Universidade Estadual de Campinas
Instituto de Computação

Vinicius Teixeira de Melo

Classificação Multirrótulos de Radiografias de Tórax Utilizando Aprendizado de Máquina Profundo

CAMPINAS
2021

Vinicius Teixeira de Melo

Classificação Multirrótulos de Radiografias de Tórax Utilizando Aprendizado de Máquina Profundo

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Zanoni Dias
Coorientador: Prof. Dr. Hélio Pedrini

Este exemplar corresponde à versão final da Dissertação defendida por Vinicius Teixeira de Melo e orientada pelo Prof. Dr. Zanoni Dias.

CAMPINAS
2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M491c

Melo, Vinicius Teixeira de, 1998-
Classificação multirrótulos de radiografias de tórax utilizando aprendizado de máquina profundo / Vinicius Teixeira de Melo. – Campinas, SP : [s.n.], 2021.

Orientador: Zanoni Dias.
Coorientador: Hélio Pedrini.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Classificação de imagem. 2. Aprendizado de máquina. 3. Tórax - Radiografia. I. Dias, Zanoni, 1975-. II. Pedrini, Hélio, 1963-. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Multi-label classification of chest x-rays using deep learning
Palavras-chave em inglês:
Image classification
Machine learning
Chest - Radiography
Área de concentração: Ciência da Computação
Titulação: Mestre em Ciência da Computação
Banca examinadora:
Zanoni Dias [Orientador]
Levy Boccato
Alexandre Mello Ferreira
Data de defesa: 08-04-2021
Programa de Pós-Graduação: Ciência da Computação

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-7790-6930
- Currículo Lattes do autor: http://lattes.cnpq.br/8024065202088399

Universidade Estadual de Campinas
Instituto de Computação

Vinicius Teixeira de Melo

Classificação Multirrótulos de Radiografias de Tórax Utilizando Aprendizado de Máquina Profundo

Banca Examinadora:

- Prof. Dr. Zanoni Dias
  IC/Unicamp
- Prof. Dr. Levy Boccato
  FEEC/Unicamp
- Prof. Dr. Alexandre Mello Ferreira
  IC/Unicamp

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 08 de abril de 2021

Agradecimentos

- Agradeço muito a todos que contribuíram direta ou indiretamente para que eu pudesse chegar neste momento da minha vida.

- Primeiramente, agradeço à minha mãe Valdelice, que sempre me apoiou em todas as escolhas que me fizeram chegar até aqui, incentivando-me a estudar desde o começo para que eu, assim, pudesse ter um futuro melhor. Ao meu pai, que sempre procurou me ajudar de toda forma que pudesse e contribuiu em parte para eu me tornar a pessoa que sou. Também agradeço ao meu irmão (o Mão) por todo apoio e ajuda durante este período do mestrado que estou longe de casa.

- Agradeço à minha namorada Maria Jêsca, por ter me ajudado a passar pelo ano mais difícil da minha vida e de muita gente (2020), por conta da pandemia e tudo mais. Sempre me apoiando e me fazendo companhia nos momentos mais necessários, sou muito grato por tê-la por perto.

- Agradeço aos meus orientadores, prof. Zanoni e prof. Hélio, por todo o apoio desde antes de eu vir para Campinas, sempre se preocupando com características além da pesquisa. Pude aprender muito com as experiências obtidas nesses dois anos, vou levar alguns aprendizados para a vida toda. Agradeço também a todos os professores que tive durante a minha vida acadêmica.

- O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001. O presente trabalho foi realizado também com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo nº 2019/20875-8.

- Agradeço a todas as pessoas que moram aqui em casa, pelos momentos de descontração e brincadeiras. Agradeço também a um casal de amigos, Juliana e Pedro, por todas as conversas que tivemos e que me ajudaram a chegar neste momento. Muito obrigado a todos.

Resumo

A radiografia de tórax é um dos exames radiológicos mais acessíveis para triagem e diagnóstico de possíveis doenças no pulmão e no coração. Além disso, esse tipo de exame é utilizado para identificar se dispositivos como marca-passos, cateteres venosos e tubos estão posicionados corretamente. Nos últimos anos, muita atenção e muitos esforços foram dedicados para melhorar os sistemas de Diagnóstico Auxiliado por Computador, sendo a classificação de imagens médicas um dos problemas principais abordados. Técnicas de Aprendizado Profundo têm sido cada vez mais utilizadas para fornecer predições de detecção e classificação de patologias e lesões em imagens de radiografias de tórax.

Considerando essas informações, propomos um método para classificação de imagens de radiografia de tórax, denominado DuaLAnet, utilizando técnicas de aprendizado profundo, como redes neurais convolucionais e módulos de atenção. O nosso método tem como objetivo explorar a complementaridade entre redes neurais convolucionais e módulos de atenção para direcionar o aprendizado sobre as classes, mostrando que a combinação de informações complementares extraídas das imagens de radiografia de tórax possui uma melhor taxa de predição do que quando utilizamos somente uma rede neural.

Para validar o nosso método, utilizamos as bases de dados ChestX-ray14 [48] e CheXpert [18], que possuem uma grande variedade de imagens de radiografias de tórax com 14 classes cada uma. Realizamos experimentos para verificar a melhor forma de inicializar os pesos das redes neurais, considerando a inicialização a partir da ImageNet [9] e a partir da base de dados de radiografia que não está sendo utilizada no treinamento. Além disso, experimentamos quatro tipos de arquiteturas e suas variações para verificar quais as redes neurais que deveríamos utilizar como extratoras de características. Depois, verificamos qual o módulo de atenção que se adequava melhor a cada extratora de característica escolhida previamente, entre as seguintes opções de módulo de atenção: *Class Activation Mapping* (CAM), *Soft Activation Mapping* (SAM) e *Feature Pyramid Attention* (FPA).

Por fim, realizamos os experimentos com o método DuaLAnet, após as escolhas de configurações que melhor se adequavam em cada base de dados. Os resultados obtidos mostram que o nosso método possui uma taxa de acerto AUROC competitiva, em comparação com os métodos do estado da arte na base de dados ChestX-ray14 [48], e vários caminhos que podemos seguir para melhorar a taxa de acerto na base de dados CheXpert [18].

Abstract

Chest X-ray is one of the most accessible radiological exams for screening and diagnosing possible lung and heart diseases. In addition, this type of examination is used to identify whether devices such as pacemakers, venous catheters and tubes are correctly positioned. In recent years, much attention and efforts have been devoted to improving Computer Aided Diagnostic systems, with the classification of medical images being one of the main problems addressed. Deep Learning techniques have been increasingly used to provide predictions for the detection and classification of pathologies and lesions in chest X-ray images.

Considering this information, we propose a method to classify chest X-ray images, called DuaLAnet, using deep learning techniques, such as convolutional neural networks and attention mechanisms. Our method aims to explore the complementarity between convolutional neural networks and attention modules to guide the learning process regarding the distinct classes, showing that the combination of complementary information extracted from chest X-ray images has a better rate of prediction when compared with the case with only a neural network.

To validate our method, we use the ChestX-ray14 [48] and CheXpert [18] datasets, which have a wide variety of chest X-ray images with 14 classes each. We carried out experiments to verify the best way to initialize the weights of the neural networks, considering the initialization from ImageNet [9] and from the radiography dataset that is not being used in the training. In addition, we experimented with four types of architectures and their variations to check which neural networks we should use as feature extractors. Then, we checked which attention mechanism was best suited to each feature extractor chosen previously, from the following attention mechanisms options: Class Activation Mapping (CAM), Soft Activation Mapping (SAM), and Feature Pyramid Attention (FPA).

Finally, we carried out the experiments with the DuaLAnet method, after choosing the settings that best fit each dataset. The obtained results indicate that our method has a competitive AUROC score, compared to state-of-the-art methods in the ChestX-ray14 [48] dataset, and several ways we can follow to improve the hit rate in the base CheXpert [18] dataset.

Lista de Figuras

1.1 Radiografia de tórax utilizada para analisar o primeiro paciente com COVID-19 nos Estados Unidos. Fonte: `https://www.healthimaging.com/media/22185`. 13
1.2 Radiologista utilizando um sistema de Diagnóstico Auxiliado por Computador. Fonte: `https://curesmb.com/neurology/`. 14
1.3 Exemplos de radiografias da base de dados ChestX-ray14 [48]. 15

2.1 Exemplo do exame de Radiografia de Tórax. 17
2.2 Exemplos de radiografia frontal e lateral da base de dados CheXpert [18]. . 18
2.3 Recortes de uma imagem 3D de radiografia de tórax utilizados no diagnóstico de COVID-19. 19
2.4 Exemplo de rede neural. 20
2.5 Exemplo de rede neural convolucional. 21
2.6 Parte da arquitetura da rede DenseNet [17]. 22
2.7 Arquitetura da rede VGGNet-16 [35]. 23
2.8 Modelo da arquitetura de uma rede ResNet [15]. 23
2.9 Método de dimensionamento composto da EfficientNet [39]. 24
2.10 Exemplo de aumentação de dados. 26
2.11 Método de acúmulo de gradiente. 27
2.12 Método de transferência de aprendizado. 28

3.1 Distribuição do número de patologias por imagem na base ChestX-ray14 [48]. 30
3.2 Ilustração de uma radiografia para cada patologia existente na base de dados ChestX-ray14 [48] e uma radiografia sem patologias (normal). 31
3.3 Distribuição do número de patologias por imagem na base CheXpert [18]. . 32
3.4 Ilustração de uma radiografia para cada classe existente na base de dados CheXpert [18]. 33

5.1 Ilustração do método DuaLAnet. Duas redes neurais são utilizadas para aprender características complementares das imagens de entrada. As saídas do *pooling* global das redes neurais A e B são concatenadas e passadas como entrada para o classificador de fusão C_f. A saída final é baseada nos resultados dos dois classificadores auxiliares C_a e C_b, e no classificador de fusão C_f. 41
5.2 Ilustração dos classificadores utilizados no método DuaLAnet: (a) módulo inserido no final das redes neurais A e B; (b) módulo usado para a classificação do vetor com a concatenação do *pooling* global das redes neurais A e B. 42
5.3 Exemplo de aplicação do método de rotação horizontal aleatória. 44
5.4 Exemplo de aplicação do método de recorte centralizado. 45

5.5 Exemplo genérico de AUROC. 47
5.6 Exemplo de AUROC obtido com a DCNN ResNet-50 [48]. 47
5.7 Média de AUROC para o método DuaLAnet variando os valores de γ_1, γ_2 e γ_3. Os valores dos parâmetros são: Config1 ($\gamma_1 = 1.0, \gamma_2 = 0.5, \gamma_3 = 0.2$), Config2 ($\gamma_1 = 1.0, \gamma_2 = 0.5, \gamma_3 = 0.5$), Config3 ($\gamma_1 = 0.5, \gamma_2 = 1.0, \gamma_3 = 0.5$) e Config4 ($\gamma_1 = 0.5, \gamma_2 = 0.5, \gamma_3 = 1.0$). 50
5.8 Taxa de perda das etapas de treinamento e validação do método DuaLAnet (ramo de fusão). 51
5.9 Média de AUROC para o método DuaLAnet variando os valores de γ_1, γ_2 e γ_3. Os valores dos parâmetros são: Config1 ($\gamma_1 = 1.0, \gamma_2 = 0.7, \gamma_3 = 0.4$), Config2 ($\gamma_1 = 1.0, \gamma_2 = 0.5, \gamma_3 = 0.5$), Config3 ($\gamma_1 = 0.5, \gamma_2 = 0.5, \gamma_3 = 0.5$) e Config4 ($\gamma_1 = 0.3, \gamma_2 = 0.3, \gamma_3 = 0.5$). 54
5.10 Taxa de perda das etapas de treinamento e validação do método DuaLAnet (ramo de fusão). 54

Lista de Tabelas

3.1 Divisão da base de dados ChestX-ray14 [48]. 29
3.2 Divisão da base de dados CheXpert [18]. 32

4.1 Informações sobre os trabalhos relacionados à base de dados ChestX-ray14 [48]. As posições com o símbolo "-" indicam que a informação não foi apresentada no trabalho. 38
4.2 Informações sobre os trabalhos relacionados à base de dados CheXpert [18]. As posições com o símbolo "-" indicam que a informação não foi informada no trabalho. 40

5.1 Resultado do pré-treinamento. 48
5.2 Resultado das redes neurais utilizadas como extratoras de características. . 48
5.3 Comparação entre os resultados da DenseNet169 (D169) e a ResNet152 (R152) considerando três módulos de atenção diferentes (CAM, SAM e FPA). 49
5.4 Comparação do nosso método DuaLAnet com as abordagens do estado da arte na base de dados ChestX-ray14 [48]. 50
5.5 Resultado do pré-treinamento. 51
5.6 Resultado das redes neurais utilizadas como extratoras de características. . 52
5.7 Comparação entre os resultados da DenseNet169 (D169) e a EfficientNetB5 (Eff) considerando três módulos de atenção diferentes (CAM, SAM e FPA). 52
5.8 Comparação do nosso método DuaLAnet com as abordagens do estado da arte na base de dados CheXpert [18]. O método ConVIRT [50] reporta somente a média final da métrica AUROC para as 5 classes. 53

Sumário

1 Introdução **13**
1.1 Motivações . . . 14
1.1.1 Diagnóstico Auxiliado por Computador . . . 14
1.1.2 Classificação de Radiografias de Tórax . . . 15
1.2 Definição do Problema . . . 15
1.3 Objetivos . . . 16
1.4 Questões de Pesquisa . . . 16
1.5 Organização da Dissertação . . . 16

2 Conceitos **17**
2.1 Radiografia de Tórax . . . 17
2.2 Visão Computacional . . . 18
2.2.1 Classificação de Imagens . . . 19
2.2.2 Redes Neurais Convolucionais . . . 21
2.2.3 Mecanismo de Atenção . . . 24
2.2.4 Aumentação de Dados . . . 26
2.2.5 Acúmulo de Gradiente . . . 27
2.2.6 Transferência de Aprendizado . . . 27

3 Bases de Dados **29**
3.1 ChestX-ray14 . . . 29
3.2 CheXpert . . . 30

4 Trabalhos Relacionados **34**
4.1 ChestX-ray14 . . . 34
4.2 CheXpert . . . 38

5 DuaLAnet **41**
5.1 Construção da DuaLAnet . . . 41
5.1.1 Estratégia de Treinamento . . . 42
5.2 Experimentos . . . 44
5.2.1 Pré-Treinamento . . . 44
5.2.2 Aumentação de Dados . . . 44
5.2.3 Extratores de Características . . . 45
5.2.4 Detalhes de Implementação . . . 45
5.2.5 Métricas de Avaliação . . . 46
5.3 Resultados e Discussões . . . 47
5.3.1 ChestX-ray14 . . . 48
5.3.2 CheXpert . . . 51

6 Conclusões e Trabalhos Futuros **55**
6.1 Contribuições 56
6.2 Trabalhos Futuros 57

Referências Bibliográficas **58**

Capítulo 1

Introdução

Radiografia de tórax é um dos exames mais comuns globalmente, sendo fundamental para a identificação, diagnóstico e manejo de diversas doenças potencialmente fatais. Esse tipo de radiografia é utilizado para identificar condições cardiopulmonares agudas e crônicas, se dispositivos como marcapassos, cateteres venosos centrais e tubos estão posicionados corretamente, e para auxiliar em exames médicos relacionados.

O diagnóstico no estágio inicial de diversas doenças pode ser de grande ajuda para que o tratamento seja bem-sucedido. A radiografia de tórax pode ser utilizada para verificar a presença de pneumonia em pacientes, a qual, segundo a Organização Mundial da Saúde (OMS), é a quarta doença mais letal no Brasil e é uma das principais causas de mortes em idosos com mais de 65 anos[1].

Figura 1.1: Radiografia de tórax utilizada para analisar o primeiro paciente com COVID-19 nos Estados Unidos. Fonte: `https://www.healthimaging.com/media/22185`.

Esse tipo de exame tem também auxiliado no diagnóstico da COVID-19. A maioria dos pacientes que são diagnosticados com COVID-19 tem algumas características peculiares em suas radiografias, como opacidades multifocais e em vidro fosco em algumas regiões do pulmão [1]. Dessa forma, torna-se possível criar abordagens para detectar esse padrão em imagens de radiografias. A Figura 1.1 apresenta uma radiografia de tórax utilizada no diagnóstico do primeiro paciente com COVID-19 nos Estados Unidos[2].

[1] `https://redepara.com.br/Noticia/217040/pneumonia-a-doenca-silenciosa-que-mata`

[2] `https://www.healthimaging.com/topics/diagnostic-imaging/chest-x-ray-coronavirus-1st-us-patient`

1.1 Motivações

Nesta seção, apresentamos as motivações que nos instigaram a pesquisar e propor um método para classificação de imagens de radiografias de tórax. Na Subseção 1.1.1, mostramos o que é e como funciona um diagnóstico auxiliado por métodos computacionais, e na Subseção 1.1.2, apresentamos uma visão geral sobre classificação de radiografias de tórax.

1.1.1 Diagnóstico Auxiliado por Computador

Sistemas de Diagnóstico Auxiliado por Computador (*Computer-Aided Diagnosis*, ou CAD) são essenciais para auxiliar os médicos na especificação de tratamentos e planos cirúrgicos [52]. No entanto, projetar sistemas CAD para detecção e interpretação de radiografias de tórax é uma tarefa desafiadora, devido à grande variação de patologias, locais e detalhes. É importante ressaltar que esses sistemas CADs são utilizados somente como uma ferramenta para obtenção de informações adicionais, sendo o diagnóstico final sempre feito pelo médico especialista.

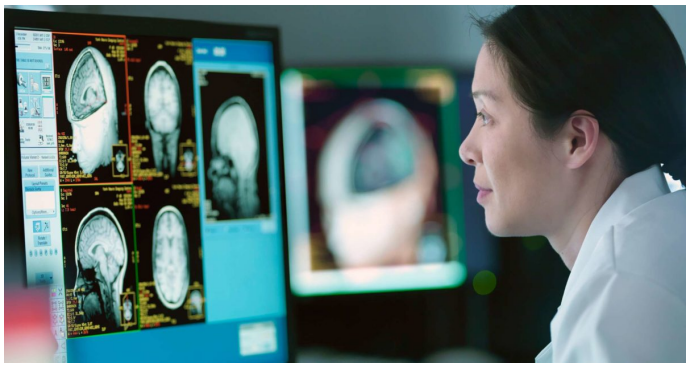


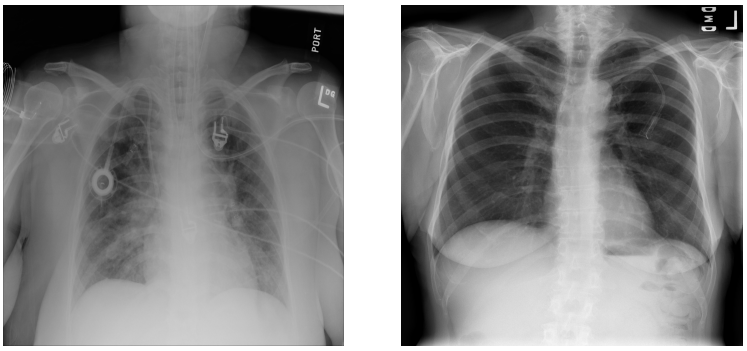
Figura 1.2: Radiologista utilizando um sistema de Diagnóstico Auxiliado por Computador. Fonte: `https://curesmb.com/neurology/`.

