

# Suporte para Diagnóstico de COVID-19 por Meio de Classificação Automática de Imagens de Raio-x e Modelos Explicáveis

Bruno Alberto Soares Oliveira  
Fundação para Inovações Tecnológicas  
Belo Horizonte, MG, Brasil  
brunooliveira@fitec.org.br

Hugo Eduardo Ziviani  
VMI Sistemas de Segurança  
Lagoa Santa, MG, Brasil  
hugo.ziviani@vmis.com.br

Juscelino Junior de Oliveira  
Fundação para Inovações Tecnológicas  
Belo Horizonte, MG, Brasil  
jjoliveira@fitec.org.br

Alan Moraes Viegas  
VMI Sistemas de Segurança  
Lagoa Santa, MG, Brasil  
alan@vmis.com.br

Daniel Calvo  
Fundação para Inovações Tecnológicas  
Belo Horizonte, MG, Brasil  
dcalvo@fitec.org.br

**Resumo**—O combate contra a pandemia causada pela COVID-19 é atualmente um dos maiores desafios enfrentados pela humanidade. Um fator chave que tende a desacelerar a disseminação do vírus é o diagnóstico rápido e o isolamento dos pacientes infectados. Pesquisadores em todo mundo vêm buscando formas para suportar os médicos no diagnóstico da doença, incluindo exames como o PCR-RT. O aprendizado profundo aplicado às radiografias de tórax de pacientes suspeitos apresenta um grande potencial para, aliado a outras técnicas, oferecer condições para um diagnóstico rápido e preciso. A falta de dados com boa qualidade e o elevado número de modelos enviesados são desafios que precisam ser enfrentados. Esse trabalho apresenta um sistema de diagnóstico de COVID-19 por meio de imagens de raio-x de tórax, com o objetivo de auxiliar a tomada de decisão do médico. Foi realizado um amplo estudo para a seleção de conjuntos de imagens de qualidade, de forma a evitar que a solução apresente vieses indesejados. Para a aplicação, primeiro os pulmões são segmentados pelo método U-Net para que, em seguida, a imagem seja classificada por um conjunto de modelos treinados para uma classificação binária. Por fim, os resultados são explicados de forma a auxiliar o médico em sua tomada de decisão. Os resultados obtidos com uma taxa de assertividade em 92% mostram que a solução proposta é promissora. Estes resultados indicam que a capacidade do modelo em avaliar os pacientes infectados por COVID-19 pode ser usada no gerenciamento de recursos referentes ao atendimento em hospitais e postos de saúde, bem como auxiliar gestores públicos na tomada de decisões relativas à restrição de atividades para promover o isolamento social.

**Palavras-chave**—Aprendizado Profundo, COVID-19, Diagnóstico, Imagens Médicas, Radiografia de Tórax.

## I. INTRODUÇÃO

No ano de 2019 o mundo foi surpreendido por uma doença provocada pelo vírus SARS-CoV-2, conhecida como COVID-19. Inicialmente, pensava-se tratar de uma doença que atacava os pulmões mas, como passar do tempo, verificou-se que é

uma doença multissistêmica que, além de afetar os pulmões, pode também provocar problemas circulatórios, neurológicos e renais, entre outros. A gravidade da COVID-19 resultou em um esforço global de saúde pública para conter a disseminação do vírus de maneira precoce, já que o mesmo apresenta uma alta taxa de transmissão [1].

Uma etapa crucial no combate à disseminação do vírus é a triagem de pacientes contaminados, de forma que os infectados possam receber o devido tratamento e cuidados imediatos, além de serem isolados para restringir a disseminação do vírus. Considerado o principal método de diagnóstico, o exame PCR-RT [2] se dá por meio de um processo manual, caro e trabalhoso, além de ser desconfortável e invasivo para o paciente. Por outro lado, máquinas de raio-x estão amplamente disponíveis e os exames possuem custos relativamente baixos.

Mesmo passado mais de um ano desde o início da pandemia, existe ainda a necessidade de se buscar soluções para um diagnóstico mais rápido e eficaz, que possa ser aplicado em um grande número de pacientes, tornando evidente a necessidade de se desenvolver e utilizar sistemas automatizados. Atualmente, observamos a disseminação de aplicações de inteligência artificial em diferentes áreas do conhecimento, como nos trabalhos: [3] que utiliza IA na pluviometria; [4] que utiliza *Machine Learning* (ML) em diagnóstico da COVID-19 com foco na rotina de exames de sangue; [5] que estuda estratégias de salvamento de embarcações náuticas; e [6] que utiliza aprendizado profundo no monitoramento de subestações de energia elétrica. Em particular, técnicas de aprendizado profundo estão sendo cada vez mais utilizadas, fazendo uso da grande capacidade de processamento disponível atualmente.

Pesquisadores e cientistas de todo o mundo têm estudado com o objetivo de minimizar as perdas humanas causadas pelos efeitos do coronavírus no corpo [7], [8]. Na área médica, conjuntos de dados de imagens de retina, raio-x de tórax e ressonância magnética apresentam qualidade suficiente para

Essa pesquisa foi amparada financeiramente pela Finep (Financiadora de Estudos e Projetos) e desenvolvida em parceria nas empresas VMI e FITec.

fornecer uma taxa de acerto relevante ao utilizar técnicas de aprendizado profundo [9], [10]. Os resultados fornecidos pela aplicação de aprendizado profundo podem ser utilizados por especialistas como ferramenta auxiliar para o diagnóstico.

Modelos de aprendizado profundo vêm alcançando resultados extremamente satisfatórios em aplicações de alto desempenho de visão computacional [11] e no reconhecimento de fala [12]. A eficiência das Redes Neurais Convolucionais (CNNs) no processo de reconhecimento foi em grande medida responsável por promover um interesse mundial em aplicações de aprendizado profundo. Atualmente muitos trabalhos empregam abordagens baseadas em CNN para as mais variadas aplicações com imagens de raio-x [13]–[15].

Esse estudo apresenta uma solução que faz uso do aprendizado profundo para suportar o diagnóstico da COVID-19 por meio de imagens de raio-x torácicas. As principais contribuições da pesquisa resumidamente são:

- Seleção criteriosa de *datasets* de qualidade.
- Segmentação pulmonar precisa em busca da diminuição do viés do modelo.
- Comparação entre diferentes modelos de classificação.
- Explicabilidade do modelo.

