

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

Exame de Qualificação de Mestrado

17 de outubro de 2019

CLASSIFICAÇÃO DE IMAGENS DE RADIOGRAFIAS DE TÓRAX
UTILIZANDO REDES NEURAS PROFUNDAS

Candidato: Vinicius Teixeira de Melo

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

Sumário

1	Introdução	1
1.1	Caracterização do Problema	1
1.2	Objetivos e Contribuições	2
1.3	Questões de Pesquisa	2
1.4	Organização do Texto	3
2	Conceitos e Trabalhos Relacionados	4
2.1	Fundamentação Teórica	4
2.1.1	Rede Neural Convolutacional Profunda	4
2.1.2	Rede Neural Recorrente	8
2.1.3	Transferência de Aprendizado	8
2.1.4	Aumentação de Dados	9
2.2	Trabalhos Correlatos	9
3	Material e Métodos	12
3.1	Metodologia	12
3.1.1	Pré-Processamento	13
3.1.2	Treinamento e Validação	13
3.1.3	Extração de Características	13
3.1.4	Classificação dos Vetores	14
3.2	Métricas de Avaliação	14
3.2.1	Acurácia	14
3.2.2	AUROC	14
3.3	Bases de Dados	15
3.4	Recursos Computacionais	18
4	Resultados Preliminares	19
5	Plano de Trabalho	21
	Bibliografia	23

Resumo

Radiografia de tórax é um dos exames mais comuns em radiologia, sendo utilizada, por exemplo, para diagnosticar e monitorar o tratamento de diversas condições pulmonares, como pneumonia, enfisema e câncer. A identificação e classificação automatizada desse tipo de exame, no mesmo nível ou melhor do que os radiologistas, poderia permitir avanços em diversos ambientes médicos, desde melhoria na velocidade do resultado até iniciativas de saúde global. Neste projeto, nós propomos uma metodologia para classificação de patologias em imagens de radiografias de tórax a partir de uma arquitetura baseada em Rede Neural Profunda. Esta é baseada na extração dos mapas de ativações da rede neural para fazer o recorte das imagens considerando somente as áreas de interesse, descartando informações irrelevantes para a sua classificação. Para validar essa metodologia desenvolvida, os experimentos serão realizados utilizando as bases de dados públicas.

Capítulo 1

Introdução

Radiografia de tórax é um dos exames mais comuns globalmente, 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.

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 [6, 19]. Um requisito fundamental para os avanços na área de análise automatizada de radiografia de tórax são as grandes bases de dados disponíveis, como a *CheXpert* [10] e a *MIMIC-CXR* [11].

Neste capítulo, nós descrevemos o problema que será abordado neste projeto, os objetivos e contribuições esperadas, as questões de pesquisa que irão nos guiar nesta pesquisa e a organização textual do projeto.

1.1 Caracterização do Problema

Radiografias de tórax são os exames mais comuns em radiologia, as quais são essenciais para o gerenciamento de vários tipos de doenças que possuem uma alta taxa de mortalidade, e também para mostrar uma gama de informações sobre potenciais patologias, muitas das quais são sutis.

A maioria das pesquisas de detecção e diagnóstico auxiliados por computador na radiografia de tórax focou a detecção de nódulos pulmonares [4]. Embora seja o alvo da atenção da maioria das pesquisas, os nódulos pulmonares são um achado relativamente raro nos pulmões. As patologias mais comuns encontradas em radiografias de tórax incluem infiltrações no pulmão, dispositivos de suporte, como os cateteres, e anormalidades de tamanho ou contorno do coração [22]. Distinguir as diferentes patologias torácicas é uma tarefa difícil, mesmo para os radiologistas. Portanto, há um interesse no desenvolvimento de sistemas computacionais para o auxílio no diagnóstico de patologias existentes em imagens de radiografias de tórax.

Dado que as radiografias de tórax são rotineiramente usadas para detectar várias anormalidades ou doenças, a interpretação de uma imagem desse tipo requer conhecimentos e recursos de tempo que nem sempre estão disponíveis, especialmente por causa do grande número de exames que precisam ser relatados todos os dias. Isso leva a erros de diagnóstico que, para algumas patologias, estima-se que estejam entre 14 e 38% [21].

Por conta das mortes e custos médicos causados por algumas dessas patologias, as fundações e universidades estão fazendo um grande esforço para, por meio de pesquisas, investigar novas técnicas que possibilitem auxiliar os radiologistas no diagnóstico ou monitoramento de doenças na região do tórax [26].

Devido à complexidade do problema, várias abordagens distintas podem ser encontradas na literatura [15, 23, 25]. Segundo Bar et al. [1], métodos de aprendizagem profunda são mais eficazes quando aplicados em grandes conjuntos de treinamento. Em geral, dado um conjunto de imagens de radiografias de tórax, uma rede de aprendizado profundo é treinada e então é aplicada em amostras de teste para verificar a qualidade dos resultados obtidos para este tipo de imagem.

1.2 Objetivos e Contribuições

O objetivo 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:

- Estudar as abordagens do estado da arte utilizadas na classificação de imagens;
- Propor uma solução baseada em extração das áreas de interesse em imagens de radiografias de tórax utilizando rede neural profunda;
- Validar o método proposto em diferentes bases de dados;
- Comparar e avaliar o método proposto com outras abordagens disponíveis;
- Publicar os resultados.

Este projeto visa contribuir com o desenvolvimento de uma metodologia para classificação de patologias em imagens de radiografias de tórax a partir da extração de áreas de interesse das imagens de entrada, descartando informações irrelevantes, e também considerando a independência entre as classes.

1.3 Questões de Pesquisa

Nesta seção, nós apresentamos as questões de pesquisa que irão nos guiar neste projeto:

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

- Considerando a classificação multirrótulo e a divisão em classificadores binários, qual método obtém melhores resultados no problema de classificação de imagens de radiografia de tórax?
- Quais são as implicações de usar imagens de alta resolução como entrada para a rede neural?
- Como os pesos de redes pré-treinadas na base *ImageNet* e, em seguida, aperfeiçoados para o domínio médico, se comparam aos pesos aprendidos somente a partir de imagens médicas?
- Como tratar a incerteza sobre os rótulos em imagens da base de dados *CheXpert* [10]?
- É possível generalizar nosso método para outras bases de dados de radiografias?

1.4 Organização do Texto

Este projeto está organizado em 5 capítulos. No Capítulo 1, nós descrevemos o problema de pesquisa a ser abordado neste projeto e também os principais objetivos e contribuições esperadas. No Capítulo 2, nós introduzimos conceitos importantes para o entendimento do projeto e apresentamos alguns trabalhos relacionados relevantes existentes na literatura. No Capítulo 3, nós apresentamos a metodologia, bases de dados e métricas que serão utilizadas para avaliação dos nossos modelos. No Capítulo 4, nós apresentamos alguns resultados preliminares obtidos. E, por fim, no Capítulo 5, apresentamos nosso plano de trabalho e cronograma.

Capítulo 2

Conceitos e Trabalhos Relacionados

Neste capítulo, nós fazemos uma breve descrição dos conceitos relacionados às redes neurais convolucionais e recorrentes, entre outras técnicas, que servem como base para o entendimento deste projeto e dos trabalhos correlatos. Além disso, nós apresentamos também alguns resultados relevantes existentes na literatura e como eles são relacionados ao nosso trabalho.

