

Sistema FIEB



PELO FUTURO DA INOVAÇÃO

CENTRO UNIVERSITÁRIO SENAI CIMATEC
Programa de Pós-Graduação em
Modelagem Computacional e Tecnologia Industrial

ADHVAN NOVAIS FURTADO

**APRENDIZAGEM PROFUNDA PARA O SUPORTE AO
DIAGNÓSTICO DA PNEUMONIA CAUSADA POR COVID-19
EM EXAMES DE RAIOS X E TOMOGRAFIA
COMPUTADORIZADA**

Salvador

2022

ADHVAN NOVAIS FURTADO

**APRENDIZAGEM PROFUNDA PARA O SUPORTE AO
DIAGNÓSTICO DA PNEUMONIA CAUSADA POR COVID-19
EM EXAMES DE RAIOS X E TOMOGRAFIA
COMPUTADORIZADA**

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial do Centro Universitário SENAI CIMATEC como requisito parcial para a obtenção do título de Doutor em Modelagem Computacional e Tecnologia Industrial.

Orientador: Prof. Dr. Erick Giovani Sperandio Nascimento.

Coorientador: Prof. Dr. Roberto José da Silva Badaró.

Salvador

2022

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

F992a Furtado, Adhvan Novais

Aprendizagem profunda para o suporte ao diagnóstico da pneumonia causada por Covid-19 em exames de raio x e tomografia computadorizada / Adhvan Novais Furtado. – Salvador, 2022.

101 f. : il. color.

Orientador: Prof. Dr. Erick Giovani Sperandio Nascimento.

Coorientador: Prof. Dr. Roberto José da Silva Badaró.

Tese (Doutorado em Modelagem Computacional e Tecnologia Industrial) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2022.

Inclui referências.

1. COVID-19. 2. Aprendizagem profunda. 3. TC de tórax. 4. Raio-X de pulmão. I. Centro Universitário SENAI CIMATEC. II. Nascimento, Erick Giovani Sperandio . III. Badaró, Roberto José da Silva. IV. Título.

CDD 616.9

Centro Universitário SENAI CIMATEC

Doutorado em Modelagem Computacional e Tecnologia Industrial

A Banca Examinadora, constituída pelos professores abaixo listados, leu e aprovou a Tese de doutorado, intitulada “**Aprendizagem Profunda para o Suporte ao Diagnóstico da Pneumonia Causada por COVID-19 em Exames de Raio X e Tomografia Computadorizada**”, apresentada no dia 23 de setembro de 2022, como parte dos requisitos necessários para a obtenção do Título de Doutor em Modelagem Computacional e Tecnologia Industrial.

Assinado eletronicamente por:
Erick Giovani Sperandio Nascimento
CPF: ***.666.177-**
Data: 28/09/2022 15:28:47 -03:00

Orientador:

Prof. Dr. Erick Giovani Sperandio Nascimento
SENAI CIMATEC

Assinado eletronicamente por:
Roberto José da Silva Badaró
CPF: ***.929.405-**
Data: 30/09/2022 15:48:39 -03:00

Coorientador:

Prof. Dr. Roberto Jose da Silva Badaró
SENAI CIMATEC

Electronically signed by:
VALTER de Senna
CPF: ***.290.367-**
Date: 9/28/2022 3:42:16 PM -03:00

Membro Interno:

Prof. Dr. Valter de Senna
SENAI CIMATEC

Assinado eletronicamente por:
Bruna Machado
CPF: ***.830.795-**
Data: 28/09/2022 17:27:11 -03:00

Membro Interno:

Profa. Dra. Bruna Aparecida Souza Machado
SENAI CIMATEC

Electronically signed by:
PEDRO MARIO CRUZ E SILVA
CPF: ***.276.774-**
Date: 9/29/2022 12:23:22 PM -03:00

Membro Externo:

Prof. Dr. Pedro Mário Cruz e Silva
NVIDIA

Assinado eletronicamente por:
Luiz Eduardo Fonteles Ritt
CPF: ***.326.605-**
Data: 20/10/2022 06:58:20 -03:00

Membro Externo:

Prof. Dr. Luiz Eduardo Fonteles Ritt
EBMSP

AGRADECIMENTOS

Agradeço à família e amigos pela paciência e insistência.

O treinamento de redes neurais profundas utilizando grande volume de imagens demanda uma infraestrutura de computação de alto desempenho. A realização deste trabalho necessitou de aceleradores de hardware e sistemas de arquivos paralelos de alta velocidade. Agradeço ao Centro de Supercomputação e Inovação Industrial do SENAI CIMATEC pela disponibilização de computadores de alto desempenho. Foram essenciais para o desenvolvimento dos algoritmos aqui descritos. O SENAI CIMATEC, entre abril e agosto de 2020, desenvolveu o projeto intitulado “Processamento Avançado e Inteligência Artificial no Combate à Covid-19”. O projeto foi uma resposta do Centro de Referência em IA do CIMATEC à chamada pública do edital de inovação da indústria com foco em soluções relacionadas à COVID-19, patrocinado pelo SENAI DN e ABDI – Agência Brasileira de Desenvolvimento Industrial. A proposta foi submetida em parceria com a empresa Repsol Sinopec Brasil e contou com recursos para três macro objetivos: disponibilização de acesso a pesquisadores de todo o mundo ao Centro de Supercomputação e Inovação Industrial, desenvolvimento de uma ferramenta de suporte ao diagnóstico da COVID utilizando IA em exames de imagem radiológica e a criação de uma ferramenta de simulação baseada em modelos que permitam prever, para os próximos dias ou semanas, a evolução dos casos da COVID-19 e os impactos econômicos para uma determinada região ou país, a partir de determinadas medidas de mitigação. Pude colaborar na concepção, elaboração e execução do projeto, que contou com a participação de uma equipe com mais de 15 pessoas, atuando com diferentes especialidades e dedicações de tempo. Embora tenha sido um projeto de tiro curto, o enorme esforço do time na coleta de dados, desenvolvimento de algoritmos e testes foi bem-sucedido e serviu de base para os algoritmos desenvolvidos nesta tese. Agradeço a cada um dos participantes. Um agradecimento especial aos bolsistas do centro de competência em IA do SENAI CIMATEC e agora profissionais de referência: Leandro Andrade Barreto e Carlos Purificação.

Agradeço ao Dr. Badaró, coorientador e inspirador de sonhos cada vez mais altos e principalmente ao Dr. Erick Giovani Sperandio Nascimento pela amizade, constante apoio e capacidade, diferenciada, de materializar ideias em realidade.

RESUMO

A disseminação exponencial da COVID-19 no mundo trouxe desafios importantes para os sistemas de saúde pública. Aprendemos que em situações de pandemia, o alto volume de pacientes sobrecarrega a capacidade de atendimento dos serviços de atenção primária à saúde. Ficou claro que nesta pandemia da COVID-19, para controlar a morbimortalidade da doença, é necessário identificar rapidamente o maior número possível de pacientes com suspeita de pneumonia. Exames de imagem têm sido utilizados com sucesso para identificação e confirmação de suspeita de pneumonia associada a COVID-19. Os pacientes diagnosticados com COVID-19 usualmente apresentam situações anormais nas imagens de tórax obtidas por exames de tomografia computadorizada (TC) e de raio-X. Embora os achados radiológicos sejam similares aos encontrados em outras doenças pulmonares, especialistas conseguem identificar padrões de características de vidro fosco e sugerir a possibilidade diagnóstica de COVID-19 pneumonia, mesmo na sua fase inicial, principalmente em situações de epidemias ou pandemias. Neste contexto, este trabalho apresenta a utilização de algoritmos de aprendizagem profunda sobre as imagens associadas ao diagnóstico da COVID-19. Foram desenvolvidos dois sistemas de código aberto, com redes neurais convolucionais, capazes de realizar a identificação de imagens sugestivas da COVID-19 presentes nos exames de raio-X e TC do tórax. Para treinamento das redes neurais foram coletados dados de bases públicas internacionais bem como dados obtidos através de parcerias realizadas com os hospitais Santa Izabel de Salvador, Bahia, hospital das Clínicas, de São Paulo, capital, e hospital Medsenior, de Vitória, Espírito Santo. Para o desenvolvimento do algoritmo de classificação de imagens de raio-X foram usados 44.031 exames durante o treinamento e validação. O modelo obteve uma sensibilidade de 0,85, especificidade de 0,82 e ROC AUC de 0,93 quando testado em um conjunto de 1.158 radiografias do tórax de um hospital de referência. O algoritmo de classificação de tomografias computadorizadas utilizou uma base de 3.000 exames e selecionou de forma inovadora as 16 imagens, por exame, mais representativas para o treinamento. O algoritmo obteve uma sensibilidade de 0,89, especificidade de 0,90, ROC AUC de 0,97 sobre uma base de testes de 414 amostras. Os dois algoritmos se apresentaram como boas opções para avaliação das imagens do tórax de pacientes da COVID-19, possibilitando separar aqueles com alterações sugestivas da doença, o que pode ser

relevante no suporte ao direcionamento de pacientes suspeitos em regiões desassistidas de modelos mais eficientes de diagnóstico. Concluímos que a utilização de algoritmos inteligentes pode auxiliar na identificação de imagens anormais no raio-X e tomografia de tórax, sobretudo na ausência de um especialista para emitir o laudo do exame, permitindo uma triagem rápida daqueles pacientes com suspeita de pneumonia que tem necessidade de atenção imediata no serviço de saúde.

Palavras-chave: COVID-19; Aprendizagem Profunda; TC de tórax; Raio-X de pulmão; Diagnóstico.

ABSTRACT

Deep Learning to Support COVID-19 Pneumonia Diagnosis in X-Ray and CT Scans

The exponential spread of COVID-19 around the world has brought important challenges to public health systems. We learned that in pandemic situations, the high volume of patients overwhelms the care capacity of primary health care services. It became clear that in this COVID-19 pandemic, to control the morbidity and mortality of the disease, it is necessary to quickly identify as many patients as possible with suspected pneumonia. Imaging tests have been successfully used to identify and confirm suspected pneumonia associated with COVID-19. COVID-19 patients usually have abnormal situations on chest images obtained by computed tomography (CT) and X-ray exams. Although the radiological findings are similar to those found in other lung diseases, specialists are able to identify images of ground-glass features and suggest the diagnostic possibility of COVID-19 pneumonia even in its initial phase, especially in pandemic or epidemic situations. In this context, this work presents the use of a deep learning algorithm on the images associated with the diagnosis of COVID-19. Two open source systems were developed, with convolutional neural networks, capable of performing the identification of images suggestive of COVID-19, present in chest's X-ray and CT scans. For the training of neural networks, data were collected from international public databases as well as data obtained through partnerships with Santa Izabel hospitals in Salvador, Bahia, Hospital das Clínicas, in São Paulo, capital, and Medsenior hospital, in Vitória, Espírito Santo. For the development of the X-ray image classification algorithm, 44,031 exams were used during training and validation.

The X-ray model obtained a sensitivity of 0.85, specificity of 0.82 and ROC AUC of 0.93 when tested on a set of 1,158 chest radiographs from a referral hospital. The computed tomography classification algorithm used a base of 3,000 exams and innovatively selected the 16 most representative images per exam for training. The algorithm achieved a sensitivity of 0.89, specificity of 0.90, ROC AUC of 0.97 over a test base of 414 samples. The two algorithms presented themselves as good options for evaluating the chest images of COVID-19 patients, making it possible to separate

those with alterations from those with no abnormalities. This is relevant in supporting the targeting of suspected patients in unassisted regions that lacks more efficient diagnostic models. We conclude that the use of intelligent algorithms can help in the identification of abnormal images in X-ray and chest tomography, especially in the absence of a specialist to report the exam, allowing a quick triage of those patients with suspected pneumonia who need immediate attention in the healthcare service.

Keywords: COVID-19; Deep Learning; Thorax CT-Scan; Pulmonary radiograph; Diagnosis.

LISTA DE FIGURAS

<i>Figura 1. Desenvolvimento de modelo de aprendizagem de máquina para classificação de imagens médicas: (a) Treinamento; (b) Classificação.</i>	<i>26</i>
<i>Figura 2. Camadas de extração de características de algoritmos de aprendizagem profunda.</i>	<i>28</i>
<i>Figura 3: Aprendizagem Profunda: Características são extraídas através de uma série de mapeamentos simples encadeados, cada um descrito por uma camada diferente do modelo.</i>	<i>29</i>
<i>Figura 4. Separação de dados para treinamento, validação e teste</i>	<i>31</i>
<i>Figura 5. Underfitting e Overfitting</i>	<i>32</i>
<i>Figura 6. Validação Cruzada 3-Fold.</i>	<i>33</i>
<i>Figura 7. Exemplo de uma Matriz de Confusão para um modelo classificador de exames anormais/COVID-19.</i>	<i>34</i>
<i>Figura 8. Ilustração de uma curva ROC</i>	<i>37</i>
<i>Figura 9. Curva Precision Recall com 20% de amostras positivas.</i>	<i>38</i>
<i>Figura 10. Exemplo de mapa de calor gerado com a técnica de Grad-CAM para explicabilidade da predição realizada por um modelo de IA para suporte a diagnóstico médico de COVID-19 em um exame de raio X.</i>	<i>39</i>
<i>Figura 11. Radiografias de tórax em um paciente idoso do sexo masculino de Wuhan, China</i>	<i>44</i>
<i>Figura 12. Características típicas de COVID-19, com alta incidência (> 70%) (a) Áreas bilaterais de opacidades em vidro fosco (setas) em uma distribuição periférica. (b) Vasos segmentares e subsegmentares dilatados</i>	<i>45</i>
<i>Figura 13. Radvid-19, um sistema nacional de suporte ao diagnóstico de COVID-19 em exames de imagem</i>	<i>49</i>
<i>Figura 14. Segmentação do pulmão com Nvidia Clara</i>	<i>51</i>

LISTA DE TABELAS

Tabela 1. Dados obtidos a partir da matriz de confusão	35
Tabela 2- Disponibilidades de aparelhos de Raios X e Tomógrafos no Brasil.	42

LISTA DE SIGLAS E ABREVIATURAS

AI – Artificial Intelligence
CIMATEC – Campus Integrado de Manufatura e Tecnologia
CNN – Convolutional Neural Network
CT – Computer Tomography
DL – Deep Learning
FN - Falso negativo
FP - Falso positivo
Grad-CAM - Gradient-weighted Class Activation Mapping
IA – Inteligência Artificial
N - Condição Negativa
OMS – Organização Mundial da Saúde
P - Condição Positiva
PACS - Picture Archiving and Communication System
PR – Precision-Recall
PR AUC –Area under the Precision-Recall Curve
RMSE – Root Mean Square Error
ROC - Receiver Operating Characteristic
ROC AUC – Area under the ROC Curve
RT-PCR - Reverse Transcription Polymerase Chain Reaction
SARS-CoV-2 - Severe Acute Respiratory Syndrome Coronavirus 2
SENAI – Serviço Nacional de Aprendizagem Industrial
TC – Tomografia Computadorizada
VN - Verdadeiros Negativos
VP - Verdadeiros Positivos
WHO – World Health Organization
XAI - Explainable AI

SUMÁRIO

1	INTRODUÇÃO	19
1.1	OBJETIVOS	22
1.1.1	<i>Objetivo Geral</i>	22
1.1.2	<i>Objetivos Específicos</i>	22
1.2	ORGANIZAÇÃO DA TESE	22
2	REVISÃO DA LITERATURA.....	24
2.1	APRENDIZAGEM DE MÁQUINA NA CLASSIFICAÇÃO DE IMAGENS	24
2.1.1	<i>Aprendizagem Profunda</i>	27
2.1.2	<i>Redes Neurais Convolucionais</i>	29
2.1.3	<i>Treinamento de redes neurais, overfitting e underfitting</i>	30
2.1.4	<i>Métricas para avaliação do modelo</i>	33
2.1.5	<i>Explicabilidade</i>	38
2.2	IA NA MEDICINA	39
2.3	SUPORTE AO DIAGNOSTICO DE COVID-19 EM EXAMES DE IMAGEM DE RAIOS X E TC	41
2.3.1	<i>Uso de Deep Learning no suporte ao diagnóstico de COVID em exames de imagem</i> ..	46
	REFERÊNCIAS	53
3	MANUSCRITO 1	59
4	MANUSCRITO 2	81
5	CONCLUSÃO.....	99

1 INTRODUÇÃO

Em 11 de março de 2020, a Organização Mundial de Saúde (OMS) decretou a pandemia da COVID-19. A doença causada pelo vírus SARS-CoV-2 foi identificada inicialmente em dezembro de 2019, na cidade de Wuhan, China, gerando infecções do trato respiratório e se espalhando rapidamente através do contágio entre pessoas. Os sistemas de saúde em todo o mundo foram sobrecarregados (HUANG et al., 2020; ZHU et al., 2020). À medida que a COVID-19 se espalhou, aumentou a pressão sobre médicos e demais profissionais de saúde para fornecer um diagnóstico eficiente, rápido e preciso aos pacientes (HUANG et al., 2020). Até o presente momento, o pico da pandemia ocorreu na segunda semana de janeiro de 2021, quando houve mais de 100.000 mortes semanais e mais de 23 milhões de casos ativos confirmados em todo mundo (“WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data,” [s.d.]). Com o avanço da vacinação e o surgimento de variantes do vírus SARS-CoV-2, a doença alterou seu curso, tornando-se menos letal e causando menores danos ao pulmão (KARIM; KARIM, 2021; TAO et al., 2021). Este trabalho foi desenvolvido no contexto das fases iniciais da pandemia.

A OMS sugere que ferramentas que auxiliem no diagnóstico precoce são essenciais para a prevenção e controle da propagação da COVID-19. Recomenda o uso de testes laboratoriais através da identificação de RNA viral na reação em cadeia da polimerase com transcriptase reversa (RT-PCR, sigla em inglês) e testes rápidos de antígeno em diferentes cenários de transmissão. Os testes devem ser seguidos por uma forte resposta de saúde pública, incluindo o fornecimento de cuidados, isolamento dos pacientes com resultados positivos, rastreamento e quarentena de pessoas que tiveram contato com infectados (WORLD HEALTH ORGANIZATION, 2021). O RT-PCR é um teste confiável e, até o momento, o “padrão ouro” para diagnóstico da COVID-19. Entretanto, ele demanda pessoas treinadas para a realização da coleta, um laboratório especializado para análise e os resultados podem demorar algumas horas ou dias. Existe ainda uma variação importante na proporção de resultados falso-negativos no RT-PCR, muito dependentes do processo de coleta e da carga viral de cada paciente (AREVALO-RODRIGUEZ et al., 2020; WOLOSHIN; PATEL; KESSELHEIM, 2020). É necessária uma busca por métodos adicionais de avaliação e gerenciamento do paciente, em especial como alternativa para regiões

pouco assistidas, onde unidades de saúde não possuem acesso a testes de RT-PCR na disponibilidade devida.

Clinicamente, os pacientes com COVID-19, no período inicial da pandemia, apresentavam febre, tosse, dispneia, dores musculares e pneumonia bilateral nas imagens (AI et al., 2020; YANG et al., 2020). Podiam ser identificadas situações anormais nas imagens do tórax com características próprias da doença. A maioria apresentava opacidades em vidro fosco no estágio inicial e consolidação pulmonar nos estágios mais avançados. Eventualmente observava-se uma morfologia arredondada e uma distribuição pulmonar periférica (CHUNG et al., 2020; HUANG et al., 2020). Exames de radiografia de tórax e TC são os métodos mais comuns para apoiar o diagnóstico de pneumonia em pacientes sintomáticos (PONTONE et al., 2021). Estes exames foram amplamente utilizados como parte da triagem inicial e em situações em que o paciente apresentava fortes sintomas respiratórios (SANDRI et al., 2021). É importante observar que existem recomendações claras da OMS e da Sociedade Americana de Radiologia para o uso de exames de imagem por raio X e TC somente em situações específicas, não sendo seu uso recomendado como único meio de diagnóstico para COVID-19 (AKL et al., 2021; RUBIN et al., 2020; SIMPSON et al., 2020a).

Embora as imagens típicas pudessem ajudar no rastreamento precoce de casos suspeitos, as imagens de diversas pneumonias virais são semelhantes e se sobrepõem a outras doenças pulmonares infecciosas e inflamatórias (AGGARWAL et al., 2022). Portanto, não é trivial para os radiologistas distinguir a pneumonia por COVID-19 de outras pneumonias virais. As imagens dos exames de raio X provaram ser de difícil análise para obtenção de diagnóstico diferencial (SMITH et al., 2020). Os resultados são muitas vezes inconclusivos, uma vez que os achados são muito sutis nas etapas iniciais da COVID-19 e se confundem com sinais de outras doenças respiratórias. Com a progressão da doença no paciente, o achado radiográfico torna-se mais evidente, o que permite o seu uso para auxiliar no diagnóstico e acompanhamento da doença.

Apesar das limitações dos sistemas de imagem, eles podem ser muito úteis, uma vez que, em regiões com pouco acesso a recursos e a médicos especializados, um diagnóstico por RT-PCR pode simplesmente não ser possível. Muitas unidades de saúde possuem equipamentos radiológicos e a radiografia de tórax torna-se uma alternativa acessível, rápida e barata para apoiar a triagem e acompanhamento de

pacientes. Os aparelhos móveis permitem, inclusive, seu uso em pacientes intubados, com dificuldade de locomoção ou em isolamento.

Neste contexto, sistemas de apoio ao diagnóstico por imagem utilizando Inteligência Artificial (IA) tornam-se ferramentas importantes para apoiar a equipe médica no direcionamento de pacientes com suspeita de pneumonia por COVID-19, especialmente em áreas onde nenhum especialista em radiologia está disponível (SHI et al., 2020). Os sistemas inteligentes podem, ainda, reduzir os riscos de falha, acelerar o processo de interpretação dos exames e permitir uma melhor operação dos serviços de radiologia em períodos de extrema demanda, como situações epidêmicas ou pandêmicas.

Contudo, a maioria dos algoritmos de IA propostos na literatura para detectar COVID-19 apresentam falhas metodológicas e embutem vieses que impedem seu uso clínico (ROBERTS et al., 2021). Neste trabalho, utilizamos técnicas modernas de IA para validar a hipótese de que algoritmos supervisionados, aplicados a radiografias de tórax e TC, podem apoiar o diagnóstico de pneumonia causada por COVID-19 para triagem de pacientes infectados. Foi preparado e disponibilizado um grande banco de dados público de exames de imagem para suportar os experimentos, incluindo uma quantidade significativa de dados oriundos de hospitais brasileiros. Este é um fator importante considerando que as características fenotípicas da população nacional, bem como os equipamentos utilizados para realização dos exames, influenciam na inferência do modelo. Observa-se um grande número de estudos que utilizam somente imagens de bases públicas internacionais (AGGARWAL et al., 2022), não aderentes à realidade do Brasil. Limitam-se também a redes treinadas com dados de poucos pacientes, a maioria com algumas dezenas ou poucas centenas de imagens. A hipótese deste trabalho é que sistemas inteligentes de suporte ao diagnóstico por imagem podem apoiar o diagnóstico de pneumonia por COVID-19, reduzir os riscos de falha, acelerar o processo de interpretação dos exames e permitir uma melhor operação dos serviços de radiologia em períodos de extrema demanda, especialmente em locais com recursos escassos.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo deste trabalho é desenvolver e validar ferramentas baseadas em aprendizagem profunda para o suporte ao diagnóstico de pneumonia causada por COVID-19 utilizando exames de imagem de raio X e tomografia computadorizada.

1.1.2 Objetivos Específicos

Para alcançar o objetivo do trabalho, foi proposto como objetivos específicos:

- Realizar levantamento e análise de bancos de dados públicos de exames de imagem de raio X e tomografia computadorizada.
- Investigar, desenvolver e testar um modelo computacional baseado em redes neurais profundas para o suporte ao diagnóstico de pneumonia causada por COVID-19 utilizando exames de imagem do tórax por raio X.
- Investigar, desenvolver e testar um modelo computacional baseado em redes neurais profundas para o suporte ao diagnóstico de pneumonia causada por COVID-19 utilizando exames de imagem do tórax por tomografia computadorizada.
- Realizar experimentos, testes e ajustes finos a fim de avaliar os modelos propostos através de métricas estatísticas e comparativas com médicos especialistas e outros algoritmos públicos.
- Estudar, analisar e investigar possíveis desvios e vies utilizando técnicas de explicabilidade e interpretabilidade de modelos de IA.

1.2 Organização da Tese

A introdução apresenta o contexto onde o projeto está inserido, a justificativa e os objetivos. O capítulo 2 apresenta uma breve fundamentação teórica em inteligência artificial e uma revisão bibliográfica com o estado da arte na área de aprendizagem profunda aplicada ao suporte ao diagnóstico de pneumonia por COVID-19 em exames de raios X e TC. O capítulo 3 apresenta a solução desenvolvida para suporte ao diagnóstico de pneumonia causada por COVID-19 em exames de raios X através de uma publicação no periódico científico MDPI *Applied Science*, com fator de impacto 2,838. O capítulo 4 apresenta o algoritmo de suporte ao diagnóstico da COVID-19 em

exames de tomografia computadorizada do tórax, através do trabalho publicado na revista MDPI *Diagnostics*, com fator de impacto 3,992. O capítulo 5 apresenta a conclusão da tese e propostas de trabalhos futuros.

2 REVISÃO DA LITERATURA

A urgência por uma resposta rápida e eficiente à pandemia da COVID-19 direcionou os esforços mundiais de pesquisa para este tema. Novos grupos de pesquisa surgiram e os que trabalhavam em temas próximos adequaram seus esforços. A IA aplicada a exames de imagem é um campo em pleno desenvolvimento. A perspectiva de seu uso como uma alternativa rápida e amplamente disponível ao diagnóstico de COVID-19 ampliou a quantidade e qualidade das pesquisas nessa área. Ao longo dos últimos dois anos muitos trabalhos foram apresentados e a cada dia surgem novas publicações relevantes. Neste capítulo serão apresentados os conceitos fundantes de IA e aprendizagem profunda necessários para a compreensão da sua utilização para classificação de imagens e uma revisão do estado da arte na área de aprendizagem profunda aplicada à identificação de pneumonia por COVID-19 em exames de raio X e TC. Revisões bibliográficas extensas já foram produzidas e serão utilizadas como base da análise aqui apresentada.

2.1 Aprendizagem de Máquina na classificação de Imagens

Embora desde a antiguidade a humanidade busque compreender a inteligência e simular o seu comportamento, a IA, como uma disciplina acadêmica surgiu em 1956, quando John McCarthy e Marvin Minsky organizaram um workshop em Dartmouth, Estados Unidos reunindo pesquisadores de várias áreas para uma discussão aberta sobre o tema. O termo “Inteligência Artificial” foi cunhado no próprio evento (ANYOHA, 2017). De acordo com (RUSSELL; NORVIG, 2002), autor de um dos livros de referência na área, a IA pode ser definida como o estudo dos agentes que existem em um ambiente, o percebem e agem. O termo é frequentemente associado ao desenvolvimento de sistemas dotados dos processos intelectuais típicos dos humanos, como a capacidade de raciocinar, descobrir significados, generalizar ou aprender com a experiência passada (COPELAND, 2021).

Ao longo dos anos, a pesquisa em IA passou por diversas fases, com maior ou menor otimismo e conseqüente maior ou menor injeção de recursos. Passou também por muitas abordagens incluindo a IA clássica, baseada em regras e lógica formal, as simulações do cérebro, os sistemas especialistas, entre outras. Atualmente, a abordagem por aprendizagem de máquina (*machine learning*) tem atraído muita atenção, pois está sendo bem-sucedida na resolução de problemas complexos.

Existem muitas empresas e grupos de pesquisas trabalhando no tema e bastante investimento associado. O método permite que os computadores adquiram conhecimento extraindo padrões de dados brutos.

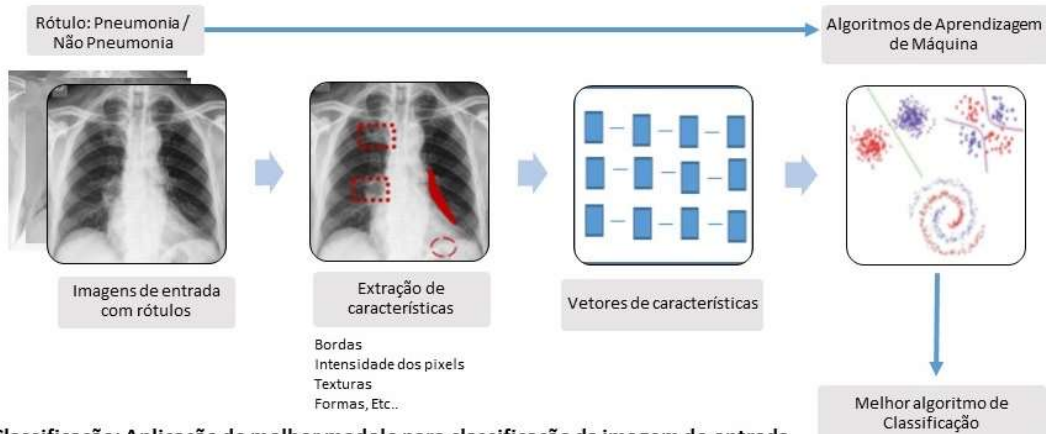
Os algoritmos aprendem com a experiência e uso de dados. Esta abordagem evita a necessidade de uma especificação formal, por um humano, de todo o conhecimento que o computador precisa para tomada de decisão (GOODFELLOW; BENGIO; COURVILLE, 2016). Alguns exemplos de técnicas de aprendizagem de máquina são: redes neurais artificiais, *K-Nearest-Neighbors*, *Support Vector Machines*, árvores de decisão, algoritmo de Bayes ingênuo e aprendizagem profunda. Os dados são fundamentais para essa classe de algoritmos. Logo, a identificação, coleta e preparação dos dados são etapas muito importantes para seu correto desempenho.

Na área médica, a aprendizagem de máquina tem sido utilizada com sucesso para reconhecer padrões em imagens auxiliando os sistemas de suporte ao diagnóstico (ERICKSON et al., 2017). A maioria desses algoritmos utiliza uma abordagem de aprendizagem supervisionada, ou seja, para cada dado de entrada é informado um rótulo com a saída esperada para ele. Um modelo preditor, em geral uma rede neural, é treinado em um processo iterativo de ajustes de parâmetros, com base nos acertos e erros do modelo para cada dado de entrada. O modelo, quando treinado com sucesso, é capaz de prever corretamente uma saída ao receber um novo dado de entrada nunca visto antes (GOODFELLOW; BENGIO; COURVILLE, 2016). Busca-se ajustar os parâmetros do modelo definindo pesos para cada uma das características relevantes extraídas dos dados. O exemplo apresentado na Figura 1 ilustra o conceito. Em uma fase inicial, um conjunto de dados associado a algum conhecimento sobre esses dados é apresentado ao algoritmo de classificação. Neste exemplo são imagens de exames de raio X associadas a um diagnóstico de pneumonia. As características mais relevantes destas imagens (*features*) são extraídas e apresentadas ao algoritmo. Em uma etapa, conhecida como “treinamento”, há um processo iterativo de apresentação dos dados, avaliação das respostas obtidas pelo algoritmo e otimização dos parâmetros do modelo, de modo que o desempenho melhore e mais casos sejam diagnosticados corretamente. Busca-se melhorar a resposta do algoritmo com o conhecimento obtido através dos dados históricos disponíveis Figura 1(a). Após treinado, o sistema pode utilizar o que aprendeu para fazer uma previsão em uma nova imagem, Figura 1(b). Somente as características

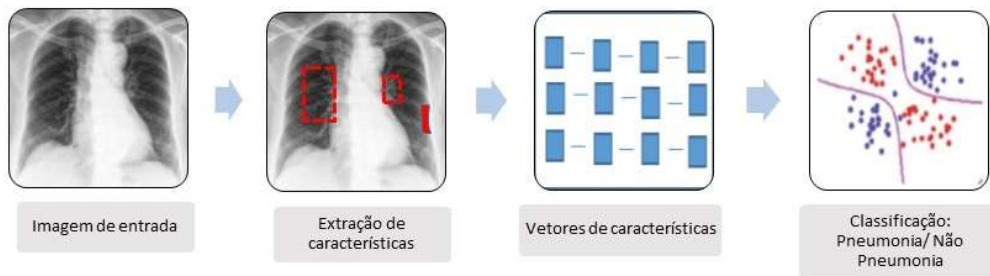
relevantes da imagem são extraídas e usadas pelo modelo para fazer a previsão ou diagnóstico de interesse.

Figura 1. Desenvolvimento de modelo de aprendizagem de máquina para classificação de imagens médicas: (a) Treinamento; (b) Classificação.