A Figura 1 apresenta o passo a passo do sistema proposto, ilustrando as várias etapas realizadas durante a inferência de uma imagem de RX.

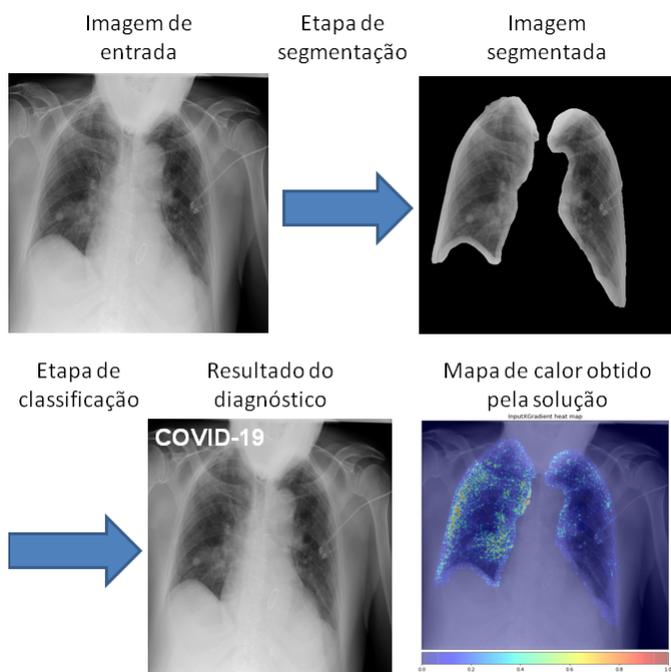


Figura 1. Sistema proposto.

Esse artigo é organizado da seguinte forma: na seção II foram analisados os principais trabalhos relacionados com essa pesquisa. Na seção III descreve os materiais empregados e a metodologia da solução proposta. Já a seção IV apresenta os resultados obtidos pelo estudo. Por fim, a seção V mostra as conclusões, limitações identificadas e trabalhos futuros.

## II. TRABALHOS RELACIONADOS

Essa pesquisa se concentra no suporte ao diagnóstico da COVID-19 por meio de imagens de radiografia, utilizando algoritmos de aprendizado de máquina. Alguns dos trabalhos a seguir podem ser considerados promissores, porém muitos apresentam resultados questionáveis.

Como o número de trabalhos relacionados é grande, foram buscadas pesquisas similares que obtiveram um número considerável de citações, entre aquelas publicadas no ano de 2020. Além disso, também foram considerados trabalhos mais recentes, publicados em 2021. A revisão bibliográfica foi realizada nas principais fontes de trabalhos acadêmicos, como a Scielo, Google Acadêmico e ERIC.

No trabalho [16] os autores defendem que embora as imagens de Tomografia Computadorizada (TC) sejam eficazes para o diagnóstico de COVID-19, as imagens de raio-x são mais acessíveis e de menor custo para aquisição. As técnicas de aprendizado profundo apresentadas se mostraram eficazes o suficiente para solucionar problemas dessa natureza. Ainda no estudo, foi proposto um novo modelo de detecção de anomalias a fim de realizar uma triagem rápida e confiável da COVID-19. Para aumentar a quantidade de dados, foram utilizadas 1.431 imagens de um *dataset* público de raio-x, que possui outras doenças além da COVID-19. Além disso, 100 imagens foram coletadas por meio de equipamentos próprios, as quais 70 eram de pacientes diagnosticados com a doença. Os resultados alcançaram 96% de sensibilidade e 70,65% de especificidade, ao ser avaliado em 10% das 1.531 imagens, as quais fizeram parte do conjunto de teste.

O trabalho de Hemdan *et al.* [17] teve como objetivo apresentar uma solução de aprendizado profundo para auxiliar os radiologistas a diagnosticar COVID-19 em imagens de raio-x. O estudo abordou sete diferentes arquiteturas de CNNs (VGG19, DenseNet121, ResNetV2, InceptionV3, Inception-ResNetV2, Xception, e MobileNetV2) treinadas em *datasets* públicos. Os modelos VGG19 e DenseNet121 apresentaram desempenhos significativos e semelhantes de classificação. Estes dois obtiveram um *F1-score* de 89% para imagens normais e 91% para imagens com suspeita de COVID-19.

O artigo [18] relata um estudo comparativo do uso de modelos de aprendizado profundo que utilizou as redes: VGG16, VGG19, DenseNet201, Inception-ResNet-V2, Inception-V3, Resnet50 e MobileNet-V2 para lidar com a detecção e classificação de COVID-19. Os experimentos utilizaram 6.087 imagens de raio-x do tórax e Tomografia Computadorizada. Dentre essas, 2.780 imagens eram de pneumonia bacteriana, 1.493 de outras doenças, 231 de COVID-19 e 1.583 saudáveis. Os resultados demonstraram que os modelos Inception-ResNet-V2 e Densnet201 obtiveram melhor desempenho comparado aos outros modelos utilizados no trabalho. Sendo assim, o modelo Inception-ResNet-V2 apresentou 92,18% de precisão e o Densnet201 88,09% de precisão.

Em [19] foi implementado um sistema de detecção como alternativa para diagnóstico rápido, com o objetivo de evitar a propagação da COVID-19 entre as pessoas. Para o desen-

volvimento da solução, cinco modelos baseados em CNNs pré-treinadas foram empregados, sendo eles: ResNet50, ResNet101, ResNet152, InceptionV3 e Inception-ResNetV2. Ao considerar os resultados obtidos, os autores constataram que o modelo ResNet50 pré-treinado apresentou melhor desempenho de classificação entre os outros quatro modelos empregados. Esse alcançou a acurácia com média de 98% de acerto.

Os autores do artigo [20] defendem que os conjuntos de dados de COVID-19 disponíveis publicamente são inadequados para treinar redes neurais profundas. Portanto, propuseram um novo conceito denominado Aprendizado por Transferência de Extensão de Domínio (DETL). A DETL foi empregada com uma CNN pré-treinada em um grande conjunto de dados de raio-x de tórax e foi ajustada para classificar quatro classes: normal, pneumonia, outra doença e COVID-19. Os resultados iniciais foram considerados promissores pelos autores, com possibilidade de replicação em conjuntos de dados maiores e mais diversos. A acurácia geral foi de 90,13%.