2.1 Fundamentação Teórica

Nesta seção, nós apresentamos alguns conceitos e técnicas utilizadas para a classificação de imagens, a fim de fornecer uma melhor fundamentação para a compreensão dos trabalhos correlatos e deste projeto.

2.1.1 Rede Neural Convolucional Profunda

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 [16].

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 2.1 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, onde os mesmos pesos são usados para diferentes partes da imagem de entrada. Normalmente, as DCNNs são formadas por várias camadas

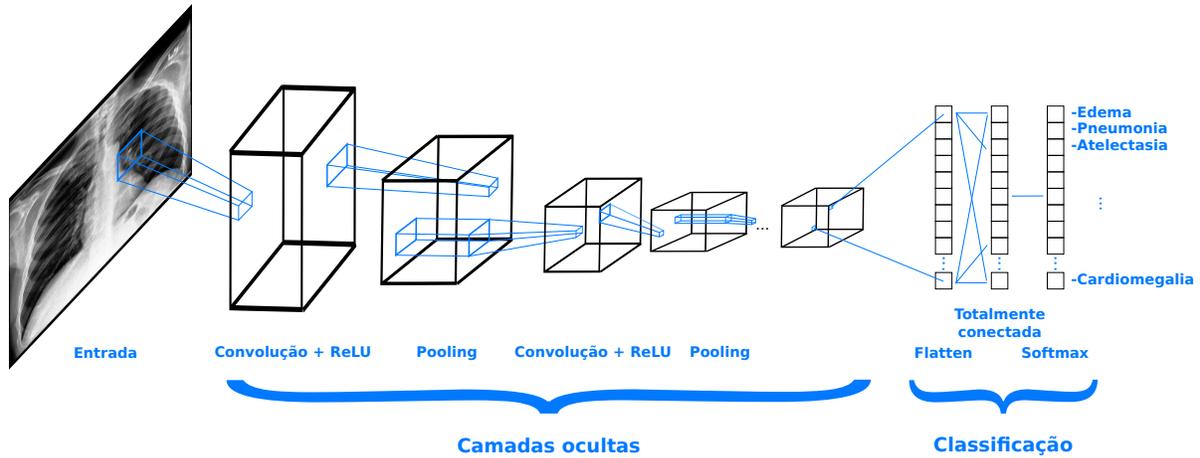


Figura 2.1: Etapas de um processo de classificação de uma Rede Neural Convolutiva Profunda.

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

A seguir, são mostradas cinco arquiteturas de DCNNs que são utilizadas por trabalhos na área de classificação de imagens de radiografia de tórax.

AlexNet

Krizhevsky et al. [12] propuseram a DCNN *AlexNet* em 2012 e ganharam a competição *ImageNet Large Scale Visual Recognition Challenge*, alcançando uma taxa de erro de 15,3% (*top-5 score*), em comparação com 26,2% alcançada pelo segundo colocado. A Figura 2.2 apresenta a arquitetura da rede, que contém 5 camadas convolucionais e 3 camadas totalmente conectadas. A função de ativação *Rectified Linear Unit* (ReLU) é aplicada depois de toda camada convolutiva e totalmente conectada. A técnica de *Dropout* [18] é aplicada antes da primeira e da segunda camada totalmente conectada.

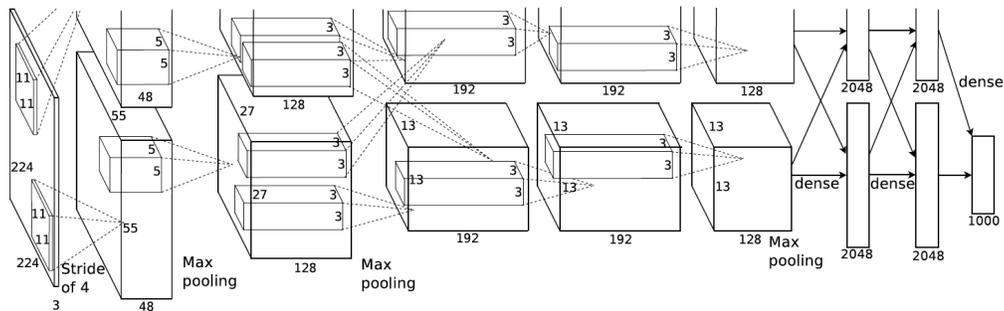


Figura 2.2: Arquitetura da rede *AlexNet* [12].

A *AlexNet* é considerada como um marco da utilização de CNNs para classificação de

imagens e, ainda hoje, técnicas utilizadas nessa rede neural são utilizadas, como por exemplo: camadas convolucionais seguidas de camadas de *pooling*, *Dropout* e ReLU. A *AlexNet* não é profunda o suficiente em comparação com modelos posteriores, como a *VGGNet*, *GoogLeNet* e *ResNet*, e vale destacar que o uso de filtros de convolução com dimensões grandes (5×5 e 11×11) não é mais encorajado hoje em dia.

GoogLeNet

Szegedy et al. [20] propuseram uma nova arquitetura de DCNN *GoogLeNet*, também conhecida como *Inception*, que foi considerada estado da arte na competição de classificação da *ImageNet* em 2014. A vantagem principal dessa rede está no módulo *inception*, mostrado na Figura 2.3, que busca cobrir uma área maior das imagens, mas também manter uma boa resolução para pequenas informações. Assim, é necessário realizar convoluções em paralelo com *filtros de Gabor* de diferentes tamanhos. A *GoogLeNet* possui 9 módulos *inception*.

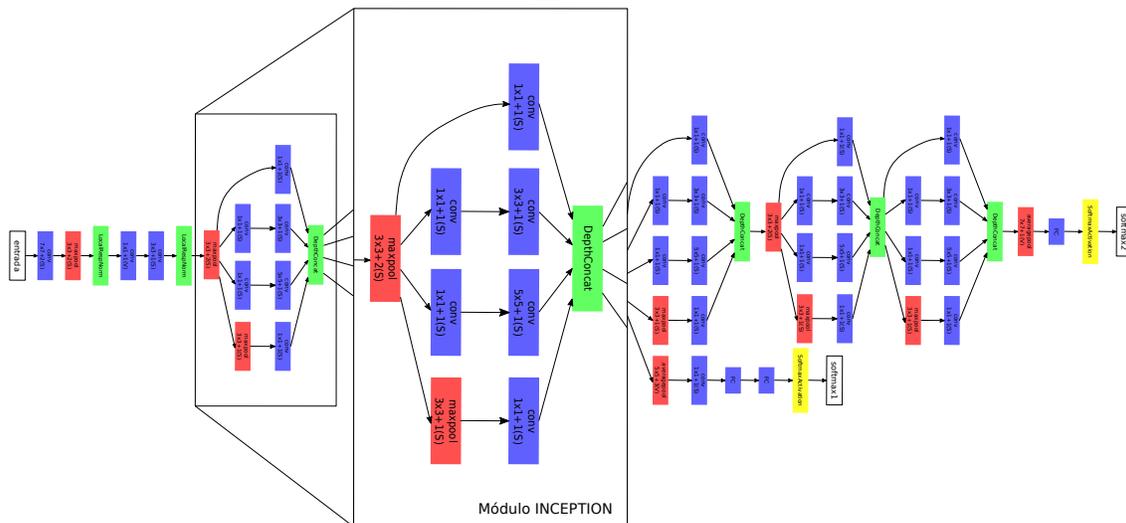


Figura 2.3: Arquitetura da rede *GoogLeNet* [20].

