



Faculdade de Medicina de São José do Rio Preto
Programa de Pós-Graduação em Ciências da Saúde

DIEGO DA COSTA BERNARDO

**INTELIGÊNCIA ARTIFICIAL APLICADA NA
DETECÇÃO DE PNEUMONIA EM IMAGENS
DE RAIO-X DO TÓRAX**

São José do Rio Preto
2023

Diego da Costa Bernardo

Inteligência Artificial Aplicada na Detecção de
Pneumonia em Imagens de Raio-X do Tórax

Dissertação apresentada à Faculdade de
Medicina de São José do Rio Preto
para obtenção do Título de Mestre no
Curso de Pós-graduação em Ciências
da Saúde, Eixo Temático: Diagnóstico
por Imagem.

Orientador: Prof. Dr. Arthur Soares Souza
Júnior

São José do Rio Preto
2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESSE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Bernardo, Diego da Costa.
Inteligência Artificial Aplicada na Detecção de Pneumonia em
Imagens de Raio-X do Tórax.
São José do Rio Preto, 2023
7 p.

Dissertação (Mestrado) – Faculdade de Medicina de São José do Rio
Preto – FAMERP

Eixo Temático: Medicina e Ciências Correlatas

Orientador: Prof. Dr. Arthur Soares Souza Júnior

1. Inteligência Artificial; 2. Pneumonia; 3. Diagnóstico por Imagem; 4.
Raio-x; 6. Pulmão.

DIEGO DA COSTA BERNARDO

Inteligência Artificial Aplicada na Detecção de
Pneumonia em Imagens de Raio-X do Tórax

BANCA EXAMINADORA

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU
DE MESTRE

Presidente e Orientador: Prof. Dr. Arthur Soares
Souza Júnior

1º Examinador: Prof. Dr. Henrique Dezani

2º Examinador: Profa. Dra. Luciana Vargas Cardoso

1º Suplente: Prof. Dr. José Fernando Vilela Martin

2º Suplente: Prof. Dr. Lucimar Sasso Vieira

São José do Rio Preto, 08/03/2023

SUMÁRIO

DEDICATÓRIA	i
AGRADECIMENTOS	ii
EPÍGRAFE	iii
LISTA DE FIGURAS	iv
LISTA DE TABELAS	v
LISTA DE ABREVIATURAS E SÍMBOLOS	vi
RESUMO	vii
ABSTRACT	viii
1. INTRODUÇÃO	1
1.1 Delimitação do Escopo	3
2. OBJETIVO	5
2.1 Objetivo Geral	5
2.2 Objetivos Específicos	5
3. MATERIAL E MÉTODO	6
3.1 Pneumonia e Raio-X	9
3.2 O conjunto de dados	10
3.3 Etapas do projeto	11
3.3.1 Análise Exploratória dos Dados	11
3.4 Construindo e Treinando o algoritmo	12
3.4.1. Criando o conjunto de dados de Treinamento e Teste	12
3.5 Arquitetura do algoritmo	13
3.5.1 Pré-processamento e aumento de dados	13
3.5.2 Treinamento	13
3.5.3 Avaliação do desempenho	14
3.6 Análise de Dados	14
3.6.1 Análise Exploratória dos Dados	14
3.6.2 Conclusões da Análise Exploratória de Dados	23
4. RESULTADOS	24
4.1. Criando o conjunto de dados de treinamento e teste	24
4.2 Pré-Processamento e Aumento de Dados	29
4.3 Imagens sintéticas para aumento de dados	29
4.4 Treinamento	31
4.5 Parâmetros para treinamento	32

4.6 Avaliação do Desempenho	37
4.7 Simulação do Sistema em Uso.	44
5. DISCUSSÃO	47
6. CONCLUSÕES	52
7. REFERÊNCIAS	53

DEDICATÓRIA

À minha esposa Luana, por cuidar de mim e dos nossos filhos, enquanto eu estudava e trabalhava. Sem o seu apoio este projeto não teria acontecido.

Aos meus filhos, Arthur e Manuela, por me inspirarem a ser uma pessoa melhor a cada dia.

Aos meus pais, Miguel e Érica, por toda criação e educação que vocês me deram. Se hoje sou um homem de bem, é por causa de vocês.

A Deus, pelo maior milagre que Ele fez em mim, concedeu-me a vida.

Eu me sinto amado por vocês.

AGRADECIMENTOS

Ao Prof. Dr. Arthur Soares Souza Júnior, pela oportunidade concedida a mim como seu orientando no programa de Pós-Graduação.

À toda equipe da FAMERP, pelo auxílio durante essa jornada.

Aos professores da FAMERP, que não mediram esforços para compartilhar o seu conhecimento.

À CAPES, por investir e acreditar neste projeto.

EPÍGRAFE

*Aquele que anda com os sábios será cada vez mais sábio, mas o companheiro dos tolos
acabará mal.
Provérbios 13:20
(Bíblia Sagrada, tradução NVI - Nova Versão Internacional)*

LISTA DE FIGURAS

Figura 1.	Ciclo de desenvolvimento de uma aplicação de IA para radiologia...	08
Figura 2.	Interação, Treinamento e Teste.....	09
Figura 3.	Número de registros por patologia.....	17
Figura 4.	Quantidade de pneumonia em conjunto com outras doenças.....	18
Figura 5.	Relação das patologias por faixa etária.....	19
Figura 6.	Homem x Mulher no conjunto de dados.....	20
Figura 7.	Relação Homem x Mulher com pneumonia.....	20
Figura 8.	Quantidade de imagens com projeção AP e PA.....	21
Figura 9.	Histograma de intensidade das imagens por patologia.....	22
Figura 10.	Histograma: Pneumonia, Consolidação e Espessamento Pleural.....	23
Figura 11.	Imagens após aplicação de técnicas de aumento de dados.....	30
Figura 12.	Evolução da acurácia na fase de treinamento.....	33
Figura 13.	Gráfico de treinamento e validação.....	39
Figura 14.	Curva ROC.....	40
Figura 15.	Curva ROC – Validação.....	41
Figura 16.	Gráfico Precision vs Recall.....	42
Figura 17.	Gráfico F1-score vs Threshold.....	44
Figura 18.	Sistema web.....	44
Figura 19.	Fluxo de utilização do sistema.....	46

LISTA DE TABELAS

Tabela 1.	Percentual de patologias no conjunto de dados.....	15
Tabela 2.	Quantidade de Cada Patologia no Conjunto de Dados.....	16
Tabela 3.	Patologia por paciente.....	16
Tabela 4.	Proporção do conjunto de dados de treinamento e teste.....	24
Tabela 5.	Arquitetura com 19 camadas da rede VGG16.....	27
Tabela 6.	Arquitetura com 8 camadas adicionadas a rede VGG16.....	28

LISTA DE ABREVIATURAS E SÍMBOLOS

AED	Análise Exploratória de Dados
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANVISA	Agência Nacional de Vigilância Sanitária
AP	Antero-Posterior
AUC	Area Under the Curve
CNN	Convolutional Neural Networks
Curva COR	Curva Característica de Operação do Receptor
DICOM	Digital Imaging and Communications in Medicine
EDA	Exploratory Data Analysis
FDA	Food and Drug Administration
IA	Inteligência Artificial
NIH	National Institutes of Health
NLP	Natural Language Processing
PA	Pósterio-Anterior
PLN	Processamento de Linguagem Natural
RGB	Red, Green, Blue
RNA	Rede Neural Artificial
ROC Curve	Receiver Operating Characteristic Curve

RESUMO

INTRODUÇÃO: Profissionais especializados existem em um número menor do que a demanda existente no mercado; isso ocorre em várias áreas de trabalho e não é diferente quando se trata de médicos radiologistas. Devido ao grande volume de exames para serem analisados, os médicos radiologistas acabam se tornando um gargalo no fluxo de liberação de exames por imagem. Quando se trata de diagnósticos por imagem em um cenário de emergência hospitalar podemos citar dois grandes problemas. O primeiro é o médico não radiologista que não possui a expertise necessária para realizar um diagnóstico a partir de uma imagem. O segundo é a ordem na qual essas imagens são analisadas, muitas vezes problemas mais sérios, que deveriam ser analisados com prioridade, são analisados posteriormente devido à ordem de chegada e ordem na qual o exame foi realizado. **OBJETIVO:** Desenvolvimento e aplicação de um sistema de Inteligência Artificial capaz de identificar pneumonia em imagens de Raio-X do tórax, como proposta de auxílio aos médicos na triagem e no diagnóstico por imagens, contribuindo na rapidez e assertividade do diagnóstico. **MÉTODOS:** Estudo de avaliação tecnológica de corte transversal com delineamento descritivo, de natureza aplicada através de procedimentos experimentais, com abordagem quantitativa do tipo analítica. O sistema de inteligência artificial desta pesquisa foi desenvolvido pelo próprio pesquisador seguindo boas práticas de desenvolvimento de software. A avaliação da performance do sistema no diagnóstico de pneumonia em imagens de Raio-x foi realizada utilizando-se os modelos de avaliação de performance mais utilizados em sistemas de IA e Machine Learning. Dentre as métricas de performance utilizada na avaliação do sistema podemos citar a curva ROC, F1 score, precision, recall e acurácia. A avaliação de performance do sistema foi realizada utilizando-se o conjunto de imagens disponíveis para a pesquisa, separados em conjunto de treinamento e conjunto de validação. **RESULTADOS:** A arquitetura final do Algoritmo ficou com um total de 27 camadas. O Algoritmo foi treinado utilizando 2.290 imagens, e para avaliar a sua performance foram utilizadas 1.430 imagens. Foram definidas inicialmente 40 rodadas de treinamento, mas ele foi interrompido na rodada 37 por estabilizar o aprendizado e não alcançar mais nenhuma melhoria na acurácia. O Algoritmo alcançou no conjunto de validação uma área sobre a curva de 0.68. O algoritmo foi disponibilizado em um sistema web capaz de receber um arquivo no formato DICOM, processar a imagem e retornar a probabilidade da imagem ser positiva para pneumonia. **CONCLUSÃO:** O sistema de inteligência artificial para identificação de pneumonia em imagens de Raio-x do tórax tem a capacidade de melhorar a eficiência dos diagnósticos por imagem, sendo uma ferramenta de auxílio ao médico no momento da triagem e priorização dos exames a serem analisados, bem como, uma segunda opinião para o diagnóstico do exame. Dessa maneira o sistema contribui para agilizar a liberação de resultados importantes ao mesmo tempo que contribui para o aumento da assertividade nos diagnósticos.

PALAVRAS-CHAVE: Inteligência Artificial; Pneumonia; Diagnóstico por Imagem; Raio-x; Pulmão.

ABSTRACT

INTRODUCTION: There are fewer specialized professionals than the available demand in the market; this is the case in many fields of work, and it is no exception when this refers to radiologists. Due to the large volume of exams to be analyzed, radiologists are considered as a bottleneck in the flow of imaging exams released. When it comes to diagnostic imaging in a hospital emergency setting; we can point out two major problems. The first is the non-radiologist physician who does not have the enough expertise to provide a diagnosis from an image. The problem is the order in which these images are analyzed; often more serious problems, which should be analyzed with priority, are analyzed later due to the order of arrival and the order in which the exam was performed. **OBJECTIVES:** Development and application of an Artificial Intelligence system capable of identifying pneumonia in chest X-ray images, as a proposal to assist physicians in screening and imaging diagnosis, contributing to the speed and assertiveness of the diagnosis. **METHODS:** A cross-sectional technological assessment study with a descriptive design; applied nature through experimental procedures, with a quantitative approach of analytical type. The artificial intelligence system in this research was developed by the researcher himself, following proper software development practices. The evaluation of the system performance in the diagnosis of pneumonia in X-ray images was performed using the most widely used performance evaluation models in AI and Machine Learning systems. Among the performance metrics used in the system evaluation we can mention the ROC curve, F1 score, precision, recall and accuracy. The system performance evaluation was carried out using the set of images available for the research, separated into training set and validation set. **RESULTS:** The final architecture of the algorithm had a total of 27 layers. The Algorithm was trained using 2,290 images and 1,430 images were used to evaluate its performance. Initially 40 training rounds were set, but it was stopped in round 37 for stabilizing learning and not achieving any further improvement in accuracy. The Algorithm achieved in the validation set an area under the curve of 0.68. The algorithm was made available in a web system capable of receiving a file in DICOM format, processing the image and returning the probability of the image being positive for pneumonia. **CONCLUSIONS:** The artificial intelligence system for identifying pneumonia in chest X-ray images has the ability to improve the efficiency of imaging diagnoses; being a tool to assist the physician at the time of screening and prioritization of exams to be analyzed, as well as a second opinion for the diagnosis of of these. In this way, the system can improve the speed in releasing important results and at the same time contributes to increase assertiveness in diagnoses.

KEYWORDS: Artificial Intelligence; Pneumonia; Medical Imaging; X-Ray; Lung.

1. INTRODUÇÃO

Quando olhamos para a evolução da tecnologia ao longo dos anos, observa-se não apenas uma maior capacidade de processar e armazenar dados; mais do que isso, verifica-se todas as áreas que serão beneficiadas com esse avanço tecnológico. Uma dessas áreas beneficiadas é a área da saúde. Dessa maneira, observamos uma relação direta entre o avanço e melhoria da saúde quando comparada ao avanço da tecnologia. Essa melhoria na saúde deve-se, em parte, ao avanço das máquinas, mais memória para processar dados mais rápidos, mais espaço de armazenamento, e tudo isso por um preço cada vez menor com o passar dos anos.

Começamos a perceber a tecnologia como uma grande aliada para melhoria dos serviços de saúde, contribuindo para reduzir os custos dos diagnósticos e tratamentos ao mesmo tempo que agiliza todo o processo. Outro grande benefício da tecnologia aplicada à saúde é a possibilidade de tornar o diagnóstico e tratamento mais acessíveis, principalmente, em regiões onde faltam médicos suficientes para atender a população; um problema latente em regiões distantes das capitais e grandes centros.

Mesmo nas capitais e nos grandes centros é possível identificar que a quantidade de médicos especialistas não é suficiente para atender a demanda e, quando pensamos no médico radiologista, isso se torna um problema grave, pois impacta diretamente no diagnóstico e início do tratamento de um paciente. Quando entendemos o fluxo de trabalho de um médico radiologista, rapidamente percebemos um gargalo no momento do diagnóstico, pois a todo instante chega um novo exame na fila para ser analisado e diagnosticado pelo médico. Se pensarmos em um ambiente de urgência e emergência de um hospital, fica evidente que a velocidade na qual um exame e diagnóstico são realizados, pode influenciar na vida ou morte de um paciente.

Dado este cenário, o presente trabalho propõe a aplicação de um sistema de inteligência artificial como forma de colaboração no diagnóstico por imagem, analisando imagens de Raio-x do tórax a partir de arquivos DICOM e retornando à probabilidade de o paciente analisado na imagem estar com pneumonia. O sistema de diagnóstico por imagem, utilizando Inteligência Artificial tem o objetivo de melhorar a eficiência, acurácia, usabilidade e confiança de serviços envolvendo imagens médicas. Aplicações utilizando Inteligência Artificial podem melhorar a eficiência dos serviços de saúde liberando o diagnóstico de um exame mais rápido do que um médico radiologista, considerando-se como já comentado uma infraestrutura robusta de tecnologia, capaz de processar em pouco tempo e armazenar um grande volume de dados médicos. Sistemas de Inteligência Artificial são considerados mais rápidos e mais baratos do que manter o trabalho de um médico radiologista; não se cansam; não adoecem e ainda podem trabalhar 24/7 além de continuar aprendendo para melhorar sua performance no diagnóstico.⁽¹⁾ Além da eficiência, sistemas de inteligência artificial podem realizar tarefas sem operadores humanos, reduzindo erros humanos. Portanto, podendo ser mais assertivos, conforme aprendem a cada novo diagnóstico e melhoram sua performance, a ponto de o sistema realizar o diagnóstico em uma imagem tão bem quanto um médico radiologista.