(a) Treinamento: Aprendizagem interativa até encontrar o melhor modelo para identificação de pneumonia



(b) Classificação: Aplicação do melhor modelo para classificação da imagem de entrada



Fonte: adaptado de ERICKSON et al. (2017)

A seleção e extração das características relevantes é uma etapa importante do método de aprendizagem por máquina. Nem todas as variáveis de um conjunto de dados são relevantes para o modelo. Para um algoritmo que classifica exames de raio X como “com pneumonia” ou “sem pneumonia”, a imagem dos ossos não é nada relevante, enquanto uma boa representação do pulmão é essencial. Modelos complexos, com muitas variáveis, são mais custosos e não necessariamente melhores (HAWKINS, 2004). O trabalho de (KHALID; KHALIL; NASREEN, 2014) apresentou uma pesquisa das técnicas de seleção e extração de características e avaliou sua eficácia na melhoria da precisão preditiva do classificador. A conclusão é que a performance dos algoritmos de aprendizagem de máquina depende muito da seleção dos atributos e da forma como esses dados são representados.

Existe uma enorme gama de problemas onde é difícil definir com clareza quais

atributos devem ser extraídos. Além disso, é necessário um grande esforço manual de pré-processamento e transformações nos dados para gerar uma representação que possa dar suporte ao aprendizado de forma eficaz. Os algoritmos de aprendizagem de máquina mais simples são incapazes de automaticamente, a partir dos dados, extrair e organizar informações que permitam uma discriminação (GOODFELLOW; BENGIO; COURVILLE, 2016). Por essa ineficiência, é natural que se use o conhecimento especializado humano e a experiência prévia para nortear o desenvolvimento dos algoritmos de extração de características. Utilizando o exemplo anterior, na ausência de uma técnica mais robusta para identificar automaticamente nos exames de raio X as características típicas de uma pneumonia, seria preciso utilizar como referência o conhecimento de uma radiologista e codificar manualmente um algoritmo capaz de extrair da imagem formas muito específicas, características de consolidações, abscessos, espessamentos, padrões nodulares, derrames, etc. e associá-las a regiões definidas da imagem. Além de trabalhoso, as variações são muito grandes e a possibilidade de erros também.

A IA busca aprimorar a percepção do ambiente que a cerca, do mundo em que está inserida. Uma tarefa importante é aprender a identificar e organizar os fatores explicativos ocultos nos dados sensoriais de baixo nível. A fim de expandir a aplicação da aprendizagem de máquina era fundamental tornar os algoritmos de aprendizado menos dependentes desta engenharia manual de atributos e a técnica de aprendizagem profunda trouxe essa capacidade (BENGIO; COURVILLE; VINCENT, 2013).

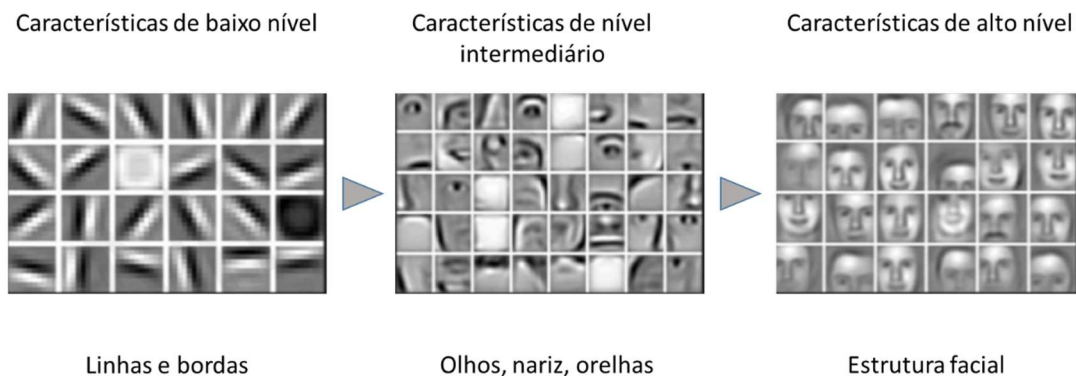
2.1.1 Aprendizagem Profunda

Em 2012, um grupo de pesquisadores da universidade de Toronto, no Canadá, ganhou uma competição acadêmica de classificação de imagens, o ILSVR - ImageNet *Large Scale Visual Recognition Challenge*. Nessa competição, as equipes utilizam um conjunto de aproximadamente 1 milhão de imagens, cada uma rotulada manualmente com uma categoria, para treinar algoritmos de classificação. Os programas são testados fazendo-os sugerir rótulos para imagens semelhantes que eles nunca viram antes. Os vencedores anteriores costumavam errar cerca de 25% das vezes. Este grupo desenvolveu um algoritmo que utilizou um método conhecido como *Deep Learning* (Aprendizagem Profunda) e obteve menos de 15% de erros na classificação de imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Desde então, variações

desta técnica ganharam todas as competições seguintes, reduzindo o erro a menos de 1% e têm sido utilizadas com enorme sucesso para tarefas de processamento de imagens, vídeos, fala e áudio.

Neste método, a seleção e extração das características relevantes dos dados faz parte do próprio processo de aprendizagem. O algoritmo busca entender o ambiente em termos de uma hierarquia de conceitos, com cada conceito definido em termos de sua relação com conceitos mais simples. Esta abordagem aproxima-se do processo como o cérebro humano reconhece padrões visuais. O nervo ótico transmite os sinais obtidos para regiões do cérebro que reconhecem características básicas como bordas e linhas, que se comunicam com outras regiões que reconhecem formas mais complexas, formando contornos, texturas e evoluindo até objetos e imagens únicas. A Figura 2 ilustra a ideia, o algoritmo possui camadas de extração de características que reconhece formas de complexidade crescente para categorizar formas mais complicadas, como faces.

Figura 2. Camadas de extração de características de algoritmos de aprendizagem profunda.

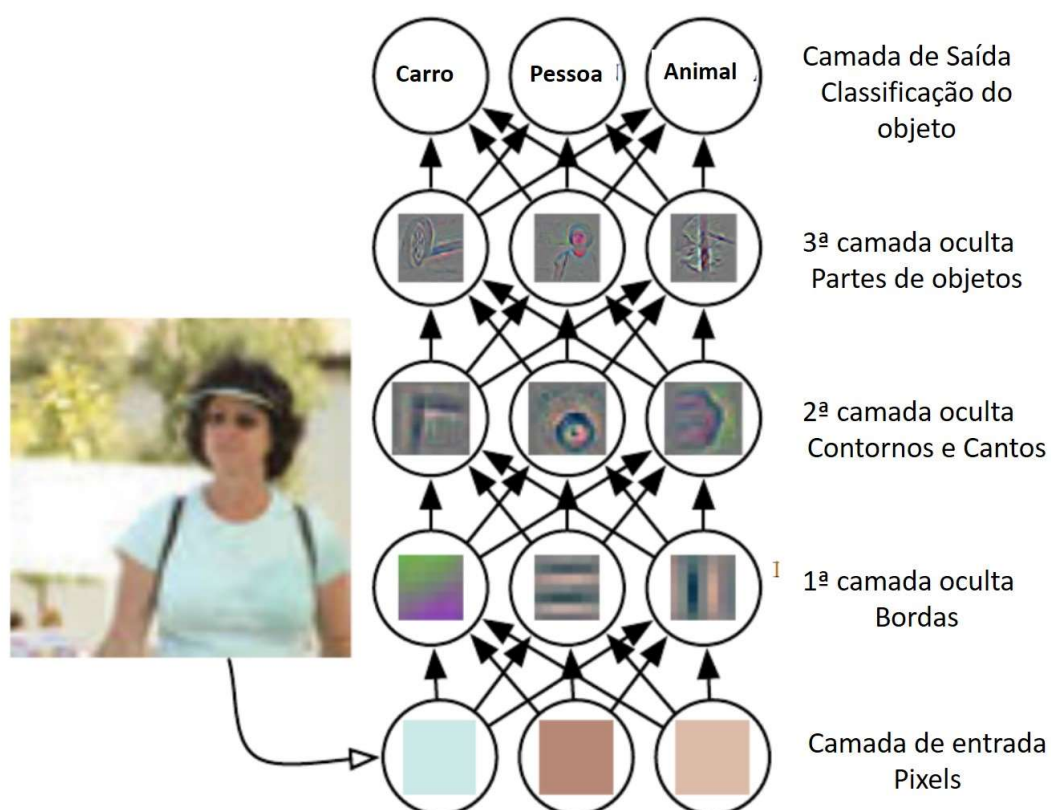


Fonte: imagem baseada no slide de Alexander Amini e Ava Soleimany, MIT 6.S191: Introduction to Deep Learning, IntroToDeepLearning.com

A técnica de aprendizagem profunda faz parte de uma abordagem conhecida como “aprendizagem por representação”. Ela utiliza vários níveis de representação, obtidos pela composição de módulos simples, mas não lineares. Começando pela entrada bruta dos dados, pixels de uma imagem por exemplo, cada módulo transforma a representação em um nível de abstração, em uma representação em um nível superior, um pouco mais abstrata. Com a composição de tais transformações, funções muito complexas podem ser aprendidas. Em um processo iterativo de aprendizagem, camadas superiores de representação amplificam os aspectos da entrada que são importantes para a discriminação e suprimem as variações que não são relevantes

(GOODFELLOW; BENGIO; COURVILLE, 2016). A Figura 3 apresenta o conceito aplicado à classificação de imagens. Os recursos aprendidos na primeira camada de representação sinalizam a presença ou ausência de bordas em orientações e locais específicos na imagem. A segunda camada detecta arranjos particulares de bordas, independente de pequenas variações nas suas posições. A terceira camada reúne combinações maiores que correspondem a parte de objetos familiares, e as camadas seguintes detectam objetos como combinações dessas partes (LECUN; BENGIO; HINTON, 2015). Um aspecto muito importante da aprendizagem profunda a ser ressaltado é que essas camadas de extração de características não foram projetadas por um analista de sistemas, elas foram aprendidas, a partir dos dados, utilizando um procedimento de aprendizado de propósito geral.

Figura 3: Aprendizagem Profunda: Características são extraídas através de uma série de mapeamentos simples encadeados, cada um descrito por uma camada diferente do modelo.



Fonte: Traduzido de (GOODFELLOW; BENGIO; COURVILLE, 2016)

2.1.2 Redes Neurais Convolucionais

Os algoritmos de aprendizagem profunda são normalmente baseados em redes neurais artificiais. As redes neurais convolucionais, em inglês: *Convolutional Neural Networks* (CNN), têm sido muito bem-sucedidas na classificação de imagens. Elas

são projetadas para receber dados de entrada organizados como múltiplos vetores. Essa é uma representação típica para imagens. Por exemplo, uma imagem colorida é composta por três matrizes de duas dimensões contendo as intensidades dos pixels para três canais de cores. Essas redes são formadas por camadas de convolução seguidas de camadas de pooling (LECUN; BENGIO; HINTON, 2015). Muitas redes convolucionais, com diferentes arquiteturas, foram desenvolvidas e tiveram bastante sucesso na classificação de imagem. As arquiteturas VGGNet (SIMONYAN; ZISSERMAN, 2014), Inception / GoogleNet (SZEGEDY et al., 2015) e ResNet (HE et al., 2016) conquistaram a competição ILSVRC em anos diferentes e acabaram servindo de inspiração e base para o desenvolvimento de muitas outras redes específicas para diversas áreas.

Como observado na Figura 3, as camadas iniciais das redes convolucionais aprendem a extrair características fundamentais das imagens. Com isso o algoritmo é capaz de identificar bordas, linhas, cantos, padrões. Todos os objetos da natureza possuem essas características, logo, surgiu a ideia de aproveitar os algoritmos já treinados em bases de dados de imagens naturais, como o da competição ILSVRC, para acelerar o treinamento de novos algoritmos. Para isso os pesos das camadas iniciais dos algoritmos pré-treinados (ex. Resnet, VGG) são congelados e as camadas finais são treinadas com os novos dados. Essa técnica é conhecida como aprendizagem por transferência (*transfer learning*) (TORREY; SHAVLIK, 2010).

2.1.3 Treinamento de redes neurais, *overfitting* e *underfitting*

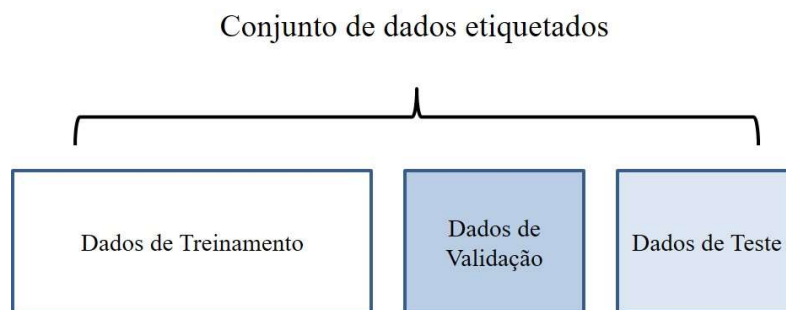
As redes convolucionais podem ter muitas camadas. Aumentando a profundidade, a rede pode aproximar melhor a função objetivo com maior não linearidade e obter melhores representações de recursos. No entanto, também aumenta a sua complexidade, tornando-a mais difícil de otimizar. Existem muitas técnicas e formas de ajustes nos parâmetros das redes neurais (hiperparâmetros) para encontrar o modelo mais adequado (CHOLLET, 2021).

Durante o processo de treinamento, os pesos dos nós da rede são continuamente ajustados buscando-se reduzir o erro. Para isso, o algoritmo de aprendizagem tenta minimizar uma função que relaciona o valor real e o predito. Essa função é conhecida como função de perda (*loss*). A cada ciclo de treinamento, ou época, os pesos são ajustados através da interação com a coleção de dados existentes, em um processo conhecido como *fitting*. O treinamento é realizado até que

o erro obtido esteja estabilizado no menor valor encontrado. Nem sempre a rede estabiliza no menor erro global. É possível que o algoritmo encontre um mínimo local da função. Isso representa uma solução sub ótima(GOODFELLOW; BENGIO; COURVILLE, 2016).

A avaliação do desempenho dos algoritmos é baseada nas probabilidades de acerto dos preditores desenvolvidos. Para avaliar o desempenho do modelo durante a fase de treinamento é comum reservar parte do conjunto de dados para validações e testes, Figura 4.

Figura 4. Separação de dados para treinamento, validação e teste



Fonte: Adaptado de (CHOLLET, 2021).

Ao longo do processo de treinamento, os dados de validação ajudam a verificar a qualidade do modelo, que ao final, depois de completamente ajustado, é avaliado sobre os dados de testes.

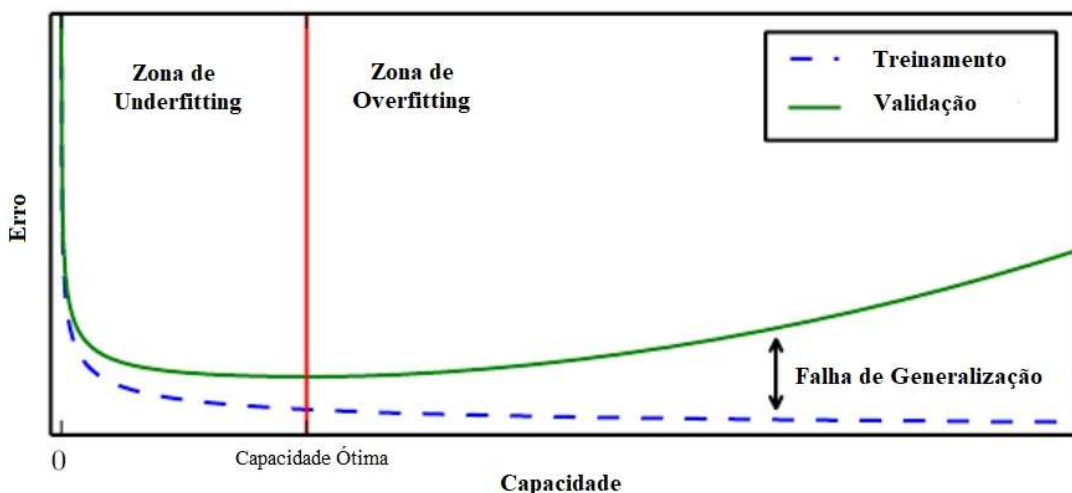
Utilizando os dados de validação é possível comparar, ao longo do treinamento, época a época, a evolução do erro do modelo em um conjunto de dados diferentes. Isso nos permite observar a capacidade de generalização do algoritmo. A capacidade de generalização é fundamental pois os dados coletados para treinamento são apenas uma amostra da realidade. Eles são incompletos e contém ruídos. O grande desafio dos modelos de aprendizagem profunda é minimizar o erro, não nos dados de treinamento, mas em dados nunca vistos antes. Uma rede com uma boa capacidade de generalização irá apresentar valores de loss convergentes nos dados de treinamento e validação.

É importante observar se o modelo está sub treinado (underfitting) ou sobre treinado (overfitting). Na primeira situação, o treinamento é realizado com uma quantidade pequena de dados ou com uma quantidade insuficiente de ciclos levando a uma baixa eficiência com previsões não confiáveis. O modelo é incapaz de capturar

a complexidade das características de entrada e conseqüentemente há uma alta quantidade de erros (loss). No segundo caso, o modelo é sobretreinado para as características dos dados de entrada, sendo incapaz de generalizar bem para dados novos. O overfitting ocorre devido à baixa variabilidade dos dados de entrada ou a modelos muito complexos que se ajustam demais ao comportamento dos dados de treinamento (HAWKINS, 2004), geram loss baixos nos dados de treinamento e não tem um bom desempenho em dados novos, gerando um loss mais alto na validação.

O gráfico da Figura 5 apresenta a relação entre o erro e complexidade do modelo. No eixo x está representada a capacidade do modelo de representar várias funções, quanto mais funções o modelo é capaz de representar, mais complexo ele é, e maior a tendência de sobretreinamento. No eixo y estão representados os erros para os dados de treinamento e validação. Um modelo com a capacidade ótima para o problema terá baixo índice de erros nos dados de teste e validação. Os modelos subtreinados, na zona de underfitting, à esquerda, apresentarão alto índice de erros nos dados de treinamento. Já os modelos sobretreinados, na zona de overfitting, apresentarão erros nos dados de validação bem maiores que nos dados de treinamento. A lacuna entre os erros de treinamento e validação representa a incapacidade de generalização do modelo.

Figura 5. Underfitting e Overfitting



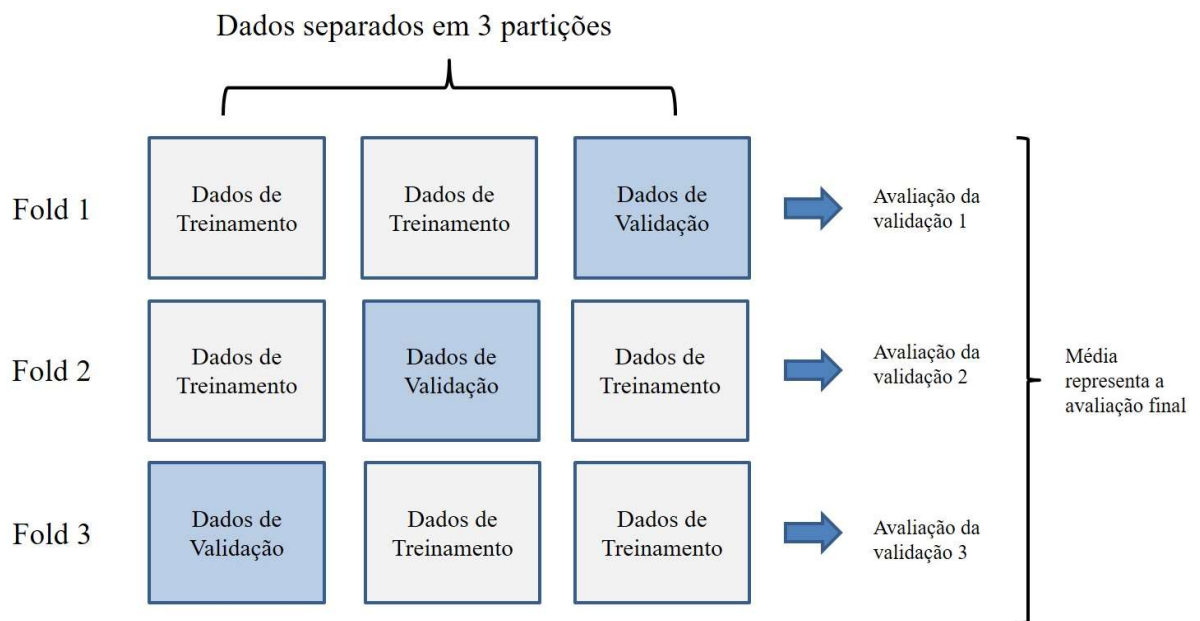
Fonte: Adaptado de (GOODFELLOW; BENGIO; COURVILLE, 2016)

Separar dados para a validação (*hold-out validation*) é uma técnica muito útil mas pode ser um problema se tivermos uma quantidade pequena de dados pois reduz a quantidade disponível de dados para treinamento. Uma técnica alternativa permite

utilizar todos os dados para a estimativa do erro médio do teste. Esse procedimento é baseado na ideia de repetir o treinamento e testar o resultado em diferentes subconjuntos dos dados. Essa técnica é conhecida como validação cruzada (*k-fold cross validation*). O k representa a quantidade de subconjuntos (não sobrepostos) que serão utilizados.

No exemplo da Figura 6, o conjunto de dados foi dividido em 3 partes, para cada uma dessas partes, o modelo utiliza 2 partes ($k-1$) para treinamento, e 1 parte para validação. Ao final do processo, quando o modelo treinar 3 vezes, a média e o desvio padrão de todos os treinos realizados definirão quão bem o modelo está generalizando.

Figura 6. Validação Cruzada 3-Fold.



Fonte: Adaptado de (CHOLLET, 2021).

Neste trabalho utilizamos a técnica de *hold-out* para a validação do modelo de suporte ao diagnóstico de COVID-19 por raio-X e a validação cruzada no modelo baseado em TC.

2.1.4 Métricas para avaliação do modelo

Na avaliação de sistemas de suporte a diagnósticos é fundamental analisar a capacidade preditiva do modelo. Uma abordagem muito útil para análise de classificadores é a construção de uma matriz de confusão. Nela são representadas a

quantidade de testes falso positivos e falso negativos, verdadeiros positivos e verdadeiros negativos, permitindo o cálculo da acurácia, sensibilidade e especificidade do modelo.

O exemplo da Figura 7 apresenta uma matriz de confusão para o classificador de radiografias entre COVID-19 e anormais e a Tabela 1, a seguir, apresenta os principais achados identificados na matriz de confusão, a fim de ilustrar seu entendimento.

Figura 7. Exemplo de uma Matriz de Confusão para um modelo classificador de exames anormais/COVID-19.

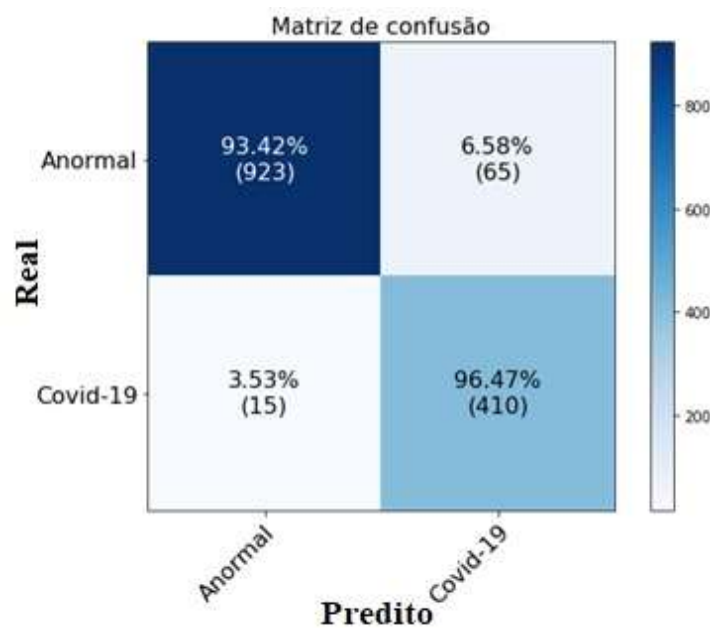


Tabela 1. Dados obtidos a partir da matriz de confusão

Métrica	Descrição	Aplicação no exemplo da Figura 7	Valor
Condição Positiva (P)	Número de casos positivos reais nos dados.	Casos com identificação de COVID-19.	P=425
Condição Negativa (N)	Número de casos reais negativos nos dados.	Exames com etiqueta Anormal.	N=988
Verdadeiros Positivos(VP)	Resultados do modelo que indicam corretamente a presença de uma condição ou característica.	Indicação de COVID-19 pelo modelo que realmente tem COVID-19.	VP=410
Verdadeiros Negativos (VN)	Resultados do modelo que indicam corretamente a ausência de uma condição ou característica.	Indicação correta de ausência de COVID-19.	VN = 923
Falso positivo (FP)	Resultados do modelo que indicam incorretamente que uma determinada condição ou atributo está presente.	Indicação incorreta da presença de COVID-19.	FP = 65
Falso negativo (FN)	Resultados do modelo que indicam erroneamente que uma determinada condição ou atributo está ausente.	Indicação incorreta como não COVID-19.	FN = 15

Fonte: Produção Própria

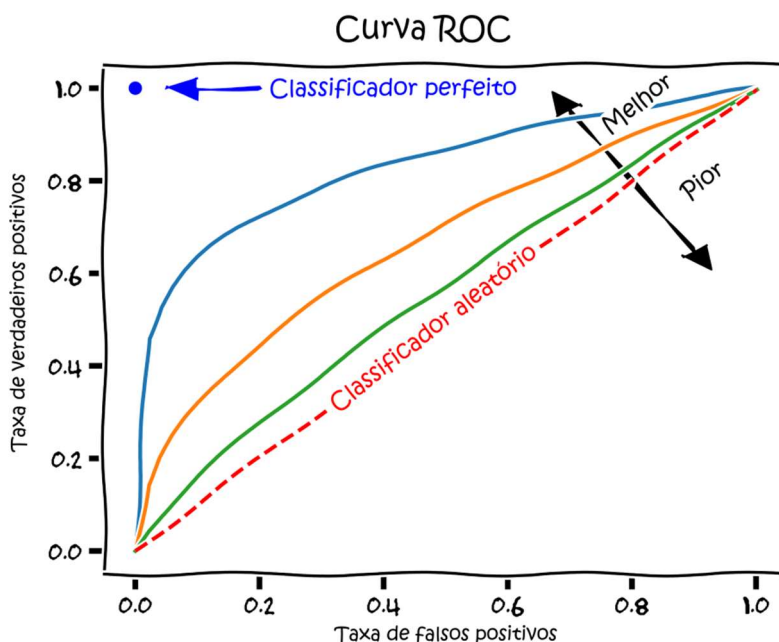
A partir da matriz podemos calcular as seguintes métricas:

- Sensibilidade (*sensitivity* ou *recall*): A probabilidade de um exame avaliado, que tem indicação real de COVID-19 ser avaliado como COVID-19.
- Sensibilidade = $VP / (VP + FN)$. Logo a sensibilidade deste modelo é de 96%.
- Especificidade (*specificity*): A probabilidade de um exame não COVID-19, ter uma avaliação de não COVID-19 pelo modelo.
- Especificidade = $VN / (VN + FP)$. Neste exemplo, 93%.
- Precisão ou valor preditivo positivo (VPP): A probabilidade de um exame avaliado como COVID-19, realmente ter COVID-19.
- Precisão = $VP / (VP + FP)$. Neste modelo, a precisão é de 86%.

Uma avaliação da qualidade do modelo não pode ser obtida com a análise de uma única métrica. É importante avaliar a relação entre elas. Existem alguns indicadores desta natureza: F1 é uma métrica que apresenta a média harmônica entre precisão e sensibilidade. O maior valor possível para a métrica é 1, indicando uma precisão e sensibilidade perfeitas. $F1 = VP / (VP + 0.5 * (FP + FN))$. Neste exemplo $F1 = 0,91$. A Curva Característica de Operação do Receptor, ou, curva ROC (do inglês Receiver Operating Characteristic) é outra métrica importante para classificadores binários. Ela apresenta graficamente o desempenho do sistema através de uma análise do *trade-off* entre sensibilidade e (1-especificidade) com relação a um limiar de discriminação variável. A Figura 8 apresenta este conceito. O gráfico apresenta a fração de previsões corretas para a classe positiva, no eixo y, versus a fração de erros para a classe negativa, no eixo x. Quanto mais alto e mais distante da linha diagonal a curva estiver, melhor. O modelo ideal teria a fração de previsões de classe positiva corretas como 1 e a fração de previsões de classe negativa incorreta como 0. Com isso, o melhor classificador possível estaria no canto superior esquerdo do gráfico, na coordenada (0,1). A curva é construída a partir da avaliação entre os verdadeiros positivos e falsos positivos para diferentes valores do discriminador. Um modelo que não tenha capacidade de discriminar entre as classes positivas e negativas formará uma linha diagonal entre uma taxa de falso positivo de 0 e uma taxa de verdadeiro positivo de 0 para uma taxa de falso positivo de 1 e uma taxa positiva verdadeira de

1, representado pela linha diagonal vermelha no gráfico. Os modelos representados por pontos abaixo desta linha são ruins pois não tem habilidade discriminadora. A análise pode ser transformada em um indicador único escalar, calculando a área sob a curva. Esta métrica chama-se ROC AUC (do inglês, *Area Under the ROC Curve*) e varia entre 0 e 1, quanto mais próximo de 1 melhor.

Figura 8. Ilustração de uma curva ROC



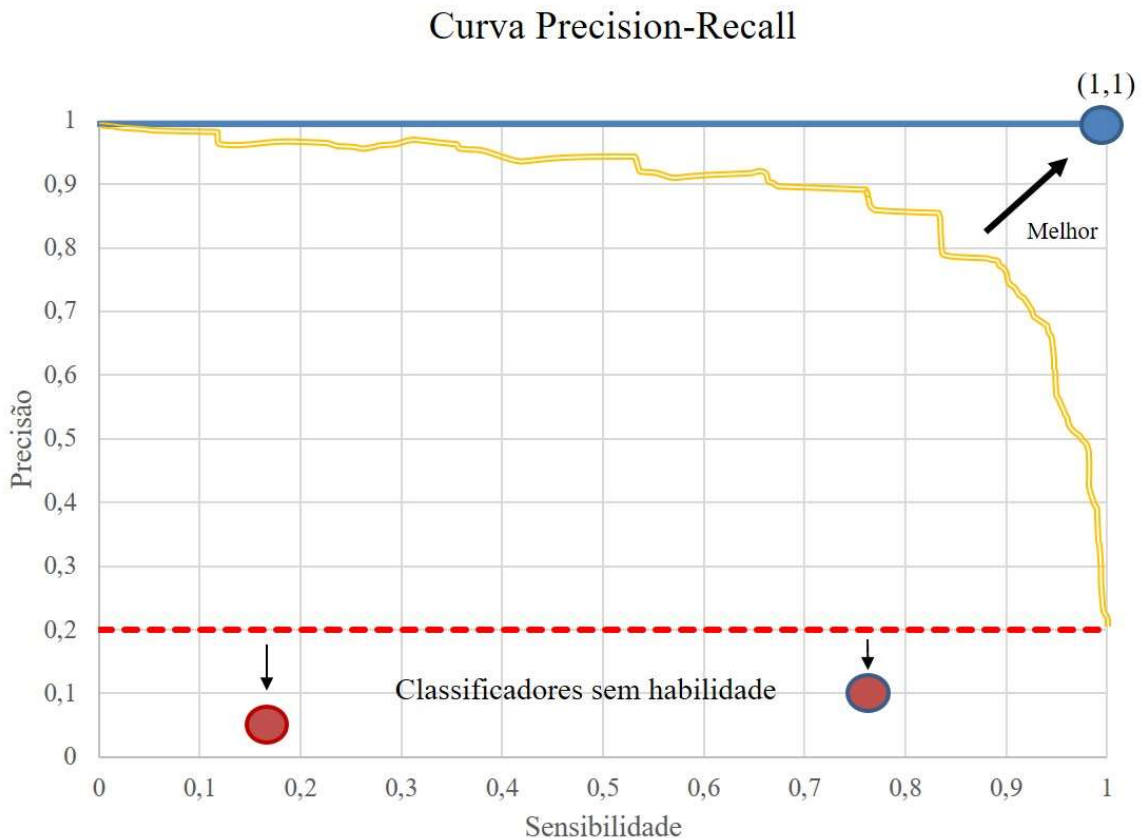
Fonte: dcbmariano adaptado de MartinThoma (Creative Commons CC0 1.0 Universal Public Domain Dedication).

