Keywords: Deep Learning, Deep Radiomics, Deep Features, Radiomics, Medical Imaging.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplos de algumas modalidades de imagens médicas.	28
Figura 2 – Fluxograma da abordagem radiômica. Etapas necessárias no processo de extração de características radiômicas em cenários clínicos.	29
Figura 3 – Processo de extração de características. Em (1) é abordado o tipo de aquisição de imagens; (2) extração das características da aquisição das imagens; e em (3) os vetores representativos de cada imagem.	32
Figura 4 – Exemplos de histogramas de níveis de cor RGB.	32
Figura 5 – Alguns exemplos de texturas com características diferentes.	34
Figura 6 – Exemplos das matrizes específicas que atribuem informação sobre a distribuição espacial dos valores de <i>pixels</i> na imagem: (a) <i>Gray Level Cooccurrence Matrix</i> (GLCM), (b) <i>Gray Level Run Length Matrix</i> (GLRLM), (c) <i>Gray Level Size Zone Matrix</i> (GLSZM) e (d) <i>Neighboring Gray Tone Difference Matrix</i> (NGTDM).	35
Figura 7 – Estrutura de uma Rede Neural Artificial com duas camadas.	36
Figura 8 – Funcionamento do Comitê com os classificadores de aprendizado de máquina utilizando voto majoritário.	38
Figura 9 – A arquitetura original de uma rede neural convolucional introduzida por LeCun <i>et al.</i> (1989).	39
Figura 10 – Funcionamento da camada de convolução.	40
Figura 11 – Arquitetura de uma rede VGG-16.	42
Figura 12 – Módulo residual.	43
Figura 13 – Arquitetura de uma rede DenseNet com três blocos densos. As camadas entre os dois blocos adjacentes são chamadas de camadas de transição e alteram os tamanhos dos mapas de características por meio da convolução e subamostragem.	43
Figura 14 – Representação de aprendizagem por transferência baseado na arquitetura da rede Inception. Esquerda: Uma rede de iniciação, modelo base. Direita: Uma nova rede baseada nas camadas pré-treinadas da rede inicial, mas com as últimas camadas modificadas por novas que serão treinadas para um novo problema.	44
Figura 15 – Abordagem radiômica baseada em modelos de aprendizado profundo. Apresentação de uma breve arquitetura de CNN simplificada na qual é realizada a extração de características radiômicas profundas.	46

Figura 16 – Funcionamento do <i>Gradient-weighted Class Activation Mapping (Grad-CAM)</i> .	47
Figura 17 – Fluxograma empregado da abordagem DEELE-Rad (<i>Deep Learning-based Radiomics</i>). Uma abordagem radiômica baseada em aprendizado profundo para detecção de COVID-19 em imagens de raio-X utilizando comitê de classificadores.	59
Figura 18 – Imagens de COVID-19 e Normal (saudável) com a aplicação do CLAHE.	61
Figura 19 – Fluxograma da metodologia de validação do DEELE-Rad.	66
Figura 20 – Exemplo de uma matriz de confusão aplicada à nossa metodologia de validação.	67
Figura 21 – Matrizes de confusão para os melhores modelos no cenário com 100 e 200 características radiômicas profundas, respectivamente.	71
Figura 22 – Matriz de confusão para o pior modelo no cenário com 300 características radiômicas profundas.	71
Figura 23 – Visualização do processo de extração de características radiômicas profundas no modelo da VGG-16 com 100 características radiômicas profundas. Cada bloco de convolução da rede apresenta diferentes níveis de cinza, forma e textura para um caso de COVID-19. As amostras com a borda na cor azul, magenta e ciano representam o nível de cinza, forma e a textura, respectivamente.	80
Figura 24 – Visualização do processo de extração de características radiômicas profundas no modelo da VGG-16 com 100 características radiômicas profundas. Cada bloco de convolução da rede apresenta diferentes níveis de cinza, forma e textura para um caso Normal. As amostras com a borda na cor azul, magenta e ciano representam o nível de cinza, forma e a textura, respectivamente.	81
Figura 25 – Avaliação visual com Grad-CAM para um caso de COVID-19, mostrando o comportamento do Grad-CAM em relação a um exemplo extraído de cada processo de convolução. Observa-se que o caso de COVID-19 é rotulado como 0 e Normal (saudável) é 1.	82
Figura 26 – Parametrização do Grad-CAM para um caso de COVID-19, apresentado a mesma perspectiva de interpretabilidade e explicabilidade com diferentes mapas de calor. Mapas de calor utilizados: Jet, Reds e RdGy reverso, respectivamente. A barra de cor varia entre 0 e 1. O valor 1 representa a intensidade mais forte identificada no Grad-CAM, e 0 a mais fraca.	82

LISTA DE TABELAS

Tabela 1 – Visão geral dos trabalhos selecionados para abordagem de radiômica baseada em aprendizado profundo e clássica em comparação com a nossa abordagem proposta.	53
Tabela 2 – Sumarização dos dados extraídos dos trabalhos relacionados.	55
Tabela 3 – Base de dados com a divisão para treino e teste.	59
Tabela 4 – Modelos de aprendizado profundo com pesos pré-treinados pela ImageNet.	62
Tabela 5 – Tipos de camadas utilizados em nosso ajuste-fino.	62
Tabela 6 – Métricas de desempenho utilizando 100 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.	69
Tabela 7 – Métricas de desempenho utilizando 128 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.	69
Tabela 8 – Métricas de desempenho utilizando 200 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.	70
Tabela 9 – Métricas de desempenho utilizando 300 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.	70
Tabela 10 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 100 características radiômicas profundas mais relevantes.	73
Tabela 11 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 128 características radiômicas profundas mais relevantes.	74
Tabela 12 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 200 características radiômicas profundas mais relevantes.	75
Tabela 13 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 300 características radiômicas profundas mais relevantes.	76
Tabela 14 – Média das métricas de desempenho do DEELE-Rad.	77
Tabela 15 – Comparação da métrica de desempenho da acurácia e a precisão com os modelos de aprendizado profundo em relação ao DEELE-Rad, em cada respectivo cenário de experimentação.	78

LISTA DE ABREVIATURAS E SIGLAS

ACE	Automated Concept-based Explanation
ACoS	Automatic COVID screening
AFS-DF	Adaptive Feature Selection Deep Forest
AHE	Adaptive Histogram Equalization
AI	Artificial Intelligence
AUC	Area Under Curve
CAD	Computer-aided Diagnosis
CLAHE	Contrast Limited Adaptive Histogram Equalization (CLAHE)
CNN	Convolutional Neural Network
ConvAE	Convolutional Deep Autoencoder
COVID-19	COronaVirus Disease 2019
CT	Computer Tomography
CXR	Chest X-Ray
DL	Deep Learning
ExtraTrees	Extremely Randomized Trees
FOS	First Order Statistical
GAN	Generative Adversarial Network
GLCM	Gray Level Cooccurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GMM	Gaussian Mixture Methods
GNB	Gaussian Naive Bayes
HOG	Histogram of Oriented Gradients
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
KNN	K-nearest Neighbor
LatLRR	Latent Low-Rank Representation
LBP	Local Binary Pattern
LIME	Local Interpretable Model-agnostic Explanations
LPP	Locality Preserving Projections
LR	Logistic Regression

LRP	Layerwise Relevance Propagation
LRR	Low-rank Representation
ML	Machine Learning
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NGTDM	Neighboring Gray Tone Diference Matrix
PET	Positron Emission Tomography Scans
ResNet	Residual Network
RF	Random Forest
RGB	Red, Green e Blue
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
SAE	Stacked Autoencoder
SARS	Severe Acute Respiratory Syndrome
SHAP	SHapely Additive exPlanations
SOF	Second Order Features
SVM	Support Vector Machine
TL	Transfer Learning
US	Ultrasound
VGG	Visual Geometry Group

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Contextualização e Motivação	23
1.2	Definição do Problema	24
1.3	Contribuições do Trabalho	25
1.4	Organização da Dissertação	26
2	FUNDAMENTAÇÃO TEÓRICA	27
2.1	Considerações Iniciais	27
2.2	Modalidades de Imagens Médicas	28
2.3	Abordagem Radiômica	29
2.4	Aprendizado de Máquina	36
2.5	Aprendizado Profundo	38
2.6	Radiômica baseada em Aprendizado Profundo	45
2.7	Interpretabilidade e Explicabilidade	46
2.8	Considerações Finais	47
3	TRABALHOS RELACIONADOS	49
3.1	Considerações Iniciais	49
3.2	Técnicas de Radiômica Clássica	50
3.3	Técnicas de Aprendizado Profundo para Radiômica	51
3.4	Sumário dos Trabalhos Selecionados	52
3.5	Considerações Finais	54
4	DEELE-RAD: EXPLORANDO CARACTERÍSTICAS RADIÔMICAS PROFUNDAS	57
4.1	Considerações Iniciais	57
4.2	A Abordagem DEELE-Rad	58
4.3	Base de dados e Pré-processamento	59
4.4	Extração de Radiômica Profunda	60
4.5	Comitê de Classificadores	62
4.6	Análise de Radiômica Profunda	63
4.7	Considerações Finais	63
5	RESULTADOS E DISCUSSÕES	65

5.1	Considerações Iniciais	65
5.2	Metodologia de Validação	65
5.3	Avaliação dos Modelos de Aprendizado Profundo	68
5.4	Avaliação dos Classificadores de Aprendizado de Máquina	72
5.5	Avaliação Visual para Análise de Radiômica Profunda	79
5.6	Discussões dos Resultados	83
5.7	Considerações Finais	84
6	CONCLUSÕES	85
6.1	Pontos Positivos e Limitações	85
6.2	Principais Contribuições	86
6.3	Trabalhos Publicados	87
6.3.1	<i>Publicação Principal</i>	87
6.3.2	<i>Publicações em Colaboração</i>	87
6.3.3	<i>Outras Atividades</i>	88
6.4	Trabalhos Futuros	88
	REFERÊNCIAS	91

INTRODUÇÃO

1.1 Contextualização e Motivação

A humanidade enfrentou surtos de coronavírus que podem causar a condição clínica da síndrome respiratória aguda grave (do inglês, *Severe Acute Respiratory Syndrome* - SARS). Entre esses surtos, pode-se citar o SARS-CoV (DROSTEN *et al.*, 2003), MERS-CoV (GROOT *et al.*, 2013) e recentemente o SARS-CoV-2 (CHENG *et al.*, 2020; HUANG *et al.*, 2020). Com a rápida disseminação do vírus SARS-CoV-2, surgiu uma nova doença chamada coronavírus 2019 (do inglês, *COronaVIrus Disease 2019* - COVID-19) em todo o mundo (HUANG *et al.*, 2020).

Os meios comumente utilizados para detecção da doença são a transcrição reversa à reação em cadeia da polimerase (do inglês, *Reverse Transcription-Polymerase Chain Reaction* - RT-PCR), tomografia computadorizada (do inglês, *Computer Tomography* - CT) e a radiografia de tórax ou raio-X (do inglês, *Chest X-Ray* - CXR). O raio-X tem sido utilizada como um dos métodos para detectar COVID-19 e avaliar sua gravidade ou monitorar seu prognóstico em hospitais devido ao baixo custo, baixa dose de radiação e ampla acessibilidade (LITMANOVICH *et al.*, 2020). Os achados observados em raio-X são a consolidação periférica¹ e opacidade em vidro fosco² (AI *et al.*, 2020; YOON *et al.*, 2020; XIE *et al.*, 2020; WONG *et al.*, 2020; GUAN *et al.*, 2020).

Embora ainda seja um desafio à precisa detecção e diferenciação da pneumonia de COVID-19 dos outros tipos de pneumonias, os métodos de aprendizado profundo (do inglês, *Deep Learning* - DL) e a radiômica podem auxiliar ou mesmo otimizar na previsão da COVID-19. Na prática, pode ser benéfico ao auxiliar médicos e radiologistas a realizar um diagnóstico ágil,

¹ De acordo com Meirelles (2020) a consolidação é identificada quando todo o órgão embranquece, impossibilitando a visualização de estruturas brônquicas e vasculares.

² De acordo com Meirelles (2020) a opacidade em vidro fosco é identificada quando o pulmão se encontra mais branco do que o normal (hiperatenuante), onde é possível enxergar vasos pulmonares que estão atrás da opacidade.

essencialmente quando os profissionais e o sistema de saúde estão sobrecarregados (ARDAKANI *et al.*, 2020; HEIDARI *et al.*, 2021).

Os sistemas de Auxílio ao Diagnóstico por Computador (do inglês, *Computer-aided Diagnosis* - CAD) apresentam potencial para auxiliar na solução do problema mencionado acima como análise quantitativa de alto rendimento (HU *et al.*, 2022). Uma possível abordagem é a utilização do CAD para análise dos exames de imagem baseado em radiômica. Isso consiste em primeiramente extrair características radiômicas como biomarcadores de um exame de imagem e posteriormente utilizar técnicas clássicas de extração de características ou utilizar algoritmos de aprendizado de máquina (do inglês, *Machine Learning* - ML), como apresentado nos trabalhos de Wang *et al.* (2021b), Moura *et al.* (2020) e Ferreira *et al.* (2021). Embora as características radiômicas sejam comumente utilizadas na análise de imagens médicas, o diagnóstico da COVID-19 ainda é limitado nos trabalhos representativos de CAD baseados em radiômica e aprendizado profundo (HU *et al.*, 2022).

1.2 Definição do Problema

As aplicações baseadas em aprendizado profundo na análise de imagens médicas, usufruem dos hiperparâmetros de um determinado tipo de arquitetura de aprendizado profundo gerados sem intervenção do conhecimento humano. A natureza caixa preta do aprendizado profundo apoiado ao CAD inibe sua interpretabilidade, prejudicando a interpretação de indicadores ou a análise de clínicos ou radiologistas (HU *et al.*, 2022).

A literatura apresenta métodos baseados em aprendizado de máquina e aprendizado profundo aplicados à detecção e diagnóstico da doença de COVID-19 de modo a criar um sistema CAD. Esses sistemas podem incluir um pré-processamento para a melhoria na qualidade da imagem (RAHMAN *et al.*, 2021) ou a remoção das partes anatômicas (como cabeça, pescoço, braços e diafragma) (CARDENAS *et al.*, 2021). Outra aplicação é a identificação de biomarcadores radiômicos para COVID-19 (FERREIRA *et al.*, 2021), diferenciação das síndromes respiratórias agudas utilizando apenas técnicas de aprendizado de máquina, redes neurais convolucionais (HEIDARI *et al.*, 2021; MARQUES; AGARWAL; DÍEZ, 2020), ou modelos que integram ambas as abordagens, denominados sistemas híbridos (SHANKAR; PERUMAL, 2021; TURKOGLU, 2021).

Desta maneira, a abordagem **Radiômica Clássica** extrai dados quantitativos de exames de imagens médicas (GILLIES; KINAHAN; HRICAK, 2016). Esta abordagem requer sucessivos passos como: aquisição de imagens, processamento de imagens, segmentação, extração de características e análise. Portanto, a intervenção humana de um especialista é necessária (KOÇAK *et al.*, 2019). Em contraste, a abordagem de **Radiômica baseada em Aprendizado Profundo** automaticamente identifica e extrai dados de alta dimensão por meio das arquiteturas de Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network* - CNN). Modelos de aprendi-

zado profundo baseado em CNN podem auxiliar no processo de tomada de decisão ou mesmo incorporar métodos explicativos.

Com intuito de fornecer subsídios para tomada de decisão, a abordagem radiômica baseada em aprendizado profundo é um desafio no contexto de aplicações para exames de imagens médicas para COVID-19, em especial para sistemas CAD. Este trabalho de Mestrado foca na extração de características radiômicas profundas e na avaliação visual no contexto médico. Como premissa, o trabalho aborda a seleção de melhores características radiômicas profundas e como podemos tornar essas características explicáveis em modelos de aprendizado profundo.

1.3 Contribuições do Trabalho

Neste trabalho de Mestrado foi concebida uma nova abordagem para extração e avaliação visual de características radiômicas em modelos de aprendizado profundo. Esta contribuição foi validada por pares e reportada em [Costa et al. \(2023\)](#). A abordagem desenvolvida, chamada de DEELE-Rad (*Deep Learning-based Radiomics*), utiliza características radiômicas profundas de modelos de aprendizado profundo, assim como provê meios para a avaliação visual dessas características. Ao desenvolver a abordagem DEELE-Rad, buscou-se responder às seguintes Questões de Pesquisa (QP):

- QP1.** *Como as principais características radiômicas extraídas de modelos de aprendizado profundo podem melhorar na predição em imagens de raio-X de COVID-19?*
- QP2.** *Como avaliar o processo de tomada de decisão dos modelos de aprendizado profundo por meio da avaliação visual do processo de extração de características radiômicas?*

A solução empregada no DEELE-Rad possui as premissas de (i) extrair as características radiômicas profundas mais importantes por meio de exames de imagens médicas de raio-X utilizando diferentes modelos/arquiteturas de aprendizado profundo, e (ii) apresentar o processo de tomada de decisão da extração de características por meio da avaliação visual. A abordagem DEELE-Rad possui três contribuições principais:

- I. A extração de características radiômicas profundas (*deep radiomics*) mais relevantes, quando abordamos os modelos de aprendizado profundo (*deep learning*);
- II. A criação de um comitê de classificadores para validação de características radiômicas profundas extraídas em diferentes modelos de aprendizado profundo;
- III. A avaliação visual do comportamento de características radiômicas profundas, de modo a visualizar o processo de extração em uma arquitetura de aprendizado profundo, em especial a VGG-16.

1.4 Organização da Dissertação

Neste capítulo foi apresentada a introdução desta Dissertação de Mestrado, abordando a contextualização, motivação, problemática e contribuições. Os demais capítulos estão organizados da seguinte maneira:

- O [Capítulo 2](#) apresenta os principais conceitos relacionados a esta pesquisa, iniciando com as modalidades de imagens médicas, abordagem radiômica, aprendizado de máquina, aprendizado profundo, a abordagem radiômica baseada em aprendizado profundo e a interpretabilidade e explicabilidade;
- No [Capítulo 3](#) são apresentados os trabalhos relacionados e suas discussões, evidenciando as problemáticas e lacunas;
- O [Capítulo 4](#) detalha a abordagem DEELE-Rad;
- O [Capítulo 5](#) apresenta os resultados com o DEELE-Rad;
- Por fim, o [Capítulo 6](#) apresenta as conclusões desta Dissertação e os trabalhos futuros a serem realizados.

FUNDAMENTAÇÃO TEÓRICA

2.1 Considerações Iniciais

A área médica enfrenta uma constante digitalização das informações geradas pela rotina clínica, em especial dos exames de imagens. Com isso, à medida que dados médicos ficam disponíveis em formato digital, surgem novos *softwares* para analisá-los, uma vez que inúmeras novas demandas clínicas surgem, conseqüentemente novos métodos manuais ou mesmo automáticos são disponibilizados ao meio médico. Ao mesmo tempo, pesquisadores em Inteligência Artificial (do inglês, *Artificial Intelligence* - AI) alcançaram resultados robustos por meio de métodos de aprendizado de máquina e aprendizado profundo, oferecendo novos meios para auxiliar ou mesmo solucionar problemas advindos de rotinas clínicas complexas, e apoiar nas tomadas de decisões dos especialistas (TIMMEREN *et al.*, 2020b).

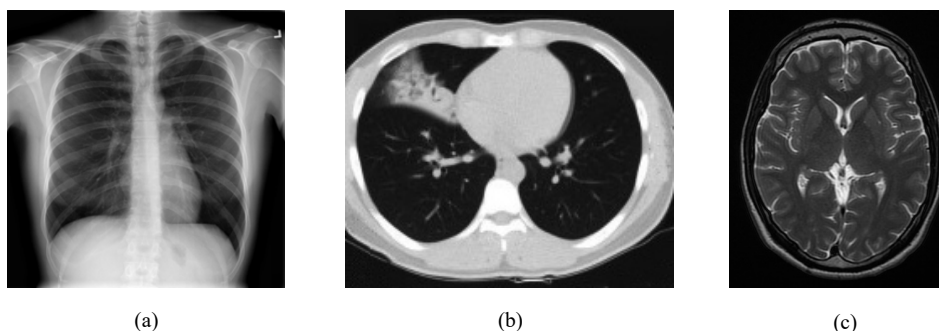
Uma das rotinas clínicas amplamente apoiadas por novas tecnologias digitais é a geração de exames por imagens, que podem oferecer informações do paciente relativas a possíveis doenças, desempenhando um papel fundamental no processo de tomada de decisão e na prática clínica. Além disso, esses exames descreverem com precisão fisiológica e anatômica as características internas dos pacientes e geralmente não são invasivos (quando não há necessidade de contraste ou outro tipo de introdução de material/equipamento). Estas tecnologias permitem que o especialista compreenda apropriadamente doenças raras e complexas (ZHAO *et al.*, 2020a; BHATTACHARYA *et al.*, 2021).

Neste capítulo serão apresentados os conceitos fundamentais para o entendimento desta pesquisa, iniciando com as modalidades de imagens médicas, abordagem radiômica, aprendizado de máquina, aprendizado profundo, e sobre o foco de contribuição principal deste trabalho, a abordagem radiômica para os modelos de aprendizado profundo, por fim sobre a interpretabilidade e explicabilidade.

2.2 Modalidades de Imagens Médicas

O termo imagem médica é a apresentação visual de regiões e órgãos de pacientes obtidas por dispositivos de aquisição de dados, podendo empregar ou não elementos de órgãos, ou tecidos internos do corpo, de modo a diagnosticar ou detectar a presença, ou mesmo evolução de doenças (FARHAT; SAKR; KILANY, 2020). A Figura 1 ilustra algumas das modalidades de imagens médicas digitais, como o Raio-X 1(a), Tomografia Computadorizada (TC) 1(b) e ressonância magnética (do inglês, *Magnetic Resonance Imaging* - MRI) 1(c). Além dessas, há outras técnicas como o ultra-som (do inglês, *Ultrasound* - US) e tomografia por emissão de pósitrons (do inglês, *Positron Emission Tomography Scans* - PET) (CHENG, 2019).

Figura 1 – Exemplos de algumas modalidades de imagens médicas.



Fonte – (CHOWDHURY *et al.*, 2020; RAHMAN *et al.*, 2021; NOVAKSEAN, 2015; ZHAO *et al.*, 2020b).

Esta Dissertação de Mestrado visou utilizar o raio-X como modalidade de imagem principal. Entre as principais vantagens da aplicação e análise referentes às imagens de raio-X estão: a sua disponibilidade; relação custo benefício; efeito de radiação inferior; velocidade de aquisição dos dados (GOEL; YADAV; SINGH, 2016; LITMANOVICH *et al.*, 2020). Geralmente, o raio-X é uma imagem bidimensional produzida por ondas eletromagnéticas que atravessam o corpo e são absorvidas uniformemente pelos tecidos e ossos (GOEL; YADAV; SINGH, 2016).

As modalidades de aquisição de imagens apresentam diferenças na forma adequada de utilizá-las. O PET é importante para monitoramento e diagnóstico em relação aos tecidos e órgãos (GOEL; YADAV; SINGH, 2016), enquanto para a detecção de nódulos pulmonares¹, a tomografia computadorizada é mais sensível (ZHANG *et al.*, 2018). Para Candemir e Antani (2019), as imagens de raio-X são convencionais para o diagnóstico de distúrbios pulmonares e cardiotorácicos, além de emitirem menos radiações do que a TC, e serem acessíveis para regiões com menos recursos. Na seção a seguir discute-se a abordagem radiômica, uma vez que com tal técnica é possível extrair informações intrínsecas das modalidades de imagens citadas anteriormente.

¹ De acordo com Zhang *et al.* (2018) o nódulo pulmonar é definido como uma opacidade arredondada e o seu diâmetro pode medir até 3 cm no peito. Além disso, um nódulo pulmonar pode ser categorizado com base no seu tamanho (grande ou pequeno), localização (bem circunscrita, justa-pleural e justa-vascular), forma (tipo bola e irregular), textura interna (sólida, parcialmente sólida e não-sólida).

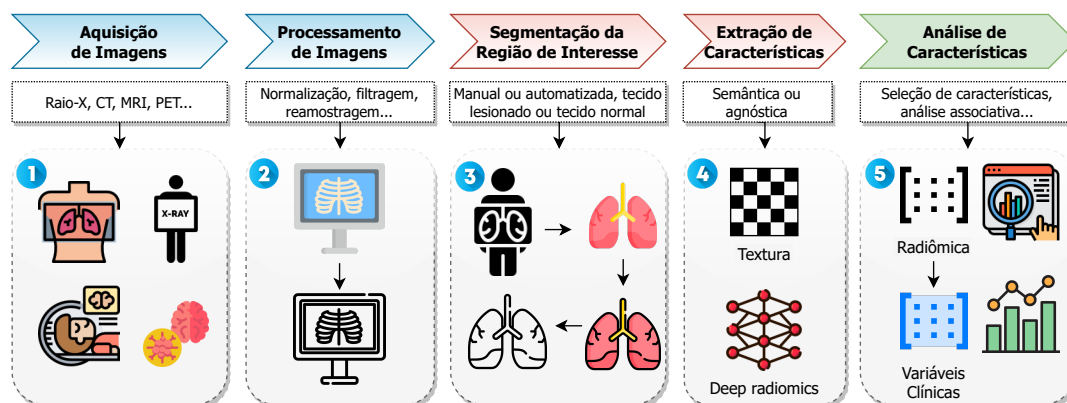
2.3 Abordagem Radiômica

A abordagem radiômica é uma técnica de extração quantitativa de características de diferentes modalidades de imagem médica (LAMBIN *et al.*, 2012; GILLIES; KINAHAN; HRICAK, 2016). Utiliza algoritmos computacionais para extrair e analisar características quantitativas (LAFATA *et al.*, 2022). O intuito é utilizar essas características no processo de tomada de decisão, dado que a heterogeneidade nos dados/*pixels* das imagens são imperceptíveis na inspeção visual de um especialista (VARGHESE *et al.*, 2019; KOÇAK *et al.*, 2019). Nesse contexto, a radiômica nos permite compreender finitas condições médicas. Para compreensão da abordagem radiômica neste projeto, a subdividimos em dois grupos: a *radiômica clássica* e a *radiômica baseada em aprendizado profundo* - que será apresentada na [Seção 2.6](#).