Ao falarmos sobre sistemas especialistas, realizando atividades humanas, é comum surgir o seguinte questionamento: Os médicos serão substituídos pelos sistemas inteligentes? E a resposta é não. Se olharmos a história, veremos que o advento de novas tecnologias trouxe a necessidade de novos conhecimentos e novas habilidades humanas, criando na maioria das vezes mais postos de trabalho. Portanto, quando pensamos em sistemas de inteligência artificial, não pensamos em substituição do ser

humano, mas pensamos como uma nova ferramenta para auxiliar o trabalho do médico. Imagine um motorista dirigindo com a ajuda de um GPS, é dessa forma que os sistemas de inteligência artificial serão incluídos no cotidiano dos médicos, podendo ser utilizados até como uma segunda opinião. Sobretudo, fazendo trabalhos repetitivos, liberando o médico para executar tarefas que são intrinsecamente humanas e não podem ser executadas por máquinas, atividades que envolvem pensamento abstrato, criativo ou ainda aqueles que envolvem empatia com o paciente.

Expertise, sabedoria, empatia, cuidado, compreensão do momento e situação são características puramente humanas e não podem ser automatizadas por uma máquina.

A inteligência artificial será capaz de revolucionar a indústria da saúde por meio dos sistemas especialistas, contudo, isso não se limita ao diagnóstico. Toda a cadeia, desde a aquisição dos dados, análise, armazenamento, distribuição, recuperação, diagnóstico, sugestão e acompanhamento de tratamento terão a presença da inteligência artificial, contribuindo em algum ponto do processo. Portanto, para que essa revolução seja possível é imprescindível o trabalho em conjunto dos profissionais de tecnologia e da saúde.

1.1 Delimitação do Escopo

O presente trabalho tem como objetivo a identificação de pneumonia em imagens de Raio-x do tórax, portanto, qualquer funcionalidade diferente desta citada estará fora do escopo deste projeto.

Podemos citar como itens fora do escopo deste projeto, não se limitando a eles, a identificação de qualquer outra patologia diferente de pneumonia. Identificação em imagens que não sejam imagens de Raio-x, por exemplo, imagens de PET-SCAN ou de ressonância magnética. Imagens de Raio-x que não sejam do tórax também são consideradas fora do escopo do projeto. O sistema desenvolvido neste projeto trabalha

com os dados no formato DICOM, portanto qualquer dado que não esteja no formato DICOM está fora do escopo deste projeto.

2. OBJETIVO

2.1 Objetivo Geral

Desenvolver um algoritmo de inteligência artificial capaz de detectar pneumonia em imagens de Raio-X do tórax.

2.2 Objetivos Específicos

Construir um sistema web para receber imagens de Raio-x do tórax no formato DICOM e retornar a probabilidade daquele paciente estar com pneumonia.

Disponibilizar o sistema em um endereço na web de forma que ele seja acessível a qualquer profissional com acesso à internet, em qualquer parte do mundo, servindo como uma ferramenta de auxílio ao médico no processo de priorização das atividades e no diagnóstico.

3. MATERIAL E MÉTODO

Estudo de avaliação tecnológica de corte transversal com delineamento descritivo, de natureza aplicada através de procedimentos experimentais, com abordagem quantitativa do tipo analítica. Trata-se de uma pesquisa aplicada que visa o desenvolvimento e avaliação de desempenho de um algoritmo de inteligência artificial para detecção de pneumonia em imagens de Raio-x do tórax como uma proposta de auxílio ao médico radiologista no processo de priorização do fluxo de trabalho e no diagnóstico. A desempenho do algoritmo na detecção da pneumonia foi calculada utilizando as métricas curva ROC, F1 score, precision, recall e acurácia.

O desenvolvimento do algoritmo seguiu o processo de desenvolvimento de aplicações de inteligência artificial para a Radiologia, este processo possui as seguintes etapas: Definição do problema clínico, Definição da tecnologia de Inteligência Artificial, Definição do conjunto de dados, Anotação do conjunto de dados, Treinamento e Teste.^(1,2)

O início do processo de desenvolvimento começa com a definição do problema clínico, esta etapa irá guiar todas as fases seguintes, pois ter a definição do problema a ser resolvido de forma clara, objetiva e bem detalhada é essencial para a construção de uma aplicação que seja usual e que atenda a necessidade especificada. Para este projeto, a definição do problema clínico é a necessidade de uma ferramenta para auxiliar o médico radiologista na priorização do seu fluxo de trabalho e no diagnóstico de patologias pulmonares em imagens de Raio-x do tórax. Dado o seguinte problema, é proposto o desenvolvimento de um algoritmo de inteligência artificial para a detecção de pneumonia em imagens de Raio-x do tórax, o qual retornará a probabilidade de determinada imagem ser positiva para pneumonia.

Uma vez que a definição do problema clínico está clara o suficiente, começamos a definir quais das soluções de inteligência artificial disponíveis hoje é a mais indicada para a aplicação da solução proposta e que irá atender da melhor maneira a necessidade especificada. Para este projeto, dentre as categorias de algoritmos disponíveis hoje para o aprendizado de máquina, é utilizada a abordagem supervisionada do tipo classificação, pois este modelo se encaixa ao conjunto de dados anotado que foi utilizado no projeto e também ao resultado esperado que é um retorno, indicando se a imagem analisada possui indícios de pneumonia ou não.

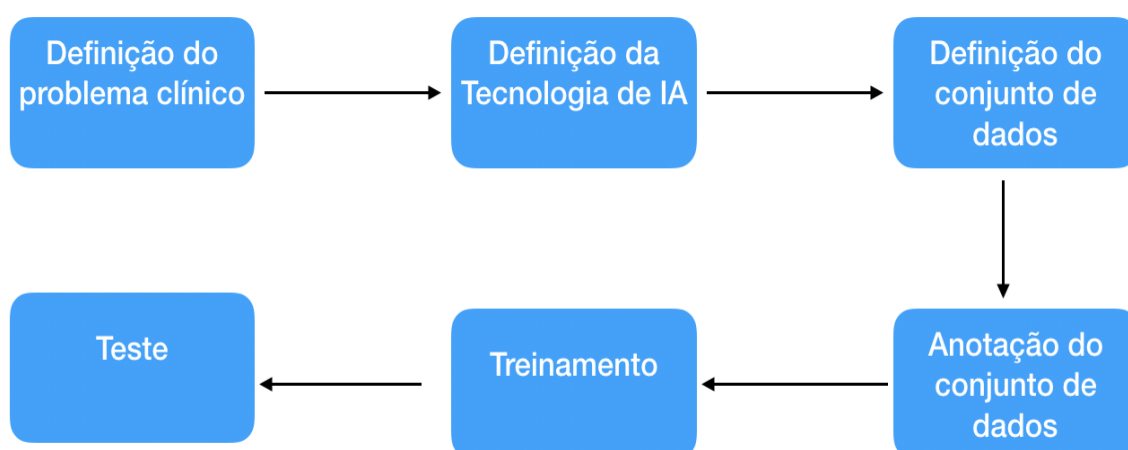
A definição do conjunto de dados é a etapa na qual são definidos os tipos de dados utilizados no desenvolvimento do projeto, todo sistema de inteligência artificial precisa de um conjunto de dados para treinar o algoritmo e entregar o resultado esperado. Para este projeto os dados utilizados para o treinamento e a avaliação do desempenho da aplicação foram cedidos pelo National Institute of Health (NIH), um dos maiores banco de dados de Raio-x do tórax disponível para a comunidade científica.⁽³⁾

Anotação do conjunto de dados é um grande problema para qualquer área que deseja desenvolver um estudo, utilizando inteligência artificial. A anotação de dados, para este projeto, significa saber se uma determinada imagem está anotada com o resultado de pneumonia, verdadeiro ou falso. Sem essa anotação no conjunto de dados, fica impossível treinar o algoritmo para que ele possa diferenciar imagens positivas para pneumonia de imagens negativas para pneumonia. O conjunto de dados que foram utilizados neste projeto já estão anotados, porém em uma situação na qual o conjunto de dados não está anotado. Portanto, é necessário a avaliação de um médico radiologista para analisar a imagem e a anotar como positiva ou negativa para a pneumonia.

O treinamento é uma das fases mais importantes durante o processo de desenvolvimento de uma aplicação de inteligência artificial. Nesta etapa, o algoritmo irá analisar cada imagem e comparar com sua respectiva anotação; no caso deste projeto, a anotação é positiva ou negativa para pneumonia, de modo a encontrar um padrão nas imagens que estão anotadas como positivas para pneumonia. Após encontrar esse padrão, a aplicação poderá receber uma imagem e, a partir do padrão que ele aprendeu no treinamento, poderá informar se essa imagem possui pneumonia ou não.

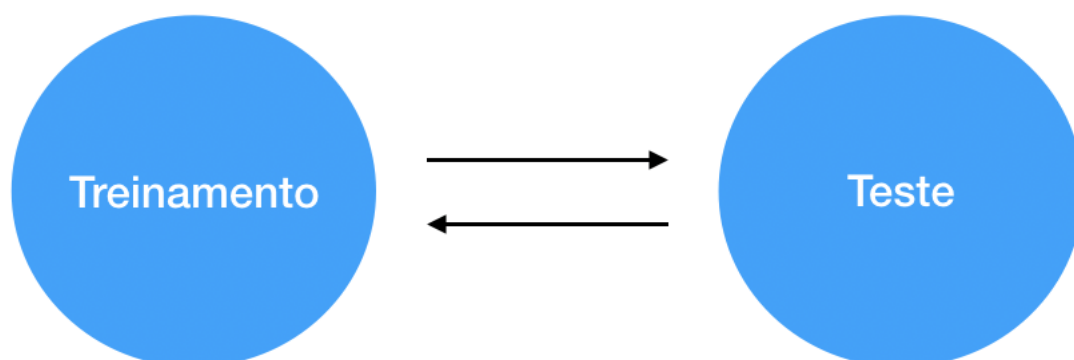
Por fim, chegamos à etapa de teste ou também conhecida como etapa de validação. Após a realização do treinamento, é feita a validação do algoritmo para avaliar o seu desempenho. Na validação, são enviadas ao algoritmo imagens com sua anotação ocultada para que a aplicação dê a sua resposta, após a aplicação retornar um resultado, é comparada a resposta da aplicação com a anotação da imagem e, a partir disso, é possível calcular a performance da aplicação e qual bem ou mal ela está indo na tarefa de classificar as imagens como positivas ou negativas para pneumonia.

Figura 1. Ciclo de desenvolvimento de uma aplicação de IA para Radiologia.



Importante ressaltarmos que durante o processo de desenvolvimento da aplicação pode haver uma interação constante entre as fases de Treinamento e Teste com o objetivo de melhorar o desempenho do algoritmo. Nessa interação são realizados ajustes na estrutura e nos parâmetros do algoritmo, a etapa de treinamento é executada novamente com o algoritmo alterado e depois de finalizado o treinamento, é realizada a etapa de teste para verificar se os ajustes realizados no algoritmo resultaram em melhora do desempenho. Essa interação pode continuar até que um nível de performance seja alcançado ou até o ponto em que os ajustes não resultem mais em melhoria na performance.

Figura 2. Interação, Treinamento e Teste



3.1 Pneumonia e Raio-X

A Radiografia de tórax é um dos exames de imagem médica mais frequentes e econômicos. A obtenção de diagnósticos clínicos a partir de radiografias de tórax pode ser um desafio, mesmo por radiologistas qualificados.⁽⁴⁾

Quando se trata de pneumonia, a Radiografia de tórax é um dos métodos disponíveis para o diagnóstico. A doença afeta por volta de 450 milhões de pessoas no mundo todo e cerca de quatro milhões morrem da doença a cada ano.⁽⁵⁻⁶⁾ A alta

prevalência de pneumonia o torna um bom candidato para o desenvolvimento de um aplicativo de aprendizado profundo por dois motivos:

1. Disponibilidade de dados em quantidade suficiente para treinar modelos de inteligência artificial para classificação de imagens;
2. Oportunidade de auxílio clínico fornecendo leituras de imagens de maior precisão de uma doença difícil de diagnosticar e/ou reduzir o esgotamento clínico realizando leituras automatizadas de exames muito comuns.

O diagnóstico de pneumonia a partir de radiografias de tórax é difícil por vários motivos:

1. O aparecimento de pneumonia em uma Radiografia de tórax pode ser muito vago, dependendo do estágio da infecção;
2. A pneumonia geralmente se sobrepõe a outros diagnósticos;
3. A pneumonia pode imitar anormalidades benignas.

Por essas razões, os métodos comuns de validação diagnóstica realizados no ambiente clínico são a obtenção de culturas de escarro para testar a presença de bactérias ou corpos virais causadores de pneumonia, leitura da história clínica do paciente e levar em consideração seu perfil demográfico. Em adição, comparar uma imagem para Radiografias de tórax anteriores para o mesmo paciente, se estiverem disponíveis.⁽⁸⁾

3.2 O conjunto de dados

O conjunto de dados utilizado neste projeto foi selecionado pelo NIH, especificamente, para resolver o problema da falta de grandes conjuntos de dados de Raio-x anotados para serem usados na criação de algoritmos de detecção de doenças.

O conjunto de dados foi obtido através do website Kaggle.⁽³⁾

Existem 112.120 imagens de Raios-x com rótulos de doenças de 30.805 pacientes únicos neste conjunto de dados. Os rótulos das doenças foram criados usando o Processamento de Linguagem Natural (NLP) para extrair os relatórios radiológicos associados. Os rótulos incluem 14 patologias torácicas comuns:

1. Atelectasia;
2. Consolidação;
3. Infiltração;
4. Pneumotórax;
5. Edema;
6. Enfisema;
7. Fibrose;
8. Efusão;
9. Pneumonia;
10. Espessamento pleural;
11. Cardiomegalia;
12. Nódulo;
13. Massa;
14. Hérnia.

3.3 Etapas do projeto

3.3.1 Análise Exploratória dos Dados

A primeira parte deste projeto envolverá análise exploratória de dados para entender e descrever o conteúdo e a natureza dos dados. Grande parte do trabalho realizado durante a análise exploratória de dados permitirá filtrar somente os dados relevantes para alcançar o objetivo proposto neste projeto, além de explicar em quais condições o algoritmo foi treinado, isso contribuirá para especificar em quais situações

e condições o algoritmo tem melhor performance e em quais condições ele não é indicado como uma ferramenta de auxílio, devido ao seu baixo desempenho com dados nos quais o algoritmo não foi treinado. Abaixo alguns pontos analisados durante a análise exploratória de dados:

1. Os dados demográficos do paciente, como sexo, idade, posição do paciente, etc;
2. As visualizações de Raios-x tiradas (ou seja, posição de visualização);
3. O número de casos, incluindo:
 - a) número de casos de pneumonia;
 - b) número de casos não-pneumonia;
4. A distribuição de outras doenças que são comórbidas com pneumonia;
5. Número de doenças por paciente;
6. Avaliações em nível de pixel dos dados de imagem para estados de interesse saudáveis e de doença (por exemplo, histogramas de valores de intensidade) e distribuições entre doenças.

3.4 Construindo e Treinando o algoritmo

3.4.1. Criando o conjunto de dados de Treinamento e Teste

A partir das descobertas na análise exploratória de dados deste projeto, foram selecionados os conjuntos de treinamento e testes apropriados para classificar pneumonia. Para que o treinamento e teste sejam efetivos, foi tomado o cuidado de levar em consideração os seguintes pontos:

1. Distribuição de outras doenças além da pneumonia que estão presentes em ambos os conjuntos de dados;
2. Informações demográficas, posições de visualização de imagens e número de imagens por paciente em cada conjunto;

3. Distribuição de casos positivos e negativos para pneumonia em cada conjunto de dados.

3.5 Arquitetura do algoritmo

Neste projeto, utilizamos uma arquitetura CNN existente para classificar imagens de Raios-x quanto à presença de pneumonia. Escolhemos utilizar uma abordagem de *Transfer Learning*, usando a arquitetura VGG16 com pesos treinados no conjunto de dados ImageNet. O ajuste fino foi realizado, congelando a rede pré-construída e adicionando várias novas camadas ao final para treinar. Essa abordagem permite um ganho de desempenho sem o custo de treinar todas as camadas da arquitetura do algoritmo, pois é usada uma rede já treinada na base do algoritmo.