Entre outras vantagens da rede *GoogLeNet*, podemos citar o baixo tempo de treinamento em comparação com a *VGGNet*, por exemplo, e o menor número de parâmetros utilizados na rede, de 60 milhões utilizados pela *AlexNet*, a *GoogLeNet* utiliza somente cerca de 7 milhões.

VGGNet-16

Simonyan e Zisserman [17] propuseram a DCNN *VGG-16* e ganharam a competição de localização da *ImageNet* em 2014. Essa rede é um aprimoramento em relação à rede *AlexNet*, 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.

ResNet

He et al. [7] propuseram a DCNN *ResNet* e ganharam a competição de classificação, localização e detecção da *ImageNet* em 2015. Diferente das redes apresentadas anteriormente, a *ResNet* busca aprender a partir dos resíduos obtidos pelas camadas anteriores, como mostra a Figura 2.4. 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.

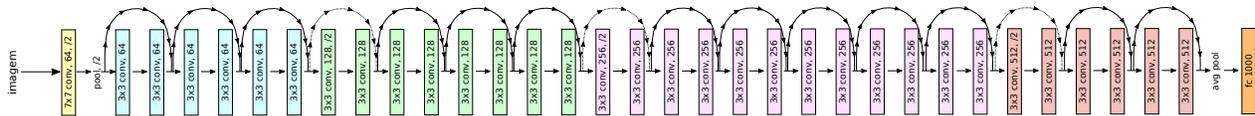


Figura 2.4: Modelo da arquitetura de uma rede *ResNet*.

DenseNet

Huang et al. [9] propuseram a DCNN *DenseNet* que tem como objetivo principal garantir o 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 2.5 ilustra esse modelo esquematicamente, onde só são utilizados os recursos que já foram calculados.

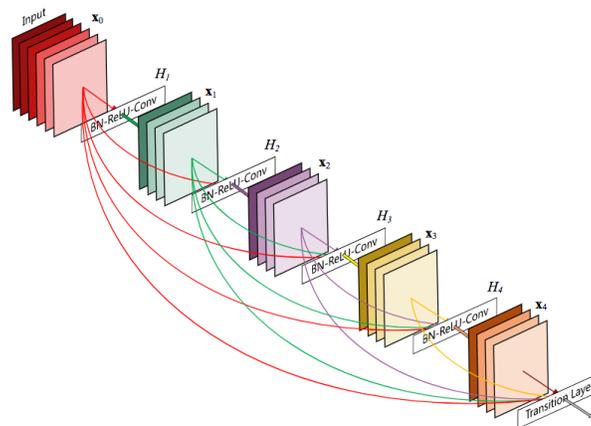


Figura 2.5: Parte da arquitetura da rede *DenseNet* [9].

Podemos citar, como vantagens da rede *DenseNet*, a redução do problema de desaparecimento do gradiente, a melhor propagação de informações tanto no *feedforward* quanto no *backpropagation*, a reutilização de recursos da rede e a redução do número de parâmetros.

2.1.2 Rede Neural Recorrente

Redes Neurais Recorrentes (*Recurrent Neural Network*, ou RNN) são algoritmos utilizados para o processamento de informações sequenciais ou dependentes. Uma RNN aplica uma computação para toda instância de entrada baseada nos resultados obtidos previamente. A Figura 2.6 apresenta uma exemplificação do funcionamento de uma RNN, onde os resultados computados anteriormente são memorizados e, então, são utilizados para computar a informação corrente.

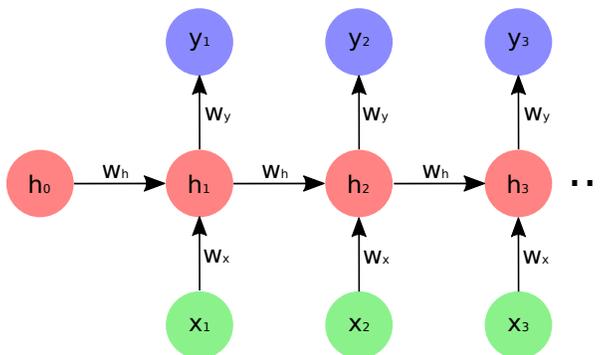


Figura 2.6: Exemplo de RNN.

As RNNs podem se mostrar melhores do que as CNNs em problemas de processamento de linguagem natural, mas não são necessariamente superiores. Isso se dá pelo fato de que o modelo escolhido deve ser baseado na semântica do problema. A entrada esperada por uma RNN é tipicamente um vetor codificado, mas em alguns casos eles são acoplados com as representações abstratas construídas por um modelo CNN.

Um dos principais problemas das RNNs é a etapa de treinamento, onde a retropropagação é geralmente custosa devido a problemas conhecidos como desaparecimento ou explosão de gradiente. O problema de desaparecimento de gradiente acontece quando o gradiente se torna tão pequeno que a RNN não converge. Por outro lado, a explosão de gradiente acontece quando o valor do gradiente torna-se muito alto.

Memória de Curto Longo Prazo

Memória de Curto Longo Prazo (*Long Short-Term Memory*, ou LSTM) é um tipo de RNN proposta por Hochreiter e Schmidhuber [8], a qual é capaz de aprender dependências de longo prazo. As LSTMs são também um tipo de célula de memória, que também possui uma estrutura de cadeia, mas o módulo de repetição possui uma estrutura diferente. A principal vantagem da LSTM é que essas operações aplicadas em cada célula de memória desse tipo permitem que as informações sejam mantidas ou esquecidas.

2.1.3 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 mo-

delos com foco em resolver outros problemas. Essa é uma abordagem muito popular em Aprendizado Profundo, onde modelos pré-treinados são utilizados como pontos de partida em problemas de Visão Computacional ou Processamento de Linguagem Natural [5].

2.1.4 Aumentação de Dados

Aumentação de Dados (*Data Augmentation*) é um método utilizado para solucionar 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 aumento de dados utilizam transformações como rotação, translação, escala e transformações de intensidade em amostras existentes [13]. A Figura 2.7 apresenta um exemplo de aplicação dessa técnica em uma imagem utilizando transformações como rotação e translação.

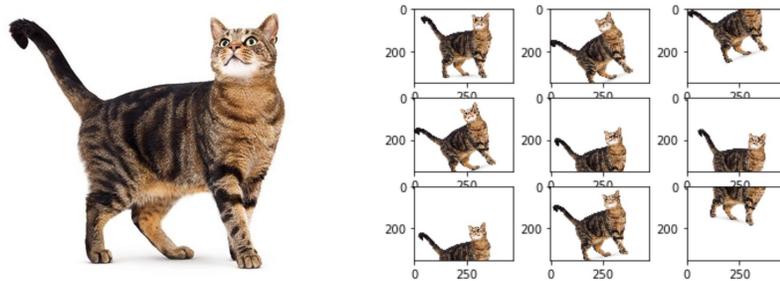


Figura 2.7: Exemplo de Aumentação de Dados.

2.2 Trabalhos Correlatos

Nesta seção, nós descrevemos alguns métodos relacionados a Aprendizado Profundo que são utilizados para a classificação de patologias em imagens de radiografias de tórax.

