



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
MESTRADO ACADÊMICO EM ENGENHARIA ELÉTRICA**

MATHEUS ARAUJO DOS SANTOS

**UMA ABORDAGEM TOTALMENTE AUTOMÁTICA PARA DIAGNÓSTICO DE
COVID-19 EM IMAGENS DE TC: INTEGRANDO SEGMENTAÇÃO DE PULMÕES,
*FINE-TUNING E VISUALIZAÇÃO EM GRAD-CAM***

FORTALEZA

2025

MATHEUS ARAUJO DOS SANTOS

UMA ABORDAGEM TOTALMENTE AUTOMÁTICA PARA DIAGNÓSTICO DE COVID-19
EM IMAGENS DE TC: INTEGRANDO SEGMENTAÇÃO DE PULMÕES, *FINE-TUNING* E
VISUALIZAÇÃO EM *GRAD-CAM*

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Elétrica do Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Elétrica. Área de Concentração: Sistemas de Energia Elétrica.

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho.

Coorientador: Prof. Dr. Solon Alves Peixoto.

FORTALEZA

2025

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

D762a dos Santos, Matheus.

Uma Abordagem Totalmente Automática para Diagnóstico de COVID-19 em Imagens de TC:
Integrando Segmentação de Pulmões, Fine-Tuning e Visualização em Grad-CAM / Matheus dos Santos. –
2025.

62 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-
Graduação em Engenharia Elétrica, Fortaleza, 2025.

Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho.

Coorientação: Prof. Dr. Solon Alves Peixoto.

1. Tomografia Computadorizada (TC). 2. COVID-19. 3. Aprendizado Profundo. 4. Segmentação
Profunda. 5. Transferência de Aprendizado. I. Título.

CDD 621.3

MATHEUS ARAUJO DOS SANTOS

UMA ABORDAGEM TOTALMENTE AUTOMÁTICA PARA DIAGNÓSTICO DE COVID-19
EM IMAGENS DE TC: INTEGRANDO SEGMENTAÇÃO DE PULMÕES, *FINE-TUNING* E
VISUALIZAÇÃO EM *GRAD-CAM*

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Elétrica do Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Elétrica. Área de Concentração: Sistemas de Energia Elétrica.

Aprovada em: 26/02/2025

BANCA EXAMINADORA

Prof. Dr. Pedro Pedrosa Rebouças Filho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Solon Alves Peixoto (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Bismark Claure Torrico
Universidade Federal do Ceará (UFC)

Prof. Dr. Aldísio Gonçalves Medeiros
Universidade Federal do Ceará (UFC)

Ao meu querido filho, Arthur Pedro.

AGRADECIMENTOS

A Deus, primeiramente, por sempre iluminar minha mente em momentos em que era necessário compreensão e paciência. Parafraseando Santo Tomás de Aquino, d'Ele vem a agudeza de entender, a sutileza de relevar, a facilidade de aprender e a graça abundante de falar e de escrever. À Santíssima Virgem, a quem sou devoto, por sua constante intercessão por mim e por seu consolo em momentos de dificuldade. Ao meu pai, Daniel, que já partiu para a eternidade e que tenho a certeza que também está sempre a interceder dos Céus por mim e por minha caminhada aqui na Terra.

À minha família, por todo seu carinho e amor ao longo dos anos. À minha mãe, Selma, que me formou como pessoa e me ensinou a sempre dar o meu melhor em todas as circunstâncias. Agradeço também a ela, que como professora, me mostrou toda a preciosidade da educação e toda a nobreza que há na docência e no ambiente acadêmico. Ademais, deixo meus agradecimentos às minhas irmãs, tios, tias, avó, cunhada, cunhados, sobrinhos e primos, por todo o convívio e presença em momentos bons e ruins.

À minha esposa, Letícia, por ser minha companheira de todas as horas e um pilar em nossa família. Deixo aqui toda a minha admiração por sua doação diária ao nosso filho e à nossa família. Agradeço ainda por todos os momentos juntos ao longo desses vários anos e por toda a alegria e simpatia que me proporciona diariamente, fazendo com que eu busque ser sempre a minha melhor versão.

Ao meu filho, Arthur Pedro, por me mostrar o mais puro amor personificado. Desde seu nascimento tem sido meu combustível para quando faltam as forças e quando o cansaço bate. Agradeço mais uma vez a Deus por tê-lo colocado em nossas vidas e tê-lo feito cheio de saúde, força, alegria e curiosidade.

Ao Laboratório de Processamento de Imagens, Sinais e Computação Aplicada (LAPISCO), e aos seus membros, por me proporcionarem diariamente desafios, suporte e aprendizado ao longo desses últimos anos. Aqui eu me encontrei como profissional e pesquisador, e fui acolhido de braços abertos por todos. Agradeço a cada membro, que não são só meus colegas de trabalho, mas amigos que fiz para a vida. Em especial deixo meus agradecimentos ao Prof. Dr. Solon Alves Peixoto, que além de coorientador é também um grande amigo que guardo no coração, e ao Prof. Dr. Pedro Pedrosa Rebouças Filho, meu eterno professor e orientador, por todas as oportunidades e ensinamentos dados, não só para a academia e no âmbito profissional, mas também para a vida.

Por fim, e não menos importante, agradeço à Universidade Federal do Ceará (UFC), ao Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) e ao Departamento de Engenharia Elétrica (DEE), por toda a estrutura e acolhimento. Agradeço a todos os mestres da instituição que contribuíram direta ou indiretamente com este trabalho e por todos os conhecimentos passados, assim como pela formação como cidadão e profissional.

"A beleza é a marca do que é bem feito, seja um universo, seja um objeto." (São Tomás de Aquino)

RESUMO

A pandemia causada pelo vírus COVID-19 destacou a necessidade de ferramentas de diagnóstico médico eficientes e automatizadas para auxiliar os profissionais de saúde na identificação rápida e precisa da doença. Este estudo apresenta uma abordagem totalmente automática para classificar COVID-19 em imagens de tomografia computadorizada (TC) torácica do conjunto de dados SARS-CoV-2, empregando aprendizado profundo, segmentação pulmonar automática e modelos de ajuste fino para o problema específico. Em um primeiro momento, foram realizados experimentos de transferência de aprendizado que revelaram que o modelo *deep* MobileNet se mostrou como a arquitetura mais eficaz para extração de características desses exames. Após escolha do modelo, a Detectron2 foi integrada para a segmentação pulmonar. Com os exames segmentados, foi realizado o ajuste fino do modelo MobileNet previamente selecionado. Com isso, o desempenho da classificação melhorou significativamente, o que foi comprovado através de uma análise visual da interpretabilidade do modelo usando Grad-CAM. Como resultado, o método totalmente automático proposto alcançou métricas de avaliação superiores em comparação a outros métodos na literatura, atingindo valores de 99,60% para acurácia, precisão e *F1-Score*, juntamente com 99,59% para *recall* e 99,20% para o Coeficiente de Correlação de Matthews (MCC). Isso demonstrou que, com imagens segmentadas e o modelo refinado, a rede se concentrou nas regiões pulmonares, melhorando a compreensão do diagnóstico do modelo.

Palavras-chave: Tomografia Computadorizada (TC); COVID-19; Aprendizado Profundo; Segmentação Profunda; Ajuste Fino; Transferência de Aprendizado.

ABSTRACT

The pandemic caused by the COVID-19 virus has highlighted the need for efficient and automated medical diagnostic tools to assist healthcare professionals in the fast and accurate identification of the disease. This study introduces a fully automatic approach for classifying COVID-19 in thoracic computed tomography (CT) images from the SARS-CoV-2 CT-scan dataset, employing deep learning, automatic lung segmentation, and fine-tuning models to the specific problem. In the first experiment, transfer learning experiments were performed, which revealed that the *deep* MobileNet model was the most effective architecture for extracting features from these exams. After choosing the model, Detectron2 was integrated for lung segmentation. With the segmented exams, the previously selected MobileNet model was fine-tuned. As a result, the classification performance improved significantly, which was confirmed through a visual analysis of the model's interpretability using Grad-CAM. As a result, the proposed fully automatic method achieved superior evaluation metrics compared to other methods in the literature, reaching values of 99.60% for accuracy, precision, and F1-Score, along with 99.59% for recall and 99.20% for the Matthews Correlation Coefficient (MCC). This demonstrated that the network focused on the lung regions with segmented images and the refined model, improving the model's diagnostic understanding.

Keywords: Computed Tomography (CT); COVID-19; Deep Learning; Deep Segmentation; Fine-Tuning; Transfer Learning.

LISTA DE FIGURAS

Figura 1 – Ilustração explicativa de como acontece a operação de convolução.	23
Figura 2 – Representação da arquitetura do <i>Framework Detectron2</i> para segmentação de imagens médicas.	27
Figura 3 – Exemplos de Grad-CAM em exames de TC torácico aplicados às imagens originais.	35
Figura 4 – Acima: Processo de <i>transfer learning</i> , no qual o melhor modelo avaliado foi selecionado para a metodologia proposta. Abaixo: Fluxograma da metodologia proposta, uma abordagem automática para classificação de COVID em imagens de TC de pulmão, integrando segmentação, <i>fine-tuning</i> do melhor modelo e validando sua interpretabilidade com Grad-CAM.	36
Figura 5 – Exemplos de imagens dos datasets utilizados neste estudo. Acima, algumas amostras de imagens do SARS-CoV-2 CT-Scan (Classe Non-Covid à esquerda e Classe Covid à direita); abaixo, imagens do dataset adquirido junto ao Hospital Walter Cantídio, para segmentação de pulmão.	37
Figura 6 – Amostras de TC dos dois datasets desse estudo segmentadas pela Detectron2. Acima, uma imagem do dataset de treinamento e a respectiva máscara em rosa. Abaixo, uma amostra do dataset SARS-CoV-2 CT-Scan e a respectiva segmentação.	46
Figura 7 – Mais mostras de TC dos dois datasets desse estudo segmentadas pela Detectron2. Acima, uma imagem do dataset de treinamento e a respectiva máscara em rosa. Abaixo, uma amostra do dataset SARS-CoV-2 CT-Scan e a respectiva segmentação.	47
Figura 8 – Curvas de acurácia para o treinamento e validação no ajuste fino do modelo MobileNet.	48
Figura 9 – Curvas de <i>loss</i> para o treinamento e validação no ajuste fino do modelo MobileNet.	49
Figura 10 – Acima e à esquerda, uma amostra da classe Non-Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.	51

Figura 11 – Acima e à esquerda, uma amostra da classe Non-Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.	51
Figura 12 – Acima e à esquerda, uma amostra da classe Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.	52
Figura 13 – Acima e à esquerda, uma amostra da classe Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.	52

LISTA DE TABELAS

Tabela 1 – Descrição das arquiteturas de CNN e suas configurações utilizadas na abordagem.	29
Tabela 2 – Descrição dos classificadores utilizados na abordagem.	31
Tabela 3 – Resumo da composição do conjunto de dados SARS-CoV-2 CT-Scan.	38
Tabela 4 – Configurações das arquiteturas de CNN utilizadas na abordagem.	39
Tabela 5 – Intervalos de parâmetros fornecidos ao algoritmo <i>Random Search</i> para cada classificador usado na abordagem de transferência aprendizado.	40
Tabela 6 – Métricas (%) obtidas por cada combinação de extrator e classificador na etapa de <i>transfer learning</i> desse estudo.	44
Tabela 7 – Tempos (ms) obtidos por cada combinação de extrator e classificador na etapa de <i>transfer learning</i> desse estudo.	45
Tabela 8 – Comparação entre as métricas obtidas pelo método proposto (com o conjunto de teste) em comparação com outros métodos da literatura que também fizeram classificação binária do dataset SARS-CoV-2 CT-scan.	49

LISTA DE ABREVIATURAS E SIGLAS

CLAHE	<i>Contrast Limited Adaptive Histogram Equalization</i>
CNN	<i>Convolutional Neural Network</i>
FPN	<i>Feature Pyramid Network</i>
Grad-CAM	<i>Gradient-weighted Class Activation Mapping</i>
GTs	<i>Ground Truths</i>
IA	Inteligência Artificial
IoT	<i>Internet of Things</i>
KNN	<i>k-Nearest Neighbors</i>
LSTM	<i>Long Short-Term Memory</i>
MLP	<i>Multilayer Perceptron</i>
OMS	Organização Mundial da Saúde
ReLU	<i>Rectified Linear Unit</i>
ROIs	<i>Regions of Interest</i>
RPN	<i>Region Proposal Network</i>
RT-PCR	<i>Reverse Transcription Polymerase Chain Reaction</i>
SVM	<i>Support Vector Machines</i>
TC	Tomografia Computadorizada
XAI	<i>Explainable Artificial Intelligence</i>
xDNN	<i>eXplainable Deep Learning</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Motivação	16
1.2	Estado da Arte	16
1.3	Objetivos	18
1.3.1	<i>Objetivo Geral</i>	18
1.3.2	<i>Objetivos Específicos</i>	18
1.4	Produção Científica	19
1.5	Organização do Texto	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Perceptron Simples e Perceptron Multicamadas	21
2.2	Rede Neural Convolucional	22
2.2.1	<i>Convolução</i>	23
2.2.2	<i>Pooling</i>	24
2.2.3	<i>Normalização de batch</i>	24
2.2.4	<i>Camadas totalmente conectadas</i>	24
2.2.5	<i>Dropout</i>	25
2.2.6	<i>Funções de ativação</i>	25
2.3	Segmentação Deep	26
2.4	Transferência de Aprendizado	28
2.5	Ajuste Fino	32
2.6	Métricas de Avaliação	32
2.7	Visualização com Grad-CAM	33
3	METODOLOGIA	36
3.1	<i>Datasets</i>	37
3.2	Transferência de Aprendizado com Modelos Deep	38
3.3	Segmentação Deep com Detectron2	40
3.4	Ajuste Fino do Melhor Modelo	41
4	RESULTADOS	43
4.1	Resultados dos Modelos com <i>Transfer Learning</i>	43
4.2	Resultados da Segmentação Pulmonar com Detectron2	45

4.3	<i>Fine Tuning</i> do Melhor Modelo	47
4.4	Análise Visual com Grad-CAM	50
5	CONCLUSÕES E TRABALHOS FUTUROS	54
	REFERÊNCIAS	56

1 INTRODUÇÃO

A doença respiratória aguda, conhecida como COVID-19 e causada pelo vírus Coronavírus tipo 2 (SARS-CoV-2), foi identificada pela primeira vez em dezembro de 2019 na cidade de Wuhan, China (HUANG *et al.*, 2020). Em questão de semanas, a COVID-19 se espalhou rapidamente por diversos países, resultando em uma pandemia global. O impacto na saúde pública, nas economias e no cotidiano das pessoas foi imenso, com sistemas de saúde de várias nações sobrecarregados, levando à escassez de recursos médicos, como ventiladores e leitos de UTI (CORONAVIRUS, 2020).

A disseminação do SARS-CoV-2 ocorre predominantemente por meio de gotículas respiratórias e contato com superfícies contaminadas, o que torna sua propagação rápida em ambientes densamente povoados (SOHRABI *et al.*, 2020). Além disso, a alta taxa de transmissibilidade, combinada com a possibilidade de transmissão por indivíduos assintomáticos, representou um enorme desafio para as autoridades de saúde e os sistemas hospitalares em todo o mundo (SOHRABI *et al.*, 2020).

A rápida disseminação do vírus levou a Organização Mundial da Saúde (OMS) a declarar oficialmente a COVID-19 uma pandemia em março de 2020 (CORONAVIRUS, 2020). Além disso, algumas medidas, como o distanciamento social, quarentenas, fechamento de fronteiras, e campanhas massivas de vacinação (a partir do final de 2020) foram reforçadas e incentivadas como medidas essenciais para conter a disseminação do vírus (ADEDOYIN; SOYKAN, 2023). Apesar desses esforços, o diagnóstico oportuno e preciso continua crucial até os dias de hoje, em decorrência de novas mutações que têm surgido, principalmente no contexto de sistemas de saúde sobrecarregados que demandam uma urgência de gerenciar os casos de forma eficaz.

Além das medidas preventivas, o diagnóstico precoce da COVID-19 tornou-se um dos pilares na luta contra a pandemia. O teste de evidência da reação em cadeia da polimerase de transcriptase reversa, do inglês *Reverse Transcription Polymerase Chain Reaction* (RT-PCR), foi amplamente adotado como o padrão ouro para a detecção do vírus, fornecendo uma confirmação da infecção (ALSHAZLY *et al.*, 2021a). Paralelamente, outros métodos de diagnóstico foram explorados para complementar a RT-PCR, particularmente em pacientes com sintomas respiratórios agudos. Dentre essas abordagens, as imagens de Tomografia Computadorizada (TC) de tórax ganharam destaque pela sua capacidade de revelar as anormalidades pulmonares típicas em pacientes com COVID-19, ou mesmo as sequelas deixadas pela doença nas regiões pulmonares,

(SHI *et al.*, 2020). Esse exame, apesar de ser relativamente rápido (por volta de 5 minutos), o diagnóstico depende de uma análise visual extensiva por parte do especialista, podendo demorar horas.

