Programa Interdisciplinar de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

Uilian Kenedi Lopes

Redes Neurais Convolucionais Aplicadas ao Diagnóstico de Tuberculose por meio de Imagens Radiológicas

São Leopoldo, 2017

Uilian Kenedi Lopes

REDES NEURAIS CONVOLUCIONAIS APLICADAS AO DIAGNÓSTICO DE TUBERCULOSE POR MEIO DE IMAGENS RADIOLÓGICAS

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos — UNISINOS

Orientador:
Prof. Dr. João Francisco Valiati

São Leopoldo
2017

L864r Lopes, Uilian Kenedi.
Redes neurais convolucionais aplicadas ao diagnóstico de tuberculose por meio de imagens radiológicas / Uilian Kenedi Lopes. – 2017.
87 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2017.
"Orientador: Prof. Dr. João Francisco Valiati."

1. Computação. 2. Tuberculose - Diagnóstico. 3. Diagnóstico por imagem. I. Valiati, João Francisco. II. Título.

CDU 004: 616.24-002.5

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Raquel Herbcz França – CRB 10/1795)

Uilian Kenedi Lopes

Redes Neurais Convolucionais Aplicadas ao Diagnóstico de Tuberculose por meio de Imagens Radiológicas

Dissertação apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Aprovado em 24 de março de 2017

BANCA EXAMINADORA

Profa. Dra. Marta Villamil - UNISINOS

Prof. Dr. Wilson Pires Galvão Neto - UNIRITTER

Prof. Dr. João Francisco Valiati (Orientador)

Visto e permitida a impressão
São Leopoldo,

Prof. Dr. Sandro José Rigo
Coordenador PPG em Computação Aplicada

Dedico este trabalho aos meus pais, Wolny e Eunice e a minha noiva Débora

AGRADECIMENTOS

Agradeço primeiramente a Deus, aos meus pais, Wolny e Eunice, cujo exemplo de trabalho, esforço, dedicação e amor tem sido grande fonte de inspiração para mim, e a minha noiva Débora, pela paciência, compreensão, carinho e por estar ao meu lado durante todo este período importante da minha vida.

Aos meus amigos e a minha família, especialmente meus irmãos, Deise e Deivis pela força, incentivo e paciência.

Aos meus antigos chefes, Eduardo Basso e Moisés Pontremoli, pelo incentivo e por todo o conhecimento que me passaram durante o período em que trabalhamos juntos.

Agradeço também aos colegas e professores da UNISINOS, especialmente ao professor e orientador Dr. João F. Valiati, pela grande dedicação, conhecimentos e atenção que foram essenciais durante o desenvolvimento deste trabalho.

"*A educação tem raízes amargas, mas os seus frutos são doces*".
(Aristóteles)

RESUMO

De acordo com a Organização Mundial de Saúde, a tuberculose (juntamente com a AIDS) é a doença infecciosa que mais causa mortes no mundo. Estima-se que em 2014 cerca de 1,5 milhão de pessoas infectadas com o *Mycobacterium Tuberculosis* morreram, a maior parte delas nos países em desenvolvimento. Muitas destas mortes poderiam ter sido evitadas caso o diagnóstico ocorresse nas fases iniciais da doença, mas infelizmente as técnicas mais avançadas de diagnóstico ainda têm custo proibitivo para adoção em massa nos países em desenvolvimento. Uma das técnicas mais populares de diagnóstico da tuberculose ainda é através da radiografia torácica frontal, entretanto este método tem seu impacto reduzido devido à necessidade de radiologistas treinados analisarem cada radiografia individualmente. Por outro lado, já existem pesquisas buscando automatizar o diagnóstico através da aplicação de técnicas computacionais às imagens radiográficas pulmonares, eliminando assim a necessidade da análise individual de cada radiografia e diminuindo grandemente o custo. Além disso, aprimoramentos recentes nas Redes Neurais Convolucionais, relacionados também à área de *Deep Learning*, obtiveram grande sucesso para classificação de imagens nos mais diversos domínios, porém sua aplicação no diagnóstico da tuberculose ainda é limitada. Assim o foco deste trabalho é produzir uma investigação que promova avanços nas pesquisas, trazendo três abordagens de aplicação de Redes Neurais Convolucionais com objetivo de detectar a doença. As três propostas apresentadas neste trabalho são implementadas e comparadas com a literatura corrente. Os resultados obtidos até o momento mostraram-se sempre competitivos com trabalhos já publicados na área, obtendo resultados superiores na maior parte dos casos, demonstrando assim o grande potencial das Redes Convolucionais como extratoras de características de imagens médicas.

Palavras-chave: Deep Learning. Redes Neurais Convolucionais. Tuberculose. Diagnóstico com Auxílio de Computadores.

ABSTRACT

According to the World Health Organization, tuberculosis (along with AIDS) is the most deadly infectious disease in the world. In 2014 it is estimated that 1.5 million people infected by the *Mycobacterium Tuberculosis* died, most of them in developing countries. Many of those deaths could have been prevented if the disease was detected at an earlier stage, but unfortunately the most advanced diagnosis methods are cost prohibitive for mass adoption in developing countries. One of the most popular tuberculosis diagnosis methods still is by analysis of frontal thoracic radiographies, however the impact of this method is diminished by the need for individual analysis of each radiography by properly trained radiologists. On the other hand, there is significant research on automating diagnosis by the application of computational techniques to lung radiographic images, eliminating the need for individual analysis of the radiographies and greatly diminishing the cost. In addition to that, recent improvements on Convolutional Neural Networks, which are related to *Deep Learning*, accomplished excellent results classifying images on diverse domains, but it's application for tuberculosis diagnosis still is limited. Thus, the focus of this work is to produce an investigation that will advance the research in the area, proposing three approaches to the application of Convolutional Neural Networks to detect the disease. The three proposals presented in this works are implemented and compared to the current literature. The obtained results are competitive with works published in the area, achieving superior results in most cases, thus demonstrating the great potential of Convolutional Networks as medical image feature extractors.

Keywords: Deep Learning. Convolutional Neural Network. Tuberculosis. Computer-Aided Diagnosis.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

Biblio.	Bibliográfica
Exp.	Experimentos
Desenv.	Desenvolvimento
Rev.	Revisão

LISTA DE SIGLAS

ANN	Artificial Neural Network
CAD	Computer Assisted Diagnosis
CNN	Convolutional Neural Network
DBN	Deep Belief Network
HOG	Histogram of Oriented Gradients
IA	Inteligência Artificial
MIL	Multiple Instance Learning
MLP	MultiLayer Perceptron
SIFT	Scale Invariant Feature Transform
SVM	Support Vector Machine

SUMÁRIO

1 INTRODUÇÃO

Nos últimos anos ocorreram grandes avanços no campo da inteligência artificial trazidos por um novo conjunto de técnicas conhecidas como *Deep Learning*. Segundo LeCun, Bengio e Hinton (2015), estes avanços foram alcançados através da combinação do uso de hardware mais poderoso (especialmente *Graphics Processing Unit* - GPUs) com a aplicação de novas técnicas no processo de treinamento, como as ReLUs (*Rectified Linear Units*), *DropOut* para regularização dos pesos e também aplicando técnicas de geração de amostras sintéticas a partir de deformações em amostras reais. Além disto, a utilização de bases de dados com milhões de amostras reais também pode ser considerado um fator importante para o sucesso do *Deep Learning* (RUSSAKOVSKY et al., 2015). Entre as técnicas de *Deep Learning* destacam-se especialmente as CNNs (*Convolutional Neural Networks*) (LECUN et al., 1998), que são as detentoras dos melhores resultados na maior parte dos *benchmarks* de classificação de imagens e que são abordadas com maior profundidade neste trabalho.

A diferença fundamental entre técnicas tradicionais de IA e técnicas de *Deep Learning* diz respeito a como é feita a extração das características relevantes dos dados. Técnicas tradicionais de IA, em geral, não produzem bons resultados quando aplicadas diretamente aos dados "crus"(pixels no caso de imagens), sendo necessário que um conhecedor do domínio ao qual a técnica está sendo aplicada planeje um extrator de características capaz de criar uma representação dos dados que seja discriminativa o suficiente para resolução do problema proposto. Já com *Deep Learning*, não existe mais a necessidade do especialista planejar a extração de características. As técnicas são capazes de aprender automaticamente (a partir dos dados "crus") representações compactas e discriminativas conforme necessário para a tarefa de classificação ou detecção (LECUN; BENGIO; HINTON, 2015).

Neste trabalho o foco é em um tipo específico de redes profundas, as Redes Neurais Convolucionais. As principais características que diferem as CNNs de outras arquiteturas de redes neurais é a utilização da operação de convolução no lugar da multiplicação de matrizes em pelo menos uma de suas camadas, o compartilhamento de pesos, a existência de camadas de *Pooling* e a presença de campos receptivos locais (GOODFELLOW; BENGIO; COURVILLE, 2016). A primeira proposta de uma arquitetura de rede que atende a todas estas características foi feita em (LECUN et al., 1998). Gradualmente, durante a primeira década do século 21, as CNNs foram evoluindo até culminar na grande revolução apresentada em (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), quando os resultados de classificação de imagens obtidos na base de dados *ImageNet* superaram por larga margem o estado da arte anterior. Desde então, as CNNs têm sido rapidamente aprimoradas e seguem estabelecendo novos recordes de *performance* a cada ano, tanto na base de dados *ImageNet* quanto em outras bases de classificação de imagens como a COCO (LIN et al., 2014), MNIST (LECUN; CORTES; BURGES, 1998), entre outras. O enorme sucesso na classificação de imagens motivou sua aplicação para resolução de problemas das mais variadas áreas como previsão de interação entre proteínas

(WALLACH; DZAMBA; HEIFETS, 2015), previsão de morfologia de galáxias (DIELEMAN; WILLETT; DAMBRE, 2015), treinamento de inteligência artificial capaz de jogar Go (SILVER et al., 2016), desenvolvimento de carros autônomos (HUVAL et al., 2015), modelagem de frases (KALCHBRENNER; GREFENSTETTE; BLUNSOM, 2014), entre outros.

No Diagnóstico Auxiliado por Computadores (*Computer Assisted Diagnosis* - CAD), área de pesquisa que busca auxiliar ou automatizar diagnósticos através de técnicas computacionais, já existem diversos trabalhos desenvolvidos utilizando Redes Neurais Convolucionais. Como principais exemplos pode-se citar (HUA et al., 2015), onde os autores buscam detectar nódulos pulmonares através de uma CNN, em (HAVAEI et al., 2015) o objetivo é segmentar tumores cerebrais em imagens obtidas através de ressonância magnética e em (DHUNGEL; CARNEIRO; BRADLEY, 2015) é feita a segmentação de nódulos um imagens de mamogramas.

Pode-se notar que no campo de CAD, a grande maioria dos trabalhos (tanto os relacionados com *Deep Learning* quanto os não-relacionados) tendem a focar a pesquisa em detecção de tumores e nódulos. Na subárea específica a ser explorada neste trabalho, a detecção da tuberculose, historicamente ocorreram poucos avanços no sentido de automatizar o diagnóstico. Felizmente nos últimos anos foram publicados trabalhos que contribuíram muito para o avanço deste campo de pesquisa. Em Jaeger et al. (2014a), foi feita uma proposta que combina características extraídas através de diferentes algoritmos voltados para detecção de bordas, formato, textura, etc. Após a extração de características, é realizada a classificação utilizando uma SVM (*Support Vector Machine*) linear. Além da apresentação e avaliação dos resultados do algoritmo, a principal contribuição do autor foi tornar públicas as bases de dados investigadas (JAEGER et al., 2014b).

No presente trabalho é realizada a aplicação das redes convolucionais na classificação de imagens de Raio-X torácico com o objetivo de diagnosticar a presença da tuberculose. Até o momento atual, o único artigo publicado aplicando CNNs na detecção de tuberculose foi Hwang et al. (2016). Neste artigo, os autores adaptam uma arquitetura de CNN para detectar a doença e realizam o ajuste dos pesos treinando-a em uma grande base de dados privada.

A importância deste trabalho se deve a necessidade da existência de um método rápido, barato e eficaz capaz de diagnosticar a tuberculose. Já existem métodos de diagnóstico de altíssima confiabilidade, entretanto o alto custo os torna inviáveis para uso em massa nos países em desenvolvimento, que são os mais atingidos pela doença (GLOBAL TUBERCULOSIS REPORT 2015, 2015). Os métodos baratos, como a Baciloscopia do Escarro, infelizmente produzem elevado número de falsos positivos (LEUNG, 2011). O diagnóstico através de imagens de Raios-X já é largamente utilizado no mundo (GLOBAL TUBERCULOSIS REPORT 2015, 2015), porém seu uso ainda é limitado pela necessidade de pessoal qualificado para analisar individualmente cada radiografia. Já existem estudos que indicam ser possível atingir índice de falsos negativos próximos de zero através da análise radiográfica (LEUNG, 2011), então aparentemente toda informação necessária para detecção está presente nas imagens. Caso seja possível eliminar a necessidade do radiologista seria possível massificar o diagnóstico da doença de forma barata

possivelmente ajudando a salvar muitas vidas (JAEGER et al., 2014a).

1.1 Objetivos

O objetivo do presente trabalho é investigar a eficiência do uso de CNNs para classificação de imagens radiológicas do tórax na intenção de demonstrar seu potencial de detecção da tuberculose. Assim, é realizada a análise comparativa deste algoritmo com alternativas presentes na literatura corrente, usando como base o trabalho de Jaeger et al. (2014a).

Pode-se definir ainda, como objetivos específicos do trabalho:

- Aplicar três abordagens diferentes utilizando Redes Neurais Convolucionais pré-treinadas para a detecção da tuberculose. A primeira aplicando diretamente diferentes arquiteturas de CNN como extratoras de características em uma versão reduzida da imagem radiográfica, a segunda utilizando as mesmas arquiteturas de CNN para extrair características em sub-janelas da imagem (em sua resolução original) que posteriormente são combinadas para gerar um único descritor global da radiografia e a terceira criando comitês de classificadores juntando os resultados obtidos nas duas primeiras abordagens;
- Produzir uma análise comparativa dos resultados obtidos nas diferentes abordagens propostas com as abordagens presentes na literatura (JAEGER et al., 2014a; HWANG et al., 2016), apresentando pontos fortes e fracos de cada uma;
- Produzir uma análise crítica sobre a eficácia de três diferentes arquiteturas de redes convolucionais como extratores de características para detecção da tuberculose.

1.2 Estrutura

O próximo capítulo apresenta a fundamentação teórica, relacionando a tuberculose e as Redes Neurais Convolucionais buscando explicitar todos os conceitos relevantes para compreensão das mesmas. No capítulo 3 são apresentados os principais trabalhos relacionados à classificação de imagens médicas para detecção de doenças, principalmente os trabalhos relacionados a tuberculose e também aqueles onde foi aplicada alguma técnica de *Deep Learning*. No capítulo 4, são descritas as técnicas de pré-processamento aplicadas nas imagens assim como as propostas de uso da CNN desenvolvidas juntamente com os resultados obtidos. No capítulo final é apresentada a conclusão e propostas de melhorias futuras para o trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os conceitos mais importantes presentes neste trabalho. A seção 2.1 explicita a tuberculose, descrevendo a doença, suas causas, efeitos e métodos atuais de diagnóstico. Na seção 2.2 são apresentadas as redes neurais e na 2.3 são apresentadas as Redes Neurais Convolucionais.

Na seção 2.4 é abordado o conceito de *Multiple Instance Learning*, que está por trás de uma das abordagens propostas neste trabalho. Na seção 2.5 é apresentada a base de dados onde foram treinadas as redes utilizadas neste trabalho: a *ImageNet*. Por fim, na seção 2.6 mostram-se as métricas definidas para avaliação dos resultados do trabalho e na 2.7 são feitas as considerações finais sobre os conceitos descritos.

2.1 Tuberculose

Juntamente com a Aids, a tuberculose é a doença infecciosa que mais causa mortes no mundo. De acordo com a Organização Mundial de Saúde em seu último relatório anual (GLOBAL TUBERCULOSIS REPORT 2016, 2016), a tuberculose se tornou a doença contagiosa com maior número de mortes registradas em 2015. Neste ano, 1 milhão e 800 mil pessoas infectadas com a doença morreram, sendo que destas 400 mil também eram HIV positivo. No mesmo ano, a Aids causou 1,1 milhão de mortes. A maior parte dos casos de tuberculose ocorrem em países de terceiro mundo, onde a pobreza e a alimentação inadequada diminuem a capacidade de resistência da população. Dos quase 10 milhões de casos estimados da doença em 2014, 58% ocorreram na Ásia e 28% na África.

Estima-se que de 2 a 3 bilhões de pessoas no mundo estejam infectadas com o *Mycobacterium Tuberculosis*, o bacilo causador da tuberculose (também conhecido como bacilo de Koch). Felizmente, apenas um pequeno percentual destas pessoas desenvolverão a doença no decorrer da vida.

A tuberculose é causada pelo bacilo *Mycobacterium Tuberculosis* que normalmente afeta os pulmões (tuberculose pulmonar) mas também pode afetar outras partes do corpo (tuberculose extrapulmonar). A principal forma de transmissão da doença é pelo ar quando pessoas infectadas espalham a bactéria, em geral pela tosse.

Sem tratamento, a taxa de mortalidade é alta. Estima-se que 70% das pessoas contaminadas com a doença morreriam em até 10 anos sem o tratamento. Com tratamento adequado a taxa de mortalidade cai para 20%.

Os primeiros tratamentos efetivos para a doença foram criados a partir de 1940. Atualmente o tratamento recomendado para a doença consiste em um coquetel composto de quatro remédios diferentes (Rifampicina, Isoniazida, Pirazinamida e Etambutol) que devem ser tomados durante seis meses. Para casos mais graves com infecções causadas por mutações do bacilo resistentes aos remédios padrão, o tratamento é feito com drogas mais caras e tóxicas podendo durar até

20 meses (GLOBAL TUBERCULOSIS REPORT 2015, 2015).