Em [21] foi mostrado que utilizar imagens de radiografia é uma das maneiras mais rápidas de diagnosticar pacientes com COVID-19. Por isso, inspirados em trabalhos anteriores, os autores abordaram a aplicação de modelos de aprendizado profundo para classificar pacientes COVID-19 a partir de imagens de radiografia de tórax. Primeiramente, advindo de fontes públicas, foi preparado um conjunto de dados de 5.000 imagens de raio-x. Quatro modelos de CNN: ResNet18, ResNet50, SqueezeNet e DenseNet-121, foram treinados com 2.000 imagens. Os autores utilizaram as 3.000 imagens restantes para avaliação, e grande parte das redes atingiu uma taxa de sensibilidade de 98%, embora tenham tido uma taxa de especificidade em torno de 90%. Além disso, para cada imagem de inferência, é gerado um mapa de calor com regiões pulmonares potencialmente infectadas pelo COVID-19.

No trabalho [22] foi proposta a CoroNet, uma rede neural profunda para detecção e diagnóstico de COVID-19 a partir de imagens de raio-x de tórax. O método é baseado na arquitetura Xception pré-treinada no banco de dados ImageNet, imagens de pacientes com COVID-19 e outras imagens de raio-x de pneumonia de dois outros conjuntos de dados públicos. Os resultados experimentais mostraram que o modelo proposto alcançou uma acurácia geral de 89,6%.

Na pesquisa [23] foram usadas as seguintes redes profundas pré-treinadas para o diagnóstico de COVID-19 em imagens de raio-x: ResNet18, ResNet50, ResNet101, VGG16 e VGG19. Com as características extraídas, foi utilizado o classificador *Support Vector Machines* (SVM). Para esse estudo foi utilizado um conjunto de dados com 180 imagens de COVID-19 e 200 imagens pacientes normais (saudáveis) de raio-x de tórax. As características extraídas do modelo ResNet50 e do classificador SVM com a função Linear *kernel* produziram 94,7% de acerto, o maior entre todos os resultados obtidos.

Em [24] os autores buscaram *datasets* públicos de imagens de raio-x de pacientes com COVID-19 e saudáveis, objetivando desenvolver uma solução automatizada utilizando aprendizado profundo para o diagnóstico da doença. Após o uso de técnicas de aumento de dados, foram empregados

modelos de CNN como: Inception V3, Xception e ResNeXt. Em seguida, os resultados de acurácia foram comparados. Para analisar o desempenho dos modelos, 6.432 amostras de radiografias de tórax foram coletadas do repositório Kaggle, das quais 5.467 foram usadas para treinamento e 965 para validação. Como resultado, o modelo Xception ofereceu o maior valor de acurácia (97,97%) para a classificação de imagens de raio-x de tórax.

No trabalho [25] foi feito um estudo sobre os conjuntos de imagens que vêm sendo utilizados nas pesquisas de COVID-19 e publicações, tendo como foco encontrar possíveis vieses. Os autores concluíram que apenas cinco dos 256 conjuntos de dados atenderam um conjunto de critérios para uma avaliação adequada do risco de viés e puderam ser analisados. Notavelmente, quase todos os conjuntos de imagens abordados em 78 artigos publicados em periódicos revisados por pares não estão entre esses cinco conjuntos de dados, levando a modelos com alto risco de viés, levantando preocupações sobre a adequação de tais modelos para uso clínico.

Em [25] foi feita uma varredura na internet sobre conjuntos de imagens de raio-x. Após a pesquisa, os autores elegeram cinco *datasets* para um estudo mais aprofundado: Brixia, Cancer Image Archive, ML Hannover, HM Hospitals e BIMCV-COVID19. Neste mesmo trabalho foi discutido sobre o problema dos *datasets* possuírem poucas imagens, além de mencionarem que usar *datasets* de diferentes fontes no conjunto de treinamento pode causar enviesamento do modelo. Além disso, outro ponto colocado no artigo [25] é o fato de alguns *datasets* apresentarem enviesamento em relação a idade e gênero dos pacientes, uma vez que podem apresentar diferentes características clínicas.

Ainda em relação aos vieses que podem ser observados nos trabalhos relacionados, outro problema é quando existem dispositivos médicos presentes na imagem. Por exemplo, para o conjunto de dados NIH ChestXray14, uma avaliação crítica mostrou que na classe do pneumotórax, 80% dos casos positivos tem drenos torácicos. Nestes exemplos, muitas vezes não havia outras características de pneumotórax [26].

O diferencial do presente trabalho em relação aos que foram citados foi o cuidado na seleção de conjuntos de dados de qualidade. A principal preocupação foi obter modelos com resultados não somente assertivos, mas também confiáveis. Além disso, a busca focou em obter modelos com boa generalização para o problema e com baixo risco de viés. Outro ponto ressaltado é a adição da etapa de segmentação do pulmão antes da tarefa de classificação, garantindo que o modelo não considere áreas da imagem que não englobem o pulmão, o que poderia implicar tendências equivocadas na tomada de decisão.

Por fim, além do resultado da classificação da imagem com COVID-19 ou não COVID-19, a solução proposta também proporciona uma interpretação do modelo, a qual pode auxiliar diretamente o médico na tomada de decisão do diagnóstico. Por mais complicado que seja tentar interpretar modelos de aprendizado profundo, julgou-se que isso é essencial para ajudar o médico na decisão final e, conseqüentemente, obter um diagnóstico mais preciso e confiável.

### III. MATERIAL E MÉTODOS

A presente pesquisa desenvolvida segue uma sequência de etapas metodológicas que estão descritas logo abaixo:

- 1) Aquisição dos *datasets*.
- 2) Pré-processamento das imagens.
- 3) Treinamento e validação da rede U-Net.
- 4) Treinamento e validação das redes para classificação.
- 5) Explicabilidade.

