

UNIVERSIDADE FEDERAL DO PARANÁ

ALINE DARTORA

RECONHECIMENTO AUTOMÁTICO DE PADRÕES RADIOLÓGICOS UTILIZANDO
APRENDIZADO PROFUNDO

CURITIBA

2018

ALINE DARTORA

RECONHECIMENTO AUTOMÁTICO DE PADRÕES RADIOLÓGICOS UTILIZANDO
APRENDIZADO PROFUNDO

Dissertação apresentada ao Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Informática.

Orientador: Prof. Dr. Lucas Ferrari de Oliveira

CURITIBA

2018

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

D226d

Dartora, Aline

Reconhecimento automático de padrões radiológicos utilizando
aprendizado profundo / Aline Dartora. – Curitiba, 2018.

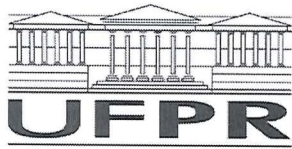
Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas,
Programa de Pós-Graduação em Informática, 2018.

Orientador: Lucas Ferrari de Oliveira .

1. Tomografia computadorizada. 2. Pulmões – Doenças. 3. Tecidos
(Anatomia e fisiologia). 4. Redes neurais (Computação) . I. Universidade
Federal do Paraná. II. Oliveira, Lucas Ferrari de. III. Título.

CDD: 616.075722

Bibliotecário: Elias Barbosa da Silva CRB-9/1894



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **ALINE DARTORA** intitulada: **RECONHECIMENTO AUTOMÁTICO DE PADRÕES RADIOLÓGICOS UTILIZANDO APRENDIZADO PROFUNDO**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 05 de Setembro de 2018.

LUCAS FERRARI DE OLIVEIRA

Presidente da Banca Examinadora (UFPR)

DAVID MENOTTI GOMES

Avaliador Interno (UFPR)

GISELLE LOPES FERRARI RONQUE

Avaliador Externo (UFPR)



RESUMO

Doenças Pulmonares Intersticiais (DPIs) envolvem vários padrões anormais observados em exames de Tomografia Computadorizada de Alta Resolução (TCAR). A caracterização automatizada desses padrões é um componente essencial de um sistema de auxílio ao diagnóstico por computador (CAD) no estudo de DPIs. A escolha e otimização de descritores de características influencia notavelmente na construção e representação de conhecimento pelo computador, a qual, se não for eficiente, compromete a precisão de um sistema CAD. Nesse contexto, técnicas de aprendizagem profunda contornam a necessidade da escolha manual de descritores de características, uma vez que aprendem características diretamente dos dados de treinamento. Neste trabalho, é investigada a eficiência de métodos baseados em aprendizagem profunda para localizar vários padrões de DPIs. A primeira etapa explora e avalia a eficiência da classificação de Regiões de Interesse (ROIs) manualmente delimitadas com Redes Neurais Convolucionais (CNNs) e o uso das técnicas de transferência de aprendizagem e produção de dados sintéticos. A segunda fase propõe a utilização do algoritmo de detecção e classificação de objetos, Faster R-CNN, em conjunto com as técnicas anteriormente citadas, e aplicadas a imagens de inteiras de TCAR, sem a necessidade de segmentação prévia de ROIs. Os resultados obtidos na classificação de blocos de imagens de TCAR atingiram uma precisão de 90,95% com o uso da rede GoogleNet. Na detecção de imagens inteiras de TCAR com a Faster R-CNN o melhor resultado apresentou uma precisão média de detecção (mAP) de 53.9%, também utilizando a arquitetura da rede GoogleNet, a qual, detectou a classe enfisema com um AP de 72.3% e a classe fibrose com AP de 74.5%. Em diferentes configurações, a Faster R-CNN apresentou os melhores valores de AP por classe de 78.4%, 72.3%, 37.6%, 95.4% e 57.4% para os padrões normal, enfisema, vidro-fosco, fibrose e micronódulos, respectivamente. Em um comparativo com a literatura existente, os resultados são próximos aos melhores resultados observados no estado da arte que utilizam a mesma base de dados.

Palavras-chave: Tomografia Computadorizada de Alta Resolução (TCAR). Doenças Pulmonares Intersticiais (DPIs). Detecção e classificação de tecido. Redes Neurais Convolucionais (CNNs). Aprendizado profundo.