2.1.1 Métodos de Diagnóstico

O diagnóstico da tuberculose pode ser realizado por diversas técnicas, algumas mais custosas que outras, entretanto com diferentes graus de precisão. Dentre elas se destacam:

- **Baciloscopia do Escarro:** É a técnica mais antiga (foi criada há mais de 100 anos) e a mais utilizada no mundo principalmente devido a seu baixo custo. Ela consiste na análise do escarro em microscópio buscando identificar visualmente a presença do bacilo de Koch. A principal vantagem desta técnica e o motivo de sua popularidade é o baixo custo de realização. Porém, segundo Leung (2011), a baciloscopia tende a produzir um número elevado de falsos negativos. Estimativas realizadas chegam a detectar menos de 50% dos casos da doença utilizando esta técnica;

- **Análise de Pele:** A principal técnica de diagnóstico através da análise da pele é chamado de teste da Tuberculina. Ele consiste em inserir uma pequena quantidade de antígenos da tuberculose (tuberculina) sob a pele do paciente. Caso a pessoa já possua a infecção, o corpo tende a reagir e formar uma inchaço vermelho no local. Esta técnica é bastante popular nos países do mediterrâneo oriental e nas Américas. Apesar de sua popularidade, segundo (GLOBAL TUBERCULOSIS REPORT 2015, 2015), o teste da tuberculina tem alto índice de falsos negativos e seu uso isoladamente não é recomendado. Recomenda-se apenas o uso em conjunto com outros testes;

- **Cultura Bacteriológica:** A técnica, considerada até recentemente como a mais confiável no diagnóstico da doença (GLOBAL TUBERCULOSIS REPORT 2015, 2015). Consiste em realizar a cultura da bactéria a partir do escarro ou de uma amostra de pus e observá-lo para buscar a presença do *Mycobacterium Tuberculosis*. Infelizmente não é viável usar este teste para diagnosticar a doença na maior parte dos casos, pois o tempo necessário para o crescimento do organismo em laboratório é grande, podendo chegar a meses em alguns casos. Além disto, o teste necessita de laboratórios bem equipados e equipes altamente qualificadas. Tudo isto encarece demasiadamente o teste tornando-o inviável como método de diagnóstico na maior parte dos casos nos países em desenvolvimento;

- **Análise de Moléculas:** Recentemente foram desenvolvidos novos testes baseados na análise da presença de moléculas existentes no DNA do bacilo de Koch. Segundo Global tuberculosis report 2015 (2015), os resultados obtidos são altamente acurados e rápidos. A OMS recomenda desde 2010 o diagnóstico através da análise molecular, entretanto estes métodos ainda não são usados em larga escala devido ao alto custo;

- **Radiografias:** A próxima subseção entrará em detalhes sobre o uso de radiografias no diagnóstico da tuberculose.

2.1.2 Diagnóstico Através de Radiografias

Segundo a Organização Mundial de Saúde (GLOBAL TUBERCULOSIS REPORT 2015, 2015), o uso de radiografias já é bastante difundido no processo de diagnóstico da tuberculose, em geral de forma conjunta com outros exames. Seu baixo custo em comparação com outras técnicas torna-o ideal para uso nos países em desenvolvimento. Entretanto, seu uso ainda é limitado pela necessidade de pessoal qualificado para inspecionar cada radiografia. Aqui pode-se notar a importância da automatização da análise radiográfica, pois assim seria possível reduzir custos e massificar sua utilização nos países em desenvolvimento, o que teria grande potencial de salvar vidas.

Em comparação com as técnicas de baixo custo, a análise radiográfica tende a produzir melhores resultados. Segundo, Leung (2011) e Global tuberculosis report 2015 (2015), o diagnóstico da tuberculose a partir de radiografias torácicas tende a obter uma sensitividade alta porém com especificidade abaixo do desejado. Porém há evidências na literatura, por exemplo em Leung (2011) e Jaeger et al. (2014a), que radiologistas treinados e instruídos a marcar qualquer tipo de anomalia presente nas imagens de Raio-X podem produzir resultados próximos a 100% de detecção. Esta constatação é muito importante, pois indica que há informação suficiente, presente nas imagens radiográficas, para detectar a doença na quase totalidade dos casos.

2.1.3 Manifestações da Tuberculose em Imagens Radiográficas

Segundo Antani (2015), as manifestações mais comuns da doença em imagens de Raios-X são:

- **Consolidação:** Manifesta-se como uma opacidade, frequentemente na região dos vasos sanguíneos pulmonares podendo se estender até a área pleural. A consolidação ocorre quando o ar de uma região é substituído por outra substância como sangue, pus, água, etc;

- **Padrão Miliar:** Aparência arenosa, similar a pequenas sementes espalhada pelo pulmão inteiro. Pode indicar outras doenças como sarcoidose ou uma infecção por fungos;

- **Formação de cavidades:** Bordas do pulmão aparecem mais densas, podendo ocorrer de forma contínua ou descontínua;

- **Alargamento das vias aéreas:** Manifesta-se como anéis tubulares de diâmetro irregular podendo também apresentar radioluminescência no centro;

- **Adenopatia:** Alargamento dos nódulos linfáticos. Pode ser indicativo da presença de um tumor ou edema;

- **Efusões pleurais:** Regiões laterais e mediais tornam-se indistintas;
- **Pleura grossa:** Pode-se notar esta variação na aparência da pleura, principalmente nas regiões periféricas dos pulmões, adjacentes as costelas e ao diafragma.

Estes padrões presentes em radiografias de pacientes com tuberculose são frequentemente encontradas em pacientes com outras doenças pulmonares como a pneumonia por exemplo. É importante notar que, com frequência, os padrões não se manifestam isoladamente, mas de forma combinada na mesma imagem.

2.2 Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são modelos matemáticos de inspiração biológica que, através de combinações de unidades computacionais simples chamadas neurônios, buscam estimar ou aproximar o valor de funções.

A primeira proposta de RNA foi feita em (MCCULLOCH; PITTS, 1943), onde já são apresentados princípios importantes que por muitos anos orientaram o desenvolvimento das RNAs como a estrutura fixa de sinapses, a existência de conexões inibitórias que previnem a ativação de neurônios, a necessidade da saída de cada neurônio ser binária, entre outros (GRAUPE, 2013). Após a proposta original de RNA, o próximo grande avanço da pesquisa na área ocorreu com a publicação de (ROSENBLATT, 1958), que trouxe a proposta do neurônio *Perceptron*. O *Perceptron* tem um funcionamento bastante simples, basicamente realizando uma soma ponderada dos valores recebidos na entrada e, baseado neste resultado, retorna uma saída binária indicando sua ativação ou não. Inicialmente este neurônio foi aplicado com sucesso para resolução de problemas simples de classificação, mas logo percebeu-se suas limitações. Minsky e Papert (1969) demonstram que o *Perceptron* consegue resolver problemas linearmente separáveis, o que o torna incapaz de aprender a solução para problemas como a função *XOR*.

Redes capazes de resolver problemas não-linearmente separáveis se tornaram possíveis após (RUMELHART; HINTON; WILLIAMS, 1986), onde é demonstrado que a técnica *BackPropagation* pode ser utilizada para o treinamento de redes neurais de múltiplas camadas. O *Back-Propagation* (abreviação de *Backward Propagation of Errors*) é usado em conjunto com alguma técnica de otimização (em geral o método de descida do gradiente) para o cálculo de uma função de erro dos pesos da rede e estes são ajustados de forma a minimizar o erro da função.

De forma simplificada pode-se resumir as etapas do treinamento de uma rede usando *Back-Propagation* da seguinte forma:

1. Inicialização dos pesos e bias, definição dos valores da taxa de aprendizado, momento e do critério de parada
2. Apresentação à rede das amostras de treinamento. Para cada amostra, é realizada a série de cálculos descrita nos passos 3 e 4

3. Propagação: supondo que cada amostra de treinamento seja representada por (x(n),d(n)), com o vetor x(n) sendo apresentado a entrada da rede e o vetor de resposta desejada d(n) apresentada a camada de saída da rede. São calculadas as saídas locais induzidas e os sinais funcionais da rede prosseguindo camada por camada

4. Retropropagação: Cálculo dos gradientes locais. Ajuste dos pesos sinápticos da rede na(s) camada(s) oculta(s) de acordo com a regra delta generalizada

5. Iteração: iterar as computações para frente e para trás de acordo com os passos 2 e 3 apresentando novas épocas de amostras de treinamento para a rede até que o critério de parada seja atingido

Dentre as RNAs treinadas com esta técnica, uma das arquiteturas mais populares é conhecida como MLP (*Multilayer Perceptron*). Nesta arquitetura existem múltiplas camadas de neurônios, sendo que cada um deles está conectado a todos os neurônios da camada seguinte. Cada neurônio é muito similar ao *Perceptron* proposto originalmente. Uma importante diferença é que a saída de cada um deles não é mais binária, ela é dada pela aplicação de uma função de ativação não-linear (historicamente normalmente foram utilizadas a função sigmóide ou a tangente hiperbólica) ao resultado da soma ponderada da entrada pelos pesos da rede. Ao contrário de uma rede com apenas uma camada, o MLP (com pelo menos uma camada intermediária) é capaz de aproximar qualquer função contínua (HORNIK; STINCHCOMBE; WHITE, 1989).

Existem outras arquiteturas importantes de RNAs presentes na literatura. Destacam-se especialmente a Rede de HopField proposta em Hopfield (1982) que é uma das primeiras Redes Neurais Recorrentes. Nela existe uma única camada e cada neurônio é conectado a todos os outros (exceto a si mesmo). Este modelo arquitetural serviu de inspiração para o desenvolvimento de outras redes bastante populares na década de 1980 como a BAM (*Bidirectional Associative Memory*) (KOSKO, 1988). Também merece destaque a arquitetura SOM (*Self Organizing Map*), proposta por (KOHONEN, 1982). O objetivo da rede SOM pode ser definido como transformar um sinal de entrada em um mapa discreto bidimensional ou unidimensional. Esta característica faz com que ela seja utilizada principalmente para visualização de dados de alta dimensionalidade em um espaço dimensional menor.

Como visto nos exemplos, existem diversos modelos de RNAs planejadas de forma diferente para realização de tarefas diferentes. No próxima seção é apresentada em detalhes o tipo de RNA que mais tem se destacado nos últimos anos: as Redes Neurais Convolucionais.

2.3 Redes Neurais Convolucionais

Esta seção está subdividida em 3 subseções. A primeira contém um breve resumo da história das redes convolucionais. A segunda traz os principais conceitos e características para

compreensão do funcionamento das mesmas e a subseção final apresenta os principais tipos de camadas presentes nas CNNs usadas na atualidade.

2.3.1 História

A inspiração para criação das redes convolucionais é o funcionamento do córtex visual dos mamíferos. Segundo estudos realizados por Hubel e Wiesel (1968), em gatos e macacos existem dois tipos de células no córtex chamadas de simples e compostas. Estas células são sensíveis a pequenas regiões do campo de visão chamadas de campos receptivos. Elas agem como filtros locais sobre o campo visual explorando a correlação espacial local presente nas imagens naturais. Ainda segundo Hubel e Wiesel (1968), as células simples agem como detectores de bordas enquanto as células complexas agem como extratores de características complexas formadas a partir de combinações de padrões simples. Além disto, as células complexas possuem campos receptivos maiores e tendem a ser invariantes a translação.

A partir das descobertas realizadas sobre o funcionamento do córtex foram propostas arquiteturas de redes neurais buscando emular o comportamento observado nos seres vivos. Fukushima (1980) propôs uma arquitetura chamada *NeoCognitron* que pode ser considerada a primeira rede convolucional da literatura. Ela buscava ter um comportamento quase que exatamente igual ao modelo proposto por Hubel e Wiesel (1968) para o córtex visual. A arquitetura *NeoCognitron* consiste em 2 tipos de neurônios artificiais (chamados *S-cells* e *C-cells*) que buscam simular o comportamento das células simples e complexas no córtex. Uma diferença importante da proposta de Fukushima para as redes convolucionais atuais é que o treinamento do *NeoCognitron* era feito de forma não-supervisionada. A primeira proposta de melhora da arquitetura incorporando o *backpropagation* para realização de um treinamento totalmente supervisionado foi feita por LeCun et al. (1989). No final da década de 1990, LeCunn propôs a famosa rede *LeNet* em (LECUN et al., 1998) que pode ser considerada a primeira arquitetura a exibir todas as características distintivas das redes convolucionais atuais. Já na década de 1990, as redes criadas por LeCun foram aplicadas a resolução de problemas práticos, como por exemplo o reconhecimento de caracteres em cheques, chegando a processar cerca de 10% dos cheques emitidos nos EUA (LECUN et al., 2010).

A partir de 2005, em grande parte devido ao constante aprimoramento das GPUs, foi possível treinar redes convolucionais com uma profundidade maior e com mais rapidez. Chellapilla, Puri e Simard (2006), Poultney et al. (2006), Uetz e Behnke (2009) e Strigl, Kofler e Podlipnig (2010) trouxeram importantes avanços neste sentido e Ciresan et al. (2011) traz uma implementação completa de CNN para GPUs reduzindo o tempo de treinamento de grandes redes, de meses para dias. Neste mesmo trabalho, se aproveitando da redução do tempo de treinamento, os autores avaliam arquiteturas diversas de CNNs em 3 bases de dados de classificação de imagens, superando o estado da arte em todas elas.

Em termos de melhoria de algoritmos destaca-se a aplicação da função de ativação ReLU

para introdução da não-linearidade (JARRETT et al., 2009) e do método de regularização chamado *DropOut* (HINTON et al., 2012).

Krizhevsky, Sutskever e Hinton (2012) trazem uma proposta de CNN (chamada *AlexNet*) treinada com GPUs onde foram inclusos também os principais avanços algorítmicos. A *AlexNet* foi treinada e testada na maior e mais importante base de dados de classificação de imagens: a *ImageNet*. Os resultados obtidos superaram por larga margem o estado da arte anterior. A partir daí, cada vez mais trabalhos tem sido publicados aplicando as CNNs nos mais diversos domínios, em geral com grande sucesso. Os aprimoramentos nas redes convolucionais continuam ocorrendo. Há vários trabalhos publicados realizando experimentos com variações nas funções de ativação (HE et al., 2015a), nas técnicas de regularização (WAN et al., 2013) e nos métodos de *Pooling* (YU et al., 2014), porém até o momento, nenhuma destas variações avaliadas das técnicas foi largamente adotada. A principal tendência nas modelagens de redes convolucionais mais recentes é que elas estão se tornando progressivamente mais profundas (GU et al., 2015). Pode-se notar esta tendência claramente verificando as CNNs vencedoras do ILSVRC (*The ImageNet Large Scale Visual Recognition Challenge*): em 2013 a rede vencedora, chamada *OverFeat* (SERMANET et al., 2013), possuía 8 camadas. Em 2014, o melhor resultado foi alcançado pela *GoogLenet* (SZEGEDY et al., 2015) com 22 camadas e em 2015 pela *ResNet* (HE et al., 2015b) que possui 152 camadas. Ainda neste mesmo trabalho são avaliadas redes (em bases de dados menores) com até 1202 camadas, o que indica que é provável que esta tendência se mantenha para os próximos anos.

2.3.2 Principais Características e Conceitos Fundamentais

De forma simplificada, pode-se considerar como Redes Neurais Convolucionais toda rede neural que utiliza a convolução no lugar da multiplicação de matrizes em pelo menos uma das suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Na prática, as redes convolucionais existentes na atualidade seguem estilos e trazem decisões de projeto importantes que foram adotadas e aprimoradas durante seu histórico de desenvolvimento. (GOODFELLOW; BENGIO; COURVILLE, 2016) enumera 3 características distintivas importantes das redes convolucionais em comparação as redes tradicionais descritas na seção anterior. Estas características são: o compartilhamento de pesos,a presença de campos receptivos locais e a existência de camadas de *Pooling*.

O compartilhamento de pesos permite que em vez da rede aprender pesos específicos para cada região da imagem, ela aprenda apenas um conjunto de filtros menores espacialmente, mas que poderá ser aplicado a todas as regiões da imagem, o que faz com que a representação aprendida tenha um maior poder de generalização (WANG et al., 2012). Ainda segundo Wang et al. (2012), outro benefício importante do compartilhamento de pesos é que ele permite diminuir grandemente a quantidade de parâmetros, simplificando o processo de treinamento e tornando-o mais eficiente.

A segunda característica distintiva é a existência de campos receptivos locais. Nas RNAs clássicas (totalmente conectadas) cada valor de entrada de cada camada é multiplicado por todos os neurônios presentes. Este tipo de arquitetura traz grandes desvantagens quando as matrizes de entrada tendem a ser grandes (como é o caso na classificação de imagens), pois o custo computacional se torna extremamente alto. Como solução para este problema, as redes convolucionais buscam explorar subestruturas presentes nas imagens para otimizar o processamento. Como em imagens naturais pixels adjacentes tendem a ser mais fortemente correlacionados do que pixels distantes, as redes convolucionais são arquitetadas para que cada filtro aprendido seja dependente de apenas uma sub-região dos dados recebidos da camada anterior (WANG et al., 2012). Ou seja, o campo receptivo dos filtros é local. Em uma rede profunda, os filtros locais acabam interagindo com regiões progressivamente maiores da imagem. Isto permite que padrões cada vez mais complexos sejam modelados a partir de combinações de operações locais simples (GOODFELLOW; BENGIO; COURVILLE, 2016).

A terceira característica distintiva importante das CNNs é a existência de camadas de *Pooling*. O objetivo principal destas camadas é a redução da dimensionalidade dos dados antes de propagá-los à próxima camada da rede (GOODFELLOW; BENGIO; COURVILLE, 2016). Isto é realizado em geral subdividindo a entrada em blocos e realizando algum tipo de sumarização estatística da informação presente em cada um deles. Os tipos de *Pooling* mais utilizados são o *Average Pooling*, onde a média de cada região é retornada e o *Max Pooling*, onde o retorno é o maior valor presente em cada vizinhança.

Também vale destacar que outro importante efeito da operação de *Pooling* é introduzir um elemento a mais de invariância a translação (GOODFELLOW; BENGIO; COURVILLE, 2016), propriedade que é extremante importante na maior parte dos casos de classificação de imagens.

Para prevenir *overfitting*, o principal método de regularização em uso na atualidade é o *Dropout*. O *DropOut* funciona da seguinte forma: a cada época de treinamento nodos individuais da rede são desativados de acordo com uma probabilidade pré-definida. Assim a cada época a rede não é treinada integralmente, mas apenas um subconjunto dela. Desta forma, os pesos individuais da rede tendem a aprender características mais robustas sendo obrigados a interagir com uma amostragem aleatória diferente de pesos a cada época pois não pode contar que outros neurônios compensem o aprendizado de pesos de baixa qualidade (SRIVASTAVA et al., 2014).

2.3.3 Tipos de Camadas

As redes convolucionais são compostas de sequências de camadas empilhadas umas sobre as outras, sendo que cada camada faz algum tipo de transformação nos dados de entrada através de uma função definida. Os anexos A, B e C trazem exemplos de arquiteturas CNNs atuais e demonstram os tipos de camadas mais comuns presentes nas redes: camadas convolucionais, camadas ReLU, camadas de *Pooling* (que foram apresentadas na seção anterior) e camadas totalmente conectadas (que são explicadas no final desta subseção).

As camadas convolucionais são responsáveis pela extração de características dos dados recebidos na entrada. A entrada de cada camada convolucional l é um conjunto de $n^{(l-1)}$ mapas de características extraídas pela camada anterior com dimensão $w^{(l-1)}$ x $h^{(l-1)}$. Caso a camada l seja a primeira da rede, sua entrada é uma imagem I composta de 1 ou mais canais. A saída da camada convolucional consiste em $n^{(l)}$ mapas de características com dimensões $w^{(l)}$ x $h^{(l)}$. O mapa de número i, na camada l (chamado aqui de $M_i^{(l)}$), é representado na seguinte equação:

$$M_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{n^{(l-1)}} K_{i,j}^{(l)} * M_j^{(l-1)}. \tag{2.1}$$

onde $B_i^{(l)}$ é uma matriz contendo o *bias* e $K_{i,j}^{(l)}$ é o filtro a ser aplicado nas entradas.