#### A. Descrição dos *datasets*

Com base no estudo [25], foram selecionados cinco *datasets* públicos para serem utilizados nesse trabalho, sendo dois para segmentação e três para classificação. São eles:

- Montgomery [27]
- Shenzhen [27]
- BIMCV [28]
- BrixIA [29]
- NIH [30]

O conjunto de imagens Montgomery foi coletado em colaboração com o Departamento de Saúde e Serviços Humanos, em Maryland, EUA. O conjunto contém 138 radiografias frontais do tórax, das quais 80 são casos normais e 58 são casos com manifestações de tuberculose. As radiografias foram capturadas com uma máquina de raio-x estacionária, em que as imagens salvas possuem o formato *Portable Network Graphics* (PNG), que contam com um nível de cinza de 12 bits. As imagens também podem ser disponibilizadas em formato DICOM mediante solicitação. As dimensões das imagens de raio-x são  $4.020 \times 4.892$  ou  $4.892 \times 4.020$  pixels.

O conjunto de dados de Shenzhen foi coletado em colaboração com o Hospital municipal da cidade, Guangdong Medical College, em Shenzhen, na China. As radiografias de tórax são de ambulatórios e foram capturadas como parte da rotina diária do hospital em um período de um mês, sendo a maior parte em setembro de 2012. O conjunto contém 662 radiografias frontais de tórax, sendo 326 casos normais e 336 casos com manifestações de tuberculose, incluindo radiografias pediátricas. Os raios-x estão no formato PNG. As dimensões de cada imagem podem variar, mas todas elas possuem aproximadamente  $3.000 \times 3.000$  pixels.

Os dois conjuntos de imagens foram rotulados por diversos médicos que marcaram manualmente o que são os pulmões em cada uma das imagens. Dessa forma, para cada imagem de raio-x, existe uma imagem binária correspondente, representando os pulmões segmentados. A Figura 2 ilustra exemplos de imagens de raio-x e suas respectivas máscaras contidas nos conjuntos de dados.

O conjunto de dados BIMCV-COVID19 possui uma grande quantidade de imagens de raio-x de tórax CXR (CR, DX) e imagens de tomografia computadorizada (CT) de pacientes COVID-19. As imagens são armazenadas em alta resolução e os arquivos são localizados com rótulos anatômicos no formato *Medical Imaging Data Structure* (MIDS). Além disso, são fornecidas informações detalhadas, incluindo dados demográficos do paciente, tipo de projeção e parâmetros de aquisição para

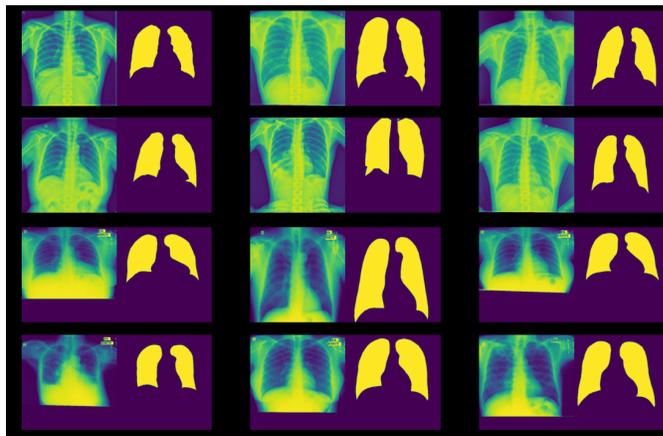


Figura 2. Imagens de raio-x e suas respectivas máscaras.

o estudo da imagem, entre outros. A quantidade de imagens disponível no banco de dados inclui 7.377 de CR, 9.463 de DX e 6.687 de CT.

O BrixIA COVID-19 é um grande conjunto de dados contendo imagens de CXR que correspondem a imagens obtidas no hospital ASST Spedali Civili di Brescia. Os dados coletados foram utilizados para triagem e monitoramento de pacientes em unidades de terapia intensiva e sub-intensiva. A coleta ocorreu durante um mês (entre quatro de março e quatro de abril de 2020), período que se teve um pico pandêmico da COVID-19. O conjunto de imagens inclui 4.707 imagens CXR de indivíduos diagnosticado com COVID-19, adquiridas nas modalidades CR e DX.

O conjunto de dados de raio-x de tórax do NIH é composto por 112.120 imagens de radiografias rotuladas para diferentes doenças em 30.805 pacientes únicos. Para criar esses rótulos, os autores usaram técnicas de processamento de linguagem natural para rotular as classificações das doenças dos relatórios radiológicos. Foi estimado que os rótulos tenham uma precisão além de 90%, fazendo com que sejam adequados para o uso em pesquisas com algoritmos de aprendizado de máquina.

#### B. Pré-processamento

Para o desenvolvimento do estudo foi empregada a linguagem de programação Python 3.6, no ambiente de desenvolvimento Colab, disponibilizado pela Google. O Colab possui uma GPU Tesla T4 com 15.109MB de memória, fazendo seu uso pertinente para o treinamento e validação de algoritmos baseados em aprendizado profundo. Além disso, conta com uma interface na Web, fazendo com que não seja necessária a instalação de ambientes ou programas localmente [31].

Primeiramente foram importadas as principais bibliotecas para se trabalhar com aprendizado profundo e visualização de dados, como a PyTorch, Numpy, OpenCV, Keras, Matplotlib entre outras. Em seguida, foi feito o download das imagens de radiografia torácica com suas respectivas máscaras (para os *datasets* Montgomery e Shenzhen), anotadas pelos médicos radiologistas.

Para a etapa de segmentação, foi feito o download de 800 imagens pertencentes aos conjuntos de dados Montgomery e Shenzhen, porém, apenas 704 máscaras estavam disponíveis. Sendo assim, optou-se por descartar imagens que não possuísem uma máscara correspondente, restando 704 imagens para divisão em treinamento e teste. Em seguida, as imagens foram divididas aleatoriamente com uma proporção de aproximadamente 90:10, sendo 633 imagens para o conjunto de treinamento e 71 imagens para o conjunto de teste.

Para a etapa de classificação, o problema é abordado de forma binária: COVID-19 e não COVID-19. O *dataset* BIMCV possui duas classes, então cada classe foi para um diretório separado. O *dataset* BrixIA conta apenas com imagens da classe COVID-19, logo, as imagens foram organizadas junto ao diretório da classe COVID-19. Por fim, o *dataset* NIH possui imagens de pacientes saudáveis e de várias outras doenças, sendo assim, fizeram parte da classe não COVID-19. A justificativa para o uso de imagens com outras doenças na classe de não COVID-19 está na possibilidade de dar mais robustez à solução, de forma que o modelo seja mais assertivo em pacientes que realmente não possuam a enfermidade.