A principal finalidade do CAD é melhorar a acurácia do diagnóstico, assim como a consistência da interpretação da imagem radiológica. As informações dos sistemas CADs podem ser utilizadas para, por exemplo, mostrar uma possível patologia existente que o médico não conseguiu detectar por meio dos métodos tradicionais. Uma visualização de um sistema CAD é apresentada na Figura 1.2.

A interpretação automatizada da radiografia de tórax, no mesmo nível dos radiologistas, proporcionaria benefícios substanciais em diversos ambientes médicos, desde melhoria da decisão clínica e priorização do fluxo de trabalho até iniciativas de saúde global da população. A combinação de algoritmos cada vez mais sofisticados com conjuntos de dados resultou em vários avanços significativos em diversas áreas da Visão Computacional [14, 37]. Um aspecto essencial para os avanços na área de análise automatizada de radiografia de tórax é a disponibilidade de grandes bases de dados rotulados, como a CheXpert [18] e a ChestX-ray14 [48].

1.1.2 Classificação de Radiografias de Tórax

Métodos de Aprendizado Profundo têm sido bastante utilizados em tarefas no domínio de imagens médicas, como detecção e classificação de patologias em radiografias [48]. A maior parte das pesquisas em detecção e diagnóstico auxiliado por computador em radiografia de tórax foi dedicada à detecção de nódulos pulmonares. No entanto, os nódulos pulmonares são um achado relativamente raro em pulmões. Os achados mais comuns em radiografias de tórax incluem infiltrações pulmonares, cateteres e anormalidades no tamanho ou contorno do coração [2]. A Figura 1.3 apresenta dois exemplos de radiografias da base de dados ChestX-ray14 [48], uma com a presença das classes Atelectasia e Infiltração (Figura 1.3a) e a outra com a ausência de patologias (Figura 1.3b).



(a) Atelectasis e Infiltration. (b) Radiografia normal.

Figura 1.3: Exemplos de radiografias da base de dados ChestX-ray14 [48].

Os métodos existentes na literatura abordam, em sua maioria, problemas de classificação multirrótulo de imagens de radiografia de tórax [5, 12, 13, 46, 48]. Wang et al. [48] propuseram um método simples para classificação de imagens de radiografia de tórax, que consistia de redes neurais pré-treinadas com a base de dados ImageNet [9] e treinadas na base de dados ChestX-ray14 [48]. Métodos mais sofisticados foram sendo criados e elevaram a taxa de acerto na predição de radiografias de tórax, como o método de Guendel et al. [13] que considera informações espaciais e a alta resolução de imagens.

1.2 Definição do Problema

Dado um conjunto de imagens de radiografia de tórax e seus respectivos rótulos, com as possíveis patologias existentes, esse problema pode ser definido como a utilização de métodos de aprendizagem profunda, como redes neurais convolucionais, para classificar essas imagens e gerar vetores de rótulos que mais se aproximem das verdadeiras classes das imagens.

Além da classificação, pode ocorrer também a geração de mapas de calor das imagens de entrada, que são utilizados para indicar as possíveis localizações das patologias, o que é de grande importância em um sistema que ajuda no diagnóstico de eventuais doenças dos pacientes.

1.3 Objetivos

O objetivo geral deste trabalho é investigar técnicas recentes de classificação de imagens e propor uma abordagem baseada em Redes Neurais Profundas para classificação de radiografias de tórax. Para atingir esse propósito, alguns objetivos específicos devem ser alcançados:

- Propor um método de classificação de imagens de radiografias de tórax explorando a complementaridade entre as redes e técnicas de atenção.
- Utilizar técnicas de Aprendizado Profundo para suprir o desbalanceamento das bases de dados e a diferença de recursos computacionais para os trabalhos do estado da arte.
- Validar o método proposto em diferentes bases de dados.
- Comparar e avaliar o método proposto com outras abordagens disponíveis.
- Publicar os resultados.

1.4 Questões de Pesquisa

A partir dos objetivos propostos, definimos as seguintes questões de pesquisa para guiar o nosso trabalho, que são listadas a seguir.

- É possível obter resultados competitivos com os trabalhos recentes em classificação de imagens de radiografias de tórax utilizando redes neurais profundas?
- Como os pesos de redes pré-treinadas na base ImageNet e, em seguida, aperfeiçoados para o domínio médico, comparam-se aos pesos aprendidos somente a partir de imagens médicas?
- Como a complementaridade entre redes neurais profundas e técnicas de atenção podem contribuir na nossa abordagem?

1.5 Organização da Dissertação

Nosso trabalho está dividido da seguinte forma. No Capítulo 2, descrevemos os conceitos utilizados nesta pesquisa; no Capítulo 3, apresentamos as bases de dados utilizadas e algumas análises sobre os dados; no Capítulo 4, apresentamos os trabalhos relacionados, destacando os pontos positivos e negativos de cada um; no Capítulo 5, descrevemos o nosso método de classificação, bem como as técnicas auxiliares que utilizamos, as configurações dos experimentos e os resultados; por fim, no Capítulo 6, apresentamos as contribuições deste mestrado e possíveis caminhos a serem explorados relacionados a esse tema de pesquisa.

Capítulo 2

Conceitos

Neste capítulo, apresentamos os conceitos utilizadas no desenvolvimento deste trabalho. Na Seção 2.1, mostramos informações médicas relacionadas ao nosso trabalho, com detalhes sobre a coleta e o processamento das radiografias de tórax. Na Seção 2.2, apresentamos os métodos e as técnicas de Visão Computacional utilizadas na nossa abordagem, como as Redes Neurais Convolucionais, os Módulos de Atenção e a técnica de Aumentação de Dados.

2.1 Radiografia de Tórax

A Radiografia de Tórax é um dos exames de radiografia mais comumente utilizados. Uma radiografia de tórax é um teste médico não invasivo que ajuda os médicos a diagnosticar e tratar possíveis patologias existentes nos pacientes. A obtenção de imagens de radiografia envolve a exposição de uma parte do corpo, neste caso, o tórax, a uma pequena dose de radiação ionizante para produzir imagens do interior do corpo. As radiografias são a forma mais antiga e mais frequente usada de imagem médica [36].

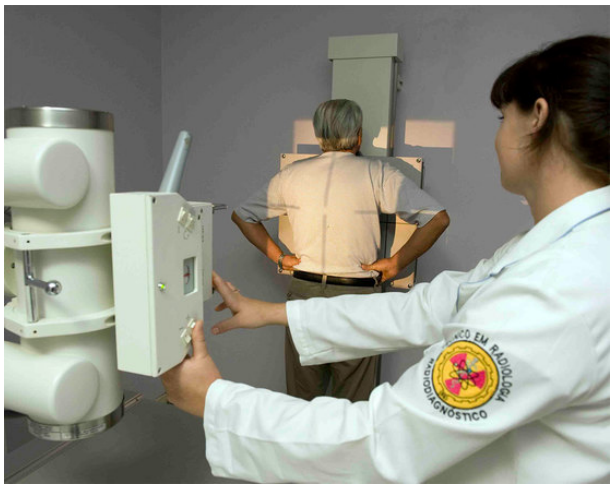
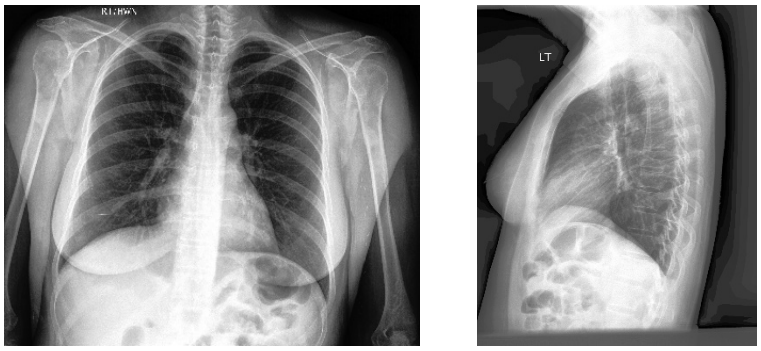


Figura 2.1: Exemplo do exame de Radiografia de Tórax.

Na medicina, as radiografias de tórax são utilizadas nas análises das condições dos

órgãos internos, busca por fraturas, tratamento de tumores, edemas, entre outras complicações. Esse tipo de exame tem a propriedade de atravessar materiais de baixa densidade, como o tecido muscular, e é atenuado por materiais de densidade mais elevada, como o cálcio, que está presente nos ossos. A Figura 2.1 apresenta um exemplo do exame de radiografia de tórax, em que o paciente fica a uma certa distância do equipamento de raio X, o qual é responsável por enviar radiação ionizante.

Normalmente, durante o exame, são capturadas imagens de vários planos e é solicitado que o paciente realize o procedimento em mais de uma posição. Dependendo do exame, é necessário que o paciente fique despido e coloque um avental. Dessa forma, o técnico de raio X, ou o médico, garante o posicionamento correto para o exame.



(a) Radiografia frontal. (b) Radiografia lateral.

Figura 2.2: Exemplos de radiografia frontal e lateral da base de dados CheXpert [18].

A radiografia física é criada a partir de uma tela de fósforo, em que a imagem é formada quando os raios X atingem a tela. A radiografia digital, que é a que utilizamos como entrada para o nosso método, é criada a partir do momento em que os raios X alcançam uma camada fotocondutora, produzindo carga positiva ou negativa. As cargas positivas são atraídas por um capacitor de carga que armazena a imagem latente, que então é lida e armazenada como uma imagem digital (formada por *pixels*). A Figura 2.2 apresenta exemplos de radiografias digitais de um mesmo paciente, do ponto de vista frontal, como mostrado na Figura 2.2a, e do ponto de vista lateral, como apresentado na Figura 2.2b.

2.2 Visão Computacional

Visão Computacional é uma subárea da Ciência da Computação que tem como objetivo replicar algumas características do sistema de visão humano e permitir que os computadores identifiquem e processem objetos em imagens e vídeos da mesma forma que os humanos. Para compreender um método de visão computacional, é necessário um conhecimento além do uso de geometria, física e teoria do aprendizado. A parte da "visão" depende de uma compreensão sólida das câmeras e do processo físico de formação das imagens, para obter inferências simples dos valores de *pixels* individuais, combinando os dados disponíveis em várias imagens, organizando em grupos para separá-los e inferir informações de forma, para, assim, reconhecer informações geométricas [10].

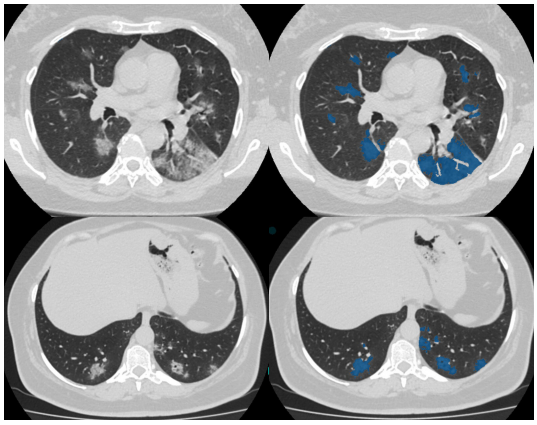


Figura 2.3: Recortes de uma imagem 3D de radiografia de tórax utilizados no diagnóstico de COVID-19. Fonte: `https://medicalxpress.com/news/2020-05-ct-scan-database-ai-covid-.html`

Uma das grandes vantagens da visão computacional é a extração de descrições do mundo a partir de imagens ou sequências de imagens. Por exemplo, uma técnica conhecida como estrutura de movimento, permite extrair uma representação do que é retratado e como a câmera se moveu a partir de uma série de imagens. Na indústria do entretenimento, as pessoas usam essa técnica para construir modelos tridimensionais (3D) de edifícios, e esses modelos são utilizados em testes que edifícios reais não podem ser submetidos, como testes de incêndios e explosões.

Uma outra aplicação importante de Visão Computacional é com imagens médicas, que nos permite fazer o uso extensivo de dados médicos para fornecer um melhor diagnóstico, tratamento e previsão de doenças [6]. A visão computacional pode explorar a textura, forma, contorno e conhecimento prévio junto com informações contextuais da sequência de imagens de entrada, fornecendo informações que ajudam no entendimento por parte dos humanos. A Figura 2.3 apresenta exemplos de recortes de uma imagem 3D de radiografia de tórax utilizados no diagnóstico de COVID-19 por meio de técnicas de visão computacional, como Classificação de Imagens (Subseção 2.2.1).

Uma subárea de Visão Computacional que tem avançado consideravelmente nos últimos anos explora técnicas de Aprendizado Profundo (*Deep Learning*), que permite que modelos computacionais de múltiplas camadas de processamento aprendam e representem os dados com diversos níveis de abstração [45]. Os métodos de aprendizado profundo representam o estado da arte em problemas desafiadores de visão computacional, como classificação de imagens, detecção de objetos e reconhecimento facial.

2.2.1 Classificação de Imagens

Quando uma pessoa olha para uma imagem, ela pode, normalmente, identificar o que ela representa com facilidade. Porém, para os computadores, essa não é uma tarefa fácil, pois as máquinas não possuem uma capacidade de associação tão desenvolvida quanto a dos seres humanos. Portanto, a tarefa de classificar uma imagem é um desafio para as máquinas e é nessa parte que entra o aprendizado.

O aprendizado pode ser realizado de duas formas: supervisionado e não supervisionado [32]. O aprendizado supervisionado é feito a partir de um conjunto de dados rotulados previamente definido e deseja-se encontrar uma função que seja capaz de predizer os rótulos de um conjunto de dados desconhecidos, mas semelhantes aos dados de entrada. A predição de rótulos pode ser feita a partir da regressão, que é quando estimamos valores reais, e a partir da classificação, que é quando existe um conjunto finito de rótulos. O aprendizado não supervisionado é utilizado quando o conjunto de dados não possui nenhum tipo de rótulo e tem como objetivo descobrir a similaridade entre os objetos analisados. Neste trabalho, utilizamos o aprendizado supervisionado e a técnica de classificação de imagens.

O reconhecimento e a classificação de imagens é o que possibilita muitas das realizações mais impressionantes da inteligência artificial. A classificação de imagens refere-se à tarefa de extrair classes de informações de uma imagem. Assim, dado um conjunto de imagens rotulados, o modelo computacional deve ser capaz de predizer os rótulos de um conjunto de imagens de teste e medir a taxa de acerto dessas predições.

A classificação de imagens pode ser dividida em três etapas. O pré-processamento, que tem como objetivo melhorar os dados das imagens, aprimorando alguns recursos importantes para que os modelos de classificação possam se beneficiar desses dados aprimorados. A etapa de pré-processamento inclui a leitura e o redimensionamento das imagens. Depois, temos a etapa de extração de características e treinamento, que é uma etapa crucial em que métodos de aprendizado são usados para identificar os padrões de imagens, informações que podem ser exclusivos de uma classe em particular e que irão, mais tarde, ajudar o modelo a diferenciar entre classes diferentes. Finalmente, a classificação, que categoriza as informações extraídas em classes predefinidas [26].

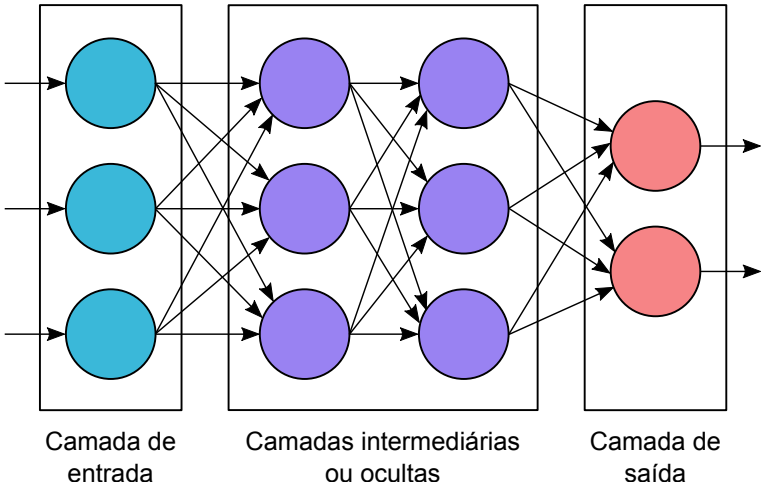


Figura 2.4: Exemplo de rede neural.

Uma abordagem bastante utilizada na tarefa de classificação de imagens é a de Redes Neurais (*Neural Networks*), que funciona de forma semelhante à rede neural do cérebro humano. Um "neurônio" em uma rede neural é uma função matemática que coleta e classifica informações de acordo com uma arquitetura predeterminada. A rede neural tem uma grande semelhança com métodos estatísticos, como ajuste de curva de acordo com o

padrão dos dados [47].

As redes neurais possuem muitos neurônios artificiais que são interconectados por nós, e esses neurônios são compostos por unidades de entrada e saída, como apresentado na Figura 2.4. As unidades de entrada recebem várias formas e estruturas de informação, e a rede neural tenta aprender sobre as informações recebidas para produzir uma informação de saída. As redes neurais podem ser de diversos tipos, como Redes Neurais Convolucionais e Redes Neurais Recorrentes, por exemplo.

2.2.2 Redes Neurais Convolucionais

Redes Neurais Convolucionais (*Convolutional Neural Networks*, ou CNNs) são uma família de modelos que foram inspirados em como o córtex visual do cérebro humano funciona ao reconhecer objetos. Primeiramente, para um bom desempenho desses algoritmos, é necessário haver uma extração satisfatória de características dos objetos de entrada. As redes neurais são capazes de aprender automaticamente o padrão de dados brutos que são mais úteis para uma determinada tarefa [29].