Outro tipo de camada presente em praticamente todas as CNNs é a camada ReLU. Na prática uma camada ReLU é apenas a aplicação de uma função de ativação. A função linear retificadora (ReLU) é a função de ativação mais utilizada em redes neurais na atualidade, sendo utilizada na quase totalidade das redes convolucionais (LECUN; BENGIO; HINTON, 2015). Em geral, nas arquiteturas de redes convolucionais, a ReLU é aplicada após a convolução com o objetivo de introduzir uma não-linearidade aos resultados obtidos. Uma parte significativa do sucesso das redes convolucionais profundas pode ser atribuído ao uso das ReLUs. Antes dela, redes treinadas com funções de ativação diferentes, como a tangente hiberbólica ou a sigmóide, encontravam problemas no treinamento que dificultavam a convergência final do resultado devido aos problema do *vanishing gradient* e *exploding gradient* (GLOROT; BORDES; BENGIO, 2011). Além das vantagens já citadas, as ReLUs são mais eficientes computacionalmente, pois ela é uma função linear trivial com truncagem em zero (a figura 1 mostra o gráfico da função):

$$\mathbf{y} = \max(\mathbf{0}, \mathbf{x}). \tag{2.2}$$

As camadas totalmente conectadas são geralmente utilizadas no final de uma CNN, com o objetivo de realizar o mapeamento da representação obtida pelas camadas anteriores (convolucionais e de *Pooling*) para os rótulos de classe. O cálculo em cada camada totalmente conectada é descrito por

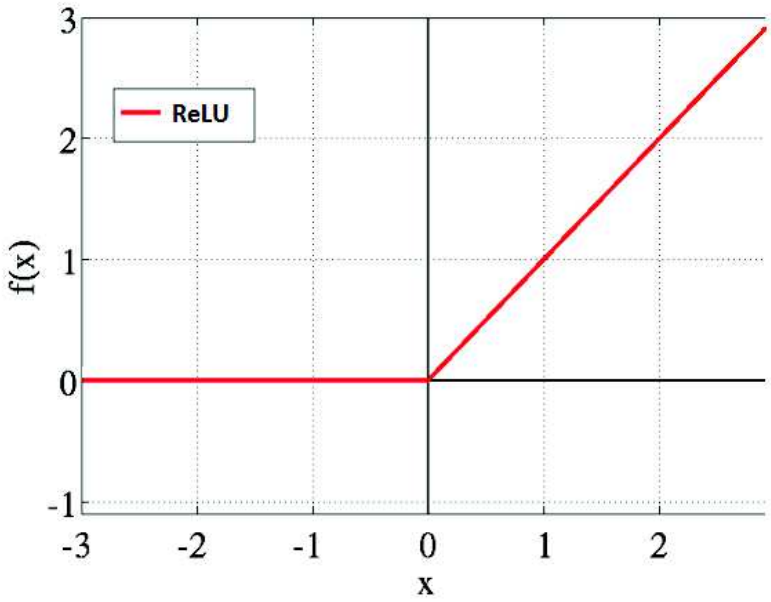
$$\vec{y_l} = W_l^T \vec{y}_{l-1} \tag{2.3}$$

onde $\vec{y_l}$ representa a saída da camada atual l. W representa os pesos da camada atual e $\vec{y}_{l-1}$ é a saída da camada anterior. Os neurônios em uma camada totalmente conectada têm conexões com todas as ativações na camada anterior, exatamente como uma rede neural MLP padrão.

2.3.4 Arquiteturas

Conforme mostrado anteriormente, existem diversos tipos de camadas que podem ser utilizadas na criação de uma CNN e estas podem ser combinadas de diferentes formas. Inspirado na rede *LeNet* (LECUN et al., 1998), as arquiteturas de CNN tendem a seguir estruturas similares

Figura 1: Função ReLU



Fonte: Adaptado de Glorot, Bordes e Bengio (2011)

formadas por pilhas de camadas convolucionais em geral seguidas por uma camada de normalização e uma de *Pooling*. Após isto, normalmente existem uma ou mais camadas totalmente conectadas que funcionam como um MLP realizando a classificação final.

Na maior parte dos trabalhos atuais, as arquiteturas de CNN utilizadas tendem a ser ou as mesmas que obtiveram os melhores resultados na base *ImageNet* ou variações destas. Nas próximas sub-seções são apresentadas de forma breve algumas das arquiteturas de rede mais importantes em uso na atualidade: a *GoogLenet*, a *ResNet* e a *VggNet*, todas utilizadas no desenvolvimento deste trabalho.

2.3.4.1 GoogLenet

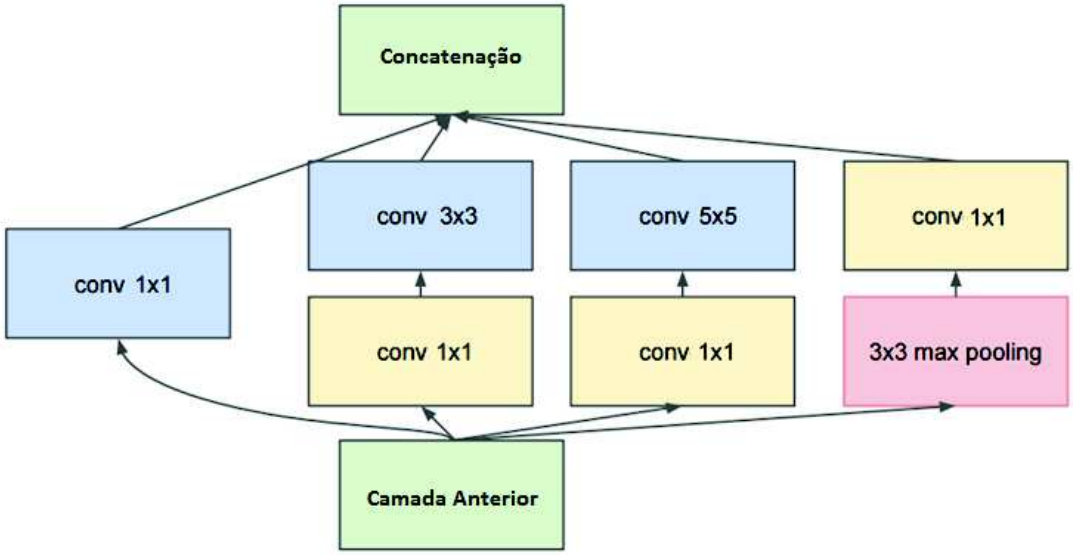
A *GoogLenet* foi a arquitetura vencedora do ILSVRC14 (*ImageNet Large Scale Visual Recognition Competion 2014*) (RUSSAKOVSKY et al., 2015) na categoria de classificação obtendo erro de 6.7%. O melhor resultado do ano anterior havia sido 11.7%, obtido por Zeiler e Fergus (2014).

A *GoogLenet* foi planejada buscando criar uma arquitetura de rede que fosse ao mesmo tempo profunda e eficiente computacionalmente, de forma que fosse possível executá-la em dispositivos com recursos limitados. Para isto, foram criados os módulos *Inception* que são o principal diferencial da *GoogLenet* em relação as demais arquiteturas de CNN.

Os módulos *Inception* são formados pela combinação em paralelo de filtros 1x1, 3x3 e 5x5 conforme mostrado na figura 2. O uso de filtros de tamanhos diferentes serve para dois propósitos principais: redução de dimensionalidade e análise de características em diferentes escalas.

Os filtros 1x1 permitem diminuir a dimensionalidade dos dados antes da aplicação de filtros maiores e mais custosos computacionalmente, acelerando assim a execução do processamento na rede. Já o uso de filtros de diferentes tamanhos permite que a informação seja analisada em escalas diferentes, teoricamente permitindo a abstração de diferentes características simultaneamente (SZEGEDY et al., 2015). Após a aplicação dos filtros os resultados de saída são concatenados e propagados até a camada seguinte.

Figura 2: Formato das Camadas *Inception*



Fonte: Adaptado de (SZEGEDY et al., 2015)

A rede *GoogLenet* é formada por um total de 22 camadas levando em conta apenas camadas convolucionais e totalmente conectadas. O total de parâmetros da rede é de 5 milhões. O anexo A traz uma figura demonstrando a estrutura arquitetural da *GoogLenet*.

2.3.4.2 ResNet

Em 2015, a vencedora do ILSVRC foi a *ResNet* de 152 camadas obtendo erro de apenas 5.7%. Criada pela Microsoft Research Asia, esta arquitetura se destaca pela capacidade de formar redes extremamente profundas, chegando a ser usada para criação de uma rede de mais de 1000 camadas (HE et al., 2015b). Percebeu-se que após certa quantidade de camadas, as CNNs encontram um problema de saturação da acurácia seguida de uma rápida degradação dos resultados (HE; SUN, 2015). Para superar estas limitações no treinamento de redes profundas, a *ResNet* aplica o que os autores chamam de *Deep Residual Learning*.

Deep Residual Learning consiste na adição de "conexões atalho", somando o vetor de entrada com a saída das camadas convolucionais. Diferentemente das camadas convolucionais, as conexões atalho estão sempre ativas e os gradientes podem ser retro-propagados facilmente através delas, o que facilita o processo de treinamento tornando-o mais rápido e simples.

Atualmente, a *ResNet* (ou variações dela) é provavelmente a arquitetura de rede mais estudada. Já existem versões mais avançadas desta arquitetura onde o resultado obtido no *ImageNet* é ainda melhor que em 2015 (HE et al., 2016). O anexo B traz um exemplo de arquitetura de uma variação da rede *ResNet* com 34 camadas.

2.3.4.3 VggNet

A arquitetura *VggNet* foi criada pelo Visual Geometry Group de Oxford e tornou-se conhecida devido a seu excelente resultado no ILSVRC 14 onde obteve a primeira colocação na tarefa de localização e a segunda posição na tarefa de classificação.

A *VggNet* foi pensada buscando criar uma arquitetura mais profunda que as existentes até então mas evitando um aumento significativo da quantidade de parâmetros da rede. Para isto, o campo receptivo foi diminuído para 3x3 em todas as camadas. Nas principais arquiteturas existentes até então os tamanhos dos filtros variavam em geral, tendo tamanhos como 7x7 e 11x11.

Existem algumas variações desta arquitetura sendo que as mais utilizadas possuem 16 ou 19 camadas. A *VggNet* apresenta um design extremamente elegante e homogêneo apresentando sempre sequências de 2 ou 3 camadas convolucionais onde é aplicado um filtro de tamanho fixo (3x3), seguido de uma camada de *Pooling*, onde os mapas de características recebidos da camada anterior tem sua altura e largura diminuídos pela metade. Após as camadas convolucionais existe uma sequencia de 3 camadas totalmente conectadas onde é realizada a classificação. O anexo C apresenta a arquitetura da *VggNet* de 19 camadas.

Como foi demonstrado pelos resultados do ILSVRC, a *VggNet* obtém uma *performance* similar a *GoogLenet*, mas possui uma desvantagem importante em relação a esta: a quantidade de parâmetros. A *VggNet* possui aproximadamente 140 milhões de parâmetros, sendo que a *GoogLenet* possui apenas 5 milhões. Isto torna a *VggNet* mais pesada, aumentando o tempo de processamento e a memória necessária para execução da mesma.

2.4 Multiple Instance Learning

No problema padrão de classificação supervisionada existe um rótulo de classe associado a cada amostra da base de dados. No problema de *Multiple Instance Learning* (MIL), os rótulos de classe são associados a conjuntos de amostras chamado de *bags*. As amostras individuais são chamadas de instâncias e o rótulo de cada uma delas não é conhecido (DONG, 2006).

De modo mais formal, Amores (2013) define o problema de MIL da seguinte forma: uma *bag* é um conjunto $X = \{\vec{x_1}, \ldots, \vec{x_N}\}$, onde os elementos $\vec{x_i}$ são os vetores de características chamados instâncias. A cardinalidade N do conjunto não precisa ser fixa, podendo variar entre as diferentes *bags*. Todas as instâncias estão em um espaço (chamado de espaço de instâncias) de dimensão *d* onde $\vec{x_i} \in \mathbb{R}^d$.

A formulação como problema de MIL é útil e natural em diversas áreas como previsão de comportamento de moléculas na fabricação de remédios, área esta que inspirou o primeiro trabalho de MIL (DIETTERICH; LATHROP; LOZANO-PÉREZ, 1997). Para a classificação de imagens de alta resolução, MIL também se encaixa naturalmente como pode-se ver em trabalhos recentes como (HOU et al., 2015) e (VATSAVAI, 2013).

Diferentes autores criaram taxonomias diversas para categorizar as possíveis implementações de MIL. As mais populares são as definidas por Foulds e Frank (2010) e Amores (2013). Aqui é apresentada a taxonomia do segundo autor por razões de coerência, simplicidade e abrangência das definições da mesma.

Amores (2013) divide os possíveis métodos de MIL em 3 paradigmas baseados no nível onde ocorre o aprendizado. Os paradigmas definidos por ele são:

- ***Instance Space:*** Neste paradigma, a informação discriminativa está presente a nível de instâncias e o aprendizado se dá apenas neste nível. Um classificador é treinado para separar as instâncias presentes nas *bags* positivas e negativas. Com base nisso, quando uma nova *bag* de amostras é apresentada ao classificador, a nova classificação é feita através da agregação dos *scores* das instâncias individuais. Este paradigma é baseado apenas em informações locais, sem levar em consideração características globais presentes na *bag*;

- ***Bag Space:*** Em grande parte dos casos, a informação local presente nas instâncias não é suficiente para fazer a separação das classes e se torna necessário utilizar informações globais, a nível de *bags*. Neste paradigma, cada *bag* é tratada como uma entidade única e o processo de classificação busca comparar *bags* inteiras. O espaço de *bags* não é vetorial neste paradigma então o aprendizado é feito a partir de uma função de distância $D(X, Y)$ que compara as duas entidades não vetoriais. Após a definição desta função D, é possível aplicar um classificador padrão na matriz de distâncias gerada por ela;

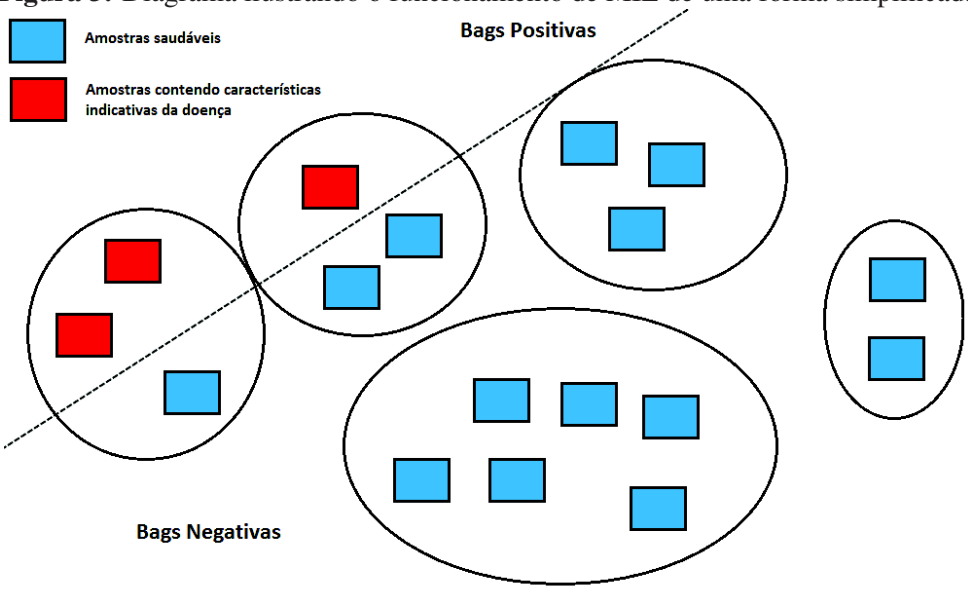
- ***Embedded Space:*** Cada *bag* é mapeada para um único vetor de características que realiza a sumarização da informação sobre toda a *bag*. Assim como no paradigma anterior, o aprendizado é baseado em informações globais da *bag*. A única diferença para o paradigma anterior é que o atual sumariza a informação presente na *bag* de forma vetorial enquanto o anterior faz isto de forma não vetorial.

A figura 3 ilustra de forma simplificada, o funcionamento de MIL. Nesta figura cada quadrado é uma amostra individual (instância) e cada região circulada é uma *bag*. Neste caso, a presença de amostras indicativas da doença dentro da *bag* faz com que ela seja classificada como positiva.

2.5 ImageNet

ImageNet é a maior base de dados de imagens categorizadas em existência na atualidade (RUSSAKOVSKY et al., 2015), contendo mais de 15 milhões de imagens anotadas que perten-

Figura 3: Diagrama ilustrando o funcionamento de MIL de uma forma simplificada



Fonte: Elaborado pelo autor, inspirado por (OUNG et al., 2015)

cem a aproximadamente 22.000 categorias diferentes (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A ideia da criação desta base de dados surgiu pela percepção no campo da visão computacional que seria necessária a existência de bases de dados maiores e melhores para que os próximos avanços na área fossem alcançados (DENG et al., 2009).

Para efetuar a categorização de todas estas imagens de forma adequada foi definida uma metodologia de trabalho apresentada em (DENG et al., 2009). A marcação das imagens foi realizada manualmente através da plataforma *Amazon Mechanical Turk* [1]. Para garantir a confiabilidade das marcações efetuadas cada imagem era marcada por múltiplos usuários independentemente e a marcação de uma imagem apenas era considerada como concluída quando uma quantidade suficiente de usuários categorizava-a da mesma forma. Estima-se que a precisão obtida na marcação seja de aproximadamente 99,7% (RUSSAKOVSKY et al., 2015).

Desde 2010 é realizada a competição *The ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) com o objetivo de avaliar e incentivar o progresso na área. Na ILSVRC é disponibilizado um subconjunto das imagens da base de dados contendo 1.000 categorias, aproximadamente 1,2 milhão de imagens de treinamento, 50.000 imagens de validação e 150.000 imagens de teste (que são substituídas a cada nova edição do desafio). Há grande diversidade de domínios entre as 1.000 categorias existentes no *ImageNet*, buscando-se reunir na mesma base de dados categorias que dificilmente seriam encontradas juntas, indo desde canecas até tanques de guerra, leões, raquetes de beisebol, entre outros.

Desde sua criação, o ILSVRC tem cumprido seu objetivo de fomentar o avanço da visão computacional. No ano inicial, o erro obtido na tarefa de classificação foi de 28,2%. Em 2015, o erro do vencedor da competição foi de apenas 3%. Obviamente, o desenvolvimento das

[1] https://www.mturk.com/

Figura 4: Amostras da *ImageNet* ilustrando a diversidade das imagens presentes na base de dados

Fonte: (T-SNE VISUALIZATION OF CNN CODES, 2017)

Redes Neurais Convolucionais pode ser considerado o motor principal neste grande avanço dos resultados. Entretanto sem a existência de uma base de dados de *benchmark*, confiável e diversa o suficiente, o progresso das redes convolucionais provavelmente seria mais lento.

