

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

Josué Yi Hon Shao

Detecção de Pneumonia em Imagens de Raio X de Tórax

Orientador: Professor Dr. Marcos Ribeiro Pereira Barretto

São Paulo

2021

Josué Yi Hon Shao

Detecção de Pneumonia em Imagens de Raio X de Tórax

Trabalho de conclusão de curso apresentado no Departamento de Engenharia Mecatrônica e Sistemas Mecânicos da Escola Politécnica da Universidade de São Paulo para obtenção do título de Engenheiro.

Área de Concentração: Engenharia Mecatrônica.

Orientador: Professor Dr. Marcos Ribeiro Pereira Barretto

São Paulo

2021

Catlogação-na-publicação

Shao, Josué

Detecção de pneumonia em imagens de raio x de tórax /

J. Shao - São Paulo, 2021.

34 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos.

1. Aprendizagem profunda 2. Rede neural convolucional
3. Transferência de aprendizado 4. Análise de imagens médicas
I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos

Resumo

A pneumonia é uma das doenças de infecção respiratória que mais causam mortes no mundo. Os principais motivos da sua alta taxa de mortalidade é o diagnóstico somente em estágios avançados da doença e a escassez de profissionais de saúde que podem detectá-la, principalmente em países subdesenvolvidos. Para minimizar os impactos da pneumonia foi desenvolvido um algoritmo de classificação binária de imagens de raio x de tórax, o exame comumente empregado para o diagnóstico da doença.

Para o desenvolvimento do modelo foi utilizada a aprendizagem profunda, mais especificamente as redes neurais convolucionais. Esta técnica para resolução de tarefas complexas de reconhecimento de padrões possui alto potencial para problemas de classificação de imagens. Além disso, foi implementada a transferência de aprendizado que reaproveita características de modelos anteriores de classificação de imagens e as adapta para uma nova aplicação.

Como avaliação do sistema desenvolvido, foram calculados os resultados das principais métricas de performance de modelos de classificação binária e comparados a resultados de outros modelos com aplicação semelhante. Além da comparação quantitativa há uma interpretação qualitativa do resultado obtido. O sistema apresentou avaliação satisfatória para o problema proposto, ainda que os testes realizados indiquem que os resultados podem ser aperfeiçoados conforme novos modelos de classificação de imagens são desenvolvidos.

Palavras-chave: aprendizagem profunda, rede neural convolucional, transferência de aprendizado, análise de imagens médicas

Abstract

Pneumonia is one of the world's leading respiratory infections. The main reasons for its high mortality rate are the diagnosis in final stages of the disease and the lack of health professionals who can detect it, especially in underdeveloped countries. In order to minimize the impact of pneumonia, a binary classification algorithm for chest x-ray images was developed, which is the most common test used for diagnosing the disease.

For the development of the model, deep learning was applied, specifically convolutional neural networks. This technique for solving complex pattern recognition tasks has high potential for image classification problems. Furthermore, transfer learning was also implemented. It reuses features from previous image classification models and adapts them for a new application.

As an evaluation of the developed system, the results of the main performance metrics of binary classification models were calculated and compared to results of other models with similar application. In addition to the quantitative comparison, there is a qualitative interpretation of the outcome. The system presented a satisfactory evaluation for the proposed problem, although the tests performed indicate that the results can be improved as new image classification models are developed.

Keywords: deep learning, convolutional neural network, transfer learning, medical image analysis

Lista de Ilustrações

Figura 1: Arquitetura de uma Rede Neural Convolutiva.....	15
Figura 2: Exemplos de imagens da base de dados disponível na plataforma Kaggle	21
Figura 3: Resultados de algoritmos anteriores para referência de comparação	22
Figura 4: Quantidade de imagens para treino	22
Figura 5: Distribuição de imagens para treino	22
Figura 6: Quantidade de imagens para teste	23
Figura 7: Distribuição de imagens para teste.....	23
Figura 8: Curva de Acurácia durante treinamento do modelo de RNC simples	24
Figura 9: Matriz de confusão da RNC simples.....	25
Figura 10: Curva ROC e AUC do modelo de RNC simples	27
Figura 11: Curva de Acurácia durante treinamento do modelo de RNC com transferência de aprendizado.....	28
Figura 12: Matriz de confusão da RNC com transferência de aprendizado	28
Figura 13: Curva ROC e AUC do modelo de RNC com transferência de aprendizado	30

Lista de Tabelas

Tabela 1: Distribuição da base de dados.....	20
Tabela 2: Resultados da RNC simples e da RNC com transferência de aprendizado	29
Tabela 3: Resultados dos dois modelos desenvolvidos e de um modelo encontrado na literatura	30

Lista de Siglas

RNC – Rede Neural Convolucional

Sumário

Lista de Ilustrações.....	7
Lista de Tabelas.....	8
Lista de Siglas.....	9
Sumário.....	10
1. Introdução.....	11
1.1 Motivação.....	11
1.2 Objetivo.....	12
1.3 Estrutura do Trabalho.....	12
2. Revisão Bibliográfica.....	13
2.1 Conceitos.....	13
2.2 Trabalhos Correlatos.....	16
3. Tecnologias.....	18
3.1 Keras.....	18
3.2 Aumento de Dados.....	19
4. Solução Proposta.....	20
4.1 Base de Dados.....	20
4.2 Categorias de Classificação.....	20
4.3 Avaliação do Modelo.....	21
5. Desenvolvimento da Solução.....	22
5.1 Explorando os Dados.....	22
5.2 Resultados.....	23
5.2.1 Resultados de uma RNC simples.....	24
5.2.2 Resultados com transferência de aprendizado.....	27
6. Conclusões.....	32
7. Referências Bibliográficas.....	33

1. Introdução

1.1 Motivação

A pneumonia é uma doença inflamatória aguda que acomete os pulmões e pode ser provocada por bactéria, vírus ou fungos. Suas principais manifestações clínicas são tosse com produção de expectoração, dor torácica, falta de ar e febre. Ela está entre as principais causas de morte no Brasil e no mundo. Seu impacto na vida e na saúde de uma população, especialmente em crianças, foi reconhecida ao ponto da Organização Mundial da Saúde (OMS) criar em 2009 o Dia Mundial da Pneumonia (12 de novembro) para conscientizar a importância de sua prevenção.