A curva ROC pode ser otimista demais para situações onde existem muitos exemplos negativos, em nosso caso, muitos exames sem indicação de COVID-19. Para lidar com conjuntos de dados com distribuição de classes desbalanceada, um outro tipo de gráfico é mais informativo, a curva Precision-Recall (PRC). O PRC permite comparar falsos positivos com verdadeiros positivos, ao invés de verdadeiros negativos, e com isso, pode capturar o efeito no desempenho do algoritmo de uma amostra com um número grande de exemplos negativos. Neste gráfico, como observado na Figura 9, a precisão é representada no eixo y e a sensibilidade no eixo x, e plota-se uma curva com todos os valores do discriminador entre 0 e 1. Um modelo perfeito seria representado como um ponto na coordenada (1,1) e os melhores modelos apresentariam uma curva direcionada a esta coordenada. Já um modelo sem habilidade seria uma linha horizontal no gráfico com uma precisão proporcional ao número de exemplos positivos no conjunto de dados, neste caso 20%. Como o foco

da curva PR é a classe minoritária, esta métrica é adequada para complementar a avaliação de modelos de classificação binária desequilibrada.

Assim como na curva ROC, pode-se calcular a área sob a curva PR para obter um indicador escalar e utilizá-lo para comparar modelos.

Figura 9. Curva Precision Recall com 20% de amostras positivas.



Fonte: Produção própria

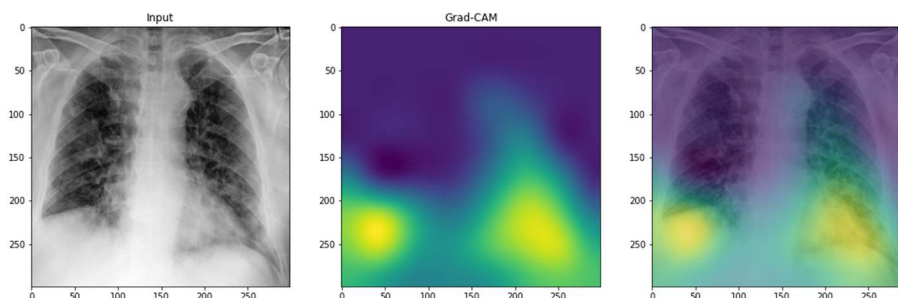
2.1.5 Explicabilidade

Os modelos de aprendizagem profunda expressam relações complexas sobre os dados e muitas vezes são tratados como modelos caixa-preta, ou seja, não há uma compreensão completa sobre como o modelo realiza a predição. Não se sabe quais atributos são mais relevantes e quais são as conexões entre eles. Essa aparente desinformação gera insegurança na adoção desses modelos em sistemas para área de saúde (SAMEK; MÜLLER, 2019), em especial, porque a ciência médica historicamente exige validações e comprovações de todo o processo físico associado ao resultado.

A área conhecida como explicabilidade em IA, em inglês XAI (*eXplainable AI*), busca entender e explicar o comportamento dos modelos e com isso aumentar a qualidade e confiança nas soluções desenvolvidas. Essas técnicas permitem detectar vieses, falhas no modelo e identificar os atributos realmente relevantes, ampliando o conhecimento sobre o modelo e auxiliando o seu desenvolvimento.

Neste trabalho utilizamos a técnica *Gradient-weighted Class Activation Mapping* (Grad-CAM) para produzir 'explicações visuais' sobre as decisões dos modelos. A técnica foi desenvolvida por (SELVARAJU et al., 2017) e utiliza os gradientes que fluem para a camada convolucional final para produzir um mapa de localização que destaca as regiões relevantes da imagem de entrada utilizadas pelo preditor. Como pode ser observado no exemplo apresentado na Figura 10, essa abordagem permite gerar uma visualização de cores que auxilia a interpretação do modelo. As características da imagem de entrada mais importantes para a decisão do modelo têm uma cor mais quente.

Figura 10. Exemplo de mapa de calor gerado com a técnica de Grad-CAM para explicabilidade da predição realizada por um modelo de IA para suporte a diagnóstico médico de COVID-19 em um exame de raio X.



Fonte: produção própria

2.2IA na Medicina

A Inteligência Artificial está causando um impacto importante na prática médica. Os sistemas de suporte ao diagnóstico, prognóstico e tratamentos facilitam a realização de planos mais personalizados e com menos custos. Apoiados por sistemas informatizados inteligentes, médicos podem realizar a detecção precoce de doenças, identificar padrões, avaliar e propor protocolos e personalizar a ação de resposta para cada paciente de forma mais barata e rápida (RAJPURKAR, 2021). O desenvolvimento de soluções habilitadas por IA dependem basicamente de três

fatores: algoritmos, disponibilidade de dados e poder computacional. Em relação aos algoritmos, os avanços recentes na área de aprendizagem profunda permitiram a realização de sistemas de classificação de imagens com bastante sucesso (LECUN; BENGIO; HINTON, 2015). Consequentemente, cresceu o estudo de soluções relacionadas ao uso da técnica nas especialidades médicas que tratam da interpretação de imagens. Foram observados avanços expressivos, especialmente nas áreas de radiologia, oftalmologia e patologia, como apontado na pesquisa realizada por (LITJENS et al., 2017). O fator primordial para o sucesso desses algoritmos é a evolução dos conjuntos de dados utilizados para seu treinamento. É cada vez mais comum a curadoria especializada de grandes conjuntos de imagens, incluindo rótulos para as segmentações de achados de interesse, associados a diagnósticos e dados clínicos dos pacientes. Entretanto, há uma grande preocupação dos sistemas de saúde com o sigilo dos dados dos pacientes por questões éticas e legais. Uma abordagem promissora para superar o desafio da governança e privacidade de dados para o treinamento de redes neurais é o “aprendizado federado” (*Federated Learning*) (RIEKE et al., 2020). Nesta técnica, o algoritmo é treinado de forma colaborativa. Múltiplos centros de saúde podem contribuir com suas bases anonimizadas, sem a necessidade de trocar os dados entre si. O estudo de (DAYAN et al., 2021) demonstrou o sucesso da técnica desenvolvendo um algoritmo para previsão de necessidades futuras de oxigênio de pacientes sintomáticos com COVID-19. O estudo utilizou a informação dos sinais vitais, dados laboratoriais e radiografias de tórax de pacientes de 20 instituições ao redor do mundo.

Apesar dos avanços, existem ainda poucos exemplos do uso de algoritmos de aprendizagem profunda para interpretação de imagens médicas implantadas com sucesso na prática clínica (KELLY et al., 2019). Até mesmo com a profusão de projetos aplicados à detecção e prognóstico da COVID-19, a maioria dos projetos apresentou falhas metodológicas importantes que limitaram seu uso à área acadêmica (ROBERTS et al., 2021).

Existem alguns desafios técnicos que precisam ser superados e que foram endereçados neste trabalho. Muitos dos algoritmos desenvolvidos para identificação de COVID-19 em exames de imagens, utilizaram dados oriundos de bases estruturadas e pré-processadas para esse objetivo específico. Estes dados, limpos, direcionados, não correspondem à realidade encontrada na medicina onde os dados são menos disponíveis e os rótulos são mais ruidosos. O tamanho do conjunto de

dados foi um outro fator limitante. Grande parte dos trabalhos utilizou um conjunto de dados pequeno e homogêneo limitando a capacidade de generalização dos algoritmos. Outro desafio importante está relacionado à metodologia adotada para avaliação dos algoritmos. Além da falta de rigor com os métodos estatísticos, muitos trabalhos publicados validaram os algoritmos no contexto das distribuições do conjunto de dados em que os algoritmos foram treinados. A avaliação visando a implantação na prática médica exigiria uma avaliação do desempenho do algoritmo com mudanças na distribuição dos dados clinicamente relevantes. Para os testes deste trabalho treinamos os modelos em imagens oriundas de diferentes partes do mundo e os testamos utilizando dados de pacientes de hospitais brasileiros, uma vez que a intenção futura é disponibilizar a solução para prática médica local.

2.3 Suporte ao Diagnóstico de COVID-19 em Exames de Imagem de Raio X e TC

A pandemia de COVID-19 tem demonstrado algumas fragilidades dos sistemas nacionais de saúde. A escassez de recursos materiais e humanos demanda soluções alternativas para acelerar os diagnósticos e encaminhamentos dos pacientes. Em algumas regiões não há exames RT-PCR amplamente disponíveis e faltam especialistas que possam apoiar o diagnóstico rápido da doença. A prática médica tem demonstrado que os exames de imagem do pulmão podem ser uma alternativa para apoiar a identificação da doença e sua evolução (SANDRI et al., 2021).

O uso de exames de imagem como suporte ao diagnóstico de COVID-19 tem sido amplamente discutido por entidades de referência na área médica. A OMS apresentou um protocolo de referência que considera a radiografia de tórax, TC e ultrassom pulmonar parte da investigação diagnóstica de pacientes com suspeita ou probabilidade de COVID-19, nos lugares em que a RT-PCR não está disponível ou em que os resultados demoram ou são inicialmente negativos mas há presença de sintomas sugestivos de COVID-19 (WORLD HEALTH ORGANIZATION, 2020). O documento orientador sobre diagnóstico, tratamento e isolamento de pacientes com COVID-19 disponibilizado pelo colégio brasileiro de radiologia (“COVID-19 - Colégio Brasileiro de Radiologia e Diagnóstico por Imagem,” [s.d.]) recomenda que não se deve realizar qualquer exame de imagem a pacientes assintomáticos ou sintomáticos leves. Orienta também que pacientes com casos moderados, sem acesso a exames laboratoriais, ou com PCR negativo, podem realizar exames com orientação clínica.

Já para os pacientes hospitalizados, sintomáticos, com quadro moderado ou grave, a tomografia computadorizada pode ser indicada, especialmente para avaliar suspeita de complicações e analisar diagnósticos diferenciais. Os exames de ultrassom são úteis e apresentam achados similares ao TC (MANNA et al., 2020), entretanto a ausência de grandes bases de dados públicas limitam o seu uso por algoritmos de aprendizagem profunda. Portanto, neste trabalho focamos nos exames radiológicos e de TC. As duas abordagens foram amplamente utilizadas para apoiar o diagnóstico de pneumonia em pacientes sintomáticos nas fases iniciais da pandemia (PONTONE et al., 2021).

Aparelhos de radiografia são amplamente disponíveis, geram baixa exposição à radiação e podem ser utilizados ao lado do leito do paciente. No Brasil, por exemplo, segundo dados demográficos disponibilizados pelo Instituto Brasileiro de Geografia e Estatística e pelo cadastro nacional de estabelecimentos de saúde, disponibilizado pelo ministério de saúde do país, existe aproximadamente um equipamento de raios X para cada 2.732 habitantes. Como pode ser observado na Tabela 2, a distribuição não é uniforme pelas regiões, mas mesmo as regiões menos assistidas possuem uma quantidade relevante de equipamentos de raios X. Os tomógrafos são mais escassos, com uma média de 1 aparelho para cada 35.666 habitantes.

Tabela 2- Disponibilidades de aparelhos de Raios X e Tomógrafos no Brasil.

Região	Equipamentos de Raios X (Abr/22)	Habitantes/Equipamento de Raios X	Tomógrafos (Abr/22)	Habitantes/Tomógrafo /	População Estimada (Jun 2021)
Norte	3.993	4.735	418	45.232	18.906.962
Nordeste	14.116	4.085	1.146	50.321	57.667.842
Sudeste	38.613	2.321	2.827	31.706	89.632.912
Sul	14.904	2.040	1.043	29.149	30.402.587
Centro-Oeste	6.455	2.588	736	22.700	16.707.336
Total	78.081	2.732	5.981	35.666	213.317.639

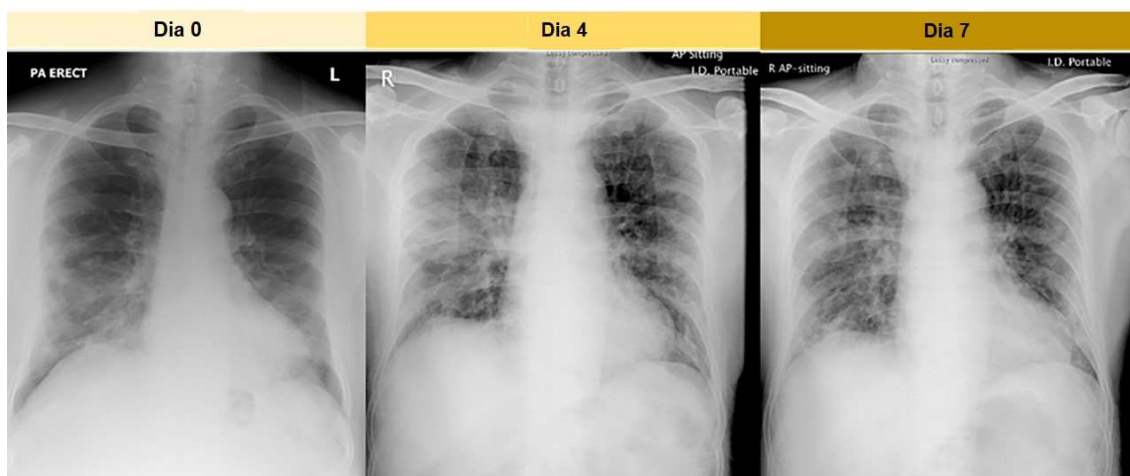
Fonte: (“Cadastro Nacional de Estabelecimentos de Saúde,” [s.d.] / (“Estimativas da população residente para os municípios e para as unidades da federação | IBGE,” [s.d.]).

Ainda mais escassa é a disponibilidade de médicos especializados em exames de imagem. Segundo os dados do estudo “Demografia Médica no Brasil”, divulgado em 2020, existem apenas 14.225 radiologistas no país (SCHEFFER et al., 2020). É um desafio disponibilizar um diagnóstico rápido nos serviços de pronto atendimento espalhados pelo país, mesmo com o uso de técnicas de telemedicina.

Os achados de lesões pulmonares na COVID-19 incluem opacidades em vidro fosco e consolidações periféricas, podendo ter aspecto nodular e predomínio basal. Derrames pleurais não são comumente vistos (ZHAO et al., 2020). Com a demonstração da existência de atributos característicos de COVID-19 em exames radiológicos e a habilidade preditiva do mesmo, abriu-se espaço para o desenvolvimento de algoritmos de aprendizagem profunda supervisionados capazes de classificar exames com COVID-19 utilizando bases de dados históricas de exames com laudo.

Na fase inicial da doença, os achados são sutis e muitas vezes se confundem com os sinais de outras doenças respiratórias. Pacientes com a doença mais evoluída possuem achados nas radiografias mais evidentes (NG et al., 2020). Uma equipe de radiologistas da LSU Health New Orleans descreveu uma aparência radiográfica de tórax característica com alta especificidade (96,6%) e valor preditivo positivo (83,8%) para infecção por COVID-19 (SMITH et al., 2020). O desempenho diagnóstico de radiografia de tórax foi confirmado em um estudo em um hospital em Milão, Itália, demonstrando alta sensibilidade para identificação da doença, com maior especificidade para radiologistas mais experientes (COZZI et al., 2020). A sensibilidade da radiografia torácica depende muito do estágio da infecção pulmonar e da extensão da doença, bem como na qualidade técnica do exame variando de 50% a 84%. A especificidade é baixa, atestada em 33% (LAINO et al., 2021). A Figura 11 apresenta imagens de um paciente com COVID-19 de Wuhan, China, com alguns achados típicos de COVID-19 em radiografias do tórax. Há uma consolidação na zona inferior direita no dia 0, que persistiu no dia 4 com novas mudanças de consolidação na periferia da zona média direita. Houve uma melhora na área média no dia 7 (NG et al., 2020).

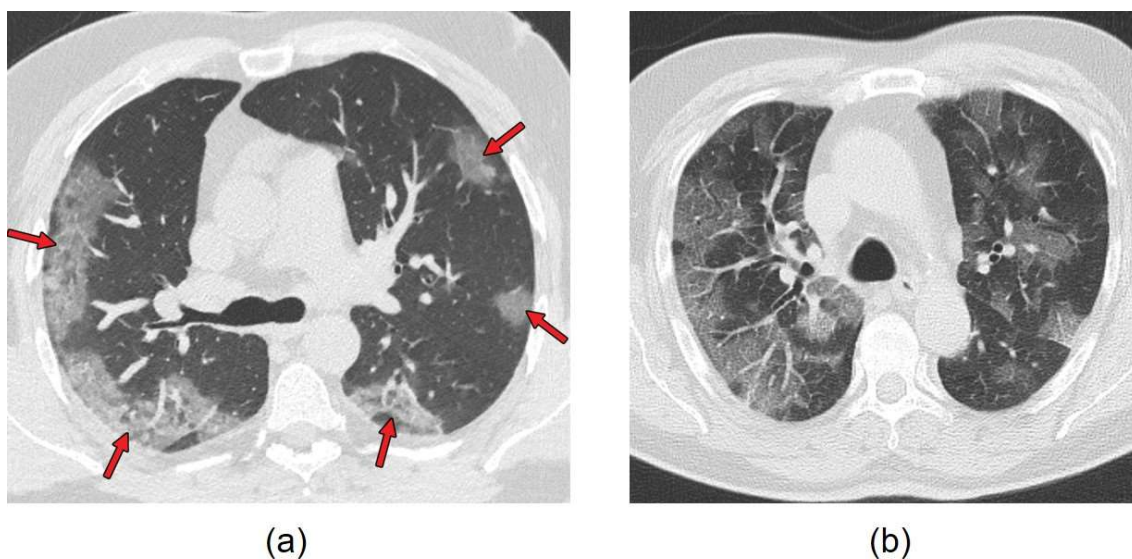
Figura 11. Radiografias de tórax em um paciente idoso do sexo masculino de Wuhan, China



Fonte: (NG et al., 2020)

Os exames de TC oferecem melhores imagens e são mais conclusivos. Muitos estudos demonstraram a existência de situações anormais nas imagens da TC de tórax com características próprias da doença, possibilitando o diagnóstico de COVID-19, inclusive em sua fase inicial (KWE; KWE, 2020). A Figura 12 apresenta dois exames de TC em pacientes com RT-PCR positivo para SARS-CoV-2 com achados típicos para COVID-19. Estes achados possuem uma incidência superior a 70%. A imagem da esquerda mostra áreas bilaterais de opacidades em vidro fosco em uma distribuição periférica em um homem de 47 anos. A imagem da direita apresenta vasos segmentares e subsegmentais dilatados, principalmente à direita, em um homem de 70 anos.

Figura 12. Características típicas de COVID-19, com alta incidência (> 70%) (a) Áreas bilaterais de opacidades em vidro fosco (setas) em uma distribuição periférica. (b) Vasos segmentares e subsegmentais dilatados



Fonte: Adaptado de (KWEE; KWEE, 2020)

Muitos pacientes com COVID-19 não apresentam sintomas pulmonares. O exame de TC é recomendado para pacientes hospitalizados sintomáticos e não deve ser utilizado como único fator de definição de diagnóstico de COVID-19. Em uma situação pandêmica, com muitos pacientes, o alto tempo de limpeza do aparelho de TC limita sua utilização para triagem de pacientes. O equipamento possui um potencial de exposição e transmissão maior que o exame de raio X. Quando indicada, deve-se realizar uma TC de alta resolução e se possível com protocolo de baixa dose, sem uso de meio de contraste endovenoso. Essa recomendação é adotada por conselhos de radiologia de muitos países, incluindo os conselhos de radiologia americano e brasileiro (AKL et al., 2021; “COVID-19 - Colégio Brasileiro de Radiologia e Diagnóstico por Imagem,” [s.d.]; SIMPSON et al., 2020b).

Há consenso que, em locais com recursos limitados, exames de imagem podem ser indicados para triagem médica de pacientes com suspeita de pneumonia por COVID-19 que apresentem características clínicas de moderadas a graves e uma alta probabilidade da doença (RUBIN et al., 2020). Há muitas regiões, no Brasil e no mundo, que não tem acesso a exames RT-PCR na quantidade e momento necessário, muito menos disponibilidade de médicos especializados. Nestes casos, alternativas que facilitem o diagnóstico de pneumonia por COVID-19 são muito importantes. Médicos chineses e espanhóis utilizaram largamente os exames de imagem de raio X

e TC como suporte ao diagnóstico da COVID-19 durante a primeira fase aguda da pandemia (FAN; LIU, 2020; “Imaging the coronavirus disease COVID-19 • healthcare-in-europe.com,” 2020).

2.3.1 Uso de Deep Learning no suporte ao diagnóstico de COVID em exames de imagem

Nos últimos anos as técnicas de aprendizagem profunda alcançaram um desempenho semelhante ao de especialistas humanos na resolução de tarefas de classificação de doenças pulmonares (ANAYA-ISAZA; MERA-JIMÉNEZ; ZEQUERA-DIAZ, 2021; LITJENS et al., 2017). A IA tem sido fundamental para ampliar a confiabilidade na tomada de decisões.

A proliferação de publicações sobre aprendizagem profunda aplicada a exames de imagem de COVID-19 é impressionante. Uma pesquisa no início de 2022, com as chaves “Deep Learning” e “CT” e “COVID-19” resultou em aproximadamente 10.300 resultados no Google Scholar, 894 na ScienceDirect, 488 na Pubmed, 371 na plos.org. O Google Scholar identificou mais de 2.000 artigos de revisão sobre o tema. Ao trocarmos “CT” por “radiographs” foram encontrados mais de 3.100 resultados e 453 artigos de revisão. Apesar da disponibilidade da informação, existem ainda muitos obstáculos para a ampla aplicação destes algoritmos na prática clínica. Um estudo publicado na Nature Machine Intelligence fez uma revisão sistemática dos novos modelos de aprendizado de máquina para o diagnóstico ou prognóstico de COVID-19 a partir de imagens de raios-X ou CT publicados entre 1 de janeiro de 2020 e 3 de outubro de 2020. A busca identificou 2.212 estudos, dos quais 415 foram incluídos após a triagem inicial e, após a triagem de qualidade, 62 estudos foram incluídos na revisão sistemática. A conclusão é impactante. Nenhum dos modelos identificados é de uso clínico potencial devido a falhas metodológicas e/ou vieses subjacentes (ROBERTS et al., 2021).

Este achado corrobora com nossa percepção. Durante a elaboração deste projeto, os principais problemas identificados nos trabalhos publicados foram, em primeiro lugar, a impossibilidade de acessar o código-fonte, treinar e testar os algoritmos. Isso limitou a possibilidade de replicar os resultados e avaliar os algoritmos desenvolvidos em diferentes conjuntos de dados. Os dados de treinamento e teste nem sempre estavam disponíveis. Em muitos casos as políticas de proteção de dados do paciente impediam a liberação de dados ou haviam interesses comerciais na

ferramenta de software desenvolvida, fornecendo apenas código-fonte parcial. Em segundo lugar, a maioria dos estudos usa um número limitado de imagens, de fontes locais, e, portanto, seus modelos não podem ser diretamente generalizados para outros fenótipos e contextos de regiões geográficas. Existem ainda limitações metodológicas como informações ausentes, uso de conjuntos de dados públicos não confiáveis para treinamento, falta de validação externa dos dados de treinamento e ausência de sensibilidade e robustez dos modelos.

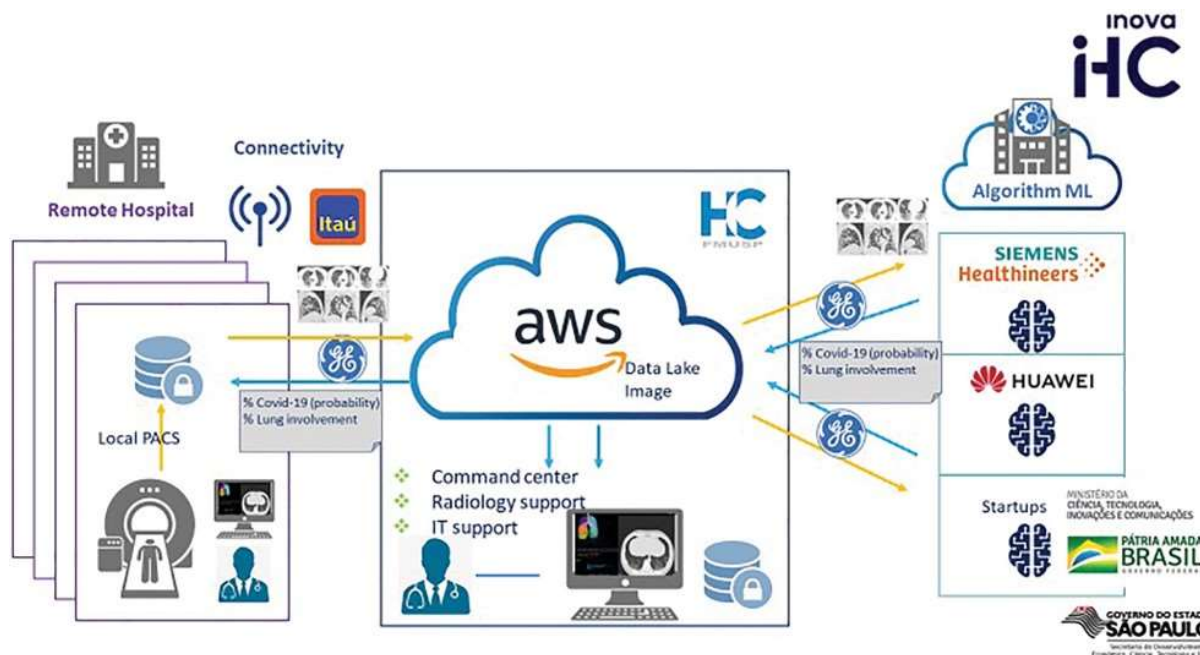
Neste trabalho evitou-se repetir as falhas mais comuns identificadas nos estudos disponíveis e buscou-se avançar o conhecimento necessário para aplicação destes algoritmos à realidade brasileira. A utilização de imagens de pacientes brasileiros foi uma premissa deste projeto. Uma pesquisa bibliográfica nas revistas de maior impacto aponta a existência de poucas publicações que utilizaram bases de dados de imagens oriundas de hospitais locais no desenvolvimento de algoritmos de aprendizagem profunda aplicado ao suporte ao diagnóstico de COVID-19. O trabalho de (Carvalho et al., 2021) utilizou dados de 130 pacientes de dois hospitais no Rio de Janeiro e um de Porto em Portugal, para desenvolver um algoritmo para identificar e quantificar a extensão do envolvimento pulmonar em pacientes com pneumonia por COVID-19. Já o estudo de (DINIZ et al., 2021) desenvolveu um algoritmo para segmentação de lesões de COVID em TC utilizando uma base de apenas 40 pacientes de um hospital do Rio de Janeiro. Uma base intitulada SARS-COV-2 CT-Scan com um conjunto de 2.482 imagens 2D de TC de 60 pacientes de hospitais de São Paulo foi disponibilizada publicamente (ANGELOV; ALMEIDA SOARES, 2020). A base não dispõe de dados clínicos dos pacientes e somente apresenta as imagens em formato JPG. Como os exames de TC geram imagens no formato DICOM e buscamos criar um modelo adequado ao fluxo padrão da prática clínica, esta base não foi utilizada neste trabalho.

Os trabalhos com dados multicêntricos de múltiplos países são os mais ricos pois permitem aos algoritmos captar as diferentes características fenotípicas das populações de geografias distintas. Um outro fator relevante é a captação pelo modelo das interferências geradas por aparelhos de imagem diferentes. A qualidade das imagens de raios-X depende de fatores como qualidade do filme, tipo e estado de conservação dos filtros e colimadores, tempo de exposição e potência (dose), distância da fonte do feixe ao alvo, entre outros, mas também varia com a marca e modelo (ano) da unidade de raio-X. Em particular, a resolução e o contraste podem

variar significativamente entre as unidades (WINSTON et al., 2001). Um trabalho liderado por pesquisadores de Stanford envolvendo 8 países, incluindo 331 pacientes do Brasil, apresentou uma importante experiência de uso de imagens de TC com diferentes fenótipos e aparelhos de distintos fabricantes (LEE et al., 2021). Nesse trabalho foi desenvolvida uma rede neural convolucional de aprendizagem profunda (CNN) 3D de 27 camadas, que usa todo o volume da TC de tórax e classifica os exames como: pneumonia causada por COVID-19, pneumonia não COVID-19 e pulmão normal.

Um projeto de suporte ao diagnóstico utilizando IA muito divulgado no Brasil foi desenvolvido pelo hospital das clínicas de São Paulo em parceria com empresas privadas e apoio do governo. Foi disponibilizada uma plataforma de telemedicina que analisa exames de raio X e tomografias de tórax procurando anomalias nos pulmões para avaliar a probabilidade de contaminação por COVID-19, intitulada RadVid-19. Divulgações públicas atestam que a plataforma recebeu mais de 33 mil acessos de médicos e analisou mais de 24 mil exames de radiografia e tomografia de tórax nos 49 hospitais de todo país cadastrados. Os hospitais cadastrados poderiam enviar suas imagens de TC de tórax para um servidor, onde dois algoritmos de IA retornariam um relatório com uma análise de probabilidade COVID-19 e a extensão do parênquima pulmonar afetado. A plataforma funcionaria 24 horas por dia sem nenhum custo para o usuário e os relatórios seriam entregues em 10 minutos. A arquitetura do sistema, indicando as tecnologias fornecidas pelas empresas é apresentada na Figura 13.

Figura 13. Radvid-19, um sistema nacional de suporte ao diagnóstico de COVID-19 em exames de imagem



Fonte: (TAN et al., 2021)

Em janeiro de 2022 o site do projeto estava indisponível e não conseguimos maiores informações sobre a plataforma. Os algoritmos utilizados pelo programa eram privados e nem o código fonte nem os resultados alcançados foram publicados. Um diferencial relevante deste trabalho de doutorado é a publicação aberta de todos os resultados e a disponibilização gratuita dos algoritmos desenvolvidos.

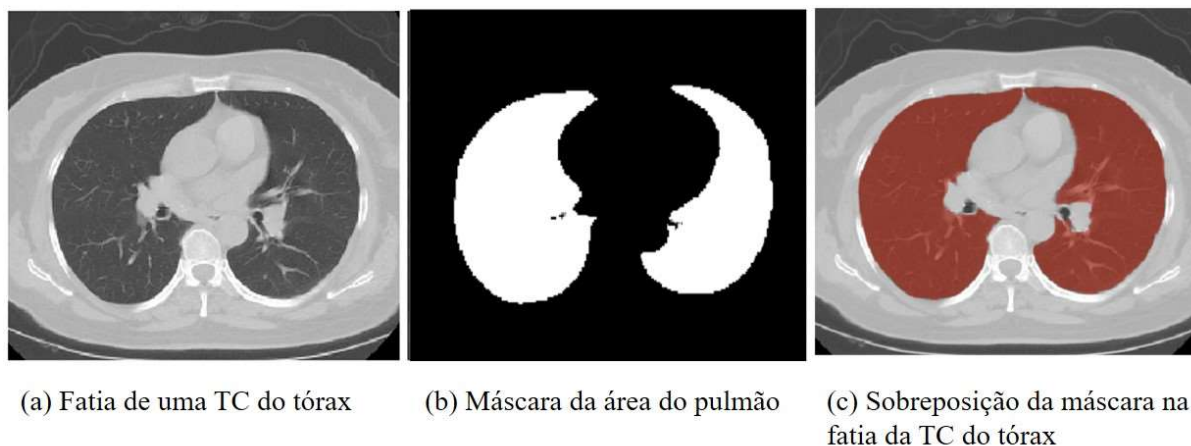
Outra avaliação importante nos trabalhos publicados na literatura recente refere-se à análise das arquiteturas de redes neurais utilizadas, otimizações e pré-processamentos que alcançaram melhor sucesso. O trabalho de (YANG et al., 2021) estudou e comparou várias técnicas de aprendizado profundo aplicadas a imagens médicas de radiografias e tomografia computadorizada do tórax para a detecção de COVID-19 e validou VGG16 e ResNet50 como boas arquiteturas para classificação. Um estudo de revisão por (OZSAHIN et al., 2020) apontou a ampla utilização de redes neurais convolucionais para extração de características relevantes dos exames de TC e observou que a maioria dos modelos de classificação para COVID-19 usa redes pré-treinadas com as arquiteturas DenseNet121, ResNet50 e ShuffleNet V2 sendo bem-sucedidas nos estágios de classificação e a UNet++ tendo um bom desempenho no estágio de segmentação. Este achado foi reforçado no trabalho de revisão de (LIU et al., 2021) que mostrou que muitas CNN 2D e 3D foram usadas para apoiar a identificação de pneumonia, principalmente as arquiteturas Inception, VGG e ResNet.