Radiômica Clássica

Nesta abordagem, as características radiômicas podem ser predefinidas ou calculadas por exames de imagem manualmente, ou semi-automaticamente (LAFATA *et al.*, 2022). A Figura 2 ilustra o fluxograma da abordagem radiômica em cenários clínicos propostos por Scapicchio *et al.* (2021). É possível observar que a radiômica é subdivida em alguns passos: (1) o passo primordial é aquisição dessas imagens, podendo variar em diferentes modalidades de imagens; (2) é abordado o processamento de imagem necessário em determinado contexto, preparando-as para as próximas etapas; (3) adquiridas as imagens pré-processadas, o próximo passo é realizar a segmentação da região de interesse, em que pode ser uma lesão no tecido, dependendo da aplicação; (4) dada a região de interesse esse passo visa a extração das características radiômicas, que podem ser baseadas em características estatísticas, filtragem ou morfológicas, que acabam gerando um espaço de características de alta dimensionalidade e em (5) essas características são classificadas por meios dos algoritmos de aprendizado de máquina ou é realizada uma análise estatística.

Figura 2 – Fluxograma da abordagem radiômica. Etapas necessárias no processo de extração de características radiômicas em cenários clínicos.



Fonte – Adaptado de Scapicchio *et al.* (2021).

Comumente, os pacientes apresentam diferenças na forma (*shape*), tamanho e na textura de determinado achado (GILLIES; KINAHAN; HRICAK, 2016; VIAL *et al.*, 2018). Desta maneira, a abordagem radiômica pode oferecer uma melhor precisão dos modelos provenientes da IA, com a capacidade de manusear dados quantitativos comparados com os métodos estatísticos tradicionais (TIMMEREN *et al.*, 2020b).

Para Koçak *et al.* (2019), a abordagem radiômica pode ser classificada em semântica ou agnóstica. As características semânticas são utilizadas por radiologistas, por descreverem as lesões, volume, morfologia e diâmetro dos achados. Já as características agnósticas são descritores quantitativos extraídos matematicamente por meio computacional. As principais definições relacionadas com a abordagem radiômica são as seguintes, de acordo com Timmeren *et al.* (2020a):

- **Biomarcadores:** são características do corpo que podem ser mensuradas, funcionando como um indicador mensurável da gravidade da presença de uma dada doença. Por exemplo, a pressão sanguínea do corpo. Um biomarcador pode ter natureza química, física ou biológica, podendo ser mensurado funcionalmente, fisiologicamente, bioquimicamente, celularmente ou molecularmente. (WU *et al.*, 2021). Idealmente, um biomarcador deve ser mensurado e permitir reprodutibilidade (ZWANENBURG *et al.*, 2020).
- **Imagem quantitativa:** assim como um biomarcador expressa uma informação derivada, por exemplo, de uma imagem, o termo de imagem quantitativa representa a extração de características quantificáveis de imagens médicas para análises clinicamente relevantes (SULLIVAN *et al.*, 2015).
- **Normalização:** há necessidade de protocolos de imagens padronizados, com intuito de evitar a variabilidade. Os protocolos não padronizados são comuns utilizando a abordagem de radiômica (LAMBIN *et al.*, 2017). Portanto, a variabilidade é notada pelos pacientes, por exemplo, a geometria, que influenciam indiretamente nos níveis de ruídos e a presença de demais artefatos em um exame de imagem (TIMMEREN *et al.*, 2020a).
- **Características radiômicas:** sua essência é a extração de características quantitativas de exames médicos. Essas características dependem dos protocolos utilizados no pré-processamento da imagem e a sua reconstrução. Contudo, as nomenclaturas de características podem variar, além da metodologia, da implementação de algoritmos e de *softwares* para obtê-las (LAMBIN *et al.*, 2017).

Para compreensão da abordagem radiômica nesse projeto, na subseção seguinte serão discutidos os aspectos de extração de características que foi utilizado para o embasamento das características radiômicas.

Extração de Características

Uma imagem digital é uma representação bidimensional de *pixels* que proporciona uma interpretação visual de uma cena ou um objeto para compreensão humana, ou computacional. Esses dados são codificados por uma função $f(x,y)$ onde x e y representam as coordenadas espaciais e a função f indica a proporção de luminância no ponto (x,y) . As imagens podem ser mapeadas em canais de cor, como o *RGB* (*Red, Green and Blue* - Vermelho, Verde e Azul), em que cada canal possui um nível de intensidade em que varia usualmente entre 0 e 255. Um *pixel* é a menor unidade de informação de uma imagem, e é uma representação visual da função $f(x,y)$ (GONZALEZ; WOODS, 2008; SUNDARARAJAN, 2017).

A extração de características é realizada sobre uma determinada cena ou imagem e que visa sintetizar (ou sumarizar) as informações relevantes seguindo um contexto específico. Esse processo acaba gerando um vetor de características que funciona como assinatura da imagem, ou seja, essa informação torna-se relevante para solucionar determinada tarefa computacional (HARALICK; SHANMUGAM; DINSTEN, 1973). Além disso, as técnicas de extração de características são divididas em *locais* e *globais*.

As técnicas de extração de características *globais* extraem informação da imagem inteira não considerando o conteúdo interno ou partes da imagem. Os histogramas são comumente utilizados nessa abordagem, pois podem representar a contagem de cores ou dos níveis de cinza da imagem. Já as técnicas de extração de características locais normalmente extraem informação sobre um processo de segmentação das imagens. Considerando as tarefas computacionais no domínio de imagens, esses extratores podem ser englobados em três grupos: características de cor, forma e textura.

Comumente, os algoritmos de extração de características pertencem a um desses grupos, sendo necessário utilizar simultaneamente diferentes extratores de características, assim gerando um novo vetor de características. Embora esta composição possivelmente melhore a qualidade do resultado, também aumenta significativamente o número de dimensões geradas para o vetor de características que irá representar a imagem.

A Figura 3 ilustra o processo de extração de características de várias imagens, em que pode ser interpretado como um caso especial de transformação e redução de dimensionalidade, já que esses extratores consistem em uma função computacional onde consegue transformar um conjunto de *pixels* da imagem em um conjunto de valores reduzidos, seja ele multidimensional ou não, denominado vetor de características (ou vetor característico da imagem).

Extratores de Cor

O histograma de cores representa a distribuição de cores ou frequência de cada cor existente nos *pixels* de uma imagem. O histograma tornou-se amplamente utilizado por demandar menos recursos computacionais, além de ser invariante à rotação e escala. As imagens coloridas

Figura 3 – Processo de extração de características. Em (1) é abordado o tipo de aquisição de imagens; (2) extração das características da aquisição das imagens; e em (3) os vetores representativos de cada imagem.



Fonte: Elaborada pelo autor.

contém três canais de cor, como já mencionando acima. A Figura 4 apresenta dois histogramas de cor de para o modelo tradicional RGB, em que cada componente do RGB possui um valor entre 0 e 255, pois são imagens de 8 bits de profundidade por canal de cor, por *pixel*. Em 4(a) apresenta-se praticamente as mesmas quantidades de contribuição nos três canais de cor, já em 4(b) há maior prevalência em um dos canais RGB.

Figura 4 – Exemplos de histogramas de níveis de cor RGB.



Fonte: Elaborada pelo autor.

A literatura apresenta algumas técnicas para aplicarmos melhoramentos nos níveis de intensidade em histogramas. Dentre essas técnicas pode-se citar a Equalização de Histograma Adaptativo (do inglês, *Adaptive Histogram Equalization* - AHE) e a Equalização de Histograma Limitado por Contraste (do inglês, *Contrast Limited Adaptive Histogram Equalization* - CLAHE).

A equalização de histogramas e o processamento de imagens é uma tarefa importante para obtermos melhores resultados em domínios de imagens diferentes. O contraste de uma imagem é a escala entre os tons mais luminosos e os tons mais escuros da imagem. Imagens capturadas com iluminação inadequada são comuns, assim como nas capturas de exames de imagens médicas. Assim, discutiremos a seguir as técnicas mencionadas acima.

Equalização de Histograma Adaptativo (AHE)

O AHE calcula diferentes histogramas da mesma imagem com cada um representando uma parte distinta da imagem (PIZER *et al.*, 1987). Isso o difere do histograma que utiliza a imagem toda. Portanto, a ideia central do AHE é melhorar o contraste local e a definição dos contornos das imagens.

Equalização de Histograma Adaptativo Limitado por Contraste (CLAHE)

O CLAHE foi desenvolvido com objetivo de melhorar o contraste de imagens médicas (PISANO *et al.*, 1998). O CLAHE é uma variante do AHE, que opera na superamplificação do contraste para evitar a amplificação de ruídos que o AHE pode originar. Esse diferencial permite controlar o limitador de contraste. O procedimento de limitação do contraste é aplicado a cada vizinhança a partir da qual é derivada uma função de transformação, consequentemente a quantidade de contraste deve ser aumentada.

Extratores de Forma

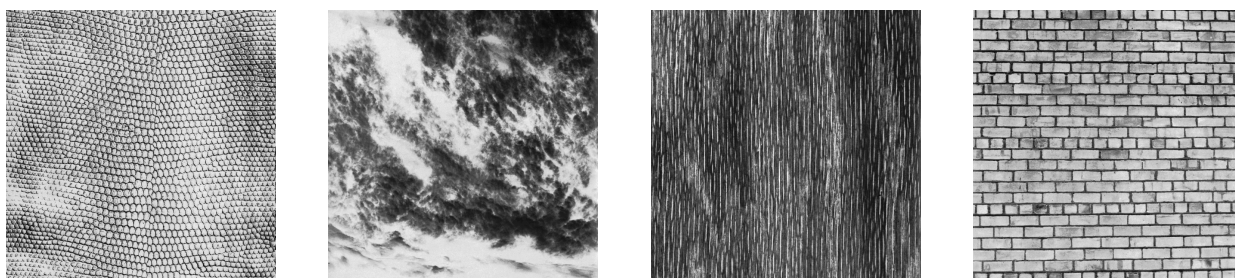
A forma visa restringir uma consulta utilizando determinado espaço ou objeto de uma imagem e, em geral, é representado por uma região de interesse (do inglês, *Region Of Interest - ROI*) que circunscreve um elemento da imagem. As formas de objetos presentes nas imagens são usualmente obtidas após um processo de segmentação de imagem. De acordo com Gonzalez e Woods (2008) uma segmentação necessita favorecer a execução correta de algoritmos de extração de características, podendo ser necessário realizar uma análise aprofundada junto ao especialista com enfoque em determinar quais regiões são mais importantes e em qual pré-processamento adequar essas regiões.

Extratores de Textura

A textura é uma abordagem para a descrição de regiões com intuito de quantificar seu conteúdo de granularidade estrutural. Ademais, não existe uma definição formal para textura, mas intuitivamente esse tipo de descritor nos fornece medidas de propriedades como suavidade, rugosidade e regularidade. As três principais abordagens para extração de características de textura, são: estatística, estrutural e espectral (GONZALEZ; WOODS, 2008).

A análise estatística é caracterizada pela variação dos atributos de intensidade, contraste, granularidade e direção, em que é considerada a distribuição de tons de cinzas e o seu inter-relacionamento. A análise estrutural lida com os arranjos de primitivas de imagens, como a descrição de textura baseada em linhas paralelas espaçadas. A Figura 5 apresenta exemplos de tipos variados de texturas.

Figura 5 – Alguns exemplos de texturas com características diferentes.



Fonte – Adaptado de [Randen \(1999\)](#).

No que tange ao domínio de imagens médicas, a literatura nos mostra que os extratores de textura auxiliam na tomada de decisão nas tarefas de classificação de imagens de raio-X pulmonar, a qual é o domínio utilizado como alvo de experimentação e coleta de resultados para esta pesquisa. Alguns trabalhos que utilizam essas técnicas são [Hu et al. \(2022\)](#), [Wang et al. \(2021a\)](#), [Chaddad, Hassan e Desrosiers \(2022\)](#), [Zhu et al. \(2021\)](#), [Koyuncu e Barstuğan \(2021\)](#), [Chandra et al. \(2021\)](#) e [Ferreira et al. \(2021\)](#). Com intuito de utilizar extratores de características nesse projeto com a abordagem radiômica, serão explicados brevemente alguns a seguir.

First Order Statistics

As características de primeira ordem (do inglês, *First Order Statistics* - FOS) caracterizam a distribuição de ocorrência de cada intensidade de *pixel* em uma imagem, ignorando as relações espaciais. Podemos citar como exemplo, a média, variância, e a entropia ([GONZALEZ; WOODS, 2008](#); [SCAPICCHIO et al., 2021](#)).

Second Order Features

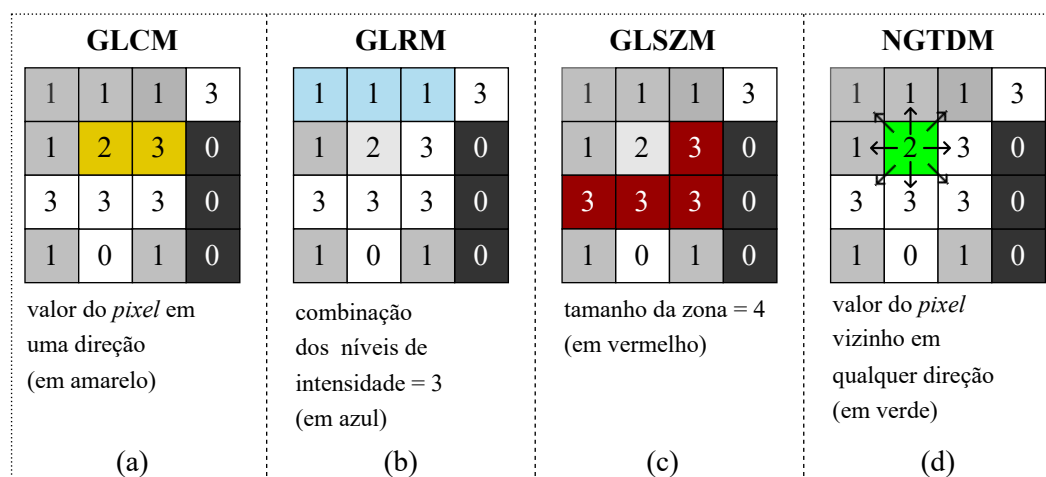
As características de segunda ordem (do inglês, *Second Order Features* - SOF) foram introduzidas por [Haralick, Shanmugam e Dinstein \(1973\)](#) e baseiam-se na distribuição conjunta de probabilidade de *pixels*, apresentando a disposição espacial dos padrões imperceptíveis ao olho humano. A análise é normalmente realizada em dois passos. No primeiro passo é definida uma matriz que atribui a informação sobre a distribuição espacial dos valores dos *pixels*; no segundo são avaliadas as métricas sobre a matriz.

As matrizes comumente utilizadas são: *Gray Level Cooccurrence Matrix* (GLCM), *Gray Level Run Length Matrix* (GLRLM), *Gray Level Size Zone Matrix* (GLSZM), *Neighboring Gray*

Tone Difference Matrix (NGTDM), *Gray Level Dependence Matrix* (GLDM) e a *Local Binary Pattern* (LBP). A Figura 6 ilustra brevemente o comportamento de algumas dessas matrizes.

Na Figura 6(a) a GLCM contém informação estatística de como os pares de *pixels* são distribuídos na imagem. Assim, caracteriza a probabilidade de encontrar dois *pixels* com valores de intensidade de cinza específicos em uma determinada distância e direção. Na 6(b) o GLRM considera a informação estatística de segunda ordem e expressa o número de vezes que a combinação de níveis de intensidades idênticas ou semelhantes ocorrem em dois *pixels* ou mais em uma direção pré-definida na imagem. Em 6(c) GLSZM quantifica as zonas de níveis de cinza numa imagem, definidos como o número de *pixels* ligados que compartilham a mesma intensidade de nível de cinza. Já em 6(d) a NGTDM quantifica a diferença entre um valor de nível de cinza e o valor médio dos seus vizinhos em uma distância estipulada. De maneira geral, a NGTDM captura informações sobre as intensidades de cinza que se alteram de *pixel* para *pixel*, incluindo as transições entre os *pixels*. A GLDM quantifica as dependências do nível de cinza em uma imagem, definidas como números de *pixels* ligados em uma certa distância (adjacentes) que dependem de uma entrada. Já o LBP é um descriptor de textura, introduzido por Ojala, Pietikäinen e Mäenpää (2001), que atribui um rótulo, ou seja, uma numeração binária, a cada *pixel* de uma imagem característica.

Figura 6 – Exemplos das matrizes específicas que atribuem informação sobre a distribuição espacial dos valores de *pixels* na imagem: (a) *Gray Level Cooccurrence Matrix* (GLCM), (b) *Gray Level Run Length Matrix* (GLRLM), (c) *Gray Level Size Zone Matrix* (GLSZM) e (d) *Neighboring Gray Tone Difference Matrix* (NGTDM).



Fonte – Adaptado de Scapicchio *et al.* (2021).

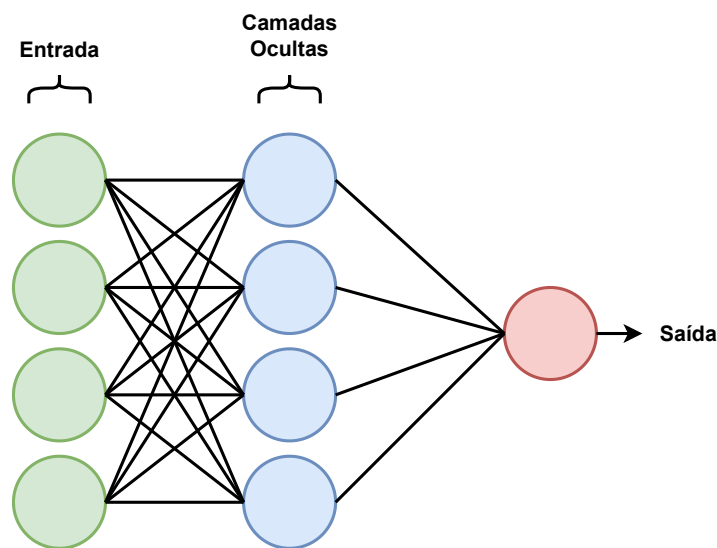
Para Scapicchio *et al.* (2021) essa classificação de características de segunda ordem não é exaustiva, devido à ampla aplicabilidade de técnicas existentes. Portanto, esta abordagem pode nos fornecer pontos de explicabilidade ou interpretabilidade quando aplicados em diferentes domínios de imagens. Na subseção a seguir será apresentado o aprendizado de máquina com o qual podemos utilizar a extração de características por meio da abordagem radiômica.

2.4 Aprendizado de Máquina

Aprendizado de Máquina (AM) é uma área de pesquisa com foco na teoria, desempenho, e nas propriedades dos sistemas e algoritmos de aprendizagem. Aprendizado de máquina apresenta sua interdisciplinaridade entre diferentes áreas de pesquisa, como a IA, teoria da otimização, teoria da informação, estatística, ciência cognitiva, engenharia, matemática, entre outras (QIU *et al.*, 2016). Portanto, aprendizado de máquina aplica-se a quase todos os domínios científicos, assim como na área médica.

Redes Neurais Artificiais (RNAs) são comumente utilizadas em aprendizado de máquina. Sua estrutura básica serve de modelo para diversos algoritmos (HAYKIN, 2009). A Figura 7 exemplifica o funcionamento de uma rede neural separada por camadas. As camadas em cor verde são responsáveis pelos dados de entrada, as camadas na cor azul são responsáveis pelo processamento das camadas ocultas, e por fim a camada de saída onde se obtém a classificação (na cor vermelha).

Figura 7 – Estrutura de uma Rede Neural Artificial com duas camadas.



Fonte: Elaborada pelo autor.

O aprendizado de máquina apresenta em sua organização diferentes tipos de abordagens como: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. O aprendizado supervisionado requer o conjunto de dados de entrada rotulados em que busca-se convergir para o melhor classificador possível. O classificador permite mapear os dados de entrada para as melhores repostas, permitindo também utilizar outros dados de entrada como dados de testes. Diferente do aprendizado supervisionado, o aprendizado não supervisionado não necessita do conjunto de dados rotulados. Assim, o classificador visa identificar os padrões por meio do conjunto de dados. O aprendizado por reforço permite aprender com a resposta recebida por meio das interações em um ambiente dinâmico, como recompensas ou penalidades conforme

as suas ações. Este tipo de aprendizado é comum em tarefas de algoritmos genéticos, comumente utilizados em carros autônomos e robótica (MELLO; PONTI, 2018; HAN; PEI; TONG, 2022). Neste trabalho, utilizaremos apenas o aprendizado supervisionado.

A seguir serão apresentados alguns algoritmos que utilizam essa abordagem, tais como os algoritmos de classificação. Algoritmos de classificação podem ser considerados conforme o seu contexto e os dados utilizados, cada um com suas características e aplicabilidades. Portanto, as aplicações são diversas por sua habilidade de aprender com diferentes exemplos de dados (AGGARWAL *et al.*, 2015).

A técnica de *Support Vector Machine* (SVM) foi proposta por Vapnik (1999) com a ideia principal de mapear os vetores característicos de entrada x em um espaço de alta dimensão ou hiperplano Z por meio de algum mapeamento não linear escolhido. Um hiperplano representa a separação do espaço dimensional em duas regiões.

O *Gaussian Naive Bayes* (GNB) é um classificador baseado no teorema de Bayes para probabilidades condicionais. A ideia principal deste teorema é quantificar a probabilidade condicional aleatória, com as observações conhecidas sobre o valor do outro conjunto de variáveis aleatórias, ou variáveis características (AGGARWAL *et al.*, 2015). O GNB assume uma forma específica da distribuição de probabilidade para cada classe.

Logistic Regression (LR) é uma técnica de análise estatística que prevê um valor contínuo para um ou mais atributos dependentes, ou seja, determina a probabilidade de um evento ocorrer (HAN; PEI; TONG, 2022).

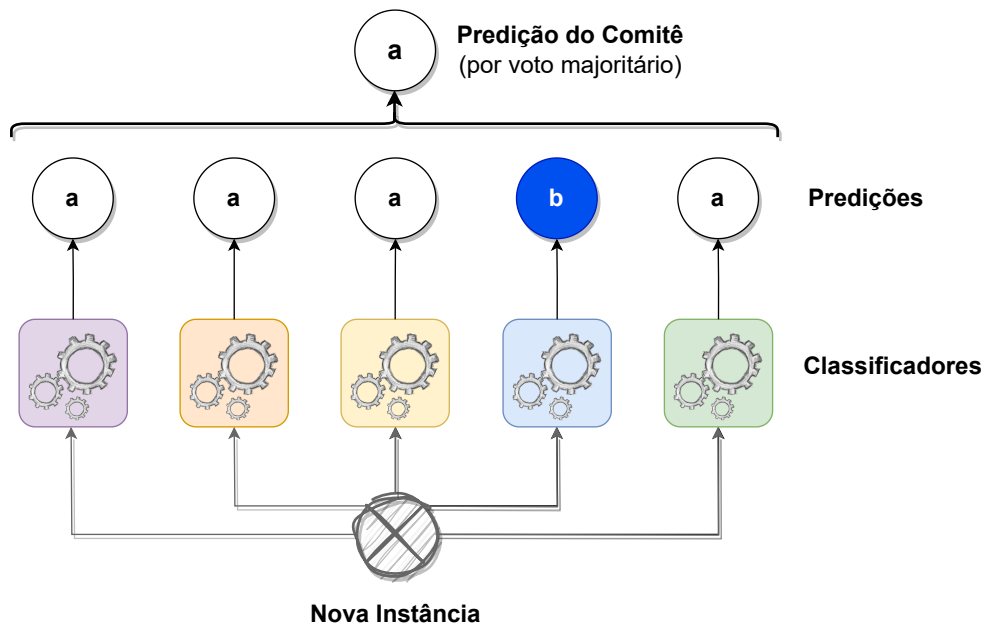
Random Forest (RF) é definida como um conjunto de árvores de decisão, no qual sua aleatoriedade foi explicitamente inserida no processo de construção do modelo de cada árvore de decisão (AGGARWAL *et al.*, 2015; BREIMAN, 2001; CUTLER; CUTLER; STEVENS, 2012). Uma RF pode ser entendida como uma generalização do método básico de ensacamento (*bagging*) com múltiplas árvores de decisão.

As *Extremely Randomized Trees* (ExtraTrees) apresentam similaridade ao que acontece no algoritmo da RF. O ExtraTrees cria árvores de decisão aleatoriamente e utiliza cada árvore para definir um resultado. O seu diferencial é na sua extrema (*extra*) aleatoriedade para criação de suas árvores. O ExtraTrees usa a média para melhorar a precisão da classificação e controlar o ajuste excessivo (*overfitting*) dos dados.

Máquinas de comitê ou aprendizado *ensemble* são motivados por uma combinação de diferentes algoritmos com intuito de realizar previsões e testar instâncias diferentes dadas as características específicas de cada algoritmo de classificação. Desta maneira, um comitê é uma abordagem que visa obter melhor desempenho com a combinação de resultados desses algoritmos (HAN; PEI; TONG, 2022). Para Haykin (2001), o comitê é a combinação de vários algoritmos especialistas, ou seja, o comitê se baseia no conhecimento dos classificadores especialistas para chegar em uma previsão/decisão global que supostamente pode ser superior àquela

alcançada por qualquer um atuando isoladamente. Os diferentes resultados dos classificadores são combinados em um classificador robusto para predições. A Figura 8 ilustra a ideia de um comitê em que utiliza-se o voto majoritário (*majority-vote*) na classificação, também chamado de *hard voting*. Pode-se observar que os diferentes classificadores possuem predições diferentes. Apenas um resultado é **b** os demais possuem a sua saída em **a**, e o voto majoritário escolhe a classe mais “votada” pelos modelos, ou seja, o resultado **a**.