3.5.1 Pré-processamento e aumento de dados

Foi realizado o pré-processamento antes de alimentar imagens em sua rede para treinamento e teste. Este procedimento serve para que os dados se adequem à arquitetura do algoritmo e/ou para aumentar seu conjunto de dados de treinamento para aumentar o desempenho. O aumento no conjunto de dados foi realizado para refletir situações reais no qual as imagens podem estar levemente inclinadas para a direita, esquerda ou ainda com um foco maior no tórax, excluindo as extremidades do corpo.

3.5.2 Treinamento

Durante o treinamento do algoritmo, alguns parâmetros foram ajustados para melhorar o desempenho, incluindo:

1. Parâmetros de aumento de imagem;
2. Tamanho do lote de treinamento;
3. Taxa de aprendizado de treinamento;
4. Inclusão e parâmetros de camadas específicas na arquitetura do algoritmo.

3.5.3 Avaliação do desempenho

À medida que o algoritmo é treinado, é realizado o monitoramento do seu desempenho em treinamento. Este procedimento serve para acompanharmos a evolução de aprendizado do algoritmo, enquanto ele está sendo treinado.

3.6 Análise de Dados

3.6.1 Análise Exploratória dos Dados

A análise exploratória dos dados é um processo aberto, ou seja, não existe uma regra definida de como ela deve ser iniciada ou realizada. O objetivo da análise exploratória dos dados é identificar as principais características do conjunto de dados, frequentemente apresentada em métodos visuais. A análise exploratória dos dados visa observar o que os dados podem nos dizer, quais informações relevantes podemos identificar ao analisarmos os dados que compõem o conjunto de dados utilizados na pesquisa.

A AED realizada neste projeto ajudou a identificar a aparência da pneumonia, com quais outros tipos de doenças a pneumonia é comumente encontrada e quais idades e gênero são mais afetados pela doença.

Um ponto importante a ser dito é que o conjunto de dados do NIH não foi adquirido especificamente para pneumonia. Portanto, o processo de AED torna-se primordial para filtrar somente os dados relevantes para este projeto, descartando outras doenças presentes no conjunto de dados, mas que estão fora do escopo do presente trabalho.

O conjunto de dados possui 112.120 registros.

Deste total de registros, 53% estão livres de doenças, ou seja, não foi encontrada neles nenhuma patologia.

Tabela 1. Percentual de patologias no conjunto de dados.

Percentage of diseases	
Atelectasis	0.103095
Cardiomegaly	0.024759
Consolidation	0.041625
Edema	0.020540
Effusion	0.118775
Emphysema	0.022440
Fibrosis	0.015037
Hernia	0.002025
Infiltration	0.177435
Mass	0.051570
No Finding	0.538361
Nodule	0.056466
Pleural Thickening	0.030191
Pneumonia	0.012763
Pneumothorax	0.047289

Abaixo se pode ver o número de vezes que cada doença aparece no conjunto de dados. (Tabela 2)

Tabela 2. Quantidade de Cada Patologia no Conjunto de Dados.

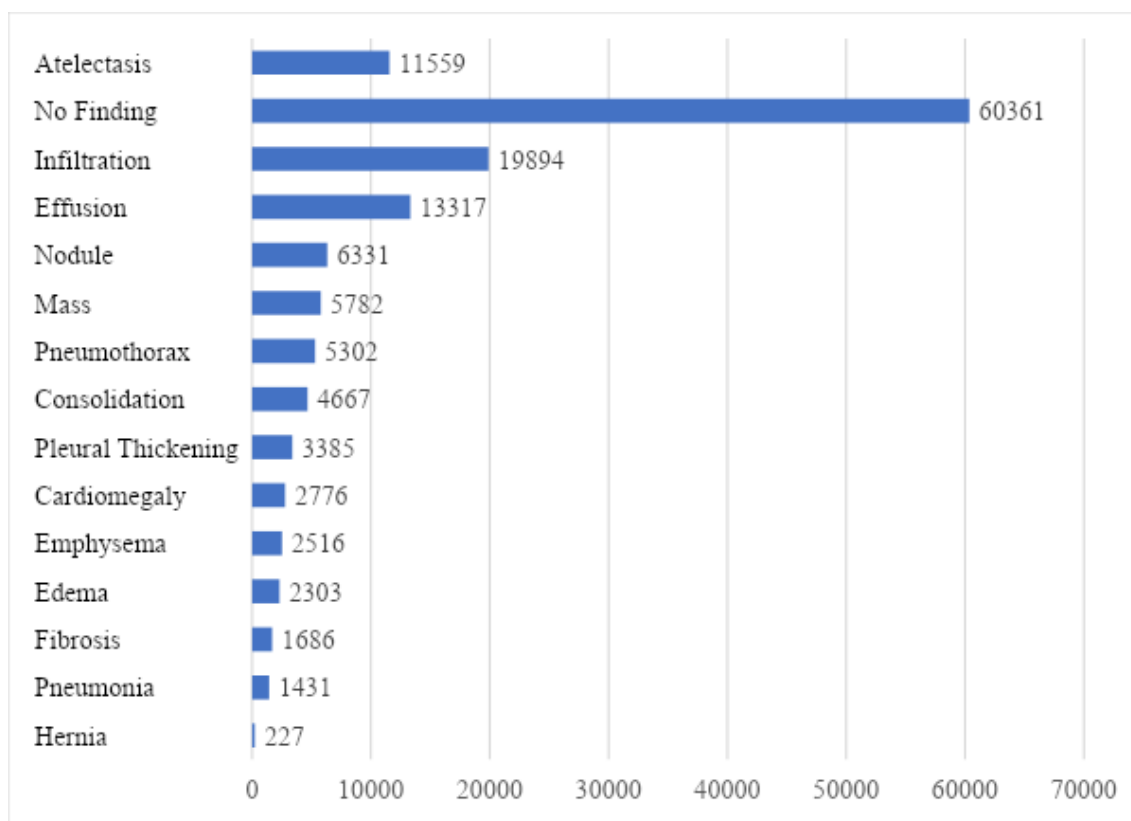
Number of each diseases	
Atelectasis	11559.0
Cardiomegaly	2776.0
Consolidation	4667.0
Edema	2303.0
Effusion	13317.0
Emphysema	2516.0
Fibrosis	1686.0
Hernia	227.0
Infiltration	19894.0
Mass	5782.0
No Finding	60361.0
Nodule	6331.0
Pleural Thickening	3385.0
Pneumonia	1431.0
Pneumothorax	5302.0

Abaixo, a relação patologia por paciente. (Tabela 3).

Tabela 3. Patologia por paciente.

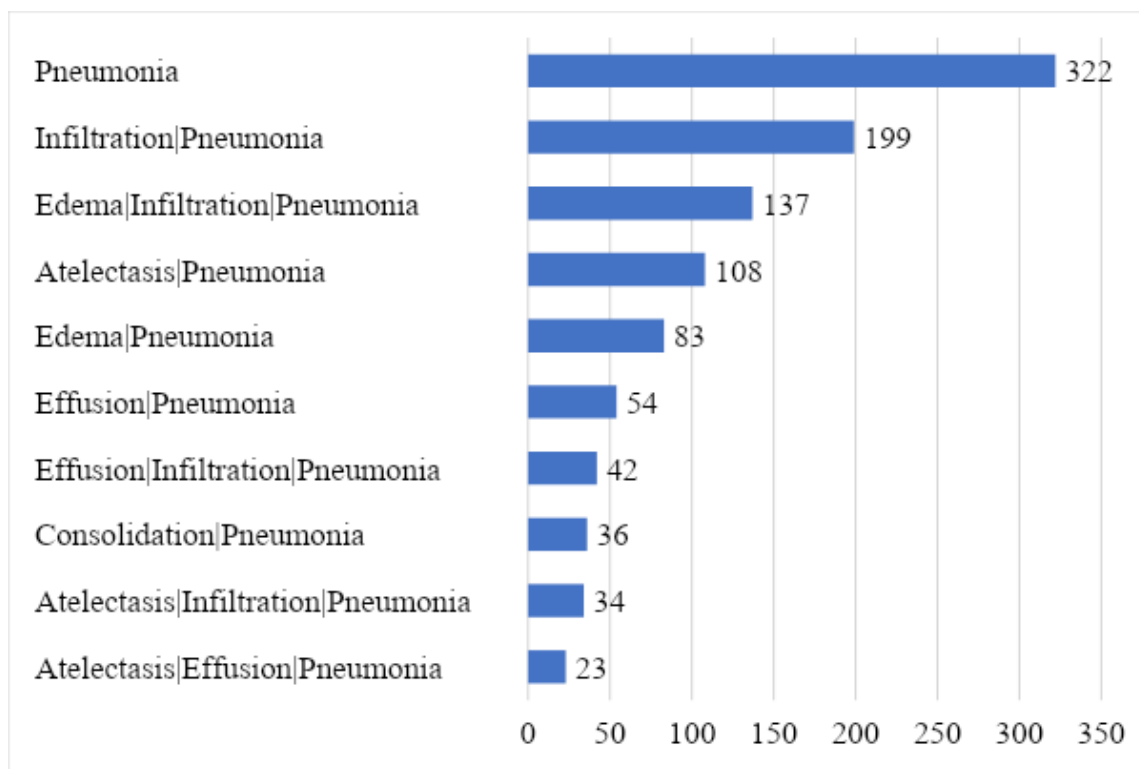
Number of Diseases	141537.0
Number of Patients	30805
Number of Diseases per Patient	4.594611264405129

Figura 3. Número de registros por patologia



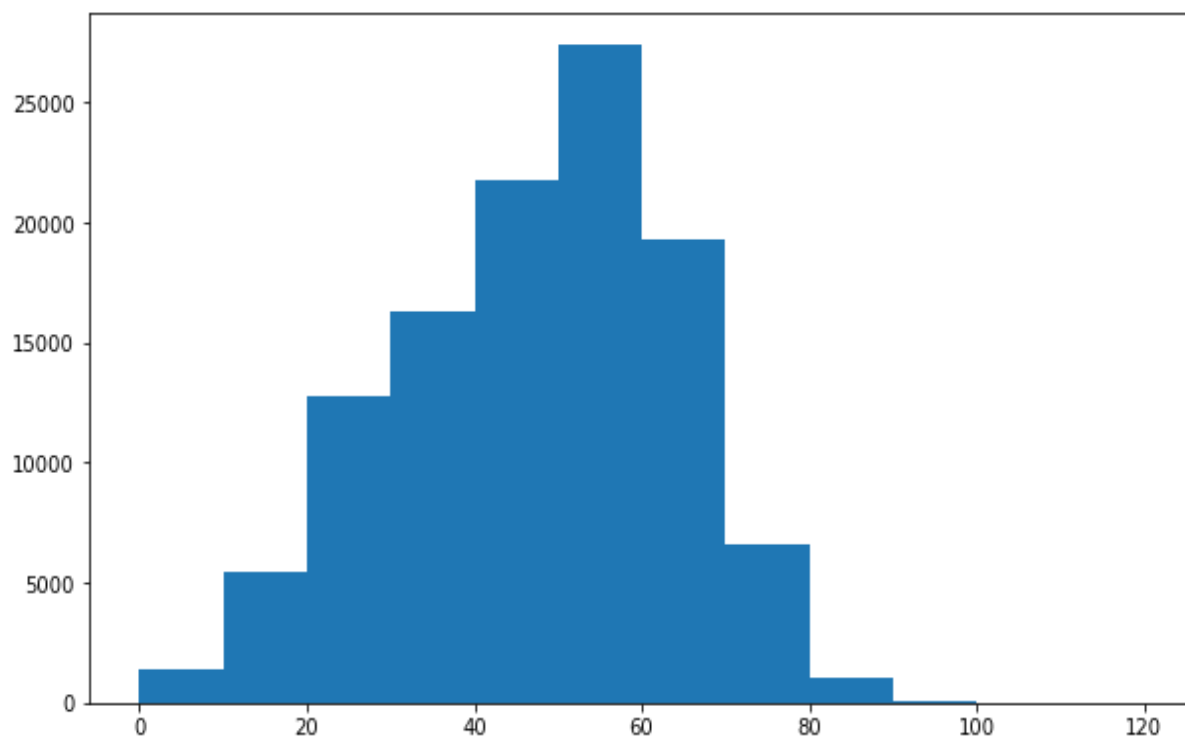
Na Figura 4 podemos notar outras patologias que frequentemente aparecem junto com a pneumonia, entre elas, nota-se a infiltração e o edema como patologias comumente apresentadas junto com a pneumonia.

Figura 4. Quantidade de pneumonia em conjunto com outras doenças.



No Gráfico abaixo, foi verificado que a maior parte das patologias encontradas em Raio-x do tórax estão presentes em pessoas com idade entre 50 e 60 anos. (Figura 5)

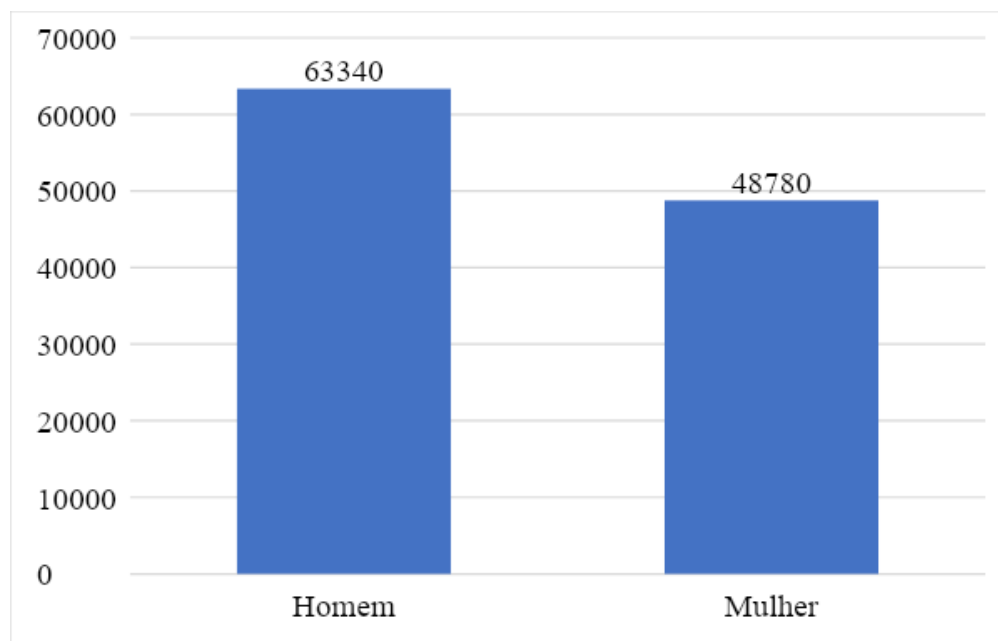
Figura 5. Relação das patologias por faixa etária.