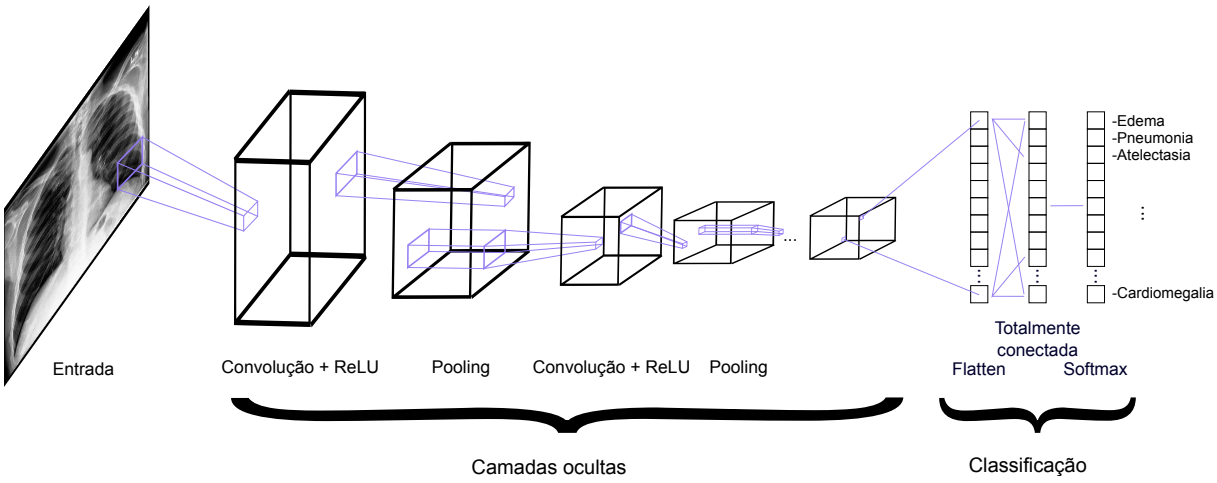


Figura 2.5: Exemplo de rede neural convolucional.

As Redes Neurais Multicamadas e, em particular, as Redes Neurais Convolucionais Profundas (*Deep Convolutional Neural Network*, ou DCNN), constroem uma hierarquia de camadas combinando dados obtidos em níveis iniciais para formar informações mais relevantes nas camadas dos níveis intermediários e finais. Por exemplo, a Figura 2.5 apresenta um exemplo de funcionamento de uma DCNN, onde nas camadas iniciais são extraídas informações mais primitivas, como bordas, cantos e ângulos, que são combinadas para formar características de alto nível, como uma possível patologia existente na imagem de entrada.

As DCNNs, em geral, funcionam muito bem em tarefas relacionadas a imagens e isso se deve, em grande parte, a duas ideias importantes: (i) conectividade esparsa, onde um único elemento no mapa de características é conectado a apenas um pequeno conjunto de *pixels*, e (ii) compartilhamento de parâmetros, em que os mesmos pesos são usados para diferentes partes da imagem de entrada. Normalmente, as DCNNs são formadas por

várias camadas convolucionais (*conv*) e de subamostragem (*pooling*), que são seguidas por uma ou mais camadas totalmente conectadas (*fully connected*) no final.

A seguir, mostramos alguns detalhes de quatro arquiteturas de DCNNs que utilizamos como extratoras de características no nosso método.

DenseNet

Huang et al. [17] propuseram a DCNN DenseNet que tem como característica marcante a garantia do fluxo máximo de informações entre as camadas da rede, ligando todas as camadas diretamente com as demais. Para preservar o *feed-forward*, cada camada obtém entradas adicionais de todas as camadas anteriores e passa seus mapas de características para todas as camadas subsequentes. A Figura 2.6 ilustra esse modelo esquematicamente, onde só são utilizados os recursos que já foram calculados.

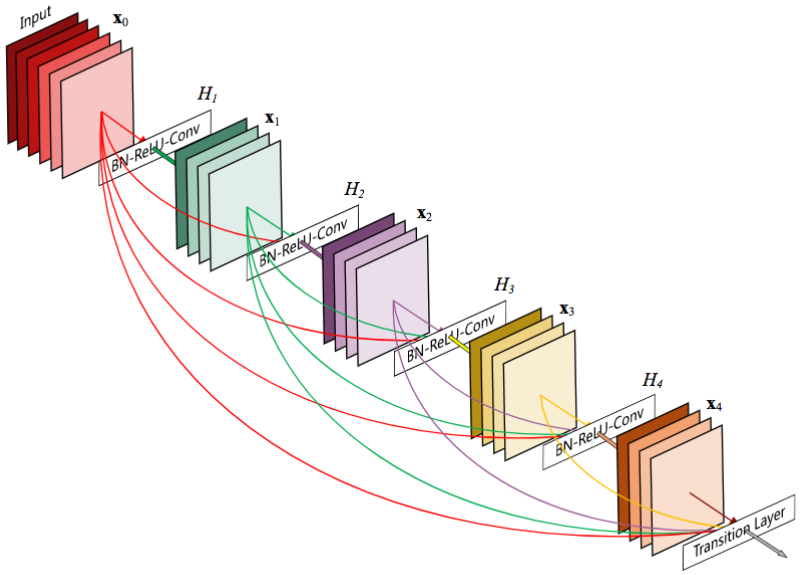


Figura 2.6: Parte da arquitetura da rede DenseNet [17].

As DenseNets possuem algumas vantagens interessantes, como uma possível solução para o problema de desaparecimento de gradiente (*vanishing gradient problem*), a propagação de informações ao longo da rede neural, reuso de recursos e a redução do número de parâmetros [17].

VGGNet

Simonyan e Zisserman [35] propuseram a DCNN VGGNet16 e ganharam a competição de localização da ImageNet [9] em 2014. Essa rede é um aprimoramento em relação à rede AlexNet [21], substituindo os filtros de tamanho 5×5 e 11×11, por vários filtros de tamanho 3×3, um após o outro. Os *kernels* de tamanhos menores empilhados são melhores do que um *kernel* maior, pois várias camadas não lineares aumentam a profundidade da

rede. Com isso, é possível a rede aprender características mais complexas e também ter um custo menor.

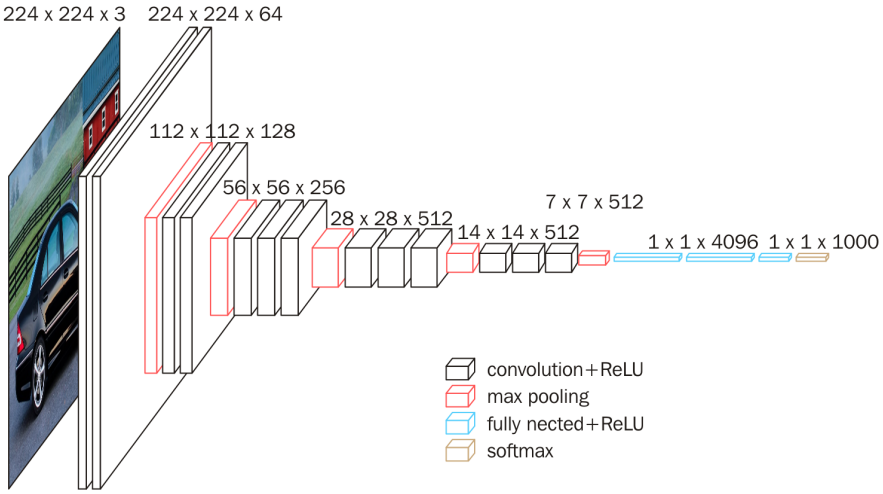


Figura 2.7: Arquitetura da rede VGGNet-16 [35].

A DCNN VGGNet16 possui a característica de ser uma boa extratora de características das imagens, por conta da sua profundidade e dos filtros. Além disso, o aumento de camadas com filtros menores resultou em um aumento da não linearidade, o que é considerado positivo para redes de aprendizado profundo. Porém, essa DCNN possui uma maior chance de apresentar o problema de desaparecimento de gradiente.

ResNet

He et al. [15] propuseram a DCNN ResNet e ganharam a competição de classificação, localização e detecção da ImageNet [9] em 2015. Diferentemente das redes apresentadas anteriormente, a ResNet busca aprender a partir dos resíduos obtidos pelas camadas anteriores, como mostra a Figura 2.8. As redes residuais buscam resolver dois principais problemas que acontecem nas redes profundas: (i) o custo do treinamento à medida que a rede fica mais profunda e (ii) a saturação da precisão, também por conta da profundidade da rede.

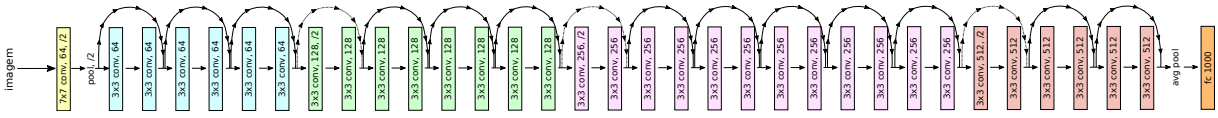


Figura 2.8: Modelo da arquitetura de uma rede ResNet [15].

Considerando a arquitetura da ResNet, os gradientes podem fluir diretamente através das conexões que vão para as camadas posteriores, por isso, o problema de desaparecimento de gradiente não é uma dificuldade nesse tipo de arquitetura. Essa arquitetura possui algumas variações de tamanho, que influencia diretamente no tempo e recursos disponíveis para o treinamento.

EfficientNet

Tan e Le [39] propuseram uma arquitetura de rede neural convolucional, denominada EfficientNet, que trabalha com um método de dimensionamento uniforme de todas as dimensões dos dados de entrada, como profundidade, largura e resolução, utilizando um coeficiente composto, como mostrado na Figura 2.9. O método de dimensionamento composto é justificado pela intuição de que, se a imagem de entrada for maior do que o previsto, a rede neural precisa de mais camadas para aumentar a área que será considerada e mais canais para capturar padrões mais refinados na imagem maior.

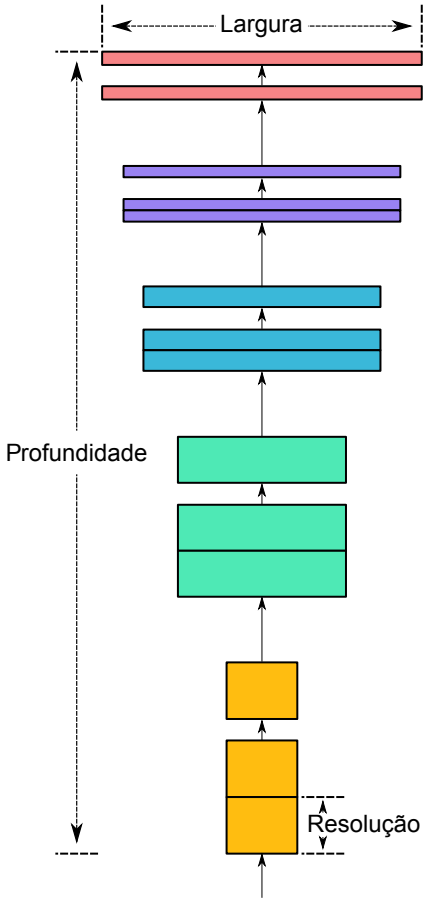


Figura 2.9: Método de dimensionamento composto da EfficientNet [39].

Tan e Le [39] também justificaram que, conforme a resolução das imagens de entrada aumenta, a profundidade e largura da rede neural também devem ser aumentadas. Conforme a profundidade é aumentada, pode-se capturar informações que incluem mais *pixels*. Além disso, conforme a largura é aumentada, recursos mais refinados são capturados.

2.2.3 Mecanismo de Atenção

Quando pensamos na palavra "atenção", imaginamos que tem relação com o foco em algo ou se concentrar mais em algo. A ideia de "atenção" veio da área de processamento de linguagem natural, com o método *seq2seq*, em que os modelos são treinados para receber uma sequência de tamanho arbitrária de entrada e gerar outra sequência, também de

tamanho arbitrário, como saída. Um exemplo da utilização desse método é na tarefa de tradução de textos. Nesse tipo de problema, geralmente há uma dependência que vai além da palavra que está sendo processada no momento e essa dependência pode variar entre as sequências de entrada.

Dado esse contexto, surge a necessidade de aprender um mecanismo de dependência que se adapte de acordo com a sequência de entrada, descobrindo quais elementos da sequência são mais importantes para construir a saída de forma precisa. Esse tipo de dependência também é importante em tarefas de Visão Computacional, onde o domínio do tempo dá lugar ao domínio espacial.

O mecanismo de atenção, na área de Aprendizado Profundo, é baseado no conceito de direcionar o foco da rede neural a certos fatores ou características ao processar os dados de entrada [44]. O módulo de atenção é um componente da arquitetura de uma rede que é responsável por gerenciar a interdependência entre os elementos de entrada e saída, denominada atenção geral, ou dentro dos elementos de entrada, chamada autoatenção.

A seguir, apresentamos a definição e alguns detalhes de três mecanismos de atenção que utilizamos nos nossos experimentos.

Class Activation Mapping

O *Class Activation Mapping*, ou CAM, é uma técnica para obter as regiões discriminativas das imagens de entrada de uma DCNN, considerando a identificação de uma classe em específico. Em outras palavras, o módulo de atenção CAM, proposto por Zhou et al. [51], permite detectar quais regiões na imagem são relevantes para cada classe.

Para utilizar um módulo de atenção CAM na rede neural, é necessário ter uma camada de Agrupamento Médio Global (*Global Average Pooling*, ou GAP) após a última camada convolucional e, em seguida, uma camada linear. Para a utilização desse método, é necessário, muitas das vezes, realizar uma adaptação na DCNN. Uma característica que podemos destacar desse método é que o módulo de atenção CAM é treinado de forma fracamente supervisionada, o que significa que os objetos não precisam ser rotulados manualmente, a localização é realizada de maneira automática.

Soft Activation Mapping

O *Soft Activation Mapping*, ou SAM, é um método utilizado para extrair informações relevantes de imagens que possuem regiões discriminativas mais locais ou específicas. Lei et al. [23] propuseram esse módulo de atenção com o objetivo de classificar nódulos pulmonares em imagens de radiografia de tórax, que é considerada uma patologia difícil de ser classificada [22].

Como o método SAM busca detectar regiões específicas que são relevantes para a classificação da imagem, é utilizado a camada de Agrupamento Médio (*AVG pooling*), ao invés do GAP. Dessa forma, pode ser calculada, separadamente, a relevância de cada área da imagem de entrada.

Feature Pyramid Attention

O *Feature Pyramid Attention*, ou FPA, é um método de atenção que extrai características de diferentes escalas de informações e aumenta a variedade, no nível de *pixels*, das áreas consideradas para obter as regiões discriminativas da imagem de entrada. O método proposto por Li et al. [24] possui camadas convolucionais 3×3, 5×5 e 7×7, que são responsáveis por uma melhor extração do contexto.

O módulo de atenção FPA utiliza informações de contexto em diferentes escalas, que contribui para uma melhor geração de mapas de atenção no nível de *pixels*. Um diferencial desse módulo é que as informações de contexto são multiplicadas com o mapa de características original *pixel* a *pixel*, o que não necessita de muito recurso computacional.

2.2.4 Aumentação de Dados

Aumentação de Dados (*Data Augmentation*) é um método utilizado para mitigar o problema de falta de dados, principalmente na etapa de treinamento. Por exemplo, para o caso de imagens, a ideia principal dessa técnica é aplicar transformações nas imagens originais para que se possa ter diversas imagens resultantes com características semelhantes as das originais. Técnicas convencionais de aumentação de dados utilizam transformações como rotação, translação, escala e transformações de intensidade em amostras existentes [27]. A Figura 2.10 apresenta um exemplo de aplicação dessa técnica em uma imagem utilizando transformações como rotação e translação.

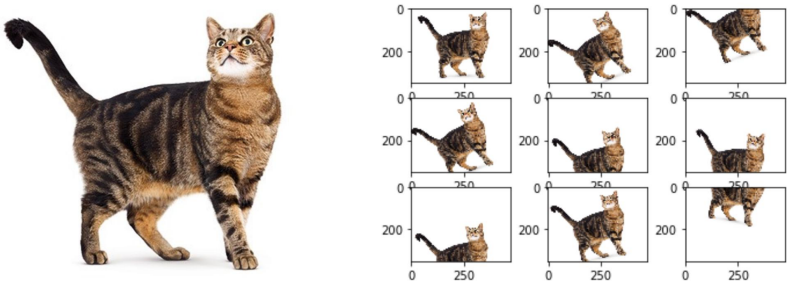


Figura 2.10: Exemplo de aumentação de dados.

A técnica de aumentação de dados é comumente utilizada como uma possível solução para o problema de subajuste/sobreajuste (*underfitting*/*overfitting*) no treinamento de uma rede neural. Esse problema causado durante o treinamento é acarretado pelo desbalanceamento de classes e/ou número reduzido de amostras de treinamento.

Temos que ter bastante cuidado ao usar a aumentação de dados, pois podem ocorrer transformações que não preservam os rótulos dos dados de entrada, prejudicando o treinamento da rede neural. Um exemplo disso, apresentado por Shorten e Khoshgoftaar [34], é a aplicação de transformações de rotação e espelhamento em bases de dados de dígitos numéricos, podendo causar confusões entre o 6 e o 9.

2.2.5 Acúmulo de Gradiente

As redes neurais convolucionais profundas são formadas por muitas camadas, conectadas umas com as outras, em que os dados de entradas são processados e as informações extraídas são passadas adiante pelas camadas. No final da rede, depois de se propagar por todas as camadas, é gerado previsões para as amostras e, em seguida, é calculado o valor de perda para cada dado de entrada, que determina o quão errada estava a rede neural para aquela determinada amostras. Após isso, os gradientes desses valores de perda são calculados e usados para a atualizar as variáveis da rede neural.

O acúmulo de gradiente significa executar um número predefinido de etapas sem atualizar as variáveis do modelo, ou seja, enquanto acumula os gradientes dessas etapas, para, em seguida, usar esses gradientes acumulados para calcular as atualizações das variáveis "treináveis" da rede neural. A técnica de acúmulo de gradiente é utilizada para dividir o *batch* de amostras, que é usado no treinamento da rede neural, em vários *mini-batches* que serão executados sequencialmente, como mostrado na Figura 2.11. Acumular os gradientes em todas as etapas é o mesmo que a soma dos gradientes como se fosse usado um tamanho maior de *batch*.

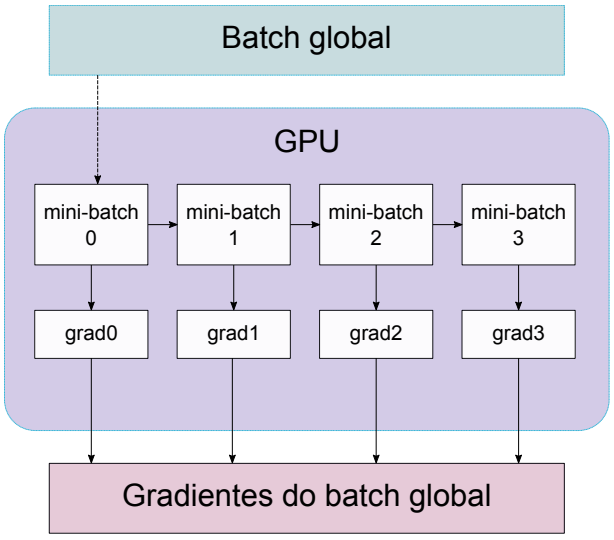


Figura 2.11: Método de acúmulo de gradiente.