1.1 Motivação

Embora o teste RT-PCR seja considerado o *ground truth* para confirmar a infecção por COVID-19 (ALSHAZLY *et al.*, 2021a), a TC torácica também desempenha um papel vital na triagem e no diagnóstico da doença. No entanto, o processo de diagnóstico manual para radiologistas, com base na análise de imagens de TC, apresenta desafios relacionados à velocidade e precisão, especialmente em emergências que exigem respostas rápidas. Portanto, há uma necessidade de uma abordagem rápida e automática que ajude os profissionais de saúde a diagnosticar a condição com precisão. Métodos de inteligência artificial (Inteligência Artificial (IA)), aprendizado profundo e aprendizado de máquina tornam-se fundamentais para auxiliar no diagnóstico médico, fornecendo aos profissionais da área diagnósticos rápidos e precisos (XU *et al.*, 2022).

1.2 Estado da Arte

No contexto de imagens médicas para classificação de COVID-19, muitos autores usaram de métodos de *deep learning*, como esse trabalho, incluindo as técnicas de *transfer learning* e *fine-tuning*, usando do mesmo conjunto de dados ou similar.

Soares et al. (SOARES *et al.*, 2020) proporam o conjunto de dados SARS-CoV-2 CT-scan, que é o principal dataset usado nesse trabalho. O resultado de referência obtido pelos autores foi usando uma abordagem de *Explainable Deep Learning* (xDNN), alcançando um F1-score de 97,31% e acurácia de 97,38%.

Alshazly et al. (ALSHAZLY *et al.*, 2021a) utilizaram arquiteturas de redes neurais, em uma estratégia de *transfer learning* adaptando entradas personalizadas para cada arquitetura. Os autores usaram dois datasets: SARS-CoV-2 CT-scan (SOARES *et al.*, 2020) e COVID19-CT (HE *et al.*, 2020). Os resultados mostraram médias de 99,4%, 99,6%, 99,1%, 99,6% e 99,4% para precisão, sensibilidade, especificidade e F1-Score, respectivamente, para o dataset SARS-CoV-2. Além de identificar casos de COVID-19, foram técnicas de visualização a partir dos modelos gerados para entender a interpretabilidade dos modelos, destacando regiões distintas nas imagens

para casos de COVID e Non-COVID.

O estudo de Panwar et al. (PANWAR *et al.*, 2020) propõe um algoritmo de transferência de aprendizado profundo para acelerar a detecção de casos de COVID-19 usando imagens de raios-X e tomografia computadorizada. O modelo alcançou uma detecção de casos positivos de COVID-19 em ≤ 2 segundos, superando em velocidade e precisão os testes RT-PCR convencionais. A pesquisa utiliza três conjuntos de dados, entre eles o SARS-COV-2 CT-scan. A técnica de visualização de *Gradient-weighted Class Activation Mapping* (Grad-CAM) também foi utilizada para analisar a interpretabilidade dos modelos às imagens radiológicas.

Wang et al. (WANG *et al.*, 2020) propõe um aprimoramento da framework COVID-Net, com o objetivo de melhorar sua acurácia e eficiência de aprendizado. Os autores abordam a variação entre domínios com normalização de características e utilizamos um objetivo de treinamento contrastivo para fortalecer a invariância de domínio. Os testes feitos foram em parte com o dataset SARS-COV-2 CT-scan, atingindo métricas de quase 91% de acurácia e curva AUC 12.16% melhor que o COVID-Net original.

O trabalho de Jaiswal et al. (JAISWAL *et al.*, 2021) utilizou de modelos de deep learning para análise automática de imagens de tomografia computadorizada torácica na detecção de COVID-19. Os autores usaram uma arquitetura baseada na rede DenseNet201, treinada com pesos em um processo de deep transfer learning. Para o dataset SARS-COV-2 CT-scan, os autores atingiram valores próximos de 96% para acurácia, precisão, recall e F1-Score.

Naeem e Bin-Salem (NAEEM; BIN-SALEM, 2021) usaram uma abordagem de extração de características em vários níveis com redes neurais convolucionais para imagens de TC e raios-X torácicos. O objetivo dessa extração é reduzir a complexidade de treinamento da rede neural convolucional, facilitando a identificação precisa e robusta da COVID-19. Por fim, os autores usaram a *Long Short-Term Memory* (LSTM) concomitante com a rede para detectar as características extraídas do COVID-19. Os resultados experimentais mostram que a abordagem proposta obteve uma acurácia de 98,94% com o conjunto de dados SARS-CoV-2 TC scan.

Em contraste, o estudo de Zhao et al. (ZHAO *et al.*, 2021), apesar de não usar o dataset usado pelos demais autores, também usa de imagens de tomografia computadorizada para complementar os testes de RT-PCR, na classificação da COVID-19. A abordagem de transfer learning proposta neste estudo mostrou-se mais bem-sucedida do que outras abordagens descritas na literatura até o momento. Além disso, a pesquisa também explora as características visuais relevantes das imagens de TC com uso de mapas de calor do tipo Grad-CAM.

Dessa forma, as abordagens usadas para classificação de COVID-19 por meio de imagens de TC reforçam o uso de *deep learning* em um contexto de imagens médicas. Destacam-se as técnicas de *transfer learning* e *fine-tuning*, além dos algoritmos Grad-CAM para visualização da interpretabilidade dos modelos utilizados.

1.3 Objetivos

O método proposto usa modelos *deep* e é formado por uma etapa de segmentação pulmonar seguida de uma etapa de extração/classificação por um modelo de rede neural convolucional, do inglês *Convolutional Neural Network* (CNN), pré-treinado e ajustado para a COVID-19. Uma análise preliminar aprofundada acerca do desempenho dos modelos foi conduzida para escolher o melhor modelo entre aqueles avaliados para o método proposto. Além disso, métodos de Inteligência Artificial Explicada, do inglês *Explainable Artificial Intelligence* (XAI), do tipo Grad-CAM, foram usados para distinguir visualmente as regiões com foco mais significativo da rede neural ao realizar a classificação.

Portanto, motivado pela importância de uma abordagem automática, rápida e precisa para o diagnóstico de COVID-19 a partir de imagens de TC, bem como pelos trabalhos similares encontrados na literatura, este estudo apresenta os seguintes objetivos.

1.3.1 *Objetivo Geral*

Este trabalho tem como objetivo principal a geração de uma *pipeline* totalmente automática para classificar a doença COVID-19 em imagens de TC torácica.

1.3.2 *Objetivos Específicos*

Além do objetivo geral supracitado, esse estudo tem alguns objetivos específicos, que serão tratados como as contribuições desse trabalho.

1. A proposta de um modelo rápido, eficaz e totalmente automático que consiste em etapas de detecção/segmentação pulmonar em imagens de TC e a subsequente extração profunda e classificação entre as classes positivo ou negativo para COVID-19.
2. A geração de um modelo de classificação binária, para as classes COVID e Non-COVID usando imagens de TC de pulmão, que seja equivalente ou superior, em termos de métricas de avaliação, aos melhores métodos encontrados na literatura.

3. Este estudo também fornece uma análise preliminar dos melhores modelos profundos em um contexto de transferência de aprendizado. Portanto, o melhor modelo avaliado neste primeiro momento foi o escolhido para compor o método proposto.
4. Este trabalho também propõe uma análise visual da interpretabilidade de redes profundas usando Grad-CAM para explicar quais são as regiões críticas dos pulmões que mais contribuem para o diagnóstico de COVID-19 com base na rede pré-treinada.
5. Este método visa auxiliar o diagnóstico médico usando uma abordagem rápida, precisa e visual.

1.4 Produção Científica

A produção científica durante o período do mestrado resultou na publicação de artigos científicos em periódicos e congressos, além de um capítulo de livro. A seguir estão listadas as publicações realizadas durante esse período:

1. CHAVES, J. M.; OHATA, E. F.; **SANTOS, M. A.**; SANTOS, J. D. A.; BERNARDES, M.; DORA, D.; REBOUCAS FILHO, P. P. Predicting Energy Consumption Data Using Deep Learning: An LSTM Approach. In: **34th Brazilian Conference on Intelligent Systems (BRACIS)**, 2024, Belém, Brazil. BRACIS 2024.
2. SANTOS, LUCAS DE O.; SILVA, IÁGSON C. L.; **DOS SANTOS, MATHEUS A.**; MEDEIROS, ALDÍSIO G.; REBOUÇAS FILHO, PEDRO P. Efficient Classification of Depression using EEG through Spectral Graph Analysis. In: **2024 IEEE 37th International Symposium on ComputerBased Medical Systems (CBMS)**, 2024, Guadalajara. 2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS), 2024. p. 537.
3. SABOIA, C.; MARQUES, A. G.; SOUZA, L. F. F.; PEIXOTO, S. A.; **SANTOS, M. A.**; BARROS, A. C. S.; REGO, P. A. L.; REBOUÇAS FILHO, P. P. Fully Automatic LPR Method Using Haar Cascade for Real Mercosur License Plates. In: **Lecture Notes in Networks and Systems. (Org.). Intelligent Systems Design and Applications**. 716ed. Switzerland AG: Springer Nature, 2023, v. 3, p. 513-522.
4. NASCIMENTO, J. J. C.; MARQUES, A. G.; **DOS SANTOS, M. A.**; SANTOS, L. O.; CHAVES, J. M.; SOUZA, L. F. F.; REBOUÇAS FILHO, P. P. New Health of Things Approach to Classification and Detection of Brain Tumors Using Transfer Learning for Segmentation in IMR Images. In: **2023 International Joint Conference on Neural**

Networks (IJCNN), 2023, Gold Coast. 2023 International Joint Conference on Neural Networks (IJCNN), 2023. p. 1.

1.5 Organização do Texto

Esse trabalho de dissertação está estruturado da seguinte forma: esse Capítulo trata da Introdução, com a contextualização do problema, estado da arte e objetivos pretendidos; o Capítulo 2 aborda a fundamentação teórica a respeito dos métodos utilizados; enquanto o Capítulo 3 detalha a metodologia proposta, incluindo a *pipeline*, a arquitetura do modelo e as métricas de avaliação utilizadas; o Capítulo 4 apresenta os resultados, sua discussão e análise. Por fim, o Capítulo 5 conclui esta pesquisa com discussões sobre as descobertas, limitações e aborda sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção aborda a fundamentação teórica desta pesquisa, assim como os métodos utilizados neste estudo. Ela abrange desde o Perceptron simples e multicamadas, que originaram os demais conceitos; a composição de uma rede neural convolucional; a segmentação *deep*, por meio do *framework* Detectron2; as abordagens de *transfer learning* e *fine tuning* usadas nesse trabalho; além da visualização Grad-CAM, por mapa de calor. As topologias de redes neurais convolucionais utilizadas no primeiro e no segundo momento também foram citadas, assim como os algoritmos de classificação e as métricas de avaliação para cada modelo formado.

2.1 Perceptron Simples e Perceptron Multicamadas

O Perceptron Simples representa um dos primeiros modelos de redes neurais artificiais. Ele realiza a combinação linear das entradas ponderadas por pesos e é utilizado para resolver problemas linearmente separáveis (ROSENBLATT, 1958). A saída do Perceptron pode ser expressa matematicamente conforme a Equação 2.1.

$$\hat{y} = \phi \left(\sum_{i=1}^p x_i w_i + b \right) \quad (2.1)$$

onde x_i é o vetor de entrada e w_i é o vetor de pesos associado a cada entrada, enquanto ϕ é a função de ativação. A dimensão dos vetores de entrada deve ser a mesma e é dada por p . O termo b corresponde ao *bias*, ou viés, que deve ser ajustado junto aos pesos iterativamente durante o processo de treinamento.

O treinamento responsável por fazer o ajuste dos pesos e do *bias* é realizado conforme as Equações 2.2 e 2.3:

$$w_i(t+1) = w_i(t) + \eta e_i(t) x_i(t),; \forall i = 1, \dots, p \quad (2.2)$$

$$e_i(t) = y - \hat{y}(t) \quad (2.3)$$

em que t denota a iteração do treinamento; $e_i(t)$ é o erro daquela predição que é dado por y , que é a classe da amostra, e $\hat{y}(t)$, que representa a predição do modelo na iteração t ; e η é a taxa de aprendizagem que controla o impacto do erro sobre a atualização dos pesos.

Embora eficaz para problemas linearmente separáveis, o Perceptron Simples não consegue lidar com problemas não lineares. Para superar essa limitação, foi proposto o Perceptron Multicamadas, ou *Multilayer Perceptron* (MLP), que se originou no Perceptron Simples, mas que passa a organizar os neurônios em múltiplas camadas interconectadas (HAYKIN *et al.*, 2009). O MLP é capaz de resolver problemas não lineares ao utilizar funções de ativação não lineares, como a função sigmoide e a função tangente hiperbólica.

A saída de um neurônio do MLP em uma camada arbitrária l pode ser descrita pela Equação 2.4:

$$\hat{y}_k = \begin{cases} \phi_k \left(\sum_{i=1}^p x_i w_{ki} + b_k \right) & \text{para a camada de entrada} \\ \phi_k \left(\sum_{i=1}^D \hat{y}_i w_{ki} + b_k \right) & \text{para as camadas subsequentes} \end{cases} \quad (2.4)$$

onde \hat{y} é a predição da camada anterior (processo chamado de *feedforward*), k denota o neurônio, x_i é o vetor de entrada de dimensão p , e D é o número de neurônios na camada anterior.

O treinamento do MLP é realizado pelo algoritmo de retropropagação do erro, ou *Backpropagation*, que ajusta os pesos minimizando o erro de predição iterativamente (RUMELHART *et al.*, 1995). Essa atualização dos pesos no MLP é conduzida pela Equação 2.5:

$$\hat{w}_{ki}(t+1) = \begin{cases} w_{ki}(t) + \eta \cdot \delta_i(t) \cdot x_k(t) & \text{para a camada de entrada} \\ w_{ki}(t) + \eta \cdot \delta_i(t) \cdot \hat{y}_k(t) & \text{para as camadas subsequentes} \end{cases} \quad (2.5)$$

onde $\delta_i(t)$ é o gradiente local no neurônio k da respectiva camada durante a iteração t . Esse gradiente que deriva da função de ativação utilizada no neurônio, ou seja, podendo ser a derivada da função sigmóide ou da função tangente hiperbólica.

O Perceptron Simples e o MLP são a base para o aprendizado de máquina e para o aprendizado profundo. Nas redes neurais convolucionais, as camadas totalmente conectadas são basicamente Perceptrons Multicamadas, o que torna possível seu funcionamento. Além disso, o MLP também pode ser usado como um classificador, como será abordado adiante.

2.2 Rede Neural Convolucional

Essa seção tem por objetivo apresentar os principais componentes de uma CNN, desde o conceito de convolução, até etapas mais avançadas de seu uso, como a *batch normalization* e o *dropout*.

2.2.1 Convolução

Uma convolução é uma operação utilizando filtros para extrair informações de um sinal de entrada. Uma imagem, por ser um sinal de duas dimensões, pode ter diversas informações extraídas, a depender do filtro utilizado, como textura, contornos, contraste, etc (ALBAWI *et al.*, 2017).

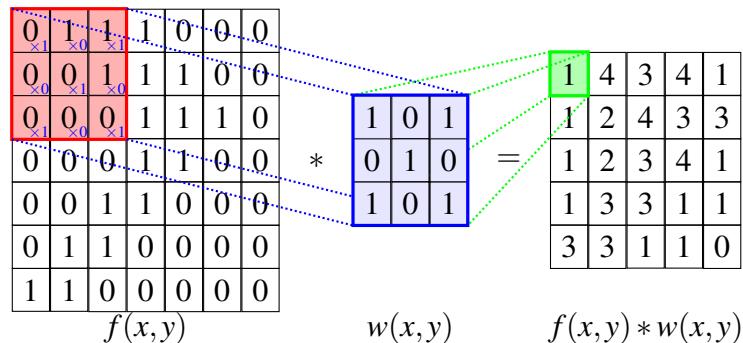
A Equação 2.6 representa a convolução de um filtro $w(x, y)$ em uma imagem $f(x, y)$, onde o filtro percorre toda a imagem, sobrepondo-a, de modo que todos os elementos do filtro percorram cada um dos pixels da imagem. Dessa forma, ocorre o somatório dos produtos de cada elemento sobreposto (GONZALEZ; WOODS, 2010).

$$w(x, y) * f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t) \quad (2.6)$$

onde a convolução de um filtro em uma imagem gera um mapa de características expressa por $f(x, y) * w(x, y)$, sendo x e y as coordenadas espaciais do pixel, enquanto $w(s, t)$ é o filtro convolucionado. Os parâmetros a e b determinam as dimensões do filtro, que se estende de $-a$ a a na direção horizontal e de $-b$ a b na direção vertical. As variáveis s e t representam os deslocamentos dentro da janela do filtro. O termo $f(x - s, y - t)$ refere-se ao valor do pixel deslocado pela posição do filtro durante a operação de convolução.