Vejamos abaixo a relação Homem x Mulher no conjunto de dados. (Figura 6)

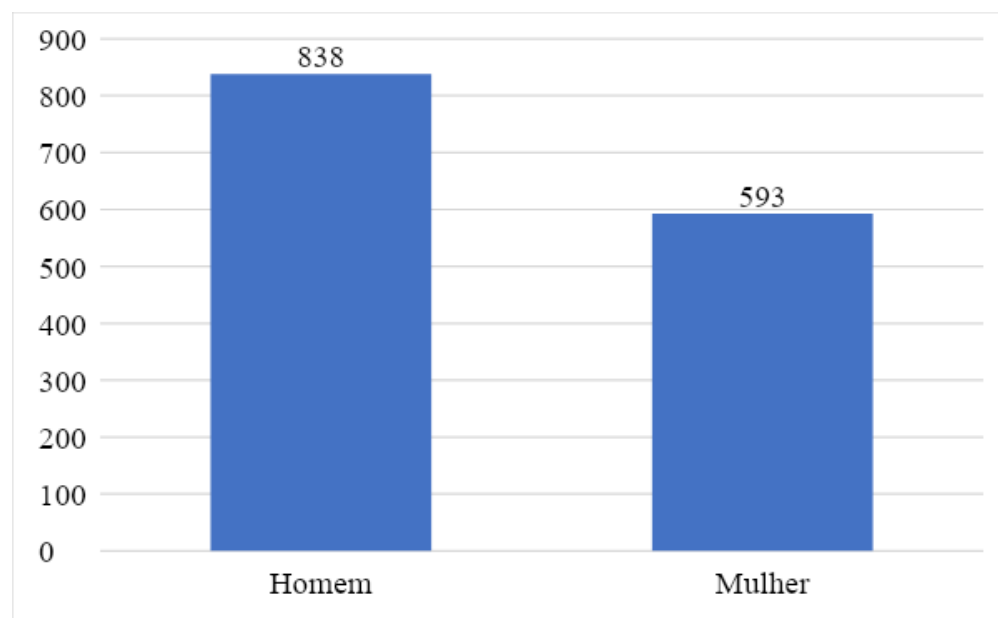
Percebe-se que no conjunto de dados há um pouco mais registros de homens do que mulheres.

Figura 6. Homem x Mulher no conjunto de dados.



Quando comparamos a relação de Homem x Mulher em registros positivos para pneumonia, constatamos que a relação se mantém próxima em relação a todo o conjunto de dados.

Figura 7. Relação Homem x Mulher com pneumonia.

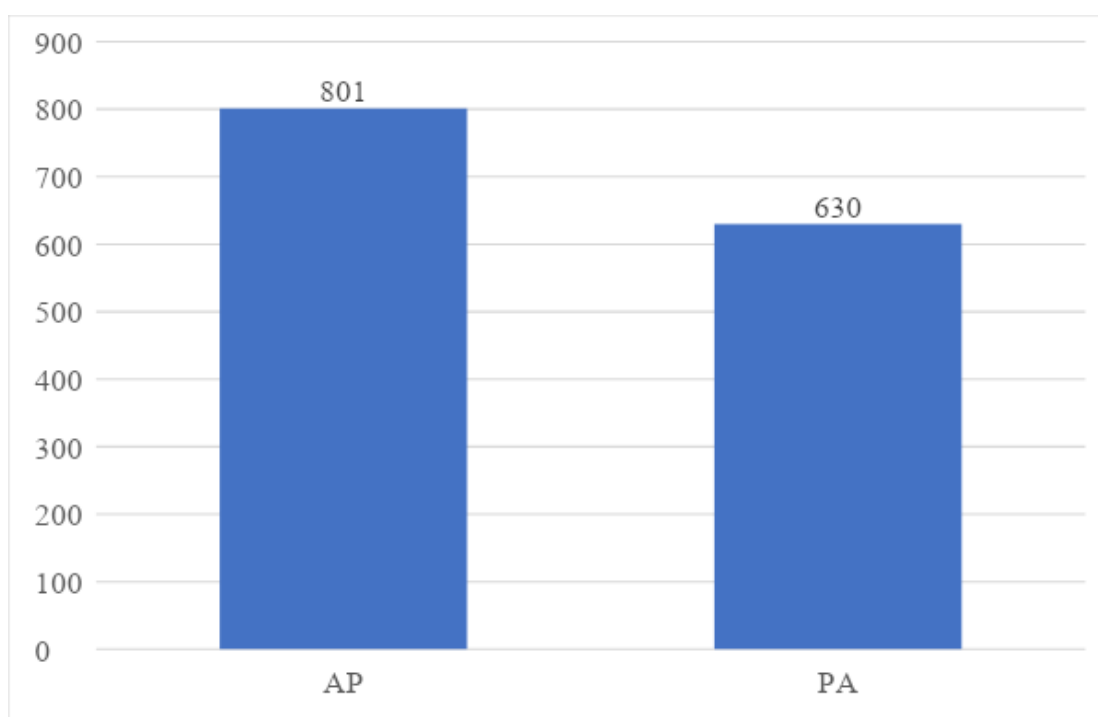


Um fator importante para o diagnóstico por imagem utilizando um sistema de inteligência artificial é saber diferenciar se a imagem de Raio-x foi obtida através da

projeção Pósterio-Anterior (PA) ou Antero-Posterior (AP). Essa mudança de posição da imagem interfere diretamente nos reconhecimentos de padrões pelo sistema, por isso, este é um fator importante a ser considerado no conjunto de dados que será utilizado para treinamento do sistema.

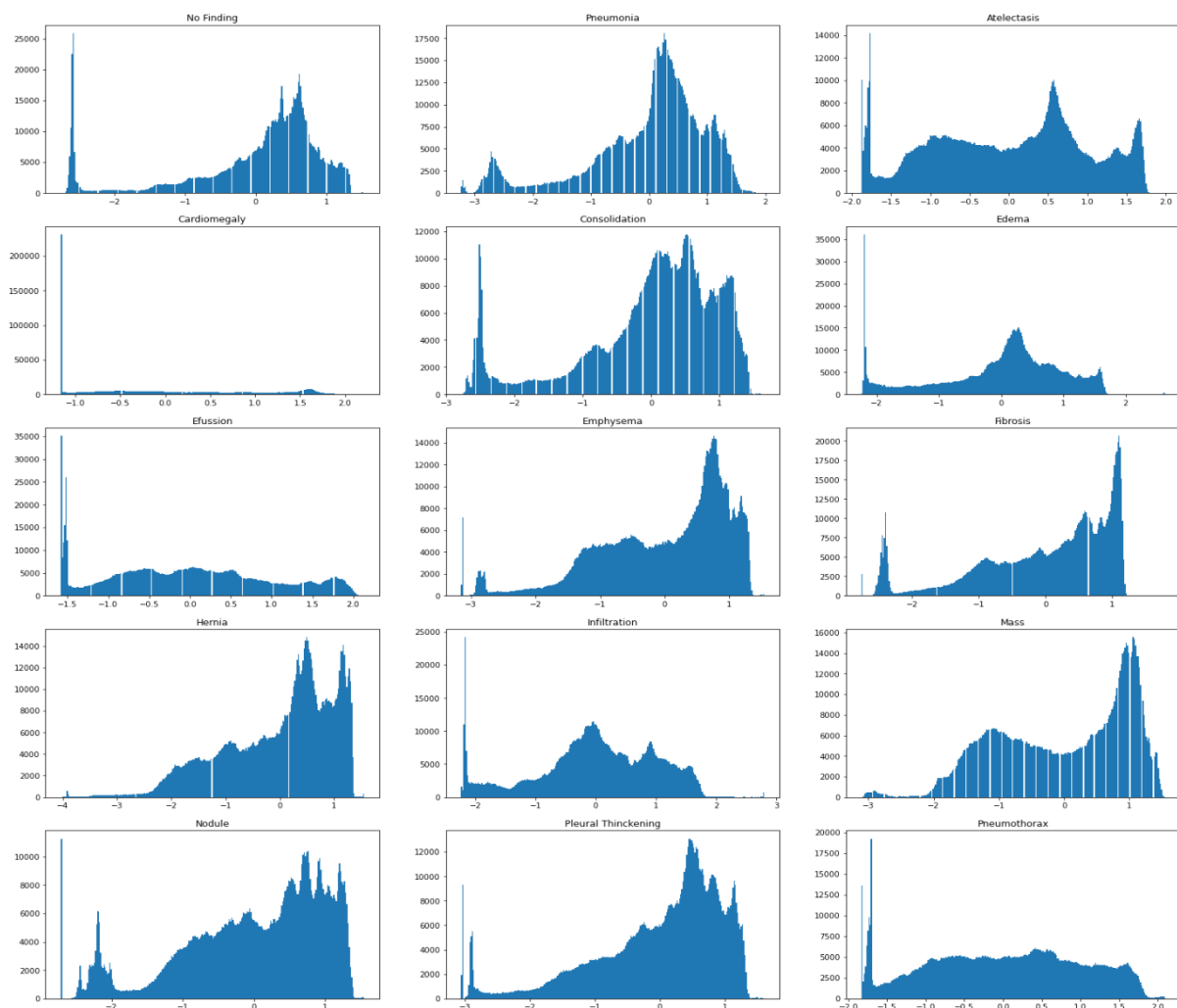
Abaixo vemos que o nosso conjunto de dados possui mais imagem AP para registros com pneumonia.

Figura 8. Quantidade de imagens com projeção AP e PA.



Na imagem abaixo podemos observar o histograma das intensidades das imagens separadas por patologia.

Figura 9. Histograma de intensidade das imagens por patologia.



Um fato curioso ao se analisar os histogramas de intensidade das patologias é identificar patologias com histogramas semelhantes. Na imagem abaixo, pode-se confirmar que o histograma de intensidade da consolidação e do espessamento pleural são semelhantes ao histograma da pneumonia. Dada essa identificação, faz-se necessário uma atenção redobrada para que o sistema não troque o diagnóstico de uma patologia por outra, devido ao seu padrão semelhante nas imagens.

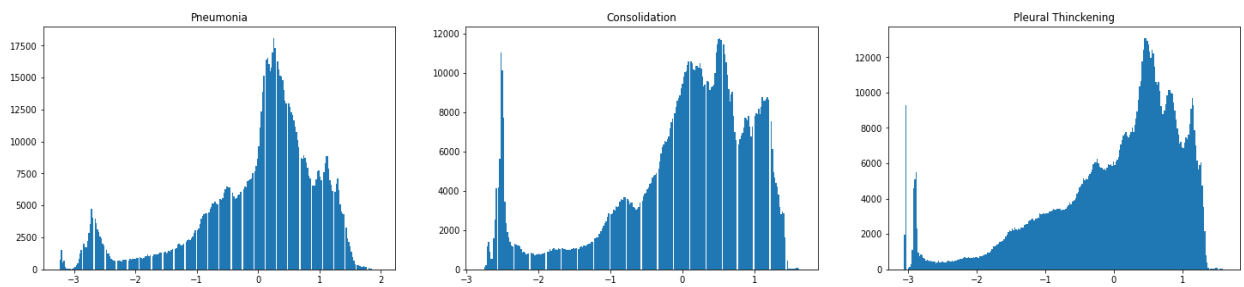


Figura 10. Histograma: Pneumonia, Consolidação e Espessamento Pleural.

3.6.2 Conclusões da Análise Exploratória de Dados

Conclui-se então que:

1. 53% do conjunto de dados não possui nenhuma patologia;
2. 1.27% do conjunto de dados possui pneumonia;
3. A maior parte das patologias estão em pessoas entre 50 e 60 anos;
4. O conjunto de dados possui um pouco mais registros de homens do que mulheres;
5. Nas imagens com patologias relação da projeção da imagem é $AP > PA$;
6. Histograma de intensidades semelhantes entre Pneumonia, Consolidação e Espessamento Pleural.

4. RESULTADOS

4.1. Criando o conjunto de dados de treinamento e teste

O conjunto de dados deste trabalho possui um total de 1.431 imagens positivas para pneumonia.

Para separar o conjunto de dados de treinamento, foram deixados 80% (1.145) das imagens com pneumonia no conjunto de treinamento e 20% (286) das imagens com pneumonia no conjunto de teste.

Conforme a Tabela abaixo, o conjunto de treinamento possui um total de 2.290 imagens, sendo que 50% dessas imagens são positivas para pneumonia e os outros 50%, negativos para pneumonia.

Para o conjunto de teste, há um total de 1.430 imagens, sendo que apenas 20% deste conjunto são de imagens positivas para pneumonia e os outros 80%, representam imagens negativas para pneumonia.

Tabela 4. Proporção do conjunto de dados de treinamento e teste.

Number of instances in train	2290
Number of instances in valid	1430
Proportion of Positive/Negative cases in train dataset	0.5
Proportion of Positive/Negative cases in valid dataset	0.2

4.2 Arquitetura do Algoritmo

Para a construção da arquitetura do algoritmo, foi utilizada a técnica de transferência de aprendizado, utilizando a rede pré-treinada Keras VGG16 treinada no conjunto de dados ImageNet.

As redes pré-treinadas Keras são modelos de aprendizado profundo que são disponibilizadas juntamente com pesos pré-treinados. Esses modelos podem ser usados para previsão, extração de recursos e ajuste fino.⁽⁹⁾

ImageNet é um banco de dados de imagens organizado, em que cada nó da hierarquia é representado por centenas e milhares de imagens. O projeto tem sido fundamental para o avanço da visão computacional e da pesquisa de aprendizado profundo. Os dados estão disponíveis gratuitamente aos pesquisadores para uso não comercial.⁽¹⁰⁾

A transferência de aprendizado é uma técnica utilizada em aprendizado de máquina que se concentra em armazenar o conhecimento adquirido ao resolver um problema e aplicá-lo a um problema diferente, mas relacionado.⁽¹¹⁾ Por exemplo, o conhecimento adquirido ao aprender a reconhecer carros pode ser aplicado ao tentar reconhecer caminhões. Esta área de pesquisa tem alguma relação com a longa história da literatura psicológica sobre transferência de aprendizagem, embora os laços práticos entre os dois campos sejam limitados. Do ponto de vista prático, reutilizar ou transferir informações de tarefas previamente aprendidas para o aprendizado de novas tarefas tem o potencial de melhorar significativamente a eficiência da amostra de um agente de aprendizado por reforço.⁽¹²⁾

Através da transferência de aprendizado, utilizou-se um modelo pré-treinado para reconhecimento de imagens e como base da arquitetura do algoritmo. Dessa forma o algoritmo não precisará ser treinado do zero, pois já foi realizado o treino anteriormente e ele possui um conhecimento em reconhecer imagens pré-adquiridas em um treinamento passado.

Um modelo pré-treinado é uma rede salva que foi previamente treinada em um grande conjunto de dados, normalmente, em uma tarefa de classificação de imagem em grande escala. Usa-se o modelo pré-treinado como ele está ou usa o aprendizado de

transferência para personalizar esse modelo para uma determinada tarefa. A intuição por trás do aprendizado de transferência para classificação de imagens é que se um modelo for treinado em um conjunto de dados grande e geral o suficiente, esse modelo servirá efetivamente como um modelo genérico do mundo visual. Pode-se aproveitar esses mapas de recursos aprendidos sem precisar começar do zero treinando um modelo grande em um grande conjunto de dados.⁽¹³⁾

A rede pré-treinada Keras VGG16 é uma rede do tipo CNN (*Convolutional Neural Network*).

No aprendizado profundo, uma rede neural convolucional (CNN ou ConvNet) é uma classe de rede neural artificial (RNN), mais comumente aplicada para analisar imagens visuais.⁽¹⁴⁾

As redes convolucionais foram inspiradas por processos biológicos em que o padrão de conectividade entre os neurônios assemelha-se à organização do córtex visual animal.⁽¹⁶⁾ Neurônios corticais individuais respondem a estímulos apenas em uma região restrita do campo visual conhecida como campo receptivo. Os campos receptivos de diferentes neurônios se sobrepõem parcialmente de modo que cobrem todo o campo visual.⁽¹⁵⁾

Como base da arquitetura do algoritmo, foram utilizadas 19 camadas da rede VGG16. Abaixo é possível verificar esta estrutura.

Tabela 5. Arquitetura com 19 camadas da rede VGG16.

Number of layers VGG16:19.		
Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
Total Params	14,714,688	
Trainable Params	14,714,688	
Non-trainable Params	0	

Com o objetivo de realizar um ajuste fino e melhorar o desempenho do algoritmo, além da rede VGG16 com 19 camadas, foram adicionadas mais oito camadas

à arquitetura da rede com o objetivo de reconhecer padrões específicos de pneumonia em imagens de Raio-x do tórax.