ABSTRACT

Interstitial Lung Diseases (ILD) present a variety of abnormal imaging patterns that can be observed in High Resolution Computed Tomography (HRCT). The automated recognition of such imaging patterns is an essential component of Computer Aided Diagnosis (CAD) systems in ILD studies. The selection and optimization of features extractors noticeably influences the construction and representation of knowledge by the computer, which if not done adequately may compromise the CAD systems accuracy. In such a context, deep learning techniques outline the need of a manual selection of features extractors, since they allow the system to recognize and learn features directly from training sample data. This paper researches the effectiveness of methods based on deep learning for identifying several ILD patterns. The effectiveness of identification of Regions of Interest (ROIs) manually delimited with Convolutional Neural Networks (CNNs) and the use of transfer learning and data augmentation techniques are explored and evaluated in the first part. Following, the use of previously mentioned techniques within Faster R-CNN framework to detect and classify ILDs in full HRCT images without performing previous ROI field segmentation is proposed. In this framework the classification results for the HRCT patch-based image representation obtained a recognition rate of 90.95% with the GoogleNet network. When using Faster R-CNN to detect full HRCT images the best result achieved a mean average precision (mAP) of 53.9% on GoogleNet network. The model achieved mAP values of 72.3% and 74.5% on the classes emphysema and fibrosis, respectively. In different scenarios the proposed algorithm performed highest AP values per class of 78.4%, 72.3%, 37.6%, 95.4% and 57.4% for normal, emphysema, ground-glass, fibrosis and micronodules, respectively. When compared with existing researches using the same database, the obtained results are similar to the most effective techniques that can be observed in the state-of-the-art systems.

Keywords: High-Resolution Computed Tomography (HRCT). Interstitial Lung Diseases (ILDs). Tissue detection and classification. Convolutional Neural networks (CNNs). Deep Learning.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – IMAGEM DE TCAR. FONTE: BAGESTEIRO et al. (2015).	15
FIGURA 2 – PADRÕES RADIOLÓGICOS MAIS COMUNS: (A) NORMAL; (B) ENFI- SEMA; (C) VIDRO-FOSCO; (D) FIBROSE; (E) MICRONÓDULOS. FONTE: DEPEURSINGE et al. (2010).	15
FIGURA 3 – MODELO DOS NEURÔNIOS UTILIZADOS NA RNA. FONTE: HAYKIN (2001).	19
FIGURA 4 – REDE NEURAL ARTIFICIAL. FONTE: HAYKIN (2001).	20
FIGURA 5 – ESTÁGIOS DE UMA REDE CONVOLUCIONAL SIMPLIFICADA. FONTE: PEEMEN et al. (2011).	22
FIGURA 6 – DENTRO DE UMA REDE CONVOLUCIONAL. FONTE: LECUN et al. (2015).	23
FIGURA 7 – CONVOLUÇÃO EM UMA IMAGEM DE ENTRADA COM UM FILTRO 5×5. FONTE: LI (2015).	23
FIGURA 8 – FILTROS APRENDIDOS NO TREINAMENTO COM A CNN. FONTE: ANTHI- MOPOULOS et al. (2016).	24
FIGURA 9 – OPERAÇÃO <i>MAX POOLING</i> . FONTE: LI (2015).	25
FIGURA 10 – ARQUITETURA IMAGENET. FONTE: KRIZHEVSKY et al. (2012).	26
FIGURA 11 – ARQUITETURA ALEXNET. FONTE: LU et al. (2016).	27
FIGURA 12 – ARQUITETURA GOOGLNET. FONTE: LU et al. (2016).	27
FIGURA 13 – EXEMPLO DA PRODUÇÃO DE DADOS SINTÉTICOS: (A) AMOSTRA ORIGINAL; (B) AMOSTRAS PRODUZIDAS ARTIFICIAL- MENTE. FONTE: LAURENT (2015).	28
FIGURA 14 – (A) CLASSIFICAÇÃO DE OBJETOS DA IMAGEM (B) CLASSIFICAÇÃO E DETECÇÃO DE OBJETOS NA IMAGEM. FONTE: GIRSHICK (2015).	29
FIGURA 15 – REDE NEURAL CONVOLUCIONAL BASEADA EM REGIÕES (R-CNN). FONTE: GIRSHICK (2015).	29
FIGURA 16 – ARQUITETURA FAST R-CNN. FONTE: GIRSHICK (2015).	30
FIGURA 17 – FASTER R-CNN. À DIREITA: REDE DE PROPOSTA DE REGIÕES. FONTE: REN (2016).	31
FIGURA 18 – DIFERENTES LIMIARES PARA IoU. FONTE: O autor (2018).	32
FIGURA 19 – ARQUITETURA PROPOSTA POR LI ET AL. (2013) PARA EXTRAÇÃO DE CARACTERÍSTICAS. FONTE: LI et al. (2013).	40
FIGURA 20 – ARQUITETURA PROPOSTA POR LI ET AL. (2014). PARA EXTRAÇÃO E CLASSIFICAÇÃO. FONTE: LI et al. (2014).	40
FIGURA 21 – ARQUITETURA PROPOSTA POR ANTHIMOPOULOS ET AL. (2016) PARA EXTRAÇÃO E CLASSIFICAÇÃO. FONTE: ANTHIMOPOULOS et al. (2016).	41
FIGURA 22 – RESUMO COMPARATIVO ENTRE AS TÉCNICAS PARA CLASSIFICAÇÃO DE PADRÕES RADIOLÓGICOS.	45