Figura 8 – Funcionamento do Comitê com os classificadores de aprendizado de máquina utilizando voto majoritário.



Fonte – Adaptado de Géron (2022).

Com a premissa de construir um modelo que obtenha melhores resultados, este trabalho de Mestrado visa se beneficiar da combinação das técnicas citadas acima. Com intuito de extrair o melhor de cada uma quando abordamos sobre as tarefas de classificação baseado em modelos de aprendizado profundo e radiômica. Desta maneira, a subseção a seguir apresentará os conceitos de aprendizado profundo.

2.5 Aprendizado Profundo

Deep Learning ou Aprendizado Profundo (AP) é uma abordagem de algoritmos de aprendizado de máquina que utilizam múltiplas camadas computacionais, no qual cada camada aprende uma representação dos dados de entrada (EKMAN, 2021). O AP se desenvolveu a partir do conhecimento do cérebro humano e inclui também as áreas da estatística e da matemática aplicada. Desta maneira, o AP tem-se popularizado pela sua utilidade e geração de resultados com precisão em diversos domínios, embora demande conjuntos de dados maiores, computadores

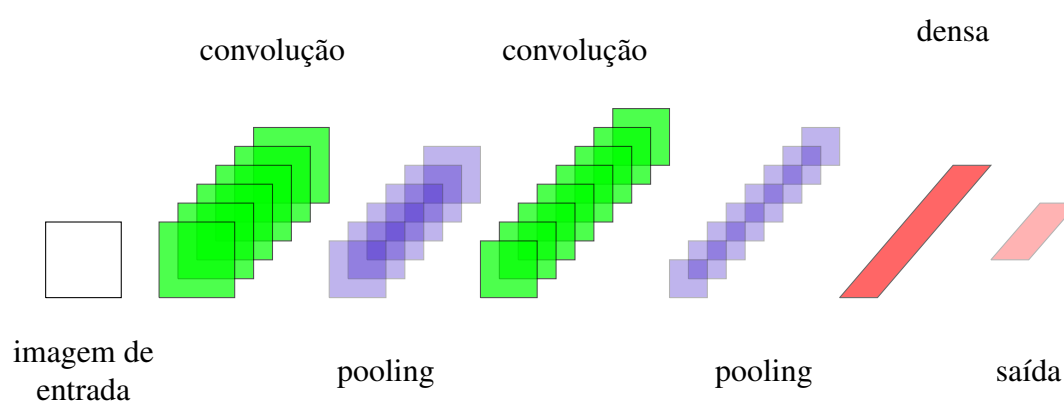
robustos em processamento e técnicas para o treinamento de redes profundas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Em suma, o AP engloba uma nova representação de aprendizagem a partir dos seus dados de entrada, na qual a ênfase da aprendizagem acontece por sucessivas camadas que contribuem para a profundidade e exploração de diferentes modelos (CHOLLET, 2018). Para compreender o processo de aprendizagem desses tipos de redes, a seguir serão apresentados detalhes sobre as Redes Neurais Convolucionais.

Redes Neurais Convolucionais

As Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network* - CNN) foram propostas por LeCun *et al.* (1989). As CNNs são provenientes de um tipo de aprendizagem profunda para processamento de dados possuindo um formato de grelha, como as imagens, inspirada na organização do córtex visual animal que compreende as hierarquias espaciais de características de baixo e alto nível (HUBEL; WIESEL, 1968; FUKUSHIMA, 1980; GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 9 – A arquitetura original de uma rede neural convolucional introduzida por LeCun *et al.* (1989).



Fonte – Adaptado de LeCun *et al.* (1989).

As redes CNNs são compostas por três tipos de camadas: convolução, subamostragem (*pooling*) e camada totalmente conectada (densa), como apresentado na Figura 9. Nas duas primeiras camadas, a de convolução e *pooling*, é realizado o processo de extração de características, enquanto na terceira camada, a densa, mapeia as características extraídas em sua saída, tal como a classificação.

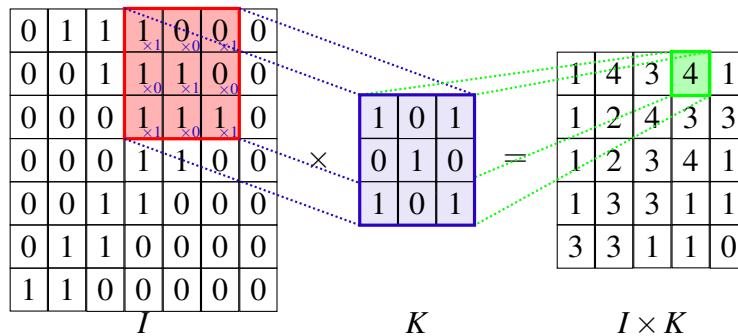
Extração de características por meio de Redes Neurais Convolucionais

Um dos principais aspectos das redes neurais convolucionais é que podem ser utilizadas como extratores de características. Assim, elas podem ser utilizadas para extrair atributos

relevantes de um determinado domínio de imagem. Além disso, as CNNs possuem filtros de diferentes tamanhos que as compõem e acabam generalizando o processo de extração global e local (GOODFELLOW; BENGIO; COURVILLE, 2016).

Para entender melhor o processo de extração de características, é apresentado na Figura 10 o processo da camada de convolução. Este processo consiste na aplicação da operação da camada de entrada e a propagação desse resultado para a próxima camada. Dado um conjunto pequeno de valores, denominado filtro (K), são feitas multiplicações sucessivas desse filtro sobre a entrada (I). Essa técnica é aplicada inúmeras vezes sobre a entrada para adquirir um mapa de características que representa a imagem resultante ($I \times K$) (YAMASHITA *et al.*, 2018). Desta maneira, essa operação pode alterar a dimensionalidade original da imagem e a quantidade de filtros utilizados.

Figura 10 – Funcionamento da camada de convolução.



Fonte – Adaptado de Manaswi, Manaswi e John (2018).

Na operação de subamostragem ou *pooling*, reduz-se a dimensionalidade da camada de entrada com intuito de propagar apenas características mais relevantes. O método comumente utilizado é o *max-pooling*, que consiste em selecionar o maior elemento do conjunto que está sendo observado, assim cria-se uma invariância e pequenas mudanças e distorções locais (GOODFELLOW; BENGIO; COURVILLE, 2016). Uma rede CNN pode ser descrita como um conjunto de camadas convolucionais e de subamostragem (*pooling*) comumente seguidas por diferentes camadas densas, até a camada de saída ou classificação.

As funções de ativação também são aplicadas depois das camadas intermediárias da rede CNN. Essas funções definem a saída com base nos dados de entrada junto ao limiar de ativação. As funções de ativação mais comuns são a Sigmóide (Equação 2.1), Tangente Hiperbólica (Equação 2.2), *Softmax* e a Função de ativação Unidade Linear Retificada - ReLU (Equação 2.4).

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

$$f(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.2)$$

$$f(x)_j = \frac{e^{x_j}}{\sum_{j=1}^K e^{x_j}} \quad \text{para } j = 1, 2, \dots, K \quad (2.3)$$

$$f(x) = \max(0, x) = \begin{cases} 0, & \text{se } x < 0 \\ x, & \text{se } x \geq 0 \end{cases} \quad (2.4)$$

Os classificadores de aprendizado de máquina não desempenham adequadamente como extratores de características, ao apresentarem problemáticas na capacidade de generalização em diferentes contextos de aplicação. Ademais, as CNNs nos permitem utilizar arquiteturas de redes distintas como extratores de características sem a necessidade de várias alterações. Na próxima subseção serão apresentadas algumas arquiteturas de CNN, utilizadas em nossos experimentos.

Redes Neurais Convolucionais muito profundas para Reconhecimento de Imagens

As Redes Neurais Convolucionais muito profundas surgiram em um desafio chamado *ImageNet Large Scale Visual Recognition Challenge*² (ILSVRC) entre os anos de 2010 a 2017, o qual é composto por tarefas que utilizam os conjuntos de dados da ImageNet³. O objetivo do desafio foi promover o desenvolvimento de técnicas para visão computacional de modo a comparar com o estado da arte. As tarefas propostas foram divididas em basicamente duas ao longo dos anos, como a classificação de imagens e a detecção de objetos. Como característica desse desafio, podemos ressaltar o conjunto de dados utilizado incluir aproximadamente um milhão de imagens e mil classes (RUSSAKOVSKY *et al.*, 2015).

Portanto, ficou evidente a importância do desafio da ILSVRC ao decorrer dos anos e refletem até os dias atuais, ao permitirem pesquisadores e entusiastas utilizarem os modelos campeões de arquitetura de CNN desenvolvidas, em diferentes domínios de imagem. Desta maneira, abordaremos algumas arquiteturas muito profundas de CNN a seguir.

Redes VGG

As redes VGG foram propostas por Simonyan e Zisserman (2014) do laboratório *Visual Geometry Group* (VGG) da Universidade de Oxford. As redes VGG partem de uma arquitetura convencional como nos trabalhos de LeCun *et al.* (1989) e Krizhevsky, Sutskever e Hinton (2012). Esta rede surgiu visando estudar o efeito da profundidade que a rede possui em relação à precisão das CNNs convencionais. O seu diferencial é sua profundidade em camadas agrupadas que diferem em tamanhos de 16 e 19 camadas, além das camadas convolucionais empilhadas e o uso de filtros menores e fixos de tamanho 3×3 , como ilustrado na Figura 10.

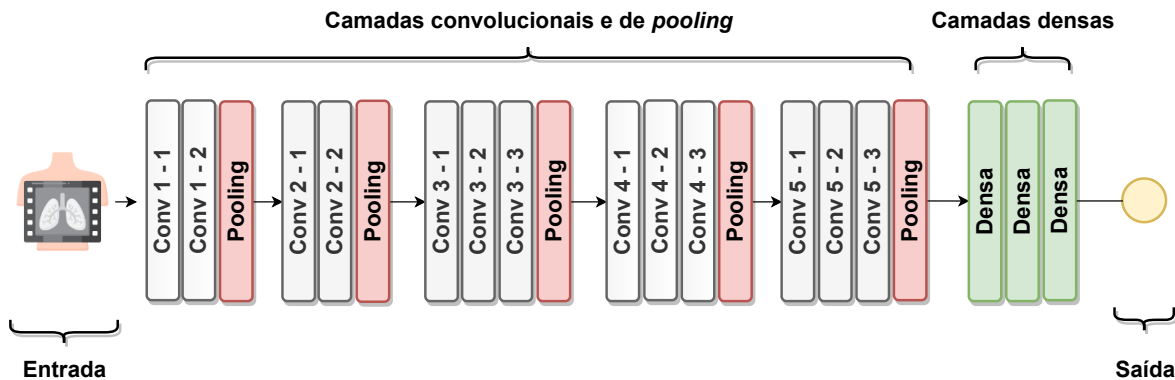
Na Figura 11 é apresentada a arquitetura de uma rede VGG-16. Alguns dos resultados foi o aumento da predição da acurácia com o modelo de profundidade de 16 camadas, que se

² <https://www.image-net.org/challenges/LSVRC/>

³ <https://www.image-net.org/>

estabilizou com a profundidade de 19 camadas, desta maneira surgiram os modelos da VGG-16 e da VGG-19 (EKMAN, 2021). A VGG-16 alcançou uma configuração diferenciada no desafio da ImageNet em 2014 obtendo um erro top-5⁴ com 7,32%, podendo ser comparado com a pioneira AlexNet que alcançou o erro top-5 de 15,3%.

Figura 11 – Arquitetura de uma rede VGG-16.



Fonte – Adaptado de Ekman (2021).

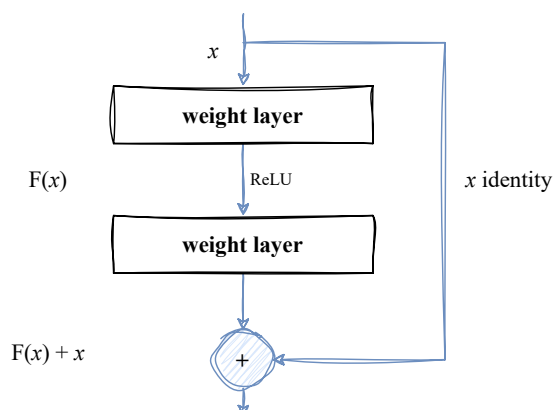
Redes ResNet

Com o intuito de desenvolver redes mais profundas, surgiram as redes ResNet (*Residual Network*) que foram propostas em 2015 por He *et al.* (2016). As ResNets se destacaram por abordar a problemática em que redes muito profundas são difíceis de treinar, devido ao seu aumento de parâmetros e de camadas. Essa problemática ficou conhecida como *vanishing gradient* (dispersão ou apagamento do gradiente), no qual a rede não consegue mais aprender. O diferencial proposto pela ResNet é apresentado na Figura 12, chamado de módulo residual, visando solucionar o problema do *vanishing gradient*. A fim de minimizar esse problema, foi aplicado o *batch normalization* e a utilização da função de ativação ReLu na própria rede. A ResNet utiliza o mecanismo de “pular conexões” o que torna simples identificar a função de aprendizagem (*x identity*).

Portanto, essa abordagem permitiu que a ResNet fosse a rede mais profunda até o momento de sua publicação, com 152 camadas evitando o problema de sobreajuste e *vanishing gradient*. A ResNet foi a rede vencedora do desafio da ImageNet em 2015, obtendo um erro top-5 de 3,57%. Assim, a ResNet foi um marco histórico para os pesquisadores desenvolverem redes mais profundas, como as redes DenseNet que serão descritas a seguir.

⁴ A taxa de erro top-5 é definida como a porcentagem de imagens de teste onde a categoria correta não se encontra entre as cinco categorias que a rede prevê como prováveis.

Figura 12 – Módulo residual.



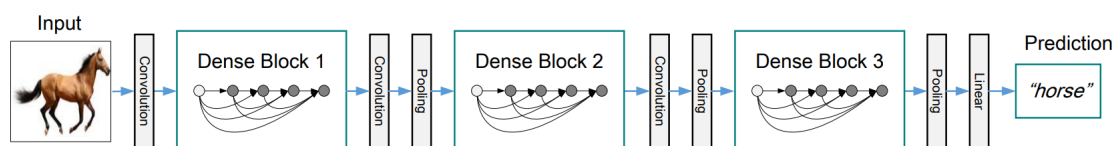
Fonte – Adaptado de [He et al. \(2016\)](#).

Redes DenseNet

As redes ResNet apresentaram uma mudança de paradigma na parametrização da rede utilizando o módulo residual. As redes DenseNet foram propostas em 2016 por [Huang et al. \(2017\)](#), por serem baseadas em parametrização de conectividade simples, no qual cada camada da rede é conectada diretamente a todas as outras camadas, o que é chamado de conexões densas ou conexões de blocos densos.

Na Figura 13 é apresentada a estrutura de uma rede DenseNet. O seu diferencial está relacionado a uma menor parametrização, ao apresentar mais conectividade entre as camadas. As conexões entre as camadas densas se auxiliam na reutilização de características, apresentando um fluxo melhor de informações para determinada tarefa de treinamento. Na DenseNet os filtros são menores e a entrada da rede é preservada, e por fim, as conexões densas evitam ajustes excessivos, o que facilita o treinamento em toda a rede. A DenseNet atingiu uma acurácia semelhante à da ResNet por meio dos desafios da ImageNet, por utilizarem metade dos parâmetros.

Figura 13 – Arquitetura de uma rede DenseNet com três blocos densos. As camadas entre os dois blocos adjacentes são chamadas de camadas de transição e alteram os tamanhos dos mapas de características por meio da convolução e subamostragem.



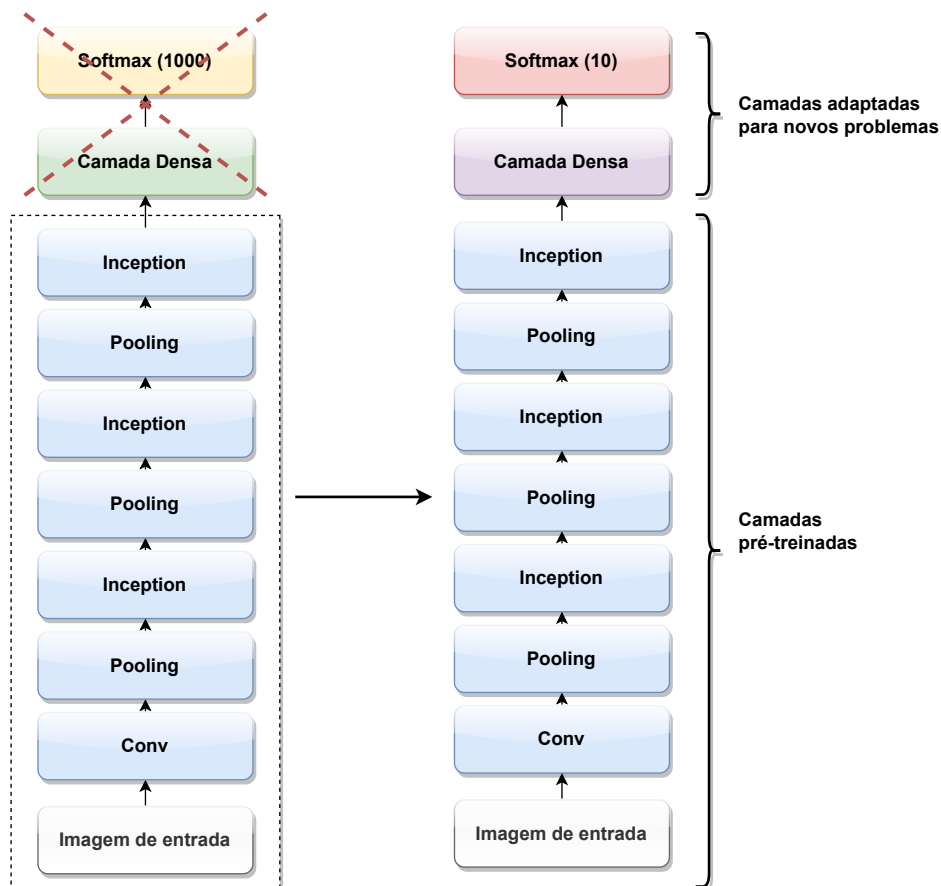
Fonte – [Huang et al. \(2017\)](#).

As redes de aprendizado muito profundo geralmente necessitam de uma técnica denominada transferência de aprendizado, que podem contribuir para a diminuição de recursos computacionais e tempo de processamento necessários para o treinamento dos modelos. Na próxima subseção será discutida essa técnica.

Transferência de Aprendizado

A Transferência de Aprendizagem (do inglês, *Transfer Learning* - TL) é uma técnica comumente utilizada em modelos de aprendizado profundo devido à sua adaptação em diferentes domínios. Dada uma aprendizagem em um determinado cenário (domínio) é explorado para melhorar sua generalização em outro domínio (GOODFELLOW; BENGIO; COURVILLE, 2016). A técnica de transferência de aprendizagem apresenta duas abordagens como ilustrado na Figura 14, em que é utilizado a arquitetura da rede Inception como exemplo. A primeira é começar com um modelo pré-treinado, treinado anteriormente em um conjunto de dados disponibilizado e depois treiná-lo com os seus próprios dados de entrada adicionando alterações nas camadas finais, realizando um processo também conhecido como *fine-tuning* (ajuste-fino). A segunda é a utilização de partes do modelo pré-treinado como um bloco de construção, com intuito de resolver um problema diferente, mas que esteja relacionado (EKMAN, 2021).

Figura 14 – Representação de aprendizagem por transferência baseado na arquitetura da rede Inception. Esquerda: Uma rede de iniciação, modelo base. Direita: Uma nova rede baseada nas camadas pré-treinadas da rede inicial, mas com as últimas camadas modificadas por novas que serão treinadas para um novo problema.



Fonte – Adaptado de Ekman (2021).

Portanto, as redes de aprendizado profundo nos apresentam características e subsídios para explorar os seus potenciais. A radiômica baseada em aprendizado profundo é o ponto de

contribuição principal deste trabalho, no qual seus conceitos serão apresentados a seguir.

2.6 Radiômica baseada em Aprendizado Profundo

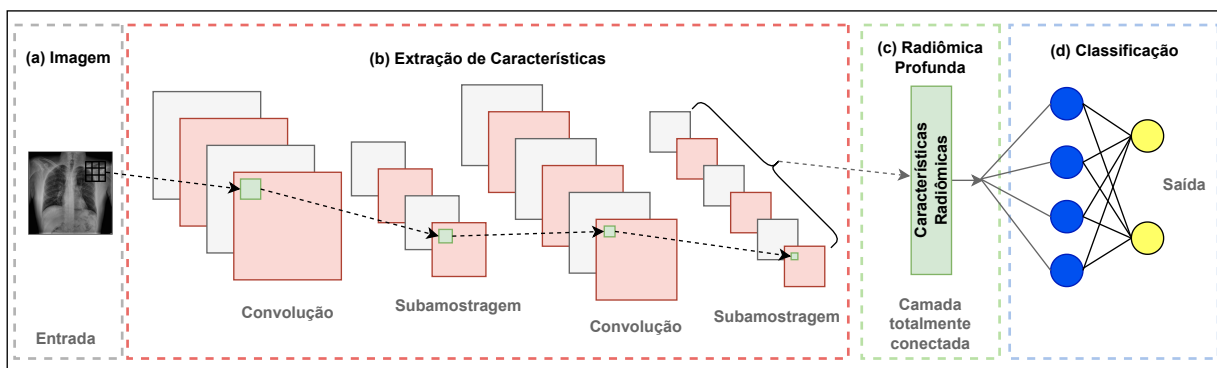
A IA tornou-se essencial para a abordagem radiômica, ao fornecer uma representação de maneira quantitativa por meio de imagens, análise e mapeamento de características. Pode-se apontar áreas como a visão computacional, que se beneficia de técnicas de aprendizado de máquina e aprendizado profundo, que estão particularmente relacionados a estudos/trabalhos de radiômica. A visão computacional torna-se responsável por entender e interpretar os padrões dos exames de imagens e o aprendizado de máquina consiste em aprender os padrões dessas características radiômicas (LAFATA *et al.*, 2022).

Para Lafata *et al.* (2022) as características radiômicas são comumente extraídas por meio de algoritmos de visão computacional (em especial, aprendizado profundo), porém analisadas aplicando os algoritmos de aprendizado de máquina. Proveniente da IA surge a abordagem radiômica baseada em aprendizado profundo. Esta abordagem identifica automaticamente e extrai características de alta dimensionalidade dos dados de entrada, aplicando arquiteturas baseadas em aprendizado profundo. O grande diferencial deste tipo de aplicação é o grande potencial em explorar diferentes níveis de escala e abstração, resultando em um reconhecimento de padrões ou de classificação de determinado domínio de imagem (LOHMANN *et al.*, 2020).

De maneira geral, a radiômica baseada em aprendizado profundo utiliza um fluxograma diferente da abordagem clássica. As arquiteturas de CNNs são utilizadas como extratores de características radiômicas. As CNNs nos permitem explorar e descobrir padrões mais complexos e características mais abstratas do que os algoritmos tradicionais de aprendizado de máquina. Portanto, as definições matemáticas de seleção de características se tornam desnecessárias (LAFATA *et al.*, 2022; VIAL *et al.*, 2018).

A Figura 15 ilustra a utilização da abordagem radiômica para modelos de aprendizado profundo. Iniciando em (a) pela entrada do domínio de imagem, (b) extração de características, (c) radiômica profunda e (d) sua classificação ou saída. Portanto, as arquiteturas de CNNs possuem os mesmos passos para extração dessas características radiômicas profundas. Nota-se que os modelos de aprendizado profundo podem variar conforme a arquitetura empregada, ou seja, os hiperparâmetros, técnicas intrínsecas de cada rede e a sua profundidade. O modelo de aprendizado profundo é utilizado como extrator de características radiômicas profundas, que o processo convolucional e de subamostragem realizam essas operações no conjunto de imagem. Seguido da camada totalmente conectada (camada densa), que são utilizadas para aprender as relações entre as combinações de características radiômicas profundas e extrair a quantidade desejada. Com intuito de apresentar perspectivas de interpretabilidade e explicabilidade por meio da avaliação visual proposta no DEELE-Rad, discutiremos os conceitos na subseção a seguir.

Figura 15 – Abordagem radiômica baseada em modelos de aprendizado profundo. Apresentação de uma breve arquitetura de CNN simplificada na qual é realizada a extração de características radiômicas profundas.



Fonte: Elaborada pelo autor.

2.7 Interpretabilidade e Explicabilidade

A abordagem radiômica baseada em aprendizado profundo nos permite explorar o processo de extração de características de diferentes arquiteturas de CNN. Desta maneira, podemos perceber como funciona o processo de propagação dessas características radiômicas. Assim, a interpretabilidade e explicabilidade por meio de métodos de IA podem nos fornecer subsídios para a tomada de decisão.

Para [Kaur et al. \(2022\)](#) e [Varshney \(2022\)](#) a interpretabilidade e a explicabilidade são maneiras de interação entre a máquina e o ser humano, ambos permitem colaborar na tomada de decisão dos algoritmos de aprendizado de máquina e dos modelos de aprendizado profundo. [Varshney \(2022\)](#) apresenta três categorias de métodos de interpretabilidade e explicabilidade: global ou local; exata ou aproximada e baseado em características ou amostras. A primeira dicotomia se baseia na explicação das decisões tomada em todo o modelo ou para alguma parte do conjunto de dados, dada a imagem de entrada. A segunda dicotomia se baseia na explicação exata do modelo subjacente ou se contém alguma aproximação. A terceira dicotomia é se a linguagem utilizada baseia-se em características ou em parte das amostras dos dados.