Os trabalhos relacionados à radiografia de tórax de (APOSTOLOPOULOS; MPESIANA, 2020; WANG; LIN; WONG, 2020) apresentaram exemplos bem sucedidos de algoritmos de classificação de COVID-19, inclusive utilizando técnicas de aprendizagem por transferência, para superar a limitação de dados de treinamento. Um estudo multicêntrico na Coreia do Sul demonstrou o uso de um algoritmo de aprendizagem profunda com resultados comparáveis ao de relatórios de radiologistas em exames de raios-X de tórax, facilitando as decisões sobre a triagem e isolamento de pacientes (JANG et al., 2020).

Nos trabalhos relacionados a TC, uma prática adotada com sucesso foi a segmentação prévia da área do pulmão para posterior entrada das imagens como treinamento do modelo de classificação (SONG et al., 2021). A segmentação do pulmão pode ser feita de diversas formas e influencia diretamente a qualidade do modelo. Neste trabalho foi utilizada uma função de segmentação do NVIDIA Clara Training Framework, parte do pacote de software Clara Imaging, que contém kits de desenvolvimento, bibliotecas aceleradas por unidades de processamento gráfico (GPU) e aplicativos de referência pré-testados para desenvolvedores de aplicativos para a área de saúde e ciências da vida (“NVIDIA Clara Imaging | NVIDIA Developer,” [s.d.]). A segmentação usa uma classificação binária em voxel para a região do pulmão. Cada voxel é previsto como primeiro plano (pulmão) ou plano de fundo. A saída é uma máscara binária, onde ao pulmão é atribuído o valor 1 e ao fundo é atribuído 0.

Como pode ser observado na Figura 14, a sobreposição da máscara (b) sobre a imagem de entrada (a) gera o segmento destacado em vermelho, somente com a área do pulmão (c). O recorte desta imagem é utilizado como entrada da rede neural.

Figura 14. Segmentação do pulmão com Nvidia Clara



Fonte: Adaptado de ("clara_train_covid19_ct_lung_seg | NVIDIA NGC," [s.d.]

Nesta revisão evitou-se comparar os algoritmos desenvolvidos com bases nas métricas comumente apresentadas, como ROC AUC, sensibilidade e especificidade, pois cada algoritmo foi desenvolvido e testado com metodologias e conjuntos de dados diferentes. Elas são válidas para avaliação estática e não necessariamente comparativa dos modelos. Para uma comparação mais correta deveríamos utilizar uma mesma base de teste. Esta avaliação foi possível, neste projeto, com um algoritmo de referência para o modelo de classificação de exames de raio X.

Nota-se que a maioria dos trabalhos de TC utilizou um número limitado de dados de pacientes. Isso restringe a capacidade de generalização dos modelos. Muitos optaram por utilizar técnicas de aprendizagem por transferência e apresentaram resultados interessantes (LI et al., 2021; SHAIK; CHERUKURI, 2021). E embora a prática de aprendizagem por transferência seja comum, o trabalho de (RAGHU et al., 2019) identificou que a aprendizagem por transferência oferece ganhos de desempenho limitados e que arquiteturas muito menores podem ter desempenhos equivalentes aos modelos baseados na base de imagens ImageNet, utilizada na competição ILSVR. Os ganhos mais significativos estão concentrados nas camadas mais baixas e abordagens híbridas de transferência podem ser exploradas. Essa conclusão está alinhada com as experimentações realizadas neste trabalho.

Os dois próximos capítulos contêm publicações que apresentam em detalhes a construção de dois sistemas que utilizam técnicas de aprendizagem profunda para classificação de COVID-19 em exames de imagem de raios X e TC do tórax. A seguir, destacam-se os resultados alcançados e a conclusão com perspectivas de trabalhos futuros.

REFERÊNCIAS

- AGGARWAL, P. et al. COVID-19 image classification using deep learning: Advances, challenges and opportunities. **Computers in Biology and Medicine**, p. 105350, 3 mar. 2022.
- Al, T. et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. **Radiology**, v. 296, n. 2, 2020.
- AKL, E. A. et al. Use of chest imaging in the diagnosis and management of COVID-19: a WHO rapid advice guide. **Radiology**, v. 298, n. 2, p. E63–E69, 2021.
- ANAYA-ISAZA, A.; MERA-JIMÉNEZ, L.; ZEQUERA-DIAZ, M. An overview of deep learning in medical imaging. **Informatics in Medicine Unlocked**, v. 26, p. 100723, 1 jan. 2021.
- ANGELOV, P.; ALMEIDA SOARES, E. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. **MedRxiv**, 2020.
- ANYOHA, R. **The History of Artificial Intelligence**. Disponível em: <<https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>>. Acesso em: 19 dez. 2021.
- APOSTOLOPOULOS, I. D.; MPESIANA, T. A. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. **Physical and Engineering Sciences in Medicine**, v. 43, n. 2, p. 635–640, 1 jun. 2020.
- AREVALO-RODRIGUEZ, I. et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. **PloS one**, v. 15, n. 12, p. e0242958, 2020.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 8, p. 1798–1828, 2013.
- Cadastro Nacional de Estabelecimentos de Saúde**. Disponível em: <<https://cnes.datasus.gov.br/>>. Acesso em: 11 jun. 2022.
- CARVALHO, A. R. S. et al. Estimating COVID-19 Pneumonia Extent and Severity From Chest Computed Tomography. **Frontiers in Physiology**, v. 12, 15 fev. 2021.
- CHOLLET, F. **Deep learning with Python**. [s.l.] Simon and Schuster, 2021.
- CHUNG, M. et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). **Radiology**, v. 295, n. 1, p. 202–207, 2020.

clara_train_covid19_ct_lung_seg | NVIDIA NGC. Disponível em: <https://catalog.ngc.nvidia.com/orgs/nvidia/teams/med/models/clara_train_covid19_ct_lung_seg>. Acesso em: 12 jan. 2022.

COPELAND, B. J. **artificial intelligence.** Disponível em: <<https://www.britannica.com/technology/artificial-intelligence>>. Acesso em: 19 dez. 2021.

COVID-19 - Colégio Brasileiro de Radiologia e Diagnóstico por Imagem. Disponível em: <<https://cbr.org.br/covid-19/>>. Acesso em: 29 dez. 2021.

COZZI, A. et al. Chest x-ray in the COVID-19 pandemic: Radiologists' real-world reader performance. **European journal of radiology**, v. 132, p. 109272, 2020.

DAYAN, I. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. **Nature Medicine** 2021 **27:10**, v. 27, n. 10, p. 1735–1743, 15 set. 2021.

DINIZ, J. O. B. et al. Segmentation and quantification of COVID-19 infections in CT using pulmonary vessels extraction and deep learning. **Multimedia Tools and Applications**, v. 80, n. 19, p. 1, 1 ago. 2021.

ERICKSON, B. J. et al. Machine learning for medical imaging. **Radiographics**, v. 37, n. 2, p. 505–515, 1 mar. 2017.

Estimativas da população residente para os municípios e para as unidades da federação | IBGE. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=31451&t=resultados>>. Acesso em: 11 jun. 2022.

FAN, L.; LIU, S. CT and COVID-19: Chinese experience and recommendations concerning detection, staging and follow-up. **European Radiology**, v. 30, n. 9, p. 5214–5216, 6 set. 2020.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning.** [s.l.] MIT press, 2016.

HAWKINS, D. M. The Problem of Overfitting. **Journal of Chemical Information and Computer Sciences**, v. 44, n. 1, p. 1–12, jan. 2004.

HE, K. et al. **Deep residual learning for image recognition.** Proceedings of the IEEE conference on computer vision and pattern recognition. **Anais...**2016.

HUANG, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. **The Lancet**, v. 395, n. 10223, p. 497–506, fev. 2020.

Imaging the coronavirus disease COVID-19 • healthcare-in-europe.com.

Disponível em: <<https://healthcare-in-europe.com/en/news/imaging-the-coronavirus-disease-covid-19.html>>. Acesso em: 15 dez. 2021.

JANG, S. B. et al. Deep-learning algorithms for the interpretation of chest radiographs to aid in the triage of COVID-19 patients: A multicenter retrospective study. **PLOS ONE**, v. 15, n. 11, p. e0242759, 1 nov. 2020.

KARIM, S. S. A.; KARIM, Q. A. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. **The Lancet**, v. 398, n. 10317, p. 2126–2128, 11 dez. 2021.

KELLY, C. J. et al. Key challenges for delivering clinical impact with artificial intelligence. **BMC Medicine**, v. 17, n. 1, p. 1–9, 29 out. 2019.

KHALID, S.; KHALIL, T.; NASREEN, S. A survey of feature selection and feature extraction techniques in machine learning. **Proceedings of 2014 Science and Information Conference, SAI 2014**, p. 372–378, 7 out. 2014.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, p. 1097–1105, 2012.

KWEE, T. C.; KWEE, R. M. Chest ct in covid-19: What the radiologist needs to know. **Radiographics**, v. 40, n. 7, p. 1848–1865, 1 nov. 2020.

LAINO, M. E. et al. The Applications of Artificial Intelligence in Chest Imaging of COVID-19 Patients: A Literature Review. **Diagnostics 2021, Vol. 11, Page 1317**, v. 11, n. 8, p. 1317, 22 jul. 2021.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 27 maio 2015.

LEE, E. H. et al. Deep COVID DeteCT: an international experience on COVID-19 lung detection and prognosis using chest CT. **NPJ Digital Medicine**, v. 4, n. 1, 1 dez. 2021.

LI, C. et al. Transfer learning for establishment of recognition of COVID-19 on CT imaging using small-sized training datasets. **Knowledge-based systems**, v. 218, 22 abr. 2021.

LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical image analysis**, v. 42, p. 60–88, 1 dez. 2017.

LIU, F. et al. The application of artificial intelligence to chest medical image analysis. **Intelligent Medicine**, v. 1, n. 3, p. 104–117, 1 set. 2021.

MANNA, S. et al. COVID-19: A multimodality review of radiologic techniques, clinical utility, and imaging features. **Radiology: Cardiothoracic Imaging**, v. 2, n. 3, 1 jun. 2020.

NG, M.-Y. et al. Imaging profile of the COVID-19 infection: radiologic findings and literature review. **Radiology: Cardiothoracic Imaging**, v. 2, n. 1, p. e200034, 2020.

NVIDIA Clara Imaging | NVIDIA Developer. Disponível em: <<https://developer.nvidia.com/clara-medical-imaging>>. Acesso em: 12 jan. 2022.

OZSAHIN, I. et al. Review on Diagnosis of COVID-19 from Chest CT Images Using Artificial Intelligence. **Computational and Mathematical Methods in Medicine**, v. 2020, p. 9756518, 2020.

PONTONE, G. et al. **Role of computed tomography in COVID-19**. **Journal of Cardiovascular Computed Tomography**, 2021.

RAGHU, M. et al. Transfusion: Understanding transfer learning for medical imaging. **arXiv preprint arXiv:1902.07208**, 2019.

RAJPURKAR, P. **DEEP LEARNING FOR MEDICAL IMAGE INTERPRETATION**. [s.l.: s.n.].

RIEKE, N. et al. The future of digital health with federated learning. **npj Digital Medicine 2020 3:1**, v. 3, n. 1, p. 1–7, 14 set. 2020.

ROBERTS, M. et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. **Nature Machine Intelligence**, v. 3, n. 3, 2021.

RUBIN, G. D. et al. The role of chest imaging in patient management during the covid-19 pandemic: A multinational consensus statement from the fleischner society. **Radiology**, v. 296, n. 1, 2020.

RUSSELL, S.; NORVIG, P. Artificial intelligence: a modern approach. 2002.

SAMEK, W.; MÜLLER, K. R. Towards Explainable Artificial Intelligence. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 11700 LNCS, p. 5–22, 2019.

SANDRI, T. L. et al. Complementary methods for SARS-CoV-2 diagnosis in times of material shortage. **Scientific Reports |**, v. 11, p. 11899, 2021.

SCHEFFER, M. et al. Demografia médica no Brasil 2020. **São Paulo: FMUSP, CFM**, p. 125, 2020.

SELVARAJU, R. R. et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2017- October, p. 618–626, 22 dez. 2017.

SHAIK, N. S.; CHERUKURI, T. K. Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans. **Computers in biology and medicine**, v. 141, p. 105127, 11 dez. 2021.

SHI, H. et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. **The Lancet Infectious Diseases**, v. 20, n. 4, p. 425–434, abr. 2020.

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

SIMPSON, S. et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA - Secondary Publication. **Journal of Thoracic Imaging**, v. 35, n. 4, 2020a.

SIMPSON, S. et al. Radiological society of North America expert consensus document on reporting chest CT findings related to COVID-19: Endorsed by the society of thoracic radiology, the American college of radiology, and RSNA. **Radiology: Cardiothoracic Imaging**, v. 2, n. 2, 1 abr. 2020b.

SMITH, D. L. et al. A Characteristic Chest Radiographic Pattern in the Setting of the COVID-19 Pandemic. **Radiology: Cardiothoracic Imaging**, v. 2, n. 5, p. e200280, 2020.

SONG, Y. et al. Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) With CT Images. **IEEE/ACM transactions on computational biology and bioinformatics**, v. 18, n. 6, p. 2775–2780, 2021.

SZEGEDY, C. et al. **Going deeper with convolutions**. Proceedings of the IEEE conference on computer vision and pattern recognition. **Anais...**2015.

TAN, B. S. et al. RSNA international trends: A global perspective on the COVID-19 pandemic and radiology in late 2020. **Radiology**, v. 299, n. 1, p. E193–E203, 1 abr. 2021.

TAO, K. et al. The biological and clinical significance of emerging SARS-CoV-2 variants. **Nature Reviews Genetics 2021 22:12**, v. 22, n. 12, p. 757–773, 17 set. 2021.

TORREY, L.; SHAVLIK, J. Transfer learning. In: **Handbook of research on machine learning applications and trends: algorithms, methods, and techniques**. [s.l.] IGI global, 2010. p. 242–264.

WANG, L.; LIN, Z. Q.; WONG, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. **Scientific Reports** 2020 10:1, v. 10, n. 1, p. 1–12, 11 nov. 2020.

WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. Disponível em: <<https://covid19.who.int/>>. Acesso em: 19 jun. 2022.

WINSTON, J. et al. **QUALITY CONTROL RECOMMENDATIONS FOR DIAGNOSTIC RADIOGRAPHY VOLUME 3 RADIOGRAPHIC OR FLUOROSCOPIC**. Radiographic or Fluoroscopic Machines, CRCPD Publication 01-6. **Anais...**2001.

WOLOSHIN, S.; PATEL, N.; KESSELHEIM, A. S. False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications. **New England Journal of Medicine**, v. 383, n. 6, p. e38, 6 ago. 2020.

WORLD HEALTH ORGANIZATION. **Use of chest imaging in COVID-19: a rapid advice guide**. [s.l.: s.n.].

WORLD HEALTH ORGANIZATION. **Recommendations for national SARS-CoV-2 testing strategies and diagnostic capacities**. .

YANG, D. et al. Detection and analysis of COVID-19 in medical images using deep learning techniques. **Scientific Reports** 2021 11:1, v. 11, n. 1, p. 1–13, 4 out. 2021.

YANG, W. et al. **The role of imaging in 2019 novel coronavirus pneumonia (COVID-19)**. **European Radiology**, 2020.

ZHAO, W. et al. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. **American Journal of Roentgenology**, v. 214, n. 5, p. 1072–1077, 2020.

ZHU, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. **New England Journal of Medicine**, v. 382, n. 8, p. 727–733, 20 fev. 2020.

3 Manuscrito 1

Deep Learning Applied to Chest Radiograph Classification—A COVID-19 Pneumonia Experience

Este artigo apresenta Cimatec_XCov-19, um algoritmo de inteligência artificial, baseado em aprendizagem profunda, capaz de apoiar o diagnóstico de pneumonia por COVID-19 em exames de raios-X. O artigo demonstra o processo de desenvolvimento do algoritmo e a realização de testes de performance, incluindo uma comparação direta com outro algoritmo de referência.

O Cimatec_Xcov19 é composto por duas redes neurais, uma capaz de classificar exames de raio X em normais e anormais e outra que identifica os casos suspeitos de pneumonia por COVID-19 entre os exames anormais. A imagem é avaliada simultaneamente nos dois modelos e um resultado suspeito de pneumonia por COVID-19 é sinalizado caso a multiplicação das probabilidades independentes seja maior que 0,5. O algoritmo foi treinado com 44.031 exames e testado sobre um banco de 1053 imagens avaliadas por 2 médicos radiologistas. Cada médico realizou duas avaliações, em momentos diferentes, sobre as imagens anonimizadas.

As principais inovações desenvolvidas no trabalho são: (i) a utilização de um conjunto de dados de treinamento com uma grande quantidade de imagens; (ii) preparação e uso de uma base de testes externa oriunda de um hospital e avaliada por uma junta médica; (iii) disponibilização de todo o código abertamente. Observou-se especial atenção ao rigor científico, seleção e preparação dos dados e adaptação do algoritmo à realidade brasileira. Conclui-se que algoritmos de IA podem apoiar de forma satisfatória a identificação de pneumonia por COVID-19 em radiografias, sendo, portanto, uma alternativa de baixo custo e eficiente no suporte à identificação da doença em locais com pouco acesso a recursos.

Além da participação deste autor e de seu orientador, Dr. Erick Giovani Sperandio Nascimento e de seu co-orientador Dr. Roberto Badaró, o trabalho contou com a participação de membros do Centro de Competência em IA do SENAI CIMATEC, notadamente os pesquisadores Dr. Leandro Machado da UFBA e Dr. Diego Frias da UNEB. Contou também com a colaboração do Dr. Thiago Maia, chefe de pesquisa da rede de hospitais Medsenior, especializada em pacientes da 3ª idade.

Ao final do artigo são identificadas as contribuições de cada autor.

Artigo publicado no periódico MDPI *Applied Sciences*. **2022**, *12*(8), 3712; <https://doi.org/10.3390/app12083712> em 07 de abril de 2022.

O artigo é de acesso livre e distribuído sob os termos e condições da licença *Creative Commons Attribution* (CC BY), disponível em <https://creativecommons.org/licenses/by/4.0/>.

Article

Deep Learning Applied to Chest Radiograph Classification—A COVID-19 Pneumonia Experience

Adhvan Furtado ¹, Leandro Andrade ², Diego Frias ³, Thiago Maia ⁴, Roberto Badaró ⁵
and Erick G. Sperandio Nascimento ^{1,*}

- ¹ Super Computing Center SENAI/CIMATEC, Av. Orlando Gomes, 1845, Piatã, Salvador 41560-010, Brazil; adhvan@fieb.org.br
- ² Escola de Administração, Universidade Federal da Bahia, Avenida Reitor Miguel Calmon s/n Vale do-Canela, Salvador 40110-903, Brazil; leandrojsa@ufba.br
- ³ Department of Natural and Earth Sciences, Universidade do Estado da Bahia, Rua Silveira Martins, 2555, Cabula 41150-000, Brazil; diegofrias@uneb.br
- ⁴ SAMEDIL—Serviços de Atendimento Médico, Rua Pedro Fonseca, 170-Monte Belo, Vitória 29053-280, Brazil; thiago.maia@medsenior.com.br
- ⁵ Instituto SENAI de Inovação em Saúde, Av. Orlando Gomes, 1845, Piatã, Salvador 41560-010, Brazil; badaro@fieb.org.br
- * Correspondence: erick.sperandio@fieb.org.br; Tel.: +55-27-992-799-651

Featured Application: The open-source deep learning algorithm presented in this work can identify anomalous chest radiographs and support the detection of COVID-19 cases. It is a complementary tool to support COVID-19 identification in areas with no access to radiology specialists or RT-PCR tests. We encourage the use of the algorithm to support COVID-19 screening, for educational purposes, as a baseline for further enhancements, and as a benchmark for different solutions. The algorithm is currently being tested in clinical practice in a hospital in Espírito Santo, Brazil.



Citation: Furtado, A.; Andrade, L.; Frias, D.; Maia, T.; Badaró, R.; Nascimento, E.G.S. Deep Learning Applied to Chest Radiograph Classification—A COVID-19 Pneumonia Experience. *Appl. Sci.* **2022**, *12*, 3712. <https://doi.org/10.3390/app12083712>

Academic Editors: Keun Ho Ryu and Nipon Theera-Umporn

Received: 25 February 2022

Accepted: 4 April 2022

Published: 7 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Due to the recent COVID-19 pandemic, a large number of reports present deep learning algorithms that support the detection of pneumonia caused by COVID-19 in chest radiographs. Few studies have provided the complete source code, limiting testing and reproducibility on different datasets. This work presents Cimatec_XCOV19, a novel deep learning system inspired by the Inception-V3 architecture that is able to (i) support the identification of abnormal chest radiographs and (ii) classify the abnormal radiographs as suggestive of COVID-19. The training dataset has 44,031 images with 2917 COVID-19 cases, one of the largest datasets in recent literature. We organized and published an external validation dataset of 1158 chest radiographs from a Brazilian hospital. Two experienced radiologists independently evaluated the radiographs. The Cimatec_XCOV19 algorithm obtained a sensitivity of 0.85, specificity of 0.82, and AUC ROC of 0.93. We compared the AUC ROC of our algorithm with a well-known public solution and did not find a statistically relevant difference between both performances. We provide full access to the code and the test dataset, enabling this work to be used as a tool for supporting the fast screening of COVID-19 on chest X-ray exams, serving as a reference for educators, and supporting further algorithm enhancements.

Keywords: deep learning; COVID-19; chest radiograph

1. Introduction

The exponential spread of COVID-19 in the world poses substantial challenges for public health services. The disease, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), initially identified in December 2019 in Wuhan, China, causes respiratory tract infections and spreads rapidly through contagion between people, thus overburdening health systems worldwide. It is necessary to evaluate the contagion scenarios and

identify as many suspicious cases as possible to define appropriate isolation and treatment strategies [1,2]. Clinically, patients infected with SARS-CoV-2 present fever, cough, dyspnea, muscle aches, and bilateral pneumonia in imaging [3,4]. Even though studies suggest that the Omicron variant has a lower replication competence in human lung, thus reducing the pneumonia occurrence [5], mechanisms for screening and monitoring the evolution of the disease in the lungs are still essential, in the sense that we still do not know how the disease will evolve in the years to come. Imaging in chest radiography or computed tomography (CT) is the most common method to support the diagnosis of pneumonia in symptomatic patients [6]. There are clear recommendations from the WHO (World Health Organization) and the American Radiology Society for the use of imaging only in particular situations, and CT as part of the initial screening stage [7–9]. With the progression of the disease in the patient, characteristic chest radiographic patterns become more evident, which allows using X-ray images to support the disease diagnosis and follow-up.

Even with limited resources, many public and private health systems have X-ray machines distributed throughout the country, which makes chest radiography an accessible, fast, and inexpensive alternative for diagnostic screening. In this scenario, an artificial intelligence (AI) system can be a tool to support radiologists or the medical staff directly in a suspected COVID-19 pneumonia patient, especially in areas where no radiology specialist is available [10], and in situations where there is a higher pressure on the health system from a higher demand caused by an epidemic or pandemic situation.

There are many deep learning (DL) algorithms proposed in the literature to detect COVID-19 in radiographs, the majority based on popular convolutional neural networks (CNN) architectures for image classification, such as VGG, Inception, Xception, and Resnet. These algorithms take benefit from the DL characteristic of automatic feature extraction. Nevertheless, learning the features normally requires training the algorithms with a huge amount of annotated images. For a thorough review, please refer to [11,12].

It is difficult to categorize CXR images for COVID-19. The images have few semantic regions (sparsity) and other pulmonary infections generate similar lesions on the lungs, so there is also an inter-class similarity in the images. Recently, some studies that were based on the VGG-16 architecture proposed new methods to enhance feature extraction in CXR images. The work by [13] adopted a novel approach based on the bag of deep visual words (BoDVW) to classify CXR images. The method removes the feature map normalization step and adds the deep features normalization step on the raw feature maps, preserving the semantics of each feature map that might have importance to differentiating COVID-19 from other forms of pneumonia. This method was improved by [14], proposing a multi-scale BoDVW, exploiting three different scales of the pooling layer's output feature map from a VGG-16 model. The study by [15] used an attention module to capture the spatial relationship between the regions of interest in CXR images. The method produced a classification accuracy of 79.58% in the 3-class problem (COVID vs. No_findings vs. Pneumonia), 85.43% in the 4-class problem (COVID vs. Normal vs. Pneumonia bacteria vs. Pneumonia viral), and 87.49% in the 5-class problem (COVID vs. No_findings vs. Normal vs. Pneumonia bacteria vs. Pneumonia viral).

Despite many algorithms being available for public use, there are still many obstacles to their wide application in clinical practice. A study published in *Nature Machine Intelligence* [16] systematically reviewed publications of machine learning models for the diagnosis or prognosis of COVID-19 from X-ray or CT images that were published between 1 January 2020 and 3 October 2020. The search identified 2212 studies, of which 415 were included after initial screening, and, after a more rigorous quality screening, 62 studies were included in the systematic review. The conclusion is impressive. None of the models identified are of potential clinical use due to methodological flaws and/or underlying biases. Our review also identified fundamental problems that limit the adoption of algorithms in clinical practice. The source code and the training and testing data are rarely publicly available. It is not possible to replicate the results and evaluate the AI algorithm on different datasets. We noticed that usually, this happens because patient data protection policies

prevent the release of data or because there are commercial interests in the developed software tool. Sometimes the researchers provide only part of the source code. In addition, most studies used a limited number of images from local sources and, therefore, their models may not generalize well to other phenotypes and geographic regions' contexts. Many works used unreliable public datasets for training, did not provide external validation or presented deficient model robustness metrics. Our observations are in line with the findings identified in the studies of [16–18]. Table 1 presents the open-source algorithms published in the major peer-reviewed publications to the best of our knowledge. Only two other studies used datasets larger than 25,000 chest X-ray images (CXR) for training, and only one had more than 2000 COVID-19 cases.

Table 1. A partial list of DL algorithms based on COVID-19 radiographs with publicly available code.

Ref.	Objective	Base Model	Training Dataset (# of CXR)	External Validation Dataset (# of CXR)
[19]	Detect common thoracic disease	DenseNet-121	120,702	24,500
[19]	Diagnose COVID-19 and multiclass classification	DenseNet-121	27,825/1571 ¹	China 1899/98 ¹ China 1034 Ecuador 650/132 ¹
[20]	Detect COVID-19 pneumonia	Ensemble of CNN: Densenet-121, Resnet-50, Inception, Inception-Resnet, Xception, EfficientNet-B2	Pre-training: NIH-CXR14 dataset >100,000 Fine-tuning: 14,788/4253 ¹	2214 images/1192 ¹
[21]	Predict COVID-19 severity and progression	VGG-11 and EfficientNet-B0	1834 all COVID-19 patients	475
[22]	Detect COVID-19 cases	COVID-Net CNN	13,975/358 ¹	300/100 ¹
[23]	Detect COVID-19 (3 binary classifiers)	ResNet-50	7406/341 ¹	N/A ²
[24]	Detect COVID-19 and Multiclass Classification	DarkNet-19	1125/125 ¹	N/A ²
This work	Detect COVID-19	Inception-V3	44,031/2917 ¹	1158/13 ¹

¹ COVID-19 infection. ² Did not use external validation. Used 20% of data for testing/5-fold cross-validation.

We avoided repeating the most common flaws identified in the available studies. We carefully prepared and used a large and multi-centric dataset for training the algorithm. We used an external validation dataset with data carefully labeled by two experienced radiologists and benchmarked our algorithm with a well-known algorithm on the same dataset. We sought to not only validate the hypothesis that supervised AI algorithms applied to chest radiographs can be an alternative for supporting COVID-19 detection, but also to share all the details related to the major methodological decisions taken to develop our proposed solution, providing full access to the code and a valuable annotated external test dataset. Thus, the main contributions of our work are:

- The proposal of a new DL system based on the Inception V3 architecture, one that supports the identification of normal and abnormal CXR examinations and the diagnosis of COVID-19.
- The preparation and publication of an annotated CXR dataset with 1158 images. It is an external validation dataset suitable not only for this but also for future works.
- The evaluation of the classification metrics of our algorithm in an external validation dataset and a comparison of the performance with a state-of-art algorithm.
- The guarantee of reproducibility.

2. Materials and Methods

In this work, we present Cimatec_XCOV19, a deep learning system to support the detection of COVID-19 in radiographs. The system is composed of two AI models: one evaluates normal and abnormal examinations, while the second is a binary classifier for being suggestive of COVID-19 or not. Both models are variations of Inception-V3 CNNs [25] trained with pre-processed CXR. Figure 1 shows the system workflow for the evaluation of an image. A CXR image, X , is pre-processed and serves as input for both models

simultaneously. The system evaluates the input image in both CNN independently. They have different box colors in the figure. One model evaluates the probability of image X being abnormal, $P_{abn}(X)$, while the other evaluates the probability of image X being COVID-19, $P_{cov}(X)$. An outcome suggestive of COVID-19 occurs only when the multiplication of the outputs of the two models is greater than 0.5.

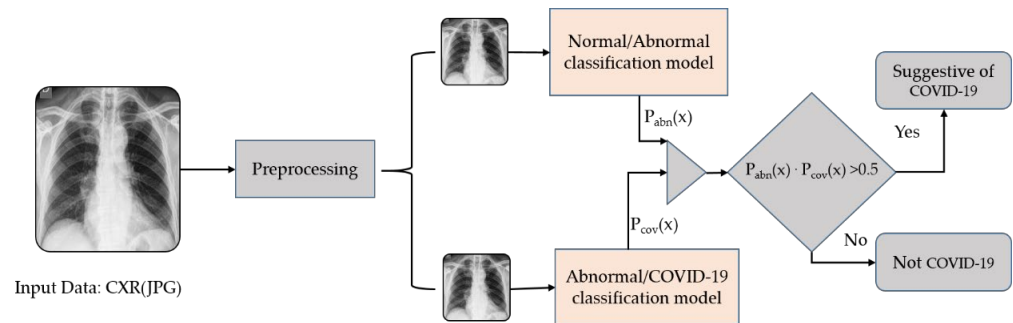


Figure 1. Cimatec_Xcov19 workflow of DL models for COVID-19 classification.

Deep CNNs are often large models and demand much computational power. The widely used Inception-V3 architecture is made of suitably factorized convolutions and aggressive regularization to scale up the networks to efficiently use the available processing capabilities. The model has both symmetric and asymmetric building blocks comprising convolutions layers, average and max pooling operations, concatenation, and fully connected layers. The model uses dropout layers and batch normalization applied to activation inputs. The loss function is a softmax. The Inception architecture innovation is the implementation of inception blocks, which splits the input into different parallel trajectories. There is a concatenation module at the end of the inception blocks to integrate these different paths, as observed in Figure 2. The Supplementary Materials details our network’s architecture, showing the structures in block diagrams. It is possible to notice the modifications they have from a traditional Inception-V3 network.

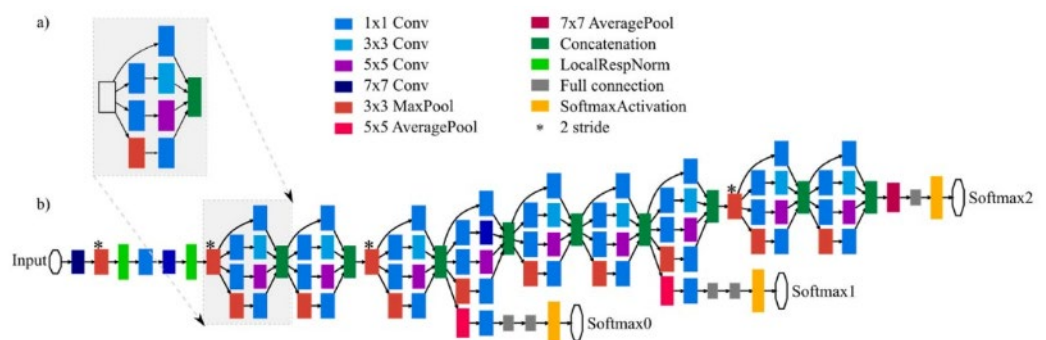


Figure 2. (a) Inception block formed by four convolutional trajectories for the same input. (b) General structure of the network with all the elements. Reprinted with permission from Ref. [26]. 2021, Andrés Anaya-Isaza, Leonel Mera-Jiménez, Martha Zequera-Diaz.

The dataset was prepared by collecting 44,031 examinations from different sources, mainly from public databases and Brazilian and Spanish healthcare institutions. We did a visual inspection of each database and manually excluded out-of-the-context images and those with bad quality. Table 2, below, details the origins of the datasets.