Com isso, foram totalizadas 7.875 imagens da classe COVID-19 e 16.841 imagens da classe não COVID-19. Para uma comparação justa entre as diferentes arquiteturas de CNN, utilizou a mesma divisão aleatória dos dados nos experimentos realizados. A divisão foi feita da seguinte forma: para a classe COVID-19, 6.300 para o treinamento e 1.575 para o teste, enquanto para a classe não COVID-19, 13.472 para o treinamento e 3.369 para o teste.

Conforme é recorrente em pesquisas na área médica, os dados estão desbalanceados. Existem mais imagens de raio-x disponíveis de pacientes saudáveis e de outras doenças que imagens de pacientes com COVID-19. Por isso, a técnica do aumento de dados foi aplicada nas imagens [32]–[34]. A operação empregada foi a inversão horizontal para a classe minoritária. Além disso, todas as imagens foram redimensionadas para o tamanho de  $256 \times 256$ . Em seguida, o filtro *Contrast Limited AHE* (CLAHE) [35] foi aplicado, com o objetivo de melhorar o contraste das imagens para a etapa de segmentação.

### C. Segmentação do Pulmão

Realizada a divisão das imagens em um conjunto de treinamento e outro de teste, o método U-Net foi implementado com o auxílio da biblioteca Keras [36], seguindo sua arquitetura conforme descrita no artigo original dos criadores da abordagem. Para o conjunto de treinamento, após aplicação do aumento de dados, foram totalizadas 2.536 imagens. Para a validação e para o teste, 140 e 280 imagens, respectivamente. O otimizador utilizado no treinamento foi o **Adam(lr=1e-4)**, com 50 épocas e tamanho de lote 4. Com isso, é possível iniciar o treinamento do algoritmo e, depois de treinado, fazer a inferência com as imagens de teste no modelo.

Para a análise quantitativa dos resultados foi empregada a *Intersection over Union* (IoU), que é uma métrica de avaliação bastante popular e usada nas competições de detecção e segmentação de objetos [37], além de ser amplamente utilizada

como um parâmetro durante o treinamento de CNNs, assim como feito no treinamento da U-Net nesse estudo. Seu cálculo é bastante simples, uma vez que é comparada à similaridade entre duas áreas, a predita pelo modelo e a real, que foi previamente anotada por um especialista.

Além da segmentação pulmonar realizada pelo método U-Net, são aplicados filtros [38] para melhorar a segmentação. Como é importante que nenhuma parte do pulmão seja excluída no processo de segmentação, primeiro é empregada a operação de abertura morfológica (erosão seguida de dilatação) com um kernel de tamanho  $5 \times 5$  com o objetivo de suavizar contornos da imagem, quebrar istmos estreitos e eliminar protuberâncias finas. Posteriormente, empregou-se somente a dilatação com um filtro  $5 \times 5$  para aumentar a área pulmonar e a imagem resultante não perder nenhuma área que corresponde ao pulmão.

### D. Classificação de COVID-19

Após a segmentação dos pulmões utilizando o modelo U-Net treinado anteriormente, as imagens contidas nos conjuntos de dados BIMCV, BrixIA e NIH foram rotuladas como pertencentes às classes COVID-19 ou não COVID-19. Gerando assim um conjunto de dados com duas pastas, separadas pela classe COVID-19 e outra não COVID-19, para posteriormente realizar a etapa de classificação. Tal etapa foi realizada com o objetivo de treinar o modelo de classificação com as imagens já segmentadas, padronizando a forma de treinamento e inferência.

Imagens de pacientes com outras enfermidades como tuberculose, pneumonia entre outras, foram utilizadas como pertencentes à classe não COVID-19, visto que o objetivo da solução é uma classificação binária e outras doenças podem ter padrões parecidos com os da COVID-19. Dessa forma, acreditou-se que o modelo estaria mais robusto para distinguir pacientes que não possuísem COVID-19 mas que apresentassem padrões similares aos de COVID-19.

Com o *dataset* preparado, foi utilizada a biblioteca PyTorch [39] para os experimentos. Foram selecionadas algumas arquiteturas de CNN citadas em [40] e suas respectivas quantidades de camadas convolucionais, de modo que fosse comparado o treinamento e os testes de cada uma delas. Além disso, alguns dos hiper parâmetros empregados para o treinamento das redes foram:

- *Early stopping*: 6
- *Batch size*: 16
- *Optimizer*: Adam (LR=5e-5)
- *Image size*: [1, 256, 256]

Dessa forma, foi possível realizar o treinamento de cada uma das redes, calculando o tempo por época, o tamanho do modelo gerado armazenado em disco, a quantidade de épocas gastas no momento que o critério de parada foi atingido, além da acurácia do conjunto de teste. Como se tomou o cuidado de utilizar um conjunto de dados balanceado, a métrica de acurácia foi o suficiente para avaliar a capacidade do modelo e comparar os resultados com as demais redes.

### E. Explicabilidade

Para explicar as decisões tomadas pelos modelos se utilizou a Captum [41], uma biblioteca para interpretar modelos utilizando o PyTorch. A Captum possui implementações genéricas de uma série de algoritmos baseados em gradiente e perturbação, também conhecidos como algoritmos de importância de recurso, neurônio e camada, bem como um conjunto de métricas de avaliação para esses algoritmos. Além disso, a biblioteca pode ser usada para modelos de classificação ou não, incluindo modelos estruturados em gráficos construídos em redes neurais.

Segundo os autores da biblioteca [41], as três principais características da Captum são a multimodalidade, extensibilidade e a facilidade de uso. A multimodalidade suporta diferentes tipos de entradas, como imagem, texto, áudio ou vídeo. A extensibilidade permite adicionar novos algoritmos e recursos. Além disso, a biblioteca também foi projetada de modo a facilitar a compreensão e sua utilização.

Por fim, a Captum também possui uma ferramenta de visualização interativa chamada Captum Insights, que foi desenvolvida sobre a biblioteca Captum, permitindo a depuração e visualização de modelos com base em amostra usando métricas de importância de recursos.