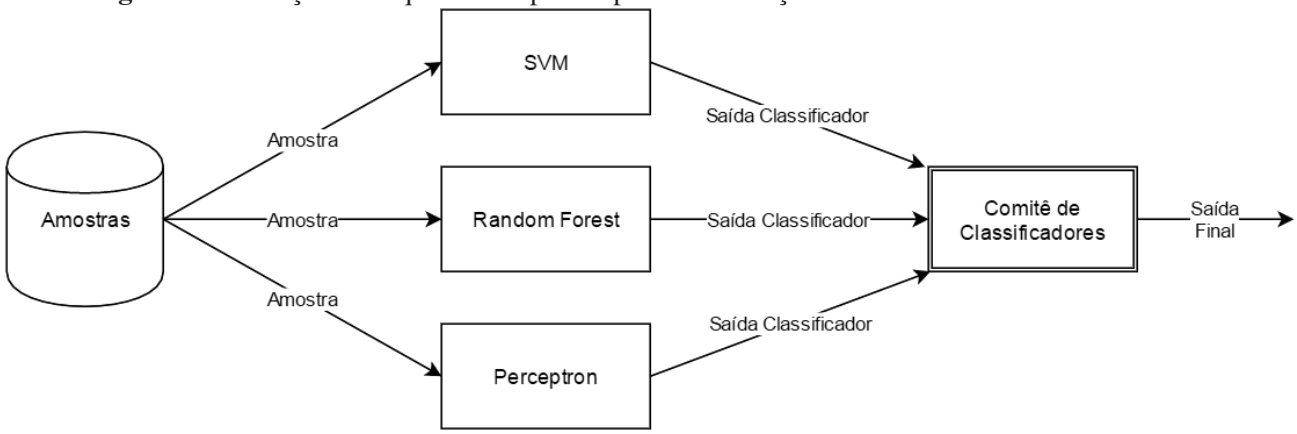
Desde Razavian et al. (2014), sabe-se que uma rede convolucional treinada na base *ImageNet* consegue obter um aprendizado tão abrangente que a torna capaz funcionar como um extrator genérico de características. Isto já foi confirmado em trabalhos onde redes treinadas na *ImageNet* foram aplicadas a bases de dados de domínios diversos obtendo resultados competitivos ou muitas vezes superando o estado da arte anterior: reconhecimento de produtos (JURASZEK, 2014), classificação de imagens aéreas (PENATTI; NOGUEIRA; SANTOS, 2015), detecção de pedestres (ANGELOVA; KRIZHEVSKY; VANHOUCKE, 2015), etc.

2.6 Comitês de Classificadores

Sabe-se que a combinação de múltiplos classificadores individuais, usando metodologias de comitê de classificadores, é capaz de gerar novos classificadores com melhores índices de acertos (OPITZ; MACLIN, 1999; DIETTERICH, 2000; POLIKAR, 2006; ROKACH, 2010).

Um dos fatores mais importantes na criação de comitês de classificadores é a diversidade na

Figura 5: Ilustração da sequência de passos para classificação usando comitês



Fonte: Elaborado pelo autor

saída dos classificadores base, isto é, os erros cometidos pelos classificadores base não sejam correlacionados (DIETTERICH, 2000).

Para garantir a diversidade de saídas nos classificadores base, estes devem ser construídos de formas variadas. As variações mais usadas na construção são a utilização de um mesmo algoritmo de aprendizado com variação de parâmetros, utilização de algoritmos de aprendizado diferentes e a utilização de amostras diferentes no treinamento de cada classificador (SANTANA, 2012; HAYKIN, 1998).

Após a execução de todos os classificadores base, o comitê deve definir sua saída através da sumarização das respostas dos mesmos. Em geral a saída do comitê é definida de uma das seguintes formas: pela soma das saídas dos classificadores base fazendo com que a classe com maior valor seja a resposta do comitê, pelo cálculo de uma média ponderada das saídas dos classificadores base ou através de votação onde a classe com maior número de votos é definida como a resposta do comitê (SANTANA, 2012). A figura 5 ilustra de forma simplificada o funcionamento de um comitê de classificadores.

2.7 Métricas de Avaliação

As métricas de avaliação utilizadas durante o desenvolvimento deste trabalho foram a acurácia, especificidade, sensitividade e AUC (*Area Under the Curve*), seguindo o apresentado em (JAEGER et al., 2014a).

A acurácia pode ser definida como a fração de casos onde a resposta obtida foi correta (METZ, 1978). A fórmula para o cálculo da acurácia é apresentada na eq. 2.4. Ela tende a ser a métrica de avaliação mais popular em *Machine Learning*, entretanto possui alguns pontos fracos que diminuem sua utilidade em casos de bases de dados desbalanceadas. Por exemplo, caso esteja sendo avaliada a presença de uma doença e existam poucas amostras de casos de pessoas contaminadas: se do total de amostras, 95% são amostras de casos saudáveis e apenas 5% de casos não-saudáveis. Neste caso um classificador que classifique todas as amostras

como saudáveis teria acurácia altíssima (95%), porém seria inútil na prática, pois seria incapaz de detectar a doença.

$$Acuracia = \frac{PositivosVerdadeiros + NegativosVerdadeiros}{TotalDeAmostras} \tag{2.4}$$

Para casos como este, é recomendado a utilização da sensitividade e da especificidade como métricas de avaliação (METZ, 1978). A sensitividade, também conhecida como taxa de positivos verdadeiros ou *recall*, mede a proporção de casos positivos que foram classificados corretamente. Já a especificidade mede a proporção de casos negativos que foram classificados corretamente. Também é conhecida como taxa de negativos verdadeiros. As fórmulas para o cálculo de ambas são mostradas em 2.5 e 2.6:

$$Sensitividade = \frac{PositivosVerdadeiros}{TotalDePositivos} \tag{2.5}$$

$$Especificidade = \frac{NegativosVerdadeiros}{TotalDeNegativos} \tag{2.6}$$

Na prática, pode-se dizer que tanto a sensitividade quanto a especificidade são tipos diferentes de acurácia: a primeira uma acurácia dos casos positivos e a segunda dos casos negativos (METZ, 1978).

Uma forma comum de combinar a sensitividade e a especificidade em uma única análise é através dos gráficos da curva ROC (*Receiving Operating Characteristic*). Nos gráficos ROC são plotados os valores da sensitividade no eixo y e da taxa de falsos positivos (cujo valor é igual a 1-especificidade) permitindo assim visualizar o *trade-off* da classificação, os benefícios (positivos verdadeiros) e custos (falsos positivos) (FAWCETT, 2004).

Para gerar o gráfico ROC é necessário utilizar um classificador que juntamente com a classe a que pertence cada instância, retorne um índice de confiança ou probabilidade da resposta obtida estar correta. Este índice pode ser utilizado como limiar para produzir uma classificação discreta. Cada possível valor de limiar forma então um ponto no gráfico ROC. Em muitos casos pode ser relevante resumir a curva ROC a um único valor escalar representando a *performance* esperada do classificador. A forma mais comum de fazer isto é calculando a área sob a curva (Area Under Curve - AUC). A AUC possui algumas propriedades matemáticas importantes. A principal delas é que seu valor é equivalente a probabilidade que o *score* obtido por uma amostra positiva escolhida aleatoriamente seja maior que o *score* de uma amostra negativa escolhida aleatoriamente (FAWCETT, 2006).

2.8 Considerações

Neste capítulo foram apresentados os principais conceitos que são necessários para o entendimento do problema tratado neste trabalho. Primeiramente foi apresentada a tuberculose, mostrou-se que a doença, apesar de estar controlada nos países desenvolvidos, ainda causa mais

de 1 milhão de mortes por ano no mundo. Foram apresentadas as principais formas de diagnóstico da doença, mostrando que as formas mais confiáveis de diagnóstico ainda têm custo elevado e que um método barato, acurado e automático, como o proposto neste trabalho, teria grande potencial de salvar vidas.

Da seção 2.2 até a 2.4 foram apresentados os conceitos-chave para entendimento das propostas presentes neste trabalho: Redes Neurais Artificiais, Redes Neurais Convolucionais, *Multiple Instance Learning* e Comitês de Classificadores. Todas as propostas presentes neste trabalho são possíveis formas de aplicar as redes convolucionais para classificação de imagens radiográficas com o objetivo de detectar a tuberculose, sendo que a segunda e a terceira proposta trazem uma combinações de técnicas (CNNs com MIL em uma e CNNs com Comitês de classificadores na outra) ainda não apresentadas na literatura.

Na seção 2.5 foi apresentada a base de dados *ImageNet*, usada para treinar as redes aplicadas no desenvolvimento do trabalho e na seção 2.6 foram mostradas as métricas de avaliação utilizadas no trabalho.

3 TRABALHOS RELACIONADOS

A literatura já apresenta uma grande quantidade de trabalhos buscando aplicar CNNs na detecção de doenças com variados graus de sucesso. Seguindo o viés já presente na área de CAD, a maior parte dos trabalhos aplicam CNNs para análise de imagens cerebrais ou na detecção de tipos de câncer.

Entre as aplicações relacionadas a diagnósticos na região cerebral destaca-se Havaei et al. (2015) onde o objetivo é detectar um tipo específico de tumor chamado Glioblastoma. A grande contribuição deste trabalho foi propor uma nova arquitetura de rede convolucional que combina características locais e globais na classificação. O sistema criado com base na proposta, além de superar o estado da arte anterior, também é 30 vezes mais rápido que os concorrentes. Em Zhang et al. (2015), uma CNN é treinada para segmentar as regiões do cérebro (matéria branca, cinzenta e fluído cérebro espinhal) em imagens de ressonância magnética infantis (*Magnetic Ressonance Imaging* - MRI). O grande diferencial deste trabalho está na criação de uma CNN capaz de receber múltiplas imagens de MRI geradas com diferentes parâmetros e analisá-las conjuntamente para produzir a segmentação. A rede criada superou com folga as alternativas presentes na literatura.

Aplicações na detecção de nódulos, até o momento, trouxeram menos avanços do que aplicações na região cerebral. Em Ginneken et al. (2015), uma rede neural convolucional pré-treinada na base de dados *ImageNet* é aplicada a detecção de nódulos em tomografias pulmonares, um domínio totalmente diferente do original. O vetor de saída da primeira camada totalmente conectada é usado como entrada para uma SVM linear que foi treinada com 10-*fold cross validation*. Foi realizada uma comparação com um sistema comercial de CAD chamado MeVis Medical Solutions aprovado pela FDA (Federal Drug Administration), aplicando ambos os sistemas em uma base de dados contendo aproximadamente 1000 imagens. O sistema comercial se mostrou consistentemente superior, porém o uso do sistema comercial em conjunto com a rede convolucional trouxe melhorias de resultado permitindo o aumento da sensitividade de 0,68 para 0,71. Em Ciompi et al. (2015), a mesma arquitetura de rede e a mesma base de dados são aplicadas para classificação de um tipo específico de nódulos pulmonares, os nódulos perifissurais. Neste caso a comparação é feita com outras propostas presentes na literatura, obtendo-se AUC de 0,868, superior às alternativas já publicadas e próximo do resultado obtido por especialistas marcando as imagens manualmente.

Apesar da radiografia da região torácica ser um dos exames mais utilizados no mundo, sistemas e publicações de CAD voltados para esta área específica são limitados. Segundo Maduskar et al. (2013), que realizaram análise comparativa de sistemas comerciais de CAD, não existem sistemas disponíveis capazes de fazer uma leitura acurada de radiografias torácicas genéricas. Já há porém sistemas capazes de identificar com sucesso anormalidades específicas presentes nas imagens radiográficas, como nódulos (GINNEKEN; HOGEWEG; PROKOP, 2009) por exemplo, e estes podem ser utilizados como ferramenta auxiliar no diagnóstico de doenças como

câncer de pulmão (DOI, 2005). Como exemplos de trabalhos buscando a detecção de anormalidades específicas pode-se citar Shen, Cheng e Basu (2010) onde é proposta uma técnica de detecção e segmentação de cavidades em radiografias torácicas com o uso do coeficiente inverso do gradiente e medidas de avaliação de circularidade, juntamente com um classificador *Bayesiano*. Em Xu, Cheng e Mandal (2011), o objetivo também é detectar cavidades porém a proposta traz um modelo de casamento de *templates* realçados através de uma técnica baseada no cálculo da matriz Hessiana. Bar et al. (2015) buscam detectar diversos tipos de possíveis anormalidades nas radiografias torácicas entre elas a efusão pleural através de *Deep Learning* e Cherry et al. (2014) propõem um técnica de detecção de alargamento dos nódulos linfáticos. Outro trabalho interessante é (HOGEWEG et al., 2015), onde é feita a comparação de técnicas de detecção de nódulos presentes na literatura com um software comercial (ClearRead+Detect v5.2; Riverain Technologies). O software comercial obteve melhor resultados com acurácia de 0,75.

Trabalhos visando especificamente a criação de sistemas de CAD voltados para a tuberculose são pouco numerosos. Van Ginneken et al. (2002) propuseram a aplicação de um banco de filtros multi-escala nas imagens pulmonares. Posteriormente é realizada a classificação através de um esquema de vizinhos mais próximos ponderados e a validação é realizada através de *Leave One Out Cross Validation* (*LOOCV*). A técnica foi aplicada em 2 base de dados privadas obtendo AUC de 0,82 e 0,986, respectivamente. Já Hogeweg et al. (2010) trazem um sistema baseado em detecção de tuberculose em texturas na região pulmonar que inclui uma supressão da região da clavícula para evitar falsos positivos. Os resultados de AUC alcançados variaram entre 0,67 e 0,86. Tan et al. (2012) trazem uma abordagem que exige a intervenção humana, selecionando as regiões da radiografia a serem analisadas. Destas regiões são extraídas estatísticas da distribuição dos pixels em tons de cinza e são classificadas por uma árvore de decisão.

A comparação destes trabalhos é difícil pois aplicam técnicas totalmente diversas a bases de dados privadas e pequenas (em pelo menos 1 caso a base continha menos de 100 amostras). As primeiras duas base de dados públicas de radiografias torácicas para detecção de tuberculose foram apresentadas em Jaeger et al. (2014b) e desde então tornou-se possível realizar comparações diretas entre trabalhos. Nas próximas subseções são analisados em detalhe 3 trabalhos publicados desde então, sendo que 2 deles já utilizam as novas bases. O terceiro trabalho foi incluído aqui devido a ser considerado relevante, já que sua proposta assim como o presente trabalho, inclui uma técnica de MIL.

3.1 Triagem Automática da Tuberculose

Jaeger et al. (2014a) propuseram combinar diversos algoritmos de visão computacional para a extração de características das imagens de Raio-X. A abordagem adotada no trabalho é dividida em 3 etapas: segmentação da região de interesse, onde os pulmões são detectados na radiografia a partir de um método de otimização *graph cut* em conjunto com uma modelagem

do pulmão. Mais detalhes sobre esta técnica podem ser obtidos no capítulo 4. Em seguida, na segunda fase são computadas as características presentes na imagem de acordo com todos os algoritmos definidos. Na fase final, as características extraídas são utilizadas como entrada para um classificador binário que indica se a radiografia é saudável ou não.

Os algoritmos usados na segunda etapa foram subdivididos em 3 categorias:

- **Algoritmos de detecção de objetos:** Foram computados o Histograma de Intensidades (IH), histogramas de Magnitudes do Gradiente (GM), histogramas do Descritor de Formas (SD), histogramas do Descritor de Curvaturas (CD), Histograma dos Gradientes Orientados (HOG) e Padrões Locais Binários (LBP);

- **Algoritmos de CBIR:** Este conjunto de técnicas extrai um total de 594 características relacionadas a intensidade, bordas, texturas e momentos. Todos os algoritmos utilizados neste passo são parte da biblioteca LIRE (LUX; CHATZICHRISTOFIS, 2008). Pode-se citar entre os principais algoritmos utilizados, o descritor de texturas de Tamura, momentos invariantes de Hu (HU, 1962), texturas de Gabor, autocorrelação, etc;

- **Algoritmos relacionadas ao formato:** As características de detecção de formato foram extraídas com a ferramenta *regionprops* do *MATLAB*. Elas descrevem a orientação, excentricidade, tamanho, centróide e *bounding box*.

O sistema foi treinado utilizando os classificadores MLP, SVM (linear, RBF e polinomial), árvores de decisão e regressão logística. Foi também aplicada uma técnica de seleção def características, não especificada, para remover características não relevantes para classificação. Para validação dos resultados obtidos foi utilizada a técnica *LOOCV*. As bases de dados onde os testes foram realizados foram a Montgomery e Shenzen, disponibilizadas pelo próprio autor.

Os melhores resultados foram obtidos com a regressão logística e com a SVM linear. Obteve-se AUC de 0,87 e acurácia 0,78 para a base Montgomery e na base Shenzen obteve-se 0,90 de AUC e 0,84 de acurácia.

3.2 Nova Abordagem para Detecção da Tuberculose Através de MIL

Melendez et al. (2015) trazem uma abordagem para detecção da doença baseada em uma adaptação das SVMs para que consigam classificar conjuntos de amostras em lugar de amostras individuais.

Na abordagem adotada, cada radiografia é dividida em sub-regiões. Os rótulo de classe de cada sub-região (indicando se a região contém indícios da tuberculose) não são conhecidos. Conhece-se apenas o rótulo da imagem inteira (o que na terminologia de MIL seria chamado de *bag*). Com o problema formulado desta maneira, uma SVM padrão não é capaz de traçar o hiperplano de separação a nível de recorte, mas sim apenas a nível de *bag*. Há várias possíveis formas de realizar a classificação, sendo que uma das mais utilizadas é baseada em (ANDREWS; TSOCHANTARIDIS; HOFMANN, 2002), onde os autores reformulam o problema

de separação da SVM para que ela passe a classificar grupos de amostras, em vez de amostras individuais isoladas (esta nova formulação é conhecida como miSVM). Melendez et al. (2015) apresentam extensões a esta formulação buscando adicionar estimativas de probabilidade, diminuir o tempo de convergência da SVM e melhorar os resultados de classificação em casos de *bags* positivas contendo amostras normalmente indicativas de casos negativos.

Foram utilizadas 3 bases de dados neste trabalho, nenhuma delas disponível publicamente. Elas são chamadas de Zâmbia, Gambia e Tanzânia. A primeira consiste em 917 imagens, sendo que 525 contém a doença e 392 são amostras saudáveis. Já a base Tanzânia possui 452 imagens de pulmões com a tuberculose e de 417 de pulmões sem ela. Por fim, a base Gambia é composta de 400 imagens com a tuberculose e 450 sem ela para um total de 850 imagens. Todas as imagens foram redimensionadas para terem 1024 pixels de largura e foram divididas em sub-regiões com espaçamento de 8 pixels e 32 pixels de raio. O vetor de características de cada sub-região é obtido através do cálculo dos momentos das distribuições de intensidade dos pixels presentes em cada uma delas.

Os resultados de AUC (infelizmente o autor não publicou os resultados em termos de acurácia) obtidos neste trabalho não são diretamente comparáveis aos obtidos nos demais trabalhos apresentados neste capítulo, devido as bases de dados utilizadas serem diferentes dos demais trabalhos. De qualquer forma, pode-se notar que a AUC obtida parece ser competitiva em relação as dos outros trabalhos: 0,86 nas bases de dados Zambia e Tanzania e 0,91 na base Gambia.