Os métodos existentes são os seguintes de acordo com [Kaur et al. \(2022\)](#) e [Velden et al. \(2022\)](#): LIME (*Local Interpretable Model-agnostic Explanations*), LRP (*Layerwise Relevance Propagation*), Automated Concept-based Explanation (ACE), SHAP (*SHapely Additive exPlanations*), e o Grad-CAM (*Gradient-weighted Class Activation Mapping*). Com intuito de explorar a explicação visual neste trabalho optamos por utilizar o Grad-CAM.

A Figura 16 apresenta o funcionamento do Grad-CAM, que visa mapear os *pixels* mais relevantes para a ativação de uma classe por meio de *heatmaps* (mapas de calor). Assim, o Grad-CAM calcula os gradientes mais relevantes para determinada classe a partir da última camada convolucional de uma arquitetura CNN. Desta maneira localizando as partes das imagens

mais relevantes após o treinamento de uma determinada arquitetura CNN. Na Figura 16(a) é apresentada a imagem original, na Figura 16(b) é ilustrada a propagação que visa mapear o cão. O vermelho do mapa de calor indica uma forte região na qual pode ter um cão, em cor amarela indica uma região moderada e a verde a mais fraca. A Figura 16(c) apresenta o mapeamento próximo ao gato.

Figura 16 – Funcionamento do *Gradient-weighted Class Activation Mapping* (Grad-CAM).



Fonte – Chollet (2023).

O Grad-CAM com a abordagem radiômica baseada em aprendizado profundo pode nos fornecer o comportamento da tomada de decisão na extração das principais características radiômicas. Assim, esse comportamento pode ser útil na rotina clínica de um especialista, podendo visualizar as decisões dos modelos de aprendizado profundo. Além disso, essa abordagem final pode ser integrada em um sistema CAD para o especialista.

2.8 Considerações Finais

Neste capítulo foram apresentados os principais conceitos que norteiam esta pesquisa. Foi possível evidenciar que os exames de imagens médicas são necessários para utilização da abordagem radiômica. Assim, os extratores de características podem ser utilizados para extrair características radiômicas mais relevantes de um conjunto de imagens. O uso de CNNs tem se apresentado como alternativa para análise, extração de características e classificação, permitindo uma maior generalização de suas aplicações, superando sucessivos passos do processo tradicional da abordagem radiômica clássica.

De maneira geral, uma rede CNN pode extrair características radiômicas e classificar grandes conjuntos de exames de imagens médicas. Permite que as soluções baseadas em aprendizado profundo possam ser utilizadas em rotina clínica de especialistas ou implementadas em sistemas CAD. Desta maneira, fornecer subsídios para interpretação visual na tomada de decisão do modelo de aprendizado profundo torna-se essencial, exemplificando as principais características para o especialista médico. Na próxima seção serão apresentados os trabalhos relacionados encontrados na literatura que motivaram a criação de uma nova abordagem para radiômica baseada em aprendizado profundo.

TRABALHOS RELACIONADOS

3.1 Considerações Iniciais

Este capítulo tem o propósito de apresentar, investigar e discutir a literatura sobre a temática de radiômica (*radiomics*) com enfoque em modelos de aprendizado profundo para COVID-19. Os trabalhos coletados possuem enfoque em **técnicas clássicas** de radiômica, assim como, em **técnicas de aprendizado profundo** na utilização da abordagem radiômica no contexto da COVID-19. A fase de investigação/levantamento dos trabalhos relacionados apresenta uma das fases metodológica desta pesquisa que visa a busca de lacunas (*gaps*) para o desenvolvimento dela. Tal investigação foi fundamentada em bases de dados e motores de busca conceituados (*Scopus, IEEE Xplore Digital Library, Web of Science, ACM Digital Library e Science Direct*). Vale ressaltar que, para o levantamento bibliográfico, foram tomados como inspiração alguns trabalhos de revisão da literatura como os de [Vial et al. \(2018\)](#), [Koçak et al. \(2019\)](#), [Lohmann et al. \(2020\)](#), [Roberts et al. \(2021\)](#), [Scapicchio et al. \(2021\)](#) e [Zhang et al. \(2022\)](#). Diante da pesquisa realizada nos repositórios de artigos, foram selecionados os principais trabalhos relacionados às técnicas de aprendizado profundo com radiômica. Por fim, na seleção dos trabalhos, os critérios adotados foram artigos de conferências e periódicos, entre os anos de 2020 e o ano de 2023.

Na [Seção 3.2](#) são apresentados os trabalhos com as técnicas de Radiômica Clássica. Seguindo da [Seção 3.3](#) que apresenta as técnicas de Aprendizado Profundo para Radiômica. A [Seção 3.4](#) traz os trabalhos relacionados que utilizam Radiômica baseada em Aprendizado Profundo, com o objetivo de enumerar os trabalhos mais relevantes no que tange à comparação com esta proposta, listando as semelhanças, diferenças e os pontos de contribuição desta pesquisa. Por fim, a [Seção 3.5](#) conclui este Capítulo.

3.2 Técnicas de Radiômica Clássica

O primeiro trabalho que será discutido é o *COVID-19 discrimination framework for X-ray images by considering radiomics, selective information, feature ranking, and a novel hybrid classifier* proposto por [Koyuncu e Barstuğan \(2021\)](#). O autor propõe um *framework* específico para detecção de COVID-19 em imagens de raio-X. O *framework* performa utilizando quatro abordagens, quatro de radiômica (FOS, GLCM, GLRLM e GLSZM), duas de normalização (*min-max* e *z-score*) de características, cinco filtros de seleção (*Bhattacharyya*, *entropy*, *ROC*, *t-test* e *Wilcoxon*) e um classificador otimizado chamado GM-CPSO-NN. Os resultados obtidos foram notáveis, especialmente na métrica de AUC da curva ROC para os conjuntos de dados utilizados (normais e desequilibrados). As características FOS e GLRLM demonstraram alta eficiência, sendo consideradas decisivas para a detecção de COVID-19 em imagens de raio-X.

O trabalho *Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble* de [Chandra et al. \(2021\)](#) é proposto um sistema automático para identificação de COVID-19 (do inglês, *automatic COVID screening - ACoS*). A ideia central do sistema é utilizar um sistema hierárquico para classificação usando um conjunto de algoritmos de aprendizado de máquina tradicionais e descritores de textura radiômica, com intuito de classificar casos de pacientes normais, com pneumonia e infectados com COVID-19. Os resultados com o ACoS foram promissores na classificação de casos normais, pneumonias e COVID-19. Na primeira fase obtiveram-se resultados com uma acurácia de 98,062% e AUC de 0,977. Na segunda fase, com acurácia de 91,329% e AUC de 0,914. Além disso, os autores evidenciaram os seguintes acontecimentos: dada a variabilidade das imagens de raio-x utilizadas no desenvolvimento da pesquisa, utilizou-se a técnica de aumento de dados para gerar transformações nas suficiente para treinar o ACoS. Os melhores descritores de textura radiômica foram os FOS, GLCM e o HOG, ao apresentarem as melhores correlações de atributos em imagens de raio-X associados com os casos de COVID-19.

O trabalho de [Ferreira et al. \(2021\)](#) intitulado de *Novel chest radiographic biomarkers for COVID-19 using radiomic features associated with diagnostics and outcomes* identifica biomarcadores radiômicos em imagens de raio-X para auxiliar na tomada de decisão da COVID-19. Para essa identificação de biomarcadores, os autores utilizaram a rede U-Net para segmentar os pulmões com as regiões de interesse criando uma máscara binária de ambos os pulmões. As características quantitativas baseadas em todos os níveis de cinza foram de primeira ordem, segunda ordem e de alta ordem, onde foram investigados potenciais biomarcadores de COVID-19. Para validar o nível de significância dos biomarcadores foi realizada uma análise estatística, com as seguintes abordagens: *Mann-Whitney*, *Kaplan-Meier* e *Log-rank test*. Os resultados obtidos pelos autores apresentam que foram identificados 57 características em imagens de raio-X que podem melhorar a especificidade do diagnóstico em pacientes com COVID-19.

Na subseção seguinte serão discutidos alguns trabalhos que abordam as técnicas de aprendizado profundo para radiômica aplicados em casos da COVID-19. Isso faz com que

tenhamos uma percepção de como utilizar as técnicas já existentes na incorporação dos modelos de aprendizado profundo.

3.3 Técnicas de Aprendizado Profundo para Radiômica

Yousefi *et al.* (2021) apresentam o *Impartially Validated Multiple Deep-Chain Models to Detect COVID-19 in Chest X-ray Using Latent Space Radiomics* que propõe um sistema automatizado para diagnóstico de COVID-19. O sistema utiliza duas abordagens radiômicas: radiômica tradicional e radiômica profunda (do inglês, *deep radiomic*) por meio dos modelos de aprendizado profundo em imagens de raio-X. O enfoque do trabalho é apresentar a redução de dimensionalidade do vetor característico junto à integração de seleção das características radiômicas. Yousefi *et al.* (2021) aborda dois experimentos para extração de características: (i) a primeira é com a utilização da arquitetura *Convolutional Deep Autoencoder* (ConvAE) para as características de baixa dimensão extraídas das camadas densas, e o (ii) segundo experimento é a extração utilizando a ferramenta py-Radiomics. Os resultados dos modelos combinados com os dois grupos de características radiômicas (radiômica tradicional - radiômica profunda) apresentam resultados promissores para acurácia, utilizando-se multiclasse e validação cruzada em 72,5% (71,2 - 73,4) e 88,8% (88,7 - 89,0) para classificação binária. As limitações do trabalho foram evidenciadas pela falta de informação clínica dos dados utilizados para a predição, assim como, propor uma estratégia mais adequada para a captura de características radiômicas com capacidade discriminativa superior e eficiente para representar os achados dos exames de imagens de raio-X.

O trabalho intitulado *A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images* de Hu *et al.* (2022) tem como proposta um modelo de aprendizado profundo que integra características de radiômica para diagnóstico de COVID-19 em imagens de raio-X. O trabalho tem como diferencial a incorporação do mapa de características de radiômica extraído no processo de convolução de CNNs, e obtém 37 descritores de características radiômicas. De modo a analisar o método, os autores propõem dois experimentos pilotos: (i) o primeiro é apenas utilizando as redes de aprendizado profundo com transferência de aprendizado e (ii) no segundo, é incorporado o mapa de características radiômicas. No segundo experimento os resultados apresentam desempenho superior comparado ao primeiro, com a inclusão da radiômica. Desta maneira, uma das principais contribuições do trabalho de Hu *et al.* (2022) é fornecer pontos de visão na interpretabilidade dos modelos de aprendizado profundo.

O trabalho de Chaddad, Hassan e Desrosiers (2022), o *Deep Radiomic Analysis for Predicting Coronavirus Disease 2019 in Computerized Tomography and X-Ray Images* propõe uma abordagem para codificar as características aprendidas em uma arquitetura de CNN utilizando a técnica de Modelos de Misturas de Gaussianas (do inglês, *Gaussian Mixture Methods*

- GMM), chamado de CNN-GMM. O CNN-GMM utiliza as características extraídas em um algoritmo de aprendizado de máquina, o *random forest*, com intuito de diferenciar casos positivos de COVID-19 de outras categorias de pneumonias. A ideia de utilizar GMM é que este método extrai uma representação compacta das características das camadas convolucionais. Com isso, os autores geram descritores para modelagem da distribuição de características das CNNs em diferentes camadas usando o GMM. Os resultados desta abordagem apresentam uma acurácia de 97% e uma AUC acima de 99%. Os fatores limitantes deste trabalho estão relacionados à falta de informação clínica, isso faz com que não se tenha um diagnóstico mais personalizado de cada paciente.

Em [Zhu et al. \(2021\)](#) é abordado o trabalho *GACDN: generative adversarial feature completion and diagnosis network for COVID-19*, que apresenta um *framework* para multi-visualização incompleta das características artesanais geradas para diagnosticar COVID-19. [Zhu et al. \(2021\)](#) propõem o uso das Redes Adversariais Generativas (do inglês, *Generative Adversarial Network* - GAN) com três redes embutidas, a geradora, discriminadora e a rede consistente com a doença. A rede geradora sintetiza o local específico da característica e a rede discriminadora avalia a qualidade da característica. Ademais, para manter os resultados de diagnóstico de características geradas consistentes com amostras complementares reais, a rede consistente com a doença é utilizada para o diagnóstico final com base nas características radiômicas reais e em características específicas de localização complementadas. Foram realizados dois experimentos, na primeira parte foram comparados os classificadores clássicos (SVM e KNN) com as redes neurais. As redes neurais apresentam os melhores resultados antes e depois com o acréscimo da incorporação de características de radiômica. Na segunda parte, foi realizada a comparação com outros métodos de extração de características (*Low-rank Representation*, *Locality Preserving Projections*, *Latent Low-Rank Representation*, *Stacked Autoencoder* e *Adaptive Feature Selection Deep Forest*), onde o método proposto pelo autor apresenta melhores resultados, dado que as características artesanais não são abandonadas, pois a informação dos dados existentes é explorada e proporcionam mais orientação no processo de diagnóstico. Como limitações, é enfatizado que a proposta atual só resolve casos de COVID-19, para casos positivos ou negativos.

3.4 Sumário dos Trabalhos Seleccionados

Nesta seção serão discutidos os trabalhos relacionados apresentados em relação aos aspectos pertinentes à abordagem proposta nesta dissertação:

- I. Utilização de **Radiômica Profunda** por meio dos modelos de aprendizado profundo;
- II. **Comitê** de classificadores, utilizando as características radiômicas mais relevantes;
- III. Utilização de métodos da **Radiômica** clássica;
- IV. **Avaliação Visual**, apresentando perspectivas de Interpretabilidade e Explicabilidade.

Na Tabela 2 são sumarizados os trabalhos relacionados conforme os aspectos seleccionados, **radiômica profunda**, **comitê** de classificadores, **radiômica** e **avaliação visual**. Além disso, são discutidos os trabalhos de acordo com cada aspecto em comparação com a abordagem de contribuição desta dissertação, intitulada de DEELE-Rad.

Tabela 1 – Visão geral dos trabalhos seleccionados para abordagem de radiômica baseada em aprendizado profundo e clássica em comparação com a nossa abordagem proposta.

Trabalhos	Radiômica Profunda	Comitê	Radiômica	Avaliação Visual
(YOUSEFI <i>et al.</i> , 2021)	✓	×	✓	×
(HU <i>et al.</i> , 2022)	✓	×	✓	×
(CHADDAD; HASSAN; DESROSIERS, 2022)	✓	×	✓	×
(ZHU <i>et al.</i> , 2021)	✓	×	✓	×
(KOYUNCU; BARSTUĞAN, 2021)	×	×	✓	×
(CHANDRA <i>et al.</i> , 2021)	×	✓	✓	×
(FERREIRA <i>et al.</i> , 2021)	×	×	✓	×
DEELE-Rad (COSTA <i>et al.</i> , 2023)	✓	✓	✓	✓

Radiômica Profunda

Características radiômicas profundas ou *deep radiomics*, apresentam a parte essencial da proposta desta dissertação. Usualmente os trabalhos que apresentaram o uso dessa abordagem utilizaram métodos para integração ou extração dessas características. Como método de integração de características podemos evidenciar o trabalho de Hu *et al.* (2022), e o de extração Yousefi *et al.* (2021) e Chaddad, Hassan e Desrosiers (2022).

Em especial, o ponto de contribuição deste trabalho é que são utilizados diferentes tipos de modelos de aprendizado profundo, já consolidados pela literatura em tarefas de classificação de COVID-19, assim como o acréscimo da técnica de transferência de aprendizado e a extração das melhores características radiômicas profundas. Isso nos permite utilizar menos recurso computacional e experimentar diferentes quantidades de características por meio desses modelos.

Comitê

Como evidenciado nos trabalhos, alguns autores após a extração de características radiômicas utilizam algoritmos de aprendizado de máquina para classificação da COVID-19. Apenas na proposta de Chandra *et al.* (2021) podemos observar que foi utilizado um comitê (*ensemble*) de classificadores (com cinco classificadores) utilizando voto majoritário.

Com intuito de validar nossa abordagem, em especial, optou-se por utilizar as características radiômicas profundas extraídas dos modelos de aprendizado profundo em diferentes classificadores. Permitindo criar um comitê de classificadores, tornando-o classificador mais flexível e robusto.

Radiômica

As características radiômicas provenientes da abordagem radiômica clássica tornam-se necessárias para a validação de todos os trabalhos relacionados. Para este trabalho, a abordagem clássica é apoiada utilizada por meio da avaliação visual provenientes dos modelos de aprendizado profundo.

Avaliação Visual

Um ponto diferencial deste trabalho, é a avaliação visual, em especial em relação a como as características radiômicas profundas se propagam nos modelos de aprendizado profundo. Desta maneira, nos permite compreender a tomada de decisão desses modelos. Assim como, fornecer aspectos de interpretabilidade e explicabilidade, podendo ser úteis na rotina clínica de um especialista. Portanto, reduzindo a natureza de “caixa preta” que os modelos de aprendizado profundo possuem.

3.5 Considerações Finais

Neste capítulo foram apresentados os trabalhos relacionados, com intuito de evidenciar os pontos de contribuição desta dissertação. A Tabela 2 apresenta algumas características dos trabalhos de aprendizado profundo com radiômica, além das técnicas clássicas. Como pode-se perceber, há poucos trabalhos que exploram radiômica em modelos de aprendizado profundo. Isso abre espaço para explorar e propor métodos, além de diferentes experimentações que possam contribuir para essa área.

Analisando os trabalhos existentes na literatura para lidar com a problemática de radiômica baseada em aprendizado profundo, destaca-se que é um tema relevante e que ainda há muitos desafios. Um dos desafios da radiômica baseada em aprendizado profundo é como podemos selecionar as melhores características radiômicas e como podemos tornar esses métodos mais interpretáveis e explicáveis para tomadas de decisões para clínicos ou especialistas, de modo a sugerir a criação de CADs baseados em aprendizado profundo com enfoque de fornecer métodos aplicáveis e utilizáveis na área médica.

No Capítulo 4 é apresentada formalmente a nossa abordagem, proposta nesta dissertação para radiômica baseada em aprendizado profundo. O nosso método lida com diferentes modelos de aprendizado profundo, diferentes quantidades de características radiômicas, a construção

de um conjunto de classificadores flexível e robusto, além da análise visual da propagação de características radiômicas em modelos de aprendizado profundo.

Tabela 2 – Sumarização dos dados extraídos dos trabalhos relacionados.

Características investigadas	Tipo de dado usado no modelo	Quantidade de dados	Modelos	Tipo de Validação	Tipo de Avaliação
	Imagens de CXR ou CT (ou ambos)?	Quantidade de imagens utilizadas?	Quais modelos foram utilizados?	K-fold, validação externa, hold out, e entre outros	Performance do modelo, AUC, sensibilidade, especificidade e entre outros
(YOUSEFI <i>et al.</i> , 2021)	CXR	2301 imagens, 417 de COVID-19	U-Net, Deep Convolutional Autoencoder, Random Forest	leave-one-out cross-validation	acurácia, ROC
(HU <i>et al.</i> , 2022)	CXR	812 imagens, 262 de COVID-19	VGG-16, VGG-19, DenseNet-121	hold out (80:20)	acurácia, AUC, sensibilidade, especificidade
(CHADDAD; HASSAN; DESROSIERS, 2022)	ambos	5254 imagens de CT, 2868 de COVID-19; 8084 imagens de CXR, 4042 de COVID-19	ResNet-50, DarkNet, SVM	hold out (80:20), cross-validation	acurácia, AUC, sensibilidade, especificidade
(ZHU <i>et al.</i> , 2021)	CT	2522 imagens de CT, 1945 de COVID-19	Generative Adversarial Network, VB-NET, handcrafted features	hold out (90:10)	acurácia, AUC, sensibilidade, especificidade, recall, precisão F1-score
(KOYUNCU; BARSUĞAN, 2021)	CXR	400 imagens, 80 de COVID-19	Gauss-map-based chaotic particle swarm optimization, handcrafted features	cross-validation	acurácia, AUC, sensibilidade, especificidade, g-mean, precisão, f-measure
(CHANDRA <i>et al.</i> , 2021)	CXR	2346 imagens, 542 de COVID-19	Automatic COVID Screening System (ACoS)	hold out (80:20)	acurácia, AUC, sensibilidade, especificidade, precisão, recall, f1-measure, Matthews correlation coefficient
(FERREIRA <i>et al.</i> , 2021)	CXR	227 imagens, 49 de COVID-19	U-Net	hold out	acurácia, AUC

DEELE-RAD: EXPLORANDO CARACTERÍSTICAS RADIÔMICAS PROFUNDAS

4.1 Considerações Iniciais

Nesta dissertação de Mestrado foi proposta e desenvolvida a abordagem DEELE-Rad, que visa a extração de características radiômicas por meio de exames de imagens médicas de raio-X. A abordagem DEELE-Rad visa colaborar para a descoberta do conhecimento na abordagem radiômica baseada em aprendizado profundo e propõe uma nova abordagem de como utilizar características radiômicas por meio de modelos de aprendizado profundo e com apoio da avaliação visual dessas características.

A motivação deste trabalho de Mestrado veio por sucessivos passos considerando que a abordagem radiômica clássica apresenta e necessita, assim como a falta de interpretabilidade e explicabilidade dessas características radiômicas em modelos de aprendizado profundo e a falta de trabalhos relacionados neste nicho de pesquisa. Desta maneira, foi proposta uma abordagem que utiliza radiômica baseada em aprendizado profundo com intuito inicial de classificar exames de imagens de raio-X de COVID-19.

Este capítulo está organizado da seguinte maneira: a [Seção 4.2](#) apresenta o fluxograma da proposta. Na [Seção 4.3](#) é abordado o *dataset* e a etapa de pré-processamento utilizado. A [Seção 4.4](#) apresenta o processo de extração de características radiômicas profundas para modelos de aprendizado profundo. Na [Seção 4.5](#) é apresentada a criação do comitê de classificadores do DEELE-Rad utilizando as características radiômicas profundas dos diferentes modelos de aprendizado profundo. A [Seção 4.6](#) fornece os meios de avaliação visual, quando são utilizados os modelos de aprendizado profundo no processo de tomada de decisão. Por fim, a [Seção 4.7](#) encerra com as considerações sobre a principal contribuição desta Dissertação de Mestrado.

4.2 A Abordagem DEELE-Rad

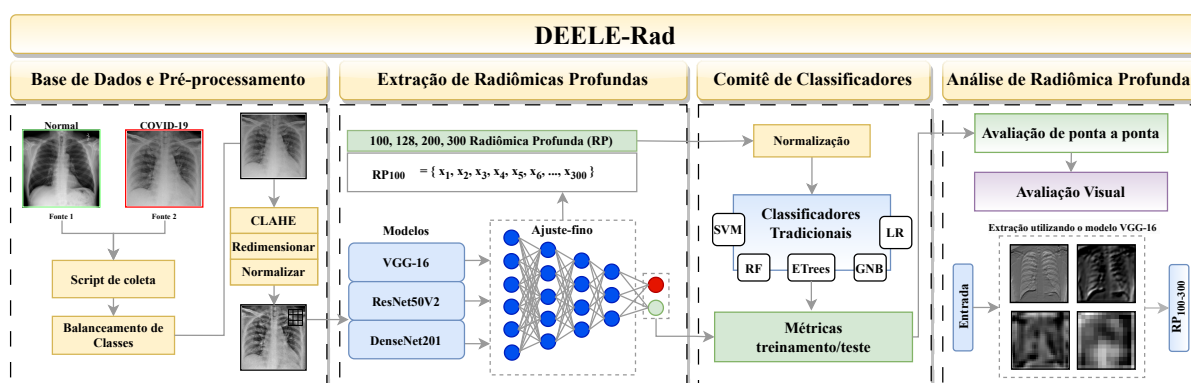
A ideia principal da abordagem DEELE-Rad (*Deep Learning-based Radiomics*) consiste em utilizar as características radiômicas profundas mais relevantes extraídas por meio dos modelos de aprendizado profundo. A abordagem DEELE-Rad utiliza imagens médicas de diferentes fontes. Essas imagens são processadas para melhorar a qualidade da extração de características. Foram utilizados modelos de aprendizado profundo baseados em arquiteturas de CNNs para extrair as características radiômicas profundas (*deep radiomics*) mais relevantes, assim como comparar o desempenho de ponta a ponta para cada modelo. Com intuito de comparar essas características, foi utilizado um comitê de classificadores tradicionais de aprendizado de máquina para validar a abordagem. Foi introduzido um comitê de classificadores de aprendizado de máquina para criar um modelo robusto, melhorando as tarefas de classificação, especialmente para cenários de COVID-19, visando propor resultados para os médicos especialistas. Além disso, disponibiliza uma avaliação visual e de ponta a ponta, com intuito de facilitar a análise da COVID-19 para a tomada de decisão.

A abordagem DEELE-Rad proposta é apresentada na Figura 17 possui quatro etapas principais: base de dados (*dataset*) e pré-processamento, extração de características radiômicas profundas, comitê de classificadores e a análise de radiômica profunda. A principal vantagem da nossa abordagem é explorar diferentes modelos de aprendizado profundo para extração de características radiômicas profundas em diferentes quantidades. Essa abordagem fornece o processo visual da extração de características dos modelos de aprendizado profundo, em especial a VGG-16, além de validar essas características radiômicas profundas em diferentes algoritmos de aprendizado de máquina. Em suma, a abordagem proposta fornece as seguintes contribuições:

- I. Um método para coletar conjuntos de dados de imagens de diferentes fontes, processando e melhorando a qualidade das imagens usando Equalização de Histograma Adaptativo Limitado por Contraste (CLAHE);
- II. A aplicação de uma abordagem que utiliza radiômica baseada em aprendizado profundo para extrair as características mais relevantes de exames de imagens, em especial a COVID-19;
- III. A construção e avaliação de um comitê (modelo de aprendizagem) de classificadores para melhorar o desempenho de tarefas de classificação;
- IV. A avaliação visual do comportamento de características radiômicas profundas, de modo a visualizar o processo de extração em uma arquitetura de aprendizado profundo, em especial a VGG-16.