## IV. RESULTADOS E DISCUSSÃO

A Figura 3 apresenta os gráficos gerados pela etapa de segmentação que mostram a perda do treinamento e da validação no gráfico à esquerda. À direita, é ilustrado o gráfico da acurácia obtida do treinamento e da validação do modelo, para as 50 épocas executadas com a U-Net durante a etapa de segmentação.

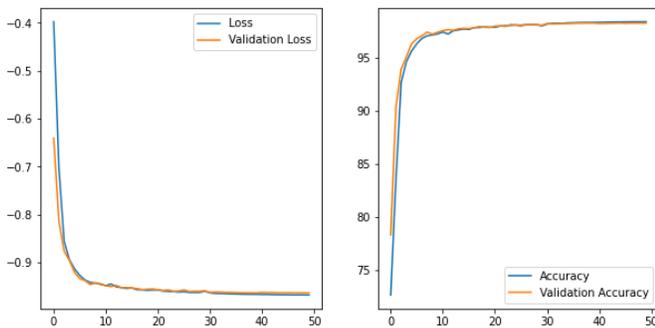


Figura 3. Treinamento da etapa de segmentação.

Como pode ser observado na Figura 3 o modelo convergiu muito rapidamente (em menos de 10 épocas). Um dos motivos que provavelmente ocasionou essa rápida convergência pode ter sido o fato de que a natureza do problema ser considerada fácil para a robustez e precisão do método U-Net. As imagens de raio-x pulmonares possuem características que diferenciam bem os pulmões do restante da imagem, implicando diretamente nos resultados alcançados.

A Figura 4 apresenta exemplos de imagens contidas no conjunto de teste utilizado em um dos experimentos para

verificar se realmente o modelo estava conseguindo segmentar os pulmões corretamente. Como anteriormente colocado, para a etapa de segmentação os *datasets* utilizados foram o Montgomery e o Shenzhen.

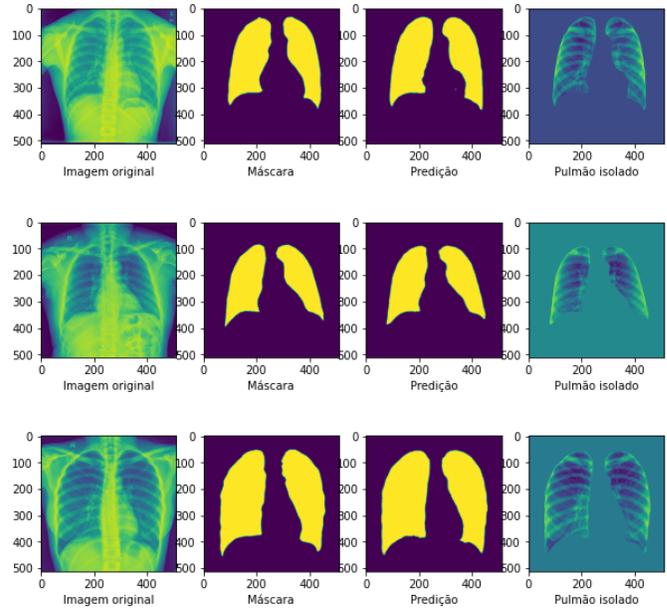


Figura 4. Imagens de raio-x segmentadas pelo método U-Net.

Na primeira coluna, têm-se as imagens originais que foram utilizadas para inferência no modelo treinado. Já a segunda coluna ilustra as máscaras anotadas pelos médicos, referentes a cada imagem original apresentada na coluna anterior. Na terceira coluna tem-se a máscara prevista pelo modelo treinado e na quarta coluna é mostrada a imagem contendo somente o pulmão, após segmentação realizada pelo método U-Net.

O *score* de teste obtido foi de 93,33%, com o tempo médio de 01:14 minutos para cada época do treinamento. O tempo médio de inferência gasto pelo modelo para cada imagem é de aproximadamente 112 milissegundos e o tamanho do arquivo gerado com o modelo treinado é de 89MB.

Apesar das imagens utilizadas apresentarem uma natureza simples para a tarefa de segmentação utilizando modelos poderosos, o método U-Net mostrou-se eficiente para extração dos pulmões em imagens de raio-x.

A Tabela I apresenta os resultados obtidos nos experimentos realizados nas imagens já com o pulmão segmentado, para a classificação da imagem em COVID-19 ou não COVID-19.

Como observado na Tabela I, de modo geral os modelos obtiveram resultados próximos em termos de acurácia. Ao avaliar os erros cometidos por cada um dos modelos foi possível perceber que os modelos não estavam errando somente para uma classe em específica, fazendo com que não fosse necessário calcular outras métricas de avaliação, tais como precisão, revocação e pontuação F1.

Para a seleção dos melhores modelos a serem levados para produção, optou-se por fazer um *ensemble* dos modelos destacados em negrito na Tabela I. O principal critério

Tabela I  
RESULTADOS DOS EXPERIMENTOS REALIZADOS COM DIFERENTES  
ARQUITETURAS DE CNN.

Modelo	Tempo/ época	Tamanho (MB)	Épocas treinadas	Acc
AlexNet	01:19	652	12	0.89
<b>VGG-11</b>	<b>03:12</b>	<b>1400</b>	<b>11</b>	<b>0.91</b>
VGG-13	04:28	1400	10	0.91
VGG-16	05:08	1500	9	0.89
VGG-19	05:48	1600	9	0.89
VGG-19_BN	06:24	1600	12	0.91
ResNet-18	01:35	128	8	0.88
ResNet-34	02:07	244	5	0.89
ResNet-50	03:39	269	9	0.90
ResNet-101	05:44	487	14	0.90
<b>ResNet-152</b>	<b>06:42</b>	<b>667</b>	<b>18</b>	<b>0.91</b>
Densenet-121	04:17	80	10	0.90
<b>Densenet-161</b>	<b>06:45</b>	<b>304</b>	<b>12</b>	<b>0.92</b>
Densenet-169	05:25	144	11	0.91
<b>Densenet-201</b>	<b>06:44</b>	<b>209</b>	<b>10</b>	<b>0.92</b>
Inception_v3	04:34	263	9	0.89
<b>GoogleNet</b>	<b>01:48</b>	<b>114</b>	<b>15</b>	<b>0.92</b>
SqueezeNet_1_1	01:16	8	44	0.85
ShuffleNet_V2_x2_0	01:54	62	4	0.83
MobileNet_V2	01:54	26	14	0.86
MobileNet_V3_large	01:59	48	4	0.84
ResNeXt-50-32x4d	06:27	263	13	0.89
ResNeXt-101-32x8d	14:18	994	12	0.90
Wide_ResNet-50-2	05:15	765	10	0.88
Wide_ResNet-101-2	08:34	1400	13	0.90

utilizado foi a acurácia no conjunto de teste, no qual foram selecionados três modelos que obtiveram um acerto de 92%. Em seguida, como houve um empate entre quatro modelos que acertaram 91% das imagens de teste, optou-se por diversificar a arquitetura da rede, uma vez que três destes quatro modelos eram da arquitetura de CNN VGG.