3.3 Detecção da Tuberculose Através de CNNs

Hwang et al. (2016) é um trabalho relevante principalmente por ser o primeiro onde CNNs foram utilizadas na detecção da tuberculose. A abordagem adotada busca treinar uma CNN de forma integral, criando uma arquitetura específica e ajustando os pesos da rede para melhor resolução do problema.

A arquitetura proposta consiste em uma variação da rede *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A principal alteração realizada na rede base foi a inclusão de uma camada convolucional e um *Max Pooling* a mais na entrada. Isto foi necessário devido a resolução da camada de entrada da rede *AlexNet* ser 225x225 e as imagens de treinamento terem resolução de 500x500 pixels.

O treinamento foi realizado de duas formas: a partir do zero, com pesos iniciais gerados aleatoriamente e também utilizando *Transfer Learning* (TL) (PAN; YANG, 2010). A utilização do TL visava facilitar o aprendizado da CNN. Assim buscou-se aproveitar as duas primeiras camadas de filtros aprendidos pela rede *AlexNet* na base *ImageNet*. Os filtros aprendidos nestas camadas tendem a ser genéricos, representando características mais abstratas como bordas e curvas que teoricamente estarão presentes em qualquer tipo de imagem natural. Desta forma, a CNN tenderá a aprender com mais facilidade pois suas primeiras camadas já trazem valores próximos do ideal. A tabela 1 mostra as características de cada camada da rede implementada

Tabela 1: Arquitetura da CNN de Hwang et al. (2016)

Camada	Tipo	Entrada	Tamanho dos filtros
C1	Convolução	(1,500,500)	(96,1,11,11)
M1	Max Pool		
C2	Convolução	(96,61,61)	(256,96,5,5)
M2	Max Pool		
C3	Convolução	(256,30,30)	(384,256,3,3)
C4	Convolução	(384,30,30)	(384,384,3,3)
C5	Convolução	(384,30,30)	(256,384,3,3)
M5	Max Pool		
C6	Convolução	(256,15,15)	(256,256,3,3)
M6	Max Pool		(3,3)
F7	Totalmente Conectada	(256,7,7)	(2048,12544)
D7	Dropout		
F8	Totalmente Conectada	(2048)	(2048,2048)
D8	Dropout		
F9	Totalmente Conectada	(2048)	(2,2048)

Fonte: Adaptado de Hwang et al. (2016)

no trabalho.

Foram utilizadas 3 bases de dados no desenvolvimento do trabalho. Além das bases de dados Montgomery e Shenzen, o autor obteve acesso a uma vasta base de dados privada criada pelo Instituto Coreano da Tuberculose (chamada KIT). Esta base consiste em 10.848 imagens, sendo que 7.020 normais e 3.828 com a doença.

Durante a fase de treinamento foram aproveitadas apenas as imagens da base KIT. As amostras foram divididas aleatóriamente em conjunto de treinamento (70%), conjunto de validação (15%) e conjunto de testes (15%). Para validação utilizou-se 3-*fold cross validation*. As demais bases de dados não foram aproveitadas durante o treinamento, mas sim apenas posteriormente para avaliação do modelo treinado.

No aprendizado a partir do zero foi constatado que a rede não conseguiu obter resultados satisfatórios, com acurácia de 0,77 e AUC de 0,816 (na base KIT). Já no caso da rede treinada com *Transfer Learning* foi possível obter números bastante superiores com acurácia de 0,903 e AUC de 0,964.

Quando o modelo treinado é aplicado em outras bases de dados os resultados são inferiores porém ainda competitivos. No conjunto de dados Montgomery foi obtida a acurácia de 0,783 e AUC de 0,884 enquanto na base Shenzen obteve-se 0,837 de acurácia e 0,926 de AUC.

Neste trabalho pode-se considerar que os grandes diferenciais foram o uso do *Transfer Learning* para facilitação do treinamento e também o aproveitamento da grande base de dados KIT. Nenhuma das bases públicas tem uma quantidade tão grande de amostras e por isso não seria possível treinar uma CNN utilizando apenas elas. Caso a KIT seja disponibilizada será um importante passo para a criação de detectores de tuberculose baseados em *Deep Learning*.

3.4 Considerações

Neste capítulo foi apresentado um breve histórico da literatura científica contendo pesquisas relacionadas a aplicações de CNNs para classificação de imagens médicas, com destaque para as regiões do cérebro e dos pulmões, além dos principais trabalhos publicados até o momento contendo pesquisas sobre a detecção da tuberculose em imagens radiográficas torácicas frontais, independentemente do tipo de método proposto. Foram apresentados também em detalhes os três trabalhos recentes mais relevantes sobre a detecção da tuberculose: (JAEGER et al., 2014a; MELENDEZ et al., 2015; HWANG et al., 2016).

No pesquisa de Jaeger, as principais contribuições não são necessariamente as propostas de técnicas de classificação da doença apresentadas. O trabalho de Jaeger faz parte de um projeto maior cujos frutos incluem a criação de algoritmos de segmentação pulmonar (CANDEMIR et al., 2014), a disponibilização das primeiras bases de dados públicas para classificação da tuberculose (JAEGER et al., 2014b), publicação de trabalho de revisão bibliográfica extenso e detalhado (JAEGER et al., 2013), entre outros.

No trabalho de Melendez é feita uma abordagem de MIL que difere consideravelmente da adotada nos testes preliminares do presente trabalho. Na abordagem adotada no presente trabalho, as amostras individuais (que representam sub-regiões da imagem) são agrupadas em um único descritor global de características que é classificado em uma SVM convencional. Já na abordagem de Melendez, o próprio algoritmo de classificação é reformulado para classificar grupos de amostras. É importante notar que o objetivo da proposta do autor não é apenas realizar a classificação correta das radiografias, mas sim também inferir o posicionamento das lesões, nódulos e demais características indicativas da tuberculose na imagem.

Em relação ao terceiro artigo, que traz uma proposta com CNN, a abordagem difere em um ponto principal das propostas do presente trabalho: a CNN foi arquitetada e os pesos foram recalculados para adaptá-la ao domínio, enquanto a rede utilizada no presente trabalho teve seus pesos calculados na base *ImageNet* e utiliza uma arquitetura já pré-definida. Esta adaptação da rede ao domínio só foi possível devido aos autores terem obtido acesso a base de dados KIT, pois sem ela não haveriam amostras suficientes para treinamento de uma CNN. Outra diferença que merece ser citada, entre o presente trabalho e a proposta de Hwang, é a resolução da imagem de entrada da rede. A rede utilizada pelos autores tem cerca de 500x500 pixels na camada de entrada, enquanto a utilizada no presente trabalho tem apenas 224x224 pixels. Esta camada de entrada maior permite que menos informação seja perdida quando for realizado o redimensionamento da imagem para processá-la, o que teoricamente ajudará a melhorar os resultados de classificação.

O principal ponto comum a todas as abordagens é que todos eles buscam maximizar a acurácia ou a AUC. Obviamente, é importante obter um classificador o mais acurado possível, entretanto para que estes sistemas sejam úteis no mundo real é indispensável que consigam obter baixo índice de falsos negativos. Com exceção de Jaeger, que traz a preocupação em diminuir

os falsos negativos em trabalhos futuros, os demais autores não citam esta necessidade.

4 MATERIAIS E MÉTODOS

Este capítulo está dividido em cinco seções apresentando detalhadamente os principais fatores relacionados aos experimentos realizados. A seção 4.1 apresenta as bases de dados que foram utilizadas durante o desenvolvimento do trabalho e a seção 4.2 expõe as principais ferramentas aplicadas no mesmo. A seção 4.3 apresenta as 3 propostas de utilização das CNNs pré-treinadas como extratores de características, assim como a técnica de pré-processamento empregada. As seções 4.4 e 4.5 trazem respectivamente a apresentação dos resultados atingidos e uma discussão dos mesmos.

4.1 Bases de Dados

Todos os experimentos realizados para este trabalho utilizaram as bases de dados Shenzen e Montgomery disponibilizados em (JAEGER et al., 2014b).

A base Montgomery consiste em 138 radiografias frontais do tórax, sendo que 80 imagens são casos de pessoas saudáveis e 58 casos possuem alguma manifestação da tuberculose. As imagens foram coletadas pelo Departamento de saúde do condado de Montgomery em Maryland nos Estados Unidos. Os tamanhos das imagens são 4020x4892 ou 4892x4020 pixels. Na base de dados Montgomery, ainda há imagens adicionais contendo máscaras de segmentação dos pulmões geradas manualmente (sob supervisão de radiologistas) para cada uma das amostras. A figura 6 exibe exemplos de radiografias pertencentes a esta base.

A base de dados Shenzen foi coletada no hospital *Guandong* em Shenzen na China. No total o Shenzen contém 662 imagens de Raio-X torácico frontal, onde 326 delas são casos normais e 336 são casos com manifestações da tuberculose. Todas as imagens possuem a resolução de aproximadamente 3000x3000 pixels e foram capturadas utilizando o aparelho *Philips DR Digital Diagnost System* (PHILIPS DR DIGITAL DIAGNOST, 2016). A figura 7 contém amostras pertencentes a esta base.

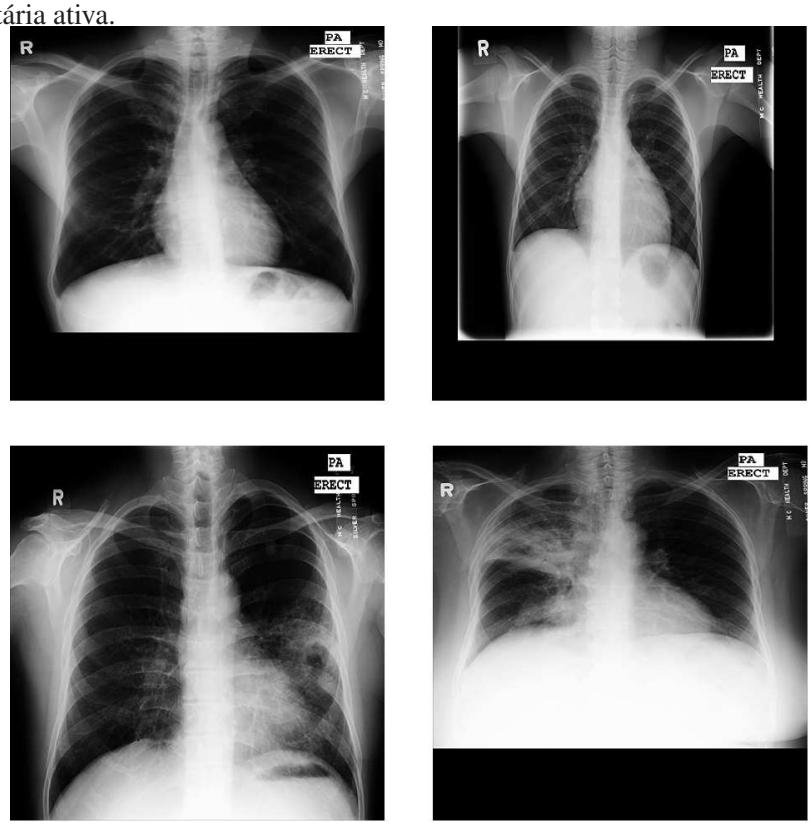
Além das imagens, foram disponibilizados arquivos de texto (que seguem os nomes das imagens, excetuando a extensão que é .txt) com informações adicionais sobre o paciente e seu status. Estas informações adicionais contidas nos arquivos de texto não foram utilizadas em nenhuma das abordagens propostas, devido a ambiguidade da mesma e a necessidade de aprofundamento maior no conhecimento do domínio.

4.2 Ferramentas

Durante o desenvolvimento deste trabalho foram avaliadas e utilizadas diversas ferramentas para o pré-processamento das imagens, para a extração de características, para classificação final e visualização dos resultados.

Todos os pré-processamentos criados especificamente para este trabalho (redimensiona-

Figura 6: Exemplos de radiografias na base Montgomery. As radiografias no canto superior direito e esquerdo exibem, respectivamente, pulmões saudáveis de um homem de 33 anos e de um garoto de 8. Ambas as radiografias na região inferior da figura são casos de infecção pela tuberculose: a radiografia a esquerda pertence a um homem com 54 anos de idade com infiltrações em ambos os pulmões e uma cavidade na língula. Na radiografia a direita, há sinais de infiltrações pulmonares consistentes com uma tuberculose cavitária ativa.



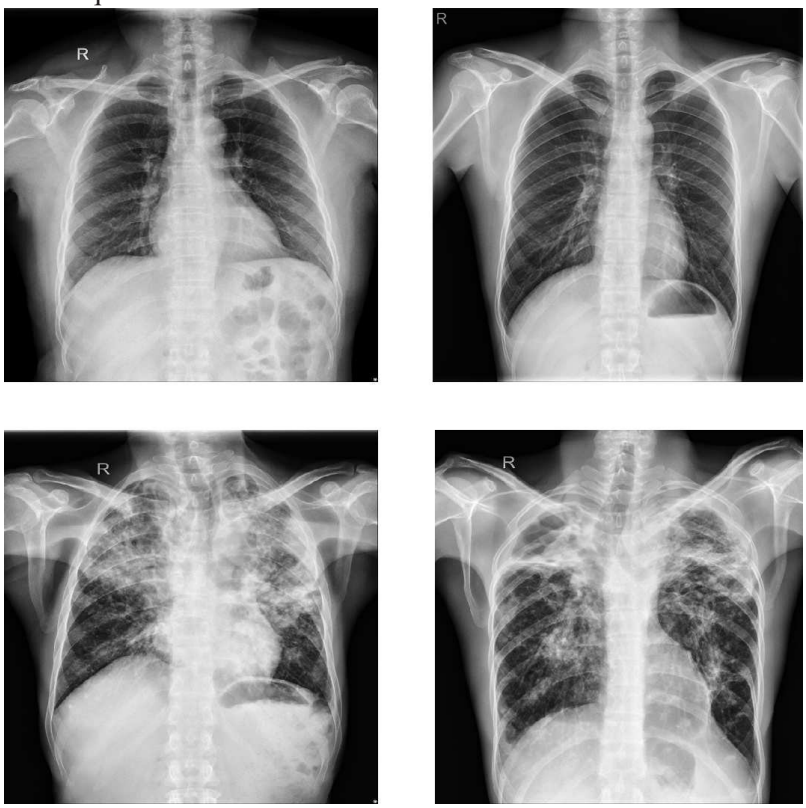
Fonte: Antani (2015)

mento de imagens, recortes, segmentação, etc) foram implementados utilizando o ambiente *Matlab 2015a* ou a linguagem C++ juntamente com a biblioteca *OpenCV* (versão 2.4.9).

Para a extração de características utilizando redes convolucionais foram avaliadas as bibliotecas *MatConvNet* (VEDALDI; LENC, 2015), *Caffe* (JIA et al., 2014) e *TensorFlow* (ABADI et al., 2015). A biblioteca *TensorFlow*, disponibilizada recentemente pelo Google possui a arquitetura mais flexível porém no momento do desenvolvimento do trabalho ela foi descartada pelo complexidade e falta de documentação. Tanto a biblioteca *Caffe* quanto a *MatConvNet*, possuem boa documentação e comunidades grandes e ativas. No final, optou-se pela *MatConvNet* devido a simplicidade e intuitividade da API.

A etapa final de classificação das imagens foi desenvolvida utilizando a biblioteca da *libsvm* (CHANG; LIN, 2012). Para visualização de resultados e geração de gráficos optou-se pela biblioteca *SciKit* de Python (PEDREGOSA et al., 2011).

Figura 7: Exemplos de radiografias da base Shenzen. As radiografias na metade superior são casos de pulmões saudáveis: a radiografia a esquerda é de uma mulher de 48 anos e a radiografia a direita pertence a um homem com 24 anos de idade. As duas imagens na metade inferior são casos de tuberculose bilateral secundária: a esquerda um caso em um homem de 56 anos e a direita em um homem de 26.



Fonte: Antani (2015)

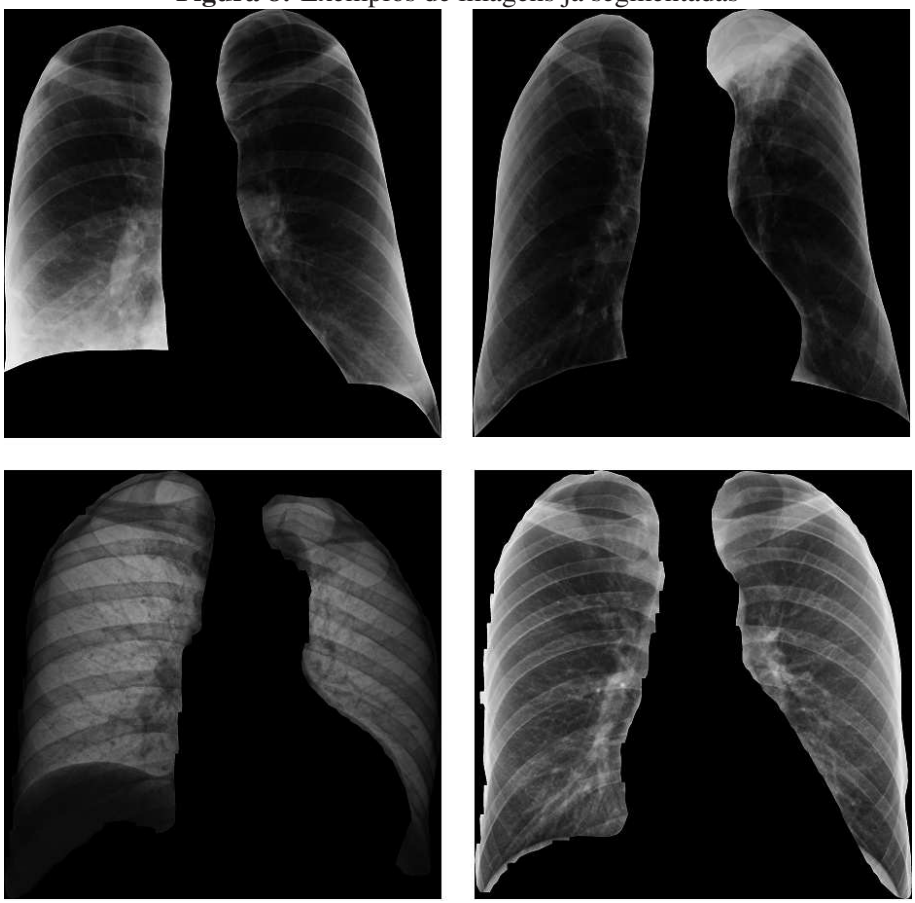
4.3 Métodos

Nesta seção são apresentadas as informações relacionadas à metodologia adotada no desenvolvimento do trabalho. A seção está dividida em quatro subseções contendo respectivamente a descrição do pré-processamento dos dados (4.3.1), da Proposta 1 (4.3.2), da Proposta 2 (4.3.3) e da Proposta 3 (4.3.4).