Wang et al. [23] propuseram um *framework* que tem como objetivo, primeiramente, a verificação da existência de uma ou mais patologias em imagens de radiografia de tórax e, caso exista alguma, localizá-las utilizando as ativações e pesos extraídos da rede neural. Para isso, foi utilizada uma DCNN que possui uma similaridade com diversos métodos de localização de objetos fracamente supervisionados [3, 14].

Como mostra a Figura 2.8, o *framework* proposto utiliza os seguintes modelos pré-treinados com a base de dados *ImageNet* [2]: *AlexNet* [12], *GoogleNet* [20], *VGGNet-16* [17] e *ResNet-50* [7]. Foram removidas as camadas totalmente conectadas e as camadas finais de classificação dessas redes, e foram inseridas uma camada de transição, uma camada de *pooling* global, uma camada de predição e uma camada de perda, após a última camada convolucional.

O método proposto foi avaliado e validado utilizando a base de dados *ChestX-ray8* [23], também proposta por Wang et al. [23], que possui 8 tipos de patologias. A base de dados é composta por 108948 imagens frontais de radiografias de tórax, as quais foram utilizadas 70% para a etapa de treinamento, 10% para validação e 20% para teste. A arquitetura que obteve melhores resultados foi a *ResNet-50*, com *Area Under the Receiver Operating Characteristics* (AUROC) média de 0,696 para classificação e acurácia média de 0,628 para localização.

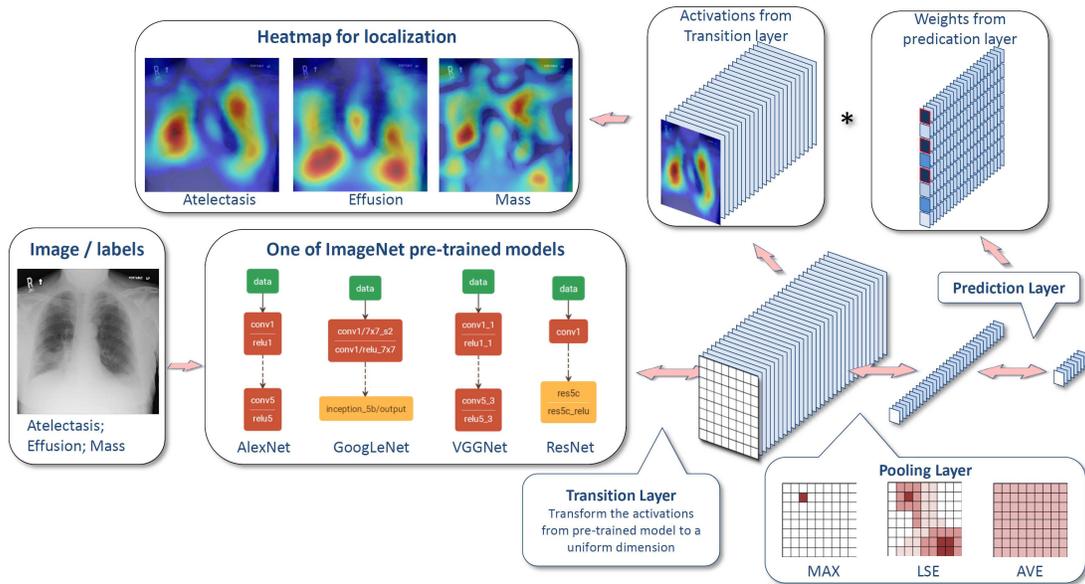


Figura 2.8: Fluxograma geral da estrutura DCNN unificada e o processo de localização das áreas de interesse, proposto por Wang et al. [23].

Yao et al. [25] propuseram uma solução parcial para a escassez de dados de treinamento e maximização da necessidade de extrair características relevantes das amostras disponíveis. Essa solução utiliza a técnica de LSTM para influenciar nas interdependências entre os rótulos na predição de 14 tipos de patologias a partir de imagens de radiografias de tórax.

O modelo proposto por Yao et al. [25] é baseado na arquitetura das redes *DenseNets* [9], levando-se em consideração as características dos problemas médicos em questão. As entradas da rede são imagens de alta resolução e o modelo proposto possui um número menor de camadas.

A Figura 2.9 apresenta a estrutura do modelo proposto. Primeiramente, a imagem de entrada é codificada por uma rede neural convolucional densamente conectada. Essa variação de *DenseNet* consiste em Blocos Densos e Blocos de Transição, onde dentro de cada Bloco Denso, existem vários Blocos Convolucionais. A representação codificada resultante da entrada é um vetor que captura a semântica de ordem mais alta que é útil para a tarefa de decodificação. E, por fim, o decodificador é baseado nas RNNs.

O método proposto foi avaliado utilizando uma base de dados estendida da utilizada no trabalho anterior, denominada *ChestX-ray14* [24], mostrada em mais detalhes na Seção 3.3. O método proposto por Yao et al. [25] obteve AUROC média de 0,802.

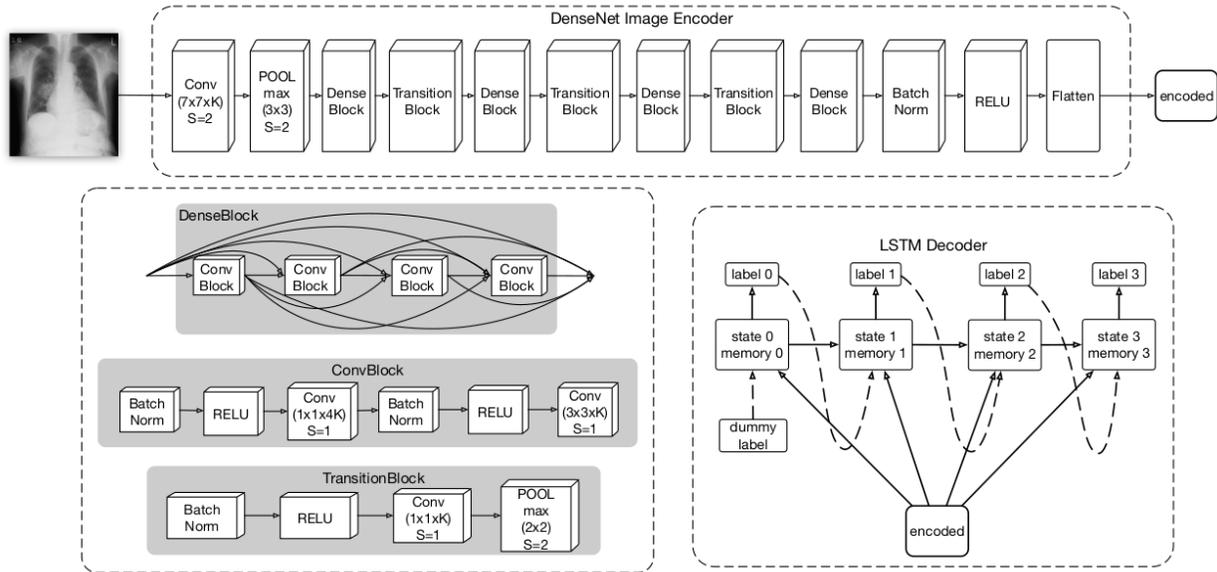


Figura 2.9: Representação do processo de codificação e decodificação baseado na arquitetura da rede *DenseNet*, proposta por Yao et al. [25].