2.2.6 Transferência de Aprendizado

Transferência de Aprendizado (*Transfer Learning*) é um método de Aprendizado de Máquina utilizado para o reuso de modelos (pesos) treinados para uma determinada tarefa em modelos com foco em resolver outros problemas [11]. Essa é uma abordagem muito popular em Aprendizado Profundo, em que modelos pré-treinados são utilizados como pontos de partida em problemas de Classificação de Imagens ou Processamento de Linguagem Natural, por exemplo.

Quando consideramos um problema de classificação em que não temos uma grande quantidade de dados rotulados, por exemplo, podemos aplicar a técnica de transferência

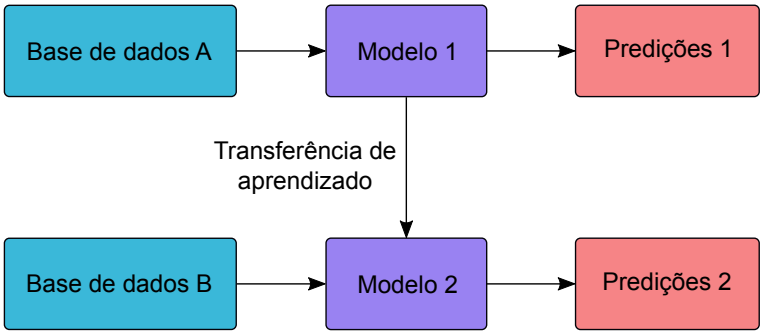


Figura 2.12: Método de transferência de aprendizado.

de aprendizado utilizando uma base de dados com características similares. Dessa forma, treinamos o modelo com a base de dados similar e, após isso, utilizamos os pesos obtidos como inicialização para a rede neural que será utilizada no treinamento da base de dados que possui poucos dados rotulados, como apresentado na Figura 2.12.

Capítulo 3

Bases de Dados

Neste capítulo, apresentamos as duas bases de dados que usamos como entrada para os nossos experimentos. Mostramos os detalhes e algumas amostras da base de dados ChestX-ray14 [48] (Seção 3.1) e da base de dados CheXpert [18] (Seção 3.2).

3.1 ChestX-ray14

A base de dados ChestX-ray14, proposta por Wang et al. [48], contém 112120 imagens frontais de radiografias de tórax de 30805 pacientes diferentes, que estão associadas à presença ou ausência de 14 patologias. Dividimos a base de dado da seguinte forma: 70% para treinamento, 10% para validação e 20% para teste. Vale ressaltar que não há imagens de um mesmo paciente em diferentes conjuntos, o que poderia enviesar os experimentos e, assim, tornar os resultados inválidos para comparações com as abordagens do estado da arte.

Tabela 3.1: Divisão da base de dados ChestX-ray14 [48].

Classe	**Treinamento**	**Validação**	**Teste**
Atelectasis	7324	956	3255
Cardiomegaly	1468	239	1065
Effusion	7631	1028	4648
Infiltration	11783	1999	6088
Mass	3569	465	1712
Nodule	4134	574	1615
Pneumonia	770	106	477
Pneumothorax	2272	365	2661
Consolidation	2524	328	1815
Edema	1237	141	925
Emphysema	1252	171	1093
Fibrosis	1185	66	435
Pleural Thickening	1973	269	1143
Hernia	122	19	86
Normal	46376	4124	9912

Mostramos na Tabela 3.1 a divisão da base de dados ChestX-ray14 [48]. Podemos destacar que essa base possui um nível considerável de desbalanceamento, por isso, aplicaremos algumas técnicas para contornar esse problema, mostradas na Subseção 5.2.2. A classe com maior número de amostras é a "Infiltration" que possui 19870 amostras, em torno de 17% da base de dados. Já a classe com menor número de amostras é a "Hernia", com apenas 227 imagens, que representa somente 0,2% da base de dados. As imagens podem ser rotuladas com mais de uma patologia, ou seja, trata-se de um problema de classificação multirrótulo. Uma informação importante sobre essa base de dados é que cerca de 53% das amostras são rotuladas com nenhuma das patologias, como mostrada na última linha da tabela.

Na Figura 3.1, apresentamos uma distribuição em ordem decrescente pelo número de imagens com mais de uma patologia. As amostras presentes nos conjuntos em azul possuem a patologia de sua respectiva coluna, além de pelo menos uma outra patologia. Apesar do gráfico dar a impressão de uma quantidade maior de imagens rotuladas com mais uma classe, isso não acontece realmente na base de dados, pois a quantidade de imagens rotuladas com mais de uma classe é 20735, que representa 18,5% da base, enquanto a quantidade de imagens rotuladas com somente uma classe é 30973, que representa 27,6% da base de dados.

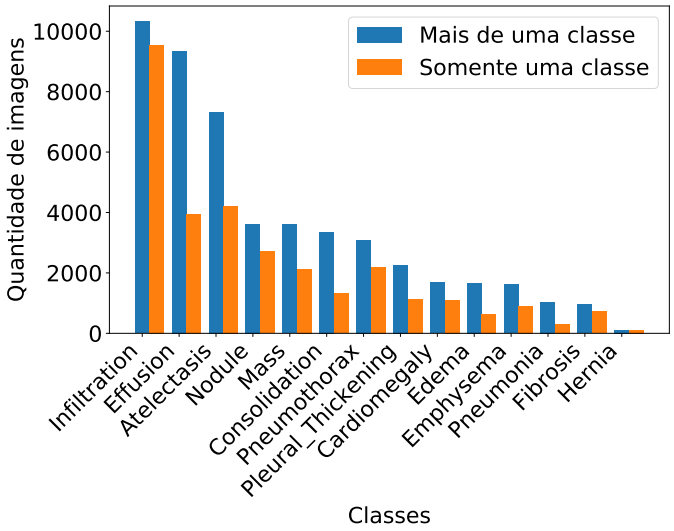


Figura 3.1: Distribuição do número de patologias por imagem na base ChestX-ray14 [48].

Para ilustrar as classes presentes na base de dados ChestX-ray14 [48], mostramos uma amostra de cada classe presente nessa base na Figura 3.2. Todas as imagens da base são em escala de cinza e as classes podem ser inferidas como positivas ou negativas. Apresentamos também um exemplo de imagem que foi rotulada com nenhuma das patologias (normal).

3.2 CheXpert

A base de dados CheXpert, proposta por Irvin et al. [18], foi disponibilizada para um desafio[1] público de Aprendizado de Máquina, consistindo em 224316 imagens de 65240

[1] `https://stanfordmlgroup.github.io/competitions/chexpert/`

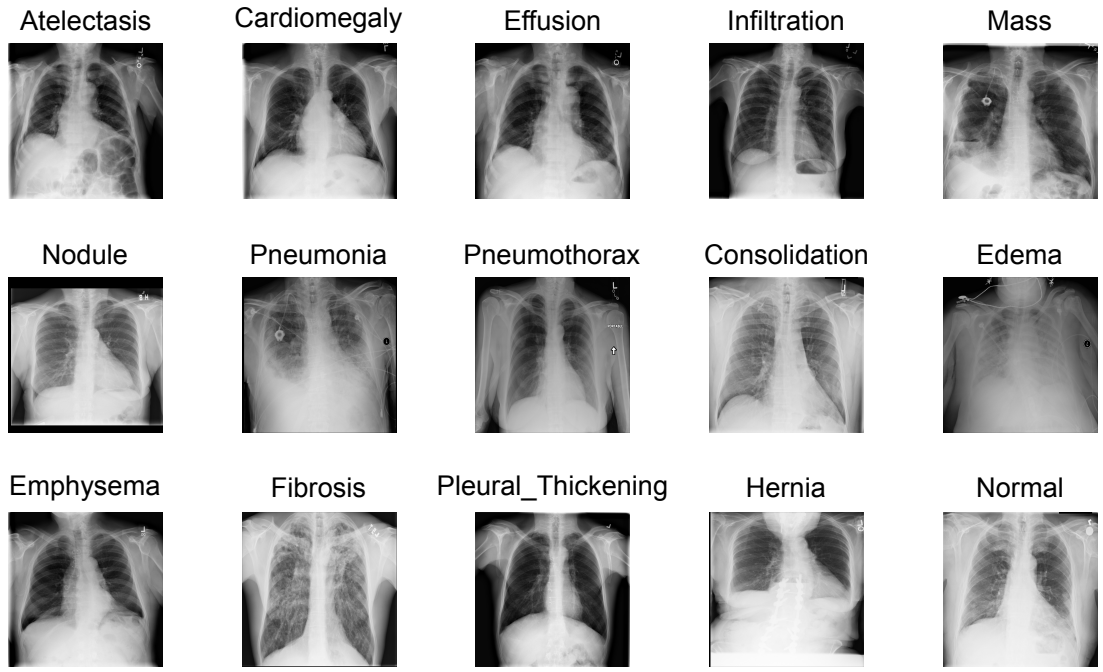


Figura 3.2: Ilustração de uma radiografia para cada patologia existente na base de dados ChestX-ray14 [48] e uma radiografia sem patologias (normal).

pacientes. A base foi coletada no *Stanford Hospital* entre outubro de 2002 e julho de 2017, tanto nos centros de internação quanto nos ambulatórios, juntamente com os relatórios radiológicos associados. Como o desafio está em andamento, estão disponíveis somente os conjuntos de treinamento e validação, que possuem 223414 e 234 imagens, respectivamente. As classes de imagens são divididas em 14 tipos de observações, como mostrado na Tabela 3.2.

Na competição, somente 5 classes são consideradas na avaliação dos resultados: Cardiomegaly, Edema, Consolidation, Atelectasis e Pleural Effusion. Considerando isso, muitos dos métodos propostos para essa base de dados só consideram essas classes em seus resultados [4, 31, 50]. Para fins de comparação, reportamos os resultados somente para as classes citadas anteriormente.

Cada classe pode ser inferida como positiva, incerta ou negativa. Não consideramos no nosso trabalho a forma como abordar essa incerteza sobre os rótulos, então realizamos um pré-processamento e consideramos todos os rótulos incertos como negativos, assim como uma das abordagens propostas por Irvin et al. [18]. Após esse pré-processamento, dividimos a base da mesma forma que a base de dados ChestX-ray14 [48]: 70% para treinamento, 10% para validação e 20% para teste. Consideramos também que as imagens de um mesmo paciente não pode aparecer em diferentes conjuntos. A classe com maior número de amostras é a "Support Devices" com 116108 imagens, e a classe com menos amostra é a "Pneumonia" com 6047 imagens.

Na Figura 3.4, apresentamos a distribuição do número de rótulos por imagem na base de dados CheXpert [18]. Podemos destacar que a quantidade de imagens com mais de um rótulo (159642) é consideravelmente maior do que a quantidade de imagens com somente um rótulo (52856). Nessa base, somente 11% das imagens são rotuladas com nenhuma das classes.

Tabela 3.2: Divisão da base de dados CheXpert [18].

Classe	Treinamento	Validação	Teste
No Finding	17061	1228	4130
Enlarged Cardiomediastinum	7603	1038	2266
Cardiomegaly	19460	2691	4917
Lung Opacity	70918	12533	22256
Lung Lesion	6892	609	1686
Edema	33715	7057	11519
Consolidation	10354	1705	2757
Pneumonia	4520	501	1026
Atelectasis	22023	3797	7636
Pneumothorax	13919	2191	3346
Pleural Effusion	59313	10507	16434
Pleural Other	2928	168	428
Fracture	6436	746	1858
Support Devices	76713	14859	24536

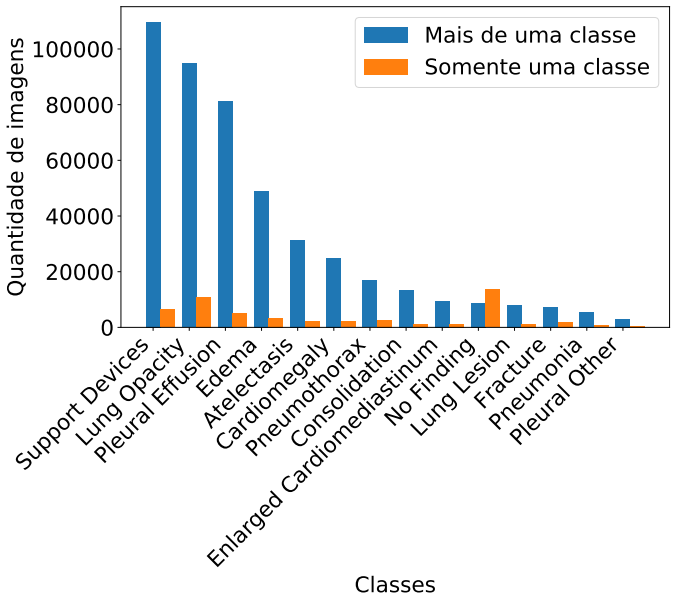


Figura 3.3: Distribuição do número de patologias por imagem na base CheXpert [18].

Ilustramos as classes presentes na base de dados CheXpert [18] com a Figura 3.4, a qual mostra uma radiografia para cada classe. Assim como as imagens da base de dados apresentada anteriormente, essa base também é composta por imagens em escala de cinza.

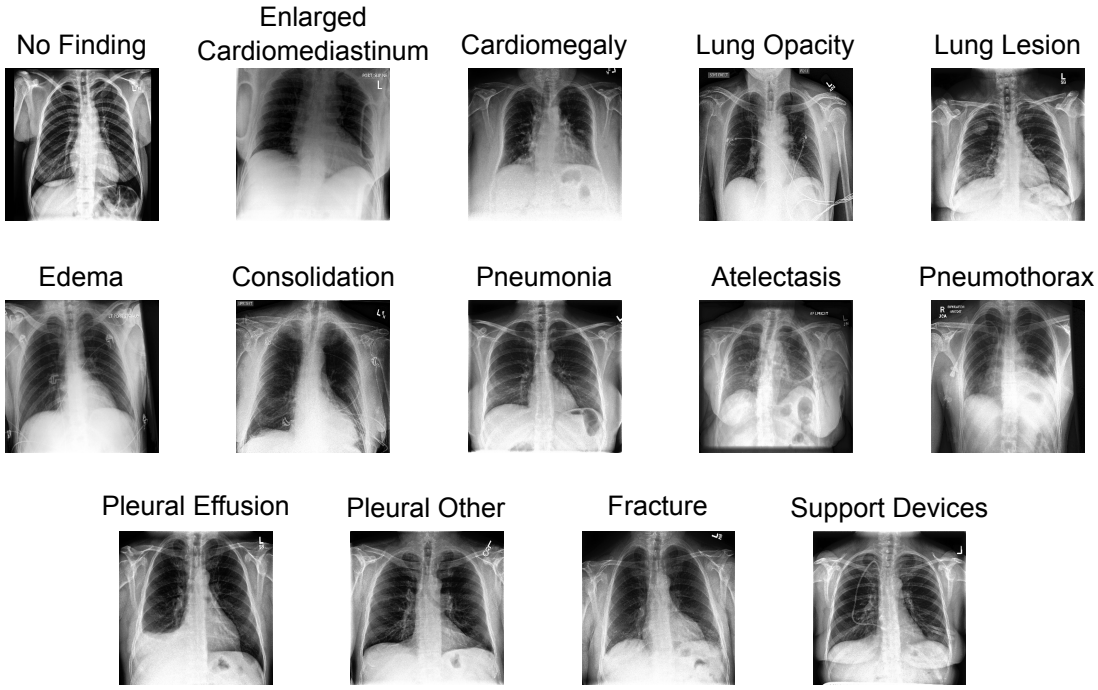


Figura 3.4: Ilustração de uma radiografia para cada classe existente na base de dados CheXpert [18].

Capítulo 4

Trabalhos Relacionados

Neste capítulo, descrevemos os métodos e detalhes de alguns artigos relacionados com o nosso tema de pesquisa, que é classificação de radiografias de tórax utilizando redes neurais profundas. Primeiramente, apresentamos trabalhos que utilizam a base de dados ChestX-ray14 [48] (Seção 4.1) e, em seguida, trabalhos que utilizam a base de dados CheXpert [18] (Seção 4.2).

4.1 ChestX-ray14

Wang et al. [48] propuseram a base de dados ChestX-ray14, com 112120 imagens frontais de radiografias de tórax de 30805 pacientes diferentes, que contém a presença ou ausência de 14 patologias. Os dados utilizados para a criação dessa base de dados foram coletados entre os anos de 1992 e 2015. Os rótulos das imagens foram criados a partir dos textos feitos por radiologistas e utilizando técnicas de Processamento de Linguagem Natural (Natural Language Processing, ou NLP) [7].

O método utilizado por Wang et al. [48] para classificar a base de dados foi bem simples. Eles utilizaram redes neurais multirrótulo pré-treinadas com a ImageNet [9] e treinadas com a base ChestX-ray14. Como a base possui uma quantidade maior de rótulos "negativos" do que "positivos", assim, eles propuseram um fator de balanceamento na função de perda W-CEL para forçar o aprendizado de amostras positivas.

Nessa abordagem, o *pooling* global e a camada de predição são usados não somente como parte da rede neural para classificação, mas também para a geração dos mapas de calor das patologias. O local com um pico no mapa de calor, geralmente, corresponde à presença de um padrão associado com alta probabilidade a uma classe. Após a classificação das imagens, são geradas as áreas delimitadoras com base nos mapas de calor.

Wang et al. [48] realizaram experimentos com as seguintes redes neurais: AlexNet [21], GoogLeNet [38], VGGNet-16 [35] e ResNet-50 [15]. A rede que obteve melhor resultado foi a ResNet-50, alcançando 0.745 de AUROC, estabelecendo o primeiro resultado quantitativo da base de dados ChestX-ray14.

Shen e Gao [33] propuseram um método para detecção de patologias em imagens de radiografias de tórax utilizando redes em cápsulas [43]. As principais características desse trabalho são baseadas nos tipos de conexões entre as camadas. Os autores introduziram

conectividades densas com roteamento dinâmico, que são obtidas por uma camada convolucional 1×1 que leva todos os mapas da camada anterior como entrada para a próxima camada.

Outra característica do trabalho de Shen e Gao [33] é a aplicação de técnicas para reduzir o tempo de treinamento e inferência. Os mapas de características precisam ser calculados apenas uma vez por camada e o coeficiente de roteamento é definido para ser treinável apenas na última iteração.

Os autores utilizaram também o módulo de atenção *Gradient-weighted Class Activation Mapping* (Grad-CAM) [30] na localização das patologias. Os mapas de calor gerados preservam a interpretabilidade do modelo como mapeamento de ativação de classes sem perder precisão de classificação [33].