4.3.1 Pré-Processamento

Como pode-se observar nas figuras contendo amostras de radiografias presentes nas bases de dados, além dos pulmões há também a presença de outras regiões do tórax que não são relevantes para a detecção da tuberculose. Na base Montgomery é possível realizar a segmentação das regiões de interesse de forma trivial (pois a base inclui máscaras criadas manualmente indicando quais pixels fazem parte da região pulmonar) porém na Shenzen esta tarefa exige passos adicionais. Para realizar a segmentação na base Shenzen utilizou-se a ferramenta disponibilizada por Candemir et al. (2012), cujo funcionamento pode ser resumido em 3 estágios:

Figura 8: Exemplos de imagens já segmentadas



Fonte: Elaborado pelo autor

- **Estágio 1:** Na primeira etapa é realizada uma busca utilizando métodos de CBIR (*Content Based Image Retrieval*) em uma base de dados anotada por especialistas (chamada de *Atlas set*). Neste passo, as 5 imagens do *Atlas set* mais similares a imagem do paciente são retornadas para uso no próximo estágio;

- **Estágio 2:** É realizado um mapeamento da transformação entre a imagem do paciente e das imagens obtidas no estágio anterior utilizando o algoritmo *SIFT Flow* (LIU et al., 2008). Primeiramente este algoritmo modela a informação dos gradientes da imagem usando SIFT. Em seguida, um algoritmo de minimização calcula a transformação entre a imagem do paciente e cada imagem retornada no estágio anterior. Os parâmetros de transformação são então utilizados para realizar o alinhamento entre as imagens. O modelo dos pulmões do paciente é obtido pela média das mascáras calculadas;

- **Estágio 3:** Como refinamento final da segmentação calcula-se uma otimização discreta com o algoritmo *Graph Cuts* e uma função de energia customizada. Busca-se assim encontrar um mínimo global para cada pixel que corresponda ao fundo (fora do pulmão) ou ao primeiro plano (pixels dentro do pulmão).

A segmentação gerada pelos passos anteriores foi avaliada pelos autores em três bases de

dados obtendo acurácia de 0,95, 0,94 e 0,92, demonstrando assim robustez para conjuntos diversos de radiografias (CANDEMIR et al., 2014). A figura 8 mostra exemplos de de segmentações obtidas aplicando a técnica.

Após a criação das máscaras, elas são utilizadas para remover todos os pixels não pertencentes aos pulmões. Em seguida, a imagem é reduzida para o menor *bounding box* contendo todos os pixels presentes nos pulmões.

Para a etapa seguinte foram implementadas 3 variações discutidas nas próximas subseções. Para a Proposta 1, ainda é necessária mais uma etapa simples de pré-processamento onde é realizado um redimensionamento da imagem para as dimensões da camada de entrada da rede (224x224 pixels) para que ela possa ser processada.

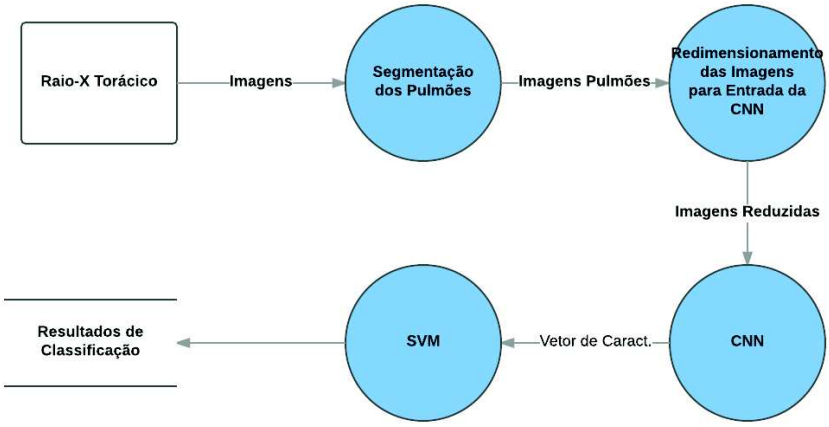
4.3.2 Proposta 1

A primeira abordagem busca avaliar de forma simples a capacidade de uma CNN pré-treinada na base *ImageNet* em extrair características relevantes para classificação das imagens radiográficas. A figura 9 mostra as etapas da Proposta 1.

A imagem contendo os pulmões segmentados é redimensionada para as dimensões da entrada de cada uma das três arquiteturas de redes convolucionais (224x224 para a *GoogLenet* e a *VggNet* e 227x227 para *ResNet*) e são então propagadas na rede. Após a propagação extrai-se o vetor de saída da última camada totalmente conectada de cada rede que será utilizado para o treinamento e classificação da SVM.

O passo seguinte é o treinamento de uma SVM. Foram avaliadas SVMs com kernel linear e RBF. Para validação utilizou-se LOOCV e para seleção dos melhores parâmetros para o classificador foi realizada uma busca em *grid* simples com o parâmetro C variando entre 1 e 1000. Para o parâmetro *gamma* a variação do valor foi entre entre 1/4096 e 1.

Figura 9: Diagrama de fluxo de dados da classificação das imagens na Proposta 1



Fonte: Elaborado pelo autor

4.3.3 Proposta 2

Conforme comentado anteriormente, a Proposta 1 tem uma grande desvantagem. Com o redimensionamento da imagem para a dimensão de entrada das redes, perde-se muita informação que provavelmente seria útil para a identificação de sinais da tuberculose.

Decidiu-se então pela modelagem como um problema de MIL, de forma a aproveitar ao máximo a informação contida nas imagens das radiografias. A forma mais simples de usar imagens de alta resolução em uma rede com entrada pequena seria dividir a imagem original em sub-janelas (com a mesma dimensão da entrada da rede) e rotular cada uma das sub-janelas indicando se em cada região da imagem existe alguma anormalidade indicativa da doença. Provavelmente esta abordagem seria mais promissora, entretanto não foi possível realizá-la devida as bases de dados utilizadas no trabalho não possuírem anotações indicando o posicionamento dos padrões indicativos da doença. As bases de dados apenas trazem a indicação se a radiografia é de um pulmão saudável ou não.

Assim sendo, foi necessária a modelagem como MIL onde cada radiografia é uma *bag* e cada sub-janela é uma instância. As sub-janelas não possuem rótulos de classe, apenas a radiografia inteira possui. Após estas definições iniciais é necessário tomar decisões sobre os detalhes da modelagem pois existem dezenas de possíveis formas de implementar MIL.

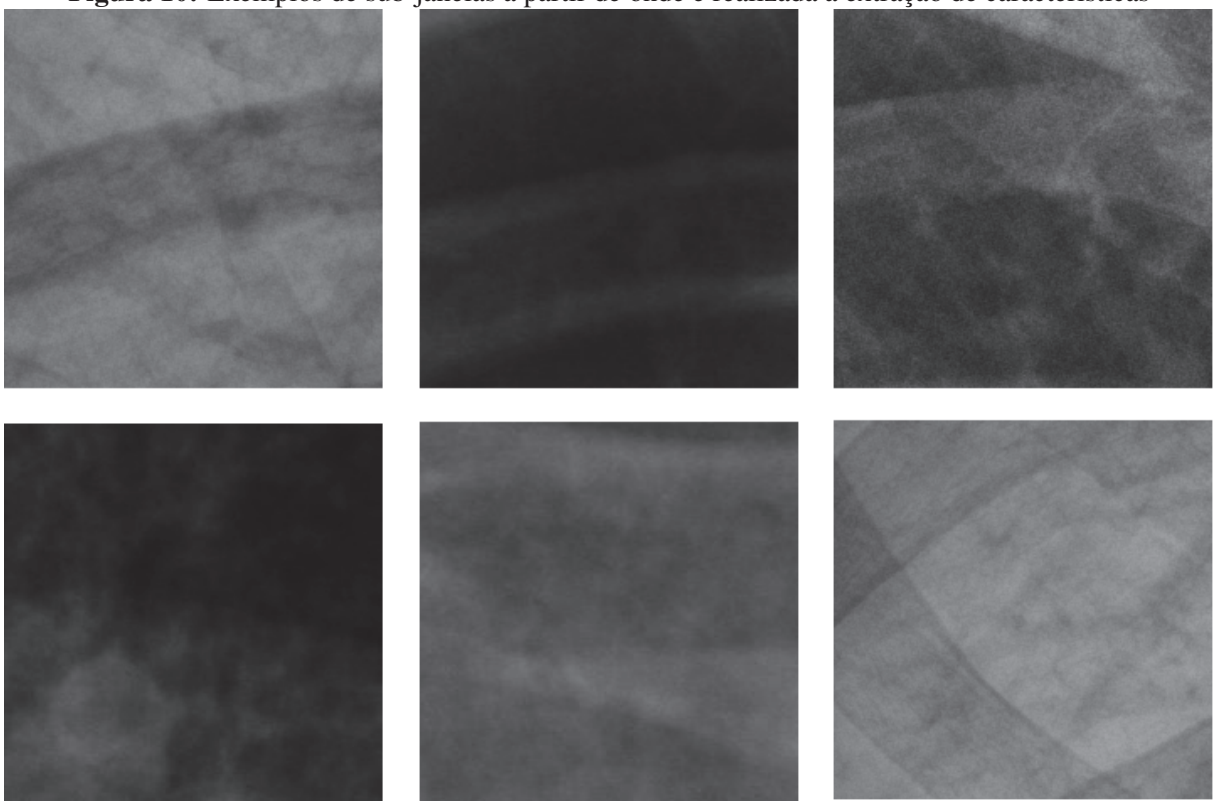
Para o trabalho atual decidiu-se modelar o problema como um caso de MIL no paradigma *Embedded Space*. Em essência, a técnica utilizada é uma aplicação do *Bag-Of-Words Model* proposto por Csurka et al. (2004) em um contexto levemente diferente do original.

O primeiro passo, que pode ser considerada uma fase de pré-treino, é a geração de um dicionário de características visuais através da clusterização dos vetores extraídos de cada uma das instâncias presentes em cada *bag*. Cada instância é na verdade uma das sub-janelas, cujas dimensões são iguais as dimensões da camada de entrada da CNN. Definiu-se deixar 50% de sobreposição entre as janelas tanto no sentido horizontal quanto vertical para evitar que características visuais importantes caíssem em janelas diferentes. Cada janela é propagada pela rede e, assim como na Proposta 1, extraí-se a saída da última camada totalmente conectada. A partir destes vetores é criado o dicionário, que é usado para gerar o descritor global da *bag*, que consiste em um histograma H cujos valores representam a quantidade de instâncias presentes na *bag* que faz parte de cada um dos *clusters* encontrados na etapa anterior. A figura 10 mostra exemplos das sub-regiões das radiografias onde as CNNs são aplicadas.

Após a obtenção de um descritor global da *bag*, o problema se torna um caso de aprendizado supervisionado convencional, pois agora cada *bag* possui um único vetor de características e um rótulo associado. Portanto pode-se realizar a classificação das *bags* utilizando qualquer classificador padrão. No presente trabalho, decidiu-se utilizar a SVM como classificador. Os parâmetros utilizados no treinamento são os mesmos da Proposta 1.

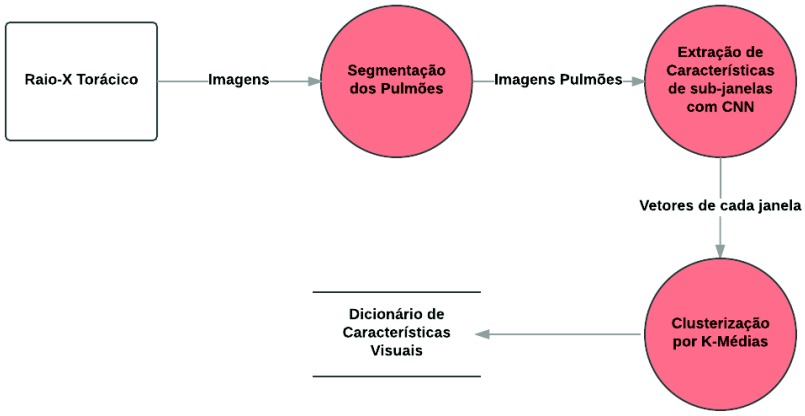
As figuras 11 e 12 mostram respectivamente as etapas de geração de dicionário de características e de classificação das imagens de acordo com o definido na Proposta 2.

Figura 10: Exemplos de sub-janelas a partir de onde é realizada a extração de características



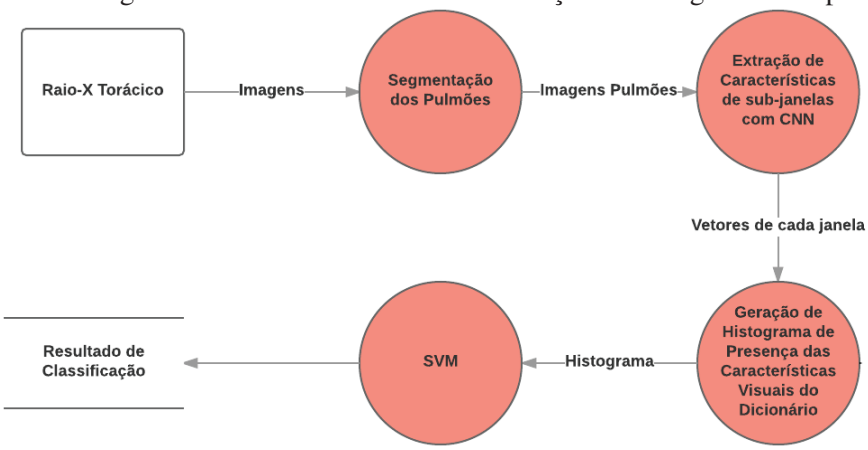
Fonte: Elaborado pelo autor

Figura 11: Diagrama de fluxo de dados de criação do dicionário de características visuais.



Fonte: Elaborado pelo autor

Figura 12: Diagrama de fluxo de dados da classificação das imagens na Proposta 2



Fonte: Elaborado pelo autor

4.3.4 Proposta 3

Como dito em seções anteriores deste trabalho, já existem trabalhos aplicando CNNs como extratores de características para classificação de imagens médicas. No entanto, o presente trabalho é o primeiro a realizar a criação de comitês a partir da combinação de múltiplos classificadores treinados a partir de características extraídas por CNNs.

A terceira e última proposta deste trabalho é a criação de comitês de classificadores combinando as melhores SVMs treinadas nas propostas anteriores. A figura 13 demonstra o funcionamento dos comitês de classificadores criados para este trabalho.

No total, 4 comitês de classificadores foram criados e avaliados neste trabalho: dois criados a partir da combinação dos classificadores da Proposta 1 (uma para cada conjunto de dados) e dois criados a partir dos melhores classificadores da Proposta 2 (também um para cada base de dados).

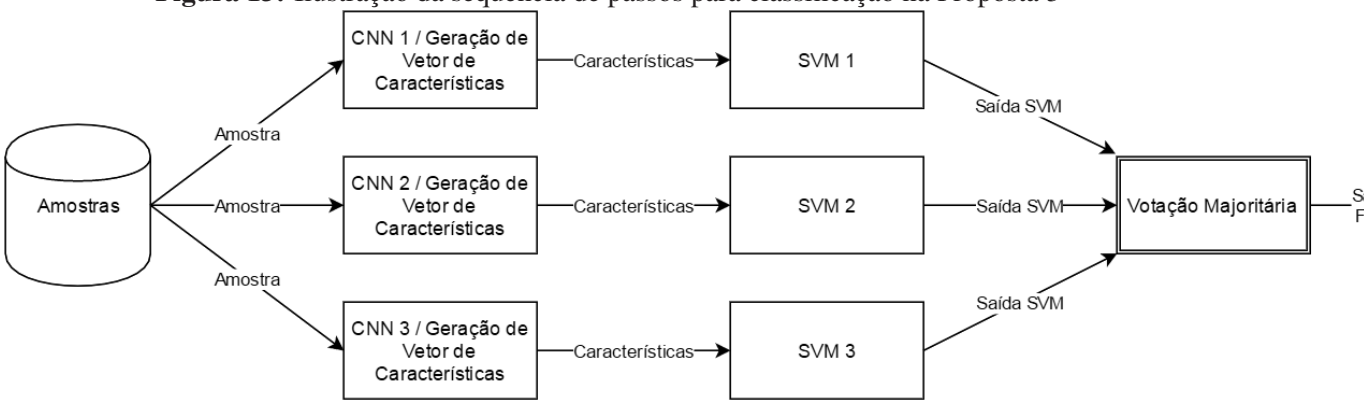
Cada um dos comitês criados combina três classificadores (um usando características extraídas usando a *GoogLenet*, um com características da *ResNet* e um com características da *VggNet*). Para obter o resultado de classificação é feita uma votação majoritária simples. Para obtenção da AUC de cada comitê é calculada a média das probabilidades de cada classificador individual.

4.4 Resultados

As tabelas 2 e 3 mostram os resultados para a Proposta 1 em termos de acurácia e AUC. Para a base de dados Montgomery os melhores resultados foram obtidos pela *GoogLenet* com acurária de 0,812 e AUC de 0,821. Na base Shenzen a rede *VggNet* se mostrou superior com acurácia de 0,856 e AUC de 0,917.

Pode-se notar nas primeiras duas tabelas que os números obtidos na base Shenzen são su-

Figura 13: Ilustração da sequência de passos para classificação na Proposta 3



Fonte: Elaborado pelo autor

Tabela 2: Proposta 1 - Montgomery

	GoogLenet	ResNet	VggNet
Acurácia	0,812	0,804	0,79
AUC	0,821	0,777	0,748

periores aos da base Montgomery. O mesmo padrão pode ser visto nas tabelas seguintes e em trabalhos similares publicados (JAEGER et al., 2014a; HWANG et al., 2016). Provavelmente isto se deve em parte ao maior desbalanceamento existente entre as classes na base Montgomery (onde 60% das amostras são negativas e 40% positivas) em comparação com a Shenzen (onde a divisão é praticamente meio a meio). Outro fator que possivelmente está prejudicando os resultados é a quantidade demasiadamente pequena de amostras na Montgomery: apenas 138. Já a base Shenzen possui 662 amostras.

As tabelas 4 a 7 exibem os resultados para a segunda proposta exibindo a melhor acurácia e AUC para cada tamanho de dicionário (parâmetro K).

Na base de dados Montgomery a arquitetura *ResNet* supera todas as outras obtendo os melhores resultados tanto em acurácia quanto em AUC para quase todos os valor de K. A melhor acurácia obtida foi 0,848 com $K = 300$ (tabela 4) e a melhor AUC foi 0,901 obtida por $K = 400$ e $K = 500$ (tabela 5).

Na base Shenzen, os resultados da *ResNet* pioraram consideravelmente sendo inferiores as outras CNNs para quase todos os valores de K. A *GoogLenet* obteve a melhor *performance* nesta base de dados com uma acurácia de 0,867 (tabela 6) e AUC de 0,921 (tabela 7). A rede *VggNet* ficou um pouco atrás com acurácia de 0,867 (tabela 6) e AUC de 0,914 (tabela 7).

Como pode ser visto nas tabelas de resultados, os melhores valores para o parâmetro K variam entre 200 e 500. Em nenhum dos resultados o valor de $K = 100$ obteve resultados competitivos. Isto indica que um dicionário de tamanho 100 é pequeno demais para representar todas as características relevantes extraídas das amostras de dados.