Rajpurkar et al. [15] propuseram um algoritmo para detectar pneumonia em imagens de radiografia de tórax com um nível maior de precisão do que os radiologistas. O modelo proposto, denominado *CheXNet*, é uma DCNN com 121 camadas que calcula a probabilidade de uma dada imagem de radiografia de tórax apresentar pneumonia e, também, apresenta um mapa de localização das áreas da imagem que são mais prováveis de ter essa patologia.

O modelo *CheXNet* é baseado na arquitetura da *DenseNet*, treinada com a base de dados *ChestX-ray14* [24]. Com a utilização da arquitetura *DenseNet*, o fluxo de informações e gradientes através da rede é menos suscetível a percas entre as camadas, tornando a otimização de redes profundas tratável. A última camada totalmente conectada foi substituída por uma de saída única, depois foi aplicada uma função *sigmoid* não-linear.

Para fins de comparação e validação do modelo, Rajpurkar et al. [15] trocaram a abordagem de classificação binária para classificação multirrótulo com os 14 tipos de patologias existentes na base de dados proposta por Wang et al. [24]. Assim, a última camada totalmente conectada da rede passou a ter como saída um vetor de 14 dimensões, onde é aplicada a função *sigmoid* não-linear em cada dimensão do vetor. A saída final da rede é um vetor de probabilidades previstas para cada tipo de patologia.

Primeiramente, as imagens da base de dados foram diminuídas para ficar com dimensão 224×224 *pixels* e normalizadas com base na média e desvio padrão das imagens do conjunto de dados *ImageNet*, e foi aplicada a técnica de Aumentação de Dados sobre as imagens do conjunto de treinamento, com a transformação inversão horizontal aleatória. O modelo proposto por Rajpurkar et al. [15] obteve o resultado de AUROC média de 0,841.

Capítulo 3

Material e Métodos

Neste capítulo, nós descrevemos a metodologia inicial proposta que será utilizada para desenvolver um modelo para classificação de imagens de radiografia de tórax, as métricas usadas para avaliação, bem como as bases de dados utilizadas para a validação do nosso método.

3.1 Metodologia

Nesta seção, nós descrevemos as etapas da metodologia proposta neste projeto, ilustrada na Figura 3.1, e também as técnicas e ferramentas que serão utilizadas em cada uma delas.

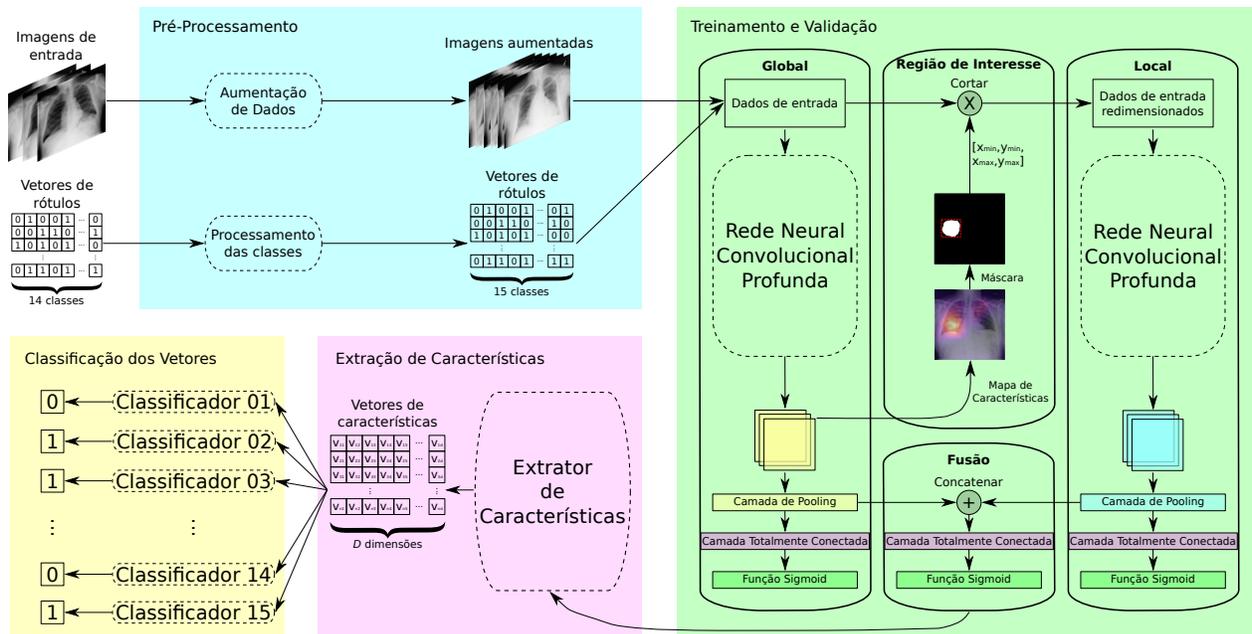


Figura 3.1: Diagrama com as principais etapas que compõem a metodologia deste projeto.

3.1.1 Pré-Processamento

Na etapa de pré-processamento dos dados, temos dois processos principais: a aumentação de dados e o processamento das classes. A aumentação de dados é aplicada somente sobre o conjunto de treinamento, utilizando a transformação Reflexão Horizontal Aleatória (*Random Horizontal Flip*). Todas as imagens são redimensionadas para ficar com dimensões 224×224 pixels e normalizadas com base na média e no desvio padrão da *ImageNet*, por conta da utilização de modelos pré-treinados.

O processamento das classes é realizado somente na base de dados *ChestX-ray14* [24], ao adicionar a classe “Normal”, já que pelo menos metade das imagens da base de dados não são rotuladas por alguma patologia, como mostrado na Seção 3.3. Esse processamento sobre as classes não é aplicado na base de dados *CheXpert* [10], pois a classe “Normal” já pertence ao conjunto de rótulos.

3.1.2 Treinamento e Validação

Na etapa de treinamento e validação, temos 4 módulos principais: o global, o de região de interesse, o local e o de fusão. O módulo global é responsável por treinar e validar uma DCNN utilizando os dados obtidos na etapa de pré-processamento. O módulo de região de interesse é responsável por receber os mapas de ativações do módulo global e gerar os *bounding boxes* das regiões das imagens em que é mais provável existir informações úteis para a classificação, descartando informações irrelevantes para este problema, como bordas e áreas que não são do tórax.

O módulo local recebe como entrada as imagens do módulo global recortadas com base nas informações obtidas no módulo de região de interesse, e redimensionadas para ficar com o mesmo tamanho das imagens de saída da etapa de pré-processamento (224×224 pixels). Ainda no módulo local, uma DCNN é treinada e validada utilizando somente as imagens recortadas. O módulo de fusão é responsável por concatenar as informações obtidas na última camada de *pooling* do módulo global e do módulo local em um único conjunto de dados, que é utilizado para gerar vetores de características que serão utilizados na classificação. Como saída do módulo de fusão, temos os vetores de rótulos das imagens de entrada, considerando o método de classificação multirrótulo.

Vale ressaltar que as DCNNs utilizadas no módulo global e no módulo local não são necessariamente as mesmas. Os otimizadores e as funções de perda serão experimentados para que possamos encontrar os parâmetros e métodos que se ajustam melhor com o problema abordado neste projeto.

3.1.3 Extração de Características

Na etapa de extração de características, iremos considerar as informações obtidas no módulo de fusão para gerar os vetores de características que serão utilizados na etapa de classificação. Para obtermos essas informações, será feito um *flatten* dos dados obtidos com a concatenação da última camada de *pooling* do módulo global e do módulo local.