Com a utilização das conexões com roteamento dinâmico entre os blocos densos da rede neural DenseNet-121 [17], Shen e Gao [33] obtiveram a taxa de 0.775 de AUROC, atingindo o estado da arte da base de dados ChestX-ray14 até então.

Wang et al. [46] propuseram o método Thorax-Net para a classificação de 14 tipos de patologias em radiografias de tórax. Esse método é composto por um ramo de classificação, que aprende representações de imagens, e um ramo de atenção, que permite aprender mapas de características discriminativos para melhorar a qualidade da classificação. Os autores utilizaram a rede neural ResNet-152 [15], como ramo de classificação, e o módulo Grad-CAM [30] incorporado em convoluções empilhadas. As saídas de ambos os ramos são utilizadas para geração da classificação de cada entrada.

As principais contribuições de Wang et al. [46] são: incorporar o Grad-CAM em operações convolucionais para formar um ramo de atenção que é capaz de converter os mapas de características aprendidos pelo ramo de classificação em um mapa de atenção que destaca as localizações de regiões específicas da patologias sob a supervisão dos rótulos no nível de imagem; e usar um ramo de classificação e um ramo de atenção em conjunto para alcançar uma taxa de classificação elevada.

Para adaptar a rede neural ResNet-152 [15] para a aplicação de classificação de imagens de radiografias de tórax, foi substituída a camada *softmax* por uma camada totalmente conectada com 14 neurônios, e aplicada a função Sigmoid [49] após essa camada. Vale destacar que a rede foi pré-treinada com a base de dados ImageNet [9].

Para a realização dos experimentos, Wang et al. [46] reduziram o tamanho das imagens de 1024×1024 para 256×256 *pixels* e recortaram essas imagens de forma aleatória, como um método de aumentação de dados, em dimensões de 224×224 *pixels*. Devido a essa manipulação das dimensões das imagens, ocorre uma perda significativa de informações, comprometendo os resultados quantitativos dos experimentos.

O método Thorax-Net alcançou a taxa de 0.787 de AUROC na tarefa de classificação da base de dados ChestX-ray14, com resultados promissores na detecção das seguintes patologias: Hernia, Cardiomegaly e Edema. Wang et al. [46] realizaram testes com outras redes neurais pré-treinadas, como a VGGNet-16 [35], GoogLeNet [38] e ResNet-50 [15], entretanto, a rede neural ResNet-152 foi a que obteve melhores resultados.

Tang et al. [40] propuseram uma abordagem baseada em aprendizagem guiada por atenção, denominada *Attention-Guided Curriculum Learning* (AGCL). Essa abordagem é utilizada para a classificação de doenças torácicas e localização fracamente supervisionada

das mesmas. Um detalhe importante é que na fase de treinamento não são utilizadas regiões delimitadoras.

A principal contribuição de Tang et al. [40] é a elaboração de um método de classificação de imagens torácicas utilizando a técnica de aprendizagem curricular [3]. A intuição dessa técnica é imitar o processo humano comum de aprendizado gradual, começando de conteúdos mais fáceis até os mais difíceis. Portanto, o treinamento do método é iniciado utilizando amostras de radiografias que contém doenças graves e, à medida que a DCNN vai convergindo, são adicionadas progressivamente amostras que contêm doenças moderadas e leves.

Além dessa técnica de aprendizagem curricular, a abordagem faz uso dos mapas de calor gerados pela DCNN para guiar o processo de treinamento iterativo. Para iniciar esse processo, são utilizadas duas sementes: imagens de radiografias que contém doenças graves e moderadas e imagens com altos valores de probabilidade de classificação gerados pelo classificador da DCNN.

Primeiramente, Tang et al. [40] realizaram um redimensionamento das imagens de 1024×1024 para 512×512 *pixels*. A rede neural utilizada como extrator de características foi a ResNet-152 [15] e as regiões de interesse de cada patologia existente nas radiografias foram geradas a partir dos mapas de calor gerados pela rede neural. Com essa abordagem, a taxa de acerto com a métrica AUROC foi de 0.803.

Rajpurkar et al. [28] propuseram, inicialmente, um método para detecção de pneumonia em imagens de radiografia de tórax, denominado CheXNet, porém, na fase experimental, os autores decidiram estender essa abordagem para a classificação de patologias torácicas existentes na base de dados ChestX-ray14 [48], para fins de comparação.

O método proposto por Rajpurkar et al. [28] é composto de uma DenseNet-121 [17] que recebe como entrada uma imagem de radiografia de tórax e retorna um vetor de 14 posições com as probabilidades das patologias existentes na base de dados. Os resultados dos experimentos apontaram para um valor de 0.807 de AUROC. Apesar desse método ser simples, por ser uma rede pré-treinada com a ImageNet [9] e treinada com a base de dados ChestX-ray14 [48], obteve um resultado significativo entre os trabalhos dessa área.

Gundel et al. [13] propuseram uma abordagem para explorar informações espaciais de patologias em imagens de radiografia de tórax de alta resolução, denominado DNetLoc. A rede neural utilizada como extratora de características é a DenseNet-121 [17], e cada saída é normalizada com uma função Sigmoid [49].

A contribuição desse trabalho está na inclusão de duas camadas convolucionais com 3 filtros 3×3 e um *stride* de 2 para explorar efetivamente a radiografia de tórax em alta resolução. Os pesos dos filtros de ambas as camadas são inicializados iguais a uma operação de redução de amostragem gaussiana. Um detalhe importante na abordagem proposta por Gundel et al. [13] é que as imagens de entrada possuem dimensões de 1024×1024 (tamanho original das imagens da base de dados ChestX-ray14 [48]), o que necessita de mais recursos para processar esses dados.

Gundel et al. [13] utilizaram duas bases de dados como entrada para o método DNetLoc, a base ChestX-ray14 [48] e a base PLCO [41]. As etapas de treinamento e validação contém imagens de ambas as bases de dados, em que muitas classes são similares. Devido ao fato de que os autores não tinham a informação das duas bases terem sido criadas a

partir da mesma definição de rótulo, decidiram tratar as classes como independentes e criar um vetor de probabilidade para cada base de dados. Portanto, a saída final consiste em 35 valores de probabilidade, 14 da ChestX-ray14 [48] e 21 da PLCO [41].

Os resultados dos experimentos foram promissores, atingindo o valor de 0.807 na métrica AUROC para a base de dados ChestX-ray14 [48], empatado com o trabalho de Rajpurkar et al. [28].

Guan e Huang [12] propuseram uma abordagem de aprendizagem de atenção residual por categoria, denominada CRAL, para a classificação de imagens de radiografias de tórax. Esse método tem como objetivo investigar a interferência entre classes não correlacionadas e preservar correlações existentes entre as classes. Um aspecto importante desse método é a execução de um mecanismo de atenção residual por categoria para atribuir diferentes pesos a diferentes regiões espaciais da imagem. Ele prevê automaticamente os pesos de atenção para realçar as características relevantes e restringir as características irrelevantes para uma patologia específica.

O método CRAL consiste em um módulo de extração de características e um módulo de atenção. O módulo de extração de características utiliza uma DCNN para extrair informações de alto nível das imagens de entrada e o módulo de atenção aprende as áreas mais relevantes extraídas da rede neural. A integração das informações extraídas dos dois módulos é feita em um bloco de atenção residual e utilizada para a classificação das imagens de entrada.

Assim como o trabalho de Wang et al. [46], Guan e Huang [12] redimensionaram as imagens de entrada para 256×256 e, em seguida, recortaram aleatoriamente em imagens de dimensões 224×224, como método de aumentação de dados. Os autores utilizaram as redes neurais ResNet-50 [15] e DenseNet-121 [17] como extratoras de características.

O método CRAL obteve uma média de 0.816 na métrica AUROC considerando a base de dados ChestX-ray14 [48], resultado que superou os trabalhos desenvolvidos até então, considerando somente essa base de dados. Em algumas classes, os resultados se mostraram muito promissores, como Hernia e Mass.

Chen et al. [5] propuseram um método assimétrico de duas redes neurais, denominado DualCheXNet, para impulsionar mais as pesquisas sobre a classificação de imagens de radiografias de tórax multirrótulo. Esse método consiste em duas DCNNs - ResNet [15] e DenseNet [17] - extraindo características das mesmas imagens de entrada e utilizando essas informações em conjunto para a classificação.

Considerando que a ResNet [15] permite o reuso dos dados de entrada de cada bloco residual no bloco posterior e que a DenseNet [17] visa estimular seus blocos densos a explorar novas características por meio do caminho densamente conectado, Chen et al. [5] inferiram que as características extraídas pela ResNet [15] e pela DenseNet [17] são diferentes e únicos. Dessa forma, espera-se que uma rede aprenda características complementares que estão faltando na outra rede.

As imagens foram redimensionadas para 556×556 *pixels* e recortadas aleatoriamente em imagens de dimensões 512×512. Os resultados do método DualCheXNet são considerados o estado da arte na classificação da base de dados ChestX-ray14 [48], com a média de 0.823 na métrica AUROC.

A Tabela 4.1 apresenta algumas informações sobre os trabalhos apresentados anterior-

Tabela 4.1: Informações sobre os trabalhos relacionados à base de dados ChestX-ray14 [48]. As posições com o símbolo "-" indicam que a informação não foi apresentada no trabalho.

Método	**Batch**	**Tamanho**	**Recursos**	**AUROC**
U-DCNN [48]	8	1024×1024	4× Titan X	0.745
Capsule-Net [33]	-	256×256	GTX-1080Ti	0.775
Thorax-Net [46]	24	224×224	Titan XP	0.787
AGCL [40]	-	512×512	-	0.803
CheXNet [28]	16	224×224	-	0.807
DNetLoc [13]	128	1024×1024	-	0.807
CRAL [12]	64	224×224	-	0.816
DualCheXNet [5]	8	512×512	4× Titan XP	0.823

mente. O método proposto por Gundel et al. [13] possui o tamanho do *batch* maior que os demais métodos, e as dimensões das imagens de entrada são as originais da base de dados ChestX-ray14 [48], o que necessita de muito recurso para processamento. O trabalho que obteve melhor taxa de acerto, considerando a métrica AUROC, foi o de Chen et al. [5], e podemos destacar a quantidade de recurso utilizado para o tamanho do *batch* e dimensões das imagens de entrada considerado.

4.2 CheXpert

Seyyed-Kalantari et al. [31] propuseram o método CheXclusion, que tem como objetivo utilizar redes neurais profundas para classificar um conjunto formado por 3 bases de dados de radiografia de tórax. Além disso, os autores utilizam também informações adicionais, como raça, sexo e idade para realizar uma análise mais profunda sobre as condições dos pacientes, sendo assim, não somente uma verificação de existência de possíveis patologias.

O primeiro passo realizado por Seyyed-Kalantari et al. [31] é o processamento das bases de dados MIMIC-CXR [19], CheXpert [18] e ChestX-ray14 [48]. O objetivo desse processamento é ter um conjunto somente com as classes em comum e associar as classes "incertas" ao rótulo negativo. No total, são 8 classes em comum nas três bases de dados.

A rede neural utilizada para classificação das imagens de radiografia de tórax foi a DenseNet121 [17] pré-treinada com a ImageNet [9]. As imagens foram redimensionadas para 256×256 e foram aplicadas várias transformações, como recorte centralizado, giro horizontal e rotação aleatória. O resultado obtido nas 5 classes consideradas da base de dados CheXpert [18] foi 0.808, utilizando a métrica AUROC.

Zhang et al. [50] propuseram um método, denominado ConVIRT, para melhorar as representações visuais de imagens médicas, combinando as características do aprendizado realizado a partir de dados textuais e imagens. Esse método é aplicado na etapa de pré-treinamento do processo de aprendizado, aumentando a qualidade das imagens e destacando recursos visuais necessários para tarefas de compreensão de imagens médicas.

O método ConVIRT recebe como entrada uma imagem e um texto que descreve as informações da imagem, e tem como objetivo aprender uma função parametrizada para codificar imagens, que mapeia uma imagem em um vetor de dimensão fixa. Após esse

processo ser concluído, a função codificadora é utilizada nas tarefas de classificação e recuperação de imagens médicas. Vale ressaltar que a função codificadora foi modelada como uma rede neural convolucional, no caso a ResNe50 [15].

Essas bases de dados com imagens e textos correspondentes são comuns em domínios médicos, e uma das bases utilizadas nos experimentos foi a CheXpert [18]. Nos experimentos de classificação de imagens, Zhang et al. [50] utilizaram duas configurações individuais de CNNs. A primeira, onde os pesos da CNN pré-treinada são congelados e apenas as camadas de classificação são treinadas, e a segunda configuração, em que todos os pesos da CNN são ajustados de acordo com o treinamento.

Para fins de comparação, como apresentado anteriormente (Seção 3.2), os trabalhos que utilizam a base de dados CheXpert [18] reportam os resultados somente em 5 classes das 14 existentes. As imagens de entrada são redimensionadas para 224×224 e passam por algumas transformações, como o recorte aleatório, giro horizontal, translação vertical e horizontal, e ajustes de brilho e contraste. A média de AUROC reportada para as 5 classes foi de 0.881.

Bressem et al. [4] propuseram um estudo mais geral do comportamento das redes neurais na classificação de duas bases de dados: CheXpert [18] e COVID-19 [8]. Esse trabalho tem como hipótese que o número de camadas de uma rede neural convolucional não é necessariamente decisivo para uma predição com alta taxa de acerto em dados médicos. Em algumas situações, CNNs com menos camadas podem ter desempenho semelhante ou até melhor que redes mais profundas/complexas.

Os experimentos foram realizados de forma sistemática com 16 CNNs, dentre elas estão a família de redes DenseNets [17], ResNets [15] e VGGNets [35], todas pré-treinadas na base dados ImageNet [9]. Além disso, foi considerado dois tamanhos de *batch* nos experimentos, 16 e 32, e os modelos foram treinados por 5 épocas só com as camadas de classificação descongeladas e, após isso, foram treinados por 3 épocas com todas as camadas descongeladas.

As imagens de entrada são redimensionadas para 320×320, utilizando interpolação bilinear, e os valores dos *pixels* são normalizados de acordo com a média e desvio padrão dos dados da ImageNet [9]. Transformações de aumentação de dados também foram aplicadas, como a inversão de colunas e linha, rotação e adição de brilho. A rede neural que obteve o melhor resultado foi a DenseNet161, com a média de AUROC 0.882.

Por fim, o último trabalho relacionado sobre a base de dados CheXpert [18] é sobre o método utilizado no próprio artigo da base de dados. Irvin et al. [18] propuseram cinco abordagens para tratar o problema da incerteza de rótulos para algumas das imagens, que são: *U-Ignore*, que simplesmente ignora os dados que possuem rótulos incertos; *U-Zeros*, que atribui o valor 0 ou "negativo" para os rótulos faltantes ou incertos; *U-Ones*, que é de forma análoga a *U-Zeros*, só que atribuindo o valor 1 ou "positivo"; *U-SelfTrained*, que utiliza um modelo treinado com a abordagem *U-Ignore* para classificar as imagens com rótulos incertos e decidir quais valores serão alocados naqueles dados faltantes ou incerto; e, por fim, a abordagem *U-MultiClass*, que trata os rótulos incertos como uma classe extra.

As arquiteturas usadas nos experimentos foram: ResNet152 [15], DenseNet121 [17], Inception-V4 [38] e SE-ResNeXt101 [16]. As imagens foram redimensionadas para

320×320 e não foi deixado claro no artigo se foi utilizado algum tipo de aumentação de dados. O melhor resultado obtido foi com a rede neural DenseNet121 [17], com média de AUROC 0.895.

Método	**Batch**	**Tamanho**	**Recursos**	**AUROC**
CheXclusion [31]	32	224×224	Titan RTX	0.808
ConVIRT [50]	48	256×256	NVIDIA GPU 16GB	0.881
DenseNet161 [4]	32	320×320	2× RTX 2080ti	0.882
CheXpert [18]	16	320×320	-	0.895

Tabela 4.2: Informações sobre os trabalhos relacionados à base de dados CheXpert [18]. As posições com o símbolo "-" indicam que a informação não foi informada no trabalho.

A Tabela 4.2 apresenta algumas informações sobre os trabalhos que consideram a base de dados CheXpert [18]. Podemos destacar o método ConVIRT [50], que utilizou o maior tamanho de *batch* dentre os trabalhos apresentados, o que mostrou ser um fator que não contribuiu para ter um resultado próximo aos demais trabalhos. Uma característica que interfere no resultado e nas comparações entre o resultado do nosso método e dos trabalhos apresentados, é a forma como é tratada a incerteza sobre os rótulos. Portanto, não podemos fazer uma comparação totalmente justa entre o nosso método e os trabalhos relacionados utilizando a base de dados CheXpert [18].

Capítulo 5

DuaLAnet

Na primeira seção deste capítulo (Seção 5.1), apresentamos o nosso método e seus detalhes, como os classificadores e estratégia de treinamento. Na Seção 5.2, definimos os experimentos realizados, os detalhes de implementação e a métrica utilizada na avaliação dos modelos. Por fim, na Seção 5.3, apresentamos os resultados dos experimentos e algumas análises sobre esses resultados.

5.1 Construção da DuaLAnet

O nosso objetivo com esse método é explorar a complementaridade de extração de características entre redes neurais. Assim, decidimos colocar duas redes em paralelo recebendo os mesmos dados de entrada, extraindo as características de acordo com cada arquitetura e concatenando essas características em um só vetor.