De forma análoga, a Figura 1 exemplifica como ocorre o processo de convolução de forma ilustrativa, para o primeiro elemento da matriz resultante, ou mapa de características, após somatório dos produtos dos elementos do filtro sobrepostos aos pixels da imagem.

Figura 1 – Ilustração explicativa de como acontece a operação de convolução.



Fonte: Elaborada pelo autor.

A operação de convolução é a base de um rede neural convolucional, pois através de repetidas convoluções, é possível extraír características da imagem de entrada em forma de

matrizes reduzidas. Além da convolução, outros conceitos utilizados em CNNs serão abordados a seguir.

2.2.2 *Pooling*

A etapa de *pooling* em uma CNN serve para reduzir a dimensionalidade de uma matriz de entrada (SCHERER *et al.*, 2010). Os tipos principais são o *max pooling* e o *average pooling*.

Semelhante ao que ocorre na convolução, um filtro de dimensões pré-definidas é sobreposto na matriz de entrada, fazendo com os pixels daquela região se tornem na matriz de saída um único pixel, que será o de maior valor no caso do *max pooling* ou a média de todos os pixels da região no caso do *average pooling* (SCHERER *et al.*, 2010).

2.2.3 *Normalização de batch*

A normalização de lote, ou normalização de *batch*, é uma técnica usada para otimizar o treinamento de redes neurais profundas (IOFFE, 2015). O método consiste em normalizar as entradas de cada camada com base nos lotes de dados processados pela rede.

A Equação 2.7 ilustra a normalização, também chamada de *z-score*, que é feita a partir do cálculo da média e do desvio padrão de cada lote.

$$X' = \frac{X - \mu}{\sigma} \quad (2.7)$$

onde X é o lote de entrada, enquanto μ e σ são a média e o desvio padrão do lote, respectivamente.

Esse método tanto acelera o processo de treinamento, ou refinamento, necessitando de menos épocas de treino, como também estabiliza esse processo, atualizando os pesos ao fim de cada lote.

2.2.4 *Camadas totalmente conectadas*

Também chamadas de camadas densas, as camadas totalmente conectadas de uma rede neural convolucional fazem parte de sua etapa de classificação. Baseadas no MLP, essas camadas são um conjunto de neurônios que fazem a propagação das informações oriúndas das camadas de convolução e *pooling* (LECUN *et al.*, 1989).

Essa camada, que fica no topo (ou na saída) da rede, é responsável pela predição do vetor de características que chega até ela, resultando na classe predita para aquela amostra.

2.2.5 Dropout

A técnica de *dropout* consiste em regularizar uma rede neural, desativando uma certa quantidade de neurônios durante sua execução, removendo-os do fluxo de alimentação direta (*feedforward*) e da retropropagação do erro (*backpropagation*) (SRIVASTAVA *et al.*, 2014).

Essa técnica diminui os casos de *overfitting* do modelo, que consiste em uma especialização da rede apenas nos dados usados no treinamento, e portanto, comprometendo sua generalização (SRIVASTAVA *et al.*, 2014).

2.2.6 Funções de ativação

Por fim, como foi visto na Seção 2.1, as funções de ativação fazem um importante papel em uma rede neural, de modo a tornar a saída não linear e permitindo seu uso em problemas complexos.

As funções de ativação mencionadas anteriormente foram a sigmóide e a tangente hiperbólica, também chamada de *tanh*, que podem ser expressas pelas Equações 2.8 e 2.9 (HAYKIN, 2001). Além das curvas serem suavizadas de forma diferente, fazendo com que a *tanh* tenha uma derivada maior próximo a sua origem, elas também costumam ter intervalos distintos, sendo o intervalo da sigmóide de $[0, 1]$ enquanto o intervalo da *tanh* é de $[-1, 1]$.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.8)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.9)$$

Outras funções de ativação bastante usadas em um contexto de redes neurais são a ReLU, a *Leaky ReLU* e a *Softmax*.

A função de ativação *Rectified Linear Unit* (ReLU) é uma função não linear e uma das mais usadas no contexto de redes neurais (AGARAP, 2018). Ela é representada pela Equação 2.10 e consiste em atribuir zero a qualquer entrada negativa, o que faz com que o neurônio não seja ativado.

$$f(x) = \max(0, x) \quad (2.10)$$

Por outro lado, a *Leaky ReLU* surgiu devido ao problema da derivada da ReLU, que nas proximidades de zero, acaba por fazer com que o gradiente dê um pico (XU, 2015). A *Leaky ReLU*, portanto, não tem saída zero para valores negativos, mas sim é proporcional a sua entrada, mas com menor inclinação que para valores positivos, como mostra a Equação 2.11.

$$f(x) = \begin{cases} x & \text{se } x \geq 0 \\ \alpha x & \text{se } x < 0 \end{cases} \quad (2.11)$$

Por fim, a função de ativação *softmax* é muito usada nas camadas densas das redes neurais convolucionais e é responsável por transformar a predição do neurônio em probabilidades de cada classe, de modo que a soma dos valores sejam sempre 1 (BRIDLE, 1989). Essa função é descrita pela Equação 2.12.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (2.12)$$

onde n é o número de classes para aquela amostra e z_i é a saída do i -ésimo neurônio da última camada densa.

2.3 Segmentação Deep

Seguindo o mesmo contexto da classificação com redes neurais convolucionais, a segmentação também pode ser feita usando *deep learning*. Para a segmentação, a saída da rede não será a classe da amostra, mas sim a máscara que abrange a região segmentada da imagem.

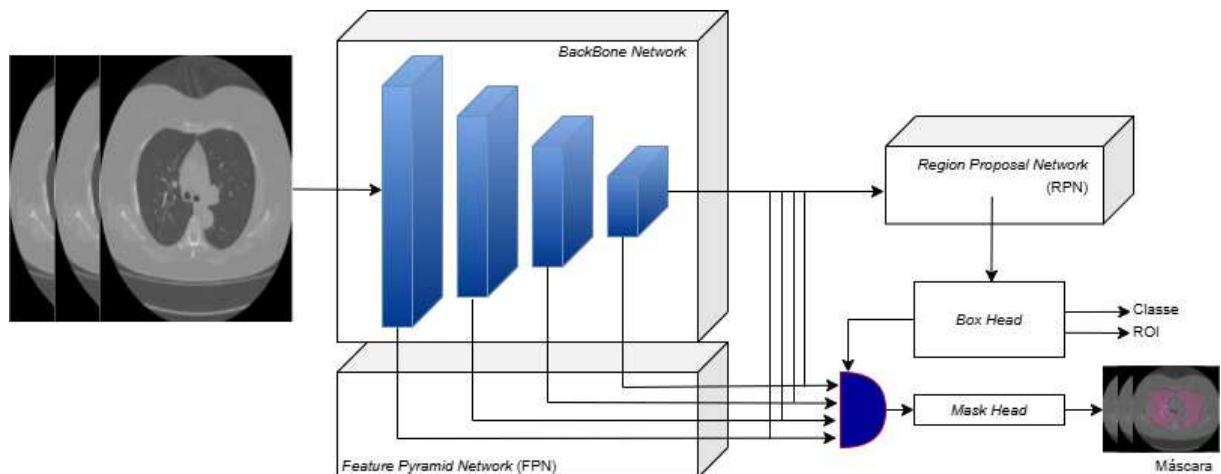
Dessa forma, o framework Detectron2 (WU *et al.*, 2019), que foi utilizado nesse trabalho como um segmentador profundo, tem sua origem na Mask R-CNN (HE *et al.*, 2017), e é a sucessora da Detectron (GIRSHICK *et al.*,). Proposta pelo Facebook Artificial Intelligence Research (FAIR), a Detectron2 é usada como uma rede neural profunda para realizar detecção e/ou segmentação de objetos.

A Detectron2 usa do método de anotações para o conjunto de dados adotado pelo COCO (LIN *et al.*, 2014), o que consiste dizer que as anotações dos objetos são compostas

por uma lista de vértices dos polígonos que compõem um objeto de interesse. A Detectron2 também recebe como entrada os *Ground Truths* (GTs) das imagens, classificando um aprendizado supervisionado para detecção e segmentação dos objetos de interesse (no caso desse estudo, são os pulmões).

Sua arquitetura é baseada em diferentes módulos distintos, compondo o modelo final de detecção e segmentação, como pode ser visto na Figura 2. O primeiro módulo crucial é conhecido como *BackBone Network* utiliza um modelo ResNet (HE *et al.*, 2016) para convoluir e gerar características hierárquicas relevantes das imagens. A ResNet é eficiente devido à sua capacidade de aprender representações profundas sem o problema do desaparecimento de gradientes, utilizando conexões residuais.

Figura 2 – Representação da arquitetura do *Framework Detectron2* para segmentação de imagens médicas.



Fonte: Elaborada pelo autor.

Seguindo a *BackBone*, a *Region Proposal Network* (RPN) é responsável por classificar regiões candidatas que possam conter os objetos de interesse. A RPN é uma rede totalmente convolucional que, a partir das características extraídas pela *BackBone*, gera *Regions of Interest* (ROIs) que são classificadas como contendo ou não um objeto de interesse.

Em seguida, a etapa de *Feature Pyramid Network* (FPN) garante tratar as informações extraídas em diferentes escalas espaciais, sendo robusta para imagens de diferentes dimensões e tamanhos. A FPN melhora a extração de características em múltiplas resoluções e é robusta, permitindo que a rede seja eficaz em imagens com diferentes dimensões. Isso é particularmente importante em imagens médicas, onde o tamanho e a proporção dos objetos podem variar significativamente.

Por fim, a Detectron2 utiliza módulos de pós-processamento com modelos regressores para refinar as previsões, incluindo a remoção de caixas delimitadoras redundantes e a aplicação de limiares para manter apenas as detecções mais confiáveis.

Dessa forma, o framework Detectron2, de arquitetura modular, é robusta e bem projetada, sendo uma ferramenta poderosa para aplicações de visão computacional, especialmente em tarefas complexas como detecção e segmentação de objetos em imagens médicas. Além disso, a sua utilização permite que o método proposto seja invariante a escala, podendo ser utilizado para diferentes máquinas de TC de forma funcional, sem necessitar de ajustes.

2.4 Transferência de Aprendizado

O primeiro método abordado nesse estudo é o de *transfer learning*, que consiste em utilizar uma rede neural convolucional, pré-treinada com um extenso banco de dados, para extrair características de forma e textura de outro conjunto de dados, e classificar esses vetores de características utilizando classificadores tradicionais (ZHU *et al.*, 2019). Em outras palavras, para ser utilizada exclusivamente como extrator de características, são removidas as camadas de classificação da CNN, chamadas de camadas totalmente conectadas originais da rede (*fully-connected layers*) (PAUL *et al.*, 2016; ORENSTEIN; BEIJBOM, 2017).

Neste trabalho, foram utilizadas topologias de CNN com pesos inicialmente treinados com o grande banco de imagens *ImageNet* (DENG *et al.*, 2009), que consiste em milhões de imagens de objetos do cotidiano em 1000 classes diferentes. As camadas totalmente conectadas foram substituídas por algoritmos de machine learning.

As arquiteturas da CNN utilizadas neste artigo foram três: a arquitetura VGG (SIMONYAN; ZISSERMAN, 2014), a arquitetura DenseNet (HUANG *et al.*, 2017) e a arquitetura MobileNet (HOWARD *et al.*, 2017). A primeira foi implementada em duas configurações diferentes (VGG16 e VGG19), enquanto a segunda foi implementada em três (DenseNet121, DenseNet169 e DenseNet201). Por fim, a MobileNet foi implementada em sua arquitetura convencional.

A Tabela 1 reúne, de forma resumida, as arquiteturas de CNN usadas nesse trabalho, assim como as configurações utilizadas e a descrição de cada arquitetura, mostrando o que as diferenciam entre si.

As configurações VGG16 e VGG19 diferem pelo número de pesos, a primeira com 16 camadas e a segunda com 19 camadas. Essa arquitetura é diferenciada por usar pequenos

Tabela 1 – Descrição das arquiteturas de CNN e suas configurações utilizadas na abordagem.

Arquitetura	Configurações	Descrição
VGG (SIMONYAN; ZISSERMAN, 2014)	VGG16 VGG19	Arquitetura que utiliza pequenos filtros convolucionais para aumentar a profundidade. As configurações diferem no número de camadas: 16 na VGG16 e 19 na VGG19.
MobileNet (HOWARD <i>et al.</i> , 2017)	MobileNet	Arquitetura leve e eficiente, projetada para ser utilizada em dispositivos embarcados, com rápido treinamento e bom desempenho em ambientes de recursos limitados.
DenseNet (HUANG <i>et al.</i> , 2017)	DenseNet121 DenseNet169 DenseNet201	Arquitetura composta por <i>DenseBlocks</i> , camadas convolucionais e de transição. As diferentes configurações variam no número total de camadas, aumentando progressivamente a profundidade e capacidade de aprendizado.

Fonte: Elaborada pelo autor.

filtros convolucionais, o que aumenta seu poder de profundidade (SIMONYAN; ZISSERMAN, 2014). Por outro lado, as diferentes configurações de DenseNet são diferenciadas pelo diferente número de camadas que as compõem, sendo formadas por camadas convolucionais, de transição e pelos chamados *DenseBlocks* (HUANG *et al.*, 2017). A MobileNet se caracteriza por sua leveza, rápido treinamento e facilidade de ser acoplada em ambientes embarcados (HOWARD *et al.*, 2017).

Após a etapa de extração de características, os vetores de atributos extraídos pelas topologias apresentadas na seção anterior são fornecidos a sete diferentes classificadores. São eles: Classificador de Naive Bayes (THEODORIDIS; KOUTROUMBAS, 2008), MLP (HAYKIN *et al.*, 2009), *k-Nearest Neighbors* (KNN) (AHA *et al.*, 1991), *Random Forest* (HO, 1998) e três versões diferentes de *Support Vector Machines* (SVM) (Linear, Polinomial e RBF) (VAPNIK, 1998). A Tabela 2 resume cada um dos classificadores utilizados e suas particularidades.

O primeiro classificador abordado, *Naive Bayes*, é um algoritmo de classificação baseado em uma análise estatística dos dados de entrada, considerando a independência entre as características dos atributos (THEODORIDIS; KOUTROUMBAS, 2008). Ele se fundamenta na aplicação da Regra de Bayes, que descreve a probabilidade de uma amostra pertencer a uma determinada classe, dado um conjunto de características observadas. A técnica utiliza a teoria da probabilidade condicional para calcular essas probabilidades, assumindo que as características são independentes entre si, o que simplifica os cálculos e torna o algoritmo computacionalmente eficiente (THEODORIDIS; KOUTROUMBAS, 2008). Para cada classe, o algoritmo calcula uma probabilidade baseada na função de densidade de probabilidade dos atributos e, ao final, rotula a amostra com a classe que apresenta a maior probabilidade de ocorrência (THEODORIDIS;

KOUTROUMBAS, 2008).

O classificador MLP, cuja matemática foi explicada na Seção 2.1, é uma rede neural composta por múltiplas camadas de neurônios artificiais, conhecidos como *perceptrons* simples, propostos por Rosenblatt (ROSENBLATT, 1958). Ele é projetado para realizar tarefas de aprendizado supervisionado, como a classificação de dados, e é composto por três componentes principais: a camada de entrada, que recebe os atributos da amostra; as camadas ocultas, que processam a informação de forma não linear; e a camada de saída, que faz a predição da classe da amostra (HAYKIN *et al.*, 2009). Cada *perceptron* nas camadas ocultas está conectado a outros *perceptrons* das camadas adjacentes por meio de conexões com pesos ajustáveis, que são atualizados durante o processo de treinamento. O aprendizado no MLP é realizado por meio de um algoritmo de retropropagação do erro, que permite otimizar os pesos das conexões de forma a minimizar o erro entre a saída predita e a real (HAYKIN *et al.*, 2009). O MLP é a origem das CNN's mais complexas e uma das arquiteturas mais utilizadas em redes neurais profundas, sendo capaz de modelar padrões complexos de dados.