3.1.4 Classificação dos Vetores

Na etapa de classificação dos vetores de características, iremos considerar as classes independentes, assim, utilizaremos classificadores binários para cada classe. Para escolhermos o classificador que melhor se adequa ao nosso projeto, iremos realizar experimentos alternando o classificador e escolher o que obtiver melhor média de avaliação, com base nas métricas que serão apresentadas na próxima seção. No final desta etapa, teremos um vetor de rótulos para cada imagem de entrada, considerando a classificação binária de cada classe separadamente.

3.2 Métricas de Avaliação

Nesta seção, nós descrevemos as métricas que serão utilizadas para avaliação dos modelos propostos neste projeto.

3.2.1 Acurácia

A acurácia é a métrica de avaliação mais comum em problemas de classificação. Seu valor pode ser calculado como mostrado na Equação 3.1, onde TP é o número de verdadeiros positivos (*True Positives*), TN equivale ao número de falsos negativos (*True Negatives*), P é o número de casos positivos e N é o número de casos negativos. A acurácia de uma classe é o número de casos preditos corretamente com relação a todos os casos existentes.

$$\text{Acurácia} = \frac{TP + TN}{P + N} \quad (3.1)$$

A acurácia usualmente considera todas as classes com o mesmo peso, assim, sendo adequada para conjuntos de dados balanceados. Porém, se o conjunto de dados utilizado não for balanceado, o mais correto é utilizar a acurácia balanceada, que considera cada classe com um peso diferente, de acordo com a quantidade de dados existentes dessa classe. A acurácia balanceada pode ser calculada como mostrado na Equação 3.2, considerando um problema com duas classes.

$$\text{Acurácia balanceada} = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right) \quad (3.2)$$

3.2.2 AUROC

AUROC (*Area Under the Receiver Operating Characteristic*) é uma métrica de desempenho para problemas de classificação em várias 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 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.

Primeiramente, precisamos definir algumas funções para entendermos a métrica AUROC. A Equação 3.3 é uma métrica que mede o sucesso da identificação de casos anormais e a Equação 3.4 é uma métrica que mede o sucesso de não identificar casos normais como anormais, onde FN é o número de casos falsos negativos (*False Negatives*) e FP é o número de casos falsos positivos (*False Positives*).

$$\text{Sensibilidade (recall)} = \frac{TP}{TP + FN} \quad (3.3)$$

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (3.4)$$

A curva ROC tem tipicamente eixo horizontal como $1 - \text{Especificidade}$ (*False Positive Rate*, ou FPR) e eixo vertical como Sensibilidade (*True Positive Rate*, ou TPR). Então, à medida que aumentamos o limiar, diminuímos a TPR e a FPR, e quando diminuimos o limiar, estamos aumentando a TPR e a FPR. Portanto, a AUROC indica o quão bem as probabilidades das classes positivas são separadas das classes negativas. A Figura 3.2 apresenta o resultado gráfico da métrica AUROC obtido no trabalho proposto por Wang et al. [23].

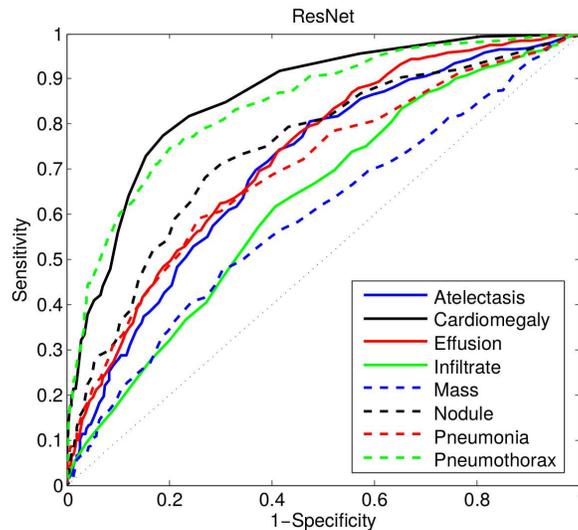


Figura 3.2: Exemplo de AUROC obtido com a DCNN *ResNet-50* [23].

3.3 Bases de Dados

Este projeto utilizará, inicialmente, duas bases de dados públicas contendo imagens de radiografias de tórax: *CheXpert* [10] e *ChestX-ray14* [24]. A seguir, são descritas as duas bases de dados.

CheXpert [10] é uma base de dados de imagens de radiografia de tórax disponibilizada

para um desafio¹ público de Aprendizado de Máquina, consistindo em 224316 imagens de 65240 pacientes. 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 3.1, sendo que cada uma pode ser inferida como positiva, incerta ou negativa. 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. A Figura 3.3 apresenta exemplos de imagens de radiografia de um mesmo paciente.

Classe	Treinamento			Validação			Teste		
	Positivo	Incerto	Negativo	Positivo	Incerto	Negativo	Positivo	Incerto	Negativo
Normal	15576	0	141051	2163	0	20408	4680	0	39770
Cardiomeiastinal aumentado	7632	8596	140399	1079	1358	20134	2196	2449	39805
Cardiomegalia	18840	5657	132130	2783	828	18960	5445	1602	37403
Opacidade pulmonar	74154	3837	78636	10766	559	11246	20787	1202	22461
Lesão pulmonar	6434	1058	149135	963	141	21467	1790	289	42371
Edema pulmonar	36775	9211	110641	5324	1241	16006	10192	2532	31726
Consolidação pulmonar	10412	19472	126743	1479	2822	18270	2925	5448	36077
Pneumonia	4277	13184	139166	625	1975	19971	1145	3611	39694
Atelectasia	23464	23588	109575	3317	3377	15877	6675	6774	31001
Pneumotórax	13564	2187	140876	1974	333	20264	3918	625	39907
Derrame pleural	60147	8217	88263	8746	1098	12727	17361	2313	24776
Outros problemas pleurais	2409	1834	152384	346	248	21977	769	571	43110
Fratura	6329	471	149827	935	61	21575	1776	110	42564
Dispositivos de suporte	81363	758	74506	11893	107	10571	22852	214	21384

Tabela 3.1: Divisão da base de dados *CheXpert* [10].

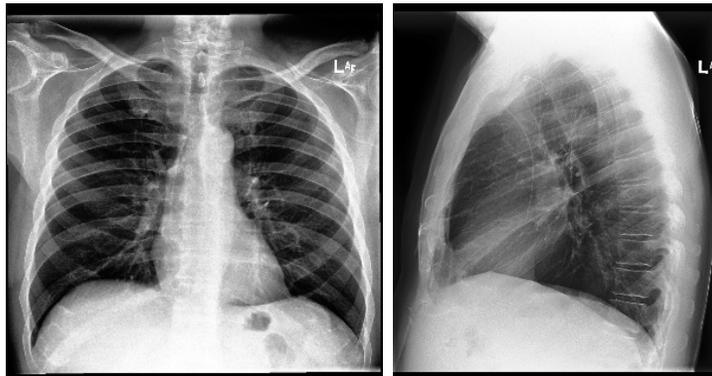


Figura 3.3: Radiografia frontal e lateral de um mesmo paciente.