Abaixo se podem ver as novas camadas adicionadas à arquitetura do modelo.

Tabela 6. Arquitetura com oito camadas adicionadas a rede VGG16

Layer (type)	Output Shape	Param #
model_1 (Model)	(None, 7, 7, 512)	14714688
flatten_1 (Flatten)	(None, 25088)	0
dense_1 (Dense)	(None, 1024)	25691136
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 1024)	1049600
dropout_2 (Dropout)	(None, 1024)	0
dense_3 (Dense)	(None, 512)	524800
dropout_3 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 1)	0
Total Params	41,980,737	
Trainable Params	34,345,473	
Non-trainable Params	7,635,264	

A arquitetura final do modelo ficou com um total de 27 camadas, sendo 19 camadas da rede pré-treinada VGG16 e mais oito camadas adicionadas como ajuste fino para melhorar a performance do modelo.

A primeira camada do modelo receberá uma imagem com dimensões (224, 224, 3), sendo 224 pixels de largura, 224 pixels de comprimento e três dimensões, representando os canais RGB (*Red, Blue, Green*) da imagem.

A saída do modelo, ou seja, a última camada será um valor decimal entre 0 e 1, representando a probabilidade de a imagem analisada possuir um padrão positivo para pneumonia.

4.2 Pré-Processamento e Aumento de Dados

O aumento de dados na análise de dados são técnicas usadas para aumentar a quantidade de dados, adicionando cópias ligeiramente modificadas de dados já existentes ou dados sintéticos recém-criados a partir de dados existentes. Ele atua como um regulador e ajuda a reduzir o *overfitting* ao treinar um modelo de aprendizado de máquina. Este tipo de técnica é utilizado principalmente quando o conjunto de dados é pequeno e precisa de mais instâncias para treinar e validar o modelo de inteligência artificial.^(22-23,26)

4.3 Imagens sintéticas para aumento de dados

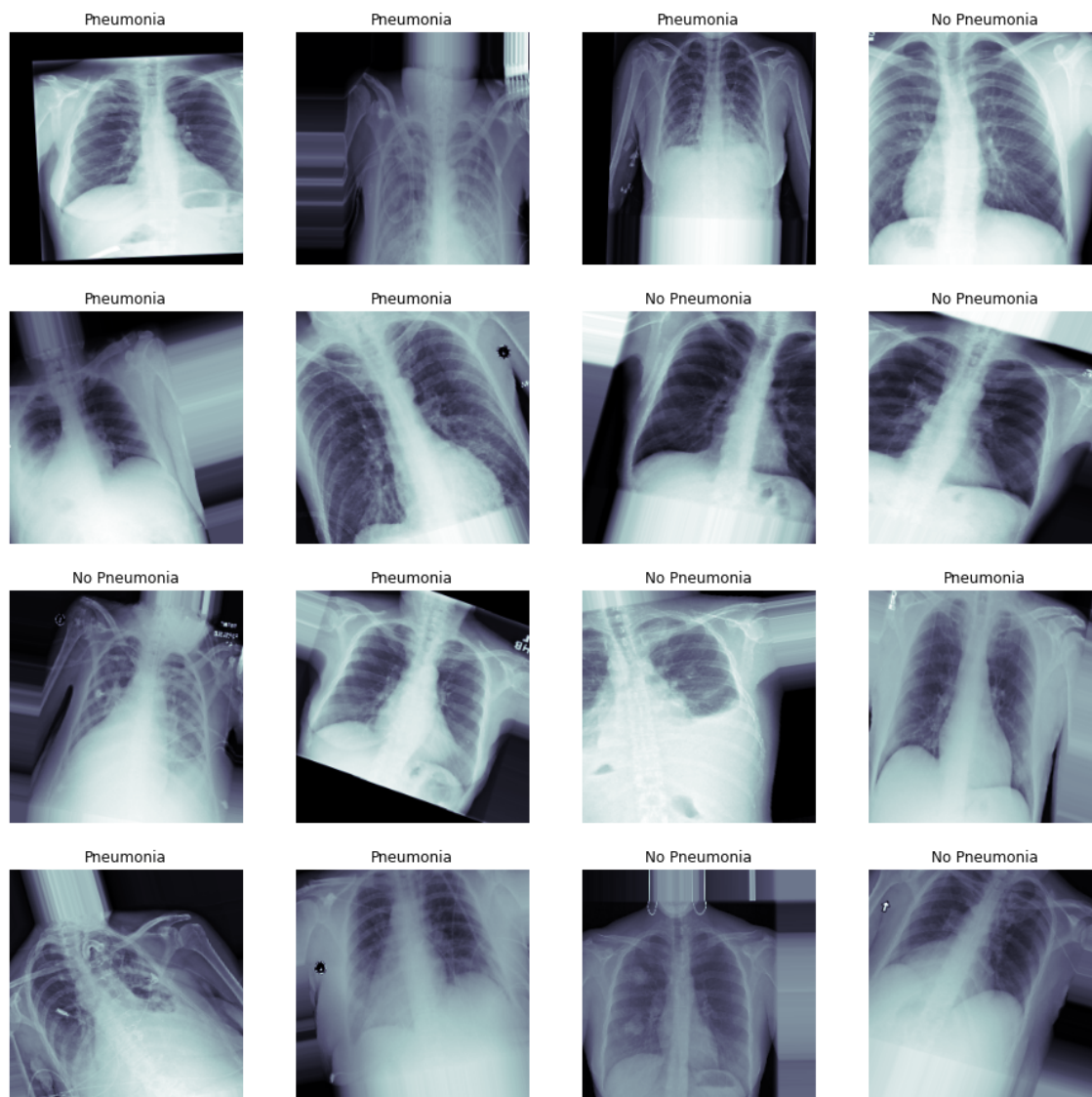
Se um conjunto de dados for muito pequeno, uma versão do conjunto de dados aumentada com rotação e espelhamento ainda pode não ser suficiente para um determinado problema. Outra solução é o fornecimento de imagens sintéticas inteiramente novas por meio de várias técnicas, por exemplo, o uso de redes adversárias generativas para criar novas imagens sintéticas para aumento de dados.⁽²⁶⁻²⁸⁾ Além disso, os algoritmos de reconhecimento de imagem mostram melhorias ao transferir de imagens renderizadas em ambientes virtuais para dados do mundo real.

Um ponto importante a ser considerado quando se fala no aumento de dados para imagens médicas, é compreender que realizar o espelhamento da imagem não é algo indicado, pois se considerar a anatomia do pulmão é percebido que existe diferença entre o lado direito e o lado esquerdo, por isso deve se ter cautela ao aplicar técnicas de aumento de dados em imagens médicas.

Para realizar o aumento de dados no nosso conjunto de dados, foram aplicadas as técnicas de rotação horizontal da imagem, zoom de 20% na imagem, aumento horizontal de 20%, aumento vertical de 20% e rotação de 20% para direita e esquerda.

Abaixo alguns exemplos das imagens alteradas através das técnicas de aumento de dados. (Figura 11).

Figura 11. Imagens, após a aplicação de técnicas de aumento de dados.



4.4 Treinamento

Após ter separado os conjuntos de dados de treinamento e teste, e montado a arquitetura do algoritmo, pode-se prosseguir para a etapa de treinamento do modelo.

Na etapa de treinamento, o algoritmo recebe exemplos de entrada de dados e a sua respectiva saída desejada. Dessa maneira, a cada interação com os dados de entrada, o modelo vai ajustando os pesos de cada nó presente em cada camada da arquitetura para identificar os padrões nos dados de entrada que justifiquem a saída desejada.^(1,32-33) Com essa abordagem, entende-se que o modelo aprende com os dados que lhe são apresentados, relacionando os dados de entrada com a sua saída esperada.

Fazendo dessa maneira, confirma-se a definição de aprendizado de máquina feita por Arthur Samuel, em 1959, “Campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados”.⁽³⁰⁾

A abordagem utilizada neste projeto é conhecida como aprendizado supervisionado, no qual algoritmo recebe os dados de entrada e também a sua respectiva saída esperada para aprender a identificar os padrões nos dados apresentados a cada rodada de treinamento.⁽³⁴⁻³⁵⁾

Sendo específico para este trabalho, o modelo irá receber como dado de entrada uma imagem de Raio-x do tórax e também o rótulo dessa imagem, que é a sua saída esperada, podendo ser 0 ou 1. Sendo 0 para não pneumonia e 1 para pneumonia.

O cenário esperado é que com o treinamento o modelo possa aprender os padrões presentes na imagem que resultam em uma saída do tipo 1, ou seja, positivo para pneumonia. Após o treinamento o modelo deverá ser capaz de receber uma entrada, ou seja, uma imagem de Raio-x do tórax e inferir se aquela imagem, de acordo com os padrões aprendidos no treinamento, é positiva ou negativa para pneumonia.

4.5 Parâmetros para treinamento

Das 27 camadas presentes na arquitetura do modelo, as 15 primeiras, que fazem parte do modelo pré-treinado VGG16, foram congeladas e somente as 12 últimas camadas foram treinadas, como um ajuste fino para encontrar os padrões nas imagens de Raio-x do tórax.

A taxa de aprendizado definido para ajuste do modelo após cada rodada de interação com os dados de treinamento foi definida em Learning Rate=0.0001

A métrica utilizada para monitoramento e avaliação de desempenho do treinamento foi binary_accuracy.

Foram definidas 40 rodadas de treinamento no conjunto de dados de teste, ou seja, o conjunto de dados será apresentado como dado de entrada para o modelo 40 vezes, com o objetivo de que a cada rodada o algoritmo aprenda um padrão novo para ajudar a diferenciar imagens com e sem pneumonia.

O algoritmo salva os padrões aprendidos e a performance obtida a cada rodada de treinamento. O treinamento é finalizado após completar as 40 rodadas ou se não houver nenhuma melhoria na métrica de desempenho, após 15 rodadas de treinamento da melhor performance registrada.

Abaixo, pode-se observar a evolução do aprendizado durante a fase de treinamento através da métrica binary_accuracy. (Figura 12).

Figura 12. Evolução da acurácia na fase de treinamento. (Continua).

Epoch 1/40
 72/72 [=====] - 73s 1s/step - loss: 0.6227 -
 binary_accuracy: 0.6437 - val_loss: 0.6031 - val_binary_accuracy: 0.6713
 Epoch 00001: val_loss improved from inf to 0.60314, saving model to
 xray_class_model.best.hdf5
 Epoch 2/40
 72/72 [=====] - 64s 889ms/step - loss: 0.5951 -
 binary_accuracy: 0.6926 - val_loss: 0.6454 - val_binary_accuracy: 0.6469
 Epoch 00002: val_loss did not improve from 0.60314
 Epoch 3/40
 72/72 [=====] - 64s 891ms/step - loss: 0.5867 -
 binary_accuracy: 0.6930 - val_loss: 0.6116 - val_binary_accuracy: 0.6692
 Epoch 00003: val_loss did not improve from 0.60314
 Epoch 4/40
 72/72 [=====] - 64s 896ms/step - loss: 0.5897 -
 binary_accuracy: 0.6869 - val_loss: 0.7529 - val_binary_accuracy: 0.5049
 Epoch 00004: val_loss did not improve from 0.60314
 Epoch 5/40
 72/72 [=====] - 64s 892ms/step - loss: 0.5723 -
 binary_accuracy: 0.7105 - val_loss: 0.6777 - val_binary_accuracy: 0.5965
 Epoch 00005: val_loss did not improve from 0.60314
 Epoch 6/40
 72/72 [=====] - 64s 887ms/step - loss: 0.5598 -
 binary_accuracy: 0.7166 - val_loss: 0.6280 - val_binary_accuracy: 0.6678
 Epoch 00006: val_loss did not improve from 0.60314
 Epoch 7/40
 72/72 [=====] - 64s 888ms/step - loss: 0.5656 -
 binary_accuracy: 0.7044 - val_loss: 0.5727 - val_binary_accuracy: 0.6832
 Epoch 00007: val_loss improved from 0.60314 to 0.57271, saving model to
 xray_class_model.best.hdf5
 Epoch 8/40
 72/72 [=====] - 64s 886ms/step - loss: 0.5698 -
 binary_accuracy: 0.7026 - val_loss: 0.6842 - val_binary_accuracy: 0.6063
 Epoch 00008: val_loss did not improve from 0.57271
 Epoch 9/40
 72/72 [=====] - 64s 888ms/step - loss: 0.5650 -
 binary_accuracy: 0.7140 - val_loss: 0.6027 - val_binary_accuracy: 0.6629
 Epoch 00009: val_loss did not improve from 0.57271

Figura 12. Evolução da acurácia na fase de treinamento. (Continua).

Epoch 10/40

72/72 [=====] - 64s 885ms/step - loss: 0.5535 -

binary_accuracy: 0.7092 - val_loss: 0.8019 - val_binary_accuracy: 0.5895

Epoch 00010: val_loss did not improve from 0.57271

Epoch 11/40

72/72 [=====] - 64s 888ms/step - loss: 0.5453 -

binary_accuracy: 0.7205 - val_loss: 0.7669 - val_binary_accuracy: 0.5839

Epoch 00011: val_loss did not improve from 0.57271

Epoch 12/40

72/72 [=====] - 65s 898ms/step - loss: 0.5309 -

binary_accuracy: 0.7384 - val_loss: 0.5907 - val_binary_accuracy: 0.7070

Epoch 00012: val_loss did not improve from 0.57271

Epoch 13/40

72/72 [=====] - 64s 887ms/step - loss: 0.5313 -

binary_accuracy: 0.7314 - val_loss: 0.5771 - val_binary_accuracy: 0.7196

Epoch 00013: val_loss did not improve from 0.57271

Epoch 00013: ReduceLROnPlateau reducing learning rate to

9.999999747378752e-06.

Epoch 14/40

72/72 [=====] - 64s 892ms/step - loss: 0.5028 -

binary_accuracy: 0.7537 - val_loss: 0.6940 - val_binary_accuracy: 0.6699

Epoch 00014: val_loss did not improve from 0.57271

Epoch 15/40

72/72 [=====] - 64s 887ms/step - loss: 0.4838 -

binary_accuracy: 0.7712 - val_loss: 0.7682 - val_binary_accuracy: 0.6364

Epoch 00015: val_loss did not improve from 0.57271

Epoch 16/40

72/72 [=====] - 64s 892ms/step - loss: 0.4815 -

binary_accuracy: 0.7725 - val_loss: 0.7359 - val_binary_accuracy: 0.6566

Epoch 00016: val_loss did not improve from 0.57271

Epoch 17/40

72/72 [=====] - 64s 892ms/step - loss: 0.4661 -

binary_accuracy: 0.7786 - val_loss: 0.8339 - val_binary_accuracy: 0.6084

Epoch 00017: val_loss did not improve from 0.57271

Epoch 18/40

72/72 [=====] - 63s 878ms/step - loss: 0.4678 -

binary_accuracy: 0.7747 - val_loss: 0.7277 - val_binary_accuracy: 0.6657

Epoch 00018: val_loss did not improve from 0.57271

Figura 12. Evolução da acurácia na fase de treinamento. (Continua)

Epoch 19/40

72/72 [=====] - 63s 876ms/step - loss: 0.4629 -

binary_accuracy: 0.7755 - val_loss: 0.7404 - val_binary_accuracy: 0.6524

Epoch 00019: val_loss did not improve from 0.57271

Epoch 00019: ReduceLROnPlateau reducing learning rate to
9.999999747378752e-07.