O KNN, ou K-Vizinhos Mais Próximos, é um algoritmo de aprendizado supervisionado simples e eficaz que realiza a classificação de uma nova amostra com base na similaridade com suas k amostras mais próximas, que já estão rotuladas (AHA *et al.*, 1991). O principal conceito do KNN é que amostras semelhantes em distribuição espacial tendem a pertencer à mesma classe. O valor de k é um parâmetro importante que influencia diretamente a acurácia do classificador; valores muito pequenos podem tornar o modelo sensível a ruídos, enquanto valores maiores podem suavizar a fronteira entre as classes (AHA *et al.*, 1991). O KNN é amplamente utilizado devido à sua simplicidade e boa performance em muitos cenários, especialmente quando os dados têm uma distribuição espacial bem definida.

O algoritmo *Random Forest*, ou Floresta Aleatória, é uma técnica de aprendizado supervisionado que utiliza uma coleção de árvores de decisão para realizar a classificação, sendo cada árvore na floresta construída a partir de um subconjunto aleatório dos dados de treinamento e, para cada divisão em uma árvore, é considerado um subconjunto aleatório de atributos (HO, 1998). Isso resulta em uma diversidade de modelos fracos, que, quando combinados, geram um modelo final robusto e preciso. O método utiliza o conceito de *bagging* (ou agregação de amostras), no qual várias versões do modelo são treinadas com diferentes subconjuntos de dados, e a previsão final é baseada na média ou na votação da maioria das árvores. Essa abordagem reduz a variabilidade do modelo, tornando-o mais resistente a problemas como o *overfitting* (HO,

1998).

As Máquinas de Vetores de Suporte, ou SVM, por sua vez, são técnicas de classificação baseadas na análise de margens de separação entre as classes. O objetivo principal do SVM é encontrar o *hiperplano* que maximiza a distância entre as amostras de diferentes classes, chamado de margem de separação ótima, o que garante uma boa generalização do modelo (VAPNIK, 1998). Em casos onde as classes não são linearmente separáveis no espaço original, o SVM pode ser estendido por meio do uso de diferentes *kernels*, que são diferentes funções matemáticas que mapeiam os dados para um espaço de maior dimensão, onde a separação entre as classes se torna mais fácil (VAPNIK, 1998). Os tipos mais comuns de *kernels* incluem o Linear, Polinomial e RBF (Função de Base Radial). A escolha do *kernel* adequado e de seus parâmetros é crucial para o desempenho do modelo. As SVMs são altamente eficientes em cenários de alta dimensionalidade e quando há poucos dados disponíveis (VAPNIK, 1998).

Tabela 2 – Descrição dos classificadores utilizados na abordagem.

Classificador	Descrição
Naive Bayes (THEODORIDIS; KOUTROUMBAS, 2008)	Algoritmo baseado em probabilidade que utiliza a Regra de Bayes para classificar amostras, assumindo independência entre os atributos. Calcula a probabilidade de uma amostra pertencer a uma classe específica e escolhe a classe com maior probabilidade. É eficiente e simples, mas a suposição de independência nem sempre reflete a realidade.
MLP (HAYKIN <i>et al.</i> , 2009)	Rede neural composta por uma camada de entrada, múltiplas camadas ocultas e uma camada de saída. Realiza aprendizado supervisionado por meio de retropropagação do erro. Capaz de modelar padrões complexos, sendo a base para redes neurais profundas.
KNN (AHA <i>et al.</i> , 1991)	Classificador baseado em similaridade que atribui a uma nova amostra a classe mais comum entre suas k vizinhas mais próximas. O valor k influencia o desempenho do modelo. É simples, mas pode ser sensível a ruídos e à escolha de k .
Random Forest (HO, 1998)	Método de aprendizado baseado em múltiplas árvores de decisão, construídas com subconjuntos aleatórios de dados e atributos. Utiliza <i>bagging</i> para reduzir a variância e melhorar a robustez do modelo, sendo eficiente contra <i>overfitting</i> .
SVM (VAPNIK, 1998)	Classificador que busca encontrar o hiperplano de separação ótima entre classes, maximizando a margem. Para casos não linearmente separáveis, utiliza <i>kernels</i> como Linear, Polinomial e RBF, que mapeiam os dados para espaços de maior dimensão. É eficaz em alta dimensionalidade.

Fonte: Elaborada pelo autor.

2.5 Ajuste Fino

O ajuste fino, ou *fine-tuning*, de modelos de *deep learning* é uma prática crucial para otimizar o desempenho de uma rede neural pré-treinada em uma tarefa específica. Inicialmente, modelos pré-treinados em conjuntos de dados massivos, como *ImageNet* (DENG *et al.*, 2009), capturam padrões em uma ampla variedade de imagens. No entanto, ao aplicar esses modelos a tarefas mais especializadas, como é o caso de imagens médicas, é necessário ajustar os pesos da rede para se alinhar com os padrões específicos do novo conjunto de dados (GUAN; LIU, 2021; TAJBAKHSH *et al.*, 2016).

A técnica de *fine-tuning* geralmente envolve dois estágios: o congelamento e o descongelamento de camadas. No primeiro estágio, as camadas convolucionais iniciais, responsáveis por características gerais, são mantidas fixas, enquanto apenas as camadas finais de classificação são treinadas para a nova tarefa (TAJBAKHSH *et al.*, 2016). Isso permite que a rede se adapte às características específicas do conjunto de dados de destino sem perder os conhecimentos prévios. No segundo estágio, camadas adicionais são gradualmente descongeladas, permitindo ajustes mais profundos nos pesos das camadas convolucionais (TAJBAKHSH *et al.*, 2016).

Para o presente trabalho, o fine-tuning foi usado tanto na segmentação, ao adaptar os pesos da Detectron2 para segmentação de pulmões, como na classificação do segundo momento, onde o melhor modelo de transfer learning foi escolhido para ser refinado para o problema em questão. Em ambos os casos, o fine-tuning gera pesos específicos para a topologia utilizada e para o dataset trabalhado.

2.6 Métricas de Avaliação

As métricas utilizadas para avaliar o desempenho dos modelos de transfer learning e fine tuning foram: acurácia (Acc), *precision* ou precisão, *recall* ou sensibilidade, *F1-Score* (F1) e Coeficiente de Correlação de Matthews (*Matthews Correlation Coefficient – MCC*) (OHATA *et al.*, 2021).

A acurácia é uma métrica que mede o número de amostras corretamente classificadas pelo número total de amostras. O *F1-Score*, por outro lado, utiliza métricas intermediárias, *precision* e *recall*, para medir o equilíbrio entre elas através de sua média harmônica. Assim, um índice *F1-Score* alto é considerado um fator de uniformidade na classificação da doença, pois apresenta um baixo número de falsos positivos e um baixo número de falsos negativos (OHATA

et al., 2021).

Por fim, o MCC é uma medida de correlação amplamente utilizada em problemas de classificação binária. Com base nos índices de correlação de Pearson, esse coeficiente pode ser tratado como uma medida balanceada mesmo quando há um conjunto de dados desbalanceado.

Todas as equações usam os verdadeiros positivos (VP), falsos negativos (FN), verdadeiros negativos (VN) e falsos positivos (FP), todos presentes na matriz de confusão para calcular as métricas de avaliação. Uma matriz de confusão é uma ferramenta que compara a classe real de cada amostra classificada e a classe prevista pelo método. As equações de cada uma das métricas usadas nesse trabalho podem ser vistas nas Equações de 2.13 à 2.17.

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.13)$$

$$Precision = \frac{VP}{VP + FP} \quad (2.14)$$

$$Recall = \frac{VP}{VP + FN} \quad (2.15)$$

$$F1 = \frac{2VP}{2VP + FP + FN} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.16)$$

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (2.17)$$

Existe um resultado métrico para cada classe em questão, por isso foram feitas médias aritméticas simples dos resultados da classe para representação em tabelas, com o objetivo de facilitar a análise dos resultados. Os valores serão representados em porcentagem.

2.7 Visualização com Grad-CAM

Como forma de analisar a interpretabilidade dos modelos desenvolvidos nesse estudo, utilizou-se mapas visuais Grad-CAM, através da biblioteca TorchCAM (FERNANDEZ, 2020). Esses mapas, construídos ao analisar os modelos com as imagens de TC de pacientes com

COVID-19, destacam as áreas de maior enfoque por parte das redes, sendo extremamente útil para o diagnóstico (ALSHAZLY *et al.*, 2021b). Eles ilustram o que poderão ser regiões críticas nas varreduras de TC, auxiliando na identificação da presença do SARS-CoV-2 e na visualização de tecido pulmonar saudável (ALSHAZLY *et al.*, 2021b).

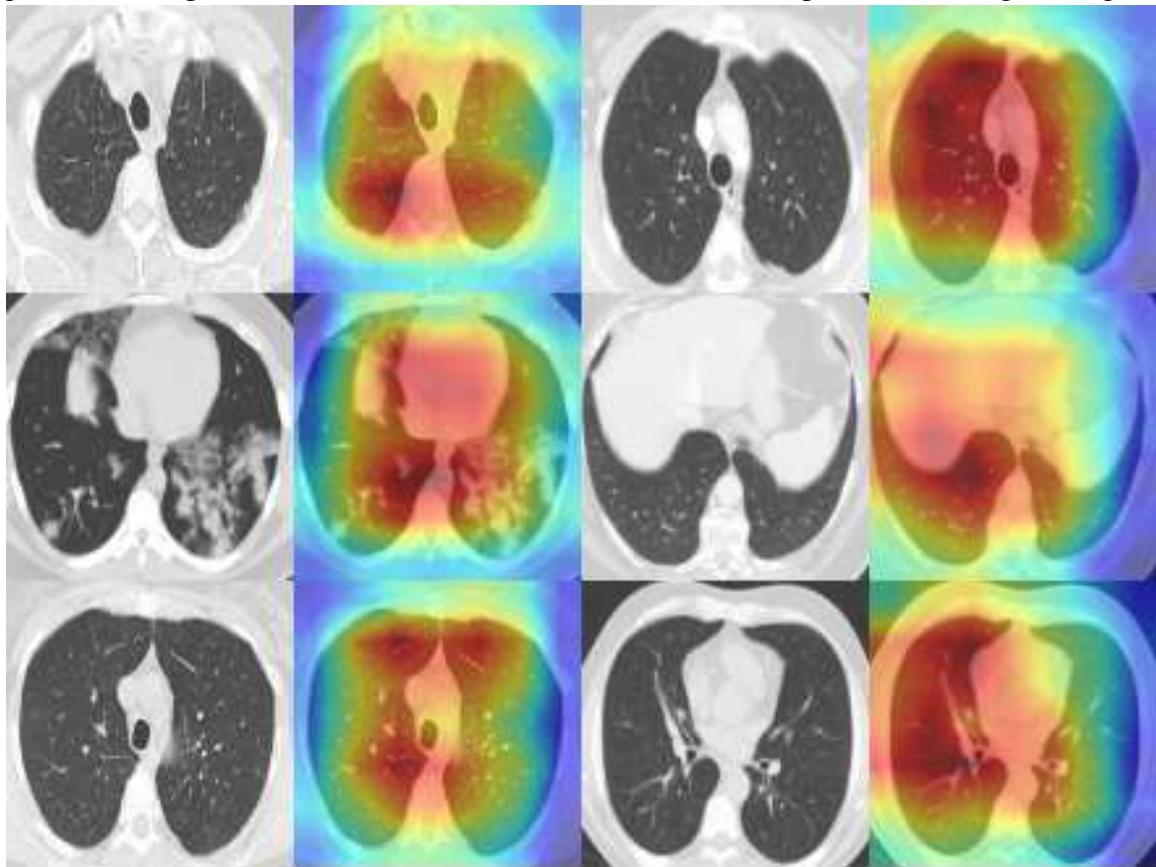
O método funciona retropropagando o gradiente da classe-alvo em relação às ativações das últimas camadas convolucionais até as camadas de entrada da rede neural. Isso resulta em um mapa de ativação ponderado que destaca as áreas da imagem que contribuíram significativamente para a decisão do modelo (SELVARAJU *et al.*, 2017). A vantagem de utilizar a última camada convolucional é que ela retém características espaciais de alto nível, preservando informações sem perder a localização dos objetos ou estruturas na imagem. Dessa forma, essa técnica pode ser utilizada combinada com qualquer CNN, já que faz uso das camadas convolucionais e das camadas de entrada.

Além disso, o Grad-CAM é uma técnica versátil, que pode ser aplicado independente da arquitetura da CNN, além de não ser necessário a modificação da estrutura do modelo ou realizar novos treinamentos. Isso o torna uma abordagem prática e não invasiva para analisar a interpretação de modelos complexos, o que é especialmente útil em ambientes clínicos e de auxílio ao diagnóstico médico, onde o entendimento das decisões do modelo é tão importante quanto sua precisão, pois o mapa demonstra as regiões de maior probabilidade da doença, podendo ser validado por especialistas da área (PANWAR *et al.*, 2020).

No contexto do SARS-CoV-2, o Grad-CAM auxilia na detecção de anomalias em exames de TC, permitindo visualizar, por exemplo, áreas de inflamação. Essa capacidade de explicabilidade contribui não apenas para o diagnóstico, mas também para o acompanhamento da progressão da doença, ajudando a monitorar a resposta do paciente ao tratamento (ALSHAZLY *et al.*, 2021b).

A Figura 3 ilustra o método de Grad-CAM, com visualização de mapa de calor, aplicado em exames de imagem de TC de pulmão. Na imagem, cores mais frias representam pixels que pouco contribuem na descrição feita pela rede, enquanto as cores mais quentes representam pontos de maior peso na predição feita pelo modelo.

Figura 3 – Exemplos de Grad-CAM em exames de TC torácico aplicados às imagens originais.



Fonte: Elaborada pelo autor.

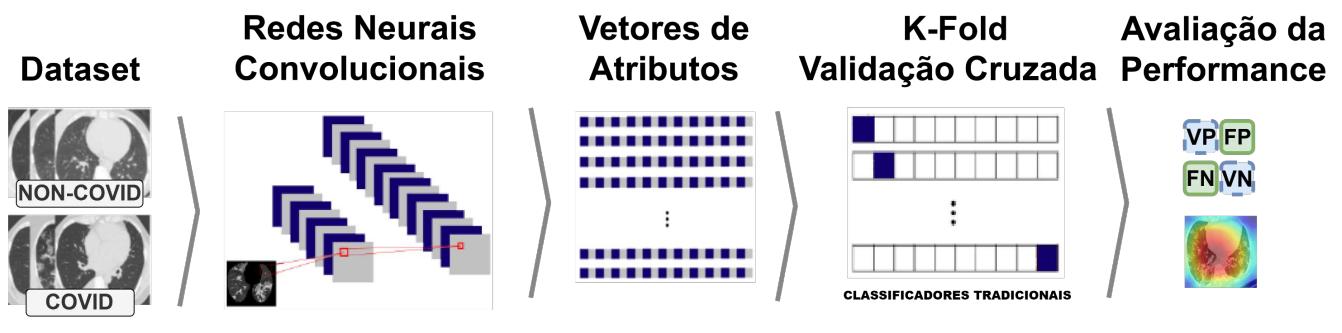
A integração de técnicas de XAI, como Grad-CAM, em sistemas de apoio ao diagnóstico médico melhora a transparência dos modelos de aprendizado profundo, criando uma ponte entre as previsões e a validação humana. Dessa forma, especialistas da área podem revisar e interpretar os resultados de maneira colaborativa, validando o desempenho da rede neural utilizada para a tarefa designada.

3 METODOLOGIA

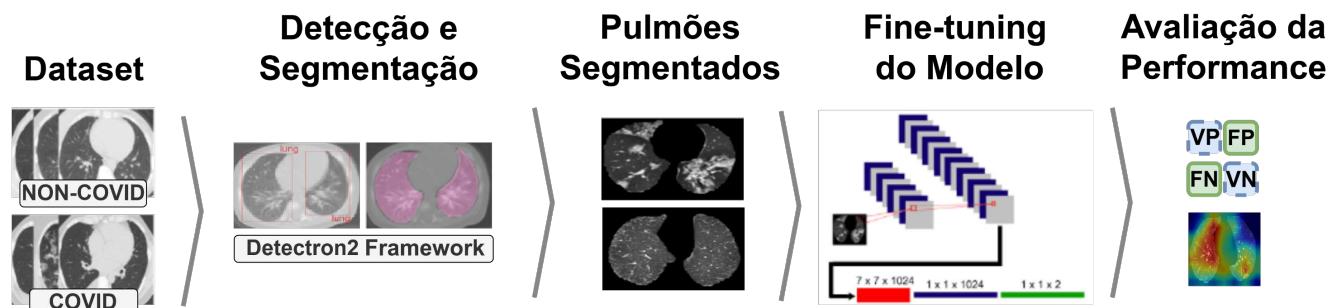
O estudo proposto se baseia em dois momentos principais: (i) a abordagem de transfer learning, a fim de comparar os desempenhos das melhores combinações entre extratores deep e classificadores tradicionais com uso do dataset original e (ii) a abordagem totalmente automática com uso do melhor modelo encontrado na etapa anterior (refinado com fine-tuning) para as imagens segmentadas pelo framework Detectron2 (também refinado para detecção e segmentação de pulmões em imagens de TC). A Figura 4 (acima) ilustra o experimento de *transfer learning*, enquanto a Figura 4 (abaixo) descreve o método proposto, incorporando etapas de detecção/segmentação e classificação. Ambos os experimentos passaram por avaliações numéricas e visuais usando métricas e mapas de calor do Grad-CAM, respectivamente.