Os detalhes de cada etapa desta abordagem e a metodologia serão apresentados nas próximas subseções.

Figura 17 – Fluxograma empregado da abordagem DEELE-Rad (*Deep Learning-based Radiomics*). Uma abordagem radiômica baseada em aprendizado profundo para detecção de COVID-19 em imagens de raio-X utilizando comitê de classificadores.



Fonte: Elaborada pelo autor.

4.3 Base de dados e Pré-processamento

Foram coletadas e combinadas imagens de dois repositórios públicos, *covid-chestxray-dataset* (COHEN *et al.*, 2020) e *Chest X-Ray Images (Pneumonia)* (KERMANY *et al.*, 2018). A base de dados resultante inclui 392 imagens de raio-X, com 196 imagens de pacientes infectados por COVID-19, e 196 imagens de pacientes Normais (saudáveis). A Tabela 3 apresenta o número exato de raio-X com projeção pósterio-anterior (PA). Além disso, a base de dados foi dividida aleatoriamente em 80% para o conjunto de treinamento e 20% para o conjunto de teste.

Tabela 3 – Base de dados com a divisão para treino e teste.

Classe	Treino	Teste
COVID-19	157	39
Normal	157	39
Total	314	78

Fonte: Elaborada pelo autor.

As imagens coletadas apresentam diferentes resoluções e contrastes. Desta maneira, foi realizado um pré-processamento da base de dados a fim de obter uma melhoria na qualidade dos exames. Assim, a técnica CLAHE foi utilizada para melhorar o contraste da imagem de entrada, operando sobre o histograma local, especialmente em imagens médicas (PISANO *et al.*, 1998). O CLAHE aplicado a base de dados é composto pelas seguintes configurações: o limite de recorte¹ de tamanho 2, e o tamanho da grelha (*grid*)² em 8×8 . Vale ressaltar que a configuração

¹ Representa o intervalo (*bin*) de corte no histograma, ou seja, esses *pixels* serão cortados e distribuídos uniformemente antes da equalização de histograma.

² A imagem é dividida em pequenos blocos de tamanho 8×8 . Cada um desses blocos é equalizado por

original do OpenCV³ é padronizada com o limite de corte de tamanho 40 e o tamanho da grelha em 8x8. O CLAHE foi aplicado conforme o fluxo abaixo:

- As imagens foram convertidas do modelo RGB para o modelo LAB⁴;
- Em seguida, foi aplicado o CLAHE para o canal de cor L;
- O canal de L foi fundido com os canais A e B, para obter uma imagem LAB melhorada;
- Por fim, a imagem LAB foi convertida para uma imagem RGB melhorada.

A Figura 18 apresenta um caso de COVID-19 e Normal (saúdavel) com o CLAHE aplicado. Pode-se notar que as imagens com o CLAHE apresentam melhorias significativas em seu contraste em relação à imagem normal, ou seja, há o realce de características específicas. Portanto, essa melhoria pode auxiliar os especialistas a encontrar determinados achados, sejam em imagens de raio-X, TC e MRI. Entretanto, vale ressaltar que os métodos de melhoramento de contraste não foram desenvolvidos para complementar ou alterar a informação estrutural de determinada imagem, sendo que o contraste tem o intuito apenas de realçar características específicas (PISANO *et al.*, 1998; PIZER *et al.*, 1987). Por fim, todas as imagens da base de dados foram redimensionadas para 224×224 pixels com três canais de cores, e normalizadas em um intervalo de [0, 1] para entrada nas arquiteturas de CNNs.

4.4 Extração de Radiômica Profunda

No processo de extração de características radiômicas profundas foram utilizadas três redes de CNNs pré-treinadas pela ImageNet⁵, sendo elas as redes da VGG-16, ResNet50V2 e a DenseNet201. A Tabela 4 apresenta a quantidade de parâmetros e a profundidade de cada rede. Esses modelos de aprendizado profundo foram escolhidos por possuírem desempenho melhor para as tarefas de diagnóstico de COVID-19 utilizando imagens de raio-X, de acordo com Roberts *et al.* (2021). Foi aplicada a técnica de ajuste-fino (*fine-tuning*) utilizando a abordagem de transferência de aprendizado, e o *global average pooling* para minimizar o sobreajuste (*overfitting*) dos modelos.

A Tabela 5 apresenta a estrutura empregada para cada arquitetura CNN, de cima para baixo. Nota-se que foi utilizada a mesma estrutura de ajuste-fino para todos os modelos de aprendizado profundo. Ressalta-se que a estrutura de cada modelo de aprendizado profundo

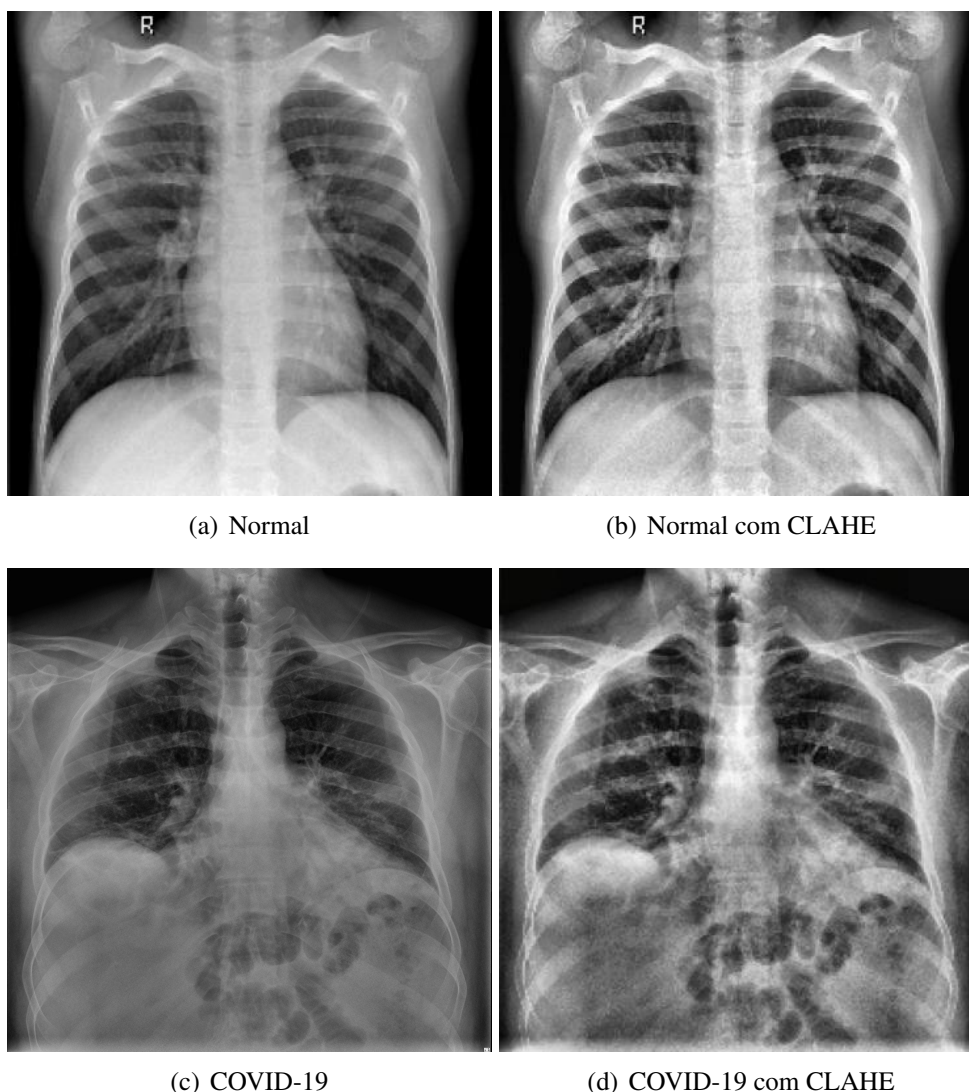
histograma.

³ https://docs.opencv.org/4.x/d5/daf/tutorial_py_histogram_equalization.html

⁴ O espaço ou modelo de cor LAB consiste em três canais de cores, expressando a cor de um *pixel* como três (L, A, B), no qual o canal L representa a luminosidade e os canais A e B representam os componentes de cor verde-vermelho e azul-amarelo, respectivamente.

⁵ <https://keras.io/api/applications/>

Figura 18 – Imagens de COVID-19 e Normal (saudável) com a aplicação do CLAHE.



Fonte: Elaborada pelo autor.

utilizado foi fundamentado e detalhado na [Seção 2.5](#). Foi proposta uma sequência de 4 (quatro) camadas densas com 2048, 1024, 512, com a última camada variando entre 128, 100, 200 e 300 neurônios, utilizando a função de ativação ReLU com o acréscimo da camada de *dropout*. Além disso, foi considerado o otimizador *Adam*⁶ para atualizar os parâmetros de cada modelo. Foi utilizado o *batch* de tamanho 32, taxa de aprendizagem (*learning rate*) em 0,0001, e a quantidade de 100 épocas. O *dropout* utilizado foi de 20%, assim como a parada precoce (do inglês, *early stopping*)⁷ foram utilizados para evitar sobreajuste da nossa abordagem. Os parâmetros foram escolhidos empiricamente por meio de experimentações, contando também com apoio dos trabalhos da literatura.

Foram extraídas 100, 128, 200 e 300 características radiômicas profundas de cada

⁶ <https://keras.io/api/optimizers/adam/>

⁷ https://keras.io/api/callbacks/early_stopping/

Tabela 4 – Modelos de aprendizado profundo com pesos pré-treinados pela ImageNet.

Modelo	Parâmetros (M)	Profundidade
VGG-16	138,4 M	16
ResNet50V2	25,6 M	103
DenseNet201	20,2 M	402

M: Milhões

Fonte – Adaptado de [Chollet et al. \(2015\)](#).

Tabela 5 – Tipos de camadas utilizados em nosso ajuste-fino.

Camada (tipo)
VGG-16, ResNet50V2, DenseNet201 (Modelos)
global_average_pooling2d
dense (Densa)
dropout (Dropout)
dense_1 (Densa)
dropout_1 (Dropout)
dense_2 (Densa)
dropout_2 (Dropout)
deep_radiomic (Densa)
dropout_3 (Dropout)

Fonte: Elaborada pelo autor.

imagem antes da camada de classificação (e.g, *softmax*). Desta maneira, a saída é estruturada como um vetor característico de radiômicas profundas com 100, 128, 200 e 300 neurônios da camada densa. A extração de radiômica profunda em diferentes quantidades nos garante as melhores características, devido à aplicação das técnicas citadas anteriormente. A parada precoce (*early stopping*) neste caso é o principal responsável, dado que o treinamento do modelo é parado se não houver melhoramento da métrica escolhida. Na abordagem proposta, optou-se pela escolha da métrica da Acurácia.

4.5 Comitê de Classificadores

Com intuito de validar a abordagem, foi proposto um comitê de classificadores para classificação de exames de raio-X de COVID-19 e Normal (saudável). Foram combinados 5 (cinco) classificadores da seguinte maneira: *Support Vector Machine* (SVM), *Gaussian Naive Bayes* (GNB), *Logistic Regression* (LR), *Random Forest* (RF) e *Extremely Randomized Trees*

(ExtraTrees). Os classificadores de aprendizado de máquina permitem a criação de um modelo flexível e robusto. Assim, o comitê apresenta uma mistura de algoritmos de aprendizagem para melhorar a qualidade, classificação e regressão de problemas. Entretanto, no comitê de classificadores um sistema de votação majoritário foi responsável pela predição final. Os parâmetros e hiperparâmetros utilizados foram os padrões disponíveis na biblioteca do *scikit-learn*⁸.

4.6 Análise de Radiômica Profunda

Na análise de radiômica profunda buscou-se apresentar as percepções que a abordagem de aprendizado profundo baseada em radiômica pode fornecer. A abordagem DEELE-Rad fornece o comportamento de tomada de decisão na extração de características radiômicas profundas a partir dos modelos de aprendizado profundo. Esse comportamento pode ser útil na rotina clínica dos especialistas. O especialista pode visualizar as decisões do modelo e apoiar o diagnóstico usando um método de aprendizado profundo baseado em radiômica. De modo a realizar uma análise, visou-se utilizar um método explicável já consolidado pela literatura, o Grad-CAM, como apresentando na [Seção 2.6](#). Desta maneira, possibilita-se explorar o aprendizado dos modelos em diferentes partes do processo de extração de características radiômicas, em especial o processo convolucional.

4.7 Considerações Finais

Este capítulo apresentou a principal contribuição desta dissertação, o método DEELE-Rad (*Deep Learning-based Radiomics*) para extração de características radiômicas profundas em modelos de aprendizado profundo. O DEELE-Rad apresenta uma nova aplicabilidade de extração de características radiômicas, assim como, a avaliação visual dessas características em modelos de aprendizado profundo, proporcionando perspectivas de interpretabilidade e explicabilidade. De modo a avaliar essas características radiômicas profundas por diferentes modelos de aprendizado profundo, foi proposto um comitê de classificadores (*ensemble*). Além disso, ao considerar modelos relacionados, podemos avaliar nosso classificador de modelos em cenários de classificação e regressão, particularmente a COVID-19. No [Capítulo 5](#) é apresentado os resultados e discussões dos experimentos referentes a abordagem proposta, além da avaliação e análises a partir da aplicabilidade do DEELE-Rad.

⁸ <https://scikit-learn.org/stable/>

RESULTADOS E DISCUSSÕES

5.1 Considerações Iniciais

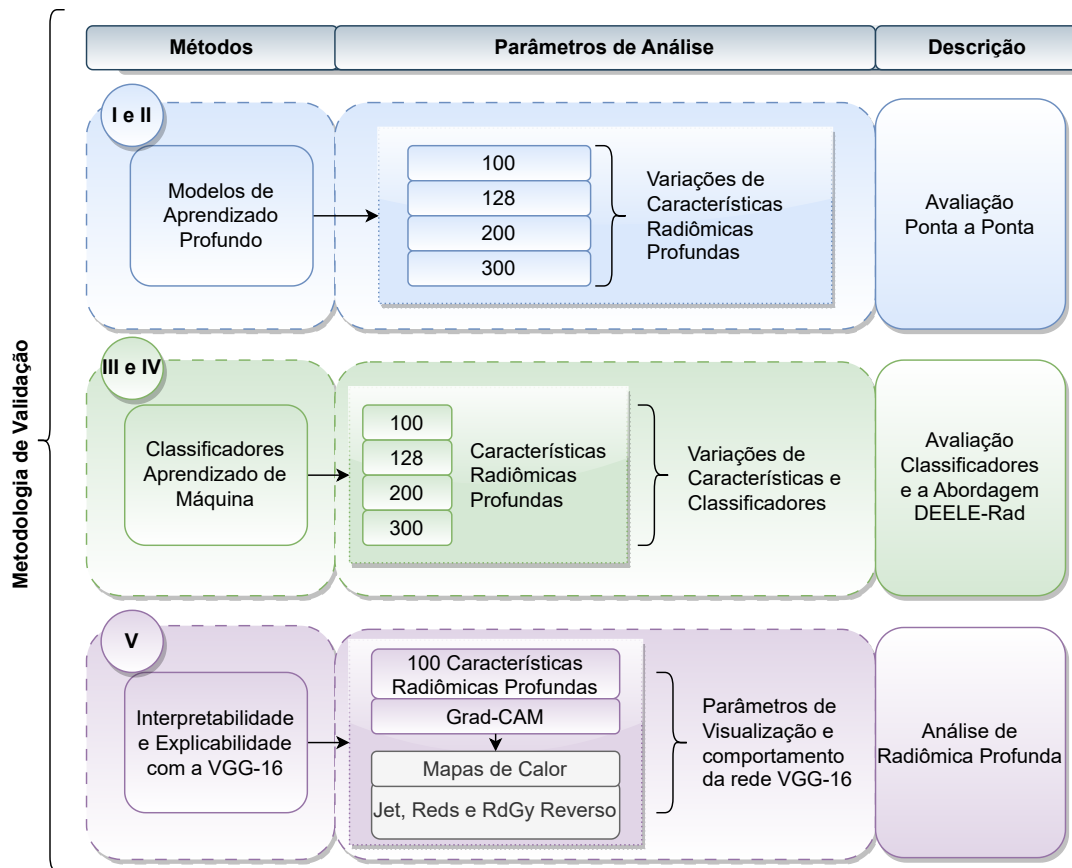
Este capítulo tem por objetivo apresentar a metodologia de validação e o processo de avaliação experimental da abordagem DEELE-Rad. São apresentados os experimentos a partir da execução do DEELE-Rad sobre a base de dados coletada. Foram utilizados três modelos de aprendizado profundo, VGG-16, ResNet50V2 e DenseNet201, assim como o comitê de classificadores do DEELE-Rad, compostos por cinco algoritmos de aprendizado de máquina: SVM, GNB, LR, RF, e a ExtraTrees. Além disso, foi proposta uma avaliação visual do DEELE-Rad com intuito de apresentar o processo de extração de características da VGG-16, e as possíveis perspectivas de interpretabilidade e explicabilidade. Ressalta-se que a abordagem do DEELE-Rad possui enfoque na classificação binária, entre casos da COVID-19 e Normal (saudável) por meio de exames de imagens de raio-X.

Este capítulo está organizado nas seguintes seções: a [Seção 5.2](#) detalha a metodologia de validação utilizada nos experimentos. A avaliação experimental do DEELE-Rad é composta por quatro partes principais. Na [Seção 5.3](#) são apresentados os resultados do desempenho dos modelos de aprendizado profundo, assim como o processo de extração de características radiômicas profundas. A [Seção 5.4](#) apresenta o desempenho dos algoritmos de aprendizado de máquina junto ao comitê de classificadores do DEELE-Rad. Na [Seção 5.5](#) é apresentada a avaliação visual visando perspectivas de interpretabilidade e explicabilidade. Por fim, na [Seção 5.6](#) são feitas discussões referentes à abordagem DEELE-Rad.

5.2 Metodologia de Validação

A Figura 19 apresenta o fluxograma da metodologia de validação para o DEELE-Rad. O desempenho da abordagem DEELE-Rad foi avaliada a partir da realização das seguintes etapas:

Figura 19 – Fluxograma da metodologia de validação do DEELE–Rad.



Fonte: Elaborada pelo autor.

- I. Comparar e avaliar o desempenho dos modelos de aprendizado profundo de ponta a ponta, utilizando a base de dados coletada;
- II. Extrair diferentes quantidades de características radiômicas profundas dos modelos de aprendizado profundo;
- III. Aplicar algoritmos de aprendizado de máquina utilizando as características radiômicas profundas retiradas dos modelos de aprendizado profundo;
- IV. Propor um comitê de classificadores por meio do voto majoritário, utilizando as características radiômicas profundas;
- V. Apresentar perspectivas de interpretabilidade por meio da avaliação visual com a abordagem DEELE–Rad.

Com intuito de validar a abordagem DEELE–Rad para análises preditivas, a matriz de confusão pode ser utilizada para apresentar a frequência de elementos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. A Figura 20 ilustra uma matriz de confusão. Abaixo segue uma breve descrição de cada elemento:

Figura 20 – Exemplo de uma matriz de confusão aplicada à nossa metodologia de validação.

		Classe Predita	
		COVID-19 (+)	Normal (-)
Classe Verdadeira	COVID-19 (+)	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Normal (-)	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte – Adaptado de [Faceli et al. \(2021\)](#).

- Verdadeiro Positivo (VP): ocorre quando o exemplo da classe de COVID-19 foi classificado corretamente.
- Verdadeiro Negativo (VN): ocorre quando o exemplo da classe Normal foi classificado corretamente.
- Falso Positivo (FP): ocorre quando a classe verdadeira do exemplo é Normal, mas o exemplo foi classificado incorretamente como pertencente à classe de COVID-19.
- Falso Negativo (FN): ocorre quando o exemplo pertence originalmente à classe de COVID-19 e foi incorretamente predito como a classe Normal.

A partir da matriz de confusão podemos derivar outras métricas de desempenho. Desta maneira, é possível obter métricas de desempenho como a Acurácia ([Equação 5.1](#)), que calcula a soma dos valores da diagonal principal dividindo os elementos da matriz. A Precisão ([Equação 5.2](#)) representa a proporção de exemplos de COVID-19 classificados corretamente entre todos aqueles preditos como COVID-19. A Revocação ([Equação 5.3](#)) corresponde à taxa de acerto na classe de COVID-19. Também é chamada de taxa de verdadeiros positivos (TVP) ([FACELI et al., 2021](#)). A precisão pode ser vista como uma medida de acerto/exatidão do modelo, e a revocação como uma medida de sua completude. A F1-Score ([Equação 5.4](#)) ou medida F, pode ser combinada com a precisão e a revocação, sendo a média harmônica ponderada da precisão e da revocação. Já a medida AUC (do inglês, *Area Under the Curve* - AUC) produz valores entre 0 e 1. Os valores mais próximos de 1 (ou 100%) são considerados melhores.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (5.1)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (5.2)$$

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (5.3)$$

$$\text{F1-Score} = \frac{2 \times \text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (5.4)$$

Todos os experimentos do DEELE-Rad foram realizados em um computador com a seguinte especificação: processador (CPU) RYZEN(R) 9, 32GB RAM, GPU NVIDIA RTX 3090 24GB VRAM. Foi utilizado o Keras¹ com *backend* em Tensorflow² para os modelos de aprendizado profundo. Para os algoritmos de aprendizado de máquina foi utilizado o *scikit-learn*³ e o Miniconda com CUDA toolkit⁴ V11.2.67 em um sistema operacional Linux Pop!OS 5.11⁵. A linguagem de programação utilizada foi o *Python*⁶.

5.3 Avaliação dos Modelos de Aprendizado Profundo

Para a avaliação de desempenho dos modelos de aprendizado profundo de ponta a ponta, foram utilizados quatro cenários de experimentação. Foram utilizados os modelos de aprendizado profundo como extratores de características radiômicas profundas, que extraem 100, 128, 200 e 300 características, respectivamente. Para melhor entendimento dos resultados, optou-se por expressar o melhor, pior e empate das métricas de desempenho, por meio da cor verde, vermelho e amarelo, respectivamente. A cor de empate foi utilizada para os cenários em que foi identificado o empate das melhores métricas.

Avaliação com 100 características

A Tabela 6 apresenta o desempenho dos modelos de aprendizado profundo com 100 características radiômicas profundas. A VGG-16 apresentou a melhor Acurácia, AUC, F1-Score, e Revocação. A DenseNet201 apresentou a Acurácia igual à da VGG-16, assim como, obteve uma melhor Precisão em relação à VGG-16 e à ResNet50V2. Já a ResNet50V2 obteve o pior desempenho quando comparada aos outros dois modelos.

Avaliação com 128 características

A Tabela 7 apresenta o desempenho dos modelos de aprendizado profundo quando utilizamos 128 características radiômicas profundas. A VGG-16 ganhou em todas as métricas de desempenho, incluindo a Acurácia, AUC, F1-Score, Revocação e Precisão. Neste cenário de 128

¹ <https://keras.io/>

² <https://www.tensorflow.org/>

³ <https://scikit-learn.org/stable/>

⁴ <https://anaconda.org/anaconda/cudatoolkit>

⁵ <https://pop.system76.com/>

⁶ <https://www.python.org/>

Tabela 6 – Métricas de desempenho utilizando 100 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.

Modelo	Acurácia	AUC	F1-Score	Revocação	Precisão
VGG-16	87,3%	87,4%	87,3%	87,4%	87,4%
ResNet50V2	84,8%	84,7%	84,6%	84,7%	86,4%
DenseNet201	87,3%	87,2%	87,2%	87,2%	88,3%
Legenda:		Melhor	Pior	Empate	

Fonte: Elaborada pelo autor.

características radiômicas profundas a DenseNet201 obteve o pior desempenho comparada aos outros dois modelos. Já a ResNet50V2 obteve resultados mais próximos da VGG-16.

Tabela 7 – Métricas de desempenho utilizando 128 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.

Modelo	Acurácia	AUC	F1-Score	Revocação	Precisão
VGG-16	83,5%	83,4%	83,3%	83,4%	85,4%
ResNet50V2	81,0%	80,8%	80,4%	80,8%	84,9%
DenseNet201	77,2%	77,1%	77,1%	77,1%	77,5%
Legenda:		Melhor	Pior	Empate	

Fonte: Elaborada pelo autor.

Avaliação com 200 características

A Tabela 8 apresenta o desempenho dos modelos de aprendizado profundo com 200 características radiômicas profundas. Neste cenário, a DenseNet201 apresenta o melhor desempenho para todas as métricas de Acurácia, AUC, F1-Score, Revocação e Precisão. Nota-se que, neste cenário de 200 características radiômicas profundas, a DenseNet201 foi a primeira a ultrapassar a faixa de mais de 90% de Acurácia assim como nas demais métricas. A VGG-16 obteve o pior desempenho em relação os outros modelos. Já a ResNet50V2 apresentou resultados satisfatórios quando comparado às radiômicas profundas anteriores.

Avaliação com 300 características

De modo a avaliar a abordagem como mais características radiômicas, optou-se por testá-la em um cenário com 300 características radiômicas profundas. A Tabela 9 apresenta o desempenho dos modelos de aprendizado profundo deste cenário. DenseNet201 obteve os melhores desempenho na Acurácia, AUC, F1-Score, Revocação. Já VGG-16 obteve a melhor Precisão. A ResNet50V2 obteve o pior desempenho em relação aos outros dois modelos, assim como nos cenários de avaliação anteriores.

Tabela 8 – Métricas de desempenho utilizando 200 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.