Epoch 20/40

72/72 [=====] - 64s 884ms/step - loss: 0.4432 -

binary_accuracy: 0.7860 - val_loss: 0.7510 - val_binary_accuracy: 0.6517

Epoch 00020: val_loss did not improve from 0.57271

Epoch 21/40

72/72 [=====] - 64s 887ms/step - loss: 0.4551 -

binary_accuracy: 0.7878 - val_loss: 0.7630 - val_binary_accuracy: 0.6441

Epoch 00021: val_loss did not improve from 0.57271

Epoch 22/40

72/72 [=====] - 64s 888ms/step - loss: 0.4534 -

binary_accuracy: 0.7769 - val_loss: 0.7655 - val_binary_accuracy: 0.6385

Epoch 00022: val_loss did not improve from 0.57271

Epoch 23/40

72/72 [=====] - 64s 882ms/step - loss: 0.4490 -

binary_accuracy: 0.7830 - val_loss: 0.7813 - val_binary_accuracy: 0.6315

Epoch 00023: val_loss did not improve from 0.57271

Epoch 24/40

72/72 [=====] - 63s 882ms/step - loss: 0.4490 -

binary_accuracy: 0.7882 - val_loss: 0.7712 - val_binary_accuracy: 0.6385

Epoch 00024: val_loss did not improve from 0.57271

Epoch 25/40

72/72 [=====] - 68s 949ms/step - loss: 0.4468 -

binary_accuracy: 0.7843 - val_loss: 0.7597 - val_binary_accuracy: 0.6420

Epoch 00025: val_loss did not improve from 0.57271

Epoch 00025: ReduceLROnPlateau reducing learning rate to
9.999999974752428e-08.

Epoch 26/40

72/72 [=====] - 64s 882ms/step - loss: 0.4437 -

binary_accuracy: 0.7865 - val_loss: 0.7613 - val_binary_accuracy: 0.6406

Epoch 00026: val_loss did not improve from 0.57271

Figura 12. Evolução da acurácia na fase de treinamento. (Continua)

Epoch 27/40

72/72 [=====] - 64s 886ms/step - loss: 0.4514 -

binary_accuracy: 0.7812 - val_loss: 0.7622 - val_binary_accuracy: 0.6392

Epoch 00027: val_loss did not improve from 0.57271

Epoch 28/40

72/72 [=====] - 64s 890ms/step - loss: 0.4402 -

binary_accuracy: 0.7930 - val_loss: 0.7613 - val_binary_accuracy: 0.6406

Epoch 00028: val_loss did not improve from 0.57271

Epoch 29/40

72/72 [=====] - 64s 888ms/step - loss: 0.4458 -

binary_accuracy: 0.7900 - val_loss: 0.7625 - val_binary_accuracy: 0.6399

Epoch 00029: val_loss did not improve from 0.57271

Epoch 30/40

72/72 [=====] - 64s 884ms/step - loss: 0.4460 -

binary_accuracy: 0.7795 - val_loss: 0.7638 - val_binary_accuracy: 0.6385

Epoch 00030: val_loss did not improve from 0.57271

Epoch 31/40

72/72 [=====] - 64s 890ms/step - loss: 0.4525 -

binary_accuracy: 0.7843 - val_loss: 0.7663 - val_binary_accuracy: 0.6378

Epoch 00031: val_loss did not improve from 0.57271

Epoch 00031: ReduceLROnPlateau reducing learning rate to

1.0000000116860975e-08.

Epoch 32/40

72/72 [=====] - 68s 940ms/step - loss: 0.4367 -

binary_accuracy: 0.7882 - val_loss: 0.7663 - val_binary_accuracy: 0.6378

Epoch 00032: val_loss did not improve from 0.57271

Epoch 33/40

72/72 [=====] - 64s 883ms/step - loss: 0.4465 -

binary_accuracy: 0.7860 - val_loss: 0.7663 - val_binary_accuracy: 0.6378

Epoch 00033: val_loss did not improve from 0.57271

Epoch 34/40

72/72 [=====] - 63s 881ms/step - loss: 0.4460 -

binary_accuracy: 0.7934 - val_loss: 0.7663 - val_binary_accuracy: 0.6378

Epoch 00034: val_loss did not improve from 0.57271

Figura 12. Evolução da acurácia na fase de treinamento. (Conclusão)

Epoch 35/40

72/72 [=====] - 63s 881ms/step - loss: 0.4564 -

binary_accuracy: 0.7830 - val_loss: 0.7662 - val_binary_accuracy: 0.6378

Epoch 00035: val_loss did not improve from 0.57271

Epoch 36/40

72/72 [=====] - 64s 887ms/step - loss: 0.4528 -

binary_accuracy: 0.7882 - val_loss: 0.7660 - val_binary_accuracy: 0.6385

Epoch 00036: val_loss did not improve from 0.57271

Epoch 37/40

72/72 [=====] - 65s 902ms/step - loss: 0.4425 -

binary_accuracy: 0.7926 - val_loss: 0.7661 - val_binary_accuracy: 0.6385

Epoch 00037: val_loss did not improve from 0.57271

4.6 Avaliação do Desempenho

Após treinar por algum tempo, observou-se o desempenho do algoritmo traçando algumas estatísticas de desempenho.

Inicialmente, foi analisada a evolução de aprendizado do algoritmo durante o treinamento. Para a fase de treinamento, o conjunto de dados foi separado em duas partes, um conjunto de dados para treinamento e outro conjunto de dados para testes ou também chamado de conjunto de validação.^(1,40)

Durante o treinamento, foram monitorados o erro e a acurácia do modelo, ou seja, a quantidade de erros e acertos no momento de rotular uma imagem como pneumonia ou não pneumonia. Esse monitoramento é feito nos dois conjuntos de dados, teste e validação.⁽⁴¹⁻⁴²⁾

Ao observar o gráfico abaixo, há quatro itens monitorados, `train_loss` representando o erro do conjunto de treinamento e o `train_acc` representando a acurácia do conjunto de treinamento, da mesma maneira, há o item `val_loss` representando o erro do conjunto de validação e o `val_acc` representando a acurácia do conjunto de validação.

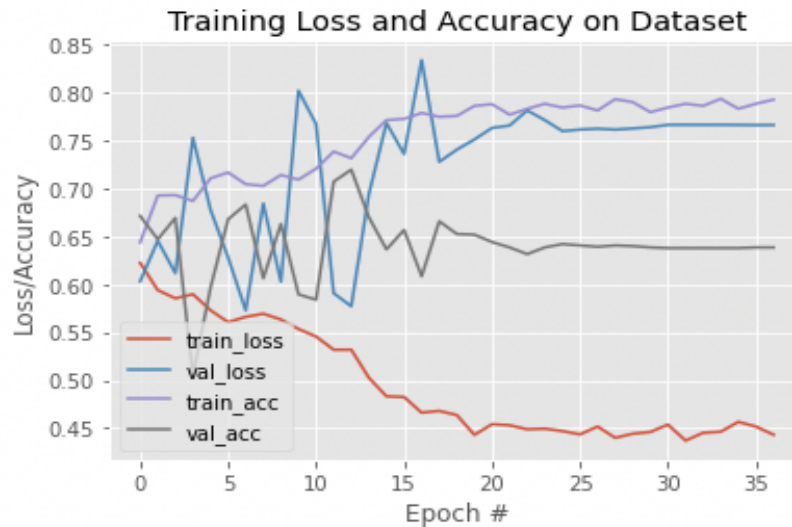
Analisando o gráfico é possível notar que o erro diminui e a acurácia aumenta, conforme as rodadas de treinamento vão sendo executadas.

Uma observação interessante a ser feita é que a partir da rodada de treinamento 25, aproximadamente, o algoritmo estabiliza-se e não oscila mais na taxa de erro e acurácia do conjunto de validação. A partir do momento que o erro e a acurácia do conjunto de validação estabilizam-se, deve-se tomar o cuidado para não continuar o treinamento por muito tempo e evitar o *overfitting*. Na modelagem matemática, entende-se *overfitting* como:

[...] produção de uma análise que corresponde muito de perto ou exatamente a um determinado conjunto de dados e, portanto, pode não se ajustar a dados adicionais ou prever observações futuras de forma confiável.⁽³⁶⁻³⁷⁾

Em outras palavras o *overfitting* é um excesso de treinamento, quando isso acontece, o algoritmo não aprendeu a identificar padrões nos dados, mas ele decorou os dados do conjunto de treinamento. Sendo assim, o desempenho do algoritmo é excelente nos dados de treinamento, porém quando lhe é apresentado um dado fora do conjunto de dados de treinamento, o algoritmo tem um baixo desempenho, pois ele não aprendeu a identificar padrões, ele apenas decorou o que lhe foi apresentado durante a fase de treinamento, sabendo reconhecer somente esses dados de teste.

Figura 13. Gráfico de treinamento e validação.



A primeira estatística a ser observada é a curva ROC.

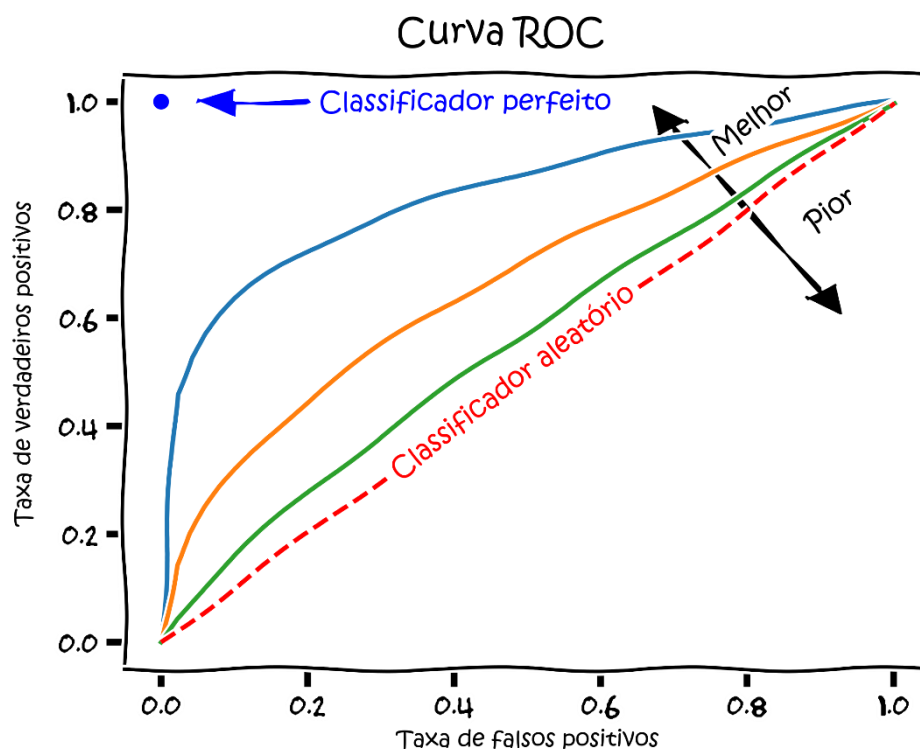
A Curva Característica de Operação do Receptor (Curva COR), ou, do inglês, *Receiver Operating Characteristic Curve* (ROC curve), ou, simplesmente, curva ROC, é uma representação gráfica que ilustra o desempenho (ou performance) de um sistema classificador binário.⁽⁴³⁾

A curva ROC é obtida pela representação da razão **RPV** = Positivos Verdadeiros / Positivos Totais *versus* a razão **RPF** = Positivos Falsos / Negativos Totais.

O RPV é também conhecido como sensibilidade (ou taxa de verdadeiros positivos), e RPF = 1-especificidade ou taxa de falsos positivos.

A análise ROC fornece ferramentas para selecionar modelos possivelmente ideais (modelos ótimos) e descartar modelos não tão ótimos, independentemente (e antes de especificar) o contexto de custos ou a distribuição de classe. A análise ROC está relacionada de forma direta e natural com a análise de custo/benefício do diagnóstico.^(32,44)

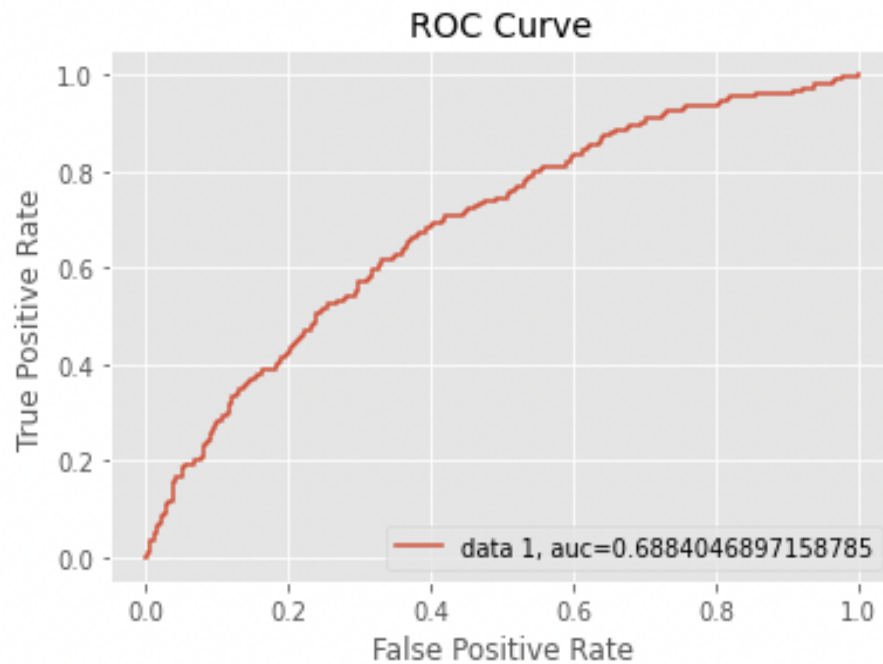
Figura 14. Curva ROC



ROC Curve. In: Wikimedia Commons: the free media repository [Internet]. [acesso em 25 abr 2021]. Disponível em: <https://commons.wikimedia.org/wiki/File:Roc-draft-xkcd-style.svg#filelinks>.

Abaixo o resultado da curva ROC, mostrando o desempenho do modelo no conjunto de teste, após o treinamento. Nota-se que o algoritmo alcançou uma área sobre a curva de 0.68. (Figura 15).

Figura 15. Curva ROC – Validação



Outro dado estatístico que foi utilizado para avaliar o desempenho do modelo foi a relação entre Precision/Recall.

No reconhecimento de padrões e classificação de informações, precisão e recuperação são métricas de desempenho que se aplicam a dados analisados de um conjunto de dados.^(1,32,43)

Em uma tarefa de classificação, a precisão para uma classe é o número de verdadeiros positivos (ou seja, o número de itens corretamente rotulados como pertencentes à classe positiva) dividido pelo número total de elementos rotulados como pertencentes à classe positiva (ou seja, a soma de verdadeiros positivos e falsos positivos, que são itens rotulados incorretamente como pertencentes à classe).^(32,43)

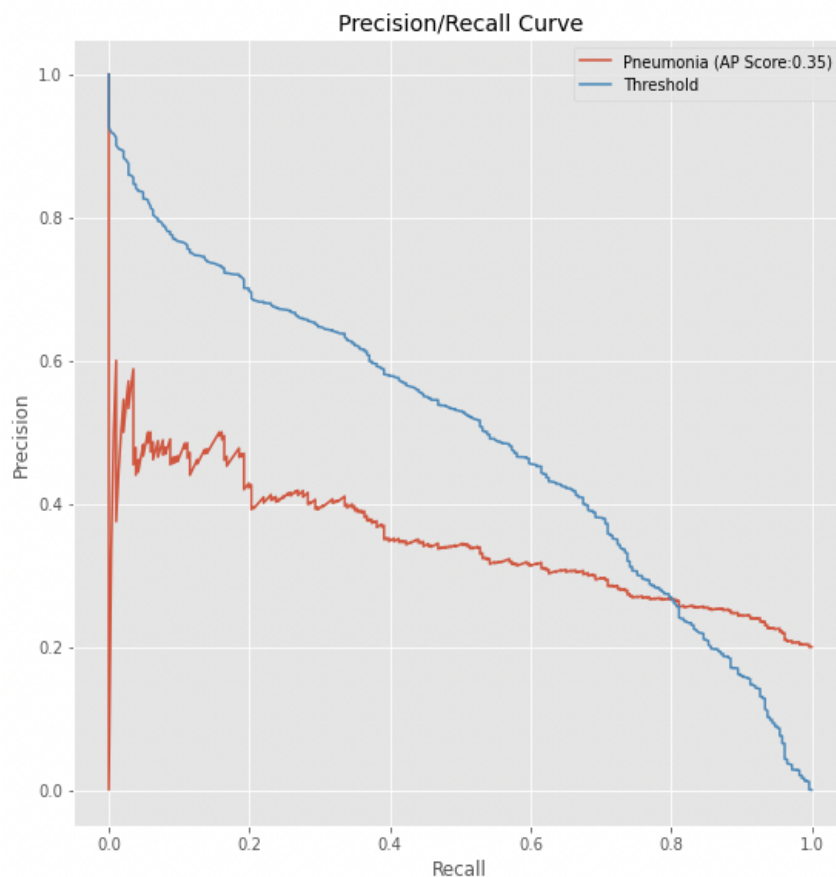
O recall neste contexto é definido como o número de verdadeiros positivos dividido pelo número total de elementos que realmente pertencem à classe positiva (ou

seja, a soma de verdadeiros positivos e falsos negativos, que são itens que não foram rotulados como pertencentes à classe positiva, mas deveriam ter sido).^(32,43)

No Gráfico abaixo, comparou-se o resultado do Precision e Recall relacionado a um limiar aplicado à resposta enviada pelo modelo.