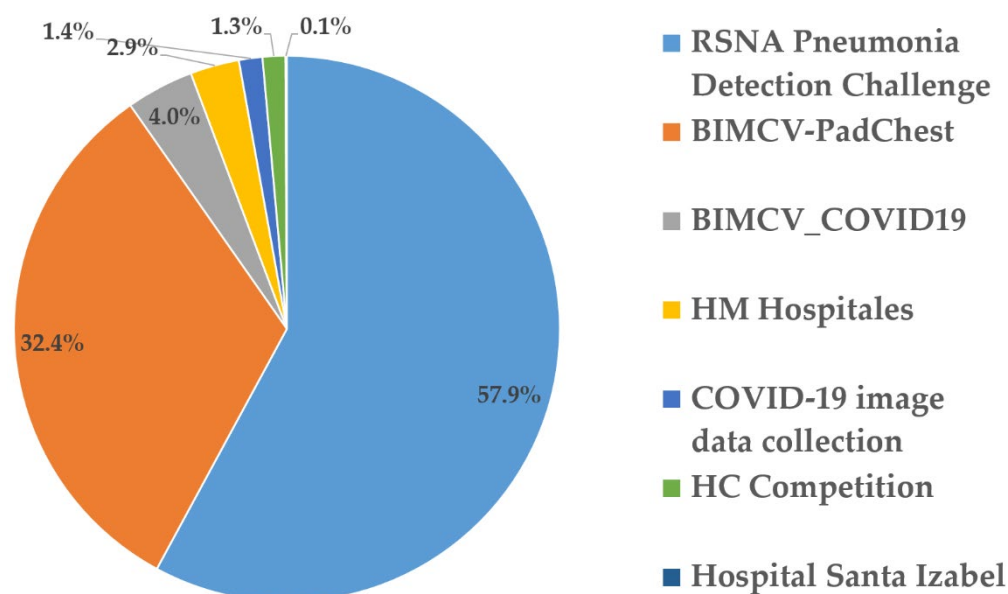
There were multiple image classifications methods in the datasets. The image tags changed according to the origin of the data. For proper use by the models, we reclassified the CXR labels into three categories: (i) normal, (ii) abnormal, but not COVID-19, and (iii) abnormal, and suggestive of COVID-19. Figures 3 and 4 represent the datasets distributions.

Table 2. Datasets description and number of images.

Dataset	Description	# of CXR
RSNA Pneumonia Detection Challenge [27]	Images labeled by the Society for Thoracic Radiology and MD.ai for pneumonia cases found in the chest radiograph database made public by the National Institutes of Health (NIH).	25,497
BIMCV PadChest [28]	Digital Medical Image Bank of the Valencian Community. Images were interpreted and reported by radiologists at Hospital San Juan (Spain) from 2009 to 2017.	14,252
BIMCV COVID-19 [29]	Digital Medical Image Bank of the Valencian Community related to COVID-19 cases.	1762
HM Hospitales	CXR images from patients from the HM Hospitales group in different cities in Spain. Private Dataset.	1277
COVID-19 Image Data Collection [30]	Data was collected from public sources, as well as through indirect collection from hospitals and doctors organized by a researcher from the University of Montreal.	613
HC USP Competition	Images obtained from patients from the HC hospital in São Paulo used for a competition. Private Dataset.	593
Hospital Santa Izabel	Images interpreted and reported by radiologists at Hospital Santa Izabel, Salvador, Bahia, Brazil. Private Dataset.	37

There were 2917 images tagged as COVID-19 (6.7%). This is one of the largest collections of images used to train COVID-19 classifiers, to our knowledge. Before inputting the data into the models, we pre-processed the images for normalization and better feature extraction. A data augmentation process included new images with variations in the gamma contrast, which generated, in total, 132,093 images.

We randomly distributed the dataset to 70% for training, 20% for validation, and 10% for testing, keeping the same distribution of classes from the original dataset. We chose to use a hold-out test dataset instead of doing cross-validation, due to hardware and time constraints. After building a stable system by training and testing it in the general dataset, we did an external validation with a new dataset of CXR from a Brazilian hospital focused on elder people and used explainable AI techniques to show how the algorithms are taking their classification decisions.

**Figure 3.** Dataset breakdown.

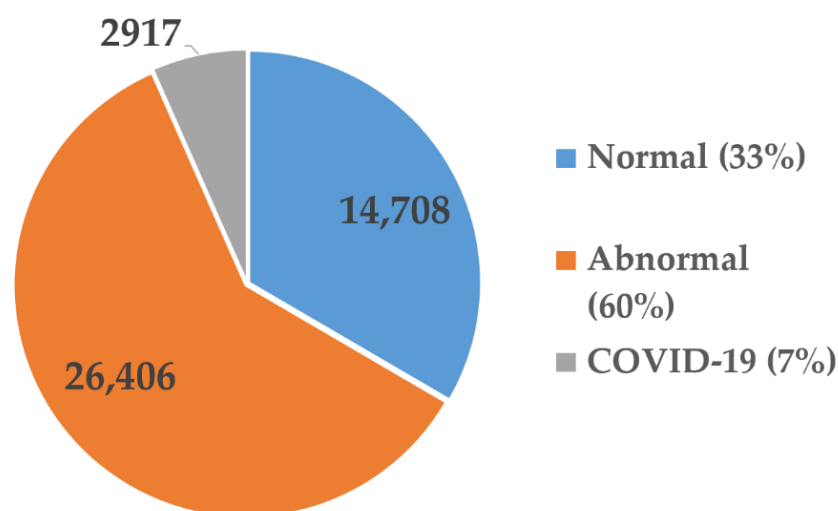


Figure 4. Dataset class distribution.

During the development of the algorithms, we used one shared computing node with four Nvidia GPUs V100 with 32 MB of memory for each.

2.1. Data Pre-Processing

The system input data are CXR in the JPG format. The source of the images is uncertain. They might come in different formats, usually DICOM or JPG. They also may have different resolutions, sizes, and qualities. To establish a standardization process for the input data, facilitate the model feature extraction and learning, and reduce training time, we perform a pre-processing routine [31]. Three preprocess routines correct the edges of the images, cut a bounding box with the lung area, resize it to 299×299 pixels, normalize the data between 0 and 1, and execute a histogram equalization to improve the contrast.

We decided to maintain the standard 299×299 pixels image input size of the Inception V3 architecture. A study on the effect of image resolution on DL in radiography by [32], identified that maximum AUCs were achieved at image resolutions between 256×256 and 448×448 pixels for binary decision networks targeting emphysema, cardiomegaly, hernias, edema, effusions, atelectasis, masses, and nodules. Although the impact of resizing the image in this work is not completely clear, we assumed this resolution had low interference in the feature detection ability of the models.

There are many images with a concentration of pixels in a reduced number of colors, which makes it difficult for the model to identify the inner region of the lung. Therefore, we apply a color histogram equalization to standardize and improve the images, as observed in Figure 5.

To expand the assertiveness of the classification models and their ability for generalization and noise tolerance, we used a technique known as data augmentation. This technique aims to expand the training database of the deep learning models by generating new images from the original dataset, with the intentional introduction of variations in color, brightness contrast, flips, rotations, or spatial distortions. After trying multiple options, we encountered better results when introducing variations in the gamma contrast. In this way, two new images were created from each original image, tripling the training and validation datasets, which generated, in total, 132,093 images.

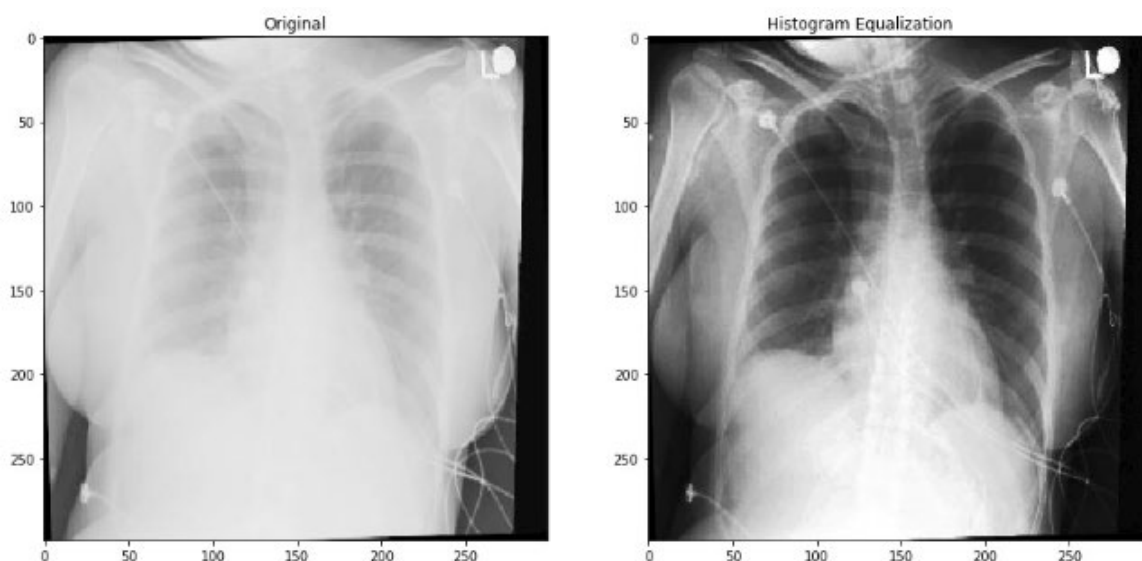


Figure 5. Histogram equalization example.

2.2. External Validation Dataset

The test dataset for external validation has frontal chest radiographs from patients from a hospital in Espírito Santo, Brazil, obtained in the period between July and September 2020, during an acute phase of the COVID-19 pandemic. The retrospective study was approved by the Escola Superior de Ciências da Santa Casa de Misericórdia de Vitória—EMESCAM institutional review board (STU# 34311720.8.0000.5065) and was granted a waiver of written informed consent. Figure 6 shows a diagram with the flow of participants. The study sample consisted of 1,158 images, being 830 (71.68%) females, 328 (28.32%) males, with a mean age of 72.56 years ± 10.02 (standard deviation), and 30 cases (2.59%) with a positive RT-PCR test for COVID-19.

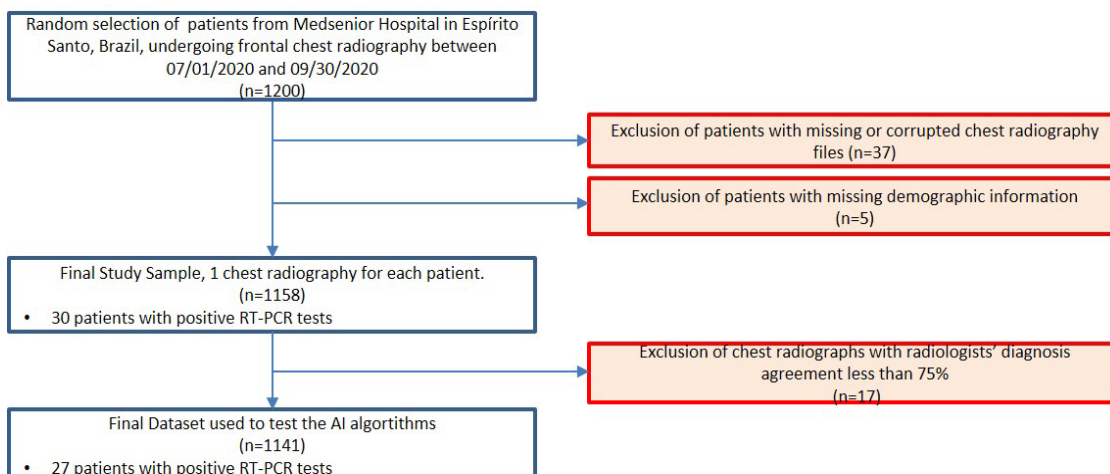


Figure 6. Flowchart for patient inclusion in the external validation dataset.

Independently, two radiologists (henceforth radiologists A and B), certified by the Brazilian Federal Council of Medicine and by the Brazilian Society of Radiology, both with at least 15 years of practical experience, evaluated the exams. The dataset was randomized and anonymized and accessed via a PACS (picture archiving and communication system), where the radiologists could review the images but had no access to any other clinical data, nor to the review of the other radiologist. They analyzed each image twice at different times and orders. Hence, each image received four diagnoses.

The radiologists issued one of the following seven possible diagnoses: (1) normal examination, (2) severe viral infection, (3) moderate viral infection, (4) mild viral infection, (5) severe bacterial infection, (6) moderate bacterial infection, or (7) mild bacterial infection. We only considered valid a diagnosis with at least three concordant analyses. Of the 1158 images, 1082 (93.43%) had 100% agreement, while 59 cases (5.09%) had 75% agreement. Seventeen images (1.46%) had less than 75% agreement and were excluded from the database. Table 3 presents the radiologists' analysis of the dataset.

Table 3. Characteristics of Patients of the External Validation Dataset.

Parameter	Number of Examinations	Age(y)	Sex	Positive RT-PCR Test
All Patients	1158	72.56 ± 10.02	830 female	30
Radiologists' Diagnosis Breakdown				
Lack of Consensus (agreement < 75%)	17	74.35 ± 9.38	10 female	3
Normal	1108	72.32 ± 9.94	802 female	12
Mild Viral Infection	1	71	1 male	1
Moderate Viral Infection	3	78.67 ± 10.01	2 female	3
Severe Viral Infection	10	81 ± 6.55	5 female	10
Mild Bacterial Infection	9	78.11 ± 11.86	6 female	1
Moderate Bacterial Infection	7	78.43 ± 9.81	4 male	0
Severe Bacterial Infection	3	85.67 ± 16.44	2 female	0

We calculated Cohen's kappa coefficient of intraobserver and interobserver agreement [33] with a 5% confidence. The intraobserver analysis of radiologist A showed a kappa of 0.847. From the first sampling to the second sampling, radiologist A changed the diagnosis for 13 images. While for radiologist B, the coefficient was 0.507, changing the diagnosis for 66 images. For the interobserver analysis, in the first round, the radiologists differed in 51 diagnoses; the kappa coefficient was 0.595. It increased to 0.699 in the second round, when they only differed in 33 diagnoses. The kappa coefficient varied between moderate and substantial agreement. A complete table with all 1158 diagnoses is available at [34]

According to the radiologists' agreed diagnosis, 1108 examinations were normal, 19 had a bacterial infection, one had a mild viral infection, and 13 had a moderate or severe viral infection. Interestingly, the 13 cases diagnosed as moderate or severe viral infection correspond to images of patients infected with COVID-19, having tested positive on the RT-PCR test. These results suggest that during a COVID-19 pandemic, it is possible to associate usual diagnoses of moderate and severe viral infection from X-ray examinations with a strong suspicion of COVID-19 infection.

2.3. Benchmark Algorithm

We used the external validation test dataset to evaluate the performance of our AI algorithm and compare it with the results obtained from the same dataset from another public COVID-19 classifier, which we will describe further. We compared the algorithm's indication of examinations suggestive of COVID-19 with the radiologists' diagnoses of moderate or severe viral infection.

We chose the DeepCOVID-XR algorithm as the public COVID-19 classifier for benchmarking. The Image and Video Processing Lab (IVPL) at Northwestern University developed the algorithm and shared the code [20]. The DeepCOVID-XR system is an ensemble of six different CNNs, as shown in Figure 7. It uses the entire chest X-ray image and a cropped image with the lung region as the input. Both images are resized to 224 × 224 and 331 × 331 pixels, which amounts to four smaller input images for each X-ray sample in the dataset. The system sends these images into each of the six different previously validated CNN architectures. A weighted average of the predictions from each model produces a single prediction of COVID-19 for each image.

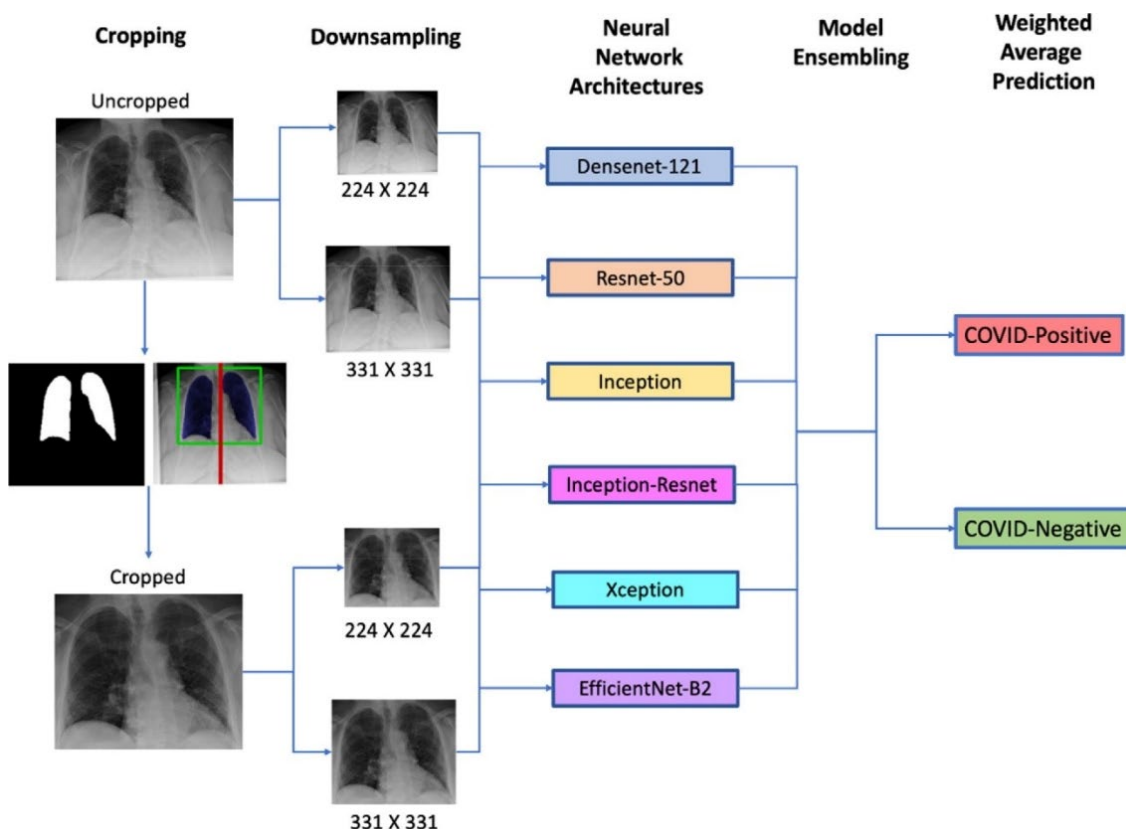


Figure 7. The general architecture of the DeepCOVID-XR deep learning weighted average prediction algorithm [20].

The CNNs were pre-trained on a public dataset with more than 100,000 images before being fitted with images collected from a clinical trial with 14,788 images (4253 positives for COVID-19) using transfer learning. The hold-out test dataset had 2214 images (1192 positives for COVID-19). It generated an 83% accuracy, 75% sensitivity, 93% specificity, and 0.90 AUC ROC (area under curve of receiver operating characteristic).

2.4. Statistical Methods

We calculated the sensitivity and specificity with a confidence interval (CI) of 95% and compared the AUC ROC of the two algorithms with the DeLong test [35]. We used the IBM SPSS 2.8[®] software to calculate Cohen's kappa coefficient and the AUC ROC. For the statistical analysis, we used the following Python libraries: sklearn, scipy, and imbalanced learn [36].

3. Results

The Cimatic_XCOV19 system, presented in this study, comprises two CNNs, one to classify the CXR images as normal or abnormal and the other to classify the CXR images as abnormal or suggestive of COVID-19.

3.1. Algorithm Evaluation

To prepare the normal and abnormal classification model, we randomly distributed 70% of the data for training, 20% for validation, and 10% for testing, keeping the same distribution of classes from the original dataset. Figure 8 shows the confusion matrix for the testing dataset.

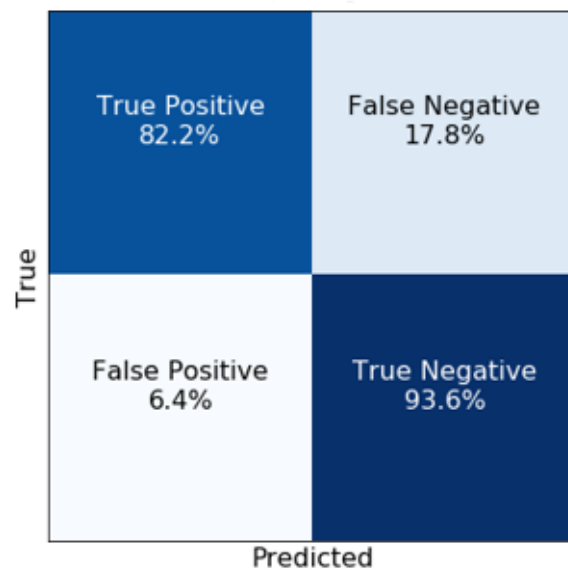


Figure 8. Normal/Abnormal model Confusion Matrix for the test dataset.

The model had, overall, an F1 score of 94%, an accuracy of 91%, a sensitivity of 94%, a specificity of 94%, and a precision of 94%. The AUC ROC and PRC (precision-recall curve) curves shown in Figures 9 and 10 complement the results that demonstrate the good performance of this approach. The model has an excellent fit as a screening tool for abnormal images since it generates few false negatives.

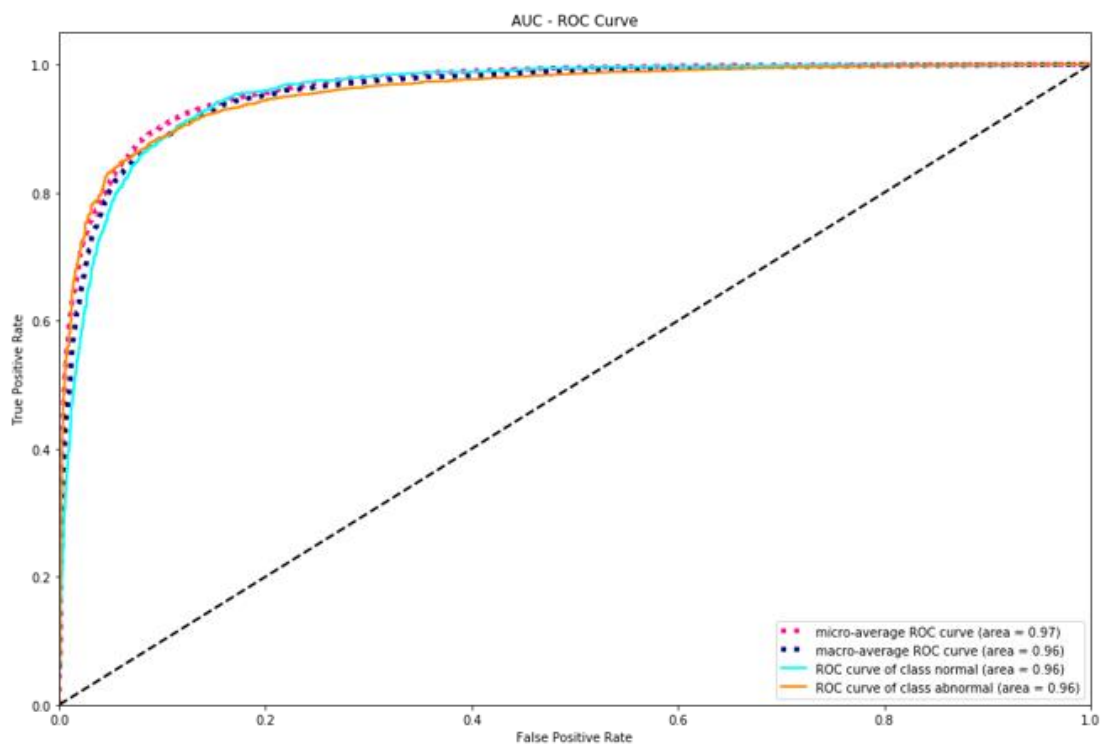


Figure 9. AUC ROC graph for the Normal/Abnormal classifier.

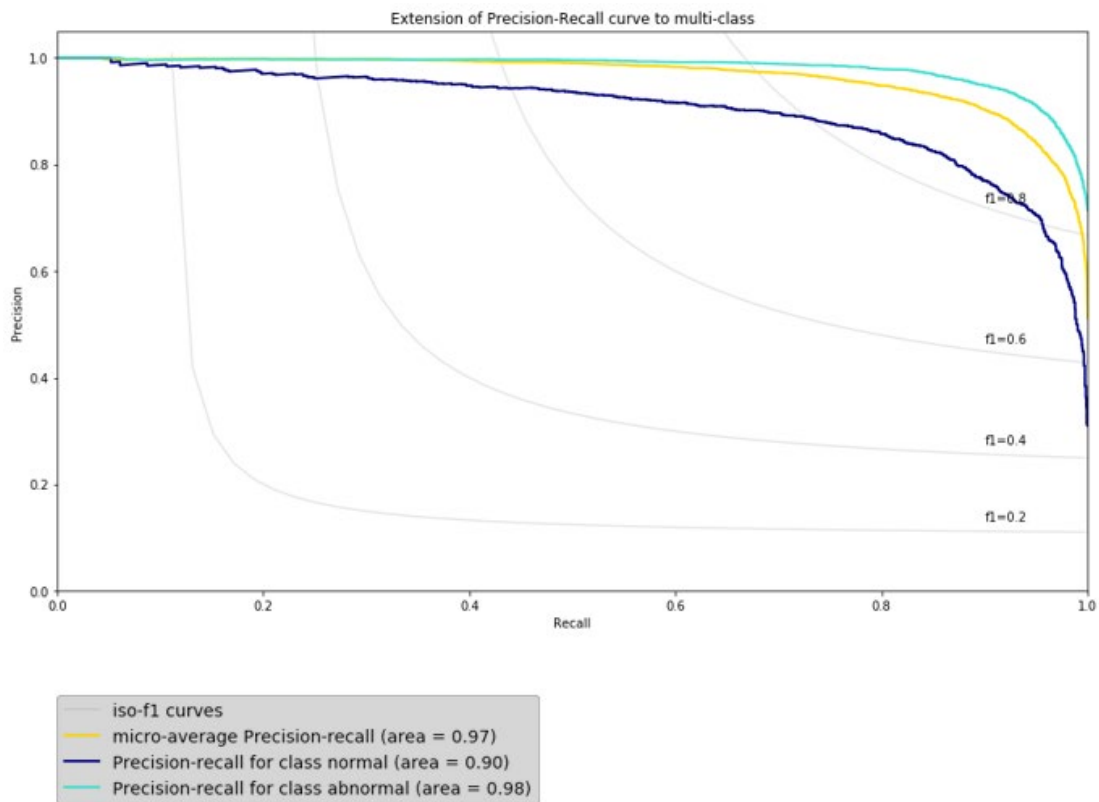


Figure 10. AUC PRC graph for the Normal/Abnormal classifier.

Table 3 CNN. We trained it to differentiate an abnormal CXR from a CXR suspicious of COVID-19. We collected the training data from multiple databases, looking to enhance variability, avoiding bias toward a specific one. We used 8493 images, being 70% for training, 20% for validation, and 10% for testing. As observed in the confusion matrix in Figure 11, the model wrongly labeled images as Abnormal in only 3.5% of the COVID-19 image examinations.

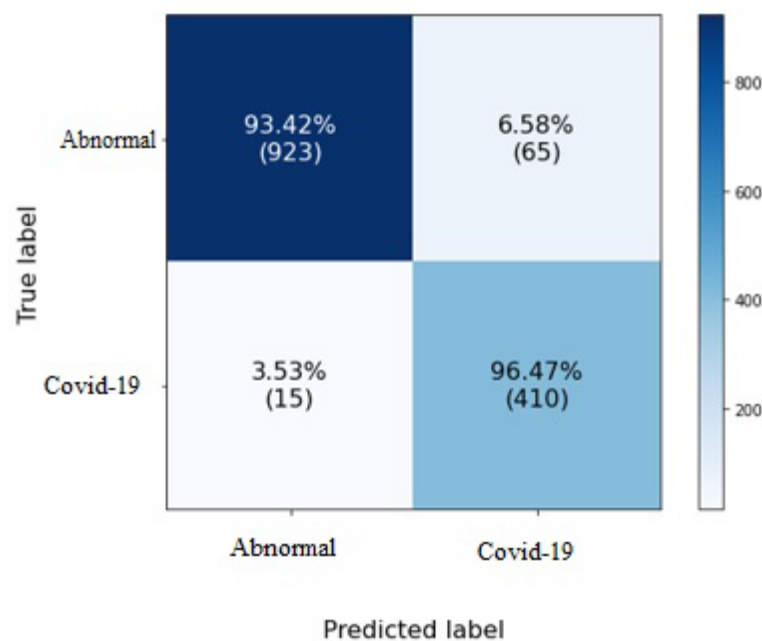


Figure 11. Confusion Matrix for COVID-19/Abnormal classification model.

The model had an average F1 score of 94%, an accuracy of 94%, a sensitivity of 93%, and a specificity of 96%, which minimizes the possibility that an anomalous image of a patient with COVID-19 is considered non-COVID-19. To complement the results that demonstrate the excellent performance of this module, Figures 12 and 13 show the AUC ROC and PRC curves.

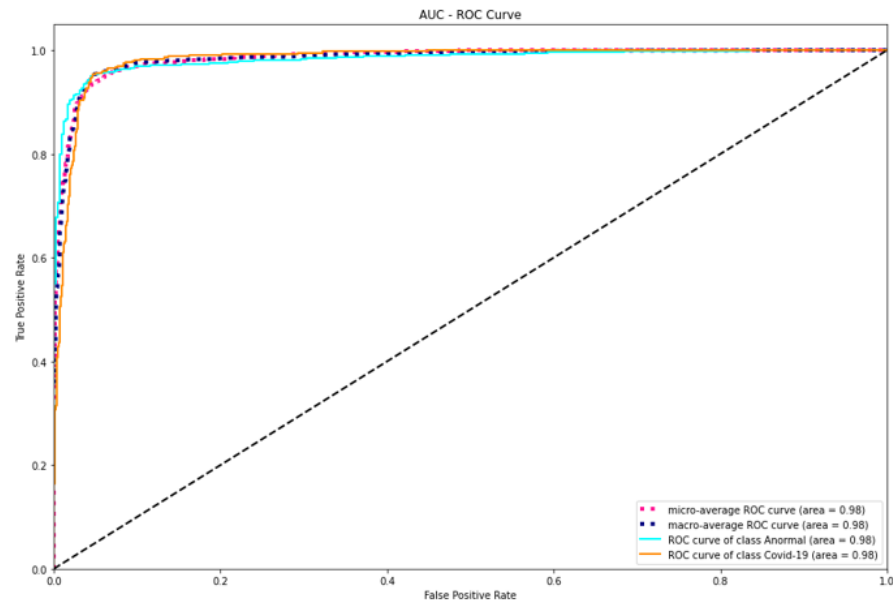


Figure 12. AUC graph for the COVID-19/Abnormal classifier.

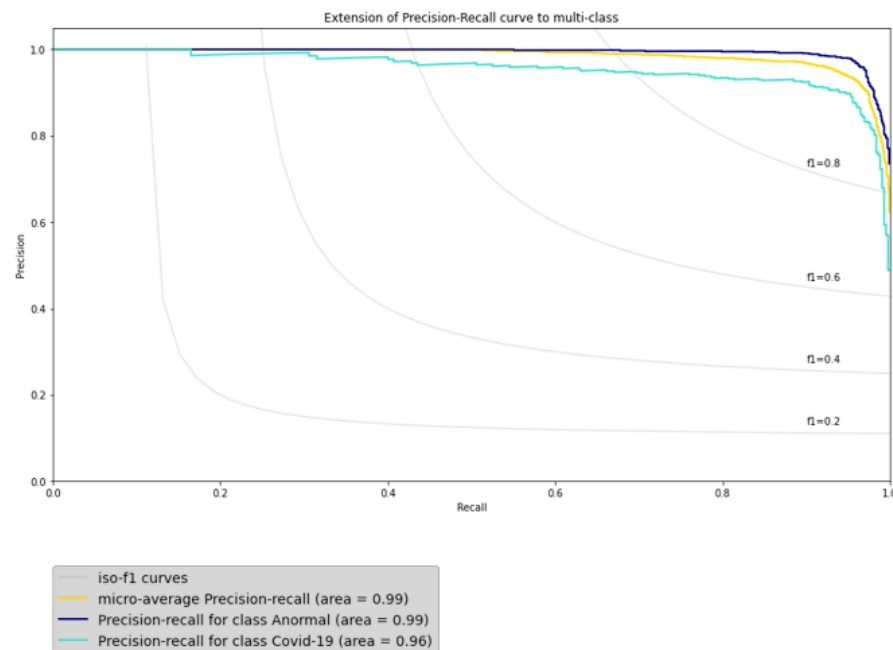


Figure 13. PRC graph for the COVID-19/Abnormal classifier.

We compared the results of Cimatec_XCOV19 with the published result of the algorithms identified in Table 1. This comparison is only a rough reference, as some of those algorithms were multiclass classifiers and all of them were trained and tested on different datasets. Table 4 shows the results.