Tabelas 8 e 9 mostram os resultados obtidos pelos comitês de classificadores construídos

Tabela 3: Proposta 1 - Shenzen

	GoogLenet	ResNet	VggNet
Acurácia	0,843	0,849	0,856
AUC	0,908	0,909	0,917

Tabela 4: Proposta 2 - Acurácia - Montgomery

	GoogLenet	ResNet	VggNet
K=100	0,812	0,797	0,768
K=200	0,826	0,833	0,826
K=300	0,804	0,848	0,775
K=500	0,819	0,841	0,833

usando as SVMs das Propostas 1 e 2 respectivamente. Como esperado seus resultados são consistentemente superiores as demais propostas. Uma constatação interessante e inesperada é que o comitê formado pelos classificadores da Proposta 1 conseguiu atingir uma AUC de 0,924 na base Shenzen, o mesmo resultado do comitê de classificadores da Proposta 2.

As últimas duas tabelas, 8 e 9, exibem a comparação entre os melhores resultados obtidos em cada proposta e os resultados de trabalhos similares publicados sobre detecção de tuberculose. As colunas P1 e P2 referem-se as Propostas 1 e 2 respectivamente. As colunas CP1 e CP2 referem-se aos comitês criados usando os classificadores treinados nas Propostas 1 e 2 respectivamente.

Em termos de acurácia, todas as propostas apresentadas no presente trabalho superam os trabalhos similares presentes na literatura. Os melhores resultados, como esperado, foram obtidos pelo comitê de classificadores da Proposta 2, com acurácia de 0,869 na base Montgomery e 0,872 na Shenzen.

Em relação a AUC, a Proposta 2 e a Proposta 3 obtiveram uma boa *performance*. Na base Montgomery o CP2 foi o grande vencedor com resultado de 0,902. Já na base Shenzen, as propostas aqui apresentadas foram ligeiramente superadas pela rede de Hwang et al que obteve AUC de 0,926. O melhor resultado para esta base entre as propostas do presente trabalho (em termos de AUC) foi 0,924.

Como pode ser visto nas tabelas comparativas, as propostas apresentadas pelo presente trabalho são superiores na grande maioria dos casos e, no pior dos casos, são competitivas com a literatura corrente.

4.5 Discussão dos Resultados

Os resultados obtidos com a Proposta 1 são bastante surpreendentes pois é inesperado que uma imagem redimensionada para menos de 1/10 do tamanho original ainda contenha infor-

Tabela 5: Proposta 2 - AUC - Montgomery

	GoogLenet	ResNet	VggNet
K=100	0,867	0,848	0,836
K=200	0,851	0,899	0,802
K=300	0,837	0,901	0,807
K=500	0,843	0,923	0,890

Tabela 6: Proposal 2 - Acurácia - Shenzen

	GoogLenet	ResNet	VggNet
K=100	0,850	0,823	0,847
K=200	0,855	0,847	0,861
K=300	0,867	0,823	0,856
K=500	0,861	0,829	0,867

mação suficiente para obter acurácia acima de 85%. A Proposta 2 trouxe resultados superiores aos da Proposta 1, mas por uma margem menor do que esperado. As principais desvantagens trazidas por esta abordagem são a maior complexidade, devido à existência de etapas extras de geração de dicionário e de histograma. Outro possível problema é o maior tempo de execução podendo chegar a ser aproximadamente 60 vezes maior do que o tempo de execução da Proposta 1. Esta última desvantagem, o tempo de execução, pode ser diminuída paralelizando a extração de características das janelas com uma GPU.

Como esperado, a Proposta 3 (Comitês de Classificadores) trouxe os melhores resultados. Um dos mais importantes requerimentos para criação de comitês de classificadores é a diversidade de erros, isto é, os erros dos classificadores base devem ter baixa correlação (DIETTERICH, 2000). Como todas as CNNs foram treinadas na mesma base de dados (*ImageNet*) havia o risco que suas saídas fossem demasiadamente similares o que prejudicaria os comitês. Felizmente este não foi o caso. Como pode-se ver, na maioria dos testes os comitês são superiores aos classificadores base mostrando a viabilidade e potencial deste tipo de técnica.

Entre as arquiteturas de CNN avaliadas, nenhuma é claramente superior as outras. A *ResNet* exibiu melhores resultados na base Montgomery enquanto a *GoogLenet* e a *VggNet* foram superiores na Shenzen. Poderia ser esperado que a *ResNet* seria a melhor em todos os testes, já que seus resultados são claramente superiores na *ImageNet*, mas não foi o caso. A arquitetura extremamente profunda da *ResNet*, com 152 camadas, parece ser exagerada para tarefas de classificação de imagens médicas. Para *benchmarks* na *ImageNet* faz bastante sentido criar redes extremamente profundas para poder abranger toda a diversidade e vastidão da base de dados. Já para imagens médicas onde a variabilidade dos dados é ordens de magnitude menor, parece ser desnecessário o uso de tantas camadas para aprimoramento dos resultados.

Tabela 7: Proposta 2 - AUC - Shenzen

	GoogLenet	ResNet	VggNet
K=100	0,907	0,887	0,913
K=200	0,915	0,901	0,912
K=300	0,918	0,891	0,912
K=500	0,921	0,902	0,914

Tabela 8: Comitê de CNNs - Extração simples de características

	Acurácia	AUC
Montgomery	0,847	0,812
Shenzen	0,865	0,924

Tabela 9: Comitê de CNNs - *Bag* de características de CNNs

	Acurácia	AUC
Montgomery	0,869	0,902
Shenzen	0,872	0,927

Tabela 10: Comparação - Acurácia

	Jaeger et al	Hwang et al	P1	P2	EP1	EP2
Montgomery	0,783	0,674	0,797	0,848	0,847	0,869
Shenzen	0,84	0,837	0,85	0,867	0,867	0,872

Tabela 11: Comparação - AUC

	Jaeger et al	Hwang et al	P1	P2	EP1	EP2
Montgomery	0,869	0,884	0,793	0,901	0,812	0,902
Shenzen	0,900	0,926	0,903	0,921	0,924	0,924

5 CONCLUSÃO

No presente trabalho foram apresentadas três diferentes propostas para aplicação de CNNs pré-treinadas para detecção de tuberculose. Na primeira proposta, três diferentes arquiteturas de CNN foram utilizadas para extração de características de radiografias pulmonares para posterior classificação através de uma SVM. Na segunda proposta, as mesmas três arquiteturas foram usadas para extrair características de sub-regiões de cada radiografia. As características extraídas então são usadas para criação de um único descritor global por radiografia que é então usado para treinar uma SVM. Na última proposta, as melhores SVMs criadas em cada uma das propostas anteriores são combinadas para gerar comitês de classificadores.

As contribuições mais importantes do presente trabalho foram a apresentação de uma análise comparativa da *performance* como extratores de características de 3 das mais importantes arquiteturas de redes convolucionais em uma base de dados de imagens médicas, a proposta de uma nova combinação de redes pré-treinadas e *Multiple Instance Learning* e a avaliação do uso de comitês de classificadores treinados em características extraídas das CNNs pré-treinadas como alternativa à combinação de características *hand-crafted* e CNNs. Até onde pode-se verificar, é a primeira vez que todas estas contribuições são apresentadas na literatura.

O uso de redes pré-treinadas normalmente não é a forma recomendada de aplicação de CNNs a tarefas de classificação de imagens médicas. Se a base de dados é grande o suficiente, em geral o melhor caminho é treinar a rede do zero ou realizar fine-tuning em uma rede existente (TAJBAKHSH et al., 2016). Para algumas tarefas se demonstrou que CNNs onde foi executado fine-tuning superam consistentemente os resultados obtidos por redes pré-treinadas (SHIN et al., 2016), o que faz sentido já que os filtros aprendidos são otimizados para maximizar a acurácia da tarefa de classificação específica. Dados estes resultados, a importante questão que fica é se ainda faz sentido utilizar redes pré-treinadas como ferramenta para classificação de imagens médicas em algum caso.

A literatura corrente sugere que características de redes pré-treinadas conseguem melhores resultados quando utilizadas em combinação com características *hand-crafted* especificamente para o domínio da base de dados (SHIN et al., 2016; BAR et al., 2015; RIBEIRO et al., 2016).

Baseado nos resultados obtidos no presente trabalho, pode-se afirmar que a criação de comitês de classificadores treinados usando diferentes arquiteturas de CNNs parece ser um promissor e sub-explorado caminho alternativo para aplicação de CNNs à classificação de imagens médicas. Os comitês de classificadores criados no presente trabalho obtiveram resultados superiores ao estado da arte anterior mesmo quando comparados com uma rede *fine-tuned*. Claro que a comparação não é totalmente justa já que as imagens das bases Montgomery e Shenzen não foram usadas no *fine-tuning* da CNN de Hwang. De qualquer forma, os resultados indicam que, no pior dos casos, comitês de classificadores construídos a partir de características extraídas de CNNs pré-treinadas podem ser mais uma poderosa ferramenta na classificação de imagens médicas.

Em relação a sistemas para detecção da tuberculose, trabalhos recentes como (ESTEVA et al., 2017), sugerem o provável melhor caminho para evolução de sistemas CAD. Neste trabalho, uma versão mais recente da *GoogLenet* (usando a *Inception* V3) é treinada em cerca de 130 mil imagens dermatológicas para detecção de câncer de pele e obtém resultados equiparáveis aos obtidos por dermatologistas treinados. Isto sugere que o melhor caminho para o avanço no desenvolvimento de sistemas CAD passa pela criação de bases de dados cada vez maiores que deverão ser usadas para o treinamento de CNNs. Como *Deep Learning* é neutro em relação ao tipo de imagem utilizada pode-se esperar com alta confiança que os bons resultados obtidos nas imagens dermatológicas se repitam em radiografias pulmonares, possibilitando assim a criação de sistemas cada vez melhores para detecção da tuberculose.

5.1 Trabalhos Futuros

Existem diversos possíveis caminhos para evolução das técnicas apresentadas neste trabalho. Entre as possíveis linhas de investigação, pode-se destacar as seguintes:

- **Comitês de classificadores:** O caminho mais promissor dentro os apresentados neste trabalho provavelmente é o aprofundamento dos estudos da aplicação de comitês de classificadores criados a partir de diferentes CNNs. Os comitês apresentados neste trabalho usam um sistema simples de votação majoritária. Pode ser interessante avaliar diferentes métodos de votação, assim como novas arquiteturas de CNN;

- **Avaliar diferentes tipos de clusterização:** Na abordagem que utiliza MIL possivelmente a substituição do K-Médias por outro algoritmo de clusterização como EM (Expectation Maximization) seja suficiente para obter melhorias na classificação. Há suporte na literatura para esta especulação (AMORES, 2013);

- **Avaliar inclusão de bases de novas bases de dados:** O trabalho utiliza apenas 2 bases de dados, pois são as únicas bases públicas de radiografias pulmonares contendo casos da tuberculose. Entretanto parecem existir outras bases como a JSRT Digital Image Database (SHIRAISHI et al., 2000) contendo radiografias torácicas com marcações indicando a presença de tumores pulmonares. Talvez seria possível incluir amostras de outras doenças e treinar um classificador melhor capaz de identificar variados tipos de anormalidades pulmonares. Espera-se que o maior número de amostras traga maior poder de generalização e melhore os resultados obtidos até aqui;

- **Diminuição dos falsos negativos:** As métricas utilizadas para avaliação até o momento são importantes porém não levam em conta o uso do algoritmo no mundo real. Para que uma técnica de detecção de tuberculose seja útil fora do mundo acadêmico é importante que ela evite ao máximo classificar uma radiografia doente como sadia. Cada falso negativo neste caso pode vir a ser uma pessoa que não recebeu tratamento adequado para a

doença a tempo e pode tornar-se um caso de óbito. Possíveis abordagens com este foco poderiam utilizar uma variação de classificadores baseados em distância (como K-NN) juntamente com heurísticas visando otimizar o resultado.

REFERÊNCIAS

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M. et al. TensorFlow: large-scale machine learning on heterogeneous systems, 2015. **Software available from tensorflow. org**, [S.l.], v. 1, 2015.

AMORES, J. Multiple Instance Classification: review, taxonomy and comparative study. **Artif. Intell.**, Essex, UK, v. 201, p. 81–105, Aug. 2013.

ANDREWS, S.; TSOCHANTARIDIS, I.; HOFMANN, T. Support vector machines for multiple-instance learning. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2002. **Anais...** [S.l.: s.n.], 2002. p. 561–568.

ANGELOVA, A.; KRIZHEVSKY, A.; VANHOUCKE, V. Pedestrian detection with a large-field-of-view deep network. In: ROBOTICS AND AUTOMATION (ICRA), 2015 IEEE INTERNATIONAL CONFERENCE ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 704–711.

ANTANI, S. Automated Detection of Lung Diseases in Chest X-Rays. **US National Library of Medicine**, [S.l.], 2015.

BAR, Y.; DIAMANT, I.; WOLF, L.; GREENSPAN, H. Deep learning with non-medical training used for chest pathology identification. In: SPIE MEDICAL IMAGING, 2015. **Anais...** [S.l.: s.n.], 2015. v. 9414.

CANDEMIR, S.; JAEGER, S.; PALANIAPPAN, K.; ANTANI, S.; THOMA, G. Graph-cut based automatic lung boundary detection in chest radiographs. In: IEEE HEALTHCARE TECHNOLOGY CONFERENCE: TRANSLATIONAL ENGINEERING IN HEALTH & MEDICINE, 2012. **Anais...** [S.l.: s.n.], 2012. p. 31–34.

CANDEMIR, S.; JAEGER, S.; PALANIAPPAN, K.; MUSCO, J. P.; SINGH, R. K.; XUE, Z.; KARARGYRIS, A.; ANTANI, S.; THOMA, G.; MCDONALD, C. J. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. **IEEE transactions on medical imaging**, [S.l.], v. 33, n. 2, p. 577–590, 2014.

CHANG, C.-C.; LIN, C.-J. LIBSVM: a library for support vector machine, 2001. **Software available at http://www. csie. ntu. edu. tw/˜ cjlin/libsvm**, [S.l.], 2012.

CHELLAPILLA, K.; PURI, S.; SIMARD, P. High performance convolutional neural networks for document processing. In: TENTH INTERNATIONAL WORKSHOP ON FRONTIERS IN HANDWRITING RECOGNITION, 2006, La Baule, France. **Anais...** [S.l.: s.n.], 2006.

CHERRY, K. M.; WANG, S.; TURKBEY, E. B.; SUMMERS, R. M. Abdominal lymphadenopathy detection using random forest. In: SPIE MEDICAL IMAGING, 2014, San Diego, California, USA. **Anais...** [S.l.: s.n.], 2014. v. 9035.

CIOMPI, F.; HOOP, B. de; RIEL, S. J. van; CHUNG, K.; SCHOLTEN, E. T.; OUDKERK, M.; JONG, P. A. de; PROKOP, M.; GINNEKEN, B. van. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. **Medical image analysis**, Amsterdam, Netherlands, v. 26, n. 1, p. 195–202, 2015.

CIRESAN, D. C.; MEIER, U.; MASCI, J.; MARIA GAMBARDELLA, L.; SCHMIDHUBER, J. Flexible, high performance convolutional neural networks for image classification. In: IJCAI PROCEEDINGS-INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2011, Barcelona, Spain. **Anais...** [S.l.: s.n.], 2011. v. 22, n. 1.

CSURKA, G.; DANCE, C.; FAN, L.; WILLAMOWSKI, J.; BRAY, C. Visual categorization with bags of keypoints. In: WORKSHOP ON STATISTICAL LEARNING IN COMPUTER VISION, ECCV, 2004. **Anais...** [S.l.: s.n.], 2004. v. 1, n. 1-22, p. 1–2.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: a large-scale hierarchical image database. In: COMPUTER VISION AND PATTERN RECOGNITION, 2009. CVPR 2009. IEEE CONFERENCE ON, 2009. **Anais...** [S.l.: s.n.], 2009. p. 248–255.

DHUNGEL, N.; CARNEIRO, G.; BRADLEY, A. P. Deep structured learning for mass segmentation from mammograms. In: IMAGE PROCESSING (ICIP), 2015 IEEE INTERNATIONAL CONFERENCE ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 2950–2954.

DIELEMAN, S.; WILLETT, K. W.; DAMBRE, J. Rotation-invariant convolutional neural networks for galaxy morphology prediction. **Monthly Notices of the Royal Astronomical Society**, [S.l.], v. 450, n. 2, p. 1441, 2015.

DIETTERICH, T. G. Ensemble methods in machine learning. In: INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEMS, 2000. **Anais...** [S.l.: s.n.], 2000. p. 1–15.

DIETTERICH, T. G.; LATHROP, R. H.; LOZANO-PÉREZ, T. Solving the multiple instance problem with axis-parallel rectangles. **Artificial intelligence**, [S.l.], v. 89, n. 1, p. 31–71, 1997.

DOI, K. Current status and future potential of computer-aided diagnosis in medical imaging. **The British Journal of Radiology**, [S.l.], v. 78, n. suppl_1, p. s3–s19, 2005. PMID: 15917443.

DONG, L. **A comparison of multi-instance learning algorithms**. 2006. Tese (Doutorado em Ciência da Computação) — The University of Waikato, 2006.

ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, [S.l.], v. 542, n. 7639, p. 115–118, 2017.

FAWCETT, T. ROC graphs: notes and practical considerations for researchers. **Machine learning**, [S.l.], v. 31, n. 1, p. 1–38, 2004.

FAWCETT, T. An introduction to ROC analysis. **Pattern recognition letters**, [S.l.], v. 27, n. 8, p. 861–874, 2006.

FOULDS, J.; FRANK, E. A review of multi-instance learning assumptions. **The Knowledge Engineering Review**, [S.l.], v. 25, n. 01, p. 1–25, 2010.

FUKUSHIMA, K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological cybernetics**, [S.l.], v. 36, n. 4, p. 193–202, 1980.

GINNEKEN, B. van; HOGEWEG, L.; PROKOP, M. Computer-aided diagnosis in chest radiography: beyond nodules. **European Journal of Radiology**, [S.l.], v. 72, n. 2, p. 226–230, 2009.

GINNEKEN, B. van; SETIO, A. A.; JACOBS, C.; CIOMPI, F. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: BIOMEDICAL IMAGING (ISBI), 2015 IEEE 12TH INTERNATIONAL SYMPOSIUM ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 286–289.

GLOBAL tuberculosis report 2015. 2015.

GLOBAL tuberculosis report 2016. 2016.

GLOROT, X.; BORDES, A.; BENGIO, Y. Deep Sparse Rectifier Neural Networks. In: AISTATS, 2011. **Anais...** [S.l.: s.n.], 2011. v. 15, n. 106.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. `http://www.deeplearningbook.org`.

GRAUPE, D. **Principles of artificial neural networks**. [S.l.]: World Scientific, 2013.

GU, J.; WANG, Z.; KUEN, J.; MA, L.; SHAHROUDY, A.; SHUAI, B.; LIU, T.; WANG, X.; WANG, G. Recent Advances in Convolutional Neural Networks. **arXiv preprint arXiv:1512.07108**, [S.l.], 2015.

HAVAEI, M.; DAVY, A.; WARDE-FARLEY, D.; BIARD, A.; COURVILLE, A.; BENGIO, Y.; PAL, C.; JODOIN, P.-M.; LAROCHELLE, H. Brain Tumor Segmentation with Deep Neural Networks. **arXiv preprint arXiv:1505.03540**, [S.l.], 2015.