O algoritmo retorna para cada imagem analisada, uma probabilidade desta imagem possuir pneumonia. Essa probabilidade retornada pelo algoritmo está entre 0 e 1. Conforme se aplicou um limiar para considerar se a imagem analisada possui ou não pneumonia, afetou-se diretamente o Precision e o Recall do algoritmo. Conforme observado no Gráfico abaixo, se considerar que a imagem possui pneumonia somente se a probabilidade for 0.6 ou superior, então o Precision será 0.4 e o Recall será 0.3.

Figura 16. Gráfico Precision vs Recall.



Uma questão observada durante a análise do desempenho do algoritmo, foi sobre qual seria um bom limiar para rotular uma imagem com pneumonia ou sem pneumonia, considerando que o algoritmo retorna uma probabilidade entre 0 e 1. Para chegar a uma conclusão do limiar ideal, considerando que existe um desbalanceamento no conjunto de dados de imagem com e sem pneumonia; foi realizada a relação entre a métrica F1-score e o limiar, de modo que foi encontrado para cada limiar o seu respectivo resultado na métrica F1-score.

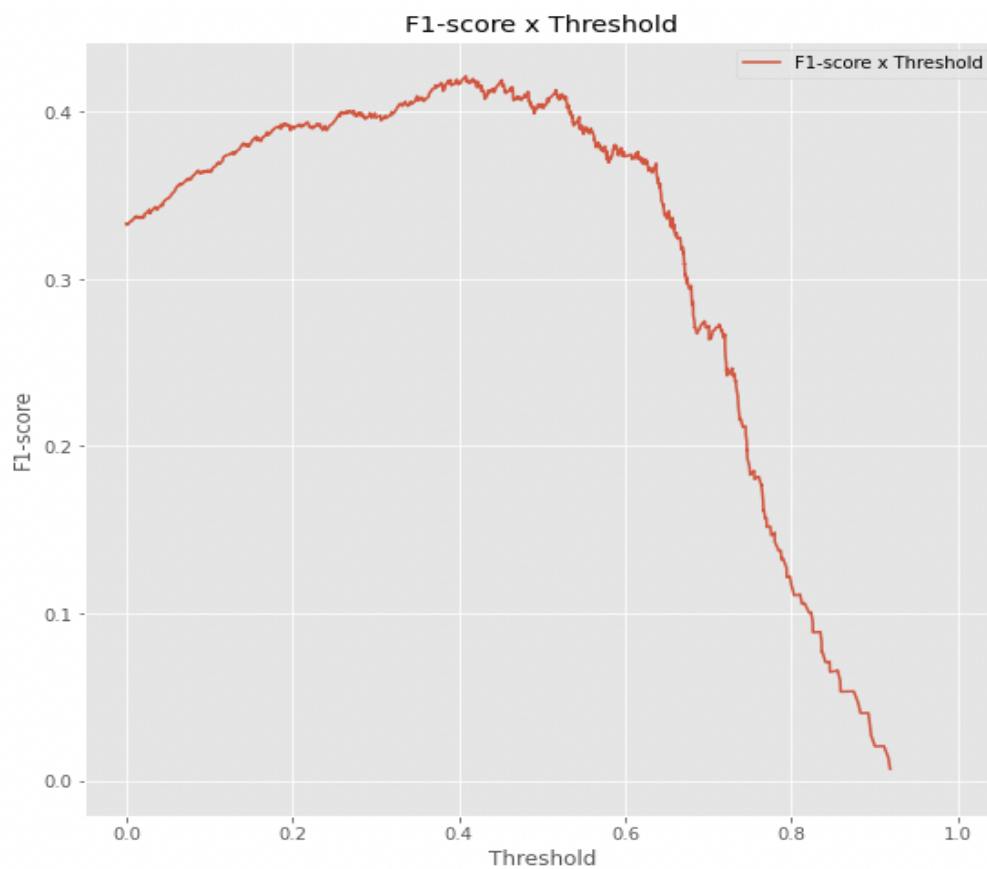
Na análise estatística de classificação binária, o F-score ou F-measure é uma medida da acurácia de um teste. É calculado a partir do Precision e do Recall.^(43,45,47)

O F1-score é a média harmônica da precisão e do recall. Aplicando pesos adicionais, valoriza um Precision ou Recall mais do que o outro.^(43,45,47)

O valor mais alto possível de um F-score é 1.0, indicando Precision e Recall perfeitos, e o valor mais baixo possível é 0 se o Precision ou o Recall for zero.^(43,45,47)

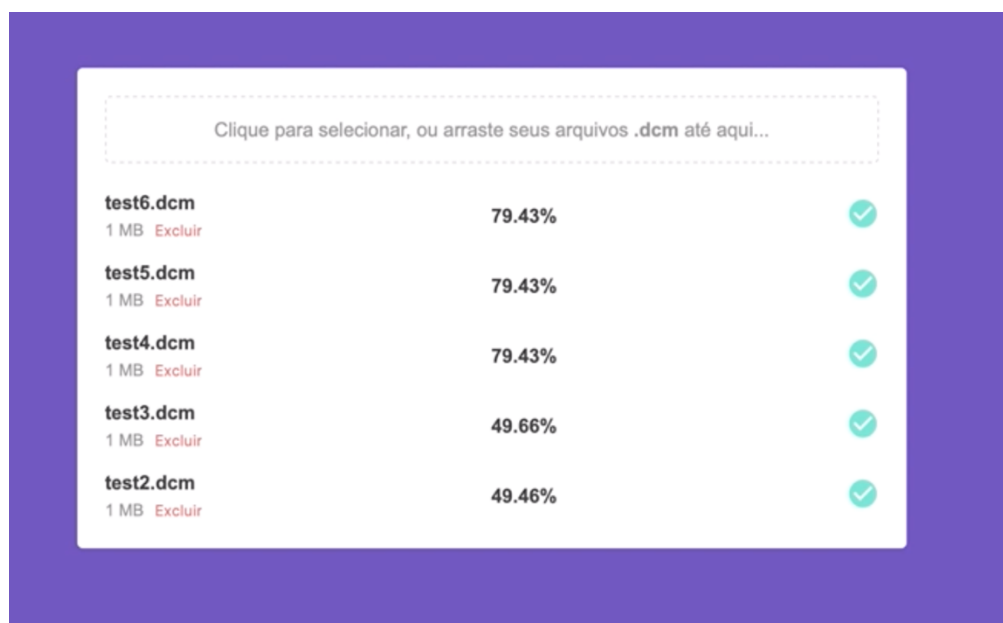
Observando o Gráfico abaixo podemos considerar o limiar = 0.5 como um bom ponto de corte para rotular resultados com pneumonia x sem pneumonia. (Figura 17)

Figura 17. Gráfico F1-score vs Threshold.



4.7 Simulação do Sistema em Uso.

Figura 18. Sistema web



Com o objetivo de liberar o algoritmo para utilização e testes de maneira simples para qualquer pessoa que tenha interesse, o algoritmo foi disponibilizado em uma página web. Dessa maneira, basta abrir um navegador de Internet e digitar o endereço do sistema para se ter acesso ao algoritmo e testá-lo em execução.

O link da página web na qual o algoritmo está disponível é:
<http://5.189.138.239:5006/>

Para testar o algoritmo em execução, basta seguir os passos abaixo:

1. Acessar o sistema;

Para acessar o sistema, basta abrir um navegador de internet no computador e digitar o seguinte endereço: <http://5.189.138.239:5006/>

2. Upload das imagens no sistema;

Com a página do sistema aberta, o usuário precisará arrastar os arquivos de Raio-x para dentro da página para realizar o upload dos arquivos para o sistema.

O sistema aceita somente arquivos no formato DICOM (.dcm) e o usuário pode realizar o upload de mais de um arquivo ao mesmo tempo.

3. Algoritmo analisa as imagens;

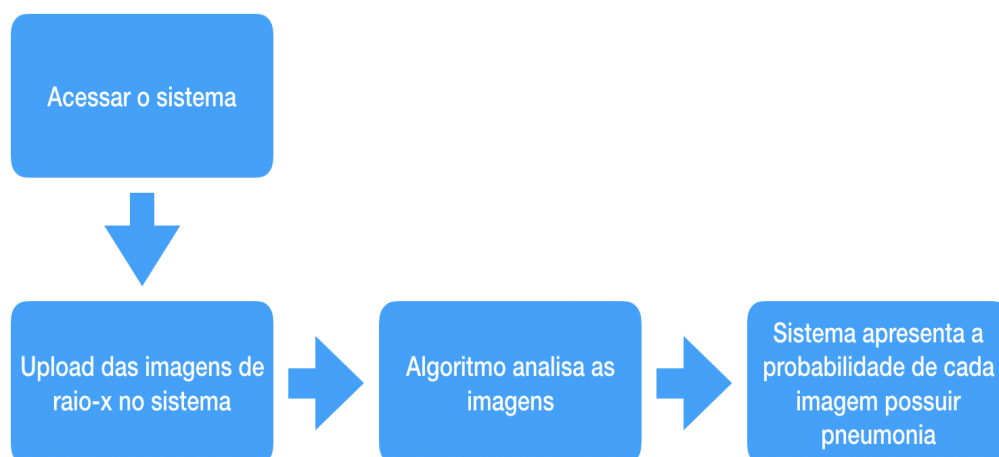
Após realizar o upload das imagens, o sistema começa a realizar a análise. Antes de começar a analisar a imagem para buscar padrões de pneumonia, o sistema verifica se os arquivos enviados são imagens de Raio-x e se o Raio-x é do tórax, caso contrário o arquivo é desconsiderado e não analisado.

Confirmando que o arquivo DICOM representa uma imagem de Raio-x do tórax, só então o algoritmo entra em execução e começa a inspecionar a imagem, procurando por um padrão de pneumonia.

4. Sistema apresenta o resultado da análise;

Por fim, o sistema apresenta na tela da página web a probabilidade de cada arquivo analisado conter pneumonia.

Figura 19. Fluxo de utilização do sistema.



5. DISCUSSÃO

O estudo desenvolvido resultou em um sistema web, sendo uma importante ferramenta para auxiliar a priorização do fluxo de trabalho do médico radiologista, bem como, auxiliá-lo no diagnóstico do exame. O sistema desenvolvido neste projeto mostra ao médico quais exames apresentam alguma anomalia e desta forma o médico poderá priorizar exames que possivelmente possuam uma patologia, definindo seu fluxo de trabalho por prioridade de risco e não mais por ordem de chegada.

O algoritmo de inteligência artificial disponibilizado em um sistema web tem a finalidade de facilitar e agilizar o fluxo de trabalho, tendo em vista que não há necessidade de nenhuma infraestrutura robusta para funcionamento, nem mesmo a aquisição de equipamentos. Basta um computador com acesso à Internet, o que já existe na maioria dos hospitais e clínicas, e o sistema desenvolvido neste projeto já poderá ser utilizado pelo médico.

A possibilidade de o sistema receber e manipular arquivos no formato DICOM é um ponto positivo, pois dessa maneira o sistema está seguindo um padrão de comunicação para imagens médicas já definido e amplamente utilizado pelo mercado.⁽⁵¹⁾ Como os equipamentos de Raio-x já geram as imagens no formato DICOM, não é necessário nenhum trabalho adicional de manipulação do arquivo de exame para transferi-lo do equipamento de Raio-x ao sistema web. Sendo assim, a inclusão do sistema web no fluxo de trabalho não gera nenhum esforço ou complexidade ou adicional para o médico.

O algoritmo de inteligência artificial para diagnóstico de pneumonia em imagens de Raio-x do tórax; objeto deste estudo, engloba em seu desenvolvimento, design da arquitetura e construção as principais técnicas de inteligência artificial aplicadas para identificação e classificação de padrões em imagens médicas.^(1,32) Um ponto diferencial

que procurou ser desenvolvido neste estudo foi que o algoritmo ficasse disponível para a utilização dos interessados de forma simples, pronto para uso, sem a necessidade de implantações complexas envolvendo equipamentos e software, para somente depois ser possível a utilizar o algoritmo e tirar proveito dos seus benefícios. O desejo inicial era que o resultado deste projeto fosse algo prático, que gerasse uma melhoria e ganho real para a saúde e os envolvidos no processo, sejam eles pacientes, profissionais ou instituições.

Acredita-se que o resultado deste projeto alcançou o objetivo inicial. O desenvolvimento de um algoritmo inteligente capaz de detectar pneumonia em imagens de Raio-x do tórax que esteja disponível de forma simples, rápida e prática para qualquer interessado em utilizar e se beneficiar deste algoritmo.

Analisando projetos semelhantes, notou-se diferenças na arquitetura do algoritmo, bem como, nos parâmetros utilizados para o seu treinamento, resultando por consequência em diferentes desempenhos. Uma técnica utilizada em outro projeto que gerou resultado com performance elevada foi a utilização *ensembles*. Essa técnica, conhecida como *ensemble*, utiliza ao invés de um único algoritmo para realizar a análise da imagem e detectar um possível padrão de patologia e gerar o resultado, une vários algoritmos diferentes que analisam a imagem em busca de um padrão de patologia e cada algoritmo gera o seu resultado da análise da imagem e, após isso, é feita uma combinação dos diagnósticos de todos os algoritmos para chegar a uma conclusão final sobre o diagnóstico da imagem analisada. Fazendo uma analogia, é como se o resultado do diagnóstico fosse realizado por um conselho de especialistas, no qual todos analisaram a imagem, apresentaram suas justificativas e por fim chegaram a uma conclusão única sobre o diagnóstico, levando em consideração e ponderando a

contribuição de cada especialista presente no conselho. Utilizando essa técnica de ensemble foi possível alcançar um desempenho na casa dos 0.8 de AUC.⁽⁴⁸⁻⁵⁰⁾

Em outro estudo foram utilizadas imagens de Tomografia Computadorizada ao invés de imagens de Raio-x. A troca do tipo de equipamento utilizado para a captura da imagem gerou um desempenho na casa dos 0.9 de AUC.⁽⁵²⁾ Este aumento se deve, em grande parte, a adição de uma nova dimensão na imagem, o que traz consigo um volume de informação muito maior para ser analisado. Imagens de Raio-x são imagens 2D, possuindo altura e largura, já imagens de tomografia computadorizada são imagens 3D, possuindo altura, largura e profundidade. Deve ser levado em consideração que a troca de uma imagem de Raio-x para uma imagem de tomografia computadorizada gera um aumento na assertividade do diagnóstico, porém em contrapartida os custos para isso também aumentam. Um equipamento de tomografia computadorizada é mais caro que um equipamento de Raio-x, o poder computacional necessário para analisar imagens 3D é muito maior do que o necessário para analisar imagens 2D e isso refletido no treinamento do algoritmo irá aumentar o tempo necessário para treinamento. Portanto, são vários fatores envolvidos que devem ser ponderados no momento de escolher qual tipo de imagem será utilizada no diagnóstico pelo sistema de inteligência artificial.