Table 4. Table comparing Cimatec_XCOV19 metrics with other algorithms.

Ref.	Name	Accuracy	Sensitivity	Specificity	ROC	PRC
[19]	Wang et al.	N/A ¹	0.93	0.87	0.97	N/A
[20]	DeepCOVID-XR	0.90	0.75	0.93	0.83	N/A
[22]	COVID-Net	0.93	0.91	N/A	N/A	N/A
[23]	Narin et al (Resnet50)	1	1	1	N/A	N/A
[24]	DarkCovidNet	0.98	0.95	0.91	N/A	N/A
This work	Cimatec_XCOV19	0.94	0.93	0.96	0.98	0.96

¹ N/A—Not available results.

3.2. External Validation

We used the 1141 CXR exams with a consensus diagnosis, detailed in Table 3, to perform an external validation. We also used this dataset to compare the performance of our algorithm with the DeepCOVID-XR published open-source algorithm. From the list in Table 1, it was the best fit because it was trained using large datasets, performed external validations, and had rigorous statistical analysis. Another good option would be the algorithm developed by [19] but it missed code documentation.

To evaluate the performance of both AI algorithms, we compared the algorithm’s indication of examinations suggestive of COVID-19 with the radiologists’ diagnoses of moderate or severe viral infection. We expected a worse performance by the AI algorithms than those presented in previous studies, given the variances between the patients’ phenotypes present in the training dataset from those present in the external validation dataset as well as the differences in X-ray images. The quality of X-ray images depends on factors, such as the film quality, type, and the state of the conservation of filters and collimators, exposure time and power (dose), the distance from the beam source to the target, among others [37], but it also varies with the brand and model (year) of the X-ray unit. In particular, resolution and contrast can vary significantly between units. For this reason, it is essential to address the ability of a trained AI to identify patients with COVID-19 using X-ray images obtained with the equipment available in each region. Figure 14 shows the confusion matrix for both algorithms.

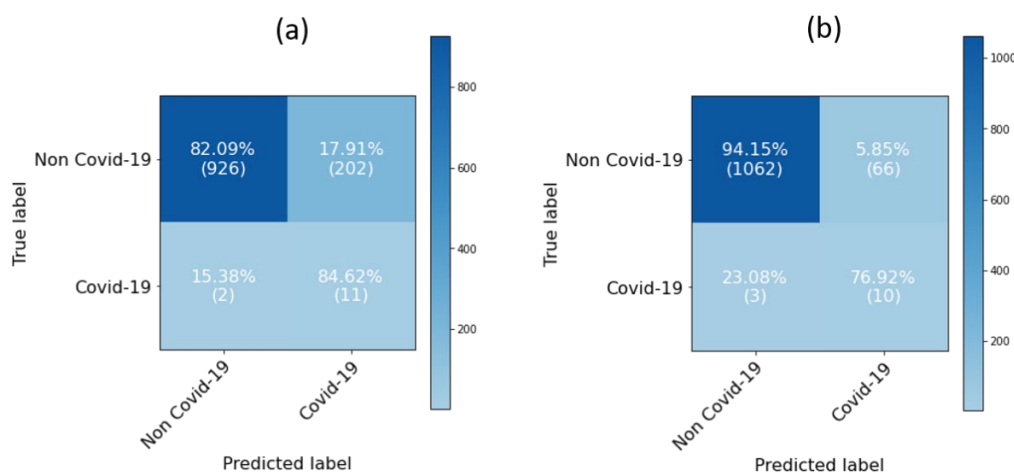


Figure 14. Confusion matrices for (a) CIMATEC_XCOV19; (b) DeepCOVID-XR.

The Cimatec_XCOV19 model had a sensitivity of 0.85 (95% CI, 0.54 to 0.97). Only two examinations were false negatives from the 13 abnormal examinations. Specificity was 0.82 (95% CI, 0.80 to 0.84) and the AUC ROC was 0.92 (95% CI, 0.84 to 1). The DeepCOVID-XR had a slightly worst sensitivity of 0.77 (95% CI, 0.46 to 0.94) with three false negatives, but it had a lower false-positive rate, generating a specificity of 0.94 (95% CI, 0.93 to 0.95) and a ROC AUC of 0.97 (95% CI, 0.93 to 0.999). Table 5 presents the algorithms’ performance in

the external validation dataset and the performance in the test dataset used in their initial training (previous performance). Both algorithms generalized well for the new dataset.

Table 5. Comparison of metrics between the Cimatec_XCOV19 and DeepCOVID-XR algorithms.

Metrics for “Suggestive of COVID-19 Infection”	Cimatec_XCOV19 Performance on the External Validation Dataset	Cimatec_XCOV19 Previous Performance	DeepCOVID-XR Performance on the External Validation Dataset	DeepCOVID-XR Previous Performance
Sensitivity	0.85	0.93	0.77	0.75
Specificity	0.82	0.96	0.94	0.93
Accuracy	0.82	0.94	0.94	0.83
AUC ROC	0.93	0.98	0.97	0.90
AUC PRC	0.48	0.96	0.7	NA

The DeepCOVID-XR improved its performance in the external validation dataset, confirming the ability to generalize well for images from different regions. We notice a performance decrease in the Cimatec_XCOV19 algorithm specificity and accuracy. Interestingly, there was an increase in sensitivity. Although there is a high number of false positives, it has few false negatives, confirming the algorithm as a good screening tool. As observed in Figure 15, according to the results of DeLong’s test of AUC ROC, $z = -0.96$ and $p = 0.34$, we can accept the null hypothesis and conclude that there is no statistically significant difference between the two AUCs.

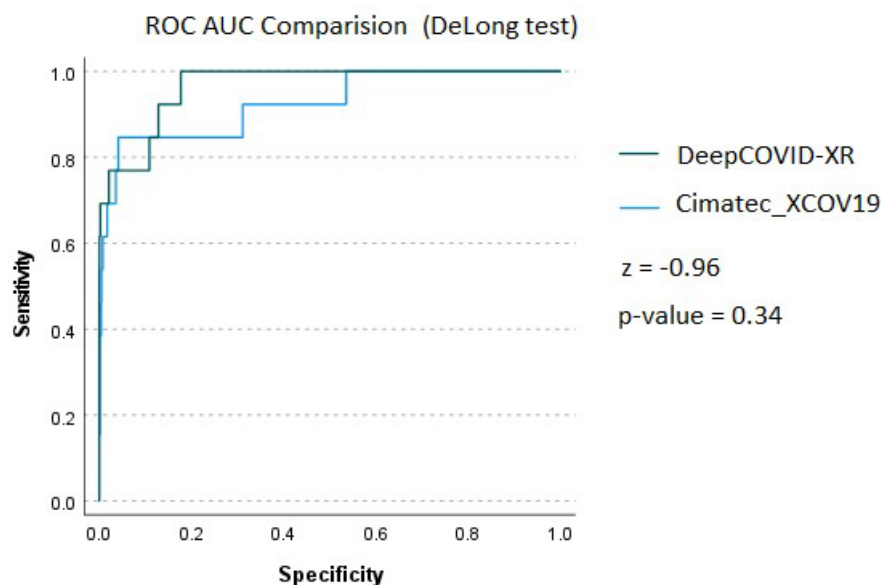


Figure 15. AUC ROC comparison using DeLong’s test.

3.3. Explainability of the AI Models

We asked a radiologist to highlight the findings in four CXRs from the external validation dataset. We compared his findings with the features extracted by the algorithms. Figure 16 provides gradient-weighted class activation mapping heat maps (Grad-CAM) of feature importance for the most representative images from each class of the algorithm’s predictions, thus helping to interpret and explain how each of the AI models performed their predictions. Figure 16a shows the heat maps for the CXR of a male patient, 73 years. It is a true-positive situation for both algorithms. The image label is suggestive of COVID-19. The bounding box highlights infiltrates, and both algorithms classified the image correctly as positive for COVID-19. Figure 16b is a false-positive situation for a 75 years old female patient. Both algorithms incorrectly identified COVID-19 findings in a patient with a moderate bacterial infection. Bounding boxes highlight infiltrates, cardiomegaly,

and atelectasis. In Figure 16c, both algorithms correctly did not identify COVID-19 in a normal examination. The patient is female, 58 years old; Figure 16d shows a false-negative situation. Both algorithms failed to identify infiltrates characteristic of viral infection. Bounding boxes highlight cardiomegaly and infiltrates. The patient is female, 83 years old. There are differences in the images' background color and size because the two AI algorithms use different image pre-processing algorithms.

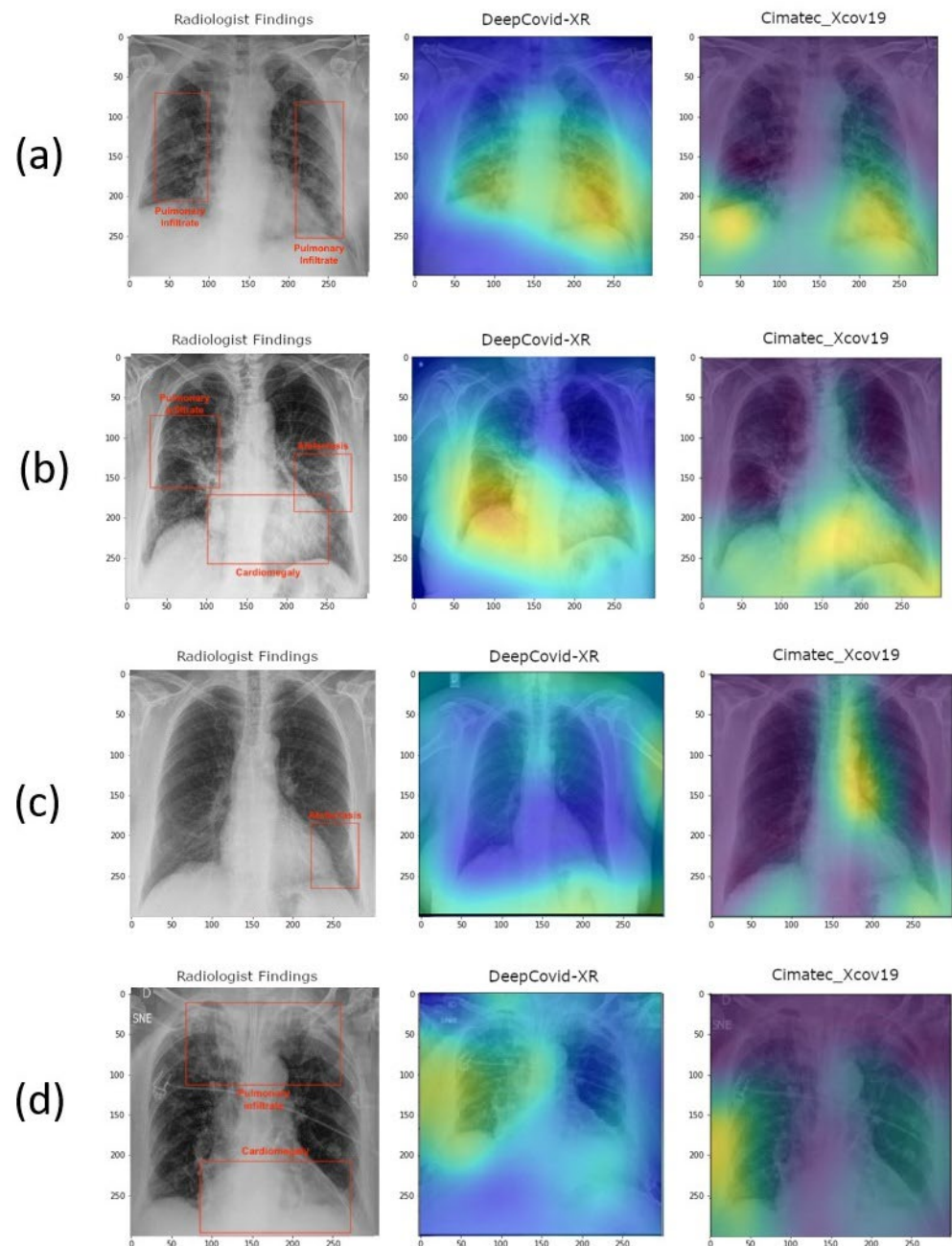


Figure 16. Heat maps for (a) True positive. Bounding box highlights infiltrates. The patient is male, 73 years old; (b) False positive. Bounding boxes highlight infiltrates, cardiomegaly, and atelectasis. The patient is female, 75 years old; (c) True negative. Bounding box highlights infiltrates. The patient is female, 58 years old; (d) False negative. Bounding boxes highlight cardiomegaly and infiltrates. The patient is female, 83 years old.

4. Discussion

The worldwide most available radiographic method to explore lung lesions is still the X-ray examination [38]. In addition, hospitalized patients in intensive care units with suspected COVID-19 pneumonia usually cannot be transported to the radiological centers in the same hospital, however, an X-ray image examination can routinely be performed on the bed of patients. Herein, we detailed the development of a new Inception-V3 based CNN system to support the identification of COVID-19 pneumonia using a chest radiograph. We examined the performance of the algorithms using a dataset from patients treated by a hospital in Espírito Santo, Brazil, during an acute phase of the pandemic and compared it with one previously published algorithm. This study validated in a controlled dataset that the two AI algorithms, Cimatec_XCOV19 and DeepCOVID-XR have, respectively, a specificity of 0.82 and 0.94, a sensitivity of 0.85 and 0.77, and an AUC ROC of 0.92 and 0.97. The performance of both algorithms is good enough to consider them reasonable tools for supporting COVID-19 pneumonia screening. The models generated too many false positives, reinforcing the limitations of the AI systems as a sole diagnostic tool for COVID-19.

The generalization of different datasets is a known problem in AI [39]. This result also reinforces the need for better techniques to adapt the algorithm to the characteristics of new datasets. Advances in the performance of both algorithms might foster the adoption of such systems in scale. In order to facilitate future works and support the development of new AI algorithms in this area, we made all the code freely available [34]. The external validation dataset with labels is also publicly available. They are a good source of images for testing and training new algorithms. The algorithm serves well for educational purposes. We believe that medical staff, under intense work pressure in a pandemic situation, can use the algorithm to help fast screening of COVID-19 cases.

One limitation of this study was the age of the population in the external validation dataset. All patients were older than 50 years and the average age was over 72 years. On one hand, this may limit the ability of the model to extrapolate the analysis to different age groups. Some patients had previous alterations in the chest, though with normal diagnosis. This might represent a bias and could lead to some misclassification of the AI algorithms. Despite this, when we consider that elderly people can be more impacted by COVID-19, these results show that these solutions can be of great help during new COVID-19 pandemic emergencies. Furthermore, all of this knowledge, methodology, and source code can be easily applied and adapted to new eventual pandemic situations, by using transfer learning with new data from CXR exams.

The importance of CXR exams is evident as an alternative for supporting COVID-19 fast screening, especially to identify severe cases, as there might be no findings on CXR exams in mild or early-stage COVID-19 patients. AI algorithms can support the detection of pneumonia caused by COVID-19 in chest radiographs, as they are fast, simple, cheap, safe, and a ubiquitous tool for the management of COVID-19 patients. In the absence of a radiologist specialist, Cimatec_XCOV19 and DeepCOVID-XR AI systems might be good tools to support the detection of COVID-19. Future studies should explore other freely available AI models, test new feature extraction techniques, and use the indications of Grad-CAM and other explainable AI techniques to understand and enhance the actual classification algorithms. Cimatec_XCOV19 is now under controlled testing in a hospital in Espírito Santo, Brazil. Feedback from clinical practice will be paramount to evolving the algorithm and mitigating adoption risks.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12083712/s1>, Figure S1: Cimatec_XCOV19 normal/abnormal classification model block diagram. Figure S2: Cimatec_XCOV19 COVID-19/abnormal classification model block diagram.

Author Contributions: Conceptualization, A.F. and E.G.S.N.; data curation, A.F., L.A. and T.M.; formal analysis, D.F. and E.G.S.N.; funding acquisition, A.F. and E.G.S.N.; investigation, A.F. and L.A.; methodology, A.F., D.F. and E.G.S.N.; project administration, A.F. and E.G.S.N.; resources, T.M.; software, A.F. and L.A.; supervision, A.F. and E.G.S.N.; validation, R.B. and E.G.S.N.; writing—original draft, A.F.; writing—review and editing, A.F., L.A., D.F., T.M., R.B. and E.G.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ABDI, SENAI, EMBRAPPII, REPSOL SINOPEC BRASIL grant “Missão contra a COVID-19 do Edital de Inovação para a Indústria”.

Institutional Review Board Statement: The retrospective study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of “Escola Superior de Ciências da Santa Casa de Misericórdia de Vitória—EMESCAM” (STU# 34311720.8.0000.5065 07/08/2020) and was granted a waiver of written informed consent.

Informed Consent Statement: Patient consent was waived in accordance with the evaluation of the Institutional Review Board considering that researchers undertake to maintain confidentiality, not disclosing the names of the participants and, using codes to identify the data generated by them to avoid violating participant privacy.

Data Availability Statement: The model’s source code, the external validation dataset with 1158 CXR images, and a complementary file sheet with the radiologists’ analysis are freely available on the research group GitHub page at <https://github.com/CRIA-CIMATEC/covid-19> (accessed on 20 February 2022).

Acknowledgments: We gratefully acknowledge the support of SENAI CIMATEC AI Reference Center and the SENAI CIMATEC/NVIDIA AI Joint Center for scientific and technical support, the SENAI CIMATEC Supercomputing Center for Industry Innovation for granting access to the necessary hardware and technical support, Repsol Sinopec Brasil, ABDI, SENAI and EMBRAPPII for providing the funding for this research, HP Brazil for providing support, and Hospital Santa Izabel, MedSenior and HM Hospitales for providing data for this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [[CrossRef](#)]
2. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, X.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [[CrossRef](#)]
3. Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* **2020**, *296*, E32–E40. [[CrossRef](#)]
4. Yang, W.; Sirajuddin, A.; Zhang, X.; Liu, G.; Teng, Z.; Zhao, S.; Lu, M. The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). *Eur. Radiol.* **2020**, *30*, 4874–4882. [[CrossRef](#)]
5. Hui, K.P.Y.; Ho, J.C.W.; Cheung, M.-C.; Ng, K.-C.; Ching, R.H.H.; Lai, K.-L.; Kam, T.T.; Gu, H.; Sit, K.-Y.; Hsin, M.K.Y.; et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature* **2022**, *603*, 715–720. [[CrossRef](#)]
6. Pontone, G.; Scafuri, S.; Mancini, M.E.; Agalbato, C.; Guglielmo, M.; Baggiano, A.; Muscogiuri, G.; Fusini, L.; Andreini, D.; Mushtaq, S.; et al. Role of computed tomography in COVID-19. *J. Cardiovasc. Comput. Tomogr.* **2020**, *15*, 27–36. [[CrossRef](#)]
7. Akl, E.A.; Blažić, I.; Yaacoub, S.; Frija, G.; Chou, R.; Appiah, J.A.; Fatehi, M.; Flor, N.; Hitti, E.; Jafri, H.; et al. Use of Chest Imaging in the Diagnosis and Management of COVID-19: A WHO Rapid Advice Guide. *Radiology* **2021**, *298*, E63–E69. [[CrossRef](#)]
8. Rubin, G.D.; Ryerson, C.J.; Haramati, L.B.; Sverzellati, N.; Kanne, J.; Raoof, S.; Schluger, N.W.; Volpi, A.; Yim, J.-J.; Martin, I.B.K.; et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Radiology* **2020**, *296*, 172–180. [[CrossRef](#)]
9. Simpson, S.; Kay, F.U.; Abbara, S.; Bhalla, S.; Chung, J.H.; Chung, M.; Henry, T.S.; Kanne, J.P.; Kligerman, S.; Ko, J.P.; et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA—Secondary Publication. *J. Thorac. Imaging* **2020**, *35*, 219–227. [[CrossRef](#)]
10. 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. *Lancet Infect. Dis.* **2020**, *20*, 425–434. [[CrossRef](#)]

11. Rahman, S.; Sarker, S.; Al Miraj, A.; Nihal, R.A.; Haque, A.K.M.N.; Al Noman, A. Deep Learning–Driven Automated Detection of COVID-19 from Radiography Images: A Comparative Analysis. *Cogn. Comput.* **2021**. [[CrossRef](#)]
12. Abelaira, M.D.C.; Abelaira, F.C.; Ruano-Ravina, A.; Fernández-Villar, A. Use of Conventional Chest Imaging and Artificial Intelligence in COVID-19 Infection. A Review of the Literature. *Open Respir. Arch.* **2021**, *3*, 100078. [[CrossRef](#)]
13. Sitaula, C.; Aryal, S. New bag of deep visual words based features to classify chest x-ray images for COVID-19 diagnosis. *Health Inf. Sci. Syst.* **2021**, *9*, 1–12. [[CrossRef](#)]
14. Sitaula, C.; Shahi, T.B.; Aryal, S.; Marzbanrad, F. Fusion of multi-scale bag of deep visual words features of chest X-ray images to detect COVID-19 infection. *Sci. Rep.* **2021**, *11*, 1–12. [[CrossRef](#)]
15. Sitaula, C.; Hossain, M.B. Attention-based VGG-16 model for COVID-19 chest X-ray image classification. *Appl. Intell.* **2020**, *51*, 2850–2863. [[CrossRef](#)]
16. Roberts, M.; Covnet, A.; Driggs, D.; Thorpe, M.; Gilbey, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.I.; Etmann, C.; McCague, C.; et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
17. Born, J.; Beymer, D.; Rajan, D.; Coy, A.; Mukherjee, V.V.; Manica, M.; Prasanna, P.; Ballah, D.; Guindy, M.; Shaham, D.; et al. On the role of artificial intelligence in medical imaging of COVID-19. *Patterns* **2021**, *2*, 100269. [[CrossRef](#)]
18. López-Cabrera, J.D.; Orozco-Morales, R.; Portal-Díaz, J.A.; Lovelle-Enriquez, O.; Pérez-Díaz, M. Current limitations to identify COVID-19 using artificial intelligence with chest X-ray imaging. *Health Technol.* **2021**, *11*, 411–424. [[CrossRef](#)]
19. Wang, G.; Liu, X.; Shen, J.; Wang, C.; Li, Z.; Ye, L.; Wu, X.; Chen, T.; Wang, K.; Zhang, X.; et al. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images. *Nat. Biomed. Eng.* **2021**, *5*, 509–521. [[CrossRef](#)]
20. Wehbe, R.M.; Sheng, J.; Dutta, S.; Chai, S.; Dravid, A.; Barutcu, S.; Wu, Y.; Cantrell, D.R.; Xiao, N.; Allen, B.D.; et al. DeepCOVID-XR: An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large U.S. Clinical Data Set. *Radiology* **2021**, *299*, E167–E176. [[CrossRef](#)]
21. Jiao, Z.; Choi, J.W.; Halsey, K.; Tran, T.M.L.; Hsieh, B.; Wang, D.; Eweje, F.; Wang, R.; Chang, K.; Wu, J.; et al. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: A retrospective study. *Lancet Digit. Health* **2021**, *3*, e286–e294. [[CrossRef](#)]
22. Wang, L.; Lin, Z.Q.; Wong, A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci. Rep.* **2020**, *10*, 19549. [[CrossRef](#)]
23. Narin, A.; Kaya, C.; Pamuk, Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Anal. Appl.* **2021**, *24*, 1207–1220. [[CrossRef](#)]
24. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **2020**, *121*, 103792. [[CrossRef](#)]
25. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826. [[CrossRef](#)]
26. AAnaya-Isaza, A.; Mera-Jiménez, L.; Zequera-Díaz, M. An overview of deep learning in medical imaging. *Informatics Med. Unlocked* **2021**, *26*, 100723. [[CrossRef](#)]
27. RSNA Pneumonia Detection Challenge. 2018. Available online: <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018> (accessed on 12 February 2022).
28. Bustos, A.; Pertusa, A.; Salinas, J.-M.; de la Iglesia-Vayá, M. PadChest: A large chest x-ray image dataset with multi-label annotated reports. *Med Image Anal.* **2020**, *66*, 101797. [[CrossRef](#)]
29. De la Iglesia Vayá, M.; Saborit, J.M.; Montell, J.A.; Pertusa, A.; Bustos, A.; Cazorla, M.; Galant, J.; Barber, X.; Orozco-Beltrán, D.; García-García, F.; et al. BIMCV COVID-19+: A Large Annotated Dataset of RX and CT Images from COVID-19 Patients. Available online: <https://arxiv.org/abs/2006.01174v3> (accessed on 12 February 2022).
30. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. COVID-19 Image Data Collection: Prospective Predictions Are the Future. Available online: <https://github.com/ieee8023/covid-chestxray-dataset> (accessed on 20 February 2022).
31. LeCun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.R. Efficient BackProp. In *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*; Montavon, G., Orr, G.B., Müller, K.R., Eds.; Springer: Berlin, Heidelberg, Germany, 2012; Volume 7700. [[CrossRef](#)]
32. Sabottke, C.F.; Spieler, B.M. The Effect of Image Resolution on Deep Learning in Radiography. *Radiol. Artif. Intell.* **2020**, *2*, e190015. [[CrossRef](#)]
33. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
34. Cimatic_XCOV19 Git Page. Available online: <https://github.com/CRIA-CIMATEC/covid-19> (accessed on 10 December 2020).
35. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* **1988**, *44*, 837–845. [[CrossRef](#)]
36. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5. Available online: <http://jmlr.org/papers/v18/16-365.html> (accessed on 28 August 2021).

37. Winston, J.; Jackson, D.; Wozniak, D.; Zeisler, J.; Farish, S.; Thoma, P. Quality Control recommendations for diagnostic radiography volume 3 radiographic or fluoroscopic. In *Radiographic or Fluoroscopic Machines*; CRCPD Publication: Frankfort, KY, USA, 2001; Volume 3.
38. Zhou, J.; Jing, B.; Wang, Z.; Xin, H.; Tong, H. SODA: Detecting COVID-19 in Chest X-rays with Semi-supervised Open Set Domain Adaptation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**. [[CrossRef](#)]
39. Rajpurkar, P.; Joshi, A.; Pareek, A.; Ng, A.Y.; Lungren, M.P. CheXternal: Generalization of deep learning models for chest X-ray interpretation to photos of chest X-rays and external clinical settings. In *Proceedings of the Conference on Health, Inference, and Learning, Virtual*, 8–9 April 2021. [[CrossRef](#)]

4 MANUSCRITO 2

A light deep learning algorithm for CT diagnosis of COVID-19 pneumonia

Este artigo apresenta Cimatec_Covnet-19, um algoritmo de inteligência artificial, baseado em aprendizagem profunda, capaz de apoiar o diagnóstico de pneumonia por COVID-19 em exames de tomografia computadorizada do tórax. O artigo demonstra o processo de desenvolvimento do algoritmo e a realização dos testes de performance. O trabalho propõe uma nova arquitetura CNN baseada em VGG 3D exigindo poucos recursos computacionais. A atuação em três dimensões permite ao algoritmo inferir o diagnóstico sobre o exame completo com múltiplas fatias de imagens (slices) simultaneamente, mimetizando a atuação do médico radiologista. Introduz uma nova técnica de pré-processamento que reduz o número de imagens necessárias para treinar o algoritmo. O projeto utiliza um conjunto de dados de bases de dados geograficamente distribuídas somadas a bases de hospitais brasileiros para treinamento e avaliação do desempenho. Todo o código foi disponibilizado abertamente.


O artigo contou com a participação deste autor e de seu orientador, Dr. Erick Giovani Sperandio Nascimento, de seu coorientador Dr. Roberto Badaró e do pesquisador Carlos Purificação do Centro de Competência em IA do SENAI CIMATEC. Ao final do artigo são identificadas as contribuições de cada autor.

Artigo publicado no periódico MDPI Diagnostics. 2022, 12(7), 1527; <https://doi.org/10.3390/diagnostics12071527> em 23 de junho de 2022.

O artigo é de acesso livre e distribuído sob os termos e condições da licença Creative Commons Attribution (CC BY), disponível em <https://creativecommons.org/licenses/by/4.0/>.

Article

A Light Deep Learning Algorithm for CT Diagnosis of COVID-19 Pneumonia

Adhvan Furtado ¹, Carlos Alberto Campos da Purificação ¹, Roberto Badaró ²
and Erick Giovanni Sperandio Nascimento ^{1,*}

¹ Supercomputing Center SENAI CIMATEC, Av. Orlando Gomes, 1845, Piatã, Salvador 41560-010, Brazil; adhvan@fieb.org.br (A.F.); carlos.purificacao@fieb.org.br (C.A.C.d.P.)

² Instituto SENAI de Inovação em Saúde, Av. Orlando Gomes, 1845, Piatã, Salvador 41560-010, Brazil; badaro@fieb.org.br

* Correspondence: erick.sperandio@fieb.org.br; Tel.: +55-2799-2799-651

Abstract: A large number of reports present artificial intelligence (AI) algorithms, which support pneumonia detection caused by COVID-19 from chest CT (computed tomography) scans. Only a few studies provided access to the source code, which limits the analysis of the out-of-distribution generalization ability. This study presents Cimatic-CovNet-19, a new light 3D convolutional neural network inspired by the VGG16 architecture that supports COVID-19 identification from chest CT scans. We trained the algorithm with a dataset of 3000 CT Scans (1500 COVID-19-positive) with images from different parts of the world, enhanced with 3000 images obtained with data augmentation techniques. We introduced a novel pre-processing approach to perform a slice-wise selection based solely on the lung CT masks and an empirically chosen threshold for the very first slice. It required only 16 slices from a CT examination to identify COVID-19. The model achieved a recall of 0.88, specificity of 0.88, ROC-AUC of 0.95, PR-AUC of 0.95, and F1-score of 0.88 on a test set with 414 samples (207 COVID-19). These results support Cimatic-CovNet-19 as a good and light screening tool for COVID-19 patients. The whole code is freely available for the scientific community.

Keywords: deep learning; COVID-19; CT; screening test



Citation: Furtado, A.; da Purificação, C.A.C.; Badaró, R.; Nascimento, E.G.S. A Light Deep Learning Algorithm for CT Diagnosis of COVID-19 Pneumonia. *Diagnostics* **2022**, *12*, 1527. <https://doi.org/10.3390/diagnostics12071527>

Academic Editor: Ivan Fan Ngai Hung

Received: 4 June 2022

Accepted: 20 June 2022

Published: 23 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

COVID-19 still affects public health services. Until 17 June 2022, there have been 535,863,950 confirmed cases of COVID-19 and 6,314,972 deaths all over the world, reported to WHO [1]. Despite the declining curve of new cases throughout the world, it is paramount to identify suspicious cases, differentiate them from other respiratory diseases, and to define appropriate isolation and treatment strategies [2]. In healthcare units, mechanisms for screening and monitoring the evolution of the disease are essential. The “Gold Standard” for diagnosing a COVID-19 infection is a reverse transcription-polymerase chain reaction (RT-PCR) test. Although RT-PCR is a reliable test, it needs trained people to perform the nasopharyngeal swab collection and a specialized laboratory for analysis. Results can take a few hours or days, and there is a significant and not yet fully explained variation in the proportion of false-negative results [3,4]. There are many healthcare facilities, especially in developing countries, where mechanisms for patient assessment and management are essential and RT-PCR is not completely available.

The SARS-CoV-2 infection generates characteristic abnormalities in chest image examinations. Chest radiography and computed tomography (CT) scans are the most common methods to support the diagnosis of pneumonia in symptomatic patients [5]. These examinations have been widely used as part of the initial screening and in situations where the patient has strong respiratory symptoms [6]. Even with the appearance of new variants less aggressive to lungs, it is still necessary to detect and monitor COVID-19 pneumonia, as we do not know how the disease will evolve in the next years to come.

An X-ray machine is the most commonly available imaging tool for patients with respiratory complaints. It is especially useful to identify severe cases of COVID-19 patients, as there might not be any findings on exams in mild or early-stage patients [7]. It is a simple, fast, and safe examination procedure. AI algorithms can support the detection of pneumonia caused by COVID-19 in chest radiographs [8]. Figure 1 presents a COVID-19 patient's radiography highlighting pulmonary infiltrates.