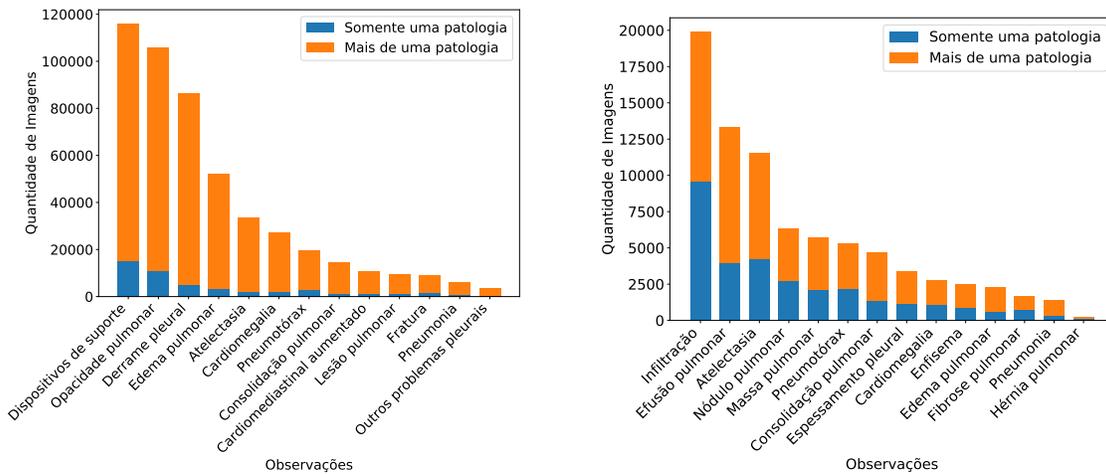
ChestX-ray14 [24] é uma base de dados com 112120 imagens frontais de radiografia de tórax de 30805 pacientes diferentes, que estão associadas à ausência ou presença de 14 patologias. A Tabela 3.2 apresenta a divisão do conjunto de dados de acordo com suas classes, porém, diferente da base anterior, as classes só podem ser inferidas como positivas ou negativas.

¹<https://stanfordmlgroup.github.io/competitions/chexpert/>

A Figura 3.4 mostra uma análise sobre a quantidade de patologias existentes nas imagens das bases de dados. Considerando a base de dados *CheXpert* [10] (Figura 3.4 (a)), 21,4% das imagens são rotuladas com uma única patologia e, em média, há 2,2 patologias por imagem. Já na base de dados *ChestX-ray14* [24] (Figura 3.4 (b)), temos que 27,6% das imagens são rotuladas com apenas uma patologia e existem 0,7 patologias por imagem, em média. Vale ressaltar que a quantidade de imagens que não são rotuladas por alguma patologia na base de dados *CheXpert* [10] é 24761, 11,0% do total de imagens, enquanto que na base de dados *ChestX-ray14* [24], esse número é de 60412, 53,8% do total de imagens da base.

Classe	Treinamento		Validação		Teste	
	Positivo	Negativo	Positivo	Negativo	Positivo	Negativo
Atelectasia	8103	70381	1128	10083	2304	20121
Cardiomegalia	1940	76544	268	10943	564	21861
Efusão pulmonar	9347	69137	1285	9926	2675	19750
Infiltração	13965	64519	1952	9259	3953	18472
Massa pulmonar	4017	74467	581	10630	1148	21277
Nódulo pulmonar	4452	74032	617	10594	1254	21171
Pneumonia	948	77536	139	11072	266	22159
Pneumotórax	3716	74768	526	10685	1056	21369
Consolidação pulmonar	3258	75226	493	10718	916	21509
Edema pulmonar	1634	76850	227	10984	442	21983
Enfisema	1769	76715	250	10961	497	21928
Fibrose pulmonar	1188	77296	146	11065	352	22073
Espessamento pleural	2397	76087	323	10888	665	21760
Hérnia pulmonar	167	78317	18	11193	42	22383

Tabela 3.2: Divisão da base de dados *ChestX-ray14* [24].



(a) *CheXpert* [10].

(b) *ChestX-ray14* [24].

Figura 3.4: Distribuição do número de patologias por classe.

As bases de dados foram divididas da mesma forma que os trabalhos correlatos [15,23,25]: 70% para treinamento, 10% para validação e 20% para teste, como mostrado nas Tabelas 3.1 e 3.2. Assim, é possível realizar uma comparação justa com outros trabalhos disponíveis na literatura. A Tabela 3.3 apresenta algumas informações relevantes sobre as bases de dados que serão utilizadas para fazer o treinamento, validação e teste neste projeto.

Base de Dados	Tipo	Dimensões	Formato	Profundidade	Tamanho
<i>CheXpert</i> Original [10]	CXR	2000 × 2000	JPG	8 bits	~439 GB
<i>CheXpert</i> Reduzida [10]	CXR	320 × 320	JPG	8 bits	~11 GB
<i>ChestX-ray14</i> [24]	CXR	1024 × 1024	PNG	8 bits	~45 GB

Tabela 3.3: Informações sobre as bases de dados que serão utilizadas neste projeto.

3.4 Recursos Computacionais

A implementação deste projeto será realizada na linguagem Python, dado que existe um grande número de ferramentas desenvolvidas com suporte para essa linguagem e com boa documentação. O projeto utilizará bibliotecas científicas, numéricas, de DCNN, de apresentação de gráficos e imagens, entre outras. Algumas bibliotecas que podemos destacar são: NumPy², scikit-learn³, TensorFlow⁴, Keras⁵, Matplotlib⁶ e PyTorch⁷.

Os experimentos deste projeto serão realizados no Laboratório de Informática Visual (LIV) do Instituto de Computação (IC-Unicamp), em uma máquina com processador Intel i7-3770, 3.50 GHz e uma GPU NVIDIA TITAN V, com 5120 núcleos e 12 GB de memória.

²<https://numpy.org/>

³<https://scikit-learn.org/>

⁴<https://www.tensorflow.org/>

⁵<https://keras.io/>

⁶<https://matplotlib.org/>

⁷<https://pytorch.org/>

Capítulo 4

Resultados Preliminares

Neste capítulo, nós apresentamos os resultados preliminares obtidos utilizando DCNNs pré-treinadas aplicadas nas bases de dados descritas na Seção 3.3.

Como DCNNs, nós escolhemos as utilizadas nos trabalhos correlatos [15, 24, 25]: *DenseNet121*, *GoogLeNet*, *VGG16*, *ResNet50* e a *AlexNet*. As configurações utilizadas nas etapas de treinamento e validação foram as mesmas para todas as DCNNs: otimizador *Adam*, taxa de aprendizagem 0,001, Entropia Cruzada Binária como função de perda e 100 épocas como duração limite da etapa de treinamento. As bases de dados foram divididas em 70% para treinamento, 10% para validação e 20% para teste, sendo que a técnica de Aumentação de Dados foi aplicada somente sobre o conjunto de treinamento, como explicado na Subseção 3.1.1. Neste momento, utilizamos somente a métrica AUROC como medida de qualidade dos resultados, pois é a mesma utilizada nos trabalhos correlatos.

A Tabela 4.1 apresenta os melhores resultados alcançados nos trabalhos correlatos e os resultados preliminares obtidos neste projeto, ambos na base de dados *ChestX-ray14* [24]. Comparando os métodos existentes na literatura as redes usadas como *baseline*, podemos destacar o melhor desempenho da DCNN *DenseNet121*, que é a mesma utilizada por Rajpurkar et al. [15]. Essa conseguiu obter o melhor resultado nas patologias atelectasia, efusão pulmonar, consolidação pulmonar e edema pulmonar. A DCNN *GoogLeNet* obteve o melhor resultado somente na patologia pneumonia. Nas demais, o método proposto por Rajpurkar et al. [15] atingiu os melhores resultados.