Um dos principais agravantes para a alta taxa de mortalidade da pneumonia é a detecção da doença somente em estágios críticos. Os sintomas são coincidentes com os de outras doenças menos agressivas e cotidianas como a gripe, elevando mais ainda a dificuldade no seu diagnóstico. Para a identificação da presença de pneumonia em um paciente, o exame comumente empregado é o raio x de tórax. Através da imagem de um pulmão obtida neste exame, um radiologista analisa cuidadosamente para verificar se o paciente está ou não infectado. Com o avanço da tecnologia e algoritmos para análises de imagem através de inteligência artificial, o diagnóstico pode ser obtido rapidamente e auxiliar o médico na detecção da doença de forma precisa e nas suas fases iniciais.

O desenvolvimento de um modelo capaz de analisar uma imagem pode ser usado para classificar exames de maior e menor complexidade, otimizando assim tempo e energia para que médicos radiologistas se concentrem nos de maior complexidade. Outra vantagem obtida com o uso de um algoritmo é a possibilidade de um médico usar o sistema como apoio para concluir um diagnóstico, conhecido como diagnóstico auxiliado por computador, considerando as precisões cada vez melhores com que esses algoritmos trabalham.

Além disso, devido ao maior volume de pacientes a serem examinados em comparação com a escassez de profissionais para detectar a pneumonia, especialmente em países subdesenvolvidos, um sistema

capaz de diagnosticar a doença pode ser implementado em regiões específicas e indicar alguma solução ou tratamento com antecedência.

1.2 Objetivo

O objetivo deste projeto é desenvolver um sistema capaz de classificar imagens de raio x de tórax e definir se a imagem representa um pulmão com pneumonia ou um pulmão saudável (sem pneumonia). O sistema deve classificar se a imagem representa um pulmão com pneumonia, mas não abrange seu agente causador (vírus, bactéria ou fungo).

O sistema deve receber como entrada uma imagem de raio x de tórax e retornar como saída uma probabilidade de o pulmão representado estar infectado ou não pela doença. Para avaliar a precisão do algoritmo, serão utilizadas 624 imagens de teste e os resultados comparados com algoritmos de outras literaturas.

1.3 Estrutura do Trabalho

Esta monografia tem como intuito explicar o contexto tecnológico de um sistema classificador de imagem radiológica e suas aplicações na detecção de pneumonia, cobrindo aspectos técnicos e teóricos, seguida de uma proposta de solução.

O capítulo 2 cita uma seleção de referências e artigos que inspiraram este projeto, seja pela tecnologia usada ou pelos tópicos abordados em análises de imagem médica através de aprendizagem profunda.

O capítulo 3 apresenta um aprofundamento das principais tecnologias utilizadas para a construção do sistema proposto neste trabalho.

O capítulo 4 discorre a respeito dos dados utilizados para construção do modelo e dos requisitos estabelecidos para o problema proposto.

O capítulo 5 apresenta os resultados obtidos com os modelos desenvolvidos e a interpretação destes resultados através dos principais indicadores de performance de algoritmos de classificação binária.

2. Revisão Bibliográfica

2.1 Conceitos

A fim de contribuir para uma melhor compreensão dos artigos selecionados em relação ao tema proposto, serão apresentados os principais termos e conceitos envolvidos. São eles análise de imagens médicas, diagnóstico auxiliado por computador, aprendizado de máquina, aprendizagem profunda, redes neurais convolucionais, transferência de aprendizado e as principais métricas de avaliação.

A análise de imagens médicas se desenvolveu em diferentes aspectos envolvendo reconhecimento de padrões, mineração de dados em imagens, visão computacional e aprendizado de máquina [1].

Em análises de imagens médicas, a precisão do diagnóstico de uma doença depende tanto da aquisição da imagem quanto da sua interpretação [2]. Nota-se, portanto, que um dos desafios enfrentados pelos radiologistas é uma qualidade baixa de exame de raio x, ou seja, quando a imagem está com baixo contraste gerando dificuldades de detectar e caracterizar lesões [1]. Para superar esta limitação, pesquisas estão sendo desenvolvidas em um ramo da análise de imagens médicas conhecido como diagnóstico auxiliado por computador [3].

Na radiologia, o diagnóstico auxiliado por computador está sendo adotado como uma “segunda opinião” para auxiliar leituras de imagem de raio x para determinar a presença de enfermidades. Uma tarefa importante no avanço do diagnóstico auxiliado por computador é detectar e classificar automaticamente uma doença através de uma imagem de raio x de tórax. Esta capacidade pode contribuir para melhorar a qualidade e a produtividade de radiologistas aumentando a precisão e a consistência de

diagnósticos ao mesmo tempo em que reduz o tempo de leitura de imagem [1].

O aprendizado de máquina é o estudo de algoritmos computacionais que são capazes de aprender relações e padrões complexos através de dados experimentais para tomadas de decisão [4]. O aprendizado de máquina mostrou-se uma forma eficaz de automatizar a análise e o diagnóstico de imagens médicas, especialmente em raios x de tórax [5]. Assim, o aprendizado de máquina contribui para o aprimoramento do diagnóstico auxiliado por computador.

Há alguns métodos de aprendizado de máquina que são aplicados em análises de imagens médicas, sendo a aprendizagem profunda a mais promissora [6, 7]. A base da maioria dos métodos de aprendizagem profunda utiliza redes neurais. Uma rede neural é composta por neurônios com algumas funções de ativação e parâmetros. Ela contém diversas camadas classificadas como camada de entrada, camada de saída e camadas ocultas (camadas entre a entrada e a saída). A arquitetura mais comum de aprendizagem profunda em análises de imagens médicas chama-se Redes Neurais Convolucionais (RNC). O principal motivo é que as RNC preservam as principais características ao filtrar imagens de entrada, o que é essencial em análises de exames de raios x de tórax [1]. Além disso, os algoritmos de RNC podem ser aplicados em paralelo, tornando-o eficiente para lidar com grandes conjuntos de informações durante sua construção e usabilidade. [1]

As Redes Neurais Convolucionais são inspiradas no córtex visual do cérebro e são usadas para resolver tarefas complexas de reconhecimento de padrões em imagens, reconhecendo tanto padrões lineares como não lineares [8]. RNC possuem bom desempenho em problemas de classificação de imagens pois utiliza menos parâmetros e conexões em comparação com outras redes neurais. As redes neurais artificiais, por exemplo, têm dificuldade com dados de imagens dado o alto grau de complexidade computacional envolvido [9].

RNC é uma rede neural sem realimentação (“feedforward”, em inglês) como mostra a figura (1). Ela consiste em 3 camadas principais de processamento: camada convolucional, camada de pooling e camada completamente conectada.