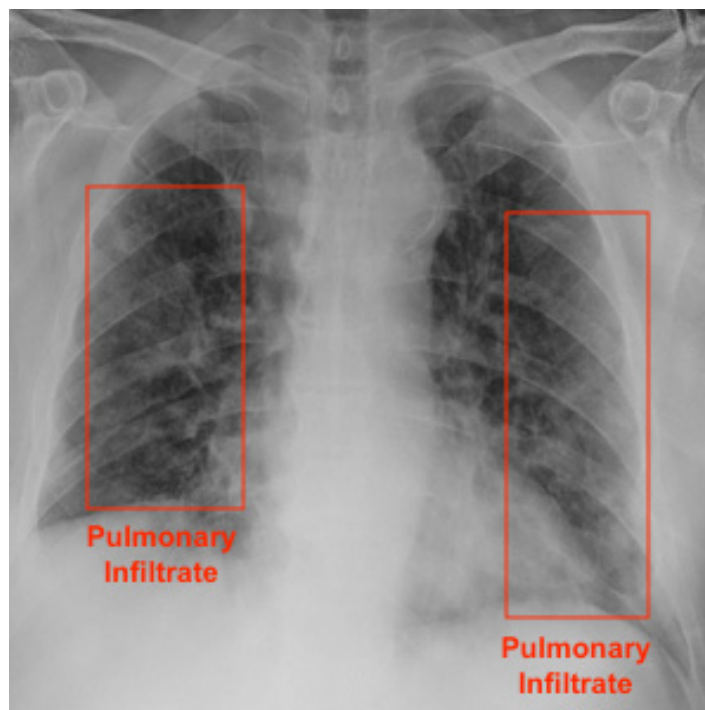


Figure 1. COVID-19 patient is male and 73 years old. Bounding box highlights infiltrates.

A chest CT scan combines data from multiple X-rays taken from different angles, which produces a detailed image of the lungs. CT scans are more effective than chest X-ray in early stages of COVID-19 disease detection. They have been used as a tool to diagnose and monitor the progression of the disease [9]. More than 70% of chest CT scans in patients with RT-PCR test-proven COVID-19 cases report ground-glass opacities, vascular enlargement, bilateral abnormalities, lower lobe involvement, and posterior predilection [10]. Figure 2 illustrates those abnormalities. Studies by [11,12] confirm that patients with COVID-19 pneumonia have ground-glass opacities in the earlier stages of the disease and pulmonary consolidation in later stages. Eventually, a rounded morphology and a peripheral pulmonary distribution are observed. Those abnormalities are analogous to those observed in other coronavirus infections, such as SARS-CoV-1 and MERS-CoV [13].

Although typical images can help in the early screening of suspected cases, images of various viral pneumonias are similar and overlap with other infectious and inflammatory lung diseases. Therefore, it is not trivial for radiologists to distinguish COVID-19 pneumonia from other viral pneumonias. AI algorithms are a valuable tool to support this task. It is important to notice that the WHO and the American Society of Radiology do not recommend the use of radiology images as the principal diagnostic method for COVID-19 [14–16].

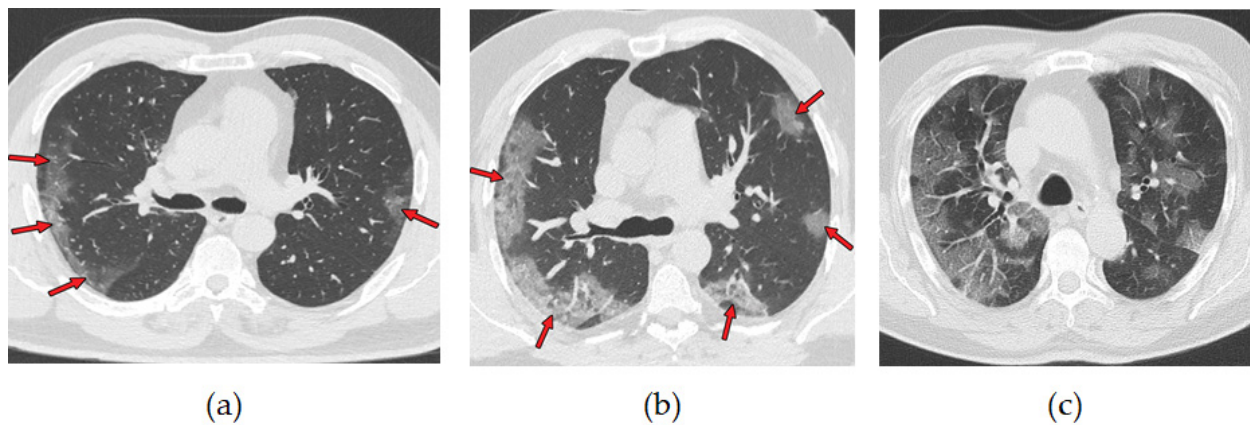


Figure 2. Axial nonenhanced chest CT images (lung window) in a 59-year-old man (a) and a 47-year-old man (b) show bilateral areas of ground-glass opacities (arrows) in a peripheral distribution; (c) shows bilateral ground-glass opacities and dilated segmental and subsegmental vessels, mainly on the right, in a 70-year-old man, each with positive RT-PCR test results for SARS-CoV-2. Adapted from [10].

The perspective of using deep learning algorithms as a fast and widely available alternative for the diagnosis of COVID-19 by RT-PCR has expanded the quantity and quality of research in this area. A research done on 1 May 2022 for articles with the words: “Deep Learning” and “CT” and “COVID-19” and “Diagnosis” in the abstract resulted in 287 findings in the *PubMed* database, being 52 in the MDPI repository. Despite the availability of studies, there are strong obstacles for the regular application of the proposed algorithms in clinical practice. A study by [17] systematically reviewed publications of machine learning models for the diagnosis or prognosis of COVID-19 from X-ray or CT images, concluding that all identified models had methodological flaws and/or underlying biases preventing their use in clinical practice. A review by [18] identified that most of the studies have utilized small datasets and lacked comparative analysis with other existing research, and the codes and data were not available. In our review, we also identified fundamental problems that limit the adoption of algorithms in healthcare centers. There is limited access to the complete source code, train, and test data. Thus, it is not possible to replicate the results and to evaluate the AI algorithm on different data sets. Most of the studies used a limited number of images from local sources or used only well-known public databases, and therefore, their models were not stressed enough to generalize properly to other phenotypes and geographic regions contexts. For instance, we only identified a few publications that used chest CT images from Brazilian hospitals. The work by [19] used data from 130 patients from two hospitals in Rio de Janeiro and one in Porto, Portugal, to develop an algorithm to identify and quantify the extent of lung involvement in patients with COVID-19 pneumonia. The study by [20] developed an algorithm for segmenting COVID lesions on CT using a base of 40 patients from a hospital in Rio de Janeiro. Both studies used small databases. In this work, we avoided repeating the most common flaws identified in the available studies and sought to advance the knowledge necessary to support the use of such algorithms in clinical practice, preparing it for use in a hospital in Brazil, a country with resources constraints to combat COVID-19. Table 1 categorizes the mapped problems and solutions developed in this work.

Table 1. Assessment of the main problems found in the literature review.

Category	Problem	Solution
Dataset	Few images for training the algorithm	Algorithm trained with 3000 chest CT examinations
Dataset	Data collected from only one geographical region	Data collected from a set of curated international public databases summed up with images from 2 Brazilian hospitals
Dataset	Poor image bank quality: non-standard scans, too many images of children, or excess of data from China patients	Attention to the selection of the best public bases; automatic and visual cleaning.
Methodology	Use of unbalanced datasets	Attention to balancing the COVID-19, non-COVID, and normal categories prior to training and testing.
Methodology	Lack of statistical rigor or bias	We used the CLAIM [10] checklist for AI in medical imaging. Available as a supplement.
Transparency	Non-replicable projects.	The whole code is open.

A review study by [21] highlighted the widespread use of convolutional neural networks for extracting relevant features from CT scans and noted that most classification models for COVID-19 use pre-trained networks. Another extensive review done by [22] showed that many 2D and 3D models were used to support the identification of pneumonia, mainly based on Inception, VGG, and ResNet architectures.

The work of [23,24] used 2D networks to analyze each CT slice image individually and adopted voting methods to classify the patient outcome. Another popular approach using 2D networks was to generate embedding feature vectors for every image, pool them to a single global feature vector, and use fully-connection layers for classification [25,26]. Some studies used 3D CNN networks, where a subset or all the available CT slice images per examinations were used as input [27,28]. Most of the 3D CNN algorithms used a fixed number of images from CT examinations as input because using all available images can be very memory-consuming. The work by [29] studied and compared various deep learning techniques applied to both chest radiographs and CT scans images for the detection of COVID-19 and validated VGG16 and ResNet50 as good architectures for classification. In order to develop a new model for the COVID-19 diagnosis, the study by [30] tested multiple architectures: DenseNet-169, VGG-16, ResNet-50, InceptionV3, and VGG-19. The VGG-19 proved to be superior with an accuracy of 94.52% when compared to all other deep learning models. The similarity of COVID-19-generated pulmonary lesions with the ones generated by other respiratory diseases reinforces the necessity of the algorithm to have an excellent feature extraction ability. A study by [31] proposed the use of a bag of deep visual words (BoDVW) on the VGG-16 architecture. The method removes the feature-map normalization step and adds a deep feature normalization step on the raw feature maps, preserving the semantics of each feature map that might have importance in differentiating COVID-19 from other forms of pneumonia on radiographies. This method was improved by including a multi-scale BoDVW [32] and an attention module to capture the spatial relationship between the regions of interest in CXR images [33].

In our work, we decided to adapt the VGG architecture for a 3D CNN. The input is a set of slices of a patient's CT. The objective is to preserve the embedded information of the CT examination on the frame stack, thus mimicking the behavior of a radiologist's analysis. We used a fixed set of 16 slices per CT scan examination to reduce hardware consumption and avoid lack of memory problems. We developed a novel pre-processing technique to choose and prepare the best slices for training and validation.

There are many regions in Brazil and in the world that do not have access to RT-PCR exams in the quantity and time needed or specialized physicians. In these cases, alternatives that facilitate the diagnosis of COVID-19 are very important. In this paper, we present Cimatic-CovNet-19, a fast, VGG-based CNN algorithm for COVID-19 diagnosis in chest CT scans. We developed our system on a set of 3000 chest CT scans, from which 734 examinations were from Brazilian hospitals. This study confirms the hypothesis that AI systems are able to correctly classify COVID-19 and non-COVID-19 classes from CT scans. We evaluated and compared the performance of the algorithm with data from geographically distributed datasets and data from a Brazilian hospital. The main innovations of this study are:

- Proposing a novel 3D VGG-based CNN architecture for accurately diagnosing COVID-19 on chest CT scans. The 3D network is able to identify correlations between adjacent slices, while 2D networks are limited to intra-slice spatial voxel information.
- Introducing a novel pre-processing technique, which reduces the number of slices required for training the algorithm: Processing fewer slices demands less computational power, prevents communications bottlenecks, and reduces time and cost constraints. Since the model only requires 16 slices per CT examination, it is also well-suited for a large number of CT machines.
- Evaluating the algorithm's diagnosis performance in both geographically distributed and Brazilian datasets: Brazil has more than 300,000,000 inhabitants. It was one of the worst-affected countries in the world by the COVID-19 pandemic. Despite that fact, there are few studies with data from this country. It was important to include images from Brazilian hospitals and confirm the algorithm's ability to generalize well for this phenotype. We plan to test the algorithm in a controlled environment in a Brazilian hospital in the near future.
- Disposing the algorithm as an open software for public use and future enhancements: This guarantees reproducibility.

2. Materials and Methods

2.1. Dataset Preparation

In the retrospective study, we gathered 5787 CT scans from nine different datasets sources. We used seven public datasets containing CT scans from all over the world: *Medical Segmentation Decathlon*, *LNDb*, *LCTSC*, *MOSMEDDATA*, *COVID-19 CT Lung and Infection Segmentation*, *COVID-19 CT Segmentation Dataset*, *BIMCV-COVID19*, and two private datasets from Brazilian hospitals: *HCUSP* and *HSI*. We included in this study only images in the axial plane and from patients with a diagnosis issued by a radiologist from well-known hospitals. All patient information was already anonymized in the data source. The ground truth for a positive COVID-19 outcome was a positive RT-PCR test associated with the CT-scan examination. We performed a visual inspection of the central slice in each of the 5787 CT scans and manually discarded all data that were in sagittal or coronal planes, had low-quality resolution, or were masks of CT scans instead of the CT scan itself. Altogether, this procedure removed 1108 samples. Table 2 presents the complete list of databases used in this work.

Considering a variety of CT scanners available worldwide, it would be natural to expect that the source datasets had different number of slices and resolutions, which, in fact, happened. Additionally, the data were unbalanced regarding the presence of COVID-19-positive CT scans. The demographic information from the patients was not consistent and thus not used in this work. From the remaining 4679 CT scans, we prepared a random, balanced subset with 3000 samples (1500 COVID-19, 1500 non-COVID-19), which were then split into training and validation sets.

Table 2. CT scans databases used in this study.

Dataset ID	Dataset Name	Public (Y/N)	Number of CT Scans after Data Cleaning	Avg Number of Slices per CT Scan	Number of CT Scans Positive for COVID-19	Training/Val Dataset	Test Dataset 1	Test Dataset 2
i	<i>Medical Segmentation Decathlon</i>	Y	94	279	0	50	0	9
ii	<i>LNDb</i>	Y	139	322	0	78	0	13
iii	<i>LCTSC</i>	Y	94	279	0	25	0	3
iv	<i>MOSMEDDATA</i>	Y	1105	42	1105	449	0	56
v	<i>HCUSP</i>	N	935	337	431	384	170	51
vi	<i>HSI</i>	N	1806	308	1294	766	0	79
vii	<i>COVID-19 CT Lung and Infection Segmentation</i>	Y	10	176	10	8	0	1
viii	<i>COVID-19 CT Segmentation Dataset</i>	Y	10	280	10	3	0	1
ix	<i>BIMCV-COVID19</i>	Y	486	288	308	222	0	31

2.2. Dataset Description

- i. *Medical Segmentation Decathlon*: The *Medical Segmentation Decathlon* is a collection of annotated medical image datasets for the development and evaluation of segmentation algorithms. The lung dataset has 96 preoperative thin-section CT scans performed without use of contrast and from patients with non-small cell lung cancer from Stanford University (Palo Alto, CA, USA) publicly available through TCIA [34].
- ii. *LNDb*: This dataset contains 294 CT scans collected retrospectively at the Centro Hospitalar e Universitário de São João (CHUSJ) in Porto, Portugal, between 2016 and 2018. All data were acquired under approval from the CHUSJ Ethical Committee and was anonymized. Among the 294 patients scanned, 164 (55.8%) were male. The average age was 66, and the minimum and maximum ages were 19 and 98, respectively [35].
- iii. *LCTSC*: This dataset was provided in association with a challenge competition and related conference session conducted at the American Association of Physicists in Medicine 2017 Annual Meeting. There are CT scans of 60 patients undergoing treatment simulation for thoracic radiotherapy from three institutions: MD Anderson Cancer Center, Memorial Sloan-Kettering Cancer Center, and the MAASTRO clinic. Each institution provided CT scans from 20 patients, including mean intensity projection (4D CT), exhale phase (4D CT), or free-breathing CT scans depending on their clinical practice. All CT scans covered the entire thoracic region with a 50 cm field of view and slice spacing of 1 mm, 2.5 mm, or 3 mm [36].
- iv. *MOSMEDDATA*: This dataset contains 1110 anonymized lung CT scans obtained between 1 March 2020 and 25 April 2020 from public medical hospitals in Moscow, Russia. Among the patients scanned, there were 42% males, 56% females, and 2% other/unknown with ages from 18 to 97 years, with an average of 47 years. They were distributed according to a classification table of the severity of lung tissue abnormalities with COVID-19 and routing rules. There were five categories ranging from CT-0, zero, not consistent with pneumonia (including COVID-19) up to CT-4, severe, with diffuse ground glass opacities, with consolidations and reticular changes, and pulmonary parenchymal involvement $\geq 75\%$. The number of cases by category was: CT-0, 254 (22.8%); CT-1, 684 (61.6%); CT-2, 125 (11.3%); CT-3, 45 (4.1%); and CT-4, 2 (0.2%) [37].
- v. *HC USP*: The data were obtained through a collaboration between SENAI CIMATEC and the Medical School of the University of São Paulo (HC USP). Altogether, we obtained 439 COVID-19-positive exams and 506 COVID-19-negative exams.

- vi. *HSI*: The data were acquired from a partnership between SENAI CIMATEC and the Santa Isabel Hospital (HSI). This database has 1294 COVID-19-positive exams and 512 COVID-19-negative exams.
- vii. *COVID-19 CT Lung and Infection Segmentation*: This dataset contains 20 labeled COVID-19 lung-infection CT scans collected from the Coronacases Initiative and Radiopaedia, which can be freely downloaded with CC BY-NC-SA license. The proportion of infections in the lungs ranges from 0.01% to 59%. The left lung, right lung, and infection segmentation were firstly delineated by junior annotators (1 to 5 years of experience), then refined by two radiologists with 5 to 10 years of experience. All the annotations were verified and refined by a senior radiologist (>10 years of experience) [38].
- viii. *COVID-19 CT Segmentation Dataset*: This dataset contains 100 axial CT images from more than 40 patients with COVID-19 converted from openly accessible JPG images provided by the Società Italiana di Radiologia Medica e Interventistica. The images were segmented by a radiologist using three labels: ground glass, consolidation, and pleural effusion [39].
- ix. *BIMCV-COVID19*: A large dataset from the Valencian Region Medical ImageBank (BIMCV) containing chest X-ray images CXR (CR, DX) and computed tomography (CT) imaging of COVID-19+ patients along with their radiological findings and locations, pathologies, radiological reports (in Spanish), DICOM metadata, polymerase chain reaction (PCR), immunoglobulin G (IgG), and Immunoglobulin M (IgM) diagnostic antibody tests was also used. This database includes 1380 CX, 885 DX, and 163 CT studies from 1311 COVID-19 patients [40].

2.3. Slice-Wise Selection

In order to normalize the input resolution, we used the Clara Training framework, part of the Clara Image software suite, to resample all DICOM and NIfTI data to a voxel spacing resolution of $1 \times 1 \times 1$ mm NIfTI format. Clara is an application framework optimized for healthcare and life sciences developers. It contains software development kits, full-stack GPU-accelerated libraries, and pre-tested reference applications [41]. We also used the Clara framework to obtain lung masks from each chest CT scan. We used the `clara_train_covid19_ct_lung_seg` model, a voxel-wise binary classification for lung region segmentation. Each voxel is predicted as either foreground (lung) or background. The output is a binary mask, where the lung is assigned 1, and the background is assigned 0. We noticed that the sum of pixels in the lung masks grows in a Gaussian-like pattern from the first to the last slice, peaking around the central slice. Using this information, we did a slice-wise selection in order to collect data from different areas of the lung. After experimenting with 64, 32 and 16 slices, the results did not have any significant statistical differences, so we used 16 slices from each CT scan in order to save computational resources. The slice-wise selection was performed according to the following expression:

$$\text{slice}_i = F + G \times i, \text{ with } i \text{ in } [0, 1, 2, \dots, 15], \quad (1)$$

where F is the first slice in the mask whose sum of pixels is greater than 1000, and G is the step size given by:

$$G = \lceil (\mu - F)/8 \rceil, \quad (2)$$

with μ being the central slice.

Before executing the described slice-wise selection, the CT scans were trimmed between -3000 and 4000 Housefield units (HU) and scaled between 0 and 1. We reshaped the 16 slices chosen from each CT scan to a $512 \times 512 \times 16 \times 1$ format. Figure 3 depicts a single slice from an exam both before and after being pre-processed.

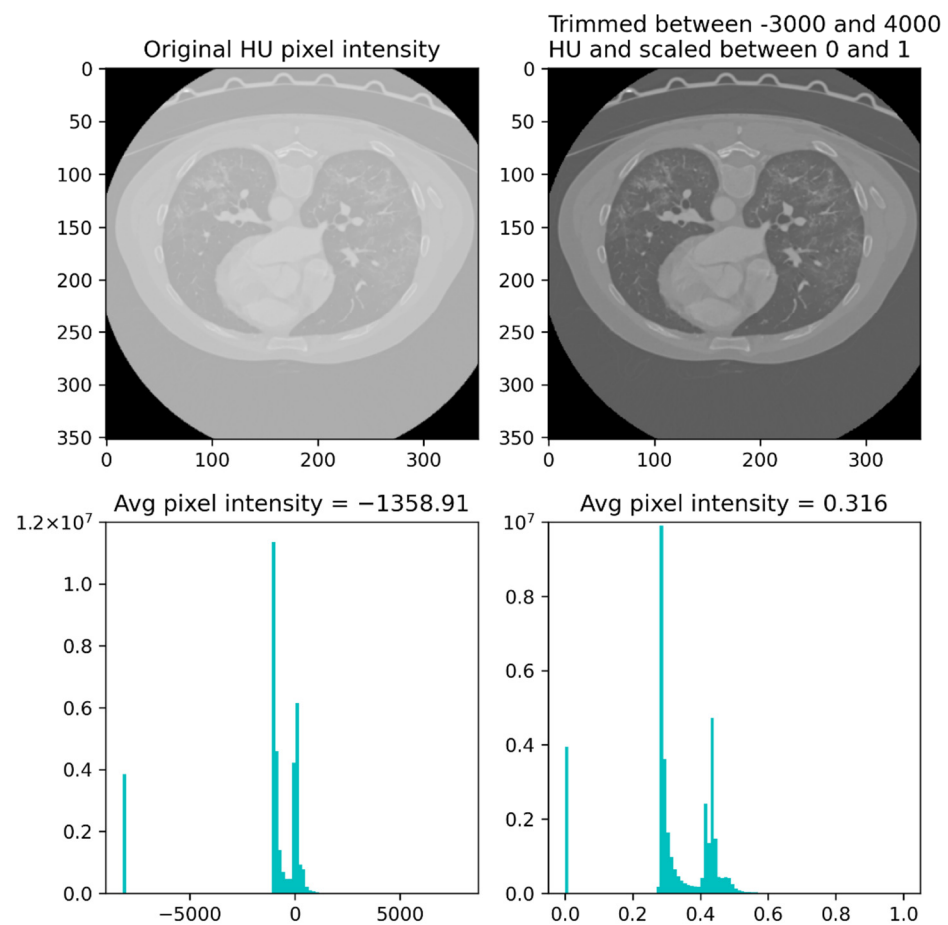


Figure 3. Example of a single slice before and after the pre-processing routine.

2.4. Algorithm Architecture

Cimatec-CovNet-19 has an architecture inspired by the VGG-16 neural network. The VGG-16 was developed in 2014 and is one of the best CNN architectures to deal with 2D large-scale image recognition tasks. The image passes through a stack of convolutional layers with very small receptive fields (3×3 kernels), which is the smallest size possible to capture pixel position notions (left/right, up/down, center). The spatial resolution is preserved with paddings. After some of the convolutional layers, there are max pooling layers (2×2 window, stride 2) to guarantee spatial pooling. The stack of convolutional layers is followed by three fully connected (FC) layers. The last layer is a softmax layer, which is a function to represent the network output as a categorical distribution [42].

In our model, there are 17 convolutional layers split into 5 convolutional blocks with different filter sizes, as can be seen in more detail in Figure 4.

The model takes CT slices as input and combines the features extracted from the slices in a sequence of convolutions and pooling operations. The number of input slices can vary. Typically, it can be 64, 32, or 16 slices. It requires an analysis and validation of the approach to select the lowest number of slices without losing accuracy, which will be presented in the following section.

There are more pooling layers in the two initial convolutional blocks than in the final ones. We chose this approach to reduce the tensors size and fit them in the available GPU memory. We also added batch normalization layers after every convolutional layer and a single dropout layer with a 0.5 dropout rate to enhance the training performance and prevent overfitting. The final feature map runs through two FC layers, the first with 4096 neurons and the second being the output layer with a sigmoid activation function to generate a binary output, namely COVID-19 or non-COVID-19. All hidden layers are built

with the rectified linear unit (ReLU) [43] activation function. The model had 47.3 million parameters, was trained in a computing node with four NVIDIA GPUs V100 32 GB SXM2, and took 9313 s to train 56 epochs.

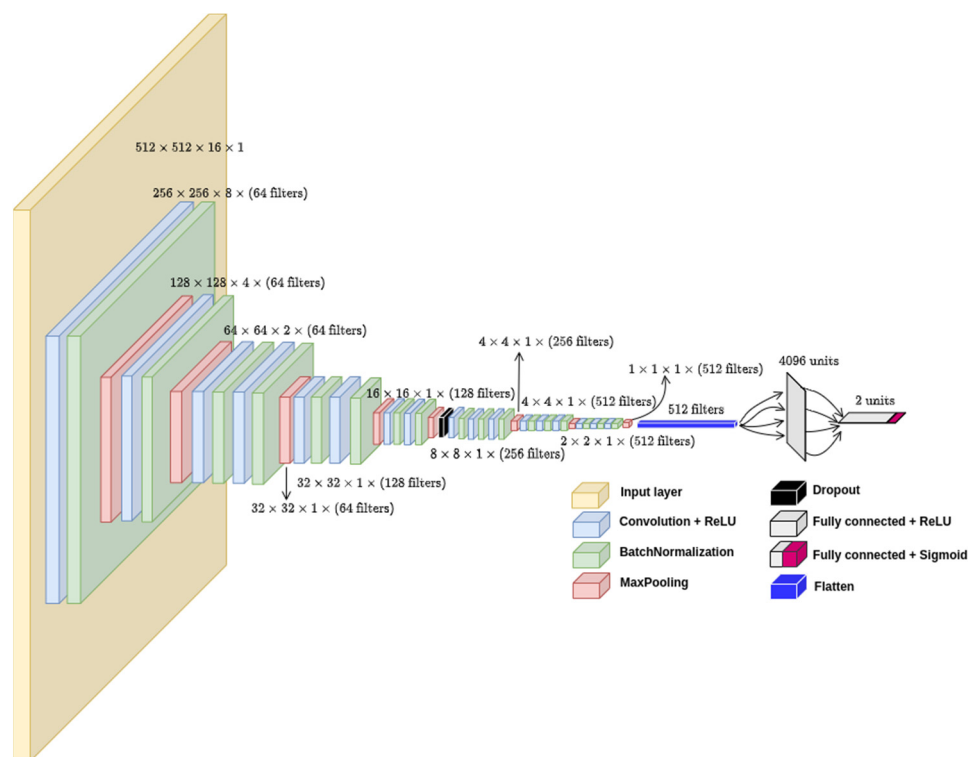


Figure 4. 3D-CNN model architecture developed on top of the standard VGG-16 2D model.

2.5. Model Training

We randomly initiated the weights and trained the neural network with a batch size of 16 using the Adamax optimizer and learning rate of 10^{-3} . We used early stopping with a patience of seven epochs based on the validation loss. During model development, 2000 samples were used for training and 1000 samples for validation as observed in Table 3.

Table 3. Dataset split during model development.

	Training	Validation	Test Dataset 1	Test Dataset 2
COVID-19	1000	500	85	122
Non-COVID-19	1000	500	85	122
Total	2000	1000	170	244

In order to fine-tune the CNN architecture, we started the experiment with a different number of convolutional and pooling layers, following the VGG-16 pattern (increasing the filter size as the layers went deeper). Then, we tried different number of neurons in the FC layers and a sequence of three FC layers. Finally, we tried different regularization techniques:

- Batch normalization layer in different positions after the convolutional layers,
- Dropout layers in different positions and different dropout rates,
- L2 regularization in different layers, resulting in regularizations to the fourth, eleventh, and fourteenth convolutional layers and to the penultimate FC layer.

All the experiments were performed with the keras tuner API [44], which is an easy-to-use, scalable, hyperparameter optimization framework. We performed the hyperparameter search with the built-in hyperband optimization algorithm [45].

We used two datasets for model assessment: (1) data from Medical School of the University of São Paulo and (2) data randomly taken from the full dataset. Both test sets were balanced (50% for each class: COVID-19, non-COVID-19). We reached a plateau for model assessment after experimenting with several different hyperparameters settings and model architectures.

In order to evaluate the model variability in different portions of the data, we used a stratified 10-fold cross-validation on the 3000 samples. Finally, we combined the training and validation datasets into a single training dataset and added data augmentation to each of the 3000 examples, bringing the total number of samples in the training dataset to 6000, as observed in Table 4.

Table 4. Final training dataset with data augmentation.

	Original	Augmented	Test Dataset 1	Test Dataset 2
COVID-19	1500	1500	85	122
Non-COVID-19	1500	1500	85	122
Total	3000	3000	170	244

Five different data augmentation techniques were tried: vertical and horizontal flip, changing brightness and contrast, shear, zoom-in and zoom-out, and small rotations. For each technique, we trained the model with a pair-wise combination of the 3000 original images with 3000 augmented images. Finally, we combined all augmented images with the original images and found that augmented rotated images showed the best results. In this technique, every image suffered small rotations. The algorithm randomly rotates the images with one of the angles in the set $(-15, -10, 10, 15)$. For the final training, there were neither validation data nor automatic early stoppage. We defined the number of epochs to train the algorithm as 56. It was the same number of epochs achieved for the best model weights reached during model development.

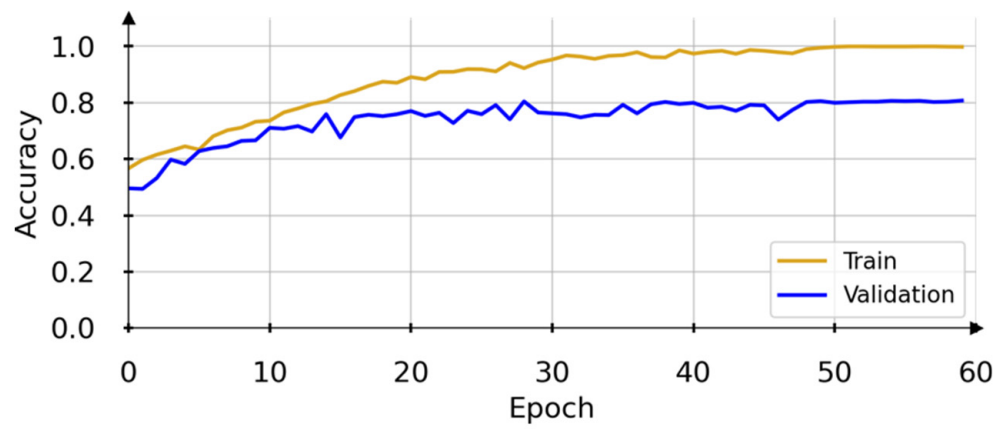
3. Results

After trying different hyperparameters setups throughout model development, we achieved the results presented on Figure 5. Notice that the validation curves reach an accuracy plateau around 0.80 by the 50th epoch. The model weights stabilize, and the accuracy for both training and validation data show little changes. The loss for the training and validation sets also stabilizes around epoch 50.

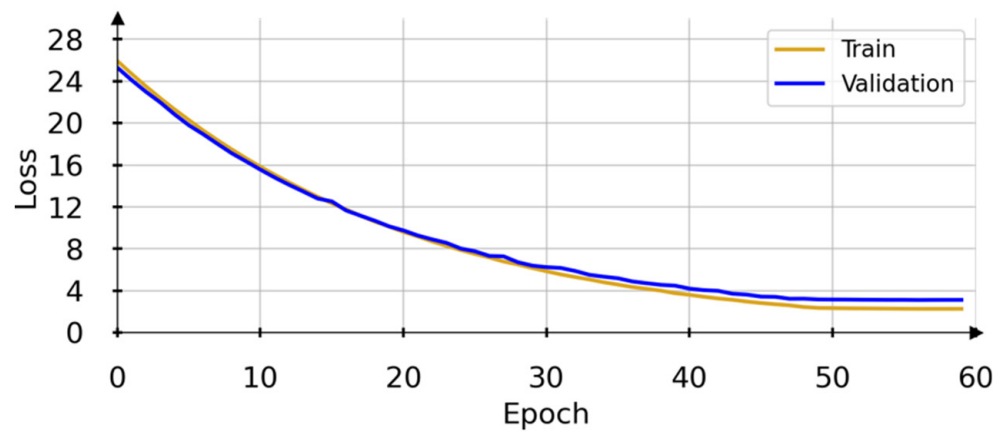
Figure 6 presents the boxplot of the stratified 10-fold cross-validation results, and Table 5 presents the evaluation results for each validation fold in more detail. We can observe that the PR-AUC varies from 0.86 to 0.96, the ROC-AUC varies from 0.87 to 0.96, and the F1-score varies from 0.80 to 0.90. Those results represent a good overall performance when compared to several recent related works [21,46,47].

The confusion matrices in Figure 7 the ROC-AUC in Figure 8 and PR-AUC in Figure 9 show the model performance in both test datasets. For test dataset 1, the model assessment shows a recall of 88.51% (95% CI, 79.88% to 94.35%), specificity of 90.36% (95% CI, 81.89% to 95.75%), accuracy of 89.41% (95% CI, 83.78% to 93.60%), and ROC-AUC and PR-AUC of 97%. Test dataset 2 shows a recall of 85.25% (95% CI, 77.69% to 91.02%), specificity 90.98% (95% CI, 84.44% to 95.41%), accuracy of 88.11% (95% CI, 83.38% to 91.89%), and ROC-AUC and PR-AUC of 93%.