Figura 4 – Acima: Processo de *transfer learning*, no qual o melhor modelo avaliado foi selecionado para a metodologia proposta. Abaixo: Fluxograma da metodologia proposta, uma abordagem automática para classificação de COVID em imagens de TC de pulmão, integrando segmentação, *fine-tuning* do melhor modelo e validando sua interpretabilidade com Grad-CAM.

METODOLOGIA DE TRANSFER LEARNING



METODOLOGIA PROPOSTA



Fonte: Elaborada pelo autor.

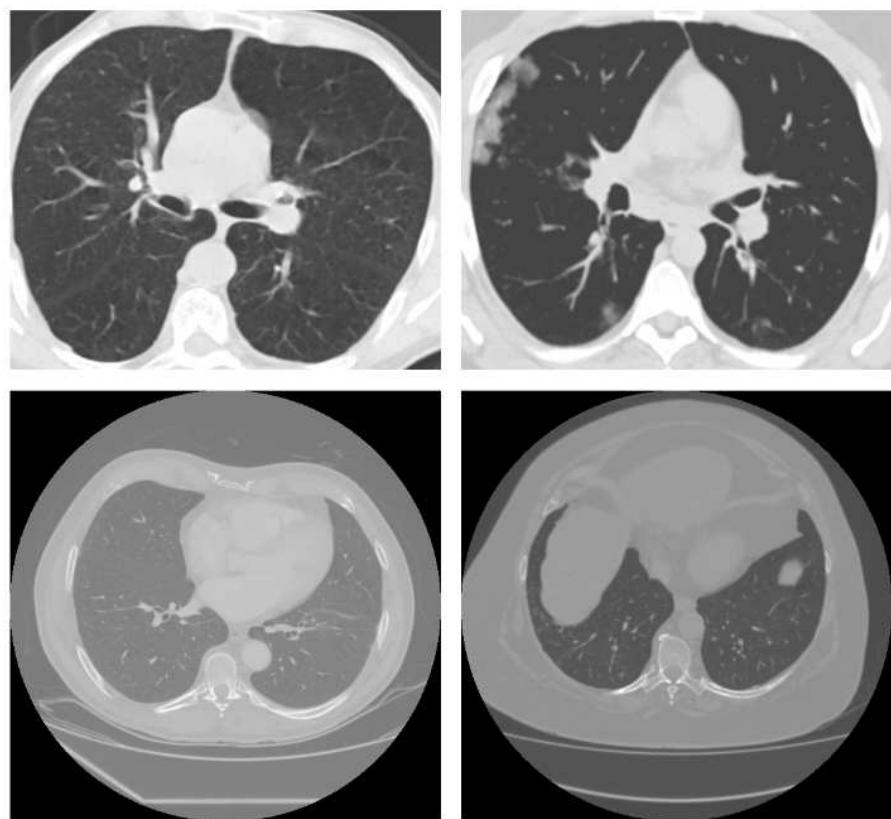
Os experimentos foram conduzidos no sistema Ubuntu 22.04, equipado com 32 GB de RAM, um processador Intel® Core™ i7-12700H e uma GPU Nvidia GeForce RTX 3050

com 6 GB de memória dedicada. Para a construção dos modelos VGG, DenseNet e MobileNet, foram empregadas as bibliotecas Python (v3.7), OpenCV 4.1.0, TensorFlow-GPU 1.14 e Keras 2.2.4. A Detectron2 foi modificada a treinada e validada usando de seu repositório oficial. Os classificadores tradicionais e as métricas de avaliação foram aplicados utilizando a biblioteca sci-kit-learn v0.20.2.

3.1 *Datasets*

Nesse estudo foram usados dois diferentes datasets compostos de imagens de TC torácico, com foco nos pulmões. Algumas amostras de imagens de ambos os datasets podem ser vistas na Figura 5.

Figura 5 – Exemplos de imagens dos datasets utilizados neste estudo. Acima, algumas amostras de imagens do SARS-CoV-2 CT-Scan (Classe Non-Covid à esquerda e Classe Covid à direita); abaixo, imagens do dataset adquirido junto ao Hospital Walter Cantídio, para segmentação de pulmão.



Fonte: Elaborada pelo autor.

O dataset usado para classificação de COVID em imagens de TC é o SARS-CoV-2 CT-Scan (SOARES *et al.*, 2023). As imagens do conjunto de dados são de tomografia

computadorizada de pulmão e estão em formato PNG com três canais RGB e 8 bits cada. As imagens não são padronizadas em suas dimensões, tendo tamanhos variados. O dataset é composto por 1252 tomografias computadorizadas de pacientes que testaram positivo para a infecção por SARS-CoV-2 (COVID-19) e 1230 TC's de pacientes não infectados pelo SARS-CoV-2, totalizando 2482 TC's.

A Tabela 3 divide o dataset de forma detalhado entre suas classes para número de pacientes, sua divisão por sexo e a quantidade de tomografias computadorizadas.

Tabela 3 – Resumo da composição do conjunto de dados SARS-CoV-2 CT-Scan.

Categoría	Número de Pacientes	Número de TC's	Sexo (Masculino/Feminino)
COVID	60	1252	32 / 28
Non-COVID	60	1230	30 / 30

Fonte: Elaborada pelo autor.

As imagens foram adquiridas nos Hospitais Públicos dos Servidores do Estado de São Paulo e do Hospital Metropolitano da Lapa, ambos em São Paulo, Brasil, ao longo do ano de 2020, marcado pela pandemia de COVID-19 (SOARES *et al.*, 2023). Ressalta-se que os dados dos pacientes foram cuidadosamente omitidos das imagens finais.

O outro dataset utilizado também foi um dataset de TC de pulmão, com o propósito de segmentação dos pulmões. Esse conjunto de dados é composto de 1265 imagens de TC de pulmão, no formato Digital Imaging and Communications in Medicine (DICOM), com dimensões de 512 x 512 pixels e profundidade de 16 bits. Cada uma das imagens possui padrão-ouro de segmentação, feito por um especialista. As imagens foram adquiridas junto ao Hospital Walter Cantídio, da Universidade Federal do Ceará (UFC), Fortaleza, Ceará, Brasil, e foram aprovadas pelo Comitê de Ética em Pesquisa - COMEPE (Protocolo Nº 35/06).

Esse dataset já foi amplamente abordado e discutido na literatura (HAN *et al.*, 2020; HU *et al.*, 2020; JÚNIOR *et al.*, ; FILHO *et al.*, 2019; FILHO *et al.*, 2017; FILHO *et al.*, 2014; RAMALHO *et al.*, 2016; REBOUCAS *et al.*, 2015), em contexto de segmentação de pulmões, inclusive junto à Detectron2, que é o modelo de segmentação que será abordado a seguir.

3.2 Transferência de Aprendizado com Modelos Deep

No primeiro momento, as imagens originais do conjunto de dados foram fornecidas aos modelos de aprendizado profundo (VGG, DenseNet e MobileNet) para extrair características

profundas de cada amostra. O único pré-processamento utilizado nas imagens foi o redimensionamento para atingir as dimensões de entrada de cada extrator profundo. Os parâmetros e a topologia da CNN não foram alterados neste experimento, e os pesos carregados eram provenientes do banco de dados ImageNet (DENG *et al.*, 2009). Os modelos foram carregados sem as camadas totalmente conectadas (usadas na classificação), uma vez que o objetivo nesta etapa é apenas extrair vetores de características que serão passados para os classificadores tradicionais.

A Tabela 4 traz um resumo das arquiteturas de CNN utilizadas nesse estudo, assim como suas configurações utilizadas. Para cada configuração de CNN, a tabela também mostra as dimensões da imagem de entrada (em pixels) e a quantidade de atributos extraídos no vetor de características resultantes da extração.

Tabela 4 – Configurações das arquiteturas de CNN utilizadas na abordagem.

Arquitetura	Configurações	Dimensões de Entrada	Atributos Extraídos
VGG (SIMONYAN; ZISSERMAN, 2014)	VGG16	224 x 224	512
	VGG19	224 x 224	512
MobileNet (HOWARD <i>et al.</i> , 2017)	MobileNet	224 x 224	1024
DenseNet (HUANG <i>et al.</i> , 2017)	DenseNet121	224 x 224	1024
	DenseNet169	224 x 224	1664
	DenseNet201	224 x 224	1920

Fonte: Elaborada pelo autor.

Após a fase de extração, com os vetores de características extraídos, inicia-se a etapa de classificação. Os conjuntos de dados formados pelos vetores de características foram utilizados em uma classificação supervisionada binária, utilizando sete diferentes classificadores. Antes da classificação em si, um algoritmo de Busca Aleatória (*Random Search*) foi utilizado para garantir os melhores hiperparâmetros para cada um desses classificadores. Os intervalos de busca para os hiperparâmetros de cada classificador podem ser vistos na Tabela 5.

Após a busca aleatória, a classificação foi realizada para cada um dos sete classificadores usando a técnica de validação cruzada, com 5 pastas, e. g., cinco rodadas de classificação com proporções de 80%/20% para treino e validação.

Por fim, avaliou-se o desempenho de cada modelo utilizando as cinco métricas citadas anteriormente: Acurácia, Precisão, Recall, F1-Score e Coeficiente de Correlação de Matthews (MCC), a fim de alcançar a melhor combinação extrator-classificador, que será usada no experimento seguinte.

Tabela 5 – Intervalos de parâmetros fornecidos ao algoritmo *Random Search* para cada classificador usado na abordagem de transferência aprendizado.

Classificador	Parâmetros de Busca	Intervalos de Busca
Naive Bayes	-	-
MLP	Camadas ocultas	[2, 1001]
KNN	Número de vizinhos	1, 3, 5, 7, 9, 11, 13, 15
	Profundidade máxima	6, Ilimitada
Random Forest	<i>Bootstrap</i>	Verdadeiro, Falso
	Critério	Gini, Entropia
SVM Linear	Parâmetro de regularização (C)	$2^x x \in [-5, 15]$
SVM Polinomial	Grau	3, 5, 7, 9
	Parâmetro de regularização (C)	$2^x x \in [-5, 15]$
SVM RBF	Gama	$2^x x \in [-15, 3]$
	Parâmetro de regularização (C)	$2^x x \in [-5, 15]$

Fonte: Elaborada pelo autor.

3.3 Segmentação Deep com Detectron2

Nessa etapa, que faz parte do segundo experimento, a Detectron2 foi treinada para fazer a detecção e segmentação de pulmões a partir de imagens de TC de tórax, usando o dataset em DICOM, do Hospital Walter Cantídio, citado anteriormente (HAN *et al.*, 2020). O objetivo nesta etapa é gerar pesos para a Detectron2, otimizando a segmentação de pulmões a partir de imagens de TC, usando esse modelo treinado na abordagem completamente automática proposta nesse estudo.

Para o treinamento da Detectron2, utilizou-se de 50 mil épocas de treinamento, com o dataset dividido em 80%/20% para treino/validação. O modelo inicial escolhido entre os suportados pelo framework da Detectron2 foi o Mask R-CNN (HE *et al.*, 2017), com uso da ResNet50 (HE *et al.*, 2016) na etapa de Backbone.

O dataset foi passado para o formato COCO (LIN *et al.*, 2014) antes do treinamento para melhor compreensão do algoritmo da Detectron2. Por fim, as classes configuradas para esse treinamento foi apenas o pulmão, e o threshold de detecção da ROI foi de 0.95.

No método completamente automático, após a etapa de detecção e segmentação usando a Detectron2, são geradas novas imagens a partir do resultado da segmentação e da imagem de entrada (original), de modo que uma seja multiplicada pela outra e que fique apenas

os pulmões na imagem que vai ser enviada para a classificação usando o modelo deep.

O método de *Contrast Limited Adaptive Histogram Equalization* (CLAHE) foi aplicado na imagem segmentada, como forma de pós-processamento (JOSHI *et al.*, 2023). O método do CLAHE é uma técnica de equalização de histograma usada para realçar o contraste da imagem, garantindo que o contraste seja amplificado (JOSHI *et al.*, 2023; OMAROVA; STAROVOITOV, 2022; HUSSAIN *et al.*, 2018).

3.4 Ajuste Fino do Melhor Modelo

Por fim, como última fase desse experimento, utilizou-se dos melhores modelos encontrados na etapa de *transfer learning* para refiná-los (usando de *fine-tuning*) a partir das imagens segmentadas pela Detectron2.

Ou seja, após o treinamento da Detectron2, que se deu com imagens de CT de pulmão, ela foi aplicada ao dataset principal desses estudo, o SARS-CoV-2 CT-Scan, para segmentar apenas os pulmões desses exames de CT e melhorar o desempenho de classificação de COVID. Nessa etapa, as imagens segmentadas são fornecidas para os métodos escolhidos, que serão treinados a partir desse novo conjunto de dados segmentados pela Detectron2.

Nessa etapa, o dataset foi dividido em 90% para treino e validação ao longo das épocas de treinamento, e os outros 10% foram usados como conjunto de teste. Esse conjunto de teste tem como objetivo validar o modelo treinado ao final do treinamento, de modo que essas imagens não tenham entrado em nenhum momento na etapa de treinamento, evitando qualquer chance de enviesamento amostral do modelo (KUHN, 2013).

Os modelos escolhidos foram carregados com os seus pesos originais, sem as camadas totalmente conectadas originais, provenientes do dataset ImageNet. Tiveram novas camadas completamente conectadas (destreinadas com ativações ReLU e Softmax) acopladas a arquitetura do modelo, assim como uma camada anterior de *global average pooling*, e assim iniciou-se o treinamento, refinando assim os pesos para o problema específico em questão (classificação de COVID). Foram usadas 50 épocas de treinamento para esse modelo, em uma combinação de 90%/10% entre os dados destinados para treino e validação.

Foi utilizado o solver Adam, com learning rate de 0,0003, e a função *loss* a ser minimizada foi a *binary cross-entropy*. O tamanho de lote (*batch*) utilizado foi definido como 32. Todas as camadas convolucionais foram refinadas, não deixando nenhuma congelada. O dataset de treinamento passou por um processo de *augmentation* usando as técnicas de flip horizontal,

zoom de 0,2, *shear* de 0,2 e deslocamentos verticais e horizontais de 0,1. As dimensões de entrada foram mantidas como 224×224 .

Assim como na etapa de *transfer learning*, após esse momento os modelos treinados foram avaliados através das métricas de avaliação e do Grad-CAM para análise de interpretabilidade do modelo.

4 RESULTADOS

Essa seção trata dos resultados obtidos nos dois principais experimentos desse trabalho. Enquanto no primeiro momento, buscou-se comparar entre métodos de transfer learning, com foco em selecionar o método com melhor performance, a fim de utilizá-lo no segundo momento, quando este modelo será refinado a partir de imagens segmentadas pela Detectron2.

4.1 Resultados dos Modelos com *Transfer Learning*

Após a primeira fase do experimento, que foi a etapa de *transfer learning*, as métricas de avaliação de cada modelo extrator-classificador foram organizadas e podem ser vistas na Tabela 6

Através dessa análise, é possível identificar as melhores combinações em termos de acurácia, precisão, *recall*, F1-score e Além disso, também é possível considerar o tempo de processamento de cada abordagem, que foi separado entre tempo de treino e tempo de predição, e podem ser vistos na Tabela 7.

Percebe-se que as melhores combinação, entre os modelos avaliados, foram as arquiteturas MobileNet e VGG19 destacando-se tanto em desempenho quanto em eficiência temporal. O MobileNet, conhecido por sua leveza e rapidez, apresentou o melhor desempenho com o classificador KNN, alcançando uma acurácia média de 97,31% e MCC de 94,67%. Essa combinação também obteve mais de 97% nas médias de *precision*, *recall* e F1-Score. Além disso, foi uma das arquiteturas mais rápidas no processamento, evidenciando sua aplicabilidade em cenários que demandam alto desempenho com restrições de tempo. O tempo médio de execução dessa combinação foi aproximadamente 12 milissegundos por predição, enquanto a VGG19 levou em média 7 milissegundos nas mesmas condições.

A VGG19, por sua vez, obteve métricas muito competitivas quando combinada com o classificador KNN, alcançando uma acurácia de 96,91% e MCC de 93,85%. Apesar de não superar o MobileNet em eficiência, ela demonstrou robustez, provavelmente devido à sua capacidade de extração de características mais detalhadas.