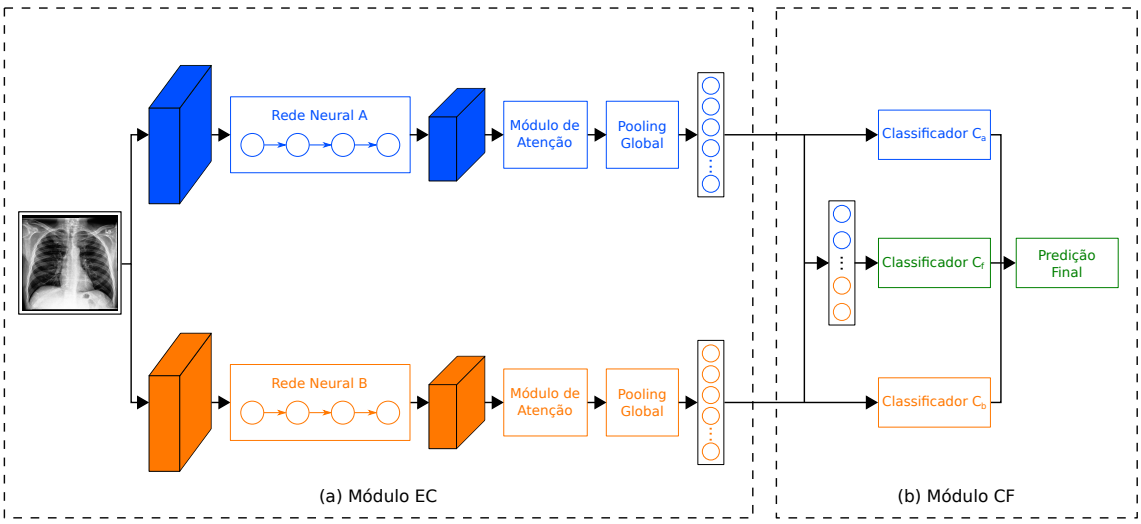
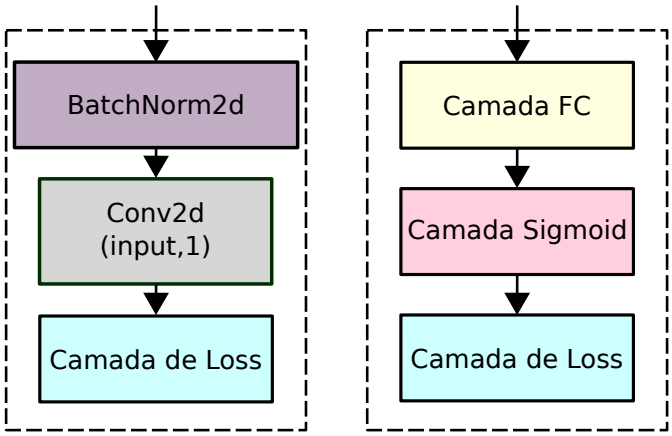


Figura 5.1: Ilustração do método DuaLAnet. Duas redes neurais são utilizadas para aprender características complementares das imagens de entrada. As saídas do *pooling* global das redes neurais A e B são concatenadas e passadas como entrada para o classificador de fusão C_f. A saída final é baseada nos resultados dos dois classificadores auxiliares C_a e C_b, e no classificador de fusão C_f.

A arquitetura do nosso método é mostrada na Figura 5.1, que contém dois módulos:

Módulo de Extração de Características (EC) e Módulo de Classificação de Fusão (CF). No Módulo EC, duas redes neurais são usadas como extrator de características para o método DuaLAnet. Após as camadas de extração de cada rede, foi adicionado um módulo de atenção. Consideramos três opções de módulos de atenção: (i) Mapeamento de Ativação de Classe (*Class Activation Mapping*, ou CAM); (ii) Mapeamento de Ativação Suave (*Soft Activation Mapping*, ou SAM); e *Feature Pyramid Attention*, ou FPA. Após o módulo de atenção de cada rede, uma camada de pooling *Average-Max* (AVG-MAX) é usada para reduzir a variância e a complexidade da computação, além de extrair características de baixo nível (AVG) e alto nível (MAX) da vizinhança.

No Módulo CF, três classificadores estão envolvidos: um classificador de fusão C_f e dois classificadores auxiliares C_a e C_b. Os classificadores auxiliares C_a e C_b tem como objetivo tornar os vetores de características produzidos por cada ramo discriminativos, como mostrado na Figura 5.2a. A saída da camada de pooling AVG-MAX é usada como entrada para a camada de normalização, que ajusta e dimensiona os mapas de ativação. Usamos uma camada convolucional 2D em vez de uma camada totalmente conectada (FC) porque consideramos a classificação de cada classe independente.



(a) Classificador Auxiliar. (b) Classificador de Fusão.

Figura 5.2: Ilustração dos classificadores utilizados no método DuaLAnet: (a) módulo inserido no final das redes neurais A e B; (b) módulo usado para a classificação do vetor com a concatenação do *pooling* global das redes neurais A e B.

5.1.1 Estratégia de Treinamento

O treinamento do método é uma etapa importante para uma boa predição sobre os dados. O nosso método de treinamento (Algoritmo 1) é dividido em três etapas principais: o treinamento do ramo da Rede Neural A, o treinamento do ramo da Rede Neural B e o treinamento da rede neural completa.

A entrada de dados é composta pelas imagens originais da base de dados (ChestX-ray14 [48] ou CheXpert [18]) e os vetores de rótulos referentes às imagens de entrada. A inicialização dos pesos das redes neurais A e B é feita a partir da base de dados ImageNet [9], e a camada totalmente conectada no classificador de fusão C_f tem seus pesos inicializados de forma aleatória.

Algoritmo 1: Algoritmo de Treinamento do Método

Entrada : Imagens originais e os vetores de rótulos.
Saída : Predição final baseada nos resultados dos classificadores C_a, C_b e C_f.
Inicialização: Inicializa as DCNNs com os pesos pré-treinados na ImageNet; a camada FC no classificador de fusão C_f tem seus pesos inicializados de forma randômica.

repita
  Atualiza os pesos do ramo da Rede Neural A: Fixa os pesos do ramo da Rede Neural B e atualiza os pesos do ramo da Rede Neural A, o classificador auxiliar C_a e o classificador de fusão C_f.
até *convergir ou o número máximo de iterações for alcançado*;
repita
  Atualiza os pesos do ramo da Rede Neural B: Fixa os pesos do ramo da Rede Neural A e atualiza os pesos do ramo da Rede Neural B, o classificador auxiliar C_b e o classificador de fusão C_f.
até *convergir ou o número máximo de iterações for alcançado*;
repita
  Atualiza a rede neural por inteira: Atualiza os pesos do ramo da Rede Neural A e do ramo da Rede Neural B, os dois classificadores auxiliares C_a e C_b, e o classificador de fusão C_f.
até *convergir ou o número máximo de iterações for alcançado*;

A primeira etapa de treinamento é realizada para atualizar os pesos do ramo da Rede Neural A. Para isso, os pesos do ramo da Rede Neural B são fixados, assim, são atualizados os pesos do ramo da Rede Neural A, o classificador auxiliar C_a e o classificador de fusão C_f. Essa atualização é feita ou até convergir a taxa de perda ou o número máximo de iterações for alcançado.

A segunda etapa de treinamento é semelhante a primeira, alterando o ramo que será treinado. Nesse caso, os pesos do ramo da Rede Neural A são fixados, e são atualizados os pesos do ramo da Rede Neural B, o classificador auxiliar C_b e o classificador de fusão C_f. Por fim, toda a rede neural tem seus pesos atualizados, considerando, inicialmente, os pesos da melhor época das etapas anteriores.

No nosso método, utilizamos a técnica de *ensemble* para treinar o modelo. Dessa forma, o modelo é otimizado durante a fase de treinamento a partir de uma função de perda L, definida a seguir:

$$L = \gamma_1 L_f + \gamma_2 L_a + \gamma_3 L_b \tag{5.1}$$

As variáveis L_f, L_a e L_b representam a taxa de perda do classificador C_f, C_a e C_b, respectivamente. Os valores das variáveis γ_1, γ_2 e γ_3 são escolhidos a partir de testes, variando essas variáveis entre os seguintes valores: $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

5.2 Experimentos

Nesta seção, apresentamos os experimentos que realizamos e suas configurações. Mostramos a etapa de pré-treinamento, considerando as bases de dados ChestX-ray14 [48] e CheXpert [18], as técnicas de aumentação de dados utilizadas, as redes neurais consideradas na extração de características, os detalhes de implementação dos métodos e a métrica utilizada para validar o modelo.

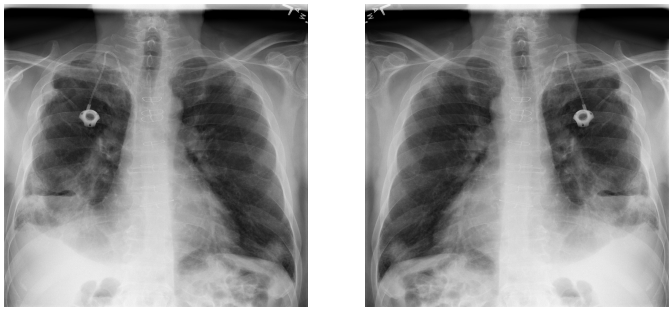
5.2.1 Pré-Treinamento

As bases de dados ChestX-ray14 [48] e CheXpert [18] possuem 14 classes cada uma, e intersecção de 4 classes, que são: Atelectasis, Pneumonia, Pneumothorax e Consolidation. Considerando essa semelhança entre as bases, decidimos realizar um treinamento cruzado para verificar a influência do pré-treinamento com imagens aleatórias (ImageNet [9]) e com imagens de radiografias de tórax.

Para esse experimento, utilizamos a rede neural DenseNet169 [17]. As configurações das imagens de entrada e da rede neural são apresentadas nas seções a seguir.

5.2.2 Aumentação de Dados

Na etapa de aumentação de dados, consideramos dois métodos principais: (i) Rotação Horizontal, que inverte as colunas de *pixels* da imagem de entrada com uma probabilidade de 50% e (ii) Recorte Centralizado, que recorta a imagem de entrada de acordo com as dimensões designadas e com foco no centro da imagem original, buscando eliminar ruídos do fundo da imagem.

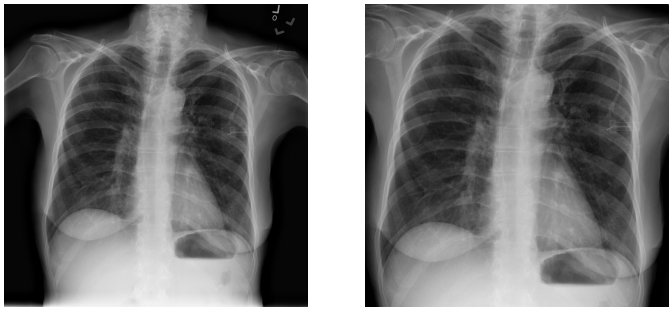


(a) Imagem original. (b) Rotação Horizontal.

Figura 5.3: Exemplo de aplicação do método de rotação horizontal aleatória.

A Figura 5.3 apresenta um exemplo de aplicação do método de aumentação de dados denominado Rotação Horizontal. Na Figura 5.3a, temos a imagem original da base de dados ChestX-ray14 [48], e na Figura 5.3b, temos a inversão das colunas de *pixels* da imagem original.

A Figura 5.4 mostra um exemplo da aplicação dá técnica de Recorte Centralizado, que é utilizada como um método de aumentação de dados para a remoção de ruídos existentes no fundo da imagem original. Na Figura 5.4a, temos a imagem original da base de dados CheXpert [18], e na Figura 5.4b, temos a imagem original recortada centralizada.



(a) Imagem original. (b) Recorte Centralizado.

Figura 5.4: Exemplo de aplicação do método de recorte centralizado.

Vale ressaltar que as imagens originais da base de dados ChestX-ray14 [48] possuem dimensões 1024×1024, enquanto as imagens da base de dados CheXpert [18] possuem dimensões 2048×2048 *pixels*. Todas as imagens de ambas as bases de dados foram redimensionadas para 556×556 e, depois, aplicamos a normalização de acordo com a média e desvio padrão da base de dados ImageNet [9]. Utilizamos os valores 512×512 como dimensões para o recorte centralizado.

5.2.3 Extratores de Características

O método DuaLAnet precisa de duas redes neurais como extratoras de características e que sejam complementares. Portanto, realizamos experimentos com ambas as bases de dados para a escolha da "Rede Neural A" e da "Rede Neural B". Consideramos as seguintes arquiteturas em nossos experimentos: DenseNet121, DenseNet161, DenseNet169, DenseNet201, VGGNet16, ResNet50, ResNet101, ResNet152 e EfficientNetB5. Ressaltamos que, como estamos procurando por redes complementares, só escolhemos a melhor rede da família DenseNet, por exemplo.

As configurações utilizadas neste experimento são as mesmas utilizadas no método principal. Após a escolha das redes neurais que serão utilizadas como extratoras de características, realizamos outros experimentos adicionando os módulos de atenção e verificando qual o que obtém melhor taxa de acerto na classificação das imagens de entrada, de acordo com cada rede neural.

5.2.4 Detalhes de Implementação

Implementamos o método DuaLAnet utilizando o *framework* Pytorch[1], com o qual é possível desenvolver métodos de Aprendizado Profundo com um alto grau de liberdade. Usamos o otimizador Adam [20] e uma taxa de aprendizagem de 1×10^{-4}, que é reduzida por um fator de 1×10^{-1} quando a taxa de perda não diminui em um intervalo de 5 épocas. Todos os experimentos são limitados a 30 épocas.

Os experimentos deste trabalho foram executados no Laboratório de Informática Visual (LIV) do Instituto de Computação (IC-Unicamp), em uma máquina com processador

[1] `https://pytorch.org/`

Intel i7-3770 de 3.50 GHz e uma GPU NVIDIA TITAN V, com 5120 núcleos e 12 GB de memória.

Como os nossos recursos são consideravelmente limitados, utilizamos a técnica de Acúmulo de Gradiente [25], que serve para dividir o *batch*, usado para o treinamento da rede neural, em vários *mini-batches* que serão executados sequencialmente e depois é realizado o processo de *backpropagation* com o gradiente de todos os *mini-batches*. Dessa forma, é possível executar métodos que necessitam de uma grande quantidade memória de GPU com menos recursos.

5.2.5 Métricas de Avaliação

Vários trabalhos existentes na literatura que utilizam as bases de dados ChestX-ray14 [48] e CheXpert [18] empregam a métrica AUROC (*Area Under the Receiver Operating Characteristic*). Essa métrica serve para avaliação de modelos em problemas de classificação variando as configurações de limiares (*thresholds*).

A ROC (*Receiver Operating Characteristic*) é uma curva de probabilidade e a AUC (*Area Under the Curve*) representa o grau ou a medida de separabilidade. Essa métrica informa quanto o modelo é capaz de distinguir entre classes, então, quanto maior a AUC, melhor o modelo está predizendo positivos como positivos e negativos como negativos. De forma análoga, quanto maior a AUC, melhor o modelo é capaz de distinguir entre pacientes com doença e sem doença.

$$\text{TPR} = \text{Sensibilidade} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.2}$$

$$\text{Especificidade} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{5.3}$$

$$\text{FPR} = 1 - \text{Especificidade} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{5.4}$$

Para entendermos a métrica AUROC, precisamos definir algumas equações. A Equação 5.2 define a métrica Taxa de Verdadeiros Positivos (*True Positive Rate*, ou TPR), também conhecida como Sensibilidade, que é responsável por medir o sucesso na identificação das patologias existentes. Por sua vez, a Equação 5.3 define a métrica Especificidade (*Specificity*), que mede a proporção de casos negativos reais que foram previstos como negativos. Finalmente, a Equação 5.4 define a métrica Taxa de Falsos Positivos (*False Positive Rate*, ou FPR), que mede o sucesso de não identificar casos normais como anormais.

Vale destacar que a Sensibilidade e a Especificidade são inversamente proporcionais, assim, quando aumentamos a sensibilidade, a especificidade diminui e vice-versa. A sigla FN é o número de casos falsos negativos (*False Negatives*), FP é o número de casos falsos positivos (*False Positives*), TN é o número de casos verdadeiros negativos (*True Negatives*) e TP é o número de casos verdadeiros positivos (*True Positives*). Uma curva ROC representa os valores de TPR em função dos valores de FPR.

Em outras palavras, AUC mede a área bidimensional abaixo da curva ROC das coordenadas $(0, 0)$ até $(1, 1)$, que é a área em cinza na Figura 5.5. A Figura 5.6 apresenta um

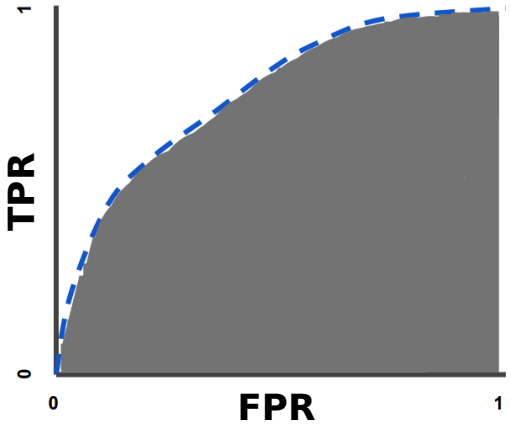


Figura 5.5: Exemplo genérico de AUROC.

exemplo real de AUROC obtido com a DCNN ResNet-50 no trabalho proposto por Wang et al. [48], em que a classe "Cardiomegaly" obteve o maior valor e a classe "Mass" obteve o menor valor.

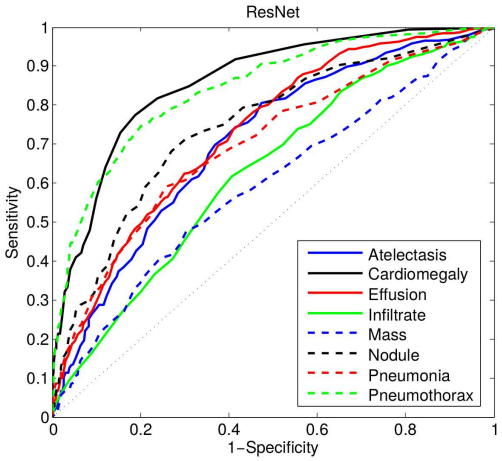


Figura 5.6: Exemplo de AUROC obtido com a DCNN ResNet-50 [48].

5.3 Resultados e Discussões

Nesta seção, apresentamos os resultados dos experimentos mostrados anteriormente, juntamente com as discussões e detalhes de cada resultado. Dividimos esta seção em duas subseções, uma para a base de dados ChestX-ray14 [48] (Seção 5.3.1) e a outra para a base de dados CheXpert [18]. Vale ressaltar que os resultados são apresentados utilizando a métrica AUROC, apresentada anteriormente na Seção 5.2.5.

5.3.1 ChestX-ray14

Primeiramente, consideramos o experimento de pré-treinamento, para verificar qual a forma de inicialização da rede neural que se comporta melhor para a classificação da base de dados ChestX-ray14 [48]. Utilizamos a rede neural DenseNet169 com o módulo de atenção CAM para esse experimento, pois foi o método que se comportou melhor nas duas bases de dados consideradas neste trabalho, como mostrado ao longo dessa seção.

Por questão de espaço, abreviamos os nomes das classes como mostrado a seguir. As 14 patologias existentes nessa base de dados são: Atelectasis (Atel), Cardiomegaly (Card), Effusion (Effu), Infiltration (Infi), Mass, Nodule (Nodu), Pneumonia (Pneu1), Pneumothorax (Pneu2), Consolidation (Cons), Edema (Edem), Emphysema (Emph), Fibrosis (Fibr), Pleural Thickening (PT) e Hernia (Hern).