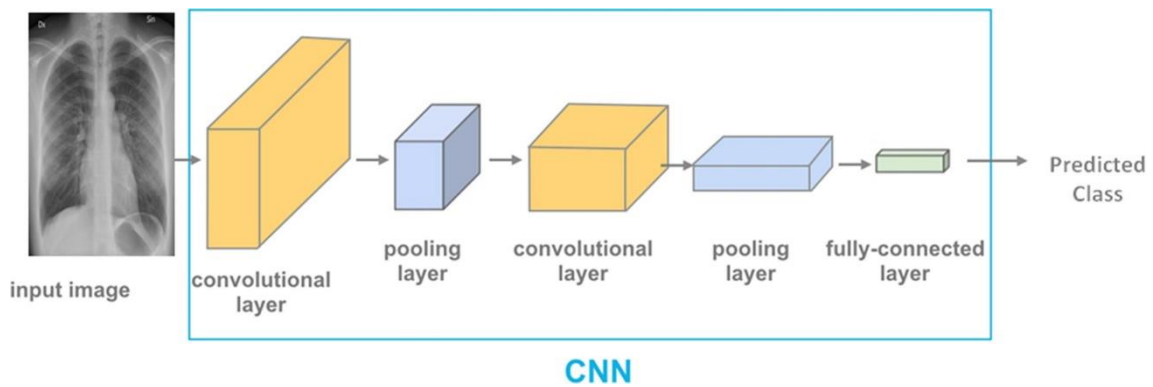


Figura 1 – Arquitetura de uma Rede Neural Convolutiva [10].

Camada convolutiva (convolutional layer): a imagem de entrada é convertida para uma forma matricial e reduzida através de um filtro mantendo as principais características da imagem. Este processo reduz as dimensões da imagem, facilitando seu processamento.

Camada de pooling (pooling layer): o objetivo desta camada é reduzir ainda mais a amostragem da imagem de entrada pelo agrupamento de pixels, assim diminuindo a complexidade computacional.

Camada completamente conectada (fully connected layer): após a imagem de entrada passar por procedimentos de redução de tamanho nas camadas anteriores, ela finalmente entra na rede neural propriamente dita. Nesta camada são calculados os pesos de acordo com a entrada para gerar uma previsão na saída.

O emprego de modelos de aprendizagem profunda, especialmente Redes Neurais Convolucionais, revelou um alto potencial de extrair características importantes em tarefas de classificação de imagens [11]. Esse processo de extração de recursos pode ser reaproveitado através de métodos de transferência de aprendizado (transfer learning), no qual modelos pré treinados de RNC aprendem características genéricas de dados de grande escala como “ImageNet”, que são posteriormente transferidos para a tarefa requerida [12]. A disponibilidade de modelos pré treinados de RNC como AlexNet [13], VGGNet [14] e Xception [15] auxiliam no procedimento de extração de características significativas e servem como modelos iniciais a serem aprimorados para uma tarefa específica.

As principais métricas para a avaliação de um modelo de aprendizado de máquina para classificação binária são acurácia, sensibilidade, especificidade, precisão e F1 score.

$$Accuracy = \frac{(TP + TN)}{(TP + FN) + (FP + TN)}$$

$$Sensitivity = \frac{(TP)}{(TP + FN)}$$

$$Specificity = \frac{(TN)}{(FP + TN)}$$

$$Precision = \frac{(TP)}{(TP + FP)}$$

$$F1\ Score = \frac{(2 * TP)}{(2 * TP + FN + FP)}$$

Neste projeto os significados das variáveis são:

True Positive (TP): número de imagens com pneumonia identificado como pneumonia.

True Negative (TN): número de imagens normais identificado como normal.

False Positive (FP): número de imagens normais incorretamente identificado como pneumonia.

False Negative (FN): número de imagens com pneumonia incorretamente identificado como normal.

2.2 Trabalhos Correlatos

A tarefa de classificar imagens de raio x de tórax tem sido significativamente explorada no campo da medicina diagnóstica. A seguir serão apresentados alguns artigos que foram publicados para tratar este assunto.

Rapjurkar et al. [16] desenvolveu um modelo de aprendizagem profunda para detecção de pneumonia em imagens de raio x de tórax na base de dados “ChestX-ray14”. Esta base de dados possui mais de 100.000 imagens de raio x de tórax com 14 classificações de doenças. Com o algoritmo desenvolvido nomeado “CheXNet”, que é uma Rede Neural Convolutiva de 121 camadas, a classificação das imagens de raio x de tórax atingiu uma performance que supera a de radiologistas ativos considerando a métrica de avaliação chamada “F1 score”. Além de detectar a pneumonia, o modelo ainda era capaz de identificar outras 13 doenças. O modelo oferece um desempenho excelente e uma expectativa de melhorar a prestação de cuidados de saúde.

Rubin et al. [17] apresentou uma RNC dupla que realiza reconhecimento automático em larga escala de imagens frontais e laterais de raio x de tórax com a base de dados “Mimic-cxr”, composta de 473.064 imagens de radiografias de tórax de 63.478 pacientes. Esta rede neural foi desenvolvida para detectar diversas doenças comuns no tórax. Primeiramente foi desenvolvido um modelo individual para a visão frontal e outro para a visão lateral dos raios x de tórax. Posteriormente foi desenvolvida uma arquitetura que processa simultaneamente as imagens frontais e laterais de raio x dos pacientes, melhorando a performance quando comparada com as anteriores de visões separadas. Os autores ainda procuram aperfeiçoar o desempenho do modelo através de técnicas de aumento de dados e normalização de imagem para auxiliar no fluxo de trabalho do processo de identificação de doenças comuns do tórax.

Lakhani et al. [18] desenvolveu uma RNC para classificação automática de tuberculose pulmonar através de radiografias de tórax. Duas RNC pré treinadas chamadas de “AlexNet” e “GoogleNet” foram usadas para classificar se as imagens mostravam manifestações de tuberculose pulmonar ou se eram imagens de pulmões saudáveis. O conjunto de dados foi pré processado com técnicas de aumento de dados e o modelo teve uma excelente performance pela métrica de área sob a curva (conhecida como AUC) de 0,99. Além disso, o algoritmo desenvolvido teve uma especificidade de 100% e sensibilidade de 97,3%.

Cicero et al. [19] desenvolveu uma RNC para detectar e classificar anormalidades em imagens de raios x de tórax frontal. A dimensão das imagens de entrada foi de 256x256 pixels. O AUC do modelo foi de 0,964

com uma média de especificidade e sensibilidade de 91%, mostrando que RNC podem ser desenvolvidas com alta acurácia de classificação e auxiliar em procedimentos de diagnóstico.