Modelo	Acurácia	AUC	F1-Score	Revocação	Precisão
VGG-16	82,3%	82,2%	82,1%	82,20%	83,10%
ResNet50V2	86,1%	86,0%	86,0%	86,0%	86,3%
DenseNet201	92,4%	92,3%	92,3%	92,3%	93,5%
Legenda:		Melhor	Pior	Empate	

Fonte: Elaborada pelo autor.

Tabela 9 – Métricas de desempenho utilizando 300 características radiômicas profundas em modelos de aprendizado profundo de ponta a ponta.

Modelo	Acurácia	AUC	F1-Score	Revocação	Precisão
VGG-16	83,5%	83,3%	83,0%	83,3%	87,7%
ResNet50V2	79,7%	79,6%	79,6%	79,6%	80,5%
DenseNet201	86,1%	86,0%	85,9%	86,0%	87,3%
Legenda:		Melhor	Pior	Empate	

Fonte: Elaborada pelo autor.

Discussão dos Resultados dos Modelos de Aprendizado Profundo

As métricas de desempenho utilizadas para mensurar a efetividade dos modelos foram a Acurácia, Precisão, e a Revocação. Foram escolhidos inicialmente os melhores modelos pelo desempenho da Acurácia, seguido da Precisão e a Revocação. A Figura 21 apresenta as matrizes de confusão para os melhores modelos.

Os experimentos com os modelos de aprendizado profundo revelaram que a VGG-16 obteve um desempenho melhor nos cenários com 100 e 128 características radiômicas profundas. A DenseNet201 obteve as melhores métricas de desempenho nos cenários com 200 e 300 características radiômicas profundas. Desta maneira, optou-se por discutir os resultados da VGG-16 e da DenseNet201 no cenário com 100 e 200 características radiômicas profundas, respectivamente. Os valores das métricas de desempenho nestes cenários de experimentação foram os melhores.

A VGG-16 obteve uma Acurácia de 87,3% no cenário com 100 características radiômicas profundas. A Precisão do modelo da VGG-16 foi de 87,4% para este cenário, refletindo a exatidão do modelo em classificar corretamente 35 dos 41 exemplos de COVID-19 (+ classificado como -) do conjunto de testes conforme a Figura 21(a). Nota-se que no cenário com 100 características radiômicas profundas a Acurácia da VGG-16 e da DenseNet201 foram iguais, e a Precisão da DenseNet201 foi de 88,3% superando os 87,4% da VGG-16. Já a Revocação foi de 87,4%, correspondendo a taxa de acertos na classe de COVID-19 (+ classificado como -).

Para a DenseNet201 no cenário com 200 características radiômicas profundas obteve uma Acurácia de 92,4%. A DenseNet201 foi responsável por ultrapassar a faixa de 90% das métricas de desempenho entre todos os cenários de experimentação. A Precisão do modelo da DenseNet201 foi de 93,5% neste cenário, em que apresenta a exatidão do modelo em classificar 33 instâncias de COVID-19 (+ classificado como -) conforme a Figura 21(b). A Revocação do modelo foi de 92,3%.

Figura 21 – Matrizes de confusão para os melhores modelos no cenário com 100 e 200 características radiômicas profundas, respectivamente.

	(+)	(-)
(+)	35	4
(-)	6	34

(a) VGG-16

	(+)	(-)
(+)	33	6
(-)	0	40

(b) DenseNet201

A ResNet50V2 obteve as piores métricas de desempenho no cenário com 300 características radiômicas profundas. A Acurácia obtida neste cenário foi de 79,7%. Já a Precisão do modelo foi de 80,5%, apresentando uma taxa de exatidão em classificar apenas 28 de 33 exemplos de COVID-19 (+ classificado como -) conforme a Figura 22. A Revocação ainda neste cenário foi de 79,6%.

Figura 22 – Matriz de confusão para o pior modelo no cenário com 300 características radiômicas profundas.

	(+)	(-)
(+)	28	11
(-)	5	35

(a) ResNet50V2

Os experimentos com os modelos de aprendizado profundo de ponta a ponta em diferentes cenários forneceu versatilidade para a validação da abordagem DEELE-Rad. Desta maneira, foi possível observar como esses modelos de aprendizado profundo se comportam no processo de extração de características radiômicas profundas, em diferentes cenários. As métricas de desempenho foram responsáveis por nortear o processo de validação.

Foi possível observar que diferentes modelos de aprendizado profundo apresentam características intrínsecas em sua arquitetura. Observa-se que a VGG-16 apresentou uma média de Acurácia de 84,15% entre os quatro cenários. A VGG-16 é constituída por poucos parâmetros treináveis quando utilizada a técnica de ajuste fino, e a sua construção em blocos é mais simples além de utilizar filtros de tamanho fixo de 3×3 , como fundamentado na [Seção 2.5](#). Já a DenseNet201 apresentou uma média de Acurácia de 85,75% entre os quatro cenários, equiparada a melhor entre os três modelos de aprendizado profundo. A DenseNet201 utiliza a ideia de convoluções separáveis em profundidade, além de conexões curtas entre as camadas de toda a rede, fortalecendo a propagação de características radiômicas profundas. Por fim, a ResNet50V2 apresentou uma média de Acurácia de 82,9% entre os quatro cenários, obtendo o pior desempenho entre os modelos de aprendizado profundo.

Portanto, a utilização dos modelos de aprendizado profundo como extratores de características fornece a versatilidade de testar e parametrizar as características radiômicas profundas com a abordagem DEELE-Rad. A próxima subseção apresenta a validação dessas características radiômicas profundas extraídas em diferentes classificadores de aprendizado de máquina.

5.4 Avaliação dos Classificadores de Aprendizado de Máquina

Para a avaliação de desempenho dos classificadores de aprendizado de máquina, foram utilizados quatro cenários de experimentação. As características radiômicas profundas extraídas no processo de experimentação dos modelos de aprendizado profundo, foram utilizados neste processo de avaliação, de modo a medir o desempenho da abordagem DEELE-Rad. Foram utilizados os mesmos cenários de experimentação para as características radiômicas profundas já extraídas, com 100, 128, 200 e 300 características, respectivamente. Para melhor entendimento dos resultados, optou-se por expressar o melhor, pior e empate das métricas de desempenho da abordagem DEELE-Rad, por meio da cor verde, vermelho e amarelo, respectivamente. A cor de empate foi utilizada para os cenários em que foi identificado o empate das melhores métricas. Considerando todos os classificadores avaliados, foi denotado para as melhores métricas de desempenho em **negrito e sublinhado**.

Avaliação com 100 características

A Tabela 10 apresenta o desempenho de cada classificador de aprendizado de máquina utilizando o cenário com 100 características radiômicas profundas. O classificador *Logistic Regression* (LR) obteve as melhores métricas de desempenho da Acurácia, AUC, F1-Score, Revocação e Precisão quando foram utilizadas as características radiômicas profundas provenientes do modelo da DenseNet201. O comitê do DEELE-Rad obteve as melhores métricas de desempenho da Acurácia, AUC, F1-Score, Revocação e Precisão quando foram utilizadas as características

radiômicas profundas também provenientes do modelo da DenseNet201. Observa-se que a abordagem DEELE-Rad utilizando as características provenientes da DenseNet201 foi superado pelo classificador LR. Isto pode ter ocorrido devido à quantidade menor de características, ou seja, a utilização de 100 características radiômicas profundas. Ao final da seção será apresentada a discussão para este cenário.

Tabela 10 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 100 características radiômicas profundas mais relevantes.

Classificador	Modelo AP	Acurácia (%)	AUC (%)	F1-Score (%)	Revocação (%)	Precisão (%)
SVM	VGG-16	81,0	81,0	81,0	81,0	81,0
	ResNet50V2	78,5	78,7	77,8	78,7	83,2
	DenseNet201	79,7	80,0	79,0	80,0	85,5
GNB	VGG-16	83,5	83,7	83,4	83,7	84,6
	ResNet50V2	78,5	78,7	77,8	78,7	83,2
	DenseNet201	82,3	82,5	81,8	82,5	86,8
LR	VGG-16	81,0	81,1	81,0	81,1	81,5
	ResNet50V2	78,5	78,5	77,8	78,7	83,2
	DenseNet201	87,3	87,5	87,2	87,5	88,9
RF	VGG-16	82,3	82,3	82,3	82,3	82,3
	ResNet50V2	79,7	80,0	79,2	80,0	83,9
	DenseNet201	83,5	83,7	83,3	83,7	86,3
ExtraTrees	VGG-16	83,5	83,6	83,5	83,6	83,7
	ResNet50V2	79,7	80,0	79,2	80,0	83,9
	DenseNet201	82,3	82,4	82,0	82,4	84,4
DEELE-Rad	VGG-16	83,5	83,6	83,5	83,6	84,1
	ResNet50V2	79,7	80,0	79,2	80,0	83,9
	DenseNet201	84,8	85,0	84,5	85,0	88,2
Legenda DEELE-Rad:		Melhor	Pior	Empate		

Fonte: Elaborada pelo autor.

Avaliação com 128 características

A Tabela 11 apresenta o desempenho para cada classificador de aprendizado de máquina utilizando o cenário com 128 características radiômicas profundas, assim como o DEELE-Rad. O classificador *Extremely Randomized Trees* (ExtraTrees) obteve as melhores métricas de desempenho da Acurácia, AUC, F1-Score, Revocação e Precisão quando foram utilizadas as características radiômicas profundas provenientes do modelo da ResNet50V2. O comitê do

DEELE-Rad obteve as melhores métricas de desempenho da Acurácia, F1-Score, Revocação, e Precisão quando foram utilizadas as características radiômicas profundas também provenientes do modelo da ResNet50V2. Observa-se que há um empate entre as métricas da ExtraTrees e o DEELE-Rad. Por ser baseado em um comitê de classificadores, o DEELE-Rad fornece uma maior confiabilidade, devido o seu processo de aprendizagem ser baseado em todos os classificadores.

Tabela 11 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 128 características radiômicas profundas mais relevantes.

Classificador	Modelo AP	Acurácia (%)	AUC (%)	F1-Score (%)	Revocação (%)	Precisão (%)
SVM	VGG-16	86,1	86,1	86,1	86,1	86,1
	ResNet50V2	89,9	89,9	89,9	89,9	90,2
	DenseNet201	87,3	87,5	87,2	87,5	89,8
GNB	VGG-16	87,3	87,5	87,2	87,5	89,8
	ResNet50V2	91,1	91,2	91,1	91,2	92,4
	DenseNet201	86,1	86,3	85,8	86,3	89,0
LR	VGG-16	86,1	86,2	85,9	86,2	88,0
	ResNet50V2	89,9	89,9	89,9	89,9	90,2
	DenseNet201	84,8	85,0	84,6	85,0	87,1
RF	VGG-16	81,0	80,9	80,8	80,9	82,0
	ResNet50V2	88,6	88,7	88,6	88,7	89,2
	DenseNet201	87,3	87,5	87,2	87,5	89,8
ExtraTrees	VGG-16	81,0	80,9	80,8	80,9	82,0
	ResNet50V2	92,4	92,5	92,4	92,5	93,3
	DenseNet201	87,3	87,5	87,2	87,5	89,8
DEELE-Rad	VGG-16	86,1	86,1	86,1	86,1	86,3
	ResNet50V2	92,4	92,5	92,4	92,5	93,3
	DenseNet201	87,3	87,5	87,2	87,5	89,8
Legenda DEELE-Rad:		Melhor	Pior	Empate		

Fonte: Elaborada pelo autor.

Avaliação com 200 características

A Tabela 12 apresenta o desempenho para cada classificador de aprendizado de máquina utilizando o cenário com 200 características radiômicas profundas. O classificador *Gaussian Naive Bayes* (GNB) obteve as melhores métricas de desempenho da Acurácia, AUC, F1-Score, Revocação e Precisão quando foram utilizadas as características radiômicas profundas provenientes do modelo da DenseNet201. O comitê da DEELE-Rad obteve as melhores métricas de

desempenho da Acurácia, AUC, F1-Score, Revocação e Precisão quando foram utilizadas as características radiômicas profundas também provenientes do modelo da DenseNet201. De maneira similar ao cenário com 100 características radiômicas profundas, o DEELE-Rad foi superado mais uma vez por um classificador baseado em probabilidades, superando-o em todas as métricas de desempenho. Em contraste, o DEELE-Rad combina diferentes classificadores especialistas, fornecendo uma tomada de decisão mais robusta para predições.

Tabela 12 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 200 características radiômicas profundas mais relevantes.

<i>Classificador</i>	<i>Modelo AP</i>	<i>Acurácia (%)</i>	<i>AUC (%)</i>	<i>F1-Score (%)</i>	<i>Revocação (%)</i>	<i>Precisão (%)</i>
SVM	VGG-16	87,3	87,4	87,3	87,4	88,2
	ResNet50V2	82,3	82,5	81,9	82,5	85,5
	DenseNet201	93,7	93,7	93,7	93,7	93,7
GNB	VGG-16	87,3	87,5	87,2	87,5	88,9
	ResNet50V2	83,5	83,7	83,4	83,7	85,3
	DenseNet201	94,9	95,0	94,9	95,0	95,0
LR	VGG-16	89,9	90,0	89,8	90,0	90,8
	ResNet50V2	83,5	83,6	83,5	83,6	84,1
	DenseNet201	94,9	94,9	94,9	94,9	94,9
RF	VGG-16	84,8	84,9	84,8	84,9	85,1
	ResNet50V2	87,3	87,5	87,2	87,5	88,9
	DenseNet201	93,7	93,7	93,7	93,7	93,7
ExtraTrees	VGG-16	86,1	86,2	86,0	86,2	86,6
	ResNet50V2	88,6	88,7	88,5	88,7	89,8
	DenseNet201	94,9	94,9	94,9	94,9	94,9
DEELE-Rad	VGG-16	87,3	87,4	87,3	87,4	88,2
	ResNet50V2	87,3	87,5	87,2	87,5	88,9
	DenseNet201	93,7	93,7	93,7	93,7	93,7
Legenda DEELE-Rad:		Melhor	Pior	Empate		

Fonte: Elaborada pelo autor.

Avaliação com 300 características

A Tabela 13 apresenta o desempenho para cada classificador de aprendizado de máquina utilizando o cenário com 300 características radiômicas profundas. Neste cenário, o comportamento dos classificadores de aprendizado de máquina apresentaram dois melhores resultados. O

Support Vector Machine (SVM) e o *Extremely Randomized Trees* (ExtraTrees) obtiveram um empate em todas as métricas de desempenho da Acurácia, AUC, F1-Score e Precisão quando foram utilizadas as características radiômicas profundas do modelo da DenseNet201. O comitê do DEELE-Rad obteve as melhores métricas de desempenho da Acurácia, AUC, F1-Score, Revocação e Precisão quando foram utilizadas as características radiômicas profundas também provenientes do modelo da DenseNet201. Abaixo será apresentada a discussão para este cenário de avaliação.

Tabela 13 – Métricas de desempenho dos classificadores de aprendizado de máquina utilizando 300 características radiômicas profundas mais relevantes.

Classificador	Modelo AP	Acurácia (%)	AUC (%)	F1-Score (%)	Revocação (%)	Precisão (%)
SVM	VGG-16	86,1	86,2	85,9	86,2	88,0
	ResNet50V2	84,8	85,0	84,5	85,0	88,2
	DenseNet201	94,9	94,9	94,9	94,9	94,9
GNB	VGG-16	87,3	87,5	87,2	87,5	89,8
	ResNet50V2	84,8	85,0	84,5	85,0	88,2
	DenseNet201	92,4	92,5	92,4	92,5	93,3
LR	VGG-16	86,1	86,2	86,0	86,2	87,2
	ResNet50V2	78,5	78,5	78,5	78,5	78,5
	DenseNet201	93,7	93,6	93,7	93,6	94,0
RF	VGG-16	84,8	84,9	84,7	84,9	85,6
	ResNet50V2	84,8	85,0	84,6	85,0	87,1
	DenseNet201	93,7	93,7	93,7	93,7	93,7
ExtraTrees	VGG-16	86,1	86,2	86,0	86,2	87,6
	ResNet50V2	88,6	88,7	88,5	88,7	89,8
	DenseNet201	94,9	94,9	94,9	94,9	94,9
DEELE-Rad	VGG-16	86,1	86,2	85,9	86,2	88,0
	ResNet50V2	86,1	86,3	85,8	86,3	89,0
	DenseNet201	94,9	94,9	94,9	94,9	94,9
Legenda:			Melhor	Pior	Empate	

Discussão dos Classificadores de Aprendizado de Máquina e o DEELE-Rad

Os resultados obtidos na Seção 5.3 parametrizam e guiam a comparação com os classificadores de aprendizado de máquina e a abordagem DEELE-Rad. O modelo de aprendizado profundo da DenseNet201 apresentou as melhores métricas de desempenho, seguindo da ResNet50V2 e a VGG-16. A Tabela 14 apresenta a média de cada métrica de desempenho do DEELE-Rad em todos os cenários de experimentação com as características radiômicas profundas

extraídas dos modelos de aprendizado profundo. Observa-se que para a abordagem DEELE-Rad baseada em comitê, o modelo de aprendizado profundo da DenseNet201 também obteve o melhor desempenho. A VGG-16 obteve a pior média das métricas de desempenho nos cenários de experimentação.

Tabela 14 – Média das métricas de desempenho do DEELE-Rad.

DEELE-Rad					
Modelo	Acurácia	AUC	F1-Score	Revocação	Precisão
VGG-16	86,10%	86,15%	86,00%	86,15%	87,15%
ResNet50V2	86,70%	86,90%	86,50%	86,90%	88,95%
DenseNet201	90,50%	90,60%	90,45%	90,60%	91,75%
Legenda:		Melhor	Pior	Empate	

Fonte: Elaborada pelo autor.

Como já fundamentado na [Seção 2.4](#), o DEELE-Rad foi baseado em um comitê de classificadores de aprendizado de máquina. Buscou-se combinar diferentes classificadores especialistas em árvores de decisão e em probabilidades estatísticas. Desta maneira, optou-se por discutir os resultados dos classificadores de aprendizado de máquina e da abordagem DEELE-Rad nos cenários com 100 e 300 características radiômicas profundas. As métricas de desempenho utilizadas para mensurar a efetividade dos classificadores e do DEELE-Rad foram a Acurácia e a Precisão.

Para o cenário com 100 características radiômicas profundas, o classificador LR obteve uma Acurácia de 87,3% e a Precisão de 88,9%, com as características provenientes do modelo de aprendizado profundo da DenseNet201. O LR superou as métricas de desempenho da abordagem DEELE-Rad neste cenário utilizando também as características provenientes da DenseNet201. Observa-se que a abordagem DEELE-Rad utilizando as características provenientes da DenseNet201 obteve a Acurácia e a Precisão de 84,8% e 88,2%, respectivamente. Embora o DEELE-Rad seja baseado em um comitê de classificadores, neste cenário o voto majoritário não desempenhou adequadamente. Isto pode ter ocorrido devido à quantidade de 100 características radiômicas profundas. Ressalta-se que a ResNet50V2 no DEELE-Rad obteve as piores métricas de desempenho.

Já para o cenário com 300 características radiômicas profundas, os classificadores SVM e o ExtraTrees obtiveram as maiores e as melhores métricas de desempenho, gerando um empate. Ambos os classificadores obtiveram uma Acurácia e Precisão de 94,9% e 94,9%, respectivamente. Neste cenário o DEELE-Rad obteve as mesmas métricas de desempenho dos classificadores, ou seja, o voto majoritário para este cenário desempenhou adequadamente em sua escolha. Já a VGG-16 e a ResNet50V2 obtiveram um empate na métrica da Acurácia em 86,1%. As outras métricas de desempenho da AUC, F1-Score e Revocação desses modelos desempenharam próximas, apresentando uma diferença suave de 0,1% e uma diferença de 1% para a métrica de

Precisão. Observa-se também que para este cenário que contém mais características radiômicas profundas, as métricas de desempenho foram melhores. Apresentando um aumento acima de 10% nos valores da métrica de desempenho da Acurácia no cenário com 100 para o com 300 características radiômicas profundas, da abordagem DEELE-Rad.

Com intuito de comparar o desempenho dos modelos de aprendizado profundo de ponta a ponta e o DEELE-Rad, a Tabela 15 apresenta a comparação da métrica de desempenho da Acurácia e a Precisão. Para um melhor entendimento, as medidas de desempenho dos modelos de aprendizado profundo (AP) da Seção 5.3 foram utilizados, assim como, os cenários de experimentações de 100, 128, 200 e 300 características radiômicas profundas. Foi possível observar que a Acurácia no cenário com 100 características os modelos de aprendizado profundo desempenharam melhor (↑) do que a abordagem DEELE-Rad. Para os demais cenários com 128, 200 e 300 características a abordagem DEELE-Rad obteve as melhores (↑) métricas de desempenho. Em contraste, os modelos de aprendizado profundo foram piores (↓). Nota-se que as características provenientes do modelo da DenseNet201 foram melhores nos cenários com 100, 200 e 300 características, na abordagem DEELE-Rad. Já a ResNet50V2 obteve um desempenho melhor no cenário com 128 características, na abordagem DEELE-Rad. A VGG-16 não superou nenhum modelo na abordagem DEELE-Rad.

Tabela 15 – Comparação da métrica de desempenho da acurácia e a precisão com os modelos de aprendizado profundo em relação ao DEELE-Rad, em cada respectivo cenário de experimentação.

Acurácia (%)								
	100		128		200		300	
Modelo	AP	DEELE-Rad	AP	DEELE-Rad	AP	DEELE-Rad	AP	DEELE-Rad
VGG-16	87,3 ↑	83,5 ↓	85,4 ↓	86,3 ↑	83,1 ↓	88,2 ↑	87,7 ↓	88,0 ↑
ResNet50V2	84,8 ↑	79,7 ↓	84,9 ↓	93,3 ↑	86,3 ↓	88,9 ↑	80,5 ↓	89,0 ↑
DenseNet201	87,3 ↑	84,8 ↓	77,5 ↓	89,8 ↑	93,5 ↓	93,7 ↑	87,3 ↓	94,9 ↑
Precisão (%)								
VGG-16	87,4 ↑	84,1 ↓	83,5 ↓	86,1 ↑	82,3 ↓	87,3 ↑	83,5 ↓	86,1 ↑
ResNet50V2	86,4 ↑	83,9 ↓	81,0 ↓	92,4 ↑	86,1 ↓	87,3 ↑	79,7 ↓	86,1 ↑
DenseNet201	88,3 ↑	88,2 ↓	77,2 ↓	87,3 ↑	92,4 ↓	93,7 ↑	86,1 ↓	94,9 ↑

↑: melhor ↓: pior

Fonte: Elaborada pelo autor.

Ainda na Tabela 15 é apresentada a comparação da métrica de desempenho da Precisão. Pode-se observar que os resultados são semelhantes aos já discutidos referentes à Acurácia, acima. Resumidamente, os modelos de aprendizado profundo só desempenharam melhor (↑) que a abordagem DEELE-Rad no cenário com 100 características. Nos demais cenários os modelos de aprendizado profundo obtiveram desempenho pior (↓). Em contraste, a abordagem DEELE-Rad

desempenhou melhor (↑) nos cenários com 128, 200 e 300 características. Nota-se também que as características provenientes do modelo da DenseNet201 foram melhores nos cenários com 100, 200 e 300 características, na abordagem DEELE-Rad. Já a ResNet50V2 obteve um desempenho melhor no cenário com 128 características, na abordagem DEELE-Rad. A VGG-16 não superou nenhum modelo na abordagem DEELE-Rad.

De maneira geral, os modelos de aprendizado profundo ponta a ponta foram essenciais para o processo de validação da abordagem DEELE-Rad, utilizando-se esses modelos como extratores de características radiômicas profundas e testando essas características em diferentes classificadores de aprendizado de máquina. Foi possível observar que a melhor configuração nos cenários de experimentações para o DEELE-Rad foi: com a utilização de 300 características radiômicas profundas, junto aos classificadores da SVM e a ExtraTrees por meio da rede DenseNet201. Portanto, tomando como parâmetro a métrica de desempenho da Precisão para abordagem DEELE-Rad, obteve-se uma melhora em até 8,8% (↑) comparada aos modelos de aprendizado profundo de ponta a ponta, para a melhor configuração. A abordagem DEELE-Rad obteve as melhores métricas por se basear em um sistema de voto majoritário por meio de um comitê de classificadores no qual os melhores resultados foram provenientes de classificadores baseados em árvores e probabilidades.

A próxima subseção a seguir apresenta o processo de avaliação visual dessas características radiômicas profundas.