A arquitetura DenseNet foi o que ficou mais abaixo entre as três arquiteturas escolhidas, enquanto a arquitetura VGG (principalmente a VGG19) obteve os melhores resultados após a MobileNet. Entre os classificadores, destacam-se principalmente o KNN, mostrando que há

Tabela 6 – Métricas (%) obtidas por cada combinação de extrator e classificador na etapa de *transfer learning* desse estudo.

Extrator	Classificador	Acurácia	Precisão	Recall	F1-Score	MCC
VGG16	Naive Bayes	82,54±1,63	82,87±1,76	82,58±1,63	82,51±1,62	65,46±3,39
	MLP	89,84±1,02	89,91±0,99	89,85±1,03	89,83±1,03	79,76±2,02
	KNN	96,87±0,97	96,90±0,94	96,86±0,98	96,87±0,97	93,76±1,91
	Random Forest	89,61±1,15	89,69±1,20	89,62±1,16	89,61±1,15	79,31±2,36
	SVM Linear	89,93±0,80	90,01±0,89	89,95±0,81	89,92±0,80	79,95±1,70
	SVM Polinomial	49,91±0,44	24,95±0,22	50,00±0,00	33,29±0,19	0,00±0,00
	SVM RBF	95,34±1,05	95,36±1,06	95,34±1,05	95,34±1,05	90,71±2,11
VGG19	Naive Bayes	82,27±1,89	82,50±1,80	82,31±1,88	82,25±1,90	64,80±3,68
	MLP	90,51±2,30	90,56±2,30	90,50±2,30	90,50±2,30	81,06±4,60
	KNN	96,91±0,58	96,94±0,58	96,90±0,57	96,91±0,58	93,85±1,15
	Random Forest	89,17±1,86	89,21±1,84	89,17±1,85	89,16±1,86	78,38±3,69
	SVM Linear	90,73±1,40	90,76±1,41	90,73±1,39	90,73±1,40	81,49±2,80
	SVM Polinomial	49,91±0,44	24,95±0,22	50,00±0,00	33,29±0,19	0,00±0,00
	SVM RBF	96,01±1,31	96,05±1,32	96,01±1,31	96,01±1,31	92,05±2,62
MobileNet	Naive Bayes	79,32±1,44	79,73±1,27	79,37±1,43	79,26±1,48	59,09±2,69
	MLP	92,03±1,16	92,06±1,17	92,04±1,16	92,03±1,16	84,10±2,33
	KNN	97,31±0,58	97,37±0,60	97,30±0,58	97,31±0,58	94,67±1,18
	Random Forest	87,74±1,07	88,24±0,92	87,79±1,06	87,70±1,09	76,03±1,99
	SVM Linear	91,32±0,87	91,36±0,84	91,33±0,86	91,32±0,87	82,69±1,70
	SVM Polinomial	95,26±0,72	95,27±0,72	95,26±0,71	95,25±0,71	90,53±1,43
	SVM RBF	95,35±0,90	95,37±0,89	95,36±0,90	95,35±0,90	90,72±1,79
DenseNet121	Naive Bayes	80,48±1,32	80,68±1,23	80,51±1,31	80,46±1,34	61,18±2,53
	MLP	88,72±0,88	88,78±0,88	88,72±0,89	88,71±0,88	77,50±1,77
	KNN	93,78±1,05	93,80±1,04	93,79±1,04	93,78±1,05	87,58±2,08
	Random Forest	88,41±0,80	88,52±0,76	88,43±0,79	88,40±0,80	76,95±1,55
	SVM Linear	88,14±0,66	88,21±0,64	88,15±0,66	88,13±0,66	76,37±1,30
	SVM Polinomial	49,91±0,44	24,95±0,22	50,00±0,00	33,29±0,19	0,00±0,00
	SVM RBF	94,00±0,56	94,01±0,55	94,00±0,56	94,00±0,56	88,01±1,12
DenseNet169	Naive Bayes	81,11±0,90	81,30±0,82	81,14±0,90	81,09±0,92	62,44±1,70
	MLP	90,46±1,91	90,48±1,88	90,47±1,90	90,46±1,91	80,95±3,78
	KNN	94,76±1,30	94,81±1,33	94,75±1,29	94,76±1,30	89,56±2,62
	Random Forest	87,60±1,59	87,86±1,47	87,63±1,58	87,58±1,61	75,49±3,04
	SVM Linear	90,06±2,29	90,13±2,26	90,07±2,28	90,06±2,29	80,21±4,54
	SVM Polinomial	49,91±0,44	24,95±0,22	50,00±0,00	33,29±0,19	0,00±0,00
	SVM RBF	94,58±1,81	94,61±1,79	94,58±1,80	94,58±1,81	89,20±3,59
DenseNet201	Naive Bayes	81,47±1,81	81,52±1,84	81,48±1,81	81,47±1,81	63,01±3,65
	MLP	91,32±0,82	91,39±0,80	91,31±0,83	91,31±0,82	82,70±1,62
	KNN	95,03±1,35	95,09±1,38	95,03±1,34	95,03±1,35	90,12±2,73
	Random Forest	88,59±1,46	88,77±1,52	88,62±1,46	88,58±1,46	77,38±2,98
	SVM Linear	90,78±0,41	90,84±0,38	90,79±0,41	90,78±0,41	81,63±0,79
	SVM Polinomial	49,91±0,44	24,95±0,22	50,00±0,00	33,29±0,19	0,00±0,00
	SVM RBF	94,09±0,22	94,10±0,23	94,09±0,23	94,09±0,22	88,19±0,45

Fonte: Elaborada pelo autor.

uma boa distribuição espacial entre os dados, e o SVM RBF, mostrando que o kernel de função de base radial também é um bom hiperplano para separação ótima dos dados.

Dessa forma, pelos resultados promissores da **MobileNet** com as imagens originais, ela foi a escolhida para ser refinada usando fine-tuning e para ser o modelo escolhido na abordagem totalmente automática proposta por esse estudo.

Tabela 7 – Tempos (ms) obtidos por cada combinação de extrator e classificador na etapa de *transfer learning* desse estudo.

Extrator	Classificador	Tempo de treino	Tempo de predição
VGG16	Naive Bayes	0.057±0.013	0.080±0.017
	MLP	75.660±4.859	0.218±0.045
	KNN	0.021±0.001	5.979±0.234
	Random Forest	20.107±0.094	2.955±0.029
	SVM Linear	2.358±0.058	1.186±0.019
	SVM Polinomial	5.653±0.189	3.733±0.056
	SVM RBF	5.632±0.047	3.355±0.058
VGG19	Naive Bayes	0.057±0.017	0.058±0.027
	MLP	25.492±4.048	0.101±0.040
	KNN	0.021±0.001	6.049±0.059
	Random Forest	17.439±0.051	2.943±0.013
	SVM Linear	2.081±0.036	0.953±0.028
	SVM Polinomial	5.687±0.039	3.820±0.129
	SVM RBF	4.533±0.051	2.075±0.027
MobileNet	Naive Bayes	0.080±0.024	0.057±0.008
	MLP	57.912±3.493	0.307±0.089
	KNN	0.059±0.015	12.442±0.093
	Random Forest	12.159±0.045	3.417±0.028
	SVM Linear	5.010±0.097	2.189±0.044
	SVM Polinomial	10.006±0.262	4.008±0.070
	SVM RBF	12.837±0.469	6.822±0.051
DenseNet121	Naive Bayes	0.083±0.018	0.121±0.089
	MLP	42.823±1.936	0.130±0.058
	KNN	0.049±0.001	12.810±0.158
	Random Forest	15.592±0.040	3.390±0.060
	SVM Linear	5.531±0.183	2.582±0.020
	SVM Polinomial	13.550±0.393	9.213±0.041
	SVM RBF	10.466±0.282	4.671±0.079
DenseNet169	Naive Bayes	0.077±0.031	0.065±0.037
	MLP	280.836±12.321	0.719±0.277
	KNN	0.086±0.001	20.423±0.034
	Random Forest	12.644±0.048	3.084±0.042
	SVM Linear	8.871±0.149	4.526±0.079
	SVM Polinomial	22.199±0.068	15.833±0.055
	SVM RBF	20.092±0.100	10.639±0.164
DenseNet201	Naive Bayes	0.067±0.027	0.098±0.039
	MLP	241.805±14.299	0.882±0.169
	KNN	0.102±0.001	23.520±0.026
	Random Forest	12.331±0.122	3.457±0.048
	SVM Linear	9.873±0.067	5.136±0.056
	SVM Polinomial	25.808±0.085	18.468±0.082
	SVM RBF	12.222±0.101	7.237±0.074

Fonte: Elaborada pelo autor.

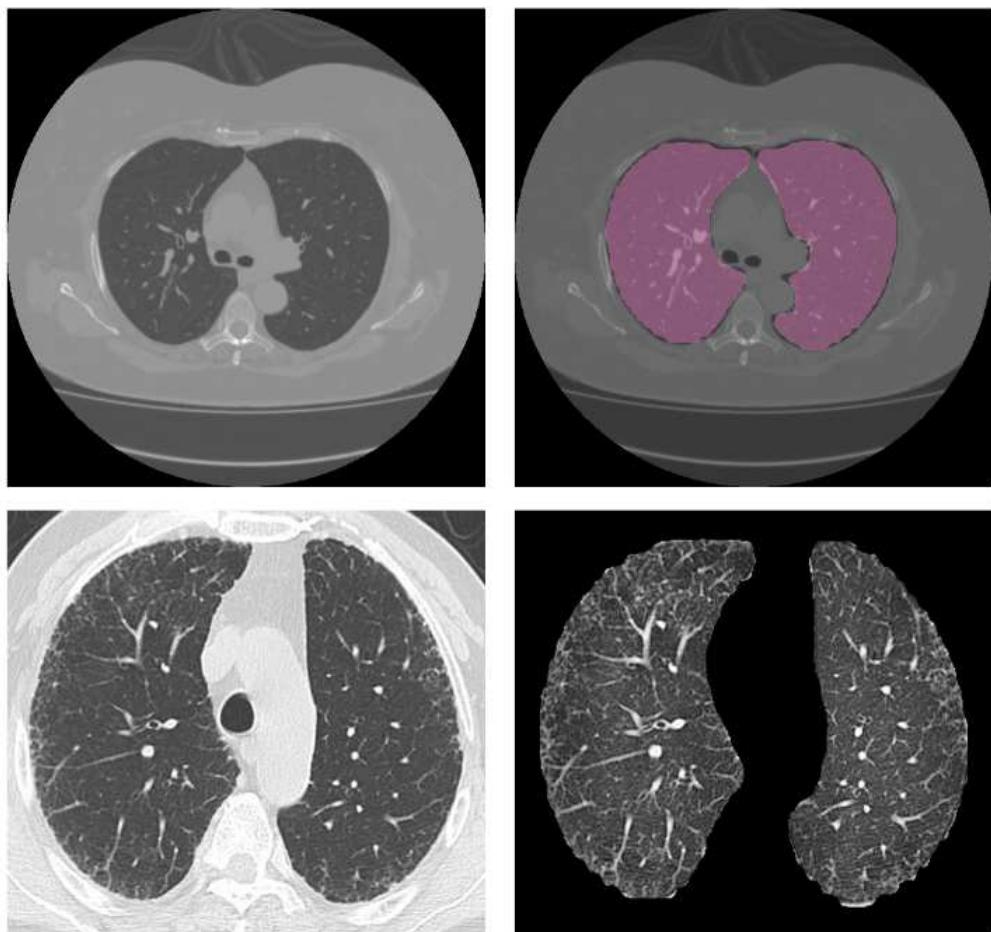
4.2 Resultados da Segmentação Pulmonar com Detectron2

O treinamento da Detectron2 se deu usando o segundo dataset abordado nesse trabalho, que consta com imagens de TC de pulmão e seus respectivos GTs, para segmentação.

As métricas para avaliação da segmentação não serão abordadas nesse trabalho, pois

o objetivo da segmentação é apenas gerar um modelo treinado para segmentação de pulmões, a fim de melhorar o desempenho da classificação do modelo. No entanto, é possível visualizar o desempenho da Detectron2 para segmentação de pulmões em imagens de TC, para os dois diferentes datasets desse estudo, tal como mostra as Figuras 6 e 7.

Figura 6 – Amostras de TC dos dois datasets desse estudo segmentadas pela Detectron2. Acima, uma imagem do dataset de treinamento e a respectiva máscara em rosa. Abaixo, uma amostra do dataset SARS-CoV-2 CT-Scan e a respectiva segmentação.

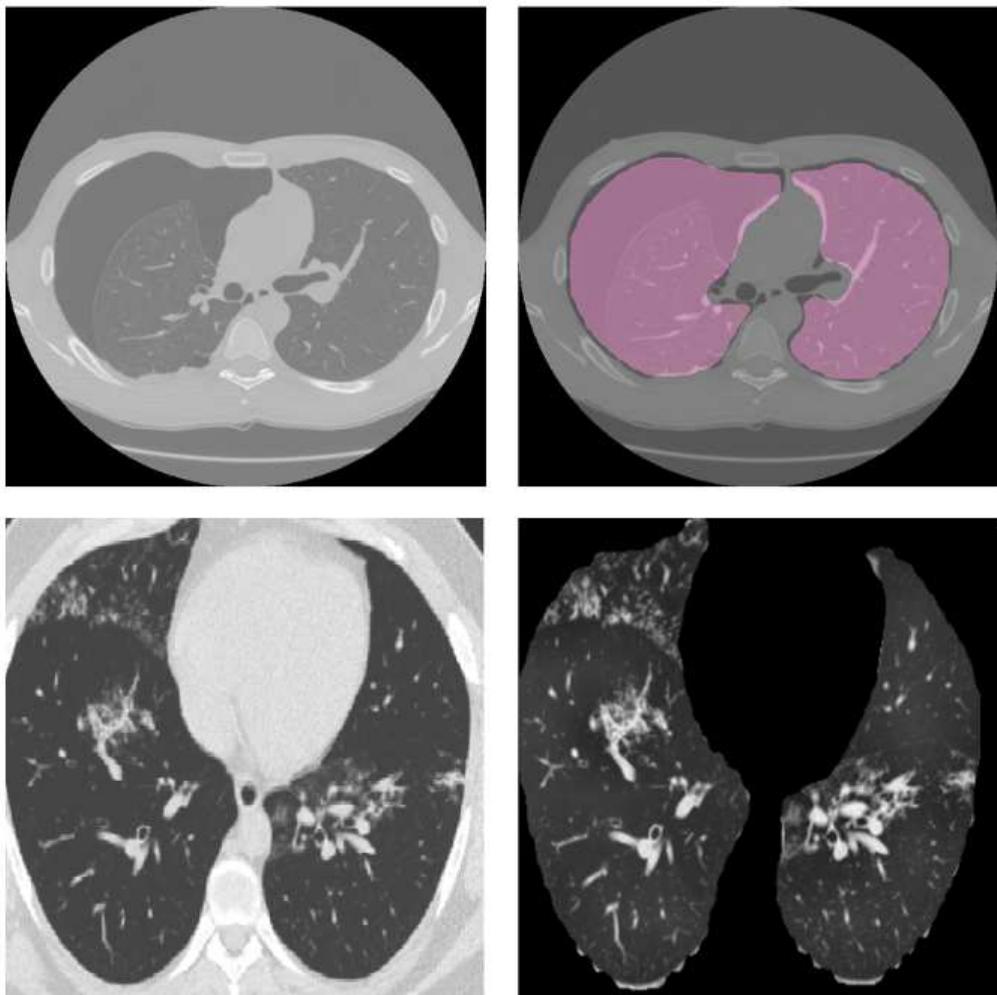


Fonte: Elaborada pelo autor.

Com a Detectron2 devidamente treinada para segmentar imagens de pulmão a partir de imagens de TC, aplicou-a ao dataset principal desse estudo, o SARS-CoV-2 CT-Scan, a fim de que, a partir das imagens segmentadas, possa-se fazer uma classificação mais acurada e eficiente na próxima etapa desse estudo, que é o ajuste fino da rede escolhida.

Como é possível ver nas imagens, a Detectron2 demonstrou boa generalização na segmentação dos pulmões para a maioria das amostras. Contudo, é notável que algumas regiões mais delgadas, próximas às bordas, não são completamente segmentadas pela rede. Apesar disso, essa limitação não compromete significativamente a classificação de COVID-19. Isso ocorre

Figura 7 – Mais mostras de TC dos dois datasets desse estudo segmentadas pela Detectron2. Acima, uma imagem do dataset de treinamento e a respectiva máscara em rosa. Abaixo, uma amostra do dataset SARS-CoV-2 CT-Scan e a respectiva segmentação.