A Tabela 4.2 mostra os resultados na base de dados *CheXpert* [10]. Assim como na base de dados anterior, a DCNN que obteve melhor resultado médio foi a *DenseNet121*, porém, a *VGG16* ganhou em quatro classes e a *ResNet50* em uma classe. Não temos resultados dos métodos da literatura para essa base de dados, já que eles não foram testados com esta base. Vale destacar que os rótulos incertos foram tratados como negativos nestes experimentos.

Classe	Wang et al. [24]	Yao et al. [25]	Rajpurkar et al. [15]	DenseNet121	GoogLeNet	VGG16	ResNet50	AlexNet
Atelectasia	0,7003	0,772	0,8094	0,8285	0,8214	0,8187	0,8141	0,7714
Cardiomegalia	0,8100	0,904	0,9248	0,9009	0,8893	0,8729	0,8819	0,8851
Efusão pulmonar	0,7585	0,859	0,8638	0,8904	0,8859	0,8884	0,8845	0,8607
Infiltração	0,6614	0,695	0,7345	0,7149	0,7065	0,7145	0,7077	0,6950
Massa pulmonar	0,6933	0,792	0,8676	0,8591	0,8506	0,8510	0,8496	0,7723
Nódulo pulmonar	0,6687	0,717	0,7802	0,7628	0,7604	0,7683	0,7514	0,6784
Pneumonia	0,6580	0,713	0,7680	0,7655	0,7713	0,7402	0,7615	0,7283
Pneumotórax	0,7993	0,841	0,8887	0,8812	0,8732	0,8780	0,8678	0,8342
Consolidação pulmonar	0,7032	0,788	0,7901	0,8099	0,8085	0,7944	0,7972	0,7802
Edema pulmonar	0,8052	0,882	0,8878	0,8971	0,8924	0,8854	0,8921	0,8767
Enfisema	0,8330	0,829	0,9371	0,9063	0,9061	0,9092	0,9080	0,8387
Fibrose pulmonar	0,7859	0,767	0,8047	0,7784	0,7978	0,7690	0,7747	0,7414
Espessamento pleural	0,6835	0,765	0,8062	0,7950	0,7848	0,7890	0,7839	0,7337
Hérnia pulmonar	0,8717	0,914	0,9164	0,8925	0,8680	0,8767	0,8512	0,8215
Média da validação	0,7451	0,8027	0,8413	0,8345	0,8298	0,8255	0,8233	0,7870
Taxa de perda	—	—	—	0,1449	0,1459	0,1456	0,1502	0,1548

Tabela 4.1: Resultados obtidos na base de dados *ChestX-ray14* [24].

Classe	DenseNet121	GoogLeNet	VGG16	ResNet50	AlexNet
Normal	0,8707	0,8703	0,8671	0,8645	0,8479
Cardiomeastinal aumentado	0,6680	0,6562	0,6458	0,6604	0,6323
Cardiomegalia	0,8506	0,8503	0,8463	0,8526	0,8259
Opacidade pulmonar	0,7299	0,7303	0,7326	0,7231	0,7157
Lesão pulmonar	0,7941	0,7754	0,7588	0,7743	0,7252
Edema pulmonar	0,8430	0,8397	0,8390	0,8371	0,8226
Consolidação pulmonar	0,7511	0,7414	0,7416	0,7397	0,7078
Pneumonia	0,7627	0,7511	0,7537	0,7525	0,7014
Atelectasia	0,6949	0,6925	0,6912	0,6919	0,6702
Pneumotórax	0,8618	0,8521	0,8653	0,8503	0,8029
Derrame pleural	0,8778	0,8781	0,8782	0,8732	0,8550
Outros problemas pleurais	0,8042	0,7990	0,7872	0,7884	0,7549
Fratura	0,7550	0,7441	0,7288	0,7373	0,6738
Dispositivos de suporte	0,8873	0,8762	0,8901	0,8811	0,8410
Média da validação	0,7965	0,7898	0,7876	0,7876	0,7555
Taxa da perda	0,2761	0,2777	0,2756	0,2803	0,2925

Tabela 4.2: Resultados obtidos na base de dados *CheXpert* [10].

Capítulo 5

Plano de Trabalho

O plano de trabalho é composto pelas seguintes atividades:

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado.
2. Revisão bibliográfica.
3. Exame de Qualificação do Mestrado (EQM).
4. Pré-processamento das bases de dados.
5. Seleção de uma rede neural inicial.
6. Definição de arquitetura para classificação de imagens de radiografias de tórax.
7. Aprimoramento da arquitetura.
8. Participação do Programa de Estágio Docente (PED).
9. Realização de testes e análise dos resultados.
10. Documentação e publicação dos resultados.
11. Escrita da dissertação de mestrado.
12. Apresentação da dissertação de mestrado.

O plano de trabalho para um período de 24 meses é mostrado na Tabela 5.1.

Bibliografia

- [1] 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. International Society for Optics and Photonics, 2015.
- [2] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *22nd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.
- [3] T. Durand, N. Thome, and M. Cord. WELDON: Weakly Supervised Learning of Deep Convolutional Neural Networks. In *29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4743–4752, 2016.
- [4] B. J. Erickson and B. Bartholmai. Computer-Aided Detection and Diagnosis at the Start of the Third Millennium. *Journal of Digital Imaging*, 15(2):59–68, 2002.
- [5] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT press, 2016.
- [6] A. Halevy, P. Norvig, and F. Pereira. The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, 24(02):8–12, 2009.
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [8] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *The Journal of Neural Computation*, 9(8):1735–1780, 1997.
- [9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. In *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.
- [10] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haggoo, 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 *arXiv:1901.07031*, 2019.

- [11] A. E. W. Johnson, T. J. Pollard, S. Berkowitz, N. R. Greenbaum, M. P. Lungren, C. Deng, R. G. Mark, and S. Horng. MIMIC-CXR: A Large Publicly Available Database of Labeled Chest Radiographs. *arXiv:1901.07042*, 2019.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105. 2012.
- [13] 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.
- [14] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Is Object Localization for Free? - Weakly-Supervised Learning With Convolutional Neural Networks. In *28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 685–694, 2015.
- [15] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv:1711.05225*, 2017.
- [16] S. Raschka and V. Mirjalili. *Python Machine Learning, 2nd Ed.* Packt Publishing, Birmingham, UK, 2nd edition, 2017.
- [17] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*, 2014.
- [18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [19] 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.
- [20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going Deeper With Convolutions. In *28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.
- [21] G. R. Tudor, D. Finlay, and N. Taub. An assessment of inter-observer agreement and accuracy when reporting plain radiographs. *Clinical Radiology*, 52(3):235–238, 1997.
- [22] B. van Ginneken, L. Hogeweg, and M. Prokop. Computer-aided Diagnosis in Chest Radiography: Beyond Nodules. *European Journal of Radiology*, 72(2):226–230, 2009.
- [23] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2097–2106, 2017.

- [24] 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.
- [25] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv:1710.10501*, 2017.
- [26] P. Ypsilantis and G. Montana. Learning what to look in chest X-rays with a recurrent visual attention model. *arXiv:1701.06452*, 2017.