Anthimopoulos et al. [20] apresentou um modelo de RNC para identificar padrões em doenças pulmonares intersticiais (grupo de doenças respiratórias caracterizadas pela progressiva cicatrização do interstício pulmonar resultando em insuficiência respiratória). Seu modelo consiste em 5 camadas convolucionais, empregando a função de ativação “ReLU”, camadas de pooling média e 3 camadas densas (camadas completamente conectadas). A base de dados no qual o modelo foi treinado contém 7 categorias num total de 14.696 imagens. O modelo alcançou a acurácia de 85,5% e demonstrou o potencial das RNC para analisar padrões em imagens de pulmão. Em trabalhos futuros, os autores pretendem estender o modelo para classificar imagens 3D para servir como ferramenta de suporte para fins de diagnóstico.

3. Tecnologias

3.1 Keras

O Keras é uma biblioteca de aprendizagem profunda para diversas linguagens de programação como Python e R. Ele é amplamente usado por estatísticos e facilita a modelagem de redes neurais. Para o projeto proposto neste trabalho, ele será usado para desenvolver e modelar o algoritmo de redes neurais convolucionais para detecção de pneumonia em imagens de raio x de tórax.

Ele contribui para reduzir o código escrito e seu processamento através de um conjunto de rotinas e padrões de programação consistentes e simples, minimizando o número de ações necessárias do usuário para casos de uso comuns e fornece resposta clara sobre os erros. Estas características facilitam o aprendizado e a implementação do Keras.

3.2 Aumento de dados (Data Augmentation)

O aumento de dados é uma técnica amplamente usada para aprimorar o treinamento de redes neurais convolucionais. Ele é utilizado para prevenir que o modelo desenvolva um desempenho satisfatório somente nas imagens utilizadas para treino, mas não alcance a mesma performance nas imagens de teste. Ou seja, ele auxilia para que o modelo se torne mais generalista utilizando a mesma base de dados, mas aumentando a sua quantidade através de diversas variações nas imagens. Dessa forma, o aumento de dados pode impedir que a rede neural aprenda padrões irrelevantes nas imagens analisadas e melhore sua capacidade preditiva.

A seguir são apresentadas as principais técnicas de aumento de dados em imagens.

1. Espelhamento (Flipping)

Espelhamento é uma técnica de inversão da imagem original, podendo ser aplicada tanto na horizontal como na vertical.

2. Rotação (Rotation)

Na rotação as imagens são giradas no sentido horário ou anti-horário em qualquer grau.

3. Redimensionamento (Scale)

As imagens podem ser aumentadas ou diminuídas para garantir que o objeto a ser estudado não precise ter exatamente as mesmas dimensões dentro da imagem.

4. Translação (Translation)

Translação envolve movimentar a posição do objeto que se deseja analisar na imagem para evitar distorções posicionais nos dados.

5. Corte (Cropping)

As imagens são cortadas conservando o objeto de estudo e redimensionadas para o tamanho original.

Neste projeto foi utilizada a função “ImageDataGenerator” disponível no Keras que aplica de forma aleatória as transformações de imagens listadas acima para aumento de dados.

4. Solução Proposta

4.1 Base de Dados

A precisão de um modelo está diretamente relacionada com o tamanho da base de dados, ou seja, o uso de grandes conjuntos de dados melhora a precisão do modelo [10]. Portanto, foi selecionada uma base de dados suficientemente grande para o problema proposto e dentro das capacidades de processamento computacionais disponíveis.

A base de dados selecionada chama-se “Chest X-Ray Images (Pneumonia)”. Ela foi montada por Paul Mooney e está disponível na plataforma Kaggle. A base de dados contém 5216 imagens para treino e 624 imagens para a fase de teste. As imagens estão em escala de cinza com a dimensão de 224x224 pixels. Os dados encontram-se distribuídos na plataforma como mostra a tabela (1).

Categoria	Treino	Teste	Total
Normal	1341	234	1575
Pneumonia	3875	390	4265
Total	5216	624	5840

Tabela 1 - Distribuição da base de dados

4.2 Categorias de classificação

A seguir na figura (2) estão alguns exemplos de imagens da base de dados.



Figura 2 - Exemplos de imagens da base de dados disponível na plataforma Kaggle

As imagens com pulmão de pacientes diagnosticados com pneumonia apresentam regiões difusas quando estas são translúcidas em comparação com um pulmão normal, tanto na pneumonia bacteriana como na viral. Nesta proposta de solução será adotada a classificação binária entre um pulmão normal e um pulmão com pneumonia. As imagens de raio x de tórax mostradas acima e da base de dados são de pacientes na faixa etária de 1 a 5 anos do Centro Médico Feminino e Infantil, localizado na cidade de Cantão na China (Women and Children's Medical Centre, Guangzhou).

O objetivo da classificação é detectar a presença ou ausência de pneumonia.

4.3 Avaliação do Modelo

Alguns modelos de trabalhos encontrados na literatura utilizando a mesma base de dados deste projeto foram VGG16, VGG19, ResNet50 e Inception-v3 [10] e foram utilizados como referência para avaliação dos resultados deste trabalho. A figura (3) mostra os principais indicadores de performance destes modelos.

Model Name	Accuracy	Recall	F1 Score
VGG16	87.18%	96%	90%
VGG19	88.46%	95%	91%
ResNet50	77.56%	97%	84%
Inception-v3	70.99%	84%	78%

Figura 3 - Resultados de algoritmos anteriores para referência de comparação [10]

5. Desenvolvimento da Solução

5.1 Explorando os Dados

Para início do desenvolvimento do modelo foram importadas as imagens e observada a distribuição dos dados no código em python. As figuras (4) e (5) representam os dados de treino e as figuras (6) e (7) representam os dados de teste.