Fonte: Elaborada pelo autor.

porque as alterações causadas pela doença, frequentemente visíveis como áreas em branco (como ilustrado na Figura 7), tendem a se localizar predominantemente nas regiões centrais do pulmão, que são adequadamente capturadas pela segmentação, como constatado empiricamente.

Por fim, com as imagens segmentadas, aplicou-se o método de CLAHE, realçando e melhorando o contraste dos pulmões segmentados. Dessa forma, as imagens segmentadas foram usadas para o refinamento do modelo treinado (seção abaixo), e a segmentação se tornou parte chave da *pipeline* do método.

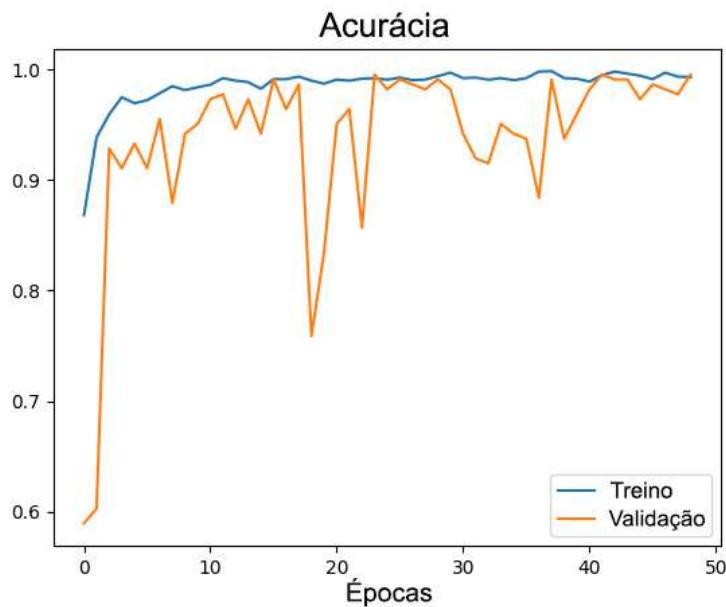
4.3 ***Fine Tuning* do Melhor Modelo**

Após a etapa de segmentação com a Detectron2, as imagens segmentadas foram usadas para o treinamento de uma MobileNet, inicializada com os pesos da ImageNet e sem suas

camadas fully-connected originais. Foram acrescidas mais três camadas ao modelo convolucional, sendo elas: uma camada de Global Average Pooling, uma camada fully-connected de ativação ReLU e uma camada fully-connected de ativação softmax, responsável pela classificação em COVID ou Non-Covid. Os resultados da validação do modelo treinado, após as 50 épocas de treinamento foram as seguintes: 99.60% de acurácia, 99.60% de precision, 99.59% de recall, 99.60% de F1-Score e 99.20% de MCC. O tempo total de treinamento se deu em 1.3 horas e o tempo de predição foi de 7 milissegundos para cada amostra da validação.

As curvas de acurácia e de *loss* do ajuste fino do modelo de MobileNet, em 50 épocas, podem ser vistos nas Figuras 8 e 9.

Figura 8 – Curvas de acurácia para o treinamento e validação no ajuste fino do modelo MobileNet.

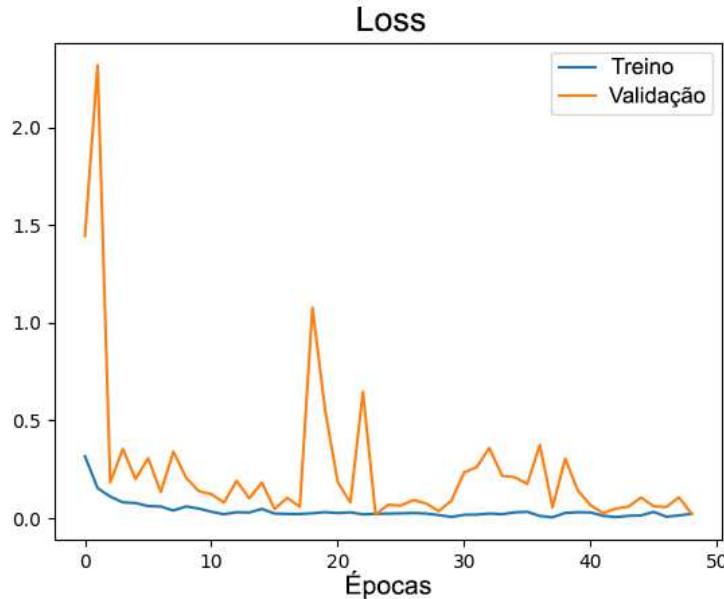


Fonte: Elaborada pelo autor.

As curvas mostram que o modelo com seus pesos tradicionais já era robusto para diferenciar os atributos de imagens da classe COVID para a classe Non-COVID. Isso é explicado pelo desempenho no experimento de *transfer learning* e é confirmado pelas curvas de acurácia e *loss*, que já começam o treinamento em valores razoáveis. Observando as curvas, também é possível ver que as 50 épocas são suficientes para o ajuste fino da rede, que converge nas últimas épocas resultando em um bom desempenho classificatório. As oscilações bruscas ao longo da curva mostram que o *batch* escolhido foi baixo, mas suficiente para o problema.

Os resultados obtidos pelo modelo proposto, em comparação com outros métodos da literatura podem ser vistos na Tabela 8. A parte do dataset usado para essa etapa de teste foram

Figura 9 – Curvas de *loss* para o treinamento e validação no ajuste fino do modelo MobileNet.



Fonte: Elaborada pelo autor.

os 10% das amostras citadas na Seção 3, segundo a abordagem de divisão dos dados entre treino, validação e teste (KUHN, 2013).

Tabela 8 – Comparação entre as métricas obtidas pelo método proposto (com o conjunto de teste) em comparação com outros métodos da literatura que também fizeram classificação binária do dataset SARS-CoV-2 CT-scan.

Modelo	Acurácia	Precisão	Recall	F1-Score	MCC
Método Proposto	99,60	99,60	99,59	99,60	99,20
ResNet101 (ALSHAZLY <i>et al.</i> , 2021a)	99,4±0,4	99,6±0,3	99,1±0,6	99,4±0,4	-
xDNN (SOARES <i>et al.</i> , 2020)	97,38	99,16	95,53	97,31	-
CNN ² -LSTM (NAEEM; BIN-SALEM, 2021)	98,94	99,00	99,00	99,00	-
DenseNet201 (JAISWAL <i>et al.</i> , 2021)	96,2	96,2	96,2	96,2	-
VGG19 Modificada (PANWAR <i>et al.</i> , 2020)	95,0	95,3	94,0	94,3	-

Fonte: Elaborada pelo autor.

Analizando a tabela, percebe-se que o método proposto teve desempenho na validação equivalente ao melhor método concorrente e superior a todos os demais métodos presentes na literatura para praticamente todas as métricas de avaliação. O MCC mostra que mesmo com o dataset não sendo balanceado, o método proposto obteve bom desempenho para ambas as classes. A alta acurácia e *recall* (ou sensitividade) mostram que o modelo tem boas taxas de acerto e minimiza as taxas de falsos negativos, o que é crucial quando se trata de diagnóstico médico.

Os autores citados usaram outras métricas de comparação que não foram utilizadas nesse trabalho como é o caso da especificidade, pois comprehende-se que o F1-Score pode fazer

essa mesma interpretação. O MCC não foi utilizado pelos outros autores, mas foi mantido aqui pois entende-se que é uma métrica útil para datasets desbalanceados (como é o caso do SARS-CoV-2 CT-scan). Alguns resultados apresentam desvio-padrão pois fizeram mais de uma rodada de teste, que não foi o caso desse trabalho.

Não foram utilizados testes estatísticos para comprovar a diferença estatística entre os métodos pois a premissa desse estudo é a *pipeline* completamente automática no diagnóstico de COVID-19, com avaliação numérica e visual (usando XAI). Dessa forma, um desempenho equivalente, ou ligeiramente superior, aos demais métodos da literatura é suficiente para os objetivos listados nesse trabalho, que usou de uma rede leve (MobileNet), com o processamento automático da segmentação (com Detectron2) e visualização rápida dos resultados (usando Grad-CAM).

4.4 Análise Visual com Grad-CAM

Por fim, na última etapa desse trabalho fez-se a análise visual da interpretabilidade dos modelos através do algoritmo de Grad-CAM. O Grad-CAM permite analisar em que áreas a rede neural convolucional mais deu enfoque para a predição daquela amostra.

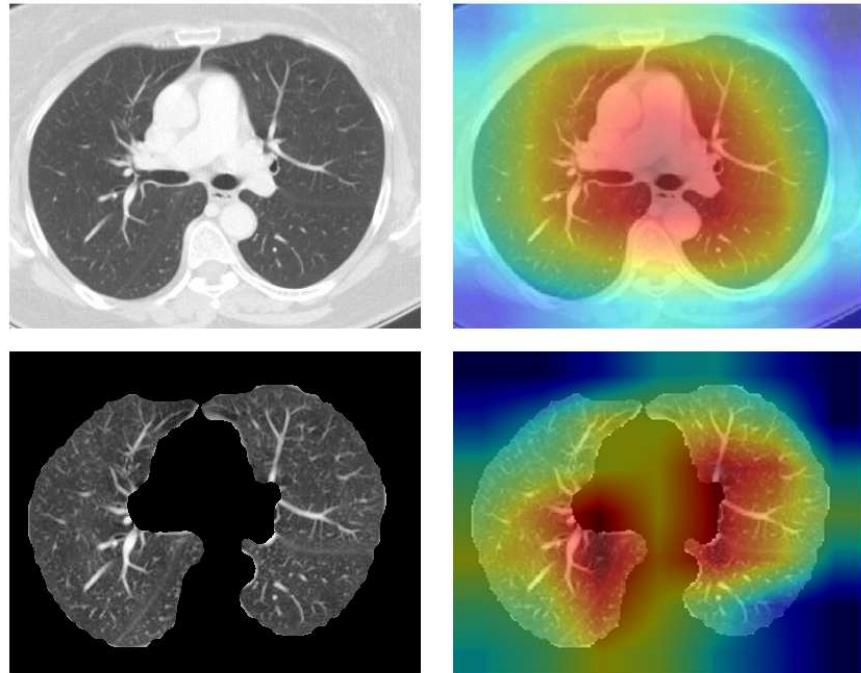
Em outras palavras, quais são as áreas da imagem que mais impactam para a rede classificar uma amostra como sendo positiva ou não para o vírus da COVID.

Como ilustrado nas Figuras 10, 11, 12 e 13, que representam dois exemplos da classe Non-Covid e dois exemplos da classe Covid, respectivamente, o padrão utilizado é sempre das imagens do dataset original no canto superior esquerdo, e seu devido Grad-CAM através da MobileNet, na etapa de *transfer learning*, à direita. Abaixo e à esquerda, tem-se as mesmas imagens segmentadas pela Detectron2, que faz parte do modelo proposto totalmente automático, enquanto à sua direita, vê-se o Grad-CAM da mesma imagem, mas agora utilizando a MobileNet do modelo proposto, que teve seus pesos refinados para essas imagens.

Pode-se ver que nos casos do modelo proposto, o Grad-CAM passou a ser bem mais localizado nos pulmões, tirando o foco do centro da imagem e comprovando que o aumento das métricas passa por uma melhor análise da rede nas características que classificam ou não a COVID-19 nos pulmões.

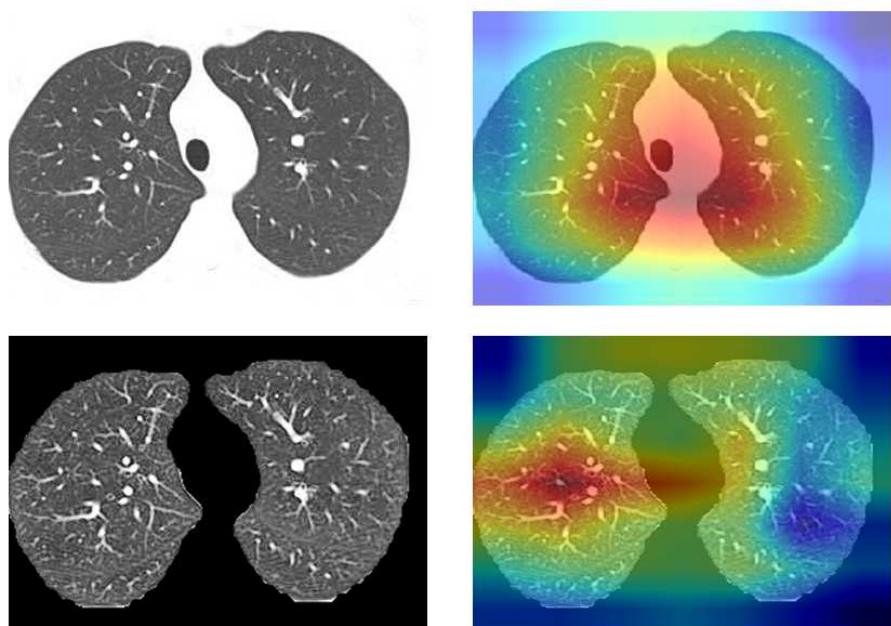
Portanto, a etapa de visualização com Grad-CAM, que representa o uso de XAI nesse estudo, desempenhou um papel crucial ao fornecer uma explicação interpretável para as decisões dos modelos. Essa abordagem validou a capacidade do modelo proposto de identificar

Figura 10 – Acima e à esquerda, uma amostra da classe Non-Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.



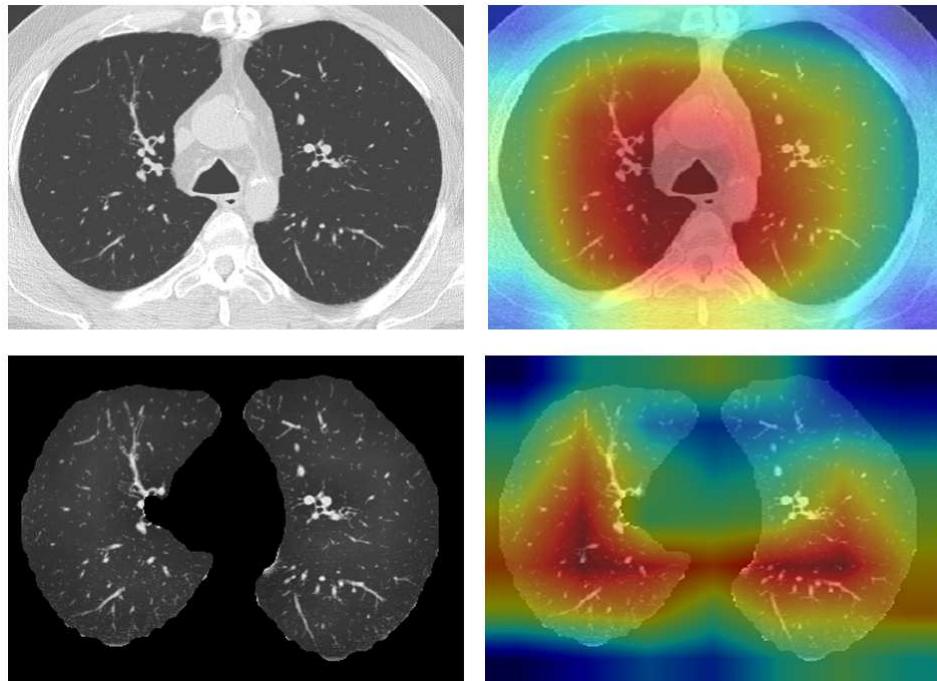
Fonte: Elaborada pelo autor.

Figura 11 – Acima e à esquerda, uma amostra da classe Non-Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.



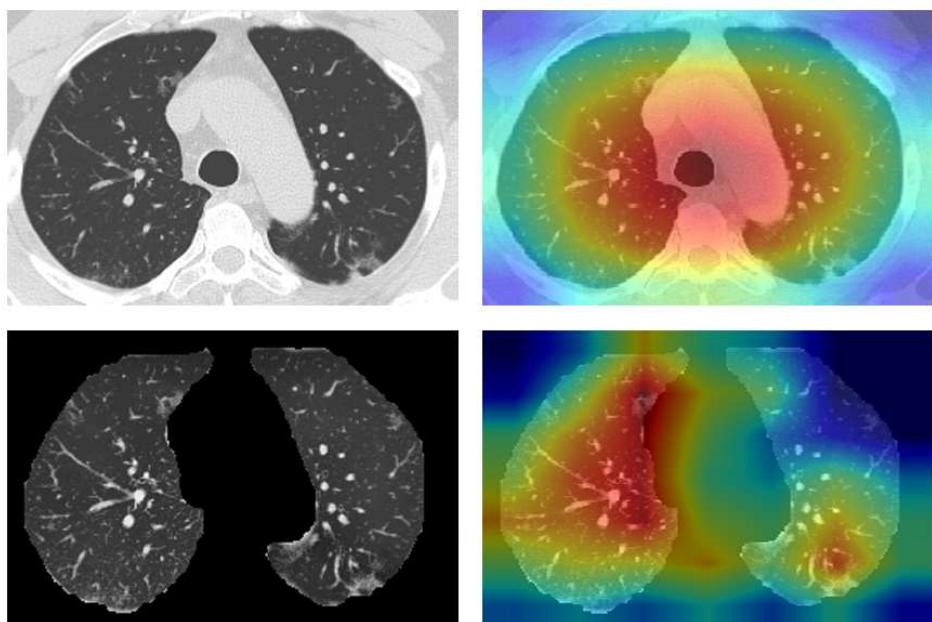
Fonte: Elaborada pelo autor.