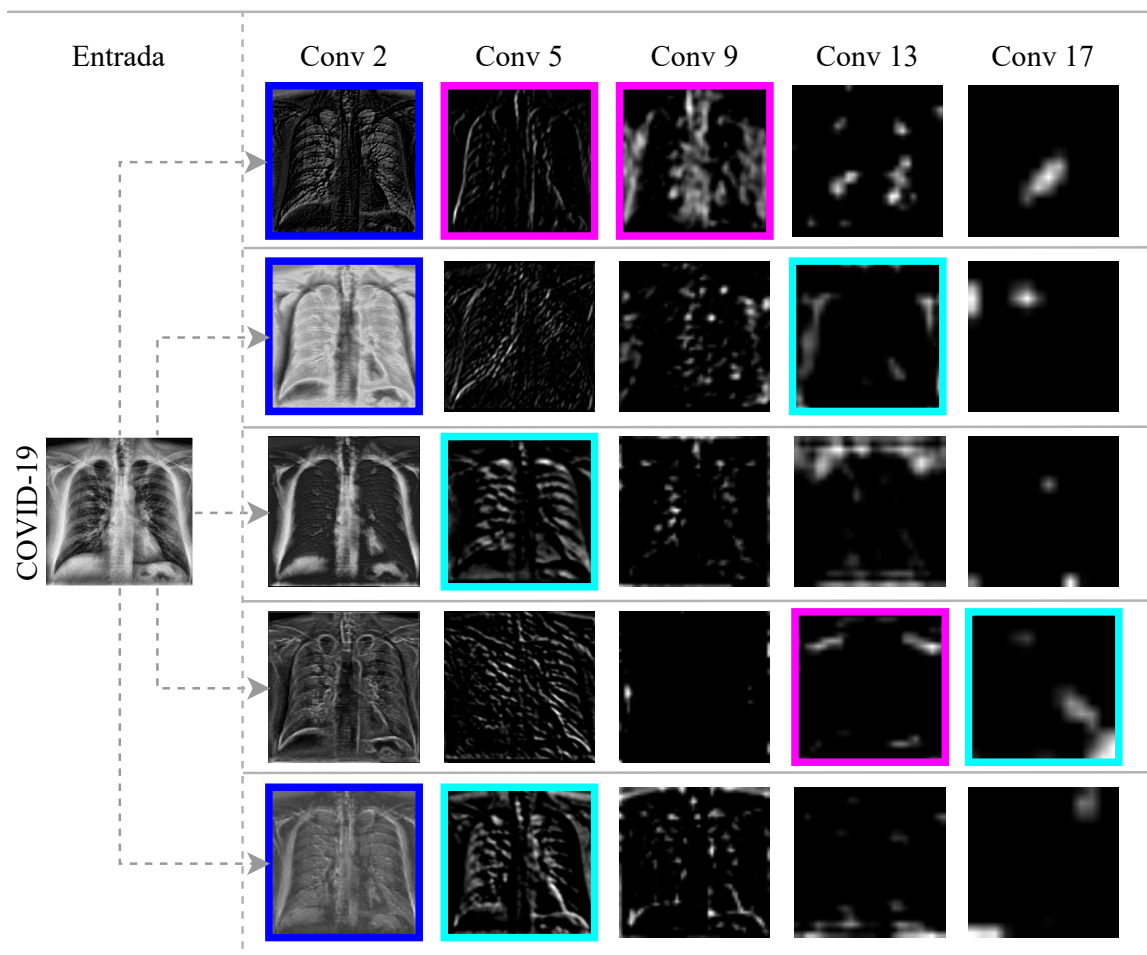
5.5 Avaliação Visual para Análise de Radiômica Profunda

Os modelos de aprendizado profundo apresentam a problemática de serem rotulados como caixa-preta (*black-box*), devido à falta de interpretabilidade na tomada de decisão desses modelos, principalmente devido à complexidade de técnicas intrínsecas que diferentes arquiteturas de aprendizado profundo apresentam. Desta maneira, a abordagem DEELE-Rad com a avaliação visual visa apresentar o funcionamento do processo de extração de características desses modelos, particularmente a VGG-16 com 100 características radiômicas profundas. A VGG-16 foi escolhida devido à simplicidade em sua arquitetura para o processo de avaliação visual.

A Figura 23 apresenta cinco amostras do processo de extração de características radiômicas profundas de casos de COVID-19 por meio da VGG-16. A Figura 24 apresenta o mesmo processo para um caso Normal. O processo convolucional (Conv) dos modelos de aprendizado profundo tentam compreender os padrões básicos por meio dos dados de entrada, ou seja, por meio dos exames de imagens de raio-X. Pode-se observar que a propagação de características ocorre ao longo de todo o processo convolucional, uma vez que os modelos de aprendizado profundo aprendem diretamente os padrões a partir das imagens de raio-X e mantêm as informações relativas às texturas, cores e formas das imagens. Note que foram escolhidas as cinco amostras

das mesmas posições no processo convolucional. As imagens com as bordas coloridas nas cores azul, magenta e ciano representam o nível de cinza (cor), forma e textura, respectivamente. Nestas amostras é possível perceber como acontece o aprendizado dos padrões pela abordagem DEELE-Rad, ressaltando os níveis de cinza, forma e textura. A fim de compreender como funciona cada camada convolucional, pode-se observar a Conv 2, e destacar que ainda existe a forma e diferentes tons de cinza no raio-X em ambas as amostras. Nota-se que as características radiômicas profundas são propagadas pela rede durante as Conv 5, 9 e 13, resultando na Conv 17 apenas mapas/regiões de ativação. Desta maneira, a rede propaga as radiômicas profundas mais relevantes com base no ajuste fino empregado na VGG-16.

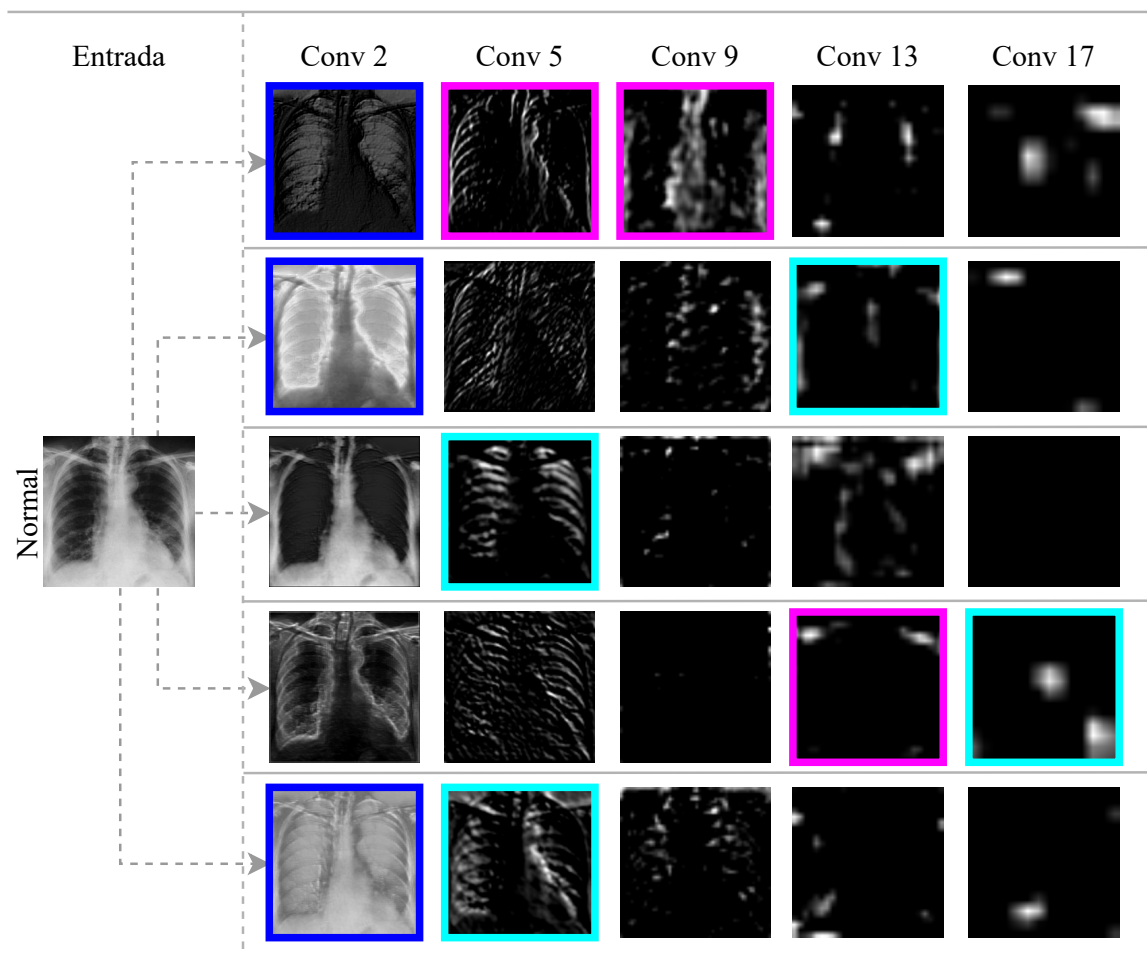
Figura 23 – Visualização do processo de extração de características radiômicas profundas no modelo da VGG-16 com 100 características radiômicas profundas. Cada bloco de convolução da rede apresenta diferentes níveis de cinza, forma e textura para um caso de COVID-19. As amostras com a borda na cor azul, magenta e ciano representam o nível de cinza, forma e a textura, respectivamente.



Fonte: Elaborada pelo autor.

De modo a investigar a abordagem DEELE-Rad, utilizou-se o Grad-CAM para realizar o seguinte experimento: foi escolhida uma amostra de cada camada convolucional para um caso de COVID-19 e analisada a explicação visual baseada no método Grad-CAM. Vale ressaltar

Figura 24 – Visualização do processo de extração de características radiômicas profundas no modelo da VGG-16 com 100 características radiômicas profundas. Cada bloco de convolução da rede apresenta diferentes níveis de cinza, forma e textura para um caso Normal. As amostras com a borda na cor azul, magenta e ciano representam o nível de cinza, forma e a textura, respectivamente.

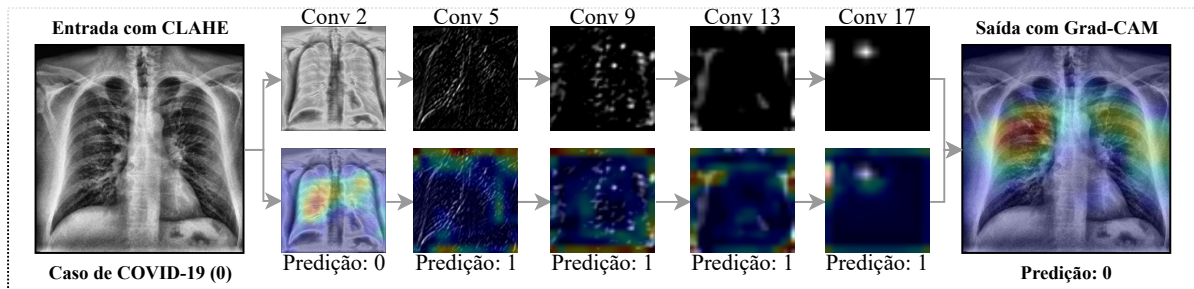


Fonte: Elaborada pelo autor.

que a mesma imagem foi utilizada/retirada do processo convolucional. A Figura 25 apresenta o comportamento do processo de extração de características da VGG-16 quando utilizado o Grad-CAM. Note que foi realizada a predição em cada camada junto ao Grad-CAM, utilizando a abordagem DEELE-Rad com 100 características radiômicas profundas treinada na Seção 5.3. Esse modelo obteve o melhor desempenho em comparação aos outros cenários com a VGG-16. Pode-se observar que na Conv 2 a predição é de COVID-19, e os mapas de calor (*heat-maps*) do Grad-CAM identificam a região pulmonar. Entretanto, esse comportamento torna-se diferente para as Conv 5, 9, 13 e 17, que realizam a predição como um caso Normal, e os mapas de calor do Grad-CAM identificam bordas em cada imagem. Por fim, a saída classifica corretamente o caso de COVID-19 e apresenta o mapa de calor na região pulmonar. Isto implica observar que, apesar de serem obtidas as imagens do processo convolucional, com o modelo proposto treinado, a tomada de decisão se torna mais explicável com o Grad-CAM por meios dos mapas

de calor apresentados. De maneira geral, o Grad-CAM serve como guia para interpretabilidade e explicabilidade para os modelos de aprendizado profundo.

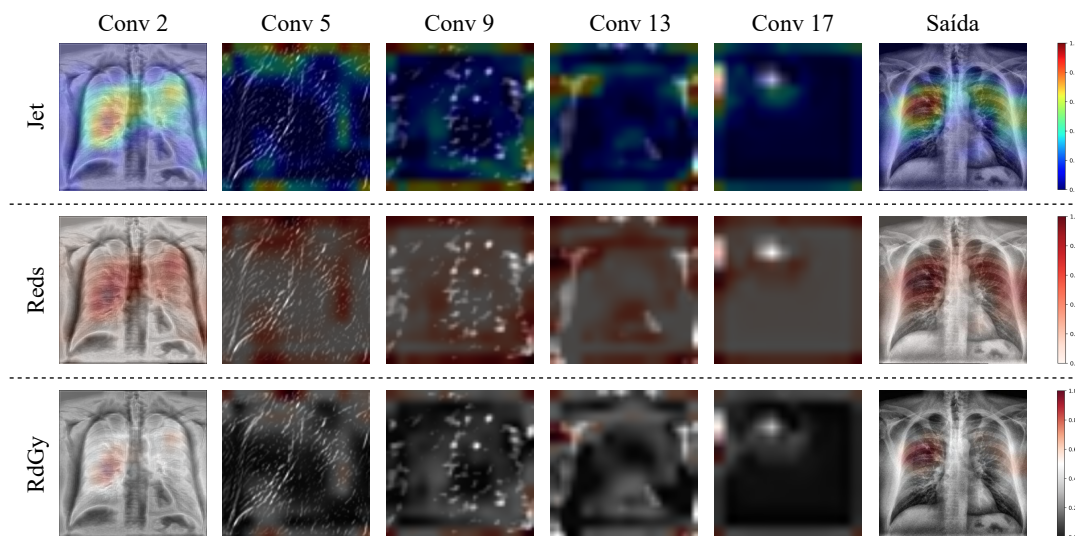
Figura 25 – Avaliação visual com Grad-CAM para um caso de COVID-19, mostrando o comportamento do Grad-CAM em relação a um exemplo extraído de cada processo de convolução. Observa-se que o caso de COVID-19 é rotulado como 0 e Normal (saudável) é 1.



Fonte: Elaborada pelo autor.

A Figura 26 apresenta a mesma perspectiva com diferentes mapas de calor. Desta maneira, é possível ter visualizações diferentes para um mesmo caso utilizando a abordagem DEELE-Rad. Os mapas de calor empregados foram o Jet, Reds e RdGy reverso⁷. O mapa de calor Jet é um dos mais conhecidos e amplamente utilizados na literatura, no qual é composto pelas cores azul, verde, amarelo e vermelho. Já o Reds é um mapa de cor sequencial, possui tons de cor vermelho predominante, transitando dos tons mais claros para os tons mais escuros. RdGy reverso é um mapa de cor divergente, inicia com tons de vermelho e indo em direção aos tons de cinza.

Figura 26 – Parametrização do Grad-CAM para um caso de COVID-19, apresentado a mesma perspectiva de interpretabilidade e explicabilidade com diferentes mapas de calor. Mapas de calor utilizados: Jet, Reds e RdGy reverso, respectivamente. A barra de cor varia entre 0 e 1. O valor 1 representa a intensidade mais forte identificada no Grad-CAM, e 0 a mais fraca.



Fonte: Elaborada pelo autor.

⁷ <https://matplotlib.org/stable/users/explain/colors/colormaps.html>

5.6 Discussões dos Resultados

Analisando os aspectos experimentais do DEELE-Rad sobre a aplicabilidade de características radiômicas profundas baseada em modelos de aprendizado profundo, são discutidos brevemente nesta seção em três aspectos de maneira complementar: (i) o desempenho dos modelos de aprendizado profundo de ponta a ponta; (ii) desempenho dos classificadores de aprendizado de máquina; e (iii) a avaliação visual proposta no DEELE-Rad.

Desempenho dos modelos de aprendizado profundo

A experimentação dos modelos de aprendizado profundo de ponta a ponta em diferentes cenários forneceu perspectivas de validação e versatilidade para abordagem DEELE-Rad. Foi possível observar que o melhor modelo foi a DenseNet201 obtendo uma média de Acurácia de 85,75% entre os quatro cenários. A DenseNet201 utiliza a ideia de convoluções separáveis em profundidade, além de conexões curtas entre as camadas de toda a rede, fortalecendo a propagação de características radiômicas profundas.

Desempenho dos algoritmos de aprendizado de máquina

Os resultados obtidos no desempenho dos modelos de aprendizado profundo parametrizam e guiam a comparação com os classificadores de aprendizado de máquina e a abordagem DEELE-Rad. Os cenários utilizados com 100, 128, 200 e 300 características radiômicas profundas extraídas dos modelos de aprendizado profundo forneceu versatilidade para testar e parametrizar os resultados com os classificadores de aprendizado de máquina e a abordagem DEELE-Rad. A abordagem DEELE-Rad obteve as melhores métricas de desempenho nos cenários com 128, 200 e 300 características, provenientes do modelo de aprendizado profundo da DenseNet201. Observou-se que no cenário com 300 características radiômicas profundas o resultado das métricas de desempenho ultrapassaram a faixa de mais de 10% (↑) de diferença na métrica da Acurácia, em comparação com o cenário com 100 características. Isto implica a observação de que, apesar dos testes em diferentes cenários, a abordagem DEELE-Rad obteve um desempenho melhor quando utilizou-se mais características radiômicas profundas provenientes do modelo da DenseNet201. De maneira geral, a abordagem DEELE-Rad melhora em 8,8% (↑) em relação aos modelos de aprendizado profundo.

Avaliação visual

O DEELE-Rad apresenta perspectivas de interpretabilidade e explicabilidade. Observou-se o comportamento do processo de extração de características radiômicas no modelo de aprendizado profundo da VGG-16. Foi empregado o Grad-CAM como método visual explicável de modo a incrementar a visualização na tomada de decisão, podendo ser utilizado como suporte para os especialistas em rotina clínica.

5.7 Considerações Finais

Neste capítulo foi apresentada a análise experimental para a validação do DEELE-Rad. O DEELE-Rad classifica casos de COVID-19 e Normal (saudável) de pacientes por meio de exames de imagens de raio-X. A abordagem DEELE-Rad utiliza dados coletados de diferentes fontes, assim como a melhora de contraste das imagens por meio do CLAHE. Foi empregado o ajuste-fino adicionando novas camadas para extração de características radiômicas mais relevantes em cada modelo de aprendizado profundo. Foram utilizados os classificadores tradicionais de aprendizado de máquina utilizando quantidades diferentes de características radiômicas profundas. A avaliação visual permitiu entender e observar os padrões do processo de extração de características radiômicas profundas em modelos de aprendizado profundo, em especial quando utilizou-se a VGG-16.

CONCLUSÕES

6.1 Pontos Positivos e Limitações

Este trabalho visou enriquecer com a aplicabilidade da área de pesquisa em Radiômica baseada em Aprendizado Profundo, utilizando avaliação visual e de desempenho. Nesta Dissertação de Mestrado, foi proposta a abordagem DEELE-Rad (*Deep Learning-based Radiomics*), que utiliza características radiômicas profundas por meio de modelos de aprendizado profundo, além de proporcionar meios para a avaliação visual dessas características.

No [Capítulo 4](#) todos os detalhes e o fluxograma da abordagem DEELE-Rad, foram apresentados. A análise experimental apresentada no [Capítulo 5](#) mostra a aplicação do DEELE-Rad em diferentes cenários de características radiômicas sobre um conjunto de exames de imagens de COVID-19. A abordagem DEELE-Rad mostra que em diferentes cenários de experimentações com as características radiômicas profundas extraídas dos modelos de aprendizado profundo, parametrizam e guiam as comparações em diferentes contextos. Foi possível observar que o DEELE-Rad obteve um melhor desempenho quando utilizou-se 300 características radiômicas profundas, tornando-o mais preciso em até 8,8% (↑) comparada aos modelos de aprendizado profundo. O DEELE-Rad têm como proposta a análise e avaliação visual de características radiômicas profundas, a fim de apresentar interpretabilidade e explicabilidade. Além disso, na avaliação visual o DEELE-Rad fornece uma parametrização de diferentes perspectivas por meio de mapas de calor, podendo ser utilizado como suporte e guia para tomada de decisão dos especialistas em rotina clínica.

As limitações da abordagem DEELE-Rad estão relacionadas ao uso de imagens coletadas de diversas fontes de repositórios, assim como à ausência de um especialista médico ou radiologista para orientar o processo de análise da avaliação visual proposta. Não foi possível realizar um teste de tempo computacional, uma vez que a abordagem DEELE-Rad não tinha nenhum concorrente equiparável para parametrização. Empiricamente, foi observado por meio

dos experimentos que a abordagem DEELE–Rad demanda um tempo considerável para a tarefa de análise visual de características radiômicas profundas. O tempo médio necessário para a análise visual é de aproximadamente 3 horas na plataforma utilizada para o desenvolvimento do DEELE–Rad (Seção 5.2), podendo variar dependendo da quantidade de características radiômicas profundas utilizadas.

Por fim, o presente trabalho abre possibilidades de trabalho em tópicos de pesquisas adicionais, para a utilização do DEELE–Rad em outros tipos de exames de imagens com a abordagem radiômica, ou áreas de aplicação, descritos a seguir na Seção 6.4.

6.2 Principais Contribuições

O DEELE–Rad destaca-se em relação aos trabalhos relacionados apresentados no Capítulo 3, sendo validado por pares e contribuindo para o estado da arte no contexto da Radiômica baseada em Aprendizado Profundo. O desenvolvimento do DEELE–Rad permitiu abordar as questões de pesquisa (QP) elencadas na Seção 1.3, sendo:

- **QP1.** *Como as principais características radiômicas extraídas de modelos de aprendizado profundo podem melhorar na predição em imagens de raio-X de COVID-19?*
- **QP2.** *Como avaliar o processo de tomada de decisão dos modelos de aprendizado profundo por meio da avaliação visual do processo de extração de características radiômicas?*

As respostas para essas QP podem ser evidenciadas nos seguintes aspectos:

- **QP1.** Foi proposta, implementada e validada uma nova abordagem para extração de características radiômicas profundas mais relevantes por meio de modelos de aprendizado profundo. A criação de um comitê de classificadores, baseado em algoritmos de aprendizado de máquina, demonstrou versatilidade ao empregar métodos como árvores de decisão e probabilidades. Em diferentes cenários de experimentação, foi possível observar o impacto e desempenho dessas características radiômicas profundas mais relevantes.
- **QP2.** Com suporte à avaliação visual, foi possível observar o processo de extração de características radiômicas profundas em uma arquitetura de aprendizado profundo, em especial a rede VGG-16. A aplicação de um método visual explicável proporcionou a análise necessária para compreender o processo de tomada de decisão por meio das amostras retiradas da rede. Além disso, a parametrização com diferentes mapas de calor foi utilizada para orientar e apresentar a tomada de decisão da rede. De maneira geral, essa abordagem pode oferecer suporte aos especialistas em uma rotina clínica, auxiliando-os na tomada de decisões.

6.3 Trabalhos Publicados

Nesta seção é apresentada a principal publicação durante o desenvolvimento deste projeto de Mestrado. Além disso, são apresentados outros trabalhos publicados, resultantes de atividades similares no contexto de aprendizado profundo realizadas em colaboração com o grupo de pesquisa, que contribuíram com o aprendizado do mestrando ao longo do desenvolvimento do seu projeto.

6.3.1 Publicação Principal

COSTA, M. V. L.; AGUIAR, E. J. de; RODRIGUES, L. S.; RAMOS, J. S.; TRAINA, C.; TRAINA, A. J. M. **A Deep Learning-based Radiomics Approach for COVID-19 Detection from CXR Images using Ensemble Learning Model.** In: *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.: s.n.], 2023. p. 517–522. DOI: 10.1109/CBMS58004.2023.00272. [QUALIS-CC A3].

Neste trabalho, foi proposta a abordagem inicial do método DEELE-Rad. O método explora a extração de características radiômicas profundas mais relevantes em modelos de aprendizado profundo. Ressalta-se que o DEELE-Rad propõe a utilização das características extraídas em um comitê de classificadores de aprendizado de máquina. Além disso, o DEELE-Rad fornece a avaliação visual no comportamento da extração das características, com intuito propor mais explicabilidade e interpretabilidade dos modelos de aprendizado profundo aos médicos e especialistas. O DEELE-Rad apresenta o primeiro passo de contribuição no estado da arte em estudos de radiômica baseada em aprendizado profundo, que utilizam a avaliação visual e as métricas de desempenho de características radiômicas profundas.

Ressalte-se que esta abordagem inicial foi convidada pelos revisores do CBMS 2023, em que foi estendida e submetida para o *Journal Health Information Science and Systems (HISS)*, com fator de impacto 6.0 em janeiro de 2024. Intitulada de: **DEELE-Rad: Exploiting deep radiomics features in deep learning models using COVID-19 chest X-ray images.**

6.3.2 Publicações em Colaboração

- AGUIAR, E. J. D.; *COSTA, M. V. L.*; TRAINA, C.; TRAINA, A. J. M. **Assessing vulnerabilities of deep learning explainability in medical image analysis under adversarial settings.** In: *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.: s.n.], 2023. p. 13–16. DOI: 10.1109/CBMS58004.2023.00184. [QUALIS-CC A3].

- RAMOS, J. S.; AGUIAR, E. J. de; BELIZARIO, I. V.; *COSTA, M. V. L.*; MACIEL, J. G.; CAZZOLATO, M. T.; TRAINA, C.; NOGUEIRA-BARBOSA, M. H.; TRAINA, A. J. M. **Analysis of vertebrae without fracture on spine mri to assess bone fragility: A comparison of traditional machine learning and deep learning.** *In: 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS).* [S.l.: s.n.], 2022. p. 78–83. DOI: 10.1109/CBMS55023.2022.00021. [QUALIS-CC A3].

6.3.3 Outras Atividades

Outras atividades acadêmicas foram realizadas no período do curso de Mestrado, sendo as seguintes:

- **Tutoria:**

- Tutor no curso de “**Recuperação de Imagens por Conteúdo**” na segunda edição do MBA em Inteligência Artificial e Big Data, realizado pelo Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, a partir de 15/05/2023, com duração de dois meses e carga horária semanal de 8 horas.

- Tutor no curso de “**Visualização de Dados**” na segunda edição do MBA em Inteligência Artificial e Big Data, realizado pelo Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, a partir de 25/06/2023, com duração de dois meses e carga horária semanal de 8 horas.