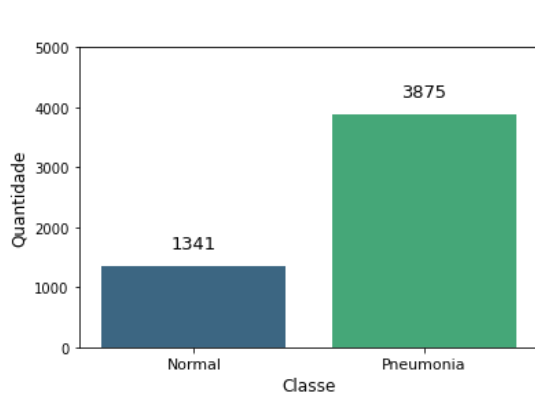


Figura 4 – Quantidade de imagens para treino

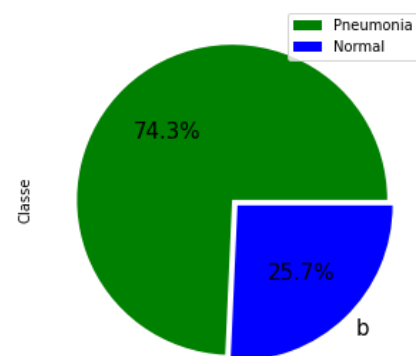


Figura 5 – Distribuição de imagens para treino

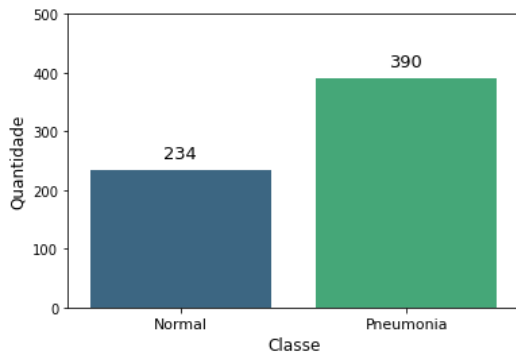


Figura 6 – Quantidade de imagens para teste

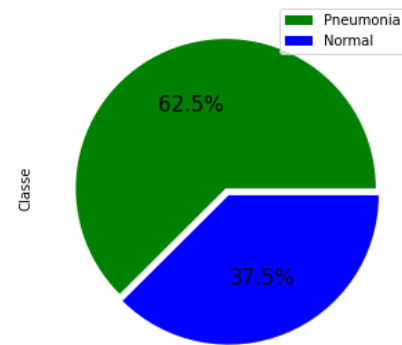


Figura 7 – Distribuição de imagens para teste

Uma dificuldade encontrada durante a construção do modelo no jupyter notebook em python foi a capacidade computacional de processar a quantidade total de imagens para testar um modelo inicial. Portanto, a amostra foi reduzida de forma arbitrária para 743 imagens de treino e 145 imagens de teste e os resultados iniciais de acurácia atingiram 68,96%.

Para continuar a construção do modelo foi sugerido pelo professor orientador (Marcos Barretto) o uso do “Google Colab” que permite usar um servidor do Google para testes e desenvolvimento. Além de melhorar o desempenho utilizando todas as imagens da base de dados, foi aplicada a transferência de aprendizado (transfer learning) exposta na revisão bibliográfica.

5.2 Resultados

Os resultados a seguir foram obtidos a partir de dois modelos de RNC. No primeiro modelo foi desenvolvido uma RNC simples e no segundo modelo foi utilizada a técnica de transferência de aprendizado (transfer learning), utilizando o modelo pré treinado “ResNet152V2” como base. Serão apresentados os resultados individuais de cada modelo e o comparativo para mostrar qual obteve a melhor performance.

Para a fase de treinamento dos dados, foi necessário separar parte dos dados de treino para validação do modelo conforme este era treinado.

Portanto, 80% desses dados foram efetivamente utilizados para o treinamento do modelo e os outros 20% para validação.

5.2.1 Resultados de uma RNC simples

A seguir na figura (8) é apresentada a evolução do modelo da RNC simples. Foram utilizadas 20 iterações (epochs) e nota-se que a partir da quinta iteração já foi obtida uma acurácia nos dados de validação próxima da alcançada pelo resultado final, em torno de 95%.

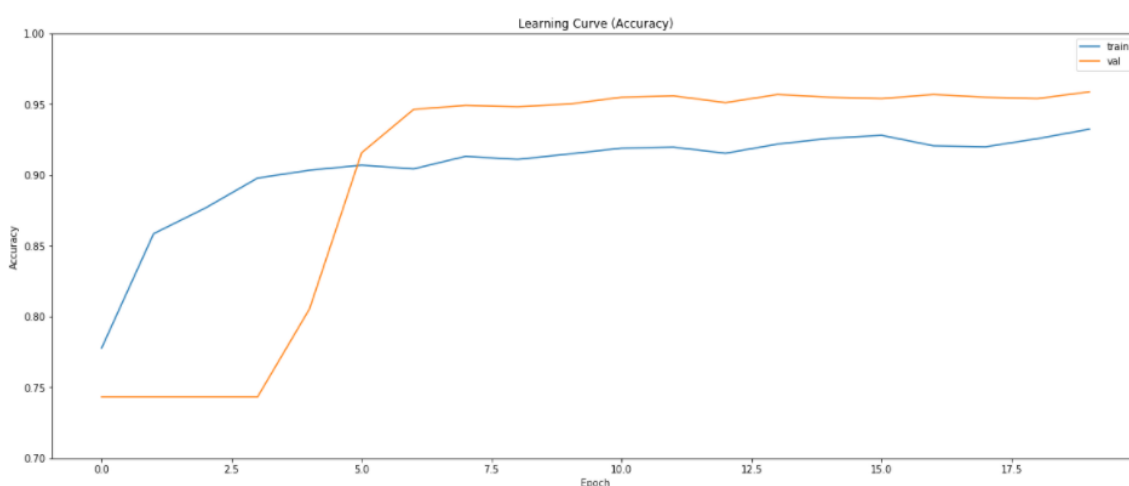


Figura 8 - Curva de Acurácia durante treinamento do modelo de RNC simples

Ao final do treinamento o modelo da RNC simples alcançou uma acurácia de 95,9%. Este valor representa a proporção da classificação binária feita corretamente nos dados de validação. O resultado é satisfatório e mostra que o modelo desenvolvido é capaz de realizar a classificação dos dados separados para validação.

O modelo foi então aplicado nos dados de teste e os resultados estão representados na matriz de confusão da figura (9).

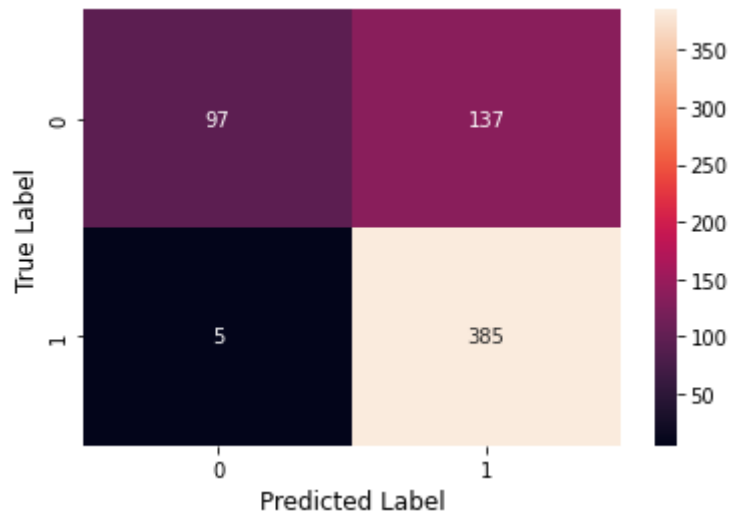


Figura 9 - Matriz de confusão da RNC simples

Através dos dados da matriz de confusão é possível avaliar o desempenho do algoritmo de classificação.