FIGURA 23 – VISÃO GERAL DA CLASSIFICAÇÃO COM A TRANSFERÊNCIA DE APRENDIZAGEM. FONTE: O autor (2018).	47
FIGURA 24 – VISÃO GERAL DA CLASSIFICAÇÃO COM A Faster R-CNN. FONTE: O autor (2018).	48
FIGURA 25 – AMOSTRA DAS PADRÕES NAS IMAGENS DA BASE DEPEURSINGE ET AL. (2010): (A) NORMAL; (B) ENFISEMA; (C) VIDRO-FOSCO; (D) FIBROSE; (E) MICRONÓDULOS. FONTE: DEPEURSINGE et al. (2010). . .	48
FIGURA 26 – EXTRAÇÃO DE BLOCOS DE ROIs FEITA PELO AUTOR BAGESTEIRO ET AL. (2015). FONTE: BAGESTEIRO et al. (2015).	49
FIGURA 27 – EXEMPLO DA CRIAÇÃO DE DADOS SINTÉTICOS NA BASE BAGESTEIRO ET AL. (2015): (A) AMOSTRA ORIGINAL; (B) AMOSTRAS SINTÉTICAS. FONTE: O autor (2018).	50
FIGURA 28 – QUANTIDADE DE BLOCOS APLICANDO A PRODUÇÃO DE DADOS SINTÉTICOS. FONTE: O autor (2018).	51
FIGURA 29 – COMPARATIVO DAS CLASSIFICAÇÕES PROPOSTAS. FONTE: O autor (2018).	54
FIGURA 30 – (a) CONVERSÃO SIMPLES. (b) CONVERSÃO NA ESCALA DE HU. FONTE: O autor (2018)	56
FIGURA 31 – ESCALAS DE ATENUAÇÃO DE TC. FONTE: O autor (2018).	57
FIGURA 32 – UTILIZAÇÃO DO ALGORITMO <i>LABEL IMAGE REGIONS</i> PARA REGIÃO DO PULMÃO (A) TCAR ORIGINAL (B) MÁSCARA: ENTRADA DO ALGORITMO; (C) LOCALIZAÇÃO DAS REGIÕES FECHADAS: SAÍDA DO ALGORITMO. FONTE: O autor (2018).	59
FIGURA 33 – UTILIZAÇÃO DO ALGORITMO <i>LABEL IMAGE REGIONS</i> PARA REGIÃO DE DPis (A) TCAR ORIGINAL (B) MÁSCARA: ENTRADA DO ALGORITMO; (C) LOCALIZAÇÃO DAS REGIÕES FECHADAS: SAÍDA DO ALGORITMO. FONTE: O autor (2018).	59
FIGURA 34 – (A) TCAR ORIGINAL (B) MÁSCARA DA ROI (C) LOCALIZAÇÃO DE BLOCOS DE ROIS DE DPis. FONTE: BAGESTEIRO et al. (2015).	60
FIGURA 35 – FASTER R-CNN ADAPTADA PARA A ARQUITETURA DA GOOGLNET. FONTE: O autor (2018).	60
FIGURA 36 – EXEMPLO DE DETECÇÃO E CLASSIFICAÇÃO DO PULMÃO COM A FASTER R-CNN. FONTE: O autor (2018).	61
FIGURA 37 – MELHOR CURVA DE PRECISÃO-RECALL UTILIZANDO A FASTER R-CNN. FONTE: O autor (2018).	65
FIGURA 38 – EXEMPLOS DE DETECÇÃO COM A FASTER R-CNN: (A) NORMAL; (B) ENFISEMA; (C) VIDRO-FOSCO; (D) FIBROSE; (E) MICRONÓDULOS. FONTE: O autor (2018).	66