Dessa forma, quando a solução recebe uma nova imagem a mesma passa por todos os cinco classificadores. Para resultado do diagnóstico, é levado em consideração o número de votos da classe mais repetida entre os cinco modelos classificadores, a qual pode ser COVID-19 ou não COVID-19.

Além do diagnóstico rotulando a imagem de raio-x como COVID-19 ou não COVID-19, a solução também apresenta ao profissional da área médica um mapa de calor que pode ser obtido por meio da biblioteca Captum. A Figura 5 apresenta alguns exemplos de mapa de calor obtidos pela biblioteca.

Como apresentado na Figura 5, o resultado da solução proposta nesse trabalho apresenta um mapa de calor com as principais áreas consideradas pelo modelo de classificação no momento da definição da classe da imagem. Dessa forma, as áreas que possuem uma tonalidade mais próxima da cor vermelha são consideradas as lesões causadas pelo vírus responsável pela doença, regiões essas que devem chamar a atenção do médico para uma análise mais criteriosa na imagem original. Como são cinco classificadores, tem-se um mapa de calor para cada. Sendo assim, é calculada a média entre os mapas de calor e consequentemente obtém-se o mapa de calor resultante (Figura 5) para ser apresentado ao médico.

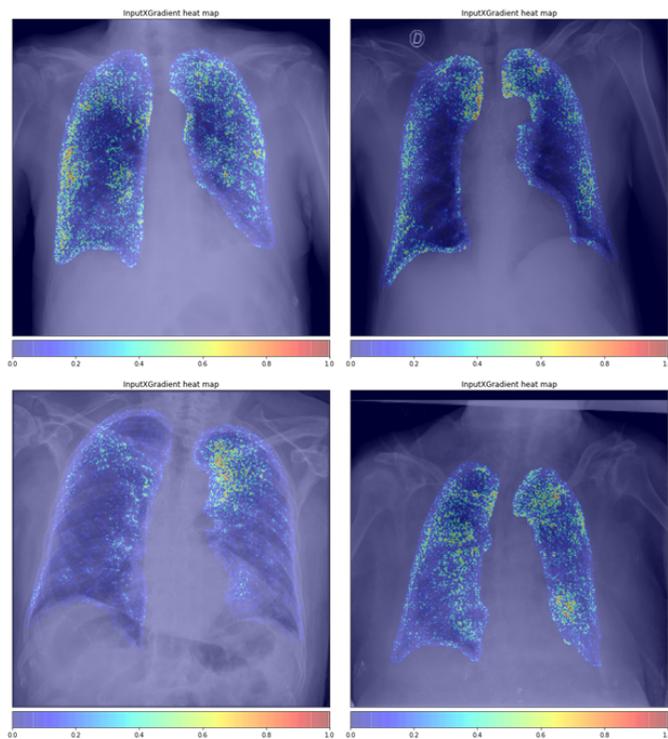


Figura 5. Exemplos de mapa de calor gerados pela biblioteca Captum.

## V. CONSIDERAÇÕES FINAIS

Nesse artigo foi apresentada uma solução para suporte ao diagnóstico de COVID-19 por meio de imagens de raio-x do tórax. Com o objetivo de se obter modelos o menos enviesados possível, tomou-se o cuidado para a seleção de *datasets* de qualidade. Para a aplicação, primeiro os pulmões são segmentados pelo método U-Net para que, em seguida, a imagem seja classificada por um conjunto de modelos treinados para uma classificação binária. Por fim, os resultados são explicados de forma a auxiliar o médico em sua tomada de decisão.

Para o conjunto de teste foi obtido um valor aproximado de 92% de acurácia, fazendo com que a atual solução proposta seja uma candidata promissora para ser incorporada em equipamentos médicos. No entanto, outras imagens com uma maior diversidade devem ser testadas para validar o sistema, antes de afirmar que a inteligência artificial realmente pode auxiliar os médicos na tarefa da classificação de COVID-19 por meio de imagens de raio-x.

Conclui-se que a pesquisa é oportuna e necessária em nível mundial. Se trabalhos nessa mesma linha forem bem-sucedidos, poderão atuar no auxílio de profissionais da saúde na luta contra a pandemia. Como trabalhos futuros, serão realizados mais testes em *datasets* heterogêneos e a solução deverá ser validada com radiologistas especializados com um maior número de imagens.