True Positive (TP): 385

True Negative (TN): 97

False Positive (FP): 137

False Negative (FN): 5

Com estes valores podemos calcular os valores das principais métricas de performance do modelo.

$$Accuracy = \frac{(TP + TN)}{(TP + FN) + (FP + TN)} = \frac{(385 + 97)}{(385 + 5) + (137 + 97)} = 77,2\%$$

$$Sensitivity = \frac{(TP)}{(TP + FN)} = \frac{(385)}{(385 + 5)} = 98,7\%$$

$$Specificity = \frac{(TN)}{(FP + TN)} = \frac{(97)}{(137 + 97)} = 41,5\%$$

$$Precision = \frac{(TP)}{(TP + FP)} = \frac{(385)}{(385 + 137)} = 73,8\%$$

$$F1\ Score = \frac{(2 * TP)}{(2 * TP + FN + FP)} = \frac{(2 * 385)}{(2 * 385 + 5 + 137)} = 84,4\%$$

Para avaliar se estes resultados são satisfatórios deve-se levar em conta que a proporção dos dados de teste é de 62,5% de imagens com pneumonia e 37,5% de imagens sem pneumonia (figura 7). Portanto, qualquer modelo desenvolvido deve obter uma acurácia superior a 62,5%, pois basta uma classificação simples de que todas as imagens de teste são de pulmão com pneumonia e a acurácia atingida seria de 62,5%.

O resultado de acurácia do modelo de RNC simples foi de 77,2% e significa que houve aprendizado pois foi superior a 62,5%. Porém, este não é um resultado aceitável pois trata-se de diagnosticar uma doença e as consequências desse erro podem ser fatais.

A sensibilidade de 98,7% demonstra que o modelo detectou essa proporção dentre os doentes na base de teste. A princípio pode parecer um resultado ótimo, mas se for considerado que uma pequena amostra dentre os doentes foi diagnosticada como tendo um pulmão sem pneumonia, estes pacientes não seriam tratados e os sintomas iriam se agravar. Neste caso de classificação binária o falso negativo possui um peso maior que o falso positivo.

A especificidade de 41,5% indica que o modelo não aprendeu de forma adequada a identificar as imagens com pulmão sem pneumonia. Outra forma de avaliar esta questão é a proporção dos resultados, no qual o modelo classificou 522 imagens com a doença (83,7%) e apenas 102 imagens sem a doença (16,3%). Esta proporção está longe da proporção real de 62,5% e 37,5%. Podemos concluir que o modelo tem uma tendência maior de classificar as imagens como representação de um pulmão com pneumonia.

A precisão de 73,8% representa a quantidade acertada pelo modelo dentre os que foram classificados como positivos. Este resultado mostra que o modelo ainda precisa ser trabalhado para conseguir selecionar de forma precisa somente as imagens que realmente possuem um pulmão com pneumonia.

O F1 Score é uma maneira de observar somente uma métrica ao invés de duas (sensibilidade e precisão). Seu cálculo é obtido pela média harmônica entre as duas. Seu valor fica mais próximo do valor menor se comparado a uma média aritmética simples. Portanto, se o F1 Score é baixo significa que uma das duas métricas está baixa.

Todas as métricas de avaliação de modelo calculadas auxiliam no entendimento e na interpretação do modelo. Há ainda outra medida utilizada para comparação entre modelos chamada de AUC (“area under the ROC curve”). O AUC é uma métrica que varia de 0 a 1 como as outras e mede a qualidade das previsões do modelo. Neste modelo de RNC simples foi alcançado um resultado de AUC de 93% como mostra a figura (10).

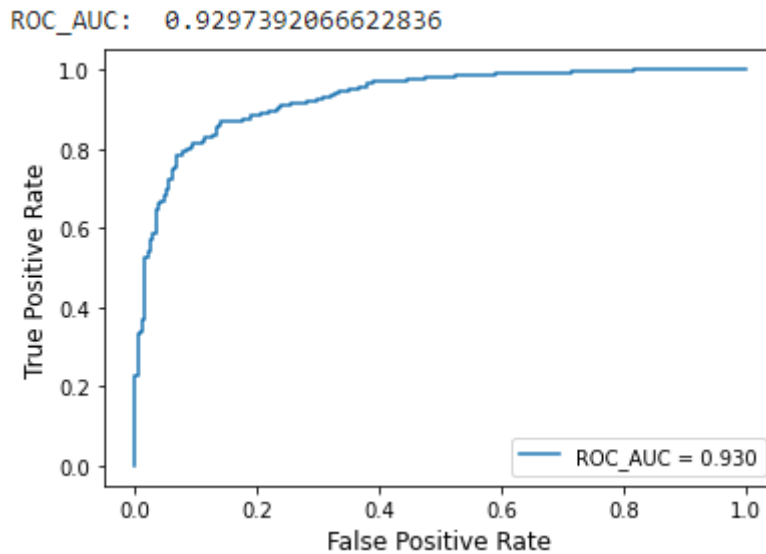


Figura 10 - Curva ROC e AUC do modelo de RNC simples

5.2.2 Resultados com transferência de aprendizado

Nesta seção são apresentados os resultados do algoritmo utilizando a transferência de aprendizado (transfer learning) através do modelo pré treinado “ResNet152V2” disponível no Keras. Esta abordagem consiste em aproveitar as características de leitura de imagem deste modelo e alimentá-lo com os dados de treinamento para utilizá-lo ao propósito deste projeto de detecção de pneumonia.

Diferentemente do modelo de RNC simples em que foram realizadas 20 iterações, o tempo de processamento para essa quantidade de repetições neste modelo se estendeu demasiadamente. Para contornar este problema a quantidade de iterações foi reduzida para cinco, pois nesta iteração o modelo de RNC simples atingiu acurácia próxima do resultado final (em torno de 95%).

Após 5 iterações o modelo atingiu a acurácia de 95,3% nos dados de validação como mostra a figura (11).