Figura 12 – Acima e à esquerda, uma amostra da classe Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.



Fonte: Elaborada pelo autor.

Figura 13 – Acima e à esquerda, uma amostra da classe Covid do dataset original, com o Grad-CAM gerado pela MobileNet (sem fine-tuning) à direita. Abaixo e à esquerda, a mesma imagem segmentada pela Detectron2, com o Grad-CAM correspondente usando a MobileNet refinada à direita.



Fonte: Elaborada pelo autor.

padrões específicos relacionados à COVID-19 nos pulmões, além de reforçar a confiabilidade dos resultados ao demonstrar que as áreas mais relevantes para a classificação eram consistentes com as regiões fisiológicas da doença. Assim, essa etapa não apenas melhora a compreensão do funcionamento do modelo, mas também agrega valor clínico ao possibilitar a validação dos achados pela perspectiva médica, ampliando a aplicabilidade do sistema em cenários reais.

5 CONCLUSÕES E TRABALHOS FUTUROS

Este estudo propôs uma abordagem totalmente automática para a classificação de COVID-19 em imagens de tomografia computadorizada (CT) torácica. Em um primeiro momento, várias topologias de CNNs foram comparadas, em combinação com diversos classificadores tradicionais, consistindo em um processo inicial de *transfer learning*, para que a melhor combinação fosse escolhida fosse escolhida. Os resultados indicaram que a combinação de MobileNet como extrator e KNN como classificador obteve desempenho promissor na fase de *transfer learning*, justificando sua escolha para o experimento subsequente.

No segundo momento, o método proposto totalmente automático começou a ser construído, priorizando sua eficácia e rapidez. A rede *deep* escolhida foi a MobileNet, pois atingiu os melhores valores de métricas para praticamente todos os classificadores, além de um tempo de predição rápido, por ser uma rede leve por natureza. Por fim, incluindo a segmentação automática dos pulmões usando o framework Detectron2, foi feito o refinamento do modelo escolhido por *fine-tuning*, com base nas imagens segmentadas.

A Detectron2 mostrou eficácia na segmentação dos pulmões, contribuindo para a melhoria da precisão na classificação. O refinamento do modelo MobileNet através de fine-tuning, utilizando as imagens segmentadas, resultou em altas métricas de avaliação, com valores de 99.60% para a acurácia, precisão e F1-Score, com resultados próximos ao do melhor modelo concorrente, e superando os demais métodos da literatura para o dataset SARS-CoV-2 CT-scan e destacando a eficácia do método proposto. O alto valor do Coeficiente de Matthew, de 99.20%, também destaca a alta eficácia mesmo com dados desbalanceados.

Ademais, no que tange ao XAI, a análise visual com Grad-CAM acerca da interpretabilidade do modelo demonstrou que o modelo refinado, utilizando das imagens segmentadas, concentrou seu foco nas regiões dos pulmões, indicando uma compreensão mais precisa das características relacionadas à classificação de COVID-19 nos pulmões em imagens de CT torácico.

Portanto, ao final desse estudo, a *pipeline* completa foi formada, constituindo um modelo totalmente automático, rápido e eficaz para classificação de COVID-19 em imagens de TC torácico, combinando desde a segmentação de pulmões, classificação binária através de uma rede *deep* refinada e a interpretabilidade do modelo através de XAI. Dessa forma, o método proposto se mostrou útil para ser usado no auxílio ao diagnóstico médico.

Para trabalhos futuros, considerando a surpreendente eficácia da MobileNet para a

classificação de COVID-19 em imagens de tomografia computadorizada, pretende-se explorar a implementação desse modelo em sistemas embarcados, com ênfase em *Internet of Things* (*Internet of Things* (IoT)). A utilização de sistemas embarcados para a execução desses modelos pode permitir diagnósticos remotos e em tempo real, facilitando a detecção precoce e a resposta rápida para COVID-19. Dessa forma, pretende-se fazer uma simulação em modelos embarcados, a fim de validar essa hipótese. A visualização através de Grad-CAM também pode ser embarcada e tem como objetivo validar o foco da rede e apontar regiões críticas diretamente a especialistas.

Além disso, como forma de trabalhos futuros, também pretende-se utilizar a mesma *pipeline* desenvolvida nesse estudo para outras doenças pulmonares que podem ser avaliadas através de imagens de TC, e que também seja conveniente a etapa de segmentar os pulmões para auxiliar na classificação do modelo, como é o caso, por exemplo, do câncer de pulmão e da embolia pulmonar.

REFERÊNCIAS

- ADEDOYIN, O. B.; SOYKAN, E. Covid-19 pandemic and online learning: the challenges and opportunities. **Interactive learning environments**, Taylor & Francis, v. 31, n. 2, p. 863–875, 2023.
- AGARAP, A. Deep learning using rectified linear units (relu). **arXiv preprint arXiv:1803.08375**, 2018.
- AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. **Machine learning**, Springer, v. 6, n. 1, p. 37–66, 1991.
- ALBAWI, S.; MOHAMMED, T.; AL-ZAWI, S. **Understanding of a convolutional neural network**. In **2017 international conference on engineering and technology (ICET) 2017 Aug 21 (pp. 1-6)**. [S. l.]: Ieee, 2017.
- ALSHAZLY, H.; LINSE, C.; BARTH, E.; MARTINETZ, T. Explainable covid-19 detection using chest ct scans and deep learning. **Sensors**, v. 21, n. 2, 2021. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/21/2/455>.
- ALSHAZLY, H.; LINSE, C.; BARTH, E.; MARTINETZ, T. Explainable covid-19 detection using chest ct scans and deep learning. **Sensors**, MDPI, v. 21, n. 2, p. 455, 2021.
- BRIDLE, J. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. **Advances in neural information processing systems**, v. 2, 1989.
- CORONAVIRUS, N. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed on, v. 10, 2020.
- DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE. **2009 IEEE conference on computer vision and pattern recognition**. [S. l.], 2009. p. 248–255.
- FERNANDEZ, F.-G. **TorchCAM: class activation explorer**. [S. l.]: GitHub, 2020. <https://github.com/frgfm/torch-cam>.
- FILHO, P. P. R.; BARROS, A. C. da S.; ALMEIDA, J. S.; RODRIGUES, J.; ALBUQUERQUE, V. H. C. de. A new effective and powerful medical image segmentation algorithm based on optimum path snakes. **Applied Soft Computing**, Elsevier, v. 76, p. 649–670, 2019.
- FILHO, P. P. R.; CORTEZ, P. C.; BARROS, A. C. da S.; ALBUQUERQUE, V. H. C. D. Novel adaptive balloon active contour method based on internal force for image segmentation—a systematic evaluation on synthetic and real images. **Expert Systems with Applications**, Elsevier, v. 41, n. 17, p. 7707–7721, 2014.
- FILHO, P. P. R.; CORTEZ, P. C.; BARROS, A. C. da S.; ALBUQUERQUE, V. H. C.; TAVARES, J. M. R. Novel and powerful 3d adaptive crisp active contour method applied in the segmentation of ct lung images. **Medical image analysis**, Elsevier, v. 35, p. 503–516, 2017.
- GIRSHICK, R.; RADOSAVOVIC, I.; GKIOXARI, G. P. Doll ar, and K. He, “Detectron,” **2018**.
- GONZALEZ, R. C.; WOODS, R. E. Processamento digital de imagem. **Pearson, ISBN-10: 8576054019**, v. 10, p. 11–27, 2010.

- GUAN, H.; LIU, M. Domain adaptation for medical image analysis: a survey. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 69, n. 3, p. 1173–1185, 2021.
- HAN, T.; NUNES, V. X.; SOUZA, L. F. D. F.; MARQUES, A. G.; SILVA, I. C. L.; JUNIOR, M. A. A. F.; SUN, J.; FILHO, P. P. R. Internet of medical things—based on deep learning techniques for segmentation of lung and stroke regions in ct scans. **IEEE Access**, IEEE, v. 8, p. 71117–71135, 2020.
- HAYKIN, S. **Redes neurais: princípios e prática**. [S. l.]: Bookman Editora, 2001.
- HAYKIN, S. S. *et al.* **Neural networks and learning machines/Simon Haykin**. [S. l.]: New York: Prentice Hall, 2009.
- HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S. l.: s. n.], 2017. p. 2961–2969.
- 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.
- HE, X.; YANG, X.; ZHANG, S.; ZHAO, J.; ZHANG, Y.; XING, E.; XIE, P. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. **medrxiv**, Cold Spring Harbor Laboratory Press, p. 2020–04, 2020.
- HO, T. K. The random subspace method for constructing decision forests. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 20, n. 8, p. 832–844, 1998.
- HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDRETTI, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.
- HU, Q.; SOUZA, L. F. d. F.; HOLANDA, G. B.; ALVES, S. S.; SILVA, F. H. d. S.; HAN, T.; FILHO, P. P. R. An effective approach for ct lung segmentation using mask region-based convolutional neural networks. **Artificial intelligence in medicine**, Elsevier, v. 103, p. 101792, 2020.
- 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.
- 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.
- HUSSAIN, K.; RAHMAN, S.; RAHMAN, M. M.; KHALED, S. M.; WADUD, M. A.-A.; KHAN, M. A. H.; SHOYAIB, M. A histogram specification technique for dark image enhancement using a local transformation method. **IPSJ Transactions on Computer Vision and Applications**, Springer, v. 10, p. 1–11, 2018.
- IOFFE, S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. **arXiv preprint arXiv:1502.03167**, 2015.

- JAISWAL, A.; GIANCHANDANI, N.; SINGH, D.; KUMAR, V.; KAUR, M. Classification of the covid-19 infected patients using densenet201 based deep transfer learning. **Journal of Biomolecular Structure and Dynamics**, Taylor & Francis, v. 39, n. 15, p. 5682–5689, 2021.
- JOSHI, K. K.; GUPTA, K.; AGRAWAL, J. An efficient transfer learning approach for prediction and classification of sars–covid-19. **Multimedia Tools and Applications**, Springer, p. 1–23, 2023.
- JÚNIOR, P. C. S.; SOUZA, L. F. de F.; NASCIMENTO, J. J. da C.; SANTOS, L. O. dos; MARQUES, A. G.; RIBEIRO, F. E. S.; FILHO, P. P. R. Detection and segmentation of lungs regions using cnn combined with levelset.
- KUHN, M. **Applied predictive modeling**. [S. l.]: Springer, 2013.
- LECUN, Y.; BOSEN, 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.
- LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft coco: Common objects in context. In: SPRINGER. **Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13**. [S. l.], 2014. p. 740–755.
- NAEEM, H.; BIN-SALEM, A. A. A cnn-lstm network with multi-level feature extraction-based approach for automated detection of coronavirus from ct scan and x-ray images. **Applied Soft Computing**, Elsevier, v. 113, p. 107918, 2021.
- OHATA, E. F.; CHAGAS, J. V. S. das; BEZERRA, G. M.; HASSAN, M. M.; ALBUQUERQUE, V. H. C. de; FILHO, P. P. R. A novel transfer learning approach for the classification of histological images of colorectal cancer. **The Journal of Supercomputing**, Springer, p. 1–26, 2021.
- OMAROVA, G.; STAROVOITOV, V. Application of the clahe method contrast enhancement of x-ray images. Science and Information Organization, 2022.
- ORENSTEIN, E. C.; BEIJBOM, O. Transfer learning and deep feature extraction for planktonic image data sets. In: IEEE. **2017 IEEE Winter Conference on Applications of Computer Vision (WACV)**. [S. l.], 2017. p. 1082–1088.
- PANWAR, H.; GUPTA, P.; SIDDIQUI, M. K.; MORALES-MENENDEZ, R.; BHARDWAJ, P.; SINGH, V. A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images. **Chaos, Solitons & Fractals**, v. 140, p. 110190, 2020. ISSN 0960-0779. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0960077920305865>.
- PAUL, R.; HAWKINS, S. H.; BALAGURUNATHAN, Y.; SCHABATH, M. B.; GILLIES, R. J.; HALL, L. O.; GOLDGOF, D. B. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. **Tomography**, Grapho Publications, v. 2, n. 4, p. 388, 2016.
- RAMALHO, G. L. B.; FERREIRA, D. S.; FILHO, P. P. R.; MEDEIROS, F. N. S. de. Rotation-invariant feature extraction using a structural co-occurrence matrix. **Measurement**, Elsevier, v. 94, p. 406–415, 2016.

REBOUCAS, E. S.; SARMENTO, R. M.; FILHO, P. P. R. 3d adaptive balloon active contour: method of segmentation of structures in three dimensions. **IEEE Latin America Transactions**, IEEE, v. 13, n. 1, p. 195–203, 2015.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; DURBIN, R.; GOLDEN, R.; CHAUVIN, Y. Backpropagation: The basic theory. In: **Backpropagation**. [S. l.]: Psychology Press, 1995. p. 1–34.

SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. In: SPRINGER. **International conference on artificial neural networks**. [S. l.], 2010. p. 92–101.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE international conference on computer vision**. [S. l.: s. n.], 2017. p. 618–626.

SHI, H.; HAN, X.; JIANG, N.; CAO, Y.; ALWALID, O.; GU, J.; FAN, Y.; ZHENG, C. Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study. **The Lancet infectious diseases**, Elsevier, v. 20, n. 4, p. 425–434, 2020.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SOARES, E.; ANGELOV, P.; BIASO, S.; FROES, M. H.; ABE, D. K. Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification. **MedRxiv**, Cold Spring Harbor Laboratory Press, p. 2020–04, 2020.

SOARES, E.; ANGELOV, P.; BIASO, S.; CURY, M.; ABE, D. A large multiclass dataset of ct scans for covid-19 identification. **Evolving Systems**, Springer, p. 1–6, 2023.

SOHRABI, C.; ALSAFI, Z.; O’NEILL, N.; KHAN, M.; KERWAN, A.; AL-JABIR, A.; IOSIFIDIS, C.; AGHA, R. World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19). **International journal of surgery**, Elsevier, v. 76, p. 71–76, 2020.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

TAJBAKHSH, N.; SHIN, J. Y.; GURUDU, S. R.; HURST, R. T.; KENDALL, C. B.; GOTWAY, M. B.; LIANG, J. Convolutional neural networks for medical image analysis: Full training or fine tuning? **IEEE transactions on medical imaging**, IEEE, v. 35, n. 5, p. 1299–1312, 2016.

THEODORIDIS, S.; KOUTROUMBAS, K. Pattern recognition, academic press. **Burlington, MA**.[Google Scholar], 2008.

VAPNIK, V. N.(1998). **Statistical learning theory**. [S. l.]: Wiley, New York, 1998.

- WANG, Z.; LIU, Q.; DOU, Q. Contrastive cross-site learning with redesigned net for covid-19 ct classification. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 24, n. 10, p. 2806–2813, 2020.
- WU, Y.; KIRILLOV, A.; MASSA, F.; LO, W.-Y.; GIRSHICK, R. Detectron2. 2019.
- XU, B. Empirical evaluation of rectified activations in convolutional network. **arXiv preprint arXiv:1505.00853**, 2015.
- XU, Y.; SANTOS, M. A. dos; SOUZA, L. F. F.; MARQUES, A. G.; ZHANG, L.; NASCIMENTO, J. J. da C.; ALBUQUERQUE, V. H. C. de; FILHO, P. P. R. New fully automatic approach for tissue identification in histopathological examinations using transfer learning. **IET Image Processing**, Wiley Online Library, v. 16, n. 11, p. 2875–2889, 2022.
- ZHAO, W.; JIANG, W.; QIU, X. Deep learning for covid-19 detection based on ct images. **Scientific Reports**, v. 11, n. 1, p. 14353, Jul 2021. ISSN 2045-2322. Disponível em: <https://doi.org/10.1038/s41598-021-93832-2>.
- ZHU, Z.; ALBADAWY, E.; SAHA, A.; ZHANG, J.; HAROWICZ, M. R.; MAZUROWSKI, M. A. Deep learning for identifying radiogenomic associations in breast cancer. **Computers in biology and medicine**, Elsevier, v. 109, p. 85–90, 2019.