HAYKIN, S. **Neural Networks**: a comprehensive foundation. 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

HE, K.; SUN, J. Convolutional neural networks at constrained time cost. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015. **Proceedings...** [S.l.: s.n.], 2015. p. 5353–5360.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2015. **Proceedings...** [S.l.: s.n.], 2015.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep Residual Learning for Image Recognition. **arXiv preprint arXiv:1512.03385**, [S.l.], 2015.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Identity mappings in deep residual networks. **arXiv preprint arXiv:1603.05027**, [S.l.], 2016.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, [S.l.], 2012.

HOGEWEG, L.; MOL, C.; JONG, P. A. de; DAWSON, R.; AYLES, H.; GINNEKEN, B. van. Fusion of local and global detection systems to detect tuberculosis in chest radiographs. In: **Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010**. [S.l.]: Springer, 2010. p. 650–657.

HOGEWEG, L.; SÁNCHEZ, C. I.; MADUSKAR, P.; PHILIPSEN, R.; STORY, A.; DAWSON, R.; THERON, G.; DHEDA, K.; PETERS-BAX, L.; GINNEKEN, B. van. Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis. **Medical Imaging, IEEE Transactions on**, [S.l.], v. 34, n. 12, p. 2429–2442, 2015.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the national academy of sciences**, [S.l.], v. 79, n. 8, p. 2554–2558, 1982.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural networks**, [S.l.], v. 2, n. 5, p. 359–366, 1989.

HOU, L.; SAMARAS, D.; KURC, T. M.; GAO, Y.; DAVIS, J. E.; SALTZ, J. H. Efficient Multiple Instance Convolutional Neural Networks for Gigapixel Resolution Image Classification. **arXiv preprint arXiv:1504.07947**, [S.l.], 2015.

HU, M.-K. Visual pattern recognition by moment invariants. **IRE transactions on information theory**, [S.l.], v. 8, n. 2, p. 179–187, 1962.

HUA, K.-L.; HSU, C.-H.; HIDAYATI, S. C.; CHENG, W.-H.; CHEN, Y.-J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. **OncoTargets and therapy**, [S.l.], v. 8, 2015.

HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. **The Journal of physiology**, [S.l.], v. 195, n. 1, p. 215–243, 1968.

HUVAL, B.; WANG, T.; TANDON, S.; KISKE, J.; SONG, W.; PAZHAYAMPALLIL, J.; ANDRILUKA, M.; CHENG-YUE, R.; MUJICA, F.; COATES, A. et al. An Empirical Evaluation of Deep Learning on Highway Driving. **arXiv preprint arXiv:1504.01716**, [S.l.], 2015.

HWANG, S.; KIM, H.-E.; JEONG, J.; KIM, H.-J. A novel approach for tuberculosis screening based on deep convolutional neural networks. In: SPIE MEDICAL IMAGING, 2016. **Anais...** [S.l.: s.n.], 2016. v. 9785.

JAEGER, S.; CANDEMIR, S.; ANTANI, S.; WÁNG, Y.-X. J.; LU, P.-X.; THOMA, G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. **Quantitative imaging in medicine and surgery**, [S.l.], v. 4, n. 6, p. 475–477, 2014.

JAEGER, S.; KARARGYRIS, A.; CANDEMIR, S.; FOLIO, L.; SIEGELMAN, J.; CALLAGHAN, F.; XUE, Z.; PALANIAPPAN, K.; SINGH, R. K.; ANTANI, S. et al. Automatic tuberculosis screening using chest radiographs. **Medical Imaging, IEEE Transactions on**, [S.l.], v. 33, n. 2, p. 233–245, 2014.

JAEGER, S.; KARARGYRIS, A.; CANDEMIR, S.; SIEGELMAN, J.; FOLIO, L.; ANTANI, S.; THOMA, G.; MCDONALD, C. J. Automatic screening for tuberculosis in chest radiographs: a survey. **Quantitative imaging in medicine and surgery**, [S.l.], v. 3, n. 2, p. 89–99, 2013.

JARRETT, K.; KAVUKCUOGLU, K.; LECUN, Y. et al. What is the best multi-stage architecture for object recognition? In: IEEE 12TH INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2009., 2009. **Anais...** [S.l.: s.n.], 2009. p. 2146–2153.

JIA, Y.; SHELHAMER, E.; DONAHUE, J.; KARAYEV, S.; LONG, J.; GIRSHICK, R.; GUADARRAMA, S.; DARRELL, T. Caffe: convolutional architecture for fast feature embedding. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 2014. **Proceedings...** [S.l.: s.n.], 2014. p. 675–678.

JURASZEK, G. D. **Reconhecimento de Produtos por Imagem Utilizando Palavras Visuais e Redes Neurais Convolucionais**. 2014. Dissertação (Mestrado em Ciência da Computação) — UDESC - Universidade do Estado de Santa Catarina, 2014.

KALCHBRENNER, N.; GREFENSTETTE, E.; BLUNSOM, P. A convolutional neural network for modelling sentences. **arXiv preprint arXiv:1404.2188**, [S.l.], 2014.

KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological cybernetics**, [S.l.], v. 43, n. 1, p. 59–69, 1982.

KOSKO, B. Bidirectional associative memories. **IEEE Transactions on Systems, Man, and Cybernetics**, [S.l.], v. 18, n. 1, p. 49–60, 1988.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2012. **Anais...** [S.l.: s.n.], 2012. p. 1097–1105.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, [S.l.], v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. **Neural computation**, [S.l.], v. 1, n. 4, p. 541–551, 1989.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [S.l.], v. 86, n. 11, p. 2278–2324, 1998.

LECUN, Y.; CORTES, C.; BURGES, C. J. **The MNIST database of handwritten digits**. 1998.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. et al. Convolutional networks and applications in vision. In: ISCAS, 2010. **Anais...** [S.l.: s.n.], 2010. p. 253–256.

LEUNG, C. C. Reexamining the role of radiography in tuberculosis case finding. **The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease**, [S.l.], v. 15, n. 10, 2011.

LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft coco: common objects in context. In: **Computer Vision–ECCV 2014**. [S.l.]: Springer, 2014. p. 740–755.

LIU, C.; YUEN, J.; TORRALBA, A.; SIVIC, J.; FREEMAN, W. T. Sift flow: dense correspondence across different scenes. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2008. **Anais...** [S.l.: s.n.], 2008. p. 28–42.

LUX, M.; CHATZICHRISTOFIS, S. A. Lire: lucene image retrieval: an extensible java cbir library. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 16., 2008. **Proceedings...** [S.l.: s.n.], 2008. p. 1085–1088.

MADUSKAR, P.; MUYOYETA, M.; AYLES, H.; HOGEWEG, L.; PETERS-BAX, L.; GINNEKEN, B. van. Detection of tuberculosis using digital chest radiography: automated reading vs. interpretation by clinical officers. **The International Journal of Tuberculosis and Lung Disease**, [S.l.], v. 17, n. 12, p. 1613–1620, 2013.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, [S.l.], v. 5, n. 4, p. 115–133, 1943.

MELENDEZ, J.; GINNEKEN, B. van; MADUSKAR, P.; PHILIPSEN, R. H.; REITHER, K.; BREUNINGER, M.; ADETIFA, I. M.; MAANE, R.; AYLES, H.; SANCHEZ, C. I. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-rays. **Medical Imaging, IEEE Transactions on**, [S.l.], v. 34, n. 1, p. 179–192, 2015.

METZ, C. E. Basic principles of ROC analysis. In: SEMINARS IN NUCLEAR MEDICINE, 1978. **Anais...** [S.l.: s.n.], 1978. v. 8, n. 4, p. 283–298.

MINSKY, M.; PAPERT, S. **Perceptrons**: an introduction to computational geometry. Cambridge, MA, USA: MIT Press, 1969.

OPITZ, D.; MACLIN, R. Popular ensemble methods: an empirical study. **Journal of Artificial Intelligence Research**, [S.l.], v. 11, p. 169–198, 1999.

OUNG, Q. W.; MUTHUSAMY, H.; LEE, H. L.; BASAH, S. N.; YAACOB, S.; SARILLEE, M.; LEE, C. H. Technologies for assessment of motor disorders in Parkinson's disease: a review. **Sensors**, [S.l.], v. 15, n. 9, p. 21710–21745, 2015.

PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on knowledge and data engineering**, [S.l.], v. 22, n. 10, p. 1345–1359, 2010.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: machine learning in python. **Journal of Machine Learning Research**, [S.l.], v. 12, n. Oct, p. 2825–2830, 2011.

PENATTI, O.; NOGUEIRA, K.; SANTOS, J. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 2015. **Proceedings...** [S.l.: s.n.], 2015. p. 44–51.

PHILIPS DR Digital Diagnost. Accessed: 2016-06-27, `http://philips.to/2lN1hxn`.

POLIKAR, R. Ensemble based systems in decision making. **IEEE Circuits and systems magazine**, [S.l.], v. 6, n. 3, p. 21–45, 2006.

POULTNEY, C.; CHOPRA, S.; CUN, Y. L. et al. Efficient learning of sparse representations with an energy-based model. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2006. **Anais...** [S.l.: s.n.], 2006. p. 1137–1144.

RAZAVIAN, A.; AZIZPOUR, H.; SULLIVAN, J.; CARLSSON, S. CNN features off-the-shelf: an astounding baseline for recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 2014. **Proceedings...** [S.l.: s.n.], 2014. p. 806–813.

RIBEIRO, E.; UHL, A.; WIMMER, G.; HÄFNER, M. Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification. **Computational and Mathematical Methods in Medicine**, [S.l.], v. 2016, 2016.

ROKACH, L. Ensemble-based classifiers. **Artificial Intelligence Review**, [S.l.], v. 33, n. 1-2, p. 1–39, 2010.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, [S.l.], v. 65, n. 6, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning internal representations by error propagation**. [S.l.]: MIT Press, 1986.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, [S.l.], v. 115, n. 3, p. 211–252, 2015.

SANTANA, L. E. A. d. S. **Otimização em comitês de classificadores**: uma abordagem baseada em filtro para seleção de subconjuntos de atributos. 2012. Tese (Doutorado em Ciência da Computação) — Universidade Federal do Rio Grande do Norte, 2012.

SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R.; LECUN, Y. Overfeat: integrated recognition, localization and detection using convolutional networks. **arXiv preprint arXiv:1312.6229**, [S.l.], 2013.

SHEN, R.; CHENG, I.; BASU, A. A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. **Biomedical Engineering, IEEE Transactions on**, [S.l.], v. 57, n. 11, p. 2646–2656, 2010.

SHIN, H.-C.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning. **IEEE transactions on medical imaging**, [S.l.], v. 35, n. 5, p. 1285–1298, 2016.

SHIRAISHI, J.; KATSURAGAWA, S.; IKEZOE, J.; MATSUMOTO, T.; KOBAYASHI, T.; KOMATSU, K.-i.; MATSUI, M.; FUJITA, H.; KODERA, Y.; DOI, K. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. **American Journal of Roentgenology**, [S.l.], v. 174, n. 1, p. 71–74, 2000.

SILVER, D.; HUANG, A.; MADDISON, C. J.; GUEZ, A.; SIFRE, L.; VAN DEN DRIESSCHE, G.; SCHRITTWIESER, J.; ANTONOGLOU, I.; PANNEERSHELVAM, V.; LANCTOT, M. et al. Mastering the game of Go with deep neural networks and tree search. **Nature**, [S.l.], v. 529, n. 7587, p. 484–489, 2016.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The Journal of Machine Learning Research**, [S.l.], v. 15, n. 1, p. 1929–1958, 2014.

STRIGL, D.; KOFLER, K.; PODLIPNIG, S. Performance and Scalability of GPU-Based Convolutional Neural Networks. In: PDP, 2010. **Anais...** [S.l.: s.n.], 2010. p. 317–324.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015. **Proceedings...** [S.l.: s.n.], 2015. p. 1–9.

T-SNE visualization of CNN codes. Accessed: 2017-01-01, `http://cs.stanford.edu/people/karpathy/cnnembed/`.

TAJBAKHSH, N.; SHIN, J. Y.; GURUDU, S. R.; HURST, R. T.; KENDALL, C. B.; GOTWAY, M. B.; LIANG, J. Convolutional Neural Networks for Medical Image Analysis: full training or fine tuning? **IEEE transactions on medical imaging**, [S.l.], v. 35, n. 5, p. 1299–1312, 2016.

TAN, J. H.; ACHARYA, U. R.; TAN, C.; ABRAHAM, K. T.; LIM, C. M. Computer-assisted diagnosis of tuberculosis: a first order statistical approach to chest radiograph. **Journal of medical systems**, [S.l.], v. 36, n. 5, p. 2751–2759, 2012.

UETZ, R.; BEHNKE, S. Large-scale object recognition with CUDA-accelerated hierarchical neural networks. In: INTELLIGENT COMPUTING AND INTELLIGENT SYSTEMS, 2009. ICIS 2009. IEEE INTERNATIONAL CONFERENCE ON, 2009. **Anais...** [S.l.: s.n.], 2009. v. 1, p. 536–541.

VAN GINNEKEN, B.; KATSURAGAWA, S.; HAAR ROMENY, B. M. ter; VIERGEVER, M. A. et al. Automatic detection of abnormalities in chest radiographs using local texture analysis. **Medical Imaging, IEEE Transactions on**, [S.l.], v. 21, n. 2, p. 139–149, 2002.

VATSAVAI, R. R. Gaussian multiple instance learning approach for mapping the slums of the world using very high resolution imagery. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 19., 2013. **Proceedings...** [S.l.: s.n.], 2013. p. 1419–1426.

VEDALDI, A.; LENC, K. MatConvNet: convolutional neural networks for matlab. In: ANNUAL ACM CONFERENCE ON MULTIMEDIA CONFERENCE, 23., 2015. **Proceedings...** [S.l.: s.n.], 2015. p. 689–692.

WALLACH, I.; DZAMBA, M.; HEIFETS, A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. **arXiv preprint arXiv:1510.02855**, [S.l.], 2015.

WAN, L.; ZEILER, M.; ZHANG, S.; CUN, Y. L.; FERGUS, R. Regularization of neural networks using dropconnect. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML-13), 30., 2013. **Proceedings...** [S.l.: s.n.], 2013. p. 1058–1066.

WANG, T.; WU, D. J.; COATES, A.; NG, A. Y. End-to-end text recognition with convolutional neural networks. In: PATTERN RECOGNITION (ICPR), 2012 21ST INTERNATIONAL CONFERENCE ON, 2012. **Anais...** [S.l.: s.n.], 2012. p. 3304–3308.

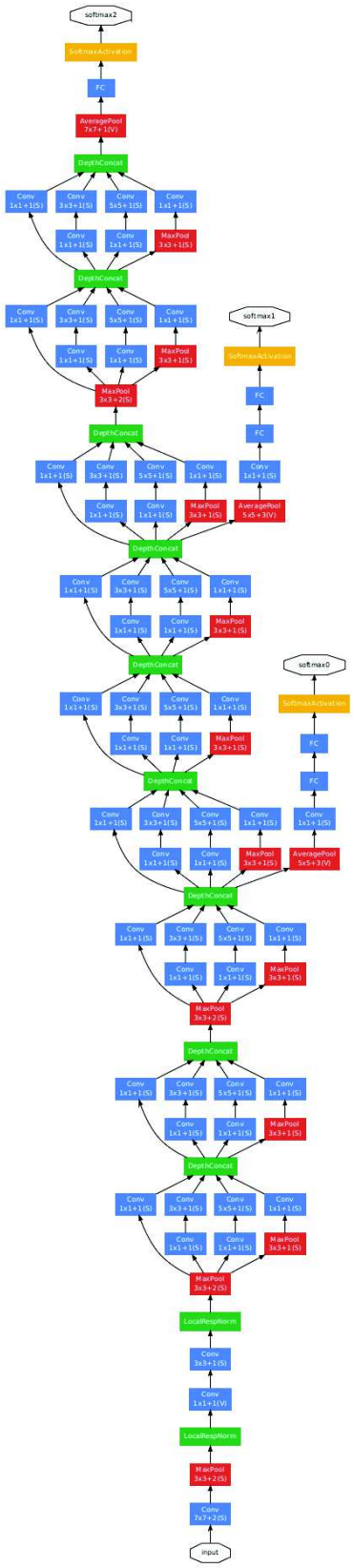
XU, T.; CHENG, I.; MANDAL, M. Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs. In: ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, EMBC, 2011 ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, 2011. **Anais...** [S.l.: s.n.], 2011. p. 5178–5181.

YU, D.; WANG, H.; CHEN, P.; WEI, Z. Mixed pooling for convolutional neural networks. In: INTERNATIONAL CONFERENCE ON ROUGH SETS AND KNOWLEDGE TECHNOLOGY, 2014. **Anais...** [S.l.: s.n.], 2014. p. 364–375.

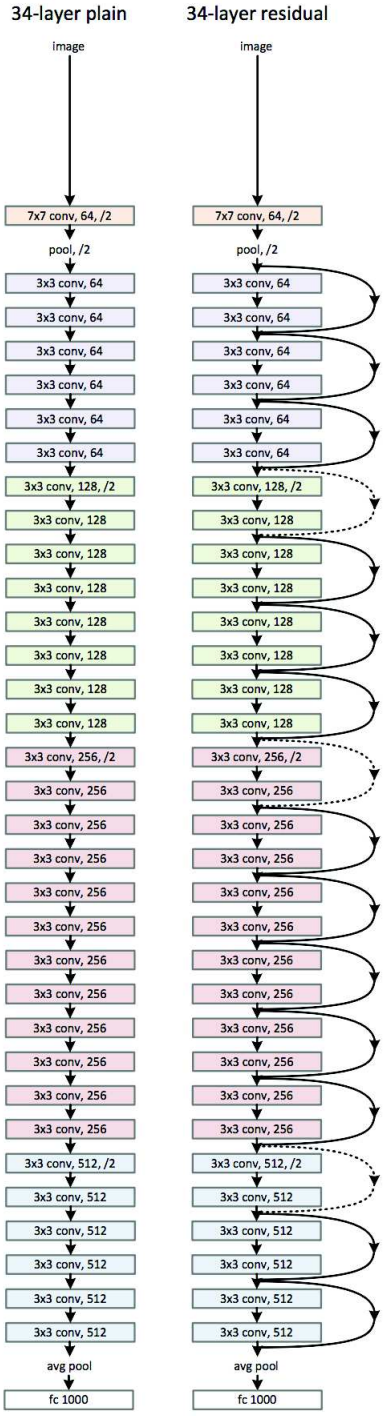
ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2014. **Anais...** [S.l.: s.n.], 2014. p. 818–833.

ZHANG, W.; LI, R.; DENG, H.; WANG, L.; LIN, W.; JI, S.; SHEN, D. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. **NeuroImage**, [S.l.], v. 108, p. 214–224, 2015.

ANEXO A ARQUITETURA DA REDE *GOOGLENET*



ANEXO B ARQUITETURA DA REDE *RESNET*



ANEXO C ARQUITETURA DA REDE *VGGNET*