Alguns estudos focaram seus esforços em identificar os efeitos da tecnologia no diagnóstico por imagem, além do desenvolvimento de novos equipamentos, a aplicação da inteligência artificial trouxe mudanças fundamentais na execução da Radiologia.⁽⁵³⁾

Conforme apresentamos neste estudo, a inteligência artificial pode ser utilizada como auxílio ao médico radiologista na triagem e interpretação dos exames em imagem,

alguns estudos mostraram que sistemas de inteligência artificial podem alcançar acurácia de médicos radiologista focados em tarefas específicas.⁽⁵⁴⁾

Um dos possíveis caminhos de derivação do projeto desenvolvido neste estudo é verificar o ganho de tempo e acurácia no diagnóstico, após a implantação de um sistema de inteligência artificial como ferramenta de auxílio no fluxo de trabalho do médico ao realizar a triagem e o diagnóstico por imagem.

Um ponto importante a ser comentado é a possibilidade de, a partir do projeto desenvolvido neste estudo, serem desenvolvidos algoritmos para detecção de outras patologias além da pneumonia. Olhando para o conjunto de dados utilizados nesse projeto notamos a disponibilidade de imagens com outros tipos de patologia que podem ser utilizadas como conjunto de dados de treinamento e validação para a construção de outros sistemas especialistas.⁽⁵⁵⁾

A aplicação da inteligência artificial no diagnóstico por imagem não se limita a atividade de classificação, no qual o sistema aponta se determinado exame possui ou não uma patologia. Alguns estudos dedicaram-se à tarefa de segmentação e mensuração da patologia pulmonar.⁽¹⁾ A segmentação e mensuração da patologia é uma das atividades chave executadas pelo radiologista durante a análise e diagnóstico do exame, ao indicar a localização da patologia e a sua medida o sistema de inteligência artificial contribui para evitar casos de falso negativo, ou seja, havia uma patologia, porém, ela não foi identificada. A segmentação pelo sistema de inteligência artificial contribui também para agilizar o processo de diagnóstico, pois já aponta ao médico, de forma prévia, a localização da patologia na imagem. O presente estudo não aplicou a segmentação como parte do escopo da pesquisa, focamos aqui na tarefa de classificação, mas entende-se que esse é um dos possíveis caminhos a serem seguidos e

implementados, pois contribui de forma significativa para a melhoria do diagnóstico por imagem.

Uma limitação importante sobre o presente estudo diz respeito à faixa etária indicada para a utilização do sistema. O conjunto de dados utilizado para treinar o algoritmo possui imagens de pacientes com idade entre 20 e 80 anos, dessa maneira não é indicado o uso do sistema para diagnóstico de pacientes que estejam fora dessa faixa etária, pois o desempenho na acurácia do diagnóstico pode ser baixo.

A informação é a base para a tomada de decisão, de modo que os médicos precisam dela para realizar suas tarefas durante sua rotina de trabalho, quanto mais informação de qualidade esses profissionais tiverem; melhor será o serviço prestado.⁽⁵⁶⁾ A tecnologia tem a capacidade de aumentar a eficiência do processamento das informações, por meio de uma coleta dados, análise e resposta em tempo real. O sistema desenvolvido neste estudo propõe esse ganho de eficiência, sendo aplicada a rotina de trabalho no diagnóstico por imagem como uma ferramenta de auxílio que trará ganho de performance na assertividade do diagnóstico, bem como, a diminuição do tempo para o diagnóstico.

6. CONCLUSÕES

Conclui-se que a utilização de um sistema de inteligência artificial para identificação de pneumonia em imagens de Raio-x do tórax mostra-se benéfica como ferramenta de auxílio na priorização do fluxo de trabalho pelo médico radiologista, bem como, uma segunda opinião para o diagnóstico do exame.

Estas observações corroboram a ideia de que o sistema desenvolvido neste estudo tem a capacidade de melhorar a eficiência dos diagnósticos por imagem, aumentando a rapidez na qual os resultados dos exames são liberados, assim como, contribui para o aumento da assertividade dos resultados.

Essa avaliação preliminar possibilitou verificar a utilidade de um sistema inteligente aplicado ao diagnóstico médico. Espera-se que este trabalho possa contribuir para o estabelecimento da utilização de sistemas computacionais inteligentes como ferramenta de auxílio à tomada de decisão eficiente para a melhoria dos diagnósticos médicos.

As futuras investigações desse estudo direcionam-se para a busca do aumento da acurácia do algoritmo. Por meio da análise e teste de outras técnicas de inteligência artificial ou mesmo estudo de outra arquitetura para o algoritmo, de modo que seja possível alcançar um novo nível de acurácia, aumentando a assertividade do diagnóstico.

O algoritmo desenvolvido neste estudo abre novas possibilidades para o estudo do diagnóstico automático de outras patologias pulmonares a partir de radiografias torácicas. A arquitetura do algoritmo desenvolvido, os dados analisados e o conjunto de dados utilizados permitem que outras patologias sejam detectadas de forma automática, a partir do algoritmo desenvolvido neste estudo. Considera-se, então, uma excelente oportunidade de aplicar a tecnologia para a melhoria da saúde.

7. REFERÊNCIAS

1. Ranschaert ER, Morozov SP, Algra PR. Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks. Springer. 2019.
2. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics*. 2017; 37(2):505-515. doi: <https://doi.org/10.1148/rg.2017160130>
3. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale chest x-ray Database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017; http://openaccess.thecvf.com/content_cvpr_2017/papers/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.pdf.
4. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500-510. doi: <https://doi.org/10.1038/s41568-018-0016-5>.
5. National Health, Lung, and Blood Institute [Internet]. How is Pneumonia Treated? [acesso em 13 nov 2021]. Disponível em: <https://web.archive.org/web/20160306030735/http://www.nhlbi.nih.gov/health/health-topics/topics/pnu/treatment>.
6. Ruuskanen O, Lahti E, Jennings LC, Murdoch DR. Viral pneumonia. *Lancet*. 2011;377(9773):1264-75. doi: [https://doi.org/10.1016/S0140-6736\(10\)61459-6](https://doi.org/10.1016/S0140-6736(10)61459-6).
7. Lodha R, Kabra SK, Pandey RM. Antibiotics for community-acquired pneumonia in children. *Cochrane Database Syst Rev*. 2013;2013(6):CD004874. doi: <https://doi.org/10.1002/14651858.CD004874.pub4>.
8. National Health, Lung, and Blood Institute [Internet]. How is Pneumonia Diagnosed? [acesso em 13 nov 2021]. Disponível em: <https://web.archive.org/web/20160307133513/http://www.nhlbi.nih.gov/health/health-topics/topics/pnu/diagnosis>.
9. Keras [Internet]. [acesso em 17 nov 2021]. Disponível em: <https://keras.io/api/applications/>
10. ImageNet [Internet]. [acesso em 17 nov 2021]. Disponível em: <https://www.image-net.org/index.php>.
11. West J, Ventura D, Warnick S. Spring research presentation: A theoretical foundation for inductive transfer. Brigham Young University, College of Physical and Mathematical Sciences, 2007;1(08).
12. Karimpanal TG, Bouffanais R. Self-organizing maps for storage and transfer of knowledge in reinforcement learning. *Adapt Behav*. 2019;27(2):111–126. doi: <https://doi.org/10.1177/1059712318818568>.

13. Tensorflow [Internet]. Aprendizado por transferências e ajustes. [acesso em 24 nov 2021]. Disponível em: https://www.tensorflow.org/tutorials/images/transfer_learning.
14. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. arXiv. 2015;1. doi: <https://doi.org/10.48550/arXiv.1511.08458>.
15. Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*. 1980;36(4):193-202. doi: <https://doi.org/10.1007/BF00344251>.
16. Matsugu M, Mori K, Mitari Y, Kaneda Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Netw*. 2003;16(5-6):555-9. doi: [https://doi.org/10.1016/S0893-6080\(03\)00115-1](https://doi.org/10.1016/S0893-6080(03)00115-1).
17. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Massachusetts: The MIT Press; 2016.
18. Aghdam HH, Heravi EJ. Guide to Convolutional Neural Networks: a practical application to traffic-sign detection and classification. Spain: Springer Cham; 2017;1. [acesso em 18 dez 2021]. doi: <https://doi.org/10.1007/978-3-319-57550-6>.
19. Venkatesan R, Li B. Convolutional neural networks in visual computing: a concise guide. Boca Raton: CRC Press; 2017. [acesso em 19 dez 2021]. doi: <https://doi.org/10.4324/9781315154282>.
20. Cireşan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J. Flexible, high performance convolutional neural networks for image classification. In: 22nd International Joint Conference on Artificial Intelligence. Barcelona: Centre Convencions Internacional Barcelona; 2011 [acesso em 25 out 2021]. doi: <https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-210>.
21. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6): 84–90. doi: <https://doi.org/10.1145/3065386>.
22. The Keras Blog [Internet]. Building powerful image classification models using very little data. [acesso em 26 nov de 2021]. Disponível em: <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>.
23. Tensorflow [Internet]. Ampliação de dados. [acesso em 05 dez 2021]. Disponível em: tensorflow.org/tutorials/images/data_augmentation.
24. Psy Image Search [Internet]. Keras ImageDataGenerator and Data Augmentation [acesso em 07 dez 2021]. Disponível em: <https://pyimagesearch.com/2019/07/08/keras-imagedatagenerator-and-data-augmentation/>
25. Allred R [Internet]. Image augmentation for deep learning using keras and histogram equalization. 10 jul 2017. [acesso em 20 dez 2021]. Disponível em:

<https://towardsdatascience.com/image-augmentation-for-deep-learning-using-keras-and-histogram-equalization-9329f6ae5085>.

26. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6(60):1-48. doi: <https://doi.org/10.1186/s40537-019-0197-0>.

27. Bird JJ, Faria DR, Ekárt A, Ayrosa PPS. from simulation to reality: CNN transfer learning for scene classification. In: 10th International Conference on Intelligent Systems (IS). Bulgaria: IEE; 2020:619-625 [acesso em 15 abr 2022]. doi: <https://doi.org/10.1109/IS48319.2020.9199968>.

28. Zanini RA, Colombini EL. Parkinson's disease EMG data augmentation and simulation with DCGANs and style transfer. *Sensors (Basel)*. 2020;20(9):2605. doi: <https://doi.org/10.3390/s20092605>.

29. Hosch WL. Machine learning [Internet]. 25 nov 2021. [acesso em 03 dez 2021]. Disponível em: <https://www.britannica.com/technology/machine-learning>.

30. Simon P. Too big to ignore: The business case for big data. 1st ed. Hoboken (NJ): John Wiley & Sons; 2013.

31. Kohavi R, Provost F. Glossary of terms. Machine Learning - special issue on Applications of Machine Learning and the Knowledge Discovery Process. *Mach Learn*. 1998;30:271-274. doi: <https://doi.org/10.1023/A:101718182689>.

32. Udacity [Internet]. [acesso em 05 dez 2021]. Disponível em: <https://www.udacity.com/course/ai-for-healthcare-nanodegree--nd320>.

33. Hardesty L. Explained: Neural networks [Internet]. 14 Apr 2017. [acesso em 05 dez 2021]. Disponível em: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>

34. Russell S, Norvig P. Artificial intelligence: a modern approach. 4th ed. Prentice Hall (NJ): Pearson; 2020.

35. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of machine learning. Cambridge (MA): MIT Press; 2012. 412p.

36. Lexico. overfitting [Internet]. [acesso em 10 dez 2021]. Disponível em: <https://www.lexico.com/definition/overfitting>.

37. Tetko IV, Livingstone DJ, Luik AI. Neural network studies, 1. Comparison of overfitting and overtraining. *J Chem Inf Comput Sci*. 1995;35:826-833. doi: <https://doi.org/10.1021/ci00027a006>.

38. Chicco D. Ten quick tips for machine learning in computational biology. *BioData Min*. 2017;10:35. doi: <https://doi.org/10.1186/s13040-017-0155-3>.

39. Leinweber DJ. Stupid data miner tricks: overfitting the S&P 500. *J Invest*. 2007;16(1):15-22. doi: <https://doi.org/10.3905/joi.2007.681820>.

40. James G, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. 1st ed. New York (NY): Springer; 2013.
41. Ripley BD. Pattern recognition and neural networks. Cambridge (UK): Cambridge University Press; 2007. doi: <https://doi.org/10.1017/CBO9780511812651>.
42. Brownlee J. What is the difference between test and validation datasets? [Internet]. 14 jul 2017. [acesso em 02 fev 2022]. Disponível em <https://machinelearningmastery.com/difference-test-validation-datasets/>
43. Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J Mach Learn Technol. 2011;2(1):37–63. doi: <https://doi.org/10.48550/arXiv.2010.16061>.
44. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006; 27(8): 861–874. doi: <https://doi.org/10.1016/j.patrec.2005.10.010>.
45. Tharwat A. Classification assessment methods. Appl Comput and Inform. 2020;17(1):168-192. doi: <https://doi.org/10.1016/j.aci.2018.08.003>.
46. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 2015;15:29. doi: <https://doi.org/10.1186/s12880-015-0068-x>.
47. Sasaki Y. The truth of the F-measure [Internet]. 2007 [acesso em 20 fev 2022]. Disponível em: <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26-Oct07.pdf>.
48. Kwon T, Lee SP, Kim D, Jang J, Lee M, Kang SU, Kim H, et al. Diagnostic performance of artificial intelligence model for pneumonia from chest radiography. PLoS One. 2021;16(4):e0249399. doi: <https://doi.org/10.1371/journal.pone.0249399>.
49. Salvatore C, Interlenghi M, Monti CB, Ippolito D, Capra D, Cozzi A, et al. Artificial intelligence applied to chest x-ray for differential diagnosis of covid-19 pneumonia. Diagnostics (Basel). 2021;11(3):530. doi: <https://doi.org/10.3390/diagnostics11030530>.
50. Wehbe RM, Sheng J, Dutta S, Chai S, Dravid A, Barutcu S, 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(1):E167-E176. doi: <https://doi.org/10.1148/radiol.2020203511>.
51. Digital Imaging and Communications in Medicine (DICOM) [Internet]. [acesso em 01 mar 2022]. Disponível em: <https://www.dicomstandard.org/>
52. Zhang Q, Chen Z, Liu G, Zhang W, Du Q, Tan J, et al. Artificial intelligence clinicians can use chest computed tomography technology to automatically diagnose coronavirus disease 2019 (COVID-19) pneumonia and enhance low-quality images. Infect Drug Resist. 2021;14:671-687. doi: <https://doi.org/10.2147/IDR.S296346>.

53. Lidströmer N, Ashrafian H. Artificial intelligence in medicine. Cham: Springer Cham; 2021. doi: https://doi.org/10.1007/978-3-030-58080-3_243-1.
54. Adams SJ, Henderson RDE, Yi X, Babyn P. Artificial intelligence solutions for analysis of x-ray images. *Can Assoc Radiol J*. 2021;72(1):60-72. doi: <https://doi.org/10.1177/0846537120941671>.
55. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *Eur J Radiol*. 2020;123:108774. doi: <https://doi.org/10.1016/j.ejrad.2019.108774>.
56. Augusto EDF. Aplicativo móvel para registro de atendimento pré-hospitalar (APH) em tempo real [dissertação]. São Paulo: Faculdade de Medicina de São José do Rio Preto (FAMERP), São José do Rio Preto; 2019.
57. Daley S. 40 AI in healthcare examples improving the future of medicine [Internet]. 10 jun 2022. [acesso 15 jun 2022]. Disponível em: <https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare>.