A Tabela 5.1 apresenta os resultados da rede neural DenseNet169 com o módulo de atenção CAM (D169-CAM) com a inicialização dos pesos feita a partir do pré-treinamento com a base de dados ImageNet [9] e a inicialização feita a partir do treinamento com a base de dados CheXpert [18] (D169-CAM-PRE). Podemos destacar que o método D169-CAM-PRE obteve resultado melhor do que o método com inicialização a partir da ImageNet [9], mas a diferença foi pequena. As classes que tiveram mais diferença nos resultados foram a Nodule e Pneumothorax, com mais de um ponto percentual a mais para o método pré-treinado com a outra base de radiografia de tórax, CheXpert [18].

Tabela 5.1: Resultado do pré-treinamento.

Método	Atel	Card	Effu	Infi	Mass	Nodu	Pneu1	Pneu2	Cons	Edem	Emph	Fibr	PT	Hern	Média
D169-CAM	0.7670	0.8733	0.8216	0.7068	0.8138	0.7610	0.7271	0.8358	0.7394	0.8363	0.9073	0.8143	0.7714	0.9046	0.8057
D169-CAM-PRE	0.7664	0.8762	0.8294	0.7058	0.8131	0.7757	0.7250	0.8497	0.7408	0.8309	0.9106	0.8158	0.7780	0.9039	**0.8090**

Como o nosso método, DuaLAnet, demanda duas redes neurais como extratoras de características, realizamos alguns experimentos com redes pré-treinadas para escolher as que obtiveram os melhores resultados. A Tabela 5.2 apresenta o resultado das redes neurais utilizadas como extratoras de características, em que todas foram inicializadas a partir da base de dados ImageNet [9]. Como a ideia do método é ter informações complementares dos dois ramos principais, consideramos somente uma rede de cada "família" de rede neural. A DenseNet169 obteve o melhor resultado em geral, e a ResNet152 obteve o segundo melhor resultado, desconsiderando o restante das redes neurais da "família" DenseNet.

A DenseNet169 obteve o melhor resultado geral em duas classes, Mass e Emphysema, mas com resultados próximos das demais redes neurais. A ResNet152 obteve o melhor

Tabela 5.2: Resultado das redes neurais utilizadas como extratoras de características.

Método	Atel	Card	Effu	Infi	Mass	Nodu	Pneu1	Pneu2	Cons	Edem	Emph	Fibr	PT	Hern	Média
DenseNet121	0.7705	0.8550	0.8283	0.6984	0.8180	0.7850	0.7314	0.8702	0.7342	0.8445	0.9195	0.8050	0.7838	0.8831	0.8091
DenseNet161	0.7771	0.8667	0.8287	0.7111	0.8126	0.7748	0.7261	0.8526	0.7414	0.8352	0.9244	0.8075	0.7810	0.8881	0.8091
DenseNet169	0.7766	0.8631	0.8258	0.7091	0.8232	0.7815	0.7300	0.8609	0.7433	0.8429	0.9335	0.8095	0.7669	0.8871	**0.8110**
DenseNet201	0.7786	0.8568	0.8276	0.7008	0.8228	0.7823	0.7230	0.8692	0.7446	0.8426	0.9137	0.8150	0.7866	0.8831	0.8105
VGG16	0.7505	0.8002	0.8204	0.7042	0.7892	0.7570	0.7094	0.8574	0.7230	0.8205	0.9172	0.7877	0.7493	0.7652	0.7822
ResNet50	0.7545	0.8421	0.8218	0.6903	0.7867	0.7465	0.7130	0.8521	0.7199	0.8296	0.9144	0.8028	0.7577	0.7915	0.7874
ResNet101	0.7597	0.8475	0.8182	0.7056	0.8090	0.7649	0.7212	0.8570	0.7177	0.8239	0.9202	0.8017	0.7633	0.8171	0.7948
ResNet152	0.7653	0.8386	0.8190	0.7122	0.8124	0.7750	0.7207	0.8598	0.7203	0.8275	0.9187	0.8033	0.7713	0.8056	0.7964
EfficientNet	0.7684	0.8357	0.8200	0.7017	0.8046	0.7863	0.7067	0.8553	0.7460	0.8240	0.9111	0.8124	0.7670	0.8023	0.7958

resultado geral somente na classe Infiltration, com menos de meio ponto percentual para o segundo melhor resultado, que é da rede DenseNet161.

Após escolhermos as redes neurais que utilizamos como extratoras de características, realizamos alguns experimentos para verificar qual módulo de atenção obtém melhor resultado com cada rede neural. Como as redes neurais que obtiveram os melhores resultados separadamente foram a DenseNet169 e a ResNet152, experimentamos os três módulos de atenção em cada uma de forma independente.

Como mostramos na Tabela 5.3, a rede neural DenseNet169 obteve melhor resultado com o módulo de atenção CAM, em comparação com os módulos de atenção SAM e FPA. Podemos destacar que o resultado foi consideravelmente melhor, com mais de um ponto percentual a mais do que o método com a DenseNet169 e o módulo FPA, segundo colocado. O método com a DenseNet169 e o módulo de atenção SAM teve o pior resultado, com mais de três pontos percentuais a menos do que o melhor resultado. Considerando a rede neural ResNet152, o método com o módulo de atenção SAM foi o que obteve melhor resultado, com menos de um ponto percentual para o segundo melhor colocado, a ResNet152 com o módulo CAM. O método com o módulo FPA obteve o pior resultado, com pelo menos cinco pontos percentuais de diferença para o melhor resultado.

Tabela 5.3: Comparação entre os resultados da DenseNet169 (D169) e a ResNet152 (R152) considerando três módulos de atenção diferentes (CAM, SAM e FPA).

Método	**Atel**	**Card**	**Effu**	**Infi**	**Mass**	**Nodu**	**Pneu1**	**Pneu2**	**Cons**	**Edem**	**Emph**	**Fibr**	**PT**	**Hern**	**Média**
D169-CAM	0.7670	0.8733	0.8216	0.7068	0.8138	0.7610	0.7271	0.8358	0.7394	0.8363	0.9073	0.8143	0.7714	0.9046	**0.8057**
D169-SAM	0.7948	0.8910	0.8734	0.5376	0.8082	0.6807	0.7066	0.7982	0.7747	0.8838	0.7525	0.6719	0.7911	0.8210	0.7704
D169-FPA	0.7699	0.8881	0.8264	0.7114	0.8202	0.7557	0.7205	0.8439	0.7241	0.8209	0.8853	0.7795	0.7347	0.8300	0.7936
R152-CAM	0.7673	0.8754	0.8164	0.6934	0.8141	0.7478	0.6988	0.8347	0.7315	0.8205	0.8921	0.8024	0.7591	0.7936	0.7891
R152-SAM	0.7648	0.8797	0.8172	0.6947	0.8240	0.7492	0.7010	0.8359	0.7346	0.8250	0.8950	0.7994	0.7606	0.8850	<u>0.7976</u>
R152-FPA	0.7408	0.8602	0.7952	0.6942	0.7379	0.7050	0.6877	0.7694	0.6882	0.8082	0.7210	0.7164	0.7039	0.8034	0.7451

Para avaliar o nosso método, DuaLAnet, comparamos essa abordagem com métodos existentes na literatura para classificação de imagens de radiografias de tórax com a base de dados ChestX-ray14 [48], todos considerando a divisão oficial da base, para uma comparação justa. Os métodos existentes que utilizamos para comparação são U-DCNN [48], Thorax-Net [46], DualCheXNet [5], DNet [13], CheXNet [28], CapsuleNet [33], AGCL [40] e CRAL [12].

A Tabela 5.4 apresenta os resultados numéricos dos métodos do estado da arte e do nosso método proposto, DuaLAnet. Em comparação com os métodos existentes, a nossa abordagem obteve uma taxa de acerto competitiva na base de dados ChestX-ray14 [48]. O nosso método obteve o valor médio da métrica AUROC de 0.820, com uma diferença de apenas 0.003 para o método DualCheXNet [5], considerado o melhor resultado. Vale ressaltar que o método proposto por Chen et al. [5] utilizou mais recursos do que utilizamos para treinar o nosso método.

Além de obter uma média das classes muito próxima do melhor resultado, a nossa abordagem obteve o melhor resultado individual nas classes Effusion, Nodule e Pleural Thickening, e o segundo melhor resultado nas classes Atelectasis, Mass, Pneumonia, Pneumothorax, Consolidation e Emphysema.

A Figura 5.7 apresenta as quatro melhores configurações de parâmetros de fusão utili-

Tabela 5.4: Comparação do nosso método DuaLAnet com as abordagens do estado da arte na base de dados ChestX-ray14 [48].

Method	Atel	Card	Effu	Infi	Mass	Nodu	Pneu1	Pneu2	Cons	Edem	Emph	Fibr	PT	Hern	Mean
U-DCNN [48]	0.700	0.810	0.759	0.661	0.693	0.669	0.658	0.799	0.703	0.805	0.833	0.786	0.684	0.872	0.745
CapsuleNet [33]	0.766	0.801	0.797	**0.751**	0.760	0.741	**0.778**	0.800	**0.787**	0.820	0.773	0.765	0.759	0.748	0.775
Thorax-Net [46]	0.750	0.871	0.818	0.681	0.799	0.714	0.693	0.825	0.741	0.835	0.842	0.804	0.746	0.902	0.787
AGCL [40]	0.756	0.887	0.819	0.689	0.814	0.755	0.729	0.850	0.728	0.848	0.906	0.818	0.765	0.875	0.803
CheXNet [28]	0.769	0.885	0.825	0.694	0.824	0.759	0.715	0.852	0.745	0.842	0.906	0.821	0.766	0.901	0.807
DNet [13]	0.767	0.883	0.828	0.709	0.821	0.758	0.731	0.846	0.745	0.835	0.895	0.818	0.761	0.896	0.807
CRAL [12]	0.781	0.880	0.829	0.702	0.834	0.773	0.729	0.857	0.754	0.850	0.908	0.830	0.778	**0.917**	0.816
DualCheXNet [5]	**0.784**	**0.888**	0.831	0.705	**0.838**	0.796	0.727	**0.876**	0.746	**0.852**	**0.942**	**0.837**	**0.796**	0.912	**0.823**
DuaLANet	0.783	0.884	**0.832**	0.708	0.837	**0.800**	0.735	0.866	0.746	0.841	0.937	0.820	**0.796**	0.895	0.820

zados no treinamento do modelo. Podemos destacar a melhor configuração de parâmetros, Config1, que considera os valores do ramo de fusão com peso 1.0, valores da DenseNet169 com o módulo de atenção CAM com peso 0.5 e os valores da ResNet152 com o módulo de atenção SAM com peso 0.2. Podemos destacar que o ramo de fusão possui um papel fundamental no método, pois o melhor resultado é considerando os valores desse ramo com maior peso. O ramo da DenseNet169 possui mais influência na classificação final do que o ramo da ResNet152, mas ambos são necessários para o bom desempenho do método.

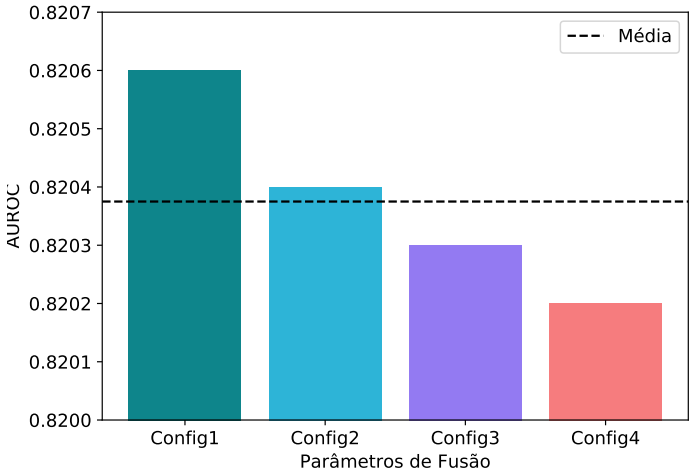


Figura 5.7: Média de AUROC para o método DuaLAnet variando os valores de γ_1, γ_2 e γ_3. Os valores dos parâmetros são: Config1 ($\gamma_1 = 1.0, \gamma_2 = 0.5, \gamma_3 = 0.2$), Config2 ($\gamma_1 = 1.0, \gamma_2 = 0.5, \gamma_3 = 0.5$), Config3 ($\gamma_1 = 0.5, \gamma_2 = 1.0, \gamma_3 = 0.5$) e Config4 ($\gamma_1 = 0.5, \gamma_2 = 0.5, \gamma_3 = 1.0$).

A Figura 5.8 apresenta a variação da taxa de perda ao longo das 30 épocas no treinamento do ramo de fusão do método DuaLAnet. A menor taxa de perda, considerando o conjunto de validação, aconteceu na época 5 e foi o modelo escolhido para a realização da etapa de teste. Vale ressaltar que a taxa de perda na etapa de validação não teve uma grande variação e não se comportou como a taxa de perda na etapa de treinamento, que foi convergindo para um valor menor.

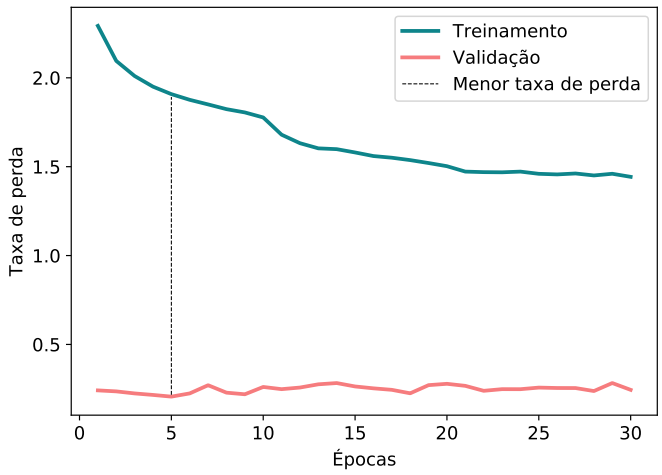


Figura 5.8: Taxa de perda das etapas de treinamento e validação do método DuaLAnet (ramo de fusão).

5.3.2 CheXpert

No experimento de pré-treinamento, considerando a base de dados CheXpert [18], consideramos a mesma arquitetura utilizada nesse mesmo experimento na base de dados anterior, que é a DenseNet169 com o módulo de atenção CAM (D169-CAM). Por motivo de espaço, abreviamos os nomes das classes como mostrado a seguir. As 14 classes existentes na base de dados CheXpert [18] são: No Finding (NFin), Enlarged Cardiom. (ECar), Cardiomegaly (Card), Lung Opacity (LOpa), Lung Lesion (LLes), Edema (Edem), Consolidation (Cons), Pneumonia (Pneu1), Atelectasis (Atel), Pneumothorax (Pneu2), Pleural Effusion (PEff), Pleural Other (POth), Fracture (Frac) e Support Devices (Supp).

A Tabela 5.5 apresenta os resultados da rede neural DenseNet169 com o módulo de atenção CAM (D169-CAM) com a inicialização dos pesos feita a partir do pré-treinamento com a base de dados ImageNet [9] e a inicialização feita a partir do treinamento com a base de dados ChestX-ray14 [48] (D169-CAM-PRE). Neste experimento, apresentamos a taxa AUROC para todas as classes existentes na base de dados CheXpert [18]. Podemos destacar que o método com inicialização feita a partir do ImageNet [9] obteve resultado superior ao método D169-CAM-PRE. As classes que apresentaram maiores diferenças (superior a um ponto percentual) nos resultados foram a Edema e Pneumonia.

Tabela 5.5: Resultado do pré-treinamento.

Método	NFin	ECar	Card	LOpa	LLes	Edem	Cons	Pneu1	Atel	Pneu2	PEff	POth	Frac	Supp	Média
D169-CAM	0.8645	0.6949	0.8501	0.7314	0.7953	0.8053	0.7417	0.7651	0.6860	0.8611	0.8712	0.8129	0.7729	0.8788	**0.7951**
D169-CAM-PRE	0.8590	0.6895	0.8455	0.7280	0.7932	0.7942	0.7332	0.7506	0.6841	0.8512	0.8675	0.8079	0.7769	0.8694	0.7893

O resultado dos experimentos realizados para escolher as duas redes neurais que são utilizadas como extratoras de características é apresentado na Tabela 5.6. Todas as redes neurais utilizadas nesse experimento foram inicializadas a partir da base de dados ImageNet [9]. A DCNN DenseNet169 obteve a maior média de AUROC, assim como na base de dados ChestX-ray14 [48], com o valor 0.7888, e a segunda melhor DCNN foi a Effici-

Tabela 5.6: Resultado das redes neurais utilizadas como extratoras de características.

Método	Cardiomegaly	Edema	Consolidation	Atelectasis	Pleural Effusion	Média
DenseNet121	0.8410	0.7998	0.7417	0.6723	0.8711	0.7852
DenseNet161	0.8393	0.8004	0.7410	0.6743	0.8687	0.7847
DenseNet169	0.8486	0.8027	0.7425	0.6792	0.8709	**0.7888**
DenseNet201	0.8471	0.8012	0.7387	0.6735	0.8712	0.7863
VGG16	0.8254	0.7589	0.7100	0.6430	0.8495	0.7574
ResNet50	0.8273	0.7756	0.7267	0.6622	0.8586	0.7701
ResNet101	0.8295	0.7774	0.7298	0.6643	0.8530	0.7708
ResNet152	0.8312	0.7802	0.7305	0.6641	0.8629	0.7738
EfficientNetB5	0.8475	0.8005	0.7411	0.6713	0.8734	0.7868

entNetB5, com média de AUROC de 0.7868. A DenseNet169 obteve o melhor resultado geral em quatro das cinco classes: Cardiomegaly, Edema, Consolidation e Atelectasis. A EfficientNetB5 obteve o melhor resultado na classe Pleural Effusion.

Após a escolha das redes neurais que são utilizadas como extratoras de características, realizamos experimentos para verificar qual módulo de atenção obtém o melhor resultado com a DCNN DenseNet169 e EfficientNetB5, considerando a base de dados CheXpert [18]. Apresentamos, na Tabela 5.7, os resultados das DCNNs escolhidas como extratoras de características com os módulos de atenção CAM, SAM e FPA. A DenseNet169 obteve melhor resultado com o módulo CAM, alcançando a média de 0.7909 de AUROC, e a EfficientNetB5 também obteve o melhor resultado com o módulo de atenção CAM, obtendo a média de 0.7884 de AUROC.

Tabela 5.7: Comparação entre os resultados da DenseNet169 (D169) e a EfficientNetB5 (Eff) considerando três módulos de atenção diferentes (CAM, SAM e FPA).