Finally, we present the results with the combined datasets (test dataset 1 and test dataset 2) in Figures 10 and 11 as an overall performance assessment. The model assessment shows a recall of 88% (95% CI, 79.88% to 94.35%), specificity of 88% (95% CI, 81.89% to 95.75%), and accuracy of 89% (95% CI, 83.78% to 93.60%). We can see a ROC AUC and PR AUC of 95% for the combined test dataset. The model's performance in both dataset and in the combined set confirms its ability to generalize well for new data.



(a)



(b)

Figure 5. Model performance achieved during the development. In (a), training and validation accuracy. In (b), training and validation loss.

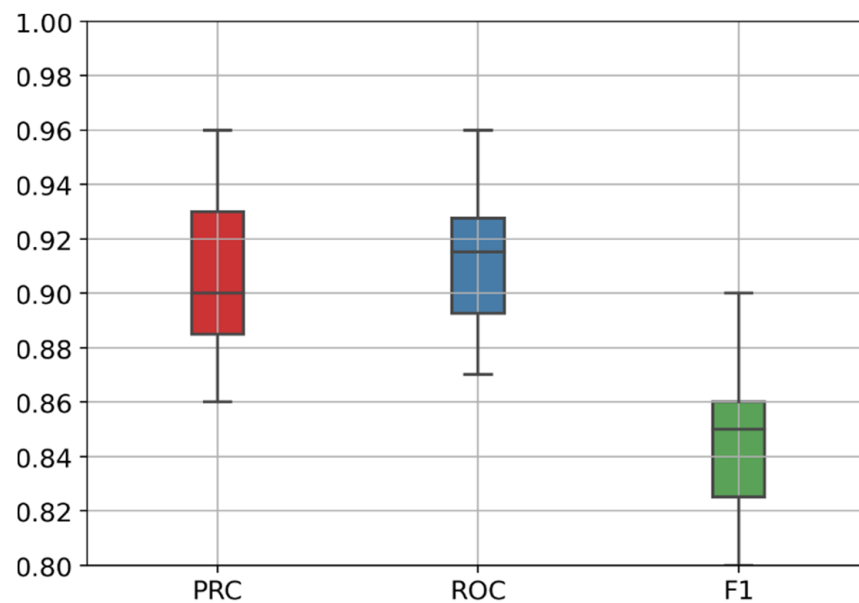


Figure 6. Boxplot of the stratified 10-fold cross-validation results showing PR-AUC, ROC-AUC, and F1-score ranges.

Table 5. Evaluation results for each validation fold.

Fold	PR-AUC	ROC-AUC	F1-Score
0	0.94	0.93	0.86
1	0.86	0.87	0.80
2	0.88	0.89	0.82
3	0.90	0.91	0.84
4	0.94	0.93	0.86
5	0.90	0.92	0.87
6	0.88	0.89	0.84
7	0.96	0.96	0.90
8	0.90	0.92	0.86
9	0.90	0.90	0.82
Average	0.906	0.912	0.847
Std	0.031	0.026	0.029

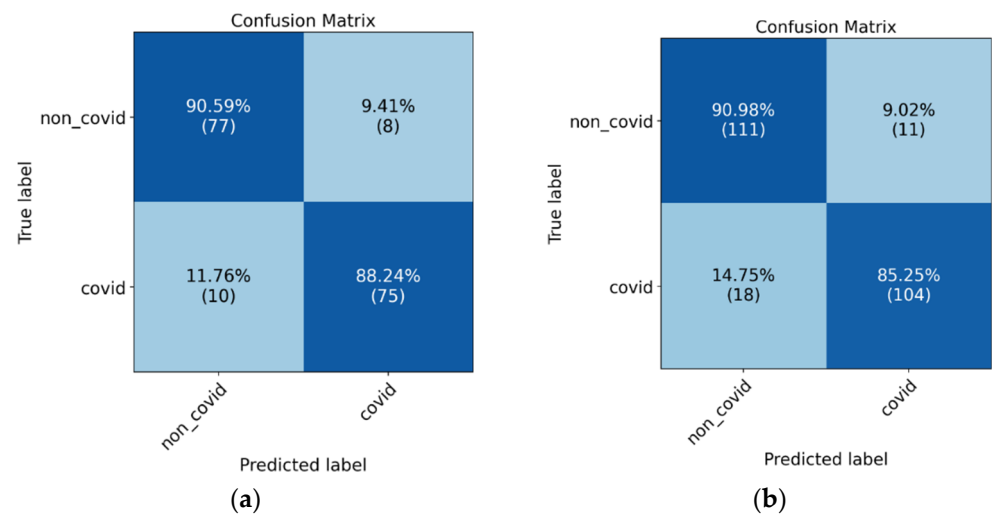


Figure 7. Confusion matrix results for model evaluation on the test datasets, (a) test dataset 1, and (b) test dataset 2.

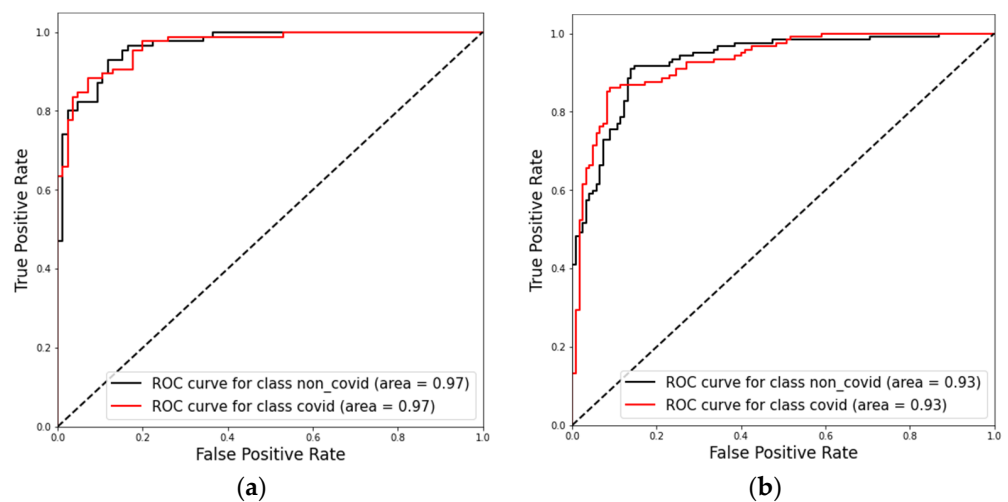


Figure 8. ROC curves obtained for model evaluation on the test dataset 1 (a) and on the test dataset 2 (b).

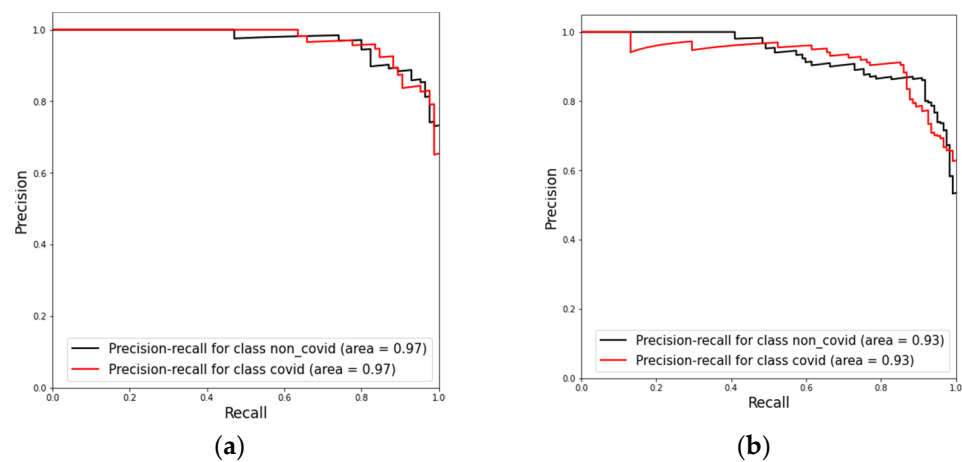


Figure 9. Precision-recall curves for model evaluation on both test datasets. (a) Results for test dataset 1 and (b) results for test dataset 2.

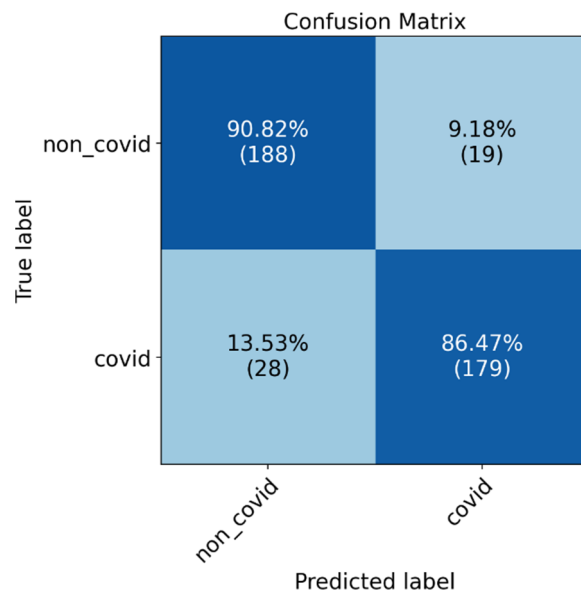


Figure 10. Confusion matrix for model evaluation on the combined test dataset.

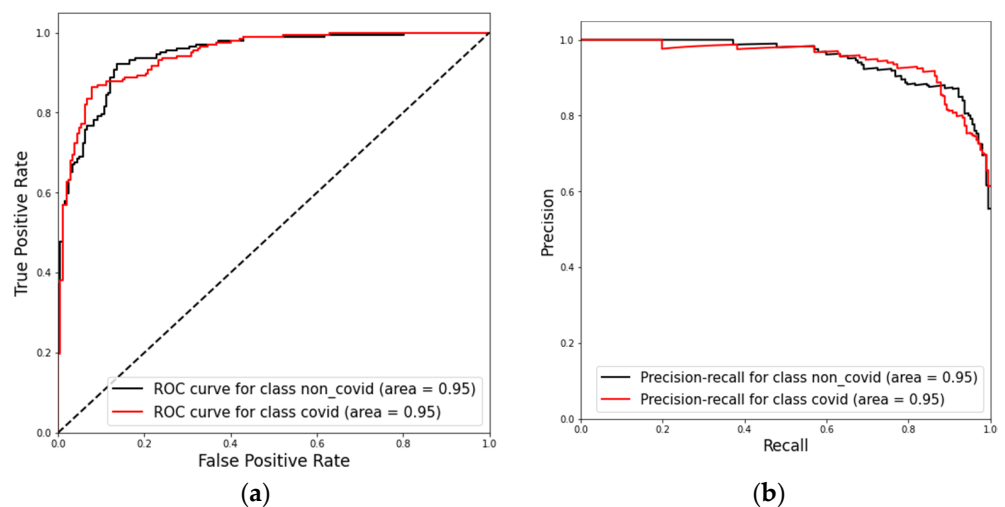


Figure 11. Present the ROC-AUC (a) and PR-AUC (b) results for model evaluations on the combined test dataset.

4. Discussion

In the lack of a specialized radiologist, AI models may support the identification of COVID-19 pneumonia characteristics in CT scans. With this objective in mind, we developed the Cimaterc-CovNet-19 neural network and evaluated its performance using two test datasets: one being a subset of a global public dataset and the other a set of 170 patients served by a hospital in São Paulo. Generalization for different datasets is a known problem in AI applied to medical images [48]. We did not observe major differences in the algorithm performance over the two tests datasets, which suggests that the algorithm generalizes well.

One limitation of this study is the use of a diverse public dataset, which lacks demographic information to train the algorithm. Those datasets might contain unknown biases and contaminate the model.

The importance of CT scans examinations to evaluate suspected COVID-19 patients and support the management of known patients is evident. The ROC-AUC and PR-AUC showed in this study validated that Cimaterc-CovNet-19 is a good screening tool for COVID-19 pneumonia from CT scans. The algorithm has a new approach for processing the images, requiring the use of fewer slices per examination and thus reducing training and inference times. This is important, especially for centers with low computing resources. The code is open for further enhancement. We encourage future works to compare this algorithm with other publicly available algorithms and explore its use in clinical practice in a controlled environment. In the near future, we plan to test Cimaterc_CovNet-19 in a hospital in Brazil.

The methodology used to build and test the algorithm and the developed model can quickly be adapted and applied to other lung infections in new potential pandemics.

Author Contributions: Conceptualization, A.F. and E.G.S.N.; data curation, A.F. and C.A.C.d.P.; formal analysis, E.G.S.N.; funding acquisition, A.F. and E.G.S.N.; investigation, A.F. and C.A.C.d.P.; methodology, A.F., C.A.C.d.P., R.B. and E.G.S.N.; project administration, A.F. and E.G.S.N.; resources, A.F. and C.A.C.d.P.; software, C.A.C.d.P.; supervision, A.F. and E.G.S.N.; validation, R.B. and E.G.S.N.; writing—original draft, A.F. and C.A.C.d.P.; writing—review and editing, A.F., C.A.C.d.P., R.B. and E.G.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ABDI, SENAI, EMBRAPII, REPSOL SINOPEC BRASIL grant “Missão contra a COVID-19 do Edital de Inovação para a Indústria”.

Institutional Review Board Statement: The retrospective study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of “Serviço Nacional de Aprendizagem Industrial—SENAI CIMATEC” (STU# 36389820.6.0000.9287) on 9 August 2020 and was granted a waiver of written informed consent.

Informed Consent Statement: Patient consent was waived in accordance with the evaluation of the Institutional Review Board considering that researchers undertake to maintain confidentiality, not disclosing the names of the participants, and using codes to identify the data generated by them to avoid violating participant privacy.

Data Availability Statement: The model’s source code is freely available on the research group GitHub page at <https://github.com/CRIA-CIMATEC/COVID-19> (accessed on 1 May 2022).

Acknowledgments: We gratefully acknowledge the support of SENAI CIMATEC AI Reference Center and the SENAI CIMATEC/NVIDIA AI Joint Center for scientific and technical support; the SENAI CIMATEC Supercomputing Center for Industry Innovation for granting access to the necessary hardware and technical support; Repsol Sinopec Brazil, ABDI, SENAI, and EMBRAPII for providing the funding for this research; HP Brazil for providing support; and Hospital Santa Izabel, MedSenior and HM Hospitales for providing data for this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard with Vaccination Data. Available online: <https://covid19.who.int/> (accessed on 19 June 2022).
2. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [[CrossRef](#)] [[PubMed](#)]
3. Arevalo-Rodriguez, I.; Buitrago-Garcia, D.; Simancas-Racines, D.; Achig, P.Z.; Del Campo, R.; Ciapponi, A.; Sued, O.; Martinez-García, L.; Rutjes, A.W.; Low, N.; et al. False-negative results of initial RT-PCR assays for COVID-19: A systematic review. *PLoS ONE* **2020**, *15*, e0242958.
4. Woloshin, S.; Patel, N.; Kesselheim, A.S. False Negative Tests for SARS-CoV-2 Infection—Challenges and Implications. *N. Engl. J. Med.* **2020**, *383*, e38. [[CrossRef](#)] [[PubMed](#)]
5. Pontone, G.; Scafuri, S.; Mancini, M.E.; Agalbatto, C.; Guglielmo, M.; Baggiano, A.; Muscogiuri, G.; Fusini, L.; Andreini, D.; Mushtaq, S.; et al. Role of computed tomography in COVID-19. *J. Cardiovasc. Comput. Tomogr.* **2020**, *15*, 27–36. [[CrossRef](#)] [[PubMed](#)]
6. Sandri, T.L.; Inoue, J.; Geiger, J.; Griesbaum, J.-M.; Heinzl, C.; Burnet, M.; Fendel, R.; Kremsner, P.G.; Held, J.; Kreidenweiss, A. Complementary methods for SARS-CoV-2 diagnosis in times of material shortage. *Sci. Rep.* **2021**, *11*, 11899. [[CrossRef](#)]
7. 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. *Lancet Infect. Dis.* **2020**, *20*, 425–434. [[CrossRef](#)]
8. Furtado, A.; Andrade, L.; Frias, D.; Maia, T.; Badaró, R.; Nascimento, E.G.S. Deep Learning Applied to Chest Radiograph Classification—A COVID-19 Pneumonia Experience. *Appl. Sci.* **2022**, *12*, 3712. [[CrossRef](#)]
9. Zu, Z.Y.; Di Jiang, M.; Xu, P.P.; Chen, W.; Ni, Q.Q.; Lu, G.M.; Zhang, L.J. Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology* **2020**, *296*, E15–E25. [[CrossRef](#)]
10. Kwee, T.C.; Kwee, R.M. Chest ct in COVID-19: What the radiologist needs to know. *Radiographics* **2020**, *40*, 1848–1865. [[CrossRef](#)]
11. Chung, M.; Bernheim, A.; Mei, X.; Zhang, N.; Huang, M.; Zeng, X.; Cui, J.; Xu, W.; Yang, Y.; Fayad, Z.A.; et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology* **2020**, *295*, 202–207. [[CrossRef](#)]
12. 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. *Lancet* **2020**, *395*, 497–506. [[CrossRef](#)]
13. Koo, H.J.; Lim, S.; Choe, J.; Choi, S.-H.; Sung, H.; Do, K.-H. Radiographic and CT features of viral pneumonia. *Radiographics* **2018**, *38*, 719–739. [[CrossRef](#)] [[PubMed](#)]
14. Akl, E.A.; Blažić, I.; Yaacoub, S.; Frija, G.; Chou, R.; Appiah, J.A.; Fatehi, M.; Flor, N.; Hitti, E.; Jafri, H.; et al. Use of chest imaging in the diagnosis and management of COVID-19: A WHO rapid advice guide. *Radiology* **2021**, *298*, E63–E69. [[CrossRef](#)] [[PubMed](#)]
15. Rubin, G.D.; Ryerson, C.J.; Haramati, L.B.; Sverzellati, N.; Kanne, J.; Raouf, S.; Schluger, N.W.; Volpi, A.; Yim, J.-J.; Martin, I.B.K.; et al. The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the fleischner society. *Radiology* **2020**, *296*, 172–180. [[CrossRef](#)]
16. Simpson, S.; Kay, F.U.; Abbara, S.; Bhalla, S.; Chung, J.H.; Chung, M.; Henry, T.S.; Kanne, J.P.; Kligerman, S.; Ko, J.P.; et al. Radiological society of North America expert consensus document on reporting chest CT findings related to COVID-19: Endorsed by the society of thoracic radiology, the American college of radiology, and RSNA. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e200152. [[CrossRef](#)]
17. Roberts, M.; Covnet, A.; Driggs, D.; Thorpe, M.; Gilbey, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.I.; Etmann, C.; McCague, C.; et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
18. Aggarwal, P.; Mishra, N.K.; Fatimah, B.; Singh, P.; Gupta, A.; Joshi, S.D. COVID-19 image classification using deep learning: Advances, challenges and opportunities. *Comput. Biol. Med.* **2022**, *144*, 105350. [[CrossRef](#)]
19. Carvalho, A.R.S.; Guimarães, A.; Garcia, T.D.S.O.; Werberich, G.M.; Ceotto, V.F.; Bozza, F.A.; Rodrigues, R.S.; Pinto, J.S.F.; Schmitt, W.R.; Zin, W.A.; et al. Estimating COVID-19 Pneumonia Extent and Severity From Chest Computed Tomography. *Front. Physiol.* **2021**, *12*, 617657. [[CrossRef](#)]
20. Diniz, J.O.B.; Quintanilha, D.B.P.; Neto, A.C.S.; da Silva, G.L.F.; Ferreira, J.L.; Netto, S.M.B.; Araújo, J.D.L.; Da Cruz, L.B.; Silva, T.F.B.; Martins, C.M.D.S.; et al. Segmentation and quantification of COVID-19 infections in CT using pulmonary vessels extraction and deep learning. *Multimed. Tools Appl.* **2021**, *80*, 29367–29399. [[CrossRef](#)]
21. Ozsahin, I.; Sekeroglu, B.; Musa, M.S.; Mustapha, M.T.; Ozsahin, D.U. Review on Diagnosis of COVID-19 from Chest CT Images Using Artificial Intelligence. *Comput. Math. Methods Med.* **2020**, *2020*, 9756518. [[CrossRef](#)]
22. Liu, F.; Tang, J.; Ma, J.; Wang, C.; Ha, Q.; Yu, Y.; Zhou, Z. The application of artificial intelligence to chest medical image analysis. *Intell. Med.* **2021**, *1*, 104–117. [[CrossRef](#)]
23. Heidarian, S.; Afshar, P.; Enshaei, N.; Naderkhani, F.; Rafiee, M.J.; Fard, F.B.; Samimi, K.; Atashzar, S.F.; Oikonomou, A.; Plataniotis, K.N.; et al. COVID-FACT: A Fully-Automated Capsule Network-Based Framework for Identification of COVID-19 Cases from Chest CT Scans. *Front. Artif. Intell.* **2021**, *4*, 65. [[CrossRef](#)] [[PubMed](#)]
24. Rahimzadeh, M.; Attar, A.; Sakhaei, S.M. A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset. *Biomed. Signal Process. Control.* **2021**, *68*, 102588. [[CrossRef](#)] [[PubMed](#)]

25. Heidarian, S.; Afshar, P.; Mohammadi, A.; Rafiee, M.J.; Oikonomou, A.; Plataniotis, K.N.; Naderkhani, F. Ct-caps: Feature extraction-based automated framework for COVID-19 disease identification from chest ct scans using capsule networks. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 1040–1044. [\[CrossRef\]](#)
26. Li, L.; Qin, L.; Xu, Z.; Yin, Y.; Wang, X.; Kong, B.; Bai, J.; Lu, Y.; Fang, Z.; Song, Q.; et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* **2020**, *296*, E65–E71. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Xue, S.; Abhayaratne, C. COVID-19 diagnostic using 3D deep transfer learning for classification of volumetric computerised tomography chest scans. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 8573–8577. [\[CrossRef\]](#)
28. Wang, X.; Deng, X.; Fu, Q.; Zhou, Q.; Feng, J.; Ma, H.; Liu, W.; Zheng, C. A Weakly-Supervised Framework for COVID-19 Classification and Lesion Localization from Chest CT. *IEEE Trans. Med. Imaging* **2020**, *39*, 2615–2625. [\[CrossRef\]](#)
29. Yang, D.; Martinez, C.; Visuña, L.; Khandhar, H.; Bhatt, C.; Carretero, J. Detection and analysis of COVID-19 in medical images using deep learning techniques. *Sci. Rep.* **2021**, *11*, 19638. [\[CrossRef\]](#)
30. Yang, D.; Martinez, C.; Visuña, L.; Khandhar, H.; Bhatt, C.; Carretero, J. Diagnosis of COVID-19 using CT scan images and deep learning techniques. *Emerg. Radiol.* **2021**, *28*, 497–505. [\[CrossRef\]](#)
31. Sitaula, C.; Aryal, S. New bag of deep visual words based features to classify chest x-ray images for COVID-19 diagnosis. *Health Inf. Sci. Syst.* **2021**, *9*, 24. [\[CrossRef\]](#)
32. Sitaula, C.; Shahi, T.B.; Aryal, S.; Marzbanrad, F. Fusion of multi-scale bag of deep visual words features of chest X-ray images to detect COVID-19 infection. *Sci. Rep.* **2022**, *11*, 23914. [\[CrossRef\]](#)
33. Sitaula, C.; Hossain, M.B. Attention-based VGG-16 model for COVID-19 chest X-ray image classification. *Appl. Intell.* **2021**, *51*, 2850–2863. [\[CrossRef\]](#)
34. Simpson, A.L.; Antonelli, M.; Bakas, S.; Bilello, M.; Farahani, K.; Van Ginneken, B.; Kopp-Schneider, A.; Landman, B.A.; Litjens, G.; Menze, B.; et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *Meml. Sloan Kettering Cancer Cent.* **2019**, *12*. [\[CrossRef\]](#)
35. Pedrosa, J.; Aresta, G.; Ferreira, C.; Rodrigues, M.; Leitão, P.; Carvalho, A.S.; Rebelo, J.; Negrão, E.; Ramos, I.; Cunha, A.; et al. LNDb: A Lung Nodule Database on Computed Tomography. *arXiv* **2019**, arXiv:1911.08434. [\[CrossRef\]](#)
36. Yang, J. Data from lung CT segmentation challenge. *Cancer Imaging Arch.* **2017**, *20*. Available online: <https://wiki.cancerimagingarchive.net/display/Public/Lung+CT+Segmentation+Challenge+2017> (accessed on 21 June 2022).
37. Morozov, S.P.; Andreychenko, A.E.; Blokhin, I.A.; Gelezhe, P.B.; Gonchar, A.P.; Nikolaev, A.E.; Pavlov, N.A.; Chernina, V.Y.; Gombolevskiy, V.A. MosMedData: Data set of 1110 chest CT scans performed during the COVID-19 epidemic. *Digit. Diagn.* **2020**, *1*, 49–59. [\[CrossRef\]](#)
38. Jun, M.; Cheng, G.; Yixin, W.; Xingle, A.; Jiantao, G.; Ziqi, Y.; Mingqing, Z.; Xin, L.; Xueyuan, D.; Shucheng, C.; et al. COVID-19 CT Lung and Infection Segmentation Dataset. 2020. Available online: <https://zenodo.org/record/3757476#.YrPlcOxByUk> (accessed on 21 June 2022).
39. MedSeg, H.; Jenssen, B.; Sakinis, T. MedSeg COVID Dataset 1. May 2021. Available online: https://figshare.com/articles/dataset/MedSeg_Covid_Dataset_1/13521488/2 (accessed on 21 June 2022).
40. de la Iglesia Vayá, M.; Saborit, J.M.; Montell, J.A.; Pertusa, A.; Bustos, A.; Cazorla, M.; Galant, J.; Barber, X.; Orozco-Beltrán, D.; García-García, F.; et al. BIMCV COVID-19+: A large annotated dataset of RX and CT images from COVID-19 patients. *IEEE Dataport* **2021**. [\[CrossRef\]](#)
41. NVIDIA Clara Imaging | NVIDIA Developer. Available online: <https://developer.nvidia.com/clara-medical-imaging> (accessed on 12 January 2022).
42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [\[CrossRef\]](#)
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [\[CrossRef\]](#)
44. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L. Keras Tuner. Available online: <https://github.com/keras-team/keras-tuner> (accessed on 21 May 2019).
45. Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *J. Mach. Learn. Res.* **2018**, *18*, 6765–6816. [\[CrossRef\]](#)
46. Laino, M.; Ammirabile, A.; Posa, A.; Cancian, P.; Shalaby, S.; Savevski, V.; Neri, E. The Applications of Artificial Intelligence in Chest Imaging of COVID-19 Patients: A Literature Review. *Diagnostics* **2021**, *11*, 1317. [\[CrossRef\]](#)
47. Abelaira, M.D.C.; Abelaira, F.C.; Ruano-Ravina, A.; Fernández-Villar, A. Use of Conventional Chest Imaging and Artificial Intelligence in COVID-19 Infection. A Review of the Literature. *Open Respir. Arch.* **2021**, *3*, 100078. [\[CrossRef\]](#)
48. Rajpurkar, P.; Joshi, A.; Pareek, A.; Ng, A.Y.; Lungren, M.P. CheXternal: Generalization of deep learning models for chest X-ray interpretation to photos of chest X-rays and external clinical settings. In Proceedings of the Conference on Health, Inference, and Learning, New York, NY, USA, 8–10 April 2021; pp. 125–132. [\[CrossRef\]](#)

5 CONCLUSÃO

Neste projeto dois algoritmos foram desenvolvidos e testados com o objetivo de validar a utilização de aprendizagem profunda no suporte ao diagnóstico de COVID-19 nos exames de imagem por raios X e tomografia computadorizada. Foram realizadas avaliações estatísticas que indicaram que os dois modelos propostos possuem a qualidade necessária para suportar a identificação de pneumonia por COVID-19 e, portanto, podem ser avaliados como ferramentas de apoio ao processo de triagem e diagnóstico da doença, especialmente dos casos mais severos e urgentes, onde as alterações nos exames de imagens são mais presentes.

O algoritmo Cimatec_XCOV19 realiza a classificação de exames de raios X como normais ou anormais e avalia a probabilidade de pneumonia por COVID-19. Quando testado sobre uma base de 1158 radiografias de tórax de um hospital do Espírito Santo obteve uma sensibilidade de 0,85, especificidade de 0,82 e AUC ROC de 0,93. Os experimentos comparativos do algoritmo Cimatec_XCOV19 com outro classificador desenvolvido por um centro de referência europeu apresentou resultados semelhantes, demonstrando a boa qualidade do algoritmo. A ferramenta pode ser particularmente relevante em locais com poucos recursos especializados para diagnóstico de COVID-19 disponíveis.

O algoritmo CIMATEC-CovNet-19 foi capaz de realizar a predição de pneumonia por COVID-19 nos exames de TC utilizando apenas 16 imagens (*slices*), uma quantidade menor de imagens que a prática observada na literatura. Esta característica é relevante pois torna o algoritmo menos dependente de recursos computacionais para o treinamento e inferência. Permite também ampliar o alcance do modelo para um número maior de aparelhos. O modelo alcançou uma sensibilidade de 0,88, especificidade de 0,88, ROC-AUC de 0,95, PR-AUC de 0,95 e F1-score de 0,88 em um conjunto de testes com 414 amostras. Esses resultados validaram o Cimatec-CovNet-19 como uma boa ferramenta de triagem para pneumonia por COVID-19.

Para a realização do projeto foram mapeados sete bancos de dados públicos de TC do tórax e cinco bases de dados públicas de exames de raios-x

do tórax. Foram desenvolvidas parceiras com os hospitais Santa Izabel de Salvador, Medsenior de Vitória do Espírito Santo e o HC de São Paulo, para coleta de imagens anonimizadas. Esses dados foram tratados e alimentaram o treinamento e testes dos sistemas. As bases estruturadas estão disponíveis no centro de supercomputação e inovação industrial do SENAI CIMATEC e poderão ser usados em projetos de pesquisas futuros. São quase 2 TB de dados disponíveis, com um contingente importante de dados brasileiros, representando o fenótipo nacional. Os códigos fonte dos algoritmos estão publicamente disponíveis. Incentivamos trabalhos futuros que utilizem os dados disponíveis, que comparem os algoritmos atuais com outros algoritmos disponíveis publicamente e avancem a qualidade dos modelos.

A utilização de técnicas de explicabilidade, em especial o GRAD-CAM, permitiu identificar ao longo do processo de desenvolvimento falhas e vieses, auxiliando a correção dos mesmos. A avaliação final indicou que os algoritmos de fato utilizaram as áreas corretas do pulmão para identificar a possibilidade de infecção.

O resultado mais expressivo e desejado desta pesquisa seria a utilização do conhecimento desenvolvido na prática clínica. Para isso algumas etapas de pesquisa e desenvolvimento ainda são necessárias. Há muitas oportunidades de melhoria, necessárias para adoção dos sistemas em ambientes reais. O treinamento dos modelos utilizando um grande volume de dados oriundos de bases públicas, sem informações demográficas associadas, pode trazer vieses desconhecidos. Seria importante ampliar o treinamento dos algoritmos com a utilização de dados mais completos contendo as informações demográficas e preferencialmente informações clínicas complementares. De fato, uma abordagem muito promissora é construir uma ferramenta de diagnóstico para COVID-19 integrando informações de imagem com dados clínicos de pacientes e informações epidemiológicas.

O algoritmo Cimatec_XCOV19 está em testes em um ambiente controlado, integrado ao sistema de PACS da rede Medsenior, onde passa pela avaliação dos médicos da rede. Esta é uma validação importante, uma vez que existem muitos desafios de integração dos modelos de IA aos sistemas e processos da prática clínica. Precisam ser avaliadas as questões técnicas relacionadas à qualidade do modelo e a interoperabilidade entre os sistemas de

imagem e os sistemas de diagnóstico e gestão. Principalmente, é uma oportunidade de observar como a ferramenta se integra aos protocolos médicos estabelecidos, verificando como os profissionais de saúde se beneficiam dela e entendendo eventuais resistências culturais.

Com a vacinação em massa e o surgimento de novas variantes da SARS-Cov-2 menos agressivas ao pulmão, a utilização de algoritmos para identificação de pneumonia por COVID-19 tornou-se menos urgente. Entretanto, a metodologia de trabalho utilizada nesta tese e as lições aprendidas podem ser utilizadas como referência para o desenvolvimento de modelos para outras enfermidades. Os algoritmos desenvolvidos podem rapidamente ser adaptados e aplicados a outras infecções de pulmão ou eventualmente a novas pandemias.

Em um país com limitações importantes de recursos como o Brasil, ferramentas de IA como as descritas nessa tese podem e devem ser pensadas como alternativas para um sistema de saúde mais eficiente.