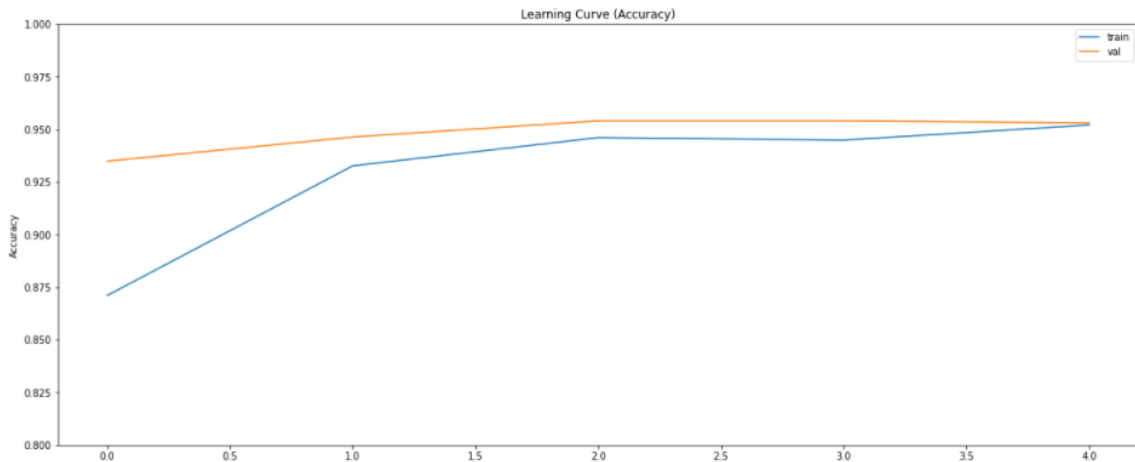


Figura 11 - Curva de Acurácia durante treinamento do modelo de RNC com transferência de aprendizado

Nota-se que no modelo de RNC simples a acurácia foi maior, no valor de 95,9%. Este resultado mostra que o primeiro modelo se adaptou a classificar melhor os dados de validação, mas isso não significa necessariamente que o primeiro modelo é melhor. Para poder realizar esta comparação deve-se medir os resultados obtidos nos dados de teste pois estes são dados novos ao modelo e podem apontar qual é o modelo mais generalista.

Na figura (12) estão apresentados os resultados da matriz de confusão do modelo com transferência de aprendizado e em seguida o racional para calcular os valores das principais métricas de performance do modelo.

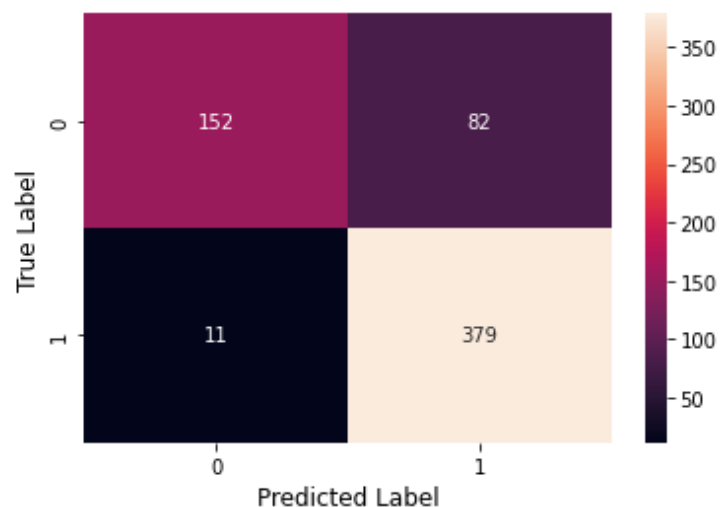


Figura 12 - Matriz de confusão da RNC com transferência de aprendizado

True Positive (TP): 379

True Negative (TN): 152

False Positive (FP): 82

False Negative (FN): 11

$$Accuracy = \frac{(TP + TN)}{(TP + FN) + (FP + TN)} = \frac{(379 + 152)}{(379 + 11) + (82 + 152)} = 85,1\%$$

$$Sensitivity = \frac{(TP)}{(TP + FN)} = \frac{(379)}{(379 + 11)} = 97,2\%$$

$$Specificity = \frac{(TN)}{(FP + TN)} = \frac{(152)}{(82 + 152)} = 65,0\%$$

$$Precision = \frac{(TP)}{(TP + FP)} = \frac{(379)}{(379 + 82)} = 82,2\%$$

$$F1\ Score = \frac{(2 * TP)}{(2 * TP + FN + FP)} = \frac{(2 * 379)}{(2 * 379 + 11 + 82)} = 89,1\%$$

Na tabela (2) estão reunidos os resultados da RNC simples e da RNC com transferência de aprendizado a partir do modelo pré treinado “ResNet152V2”.

Métrica	Simple	ResNet
Accuracy	77,2%	85,1%
Sensitivity	98,7%	97,2%
Specificity	41,5%	65,0%
Precision	73,8%	82,2%
F1 Score	84,4%	89,1%

Tabela 2 - Resultados da RNC simples e da RNC com transferência de aprendizado

Com exceção da sensibilidade em que foi alcançada 97,2% no segundo modelo, em todas as outras métricas o algoritmo com transferência de aprendizado performou melhor.

A maior diferença foi no indicador de especificidade no qual houve um aprimoramento significativo de 41,5% para 65%. Este foi um resultado crítico da RNC simples discutido anteriormente e o segundo modelo foi capaz de classificar as imagens de pulmão sem pneumonia de forma muito mais adequada.

Para definir qual modelo desempenhou o melhor resultado, segue a curva ROC e o AUC obtido no segundo modelo (figura 13).

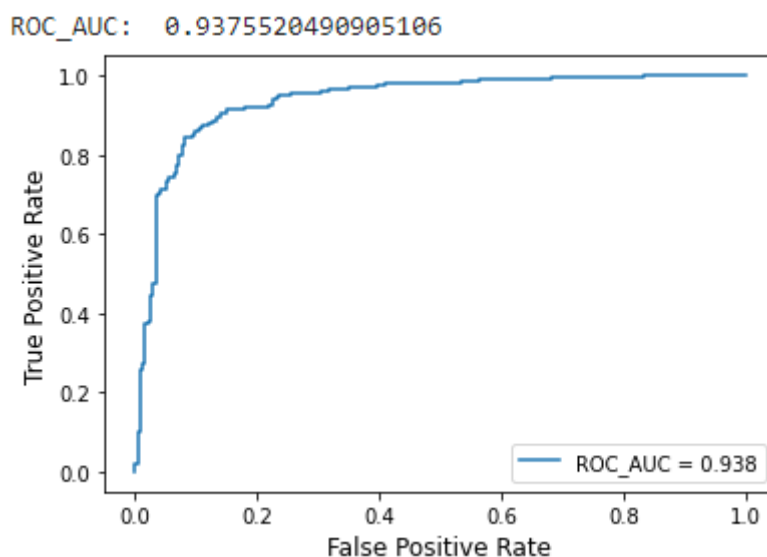


Figure 13 - Curva ROC e AUC do modelo de RNC com transferência de aprendizado

O AUC obtido foi de 93,8% e, portanto, o segundo modelo é capaz de classificar de forma mais adequada que o primeiro modelo.