## REFERÊNCIAS

- [1] A. H. Davarpanah, A. Mahdavi, A. Sabri, T. F. Langroudi, S. Kahkouee, S. Haseli, M. A. Kazemi, P. Mehran, A. Mahdavi, F. Falahati *et al.*, “Novel screening and triage strategy in iran during deadly coronavirus disease 2019 (covid-19) epidemic: value of humanitarian teleconsultation service,” *Journal of the American College of Radiology*, vol. 17, no. 6, pp. 734–738, 2020.
- [2] T. Zitek, “The appropriate use of testing for covid-19,” *Western Journal of Emergency Medicine*, vol. 21, no. 3, p. 470, 2020.
- [3] B. A. S. Oliveira, M. S. Nametala, B. R. Castoril, M. D. Catalano, and C. A. L. Nametala, “Avaliação de uma rede neural artificial como estimador temporal pluviométrico no sistema de abastecimento cantareira,” *Revista de Informática Aplicada*, vol. 14, no. 1, 2018.
- [4] M. A. Alves, G. Z. de Castro, B. A. S. Oliveira, L. A. Ferreira, J. A. Ramírez, R. Silva, and F. G. Guimarães, “Explaining machine learning based diagnosis of covid-19 from routine blood tests with decision trees and criteria graphs,” *Computers in Biology and Medicine*, p. 104335, 2021.
- [5] F. H. W. dos Santos, L. A. M. Pessôa, B. A. S. Oliveira, L. M. Vieira, and G. M. M. Cardoso, “Rastreamento de embarcações em imagens satelitais utilizando metodologia multicritério para a priorização em tarefas de busca e salvamento,” *Brazilian Journal of Development*, vol. 6, no. 5, pp. 28 245–28 257, 2020.
- [6] B. A. S. Oliveira, A. P. D. F. Neto, R. M. A. Fernandino, R. F. Carvalho, A. L. Fernandes, and F. G. Guimarães, “Automated monitoring of construction sites of electric power substations using deep learning,” *IEEE Access*, vol. 9, pp. 19 195–19 207, 2021.
- [7] R. Manne and S. Kantheti, “Coronavirus impact on cardiovascular system of body-review,” *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol. 8, no. 11, pp. 276–280, 2020.
- [8] A. Budhrani, “A review: Coronavirus, its types, and impact of covid-19 on global wealth,” *International Journal of Research in Pharmaceutical Sciences*, pp. 455–461, 2020.
- [9] J. Olczak, N. Fahlberg, A. Maki, A. S. Razavian, A. Jilert, A. Stark, O. Sköldenberg, and M. Gordon, “Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures?” *Acta orthopaedica*, vol. 88, no. 6, pp. 581–586, 2017.
- [10] A. Majkowska, S. Mittal, D. F. Steiner, J. J. Reicher, S. M. McKinney, G. E. Duggan, K. Eswaran, P.-H. Cameron Chen, Y. Liu, S. R. Kalidindi *et al.*, “Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation,” *Radiology*, vol. 294, no. 2, pp. 421–431, 2020.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [12] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649.
- [13] M. Murata, Y. Arijji, Y. Ohashi, T. Kawai, M. Fukuda, T. Funakoshi, Y. Kise, M. Nozawa, A. Katsumata, H. Fujita *et al.*, “Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography,” *Oral radiology*, vol. 35, no. 3, pp. 301–307, 2019.
- [14] Y. Sim, M. J. Chung, E. Kotter, S. Yune, M. Kim, S. Do, K. Han, H. Kim, S. Yang, D.-J. Lee *et al.*, “Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs,” *Radiology*, vol. 294, no. 1, pp. 199–209, 2020.
- [15] K. Üreten, H. Erbay, and H. H. Maras, “Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network,” *Clinical rheumatology*, vol. 39, no. 4, pp. 969–974, 2020.
- [16] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, “Covid-19 screening on chest x-ray images using deep learning based anomaly detection,” *arXiv preprint arXiv:2003.12338*, vol. 27, 2020.
- [17] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, “Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images,” *arXiv preprint arXiv:2003.11055*, 2020.
- [18] K. El Asnaoui and Y. Chawki, “Using x-ray images and deep learning for automated detection of coronavirus disease,” *Journal of Biomolecular Structure and Dynamics*, pp. 1–12, 2020.
- [19] A. Narin, C. Kaya, and Z. Pamuk, “Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks,” *arXiv preprint arXiv:2003.10849*, 2020.
- [20] S. Basu, S. Mitra, and N. Saha, “Deep learning for screening covid-19 using chest x-ray images,” in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 2521–2527.
- [21] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, “Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning,” *Medical image analysis*, vol. 65, p. 101794, 2020.
- [22] A. I. Khan, J. L. Shah, and M. M. Bhat, “Coronet: A deep neural network for diagnosis of covid-19 from chest x-ray images,” *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105581, 2020.
- [23] A. M. Ismael and A. Şengür, “Deep learning approaches for covid-19 detection based on chest x-ray images,” *Expert Systems with Applications*, vol. 164, p. 114054, 2021.
- [24] R. Jain, M. Gupta, S. Taneja, and D. J. Hemanth, “Deep learning based detection and analysis of covid-19 on chest x-ray images,” *Applied Intelligence*, vol. 51, no. 3, pp. 1690–1700, 2021.
- [25] B. G. Santa Cruz, M. N. Bossa, J. Soelter, and A. D. Husch, “Public covid-19 x-ray datasets and their impact on model bias—a systematic review of a significant problem,” *medRxiv*, 2021.
- [26] L. Oakden-Rayner, “Exploring large-scale public medical image datasets,” *Academic radiology*, vol. 27, no. 1, pp. 106–112, 2020.
- [27] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, “Two public chest x-ray datasets for computer-aided screening of pulmonary diseases,” *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, p. 475, 2014.
- [28] M. De La Iglesia Vayá, J. M. Saborit, J. A. Montell, A. Pertusa, A. Bustos, M. Cazorla, J. Galant, X. Barber, D. Orozco-Blatrán, F. García-García *et al.*, “Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients,” *arXiv preprint arXiv:2006.01174*, 2020.
- [29] A. Signoroni, M. Savardi, S. Benini, N. Adami, R. Leonardi, P. Gibellini, F. Vaccher, M. Ravanelli, A. Borghesi, R. Maroldi *et al.*, “Bs-net: learning covid-19 pneumonia severity on a large chest x-ray dataset,” *Medical Image Analysis*, p. 102046, 2021.
- [30] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.
- [31] E. Bisong, “Google colab,” in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Springer, 2019, pp. 59–64.
- [32] Z. Eaton-Rosen, F. Bragman, S. Ourselin, and M. J. Cardoso, “Improving data augmentation for medical image segmentation,” *MIDL*, 2018.
- [33] A. Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.
- [34] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [35] A. M. Reza, “Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement,” *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, no. 1, pp. 35–44, 2004.
- [36] F. Chollet *et al.*, “Keras,” <https://github.com/fchollet/keras>, 2015.
- [37] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.
- [38] B. Kurt, V. V. Nabyev, and K. Turhan, “Medical images enhancement by using anisotropic filter and clahe,” in *2012 International symposium on innovations in intelligent systems*. IEEE, 2012, pp. 1–4.
- [39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” *arXiv preprint arXiv:1912.01703*, 2019.
- [40] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.
- [41] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan *et al.*, “Captum: A unified and generic model interpretability library for pytorch,” *arXiv preprint arXiv:2009.07896*, 2020.