Método	Cardiomegaly	Edema	Consolidation	Atelectasis	Pleural Effusion	Média
D169-CAM	0.8501	0.8053	0.7417	0.6860	0.8712	**0.7909**
D169-SAM	0.8396	0.7985	0.7433	0.6741	0.8664	0.7844
D169-FPA	0.8360	0.7914	0.7411	0.6754	0.8608	0.7809
Eff-CAM	0.8499	0.8031	0.7453	0.6722	0.8717	0.7884
Eff-SAM	0.8434	0.8001	0.7396	0.6685	0.8714	0.7846
Eff-FPA	0.8385	0.7966	0.7391	0.6689	0.8733	0.7833

Podemos destacar que o módulo de atenção CAM se destacou mais com a DCNN DenseNet169 do que com a DCNN EfficientNetB5, alcançando uma diferença de, pelo menos, meio ponto percentual para o segundo melhor resultado, enquanto que com a EfficientNetB5, a diferença foi menor que meio ponto percentual para o segundo melhor resultado.

Após a escolha do módulo de atenção de cada extratora de características, realizamos o experimento com o nosso método por completo, DuaLAnet, e comparamos com os métodos da base de dados CheXpert [18] apresentados anteriormente (Capítulo 4). Temos que considerar que, como a base de dados é de uma competição ainda em andamento, o conjunto de teste não está disponível, então não podemos realizar uma comparação justa, pois os conjuntos utilizados para treinamento, validação e teste não são iguais. Além

disso, os trabalhos relacionados que utilizamos sobre essa base de dados não deixam claro se o treinamento foi realizado considerando todas as 14 classes ou somente as 5 classes reportadas.

A Tabela 5.8 apresenta os resultados numéricos dos métodos CheXclusion [31], ConVIRT [50], DenseNet161 [4], CheXpert [18] e da nossa abordagem proposta, DuaLAnet. A nossa abordagem não obteve resultados próximos aos trabalhos com maior média de AUROC, possivelmente devido ao fato de não termos acesso aos mesmos conjuntos de dados. Podemos destacar a classe Cardiomegaly, que alcançamos o resultado praticamente empatado com os demais trabalhos, porém, nas demais classes obtivemos resultados inferiores.

Tabela 5.8: Comparação do nosso método DuaLAnet com as abordagens do estado da arte na base de dados CheXpert [18]. O método ConVIRT [50] reporta somente a média final da métrica AUROC para as 5 classes.

Método	**Cardiomegaly**	**Edema**	**Consolidation**	**Atelectasis**	**Pleural Effusion**	**Média**
CheXclusion [31]	**0.855**	0.849	0.734	0.717	0.885	0,808
ConVIRT [50]	-	-	-	-	-	0.881
DenseNet161 [4]	0.836	0.920	0.917	0.802	**0.937**	0.882
CheXpert [18]	0.854	**0.928**	**0.937**	**0.821**	0.936	**0.895**
DuaLAnet	0.853	0.829	0.784	0.712	0.886	0.813

A Figura 5.9 apresenta as quatro melhores configurações de parâmetros que experimentamos no treinamento do nosso modelo. Podemos destacar a configuração que obteve o melhor resultado, Config1, que considera os valores do ramo de fusão com peso 1.0, os valores da DenseNet169 com o módulo de atenção CAM com peso 0.7 e os valores da EfficientNetB5 com o módulo de atenção CAM com peso 0.4. Assim como mostrado no resultado da base de dados ChestX-ray14 [48], o módulo de fusão teve um papel fundamental no método.

A Figura 5.10 mostra a variação da taxa de perda ao longo das 30 épocas no treinamento do ramo de fusão do nosso método, que leva em consideração a taxa de perda dos dois ramos de extração de características. A menor taxa de perda, no conjunto de validação, foi alcançada na época 9 e foi o modelo que escolhemos para a realização da etapa de teste. Assim como na Figura 5.8 (base de dados ChestX-ray14 [48]), a taxa de perda da validação não teve uma grande variação, diferente da taxa de perda na etapa de treinamento, que foi convergindo.

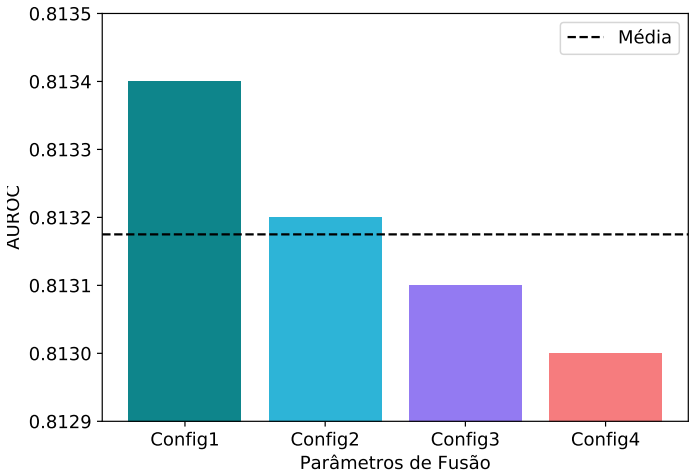


Figura 5.9: Média de AUROC para o método DuaLAnet variando os valores de γ_1, γ_2 e γ_3. Os valores dos parâmetros são: Config1 ($\gamma_1 = 1.0, \gamma_2 = 0.7, \gamma_3 = 0.4$), Config2 ($\gamma_1 = 1.0, \gamma_2 = 0.5, \gamma_3 = 0.5$), Config3 ($\gamma_1 = 0.5, \gamma_2 = 0.5, \gamma_3 = 0.5$) e Config4 ($\gamma_1 = 0.3, \gamma_2 = 0.3, \gamma_3 = 0.5$).

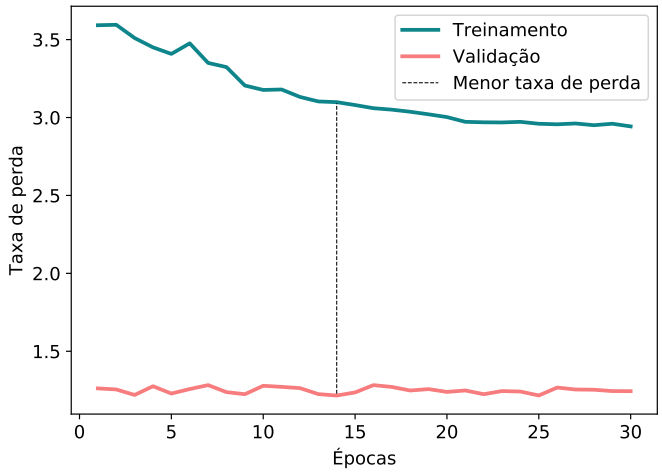


Figura 5.10: Taxa de perda das etapas de treinamento e validação do método DuaLAnet (ramo de fusão).

Capítulo 6

Conclusões e Trabalhos Futuros

Considerando o contexto desta pesquisa, que trata do exame de radiografias de tórax, um dos mais comumente utilizados para a identificação e diagnóstico de diversas doenças potencialmente fatais, desenvolvemos o método DuaLAnet que utiliza técnicas de aprendizado profundo para obter resultados competitivos com o estado da arte na classificação de bases de dados públicas de radiografias de tórax. Este método foi o foco do artigo intitulado "*DuaLAnet: Dual Lesion Attention Network for Thoracic Disease Classification in Chest X-Rays*" [42], aceito na *27th International Conference on Systems, Signals and Image Processing* (IWSSIP). O trabalho foi apresentado na conferência, em julho de 2020.

O nosso método, DuaLAnet, é composto de três ramos principais, sendo dois deles responsáveis pela extração de características das imagens de entrada e um ramo utilizado para a fusão dos vetores de características dos ramos anteriores. Além disso, utilizamos módulos de atenção nos ramos de extração de características para direcionar o aprendizado das redes neurais e para que o método pudesse obter uma taxa de acerto mais elevada. Utilizamos as bases de dados ChestX-ray14 [48], a principal base de nosso estudo, e CheXpert [18], que consideramos como um estudo de caso extra, para validar e comparar com os resultados dos métodos do estado da arte.

Para cada base de dados, realizamos experimentos para obter a configuração, entre as opções pré-definidas de redes neurais e módulos de atenção, que melhor se ajusta com o método DuaLAnet. Primeiramente, verificamos qual o método de inicialização dos pesos das redes neurais que obtém o melhor resultado nas duas bases de dados. Para isso, treinamos uma rede neural com os pesos inicializados a partir da ImageNet [9] e a partir de uma das bases de radiografias de tórax (a base que não utilizamos no treinamento do método) e verificamos qual inicialização obtinha maior taxa de acerto.

Após a escolha do método de inicialização dos pesos, verificamos quais as arquiteturas obtêm os melhores resultados separadamente, utilizando a métrica AUROC. Assim, definimos as redes neurais que seriam utilizadas nos ramos de extração de características para cada base de dados. Após isso, experimentamos os três tipos de módulos de atenção em cada rede neural para escolher o que melhor se adequava. Por fim, realizamos os experimentos com o método DuaLAnet completo em cada base de dados.

A seguir, apresentamos as respostas das questões de pesquisa introduzidas no Capítulo 1:

1. É possível obter resultados competitivos com os trabalhos recentes em classificação de imagens de radiografias de tórax utilizando redes neurais profundas?

   Com a utilização do método DuaLAnet, obtivemos resultados competitivos com os trabalhos do estado da arte na classificação de imagens de radiografias de tórax na base de dados ChestX-ray14 [48], praticamente empatado com o melhor método. Na base de dados CheXpert [18], vimos que temos alguns caminhos que podemos seguir para obter melhores resultados.

2. Como os pesos de redes pré-treinadas na base ImageNet e, em seguida, aperfeiçoados para o domínio médico, comparam-se aos pesos aprendidos somente a partir de imagens médicas?

   Com relação à inicialização dos pesos, destacamos que não obtivemos uma resposta definitiva sobre qual o melhor método, considerando a configuração dos nossos experimentos, pois na base de dados ChestX-ray14 [48], a inicialização a partir do pré-treinamento com a base CheXpert [18] obteve melhor resultado, entretanto, este fato não foi constatado no experimento com a outra base, que obteve melhor resultado com inicialização dos pesos realizada a partir da ImageNet [9].

3. Como a complementaridade entre redes neurais profundas e técnicas de atenção podem contribuir na nossa abordagem?

   Destacamos que a complementaridade entre as redes neurais e os módulos de atenção contribuiu para que o nosso método obtivesse resultados competitivos com os trabalhos do estado da arte. Isto pode ser observado na evolução dos experimentos e na diferença da taxa de acerto AUROC da utilização de uma rede neural pré-treinada, para a rede neural com o módulo de atenção e, por fim, o método completo.

6.1 Contribuições

As principais contribuições do trabalho desenvolvido durante o mestrado são listadas a seguir:

- Proposição de um método de aprendizado profundo, denominado DuaLAnet, que usa a complementaridade entre redes neurais profundas e métodos de atenção para classificar imagens de radiografia de tórax.
- Proposição de um método de treinamento baseado na unificação da taxa de perda de três ramos de classificação em um só valor.
- Comparação entre três módulos de atenção para verificar o que melhor se adequa ao método principal.
- Comparação entre métodos de inicialização dos pesos das redes neurais.

6.2 Trabalhos Futuros

Como sugestões para trabalhos futuros, pretendemos utilizar mais variações de aumentação de dados, como os utilizados pelos trabalhos relacionados à base de dados CheXpert [18], que fazem uso de transformações como ajustes de brilho e contraste. Além disso, um pré-processamento com a base de dados CheXpert [18] poderia ser realizado, no treinamento, validação e teste, para considerar apenas as cinco classes que são reportadas nos trabalhos existentes na literatura. Após esse pré-processamento, pode-se verificar o comportamento do nosso método com esse conjunto de dados.

Ainda sobre trabalhos futuros utilizando a base de dados CheXpert [18], pretendemos realizar um estudo sobre o tratamento dos rótulos incertos. Nos experimentos apresentados nesta dissertação, definimos os rótulos incertos como negativos. Entretanto, experimentos com outras técnicas de substituição devem ser realizados para verificar qual delas que melhor se adapta a esse tipo de informação.

Considerando o método de forma geral, seria interessante realizar experimentos utilizando as imagens de entrada em tamanho original ou, pelo menos, 1024×1024 *pixels*, para verificar se o redimensionamento está fazendo com que informações importantes para a classificação sejam perdidas ou distorcidas a ponto de interferir consideravelmente no resultado final de predição das classes. Os recursos computacionais disponíveis não foram suficientes para a realização desses experimentos.

Referências Bibliográficas

[1] S. Asif, Y. Wenhui, H. Jin, Y. Tao, and S. Jinhai. Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Networks. *medRxiv*, 2020.

[2] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan. Deep learning with non-medical training used for chest pathology identification. In *Medical Imaging 2015: Computer-Aided Diagnosis*, volume 9414, page 94140V, 2015.

[3] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum Learning. In *26th Annual International Conference on Machine Learning*, pages 41–48, 2009.

[4] K. K. Bressem, L. C. Adams, C. Erxleben, B. Hamm, S. M. Niehues, and J. L. Vahldiek. Comparing different deep learning architectures for classification of chest radiographs. *Scientific Reports*, 10(1):1–16, 2020.

[5] B. Chen, J. Li, X. Guo, and G. Lu. DualCheXNet: Dual Asymmetric Feature Learning for Thoracic Disease Classification in Chest X-rays. *Biomedical Signal Processing and Control*, 53:101554, 2019.

[6] C. Chen. *Computer Vision in Medical Imaging*, volume 2. World Scientific, 2014.

[7] K. Chowdhary. Natural Language Processing. In *Fundamentals of Artificial Intelligence*, pages 603–649. Springer, 2020.

[8] J. P. Cohen, P. Morrison, and L. Dao. COVID-19 Image Data Collection, 2020.

[9] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE, 2009.

[10] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Pearson,, 2012.

[11] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT press, 2016.

[12] Q. Guan and Y. Huang. Multi-Label Chest X-ray Image Classification via Category-wise Residual Attention Learning. *Pattern Recognition Letters*, 2018.

[13] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, and D. Comaniciu. Learning to Recognize Abnormalities in Chest X-Rays with Location-Aware Dense Networks. In *Iberoamerican Congress on Pattern Recognition*, pages 757–765, 2018.

[14] A. Halevy, P. Norvig, and F. Pereira. The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, 24(02):8–12, 2009.

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[16] J. Hu, L. Shen, and G. Sun. Squeeze-and-Excitation Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141, 2018.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[18] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In *AAAI Conference on Artificial Intelligence*, volume 33, pages 590–597, 2019.

[19] A. E. W. Johnson, T. J. Pollard, N. R. Greenbaum, M. P. Lungren, C. Deng, Y. Peng, Z. Lu, R. G. Mark, S. J. Berkowitz, and S. Horng. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv:1901.07042*, 2019.

[20] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*, 2014.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6):84–90, 2017.

[22] D. Kumar, A. Wong, and D. A. Clausi. Lung Nodule Classification Using Deep Features in CT Images. In *12th Conference on Computer and Robot Vision*, pages 133–138. IEEE, 2015.

[23] Y. Lei, Y. Tian, H. Shan, J. Zhang, G. Wang, and M. K. Kalra. Shape and Margin-aware Lung Nodule Classification in Low-Dose CT Images via Soft Activation Mapping. *Medical Image Analysis*, 60:1–13, 2020.

[24] H. Li, P. Xiong, J. An, and L. Wang. Pyramid Attention Network for Semantic Segmentation. *arXiv:1805.10180*, 2018.

[25] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally. Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training. *arXiv:1712.01887*, 2017.

[26] D. Lu and Q. Weng. A Survey of Image Classification Methods and Techniques for Improving Classification Performance. *International Journal of Remote Sensing*, 28(5):823–870, 2007.

[27] L. Lu, Y. Zheng, G. Carneiro, and L. Yang. Deep Learning and Convolutional Neural Networks for Medical Image Computing. *Advances in Computer Vision and Pattern Recognition*, 2017.

[28] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, and K. Shpanskaya. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv:1711.05225*, 2017.

[29] S. Raschka and V. Mirjalili. *Python Machine Learning, 2nd Ed.* Packt Publishing, Birmingham, UK, 2nd edition, 2017.

[30] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In *IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[31] L. Seyyed-Kalantari, G. Liu, M. McDermott, and M. Ghassemi. CheXclusion: Fairness gaps in deep chest X-ray classifiers. *arXiv:2003.00827*, 2020.

[32] J. Shen, W. Li, S. Deng, and T. Zhang. Supervised and unsupervised learning of directed percolation. *preprint arXiv:2101.06392*, 2021.

[33] Y. Shen and M. Gao. Dynamic Routing on Deep Neural Network for Thoracic Disease Classification and Sensitive Area Localization. In *International Workshop on Machine Learning in Medical Imaging*, pages 389–397, 2018.

[34] C. Shorten and T. M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):1–48, 2019.

[35] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*, 2014.

[36] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva. An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Computer Methods and Programs in Biomedicine*, 177:285–296, 2019.

[37] C. Sun, A. Shrivastava, S. Singh, and A. Gupta. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In *12th International Conference on Computer Vision (ICCV)*, pages 843–852, 2017.

[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going Deeper with Convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[39] M. Tan and Q. Le. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[40] Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers. Attention-Guided Curriculum Learning for Weakly Supervised Classification and Localization of Thoracic Diseases on Chest Radiographs. In *International Workshop on Machine Learning in Medical Imaging*, pages 249–258, 2018.

[41] P. P. Team, J. K. Gohagan, P. C. Prorok, R. B. Hayes, and B. Kramer. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: History, organization, and status. *Controlled Clinical Trials*, 21(6):251S–272S, 2000.

[42] V. Teixeira, L. Braz, H. Pedrini, and Z. Dias. DuaLAnet: Dual Lesion Attention Network for Thoracic Disease Classification in Chest X-Rays. In *International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 69–74. IEEE, 2020.

[43] N. Van Quang, J. Chun, and T. Tokuyama. CapsuleNet for Micro-Expression Recognition. In *14th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–7. IEEE, 2019.

[44] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need. *arXiv:1706.03762*, 2017.

[45] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, 2018.

[46] H. Wang, H. Jia, L. Lu, and Y. Xia. Thorax-Net: An Attention Regularized Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography. *Journal of Biomedical and Health Informatics*, pages 475–485, 2019.

[47] S. Wang. Artificial Neural Network. In *Interdisciplinary Computing in Java Programming*, pages 81–100. Springer, 2003.

[48] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. ChestX-ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *arXiv:1705.02315*, 2017.

[49] X. Yin, J. Goudriaan, E. A. Lantinga, J. Vos, and H. J. Spiertz. A Flexible Sigmoid Function of Determinate Growth. *Annals of Botany*, 91(3):361–371, 2003.

[50] Y. Zhang, H. Jiang, Y. Miura, C. D. Manning, and C. P. Langlotz. Contrastive Learning of Medical Visual Representations from Paired Images and Text. *arXiv:2010.00747*, 2020.

[51] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2921–2929, 2016.

[52] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, D. Shen, and P. Yap. Multi-Task Learning for Segmentation and Classification of Tumors in 3D Automated Breast Ultrasound Images. *Medical Image Analysis*, page 101918, 2020.