Como comparação final, a tabela (3) apresenta os resultados obtidos pela RNC simples, RNC com transferência de aprendizado utilizando o modelo “ResNet152V2” e o resultado de outro trabalho mostrado na figura (3) que utilizou o modelo “ResNet50” como base de transferência de aprendizado.

Métrica	Simples	ResNet152V2	ResNet50
Accuracy	77,2%	85,1%	77,6%
Sensitivity	98,7%	97,2%	97,0%
F1 Score	84,4%	89,1%	84,0%

Tabela 3 - Resultados dos dois modelos desenvolvidos e de um modelo encontrado na literatura

Dentre os 3 modelos o que obteve o melhor resultado foi a RNC desenvolvida com transferência de aprendizado baseada no modelo “ResNet152V2” que alcançou acurácia de 85,1%. Considerando que a “ResNet152V2” é uma versão aperfeiçoada da “ResNet50”, a expectativa deste resultado era esperada. Além disso, a proximidade dos resultados significa que o modelo desenvolvido está alinhado com o de trabalhos anteriores, demonstrando um resultado satisfatório para a solução do problema proposto.

6. Conclusões

Através de métodos de transferência de aprendizado e técnicas de aumento de dados, o resultado obtido no modelo final com acurácia de 85,1% é coerente com os resultados observados em outros trabalhos citados na revisão bibliográfica. Entretanto, considerando o risco de um diagnóstico equivocado de uma doença que pode ser fatal como a pneumonia, o ideal é que seja alcançada uma acurácia próxima de 100%. Portanto, ainda há um longo percurso de pesquisa e desenvolvimento no ramo de análise de imagens médicas para diagnóstico auxiliado por computador.

Neste trabalho algumas limitações que impediram um resultado melhor ainda foram a quantidade de imagens na base de dados e a capacidade de processamento computacional. Inicialmente o código estava sendo escrito no jupyter notebook e posteriormente foi passado para o Google Colab para obter uma velocidade de processamento mais rápida. Com aproximadamente cinco mil imagens como base para treinamento dos modelos e a utilização de técnicas de aumento de dados, mesmo o modelo mais simples levou de 4 a 5 horas para rodar 20 iterações. O modelo com transferência de aprendizado levou de 6 a 7 horas para rodar 5 iterações. Com uma base de dados maior, velocidade de processamento mais veloz e uma quantidade maior de iterações, provavelmente os modelos desenvolvidos teriam uma performance significativamente melhor.

Os algoritmos para classificação de imagens continuam sendo desenvolvidos e aprimorados conforme aumentam a capacidade computacional e a qualidade e quantidade de dados. Por enquanto para fins de diagnóstico esses algoritmos com acurácia de cerca de 90% não podem substituir o diagnóstico de um profissional de saúde. Atualmente eles são utilizados apenas como auxílio para tomada de decisão de diagnóstico e também podem ser aplicados para classificar níveis de complexidade de diagnóstico, segregando os de maior complexidade para profissionais mais experientes e, assim, otimizando o tempo destes deixando os diagnósticos mais simples para aqueles com menos experiência.

7. Referências Bibliográficas

- [1] RAHMAT, Taufik; ISMAIL, Azlan; Aliman, Sharifah. **Chest X-rays Image Classification in Medical Image Analysis**. Applied Medical Informatics, v. 40, n. 3-4, p. 63–73, 2018.
- [2] GREENSPAN, Hayit; GINNEKEN, Bram; SUMMERS, Ronald. Guest Editorial **Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique**. IEEE transactions on medical imaging, v. 35, n. 5, 2016.
- [3] DOI, Kunio. **Computer-aided diagnosis in medical imaging: historical review, current status and future potential**. Computarized Medical Imaging Graphics, v. 31, n. 4-5, p. 198-211, 2007.
- [4] BISHOP, Christopher. **Pattern Recognition and Machine Learning**. Journal of Electronic Imaging, 2007.
- [5] WANG, Shijun; SUMMERS, Ronald. **Machine learning and radiology**. Medical Image Analysis, v. 16, n. 5, p. 933-951, 2012.
- [6] LITJENS, Geert et al. **A survey on deep learning in medical image analysis**. Medical Image Analysis, v. 42, p. 60-88, 2017.
- [7] KER, Justin et al. **Deep Learning Applications in Medical Image Analysis**. Special section on soft computing techniques for image analysis in the medical industry current trends, challenges and solutions, 2017.
- [8] CIRESAN, Dan et al. **High-Performance Neural Networks for Visual Object Classification**. arXiv preprint arXiv:1102.0183, 2011.
- [9] O'SHEA, Keiron; NASH, Ryan. **An introduction to convolutional neural networks**. arXiv preprint arXiv:1511.08458, 2015.
- [10] JAIN, Rachna et al. **Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning**. Measurement, v. 165, 2020.
- [11] RAZAVIAN, Ali et al. **CNN features off-the-shelf: an astounding baseline for recognition**. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, p. 806-813, 2014.
- [12] VARSHNI, Dimpy et al. **Pneumonia Detection Using CNN based Feature Extraction**. IEEE International Conference on Electrical, Computer and

Communication Technologies (ICECCT), p. 1-7, 2019.

- [13] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey. **Imagenet classification with deep convolutional neural networks**. Advances in neural information processing systems, v. 25, p. 1097-1105, 2012.
- [14] SIMONYAN, Karen; ZISSERMAN, Andrew. **Very deep convolutional networks for large-scale image recognition**. arXiv preprint arXiv:1409.1556, 2014.
- [15] CHOLLET, François. **Xception: Deep learning with depthwise separable convolutions**. Proceedings of the IEEE conference on computer vision and pattern recognition, p. 1251-1258, 2017.
- [16] RAJPURKAR, Pranav et al. **CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning**. arXiv preprint arXiv:1711.05225, 2017.
- [17] RUBIN, Jonathan et al. **Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Networks**. arXiv preprint arXiv:1804.07839, 2018.
- [18] LAKHANI, Paras; SUNDARAM, Baskaran. **Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks**. Radiology, v. 284, n. 2, p. 574-582, 2017.
- [19] CICERO, Mark et al. **Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs**. Investigative Radiology, 2017.
- [20] ANTHIMOPOULOS, Marios et al. **Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network**. IEEE Transactions on Medical Imaging, v. 35, n. 5, p. 1207-1216, 2016.