6.4 Trabalhos Futuros

O trabalho de Mestrado desenvolvido possibilita a realização de pesquisa sobre diferentes aspectos em trabalhos futuros. Algumas sugestões principais incluem:

- I. Ampliação do escopo de aplicação do DEELE–Rad:

- Explorar a aplicação do DEELE–Rad em outros tipos de exames de imagens médicas, como tomografia computadorizada ou ressonância magnética, e considerar a avaliação do DEELE–Rad em bases de dados multiclases.

- I. Variação de arquiteturas/modelos de aprendizado profundo:

- Investigar e explorar a avaliação visual em diferentes arquiteturas de aprendizado profundo de modo a incrementar as perspectivas de interpretabilidade e explicabilidade em diferentes modelos.

I. Estudo de caso com profissionais:

- Desenvolver ou aplicar um estudo de caso envolvendo profissionais da área, visando analisar e validar os resultados visuais obtidos pelo DEELE-Rad.

REFERÊNCIAS

- AGGARWAL, C. C. *et al.* **Data mining: the textbook**. [S.l.]: Springer, 2015. v. 1. Citado na página 37.
- AI, T.; YANG, Z.; HOU, H.; ZHAN, C.; CHEN, C.; LV, W.; TAO, Q.; SUN, Z.; XIA, L. Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. **Radiology**, Radiological Society of North America, v. 296, n. 2, p. E32–E40, 2020. Citado na página 23.
- ARDAKANI, A. A.; KANAFI, A. R.; ACHARYA, U. R.; KHADEM, N.; MOHAMMADI, A. Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks. **Computers in biology and medicine**, Elsevier, v. 121, p. 103795, 2020. Citado na página 24.
- BHATTACHARYA, S.; MADDIKUNTA, P. K. R.; PHAM, Q.-V.; GADEKALLU, T. R.; CHOWDHARY, C. L.; ALAZAB, M.; PIRAN, M. J. *et al.* Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey. **Sustainable cities and society**, Elsevier, v. 65, p. 102589, 2021. Citado na página 27.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001. Citado na página 37.
- CANDEMIR, S.; ANTANI, S. A review on lung boundary detection in chest x-rays. **International journal of computer assisted radiology and surgery**, Springer, v. 14, n. 4, p. 563–576, 2019. Citado na página 28.
- CARDENAS, D. A. C.; JR, J. R. F.; MORENO, R. A.; REBELO, M. d. F. de S.; KRIEGER, J. E.; GUTIERREZ, M. A. Automated radiographic bone suppression with deep convolutional neural networks. v. 11600, p. 116001D, 2021. Citado na página 24.
- CHADDAD, A.; HASSAN, L.; DESROSIERS, C. Deep Radiomic Analysis for Predicting Coronavirus Disease 2019 in Computerized Tomography and X-Ray Images. **IEEE Transactions on Neural Networks and Learning Systems**, v. 33, n. 1, p. 3–11, jan 2022. ISSN 2162-237X. Disponível em: <<https://ieeexplore.ieee.org/document/9582743/>>. Citado nas páginas 34, 51, 53 e 55.
- CHANDRA, T. B.; VERMA, K.; SINGH, B. K.; JAIN, D.; NETAM, S. S. Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble. **Expert systems with applications**, Elsevier, v. 165, p. 113909, 2021. Citado nas páginas 34, 50, 53 e 55.
- CHENG, C. P. Medical imaging modalities and protocols. In: **Handbook of Vascular Motion**. [S.l.]: Elsevier, 2019. p. 23–43. Citado na página 28.
- CHENG, S. C.; CHANG, Y. C.; Fan Chiang, Y. L.; CHIEN, Y. C.; CHENG, M.; YANG, C. H.; HUANG, C. H.; HSU, Y. N. First case of Coronavirus Disease 2019 (COVID-19) pneumonia in

Taiwan. **Journal of the Formosan Medical Association**, Elsevier Ltd, v. 119, n. 3, p. 747–751, 2020. ISSN 18760821. Disponível em: <<https://doi.org/10.1016/j.jfma.2020.02.007>>. Citado na página 23.

CHOLLET, F. Deep learning with python. 2018. Citado na página 39.

_____. **Grad-CAM class activation visualization**. 2023. <https://keras.io/examples/vision/grad_cam/>. Citado na página 47.

CHOLLET, F. *et al.* **Keras**. 2015. <<https://keras.io>>. Citado na página 62.

CHOWDHURY, M. E.; RAHMAN, T.; KHANDAKAR, A.; MAZHAR, R.; KADIR, M. A.; MAHBUB, Z. B.; ISLAM, K. R.; KHAN, M. S.; IQBAL, A.; EMADI, N. A. *et al.* Can ai help in screening viral and covid-19 pneumonia? **IEEE Access**, IEEE, v. 8, p. 132665–132676, 2020. Citado na página 28.

COHEN, J. P.; MORRISON, P.; DAO, L.; ROTH, K.; DUONG, T. Q.; GHASSEMI, M. COVID-19 image data collection: Prospective predictions are the future. **arXiv 2006.11988**, 2020. Disponível em: <<https://github.com/ieee8023/covid-chestxray-dataset>>. Citado na página 59.

COSTA, M. V. L.; AGUIAR, E. J. de; RODRIGUES, L. S.; RAMOS, J. S.; TRAINA, C.; TRAINA, A. J. M. A Deep Learning-based Radiomics Approach for COVID-19 Detection from CXR Images using Ensemble Learning Model. In: **2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.: s.n.], 2023. p. 517–522. Citado nas páginas 25 e 53.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random forests. **Ensemble machine learning: Methods and applications**, Springer, p. 157–175, 2012. Citado na página 37.

DROSTEN, C.; GÜNTHER, S.; PREISER, W.; WERF, S. van der; BRODT, H.-R.; BECKER, S.; RABENAU, H.; PANNING, M.; KOLESNIKOVA, L.; FOUCHIER, R. A.; BERGER, A.; BURGUIÈRE, A.-M.; CINATL, J.; EICKMANN, M.; ESCRIOU, N.; GRYWNA, K.; KRAMME, S.; MANUGUERRA, J.-C.; MÜLLER, S.; RICKERTS, V.; STÜRMER, M.; VIETH, S.; KLENK, H.-D.; OSTERHAUS, A. D.; SCHMITZ, H.; DOERR, H. W. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. **New England Journal of Medicine**, v. 348, n. 20, p. 1967–1976, 2003. ISSN 0028-4793. Citado na página 23.

EKMAN, M. **Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, NLP, and Transformers Using TensorFlow**. [S.l.]: Addison-Wesley Professional, 2021. Citado nas páginas 38, 42 e 44.

FACELI, K.; LORENA, A. C.; GAMA, J.; ALMEIDA, T. A. d.; CARVALHO, A. C. P. d. L. F. d. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [S.l.: s.n.], 2021. Citado na página 67.

FARHAT, H.; SAKR, G. E.; KILANY, R. Deep learning applications in pulmonary medical imaging: recent updates and insights on covid-19. **Machine vision and applications**, Springer, v. 31, n. 6, p. 1–42, 2020. Citado na página 28.

FERREIRA, J. R.; CARDENAS, D. A. C.; MORENO, R. A.; REBELO, M. d. F. de S.; KRIEGER, J. E.; GUTIERREZ, M. A. Novel chest radiographic biomarkers for covid-19 using radiomic features associated with diagnostics and outcomes. **Journal of Digital Imaging**, Springer, v. 34, n. 2, p. 297–307, 2021. Citado nas páginas 24, 34, 50, 53 e 55.

FUKUSHIMA, K. **Biological Cybernetics Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position**. 1980. 202 p. Citado na página 39.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [S.l.]: "O'Reilly Media, Inc.", 2022. Citado na página 38.

GILLIES, R. J.; KINAHAN, P. E.; HRICAK, H. Radiomics: images are more than pictures, they are data. **Radiology**, Radiological Society of North America, v. 278, n. 2, p. 563–577, 2016. Citado nas páginas 24, 29 e 30.

GOEL, N.; YADAV, A.; SINGH, B. M. Medical image processing: a review. In: IEEE. **2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)**. [S.l.], 2016. p. 57–62. Citado na página 28.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing (preview)**. Prentice Hall, 2008. Citado nas páginas 31, 33 e 34.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado nas páginas 39, 40 e 44.

GROOT, R. J. de; BAKER, S. C.; BARIC, R. S.; BROWN, C. S.; DROSTEN, C.; ENJUANES, L.; FOUCHIER, R. A. M.; GALIANO, M.; GORBALENYA, A. E.; MEMISH, Z. A.; PERLMAN, S.; POON, L. L. M.; SNIJDER, E. J.; STEPHENS, G. M.; WOO, P. C. Y.; ZAKI, A. M.; ZAMBON, M.; ZIEBUHR, J. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. **Journal of Virology**, v. 87, n. 14, p. 7790–7792, 2013. ISSN 0022-538X. Citado na página 23.

GUAN, W.-j.; NI, Z.-y.; HU, Y.; LIANG, W.-h.; OU, C.-q.; HE, J.-x.; LIU, L.; SHAN, H.; LEI, C.-l.; HUI, D. S. *et al.* Clinical characteristics of coronavirus disease 2019 in china. **New England journal of medicine**, Mass Medical Soc, v. 382, n. 18, p. 1708–1720, 2020. Citado na página 23.

HAN, J.; PEI, J.; TONG, H. **Data mining: concepts and techniques**. [S.l.]: Morgan kaufmann, 2022. Citado na página 37.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. H. Textural features for image classification. **IEEE Transactions on systems, man, and cybernetics**, Ieee, n. 6, p. 610–621, 1973. Citado nas páginas 31 e 34.

HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2001. Citado na página 37.

_____. **Neural networks and learning machines, 3/E**. [S.l.]: Pearson Education India, 2009. Citado na página 36.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778. Citado nas páginas 42 e 43.

HEIDARI, M.; MIRNIAHARIKANDEHEI, S.; KHUZANI, A. Z.; DANALA, G.; QIU, Y.; ZHENG, B. Detecting covid-19 infected pneumonia from x-ray images using a deep learning model with image preprocessing algorithm. v. 11597, p. 115970V, 2021. Citado na página 24.

- HU, Z.; YANG, Z.; LAFATA, K. J.; YIN, F.-F.; WANG, C. A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images. **Med. Phys.**, Wiley, v. 49, n. 5, p. 3213–3222, maio 2022. Citado nas páginas 24, 34, 51, 53 e 55.
- HUANG, C.; WANG, Y.; LI, X.; REN, L.; ZHAO, J.; HU, Y.; ZHANG, L.; FAN, G.; XU, J.; GU, X. *et al.* Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. **The lancet**, Elsevier, v. 395, n. 10223, p. 497–506, 2020. Citado na página 23.
- HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 4700–4708. Citado na página 43.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. **The Journal of physiology**, Wiley Online Library, v. 195, n. 1, p. 215–243, 1968. Citado na página 39.
- KAUR, D.; USLU, S.; RITTICHER, K. J.; DURRESI, A. Trustworthy artificial intelligence: a review. **ACM Computing Surveys (CSUR)**, ACM New York, NY, v. 55, n. 2, p. 1–38, 2022. Citado na página 46.
- KERMANY, D. S.; GOLDBAUM, M.; CAI, W.; VALENTIM, C. C.; LIANG, H.; BAXTER, S. L.; MCKEOWN, A.; YANG, G.; WU, X.; YAN, F. *et al.* Identifying medical diagnoses and treatable diseases by image-based deep learning. **Cell**, Elsevier, v. 172, n. 5, p. 1122–1131, 2018. Citado na página 59.
- KOÇAK, B.; DURMAZ, E. Ş.; ATEŞ, E.; KILIÇKESMEZ, Ö. Radiomics with artificial intelligence: a practical guide for beginners. **Diagnostic and interventional radiology**, Turkish Society of Radiology, v. 25, n. 6, p. 485, 2019. Citado nas páginas 24, 29, 30 e 49.
- KOYUNCU, H.; BARSTUĞAN, M. Covid-19 discrimination framework for x-ray images by considering radiomics, selective information, feature ranking, and a novel hybrid classifier. **Signal Processing: Image Communication**, Elsevier, p. 116359, 2021. Citado nas páginas 34, 50, 53 e 55.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, 2012. Citado na página 41.
- LAFATA, K. J.; WANG, Y.; KONKEL, B.; YIN, F. F.; BASHIR, M. R. Radiomics: a primer on high-throughput image phenotyping. **Abdominal Radiology**, Springer, v. 47, p. 2986–3002, 9 2022. ISSN 23660058. Citado nas páginas 29 e 45.
- LAMBIN, P.; LEIJENAAR, R. T.; DEIST, T. M.; PEERLINGS, J.; JONG, E. E. D.; TIMMEREN, J. V.; SANDULEANU, S.; LARUE, R. T.; EVEN, A. J.; JOCHEMS, A. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. **Nature reviews Clinical oncology**, Nature Publishing Group, v. 14, n. 12, p. 749–762, 2017. Citado na página 30.
- LAMBIN, P.; RIOS-VELAZQUEZ, E.; LEIJENAAR, R.; CARVALHO, S.; STIPHOUT, R. G. V.; GRANTON, P.; ZEGERS, C. M.; GILLIES, R.; BOELLARD, R.; DEKKER, A. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. **European journal of cancer**, Elsevier, v. 48, n. 4, p. 441–446, 2012. Citado na página 29.

- LECUN, Y.; BOSER, B.; DENKER, J.; HENDERSON, D.; HOWARD, R.; HUBBARD, W.; JACKEL, L. Handwritten digit recognition with a back-propagation network. **Advances in neural information processing systems**, v. 2, 1989. Citado nas páginas 15, 39 e 41.
- LITMANOVICH, D. E.; CHUNG, M.; KIRKBRIDE, R. R.; KICSKA, G.; KANNE, J. P. Review of chest radiograph findings of covid-19 pneumonia and suggested reporting language. **Journal of thoracic imaging**, Wolters Kluwer, v. 35, n. 6, p. 354–360, 2020. Citado nas páginas 23 e 28.
- LOHMANN, P.; BOUSABARAH, K.; HOVELS, M.; TREUER, H. Radiomics in radiation oncology—basics, methods, and limitations. **Strahlentherapie und Onkologie**, Springer Science and Business Media Deutschland GmbH, v. 196, p. 848–855, 10 2020. ISSN 0179-7158. Disponível em: <<https://link.springer.com/10.1007/s00066-020-01663-3>>. Citado nas páginas 45 e 49.
- MANASWI, N. K.; MANASWI, N. K.; JOHN, S. **Deep learning with applications using python**. [S.l.]: Springer, 2018. Citado na página 40.
- MARQUES, G.; AGARWAL, D.; DÍEZ, I. de la T. Automated medical diagnosis of covid-19 through efficientnet convolutional neural network. **Applied soft computing**, Elsevier, v. 96, p. 106691, 2020. Citado na página 24.
- MEIRELLES, G. d. S. P. Covid-19: uma breve atualização para radiologistas. **Radiologia Brasileira**, SciELO Brasil, v. 53, p. 320–328, 2020. Citado na página 23.
- MELLO, R. F. d.; PONTI, M. A. **Machine learning: a practical approach on the statistical learning theory**. [S.l.]: Springer, 2018. Citado na página 37.
- MOURA, L. V. de; DARTORA, C. M.; OLIVEIRA, C. M. de; BARROS, R. C.; SILVA, A. M. M. da. A novel approach to differentiate covid-19 pneumonia in chest x-ray. In: IEEE. **2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)**. [S.l.], 2020. p. 446–451. Citado na página 24.
- NOVAKSEAN. **File:Normal axial T2-weighted MR image of the brain.jpg**. 2015. <https://commons.wikimedia.org/wiki/File:Normal_axial_T2-weighted_MR_image_of_the_brain.jpg>. (Acessado em: 05 de jun. de 2022). Citado na página 28.
- OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In: SPRINGER. **International conference on advances in pattern recognition**. [S.l.], 2001. p. 399–408. Citado na página 35.
- PISANO, E. D.; ZONG, S.; HEMMINGER, B. M.; DELUCA, M.; JOHNSTON, R. E.; MULLER, K.; BRAEUNING, M. P.; PIZER, S. M. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. **Journal of Digital imaging**, Springer, v. 11, p. 193–200, 1998. Citado nas páginas 33, 59 e 60.
- PIZER, S. M.; AMBURN, E. P.; AUSTIN, J. D.; CROMARTIE, R.; GESELOWITZ, A.; GREER, T.; ROMENY, B. ter H.; ZIMMERMAN, J. B.; ZUIDERVELD, K. Adaptive histogram equalization and its variations. **Computer vision, graphics, and image processing**, Elsevier, v. 39, n. 3, p. 355–368, 1987. Citado nas páginas 33 e 60.
- QIU, J.; WU, Q.; DING, G.; XU, Y.; FENG, S. A survey of machine learning for big data processing. **EURASIP Journal on Advances in Signal Processing**, Springer, v. 2016, p. 1–16, 2016. Citado na página 36.

- RAHMAN, T.; KHANDAKAR, A.; QIBLAWEY, Y.; TAHIR, A.; KIRANYAZ, S.; KASHEM, S. B. A.; ISLAM, M. T.; MAADEED, S. A.; ZUGHAIER, S. M.; KHAN, M. S. *et al.* Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. **Computers in biology and medicine**, Elsevier, v. 132, p. 104319, 2021. Citado nas páginas 24 e 28.
- RANDEN, T. **Brodatz Textures**. 1999. <<https://www.ux.uis.no/~tranden/brodatz.html>>. (Acessado em: 30 de jun. de 2022). Citado na página 34.
- ROBERTS, M.; DRIGGS, D.; THORPE, M.; GILBEY, J.; YEUNG, M.; URSPRUNG, S.; AVILES-RIVERO, A. I.; ETMANN, C.; MCCAGUE, C.; BEER, L. *et al.* Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. **Nature Machine Intelligence**, Nature Publishing Group, v. 3, n. 3, p. 199–217, 2021. Citado nas páginas 49 e 60.
- RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M. *et al.* Imagenet large scale visual recognition challenge. **International journal of computer vision**, Springer, v. 115, p. 211–252, 2015. Citado na página 41.
- SCAPICCHIO, C.; GABELLONI, M.; BARUCCI, A.; CIONI, D.; SABA, L.; NERI, E. A deep look into radiomics. **La radiologia medica**, Springer, v. 126, n. 10, p. 1296–1311, 2021. Citado nas páginas 29, 34, 35 e 49.
- SHANKAR, K.; PERUMAL, E. A novel hand-crafted with deep learning features based fusion model for covid-19 diagnosis and classification using chest x-ray images. **Complex & Intelligent Systems**, Springer, v. 7, n. 3, p. 1277–1293, 2021. Citado na página 24.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Citado na página 41.
- SULLIVAN, D. C.; OBUCHOWSKI, N. A.; KESSLER, L. G.; RAUNIG, D. L.; GATSONIS, C.; HUANG, E. P.; KONDRATOVICH, M.; MCSHANE, L. M.; REEVES, A. P.; BARBORIAK, D. P. *et al.* Metrology standards for quantitative imaging biomarkers. **Radiology**, Radiological Society of North America, v. 277, n. 3, p. 813–825, 2015. Citado na página 30.
- SUNDARARAJAN, D. **Digital image processing: a signal processing and algorithmic approach**. [S.l.]: Springer, 2017. Citado na página 31.
- TIMMEREN, J. E. V.; CESTER, D.; TANADINI-LANG, S.; ALKADHI, H.; BAESSLER, B. Radiomics in medical imaging—“how-to” guide and critical reflection. **Insights into imaging**, SpringerOpen, v. 11, n. 1, p. 1–16, 2020. Citado na página 30.
- TIMMEREN, J. E. van; CESTER, D.; TANADINI-LANG, S.; ALKADHI, H.; BAESSLER, B. Radiomics in medical imaging—“how-to” guide and critical reflection. **Insights into Imaging**, SpringerOpen, v. 11, n. 1, p. 1–16, 2020. Citado nas páginas 27 e 30.
- TURKOGLU, M. Covidetectionet: Covid-19 diagnosis system based on x-ray images using features selected from pre-learned deep features ensemble. **Applied Intelligence**, Springer, v. 51, n. 3, p. 1213–1226, 2021. Citado na página 24.
- VAPNIK, V. **The nature of statistical learning theory**. [S.l.]: Springer science & business media, 1999. Citado na página 37.

VARGHESE, B. A.; CEN, S. Y.; HWANG, D. H.; DUDDALWAR, V. A. Texture analysis of imaging: what radiologists need to know. **American Journal of Roentgenology**, Am Roentgen Ray Soc, v. 212, n. 3, p. 520–528, 2019. Citado na página 29.

VARSHNEY, K. R. **Trustworthy Machine Learning**. Chappaqua, NY, USA: Independently Published, 2022. Citado na página 46.

VELDEN, B. H. Van der; KUIJF, H. J.; GILHUIJS, K. G.; VIERGEVER, M. A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. **Medical Image Analysis**, Elsevier, v. 79, p. 102470, 2022. Citado na página 46.

VIAL, A.; STIRLING, D.; FIELD, M.; ROS, M.; RITZ, C.; CAROLAN, M.; HOLLOWAY, L.; MILLER, A. A. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. **Transl Cancer Res**, v. 7, n. 3, p. 803–816, 2018. Citado nas páginas 30, 45 e 49.

WANG, H.; WANG, L.; LEE, E. H.; ZHENG, J.; ZHANG, W.; HALABI, S.; LIU, C.; DENG, K.; SONG, J.; YEOM, K. W. Decoding covid-19 pneumonia: comparison of deep learning and radiomics ct image signatures. **European journal of nuclear medicine and molecular imaging**, Springer, v. 48, n. 5, p. 1478–1486, 2021. Citado na página 34.

WANG, L.; KELLY, B.; LEE, E. H.; WANG, H.; ZHENG, J.; ZHANG, W.; HALABI, S.; LIU, J.; TIAN, Y.; HAN, B. *et al.* Multi-classifier-based identification of covid-19 from chest computed tomography using generalizable and interpretable radiomics features. **European journal of radiology**, Elsevier, v. 136, p. 109552, 2021. Citado na página 24.

WONG, H. Y. F.; LAM, H. Y. S.; FONG, A. H.-T.; LEUNG, S. T.; CHIN, T. W.-Y.; LO, C. S. Y.; LUI, M. M.-S.; LEE, J. C. Y.; CHIU, K. W.-H.; CHUNG, T. W.-H. *et al.* Frequency and distribution of chest radiographic findings in patients positive for covid-19. **Radiology**, Radiological Society of North America, v. 296, n. 2, p. E72–E78, 2020. Citado na página 23.

WU, G.; JOCHEMS, A.; IBRAHIM, A.; YAN, C.; SANDULEANU, S.; WOODRUFF, H. C.; LAMBIN, P. Structural and functional radiomics for lung cancer. **European Journal of Nuclear Medicine and Molecular Imaging**, Springer, p. 1–14, 2021. Citado na página 30.

XIE, X.; ZHONG, Z.; ZHAO, W.; ZHENG, C.; WANG, F.; LIU, J. Chest ct for typical coronavirus disease 2019 (covid-19) pneumonia: relationship to negative rt-pcr testing. **Radiology**, Radiological Society of North America, v. 296, n. 2, p. E41–E45, 2020. Citado na página 23.

YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, Springer, v. 9, n. 4, p. 611–629, 2018. Citado na página 40.

YOON, S. H.; LEE, K. H.; KIM, J. Y.; LEE, Y. K.; KO, H.; KIM, K. H.; PARK, C. M.; KIM, Y.-H. Chest radiographic and ct findings of the 2019 novel coronavirus disease (covid-19): analysis of nine patients treated in korea. **Korean journal of radiology**, The Korean Society of Radiology, v. 21, n. 4, p. 494–500, 2020. Citado na página 23.

YOUSEFI, B.; KAWAKITA, S.; AMINI, A.; AKBARI, H.; ADVANI, S. M.; AKHLOUFI, M.; MALDAGUE, X. P.; AHADIAN, S. Impartially validated multiple deep-chain models to detect covid-19 in chest x-ray using latent space radiomics. **Journal of Clinical Medicine**, MDPI, v. 10, n. 14, p. 3100, 2021. Citado nas páginas 51, 53 e 55.

- ZHANG, G.; JIANG, S.; YANG, Z.; GONG, L.; MA, X.; ZHOU, Z.; BAO, C.; LIU, Q. Automatic nodule detection for lung cancer in ct images: A review. **Computers in biology and medicine**, Elsevier, v. 103, p. 287–300, 2018. Citado na página 28.
- ZHANG, X.; ZHANG, Y.; ZHANG, G.; QIU, X.; TAN, W.; YIN, X.; LIAO, L. **Deep Learning With Radiomics for Disease Diagnosis and Treatment: Challenges and Potential**. [S.l.]: Frontiers Media S.A., 2022. Citado na página 49.
- ZHAO, G.; BAI, J.; WANG, P. P.; YANG, G.; SHI, L.; TAO, Y.; LIN, Y.; CHENG, J. Hs-gs: A method for multicenter mr image standardization. **IEEE Access**, IEEE, v. 8, p. 158512–158522, 2020. Citado na página 27.
- ZHAO, J.; ZHANG, Y.; HE, X.; XIE, P. Covid-ct-dataset: a ct scan dataset about covid-19. **arXiv preprint arXiv:2003.13865**, 2020. Citado na página 28.
- ZHU, Q.; YE, H.; SUN, L.; LI, Z.; WANG, R.; SHI, F.; SHEN, D.; ZHANG, D. Gacdn: generative adversarial feature completion and diagnosis network for covid-19. **BMC Medical Imaging**, BioMed Central, v. 21, n. 1, p. 1–13, 2021. Citado nas páginas 34, 52, 53 e 55.
- ZWANENBURG, A.; VALLIÈRES, M.; ABDALAH, M. A.; AERTS, H. J.; ANDREARCZYK, V.; APTE, A.; ASHRAFINIA, S.; BAKAS, S.; BEUKINGA, R. J.; BOELLAARD, R. *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. **Radiology**, Radiological Society of North America, v. 295, n. 2, p. 328–338, 2020. Citado na página 30.