LISTA DE TABELAS

TABELA 1 – VALORES DAS UNIDADES DE HOUNSFIELD. FONTE: DEPEURSINGE et al. (2011).	16
TABELA 2 – MATRIZ DE CONFUSÃO GENÉRICA DE DUAS CLASSES. FONTE: VISA B. HAMSAY (2011).	33
TABELA 3 – RESULTADOS DO TRABALHO AGARWALA ET AL. (2018) PARA DETECÇÃO DE PADRÕES PULMONARES. FONTE: AGARWALA et al. (2018). . .	46
TABELA 4 – IMAGENS COM MAIS DE UMA CLASSE. FONTE: O autor (2018).	46
TABELA 5 – BLOCOS DE IMAGENS. FONTE: O autor (2018).	50
TABELA 6 – DIVISÃO DOS GRUPOS PARA CNN NA BASE DESBALANCEADA. FONTE: O autor (2018).	51
TABELA 7 – DIVISÃO DOS GRUPOS PARA CNN NA BASE BALANCEADA. FONTE: O autor (2018).	51
TABELA 8 – DIVISÃO DOS PACIENTES POR GRUPO. FONTE: O autor (2018).	51
TABELA 9 – ARQUITETURA PRÓPRIA ADAPTADA DA ALEXNET. FONTE: O autor (2018).	52
TABELA 10 – REDE COM OS PESOS ZERADOS E BASE DESBALANCEADA. FONTE: O autor (2018).	53
TABELA 11 – REDE COM OS PESOS ZERADOS E BASE BALANCEADA. FONTE: O autor (2018).	54
TABELA 12 – REDE COM OS PESOS APRENDIDOS E BASE DESBALANCEADA. FONTE: O autor (2018).	54
TABELA 13 – REDE COM OS PESOS APRENDIDOS E BASE BALANCEADA. FONTE: O autor (2018).	54
TABELA 14 – MATRIZ DE CONFUSÃO DA MELHOR CONFIGURAÇÃO PARA CLASSIFICAÇÃO POR BLOCOS. FONTE: O autor (2018).	55
TABELA 15 – QUANTIDADE DE IMAGENS POR CLASSE. FONTE: O autor (2018).	57
TABELA 16 – PACIENTES COM MAIS DE UMA CLASSE. FONTE: O autor (2018).	58
TABELA 17 – DETECÇÃO E CLASSIFICAÇÃO DO PULMÃO COM A FASTER R-CNN NA CONFIGURAÇÃO ORIGINAL. FONTE: O autor (2018).	61
TABELA 18 – DIVISÃO DOS GRUPOS PARA FASTER R-CNN. FONTE: O autor (2018). . .	62
TABELA 19 – CONFIGURAÇÕES TESTADAS NA FASTER R-CNN PARA ANOTAÇÃO POR ROI INTEIRA. FONTE: O autor (2018).	63
TABELA 20 – CONFIGURAÇÕES TESTADAS NA FASTER R-CNN PARA ANOTAÇÃO POR BLOCOS DE ROIS. FONTE: O autor (2018).	64

LISTA DE ABREVIATURAS E SIGLAS

CAD	Diagnóstico Auxiliado por Computador
CLBP	<i>Complete Local Binary Pattern</i>
CNN	Rede Neural Convolutacional
CPUs	Unidades de Processamento Central
DCT	Transformada do Cosseno Discreto
DoG	Diferença de Gaussianos
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DLD	Doenças Pulmonares Difusas
DNN	Rede Neural Profunda (<i>Deep Neural Network</i>)
DPD	Doenças Pulmonares Difusas
DPI	Doenças Pulmonares Intersticiais (<i>Interstitial Lung Disease</i>)
DPLD	Doenças Pulmonares do Parênquima Difuso
E	Enfisema
F	Fibrose
FC	<i>Fuzzy-Connectedness</i>
GG	Vidro-Fosco
GPUs	Unidades de Processamento Gráfico
HU	Unidade Hounsfield (<i>Hounsfield Unit</i>)
IoU	Intersecção sobre a União
k-NN	K-Vizinhos mais próximos (<i>K-Nearest Neighbors</i>)
LBP	Padrões Binários Locais
LMLE	<i>Large Margin Local Estimate</i>
LN	Linfonodos
LP	Parênquima Pulmonar

M	Micronódulos
MCHOG	<i>Multi-coordinate histogram of oriented gradients</i>
MLP	Perceptron Multi-Camada (<i>Multi-Layer Perceptron</i>)
N	Normal
NMS	Supressão não máxima (<i>Non-Max Suppression</i>)
NP	Não-pulmão
OAA	<i>One-Against-All</i>
OMP	<i>Orthogonal Matching Pursuit</i>
PASA	<i>Patch-Adaptive Aparse Approximation</i>
PI	Pneumonia Intersticial
RBM	Máquina de Boltzmann Restrita
RF	Floresta Rândomica
RGB	Vermelho, Verde e Azul
RGLBP	Gabor-LBP Invariante a Rotação
RNA	Rede Neural Artificial
ROIs	Regiões de Interesse
RPN	Rede de Proposta de Regiões (<i>Region Proposal Network</i>)
R-CNN	<i>Region-based Convolutional Network</i>
SFS	<i>Sequential Forward Selection</i>
SVD	<i>Singular Value Decomposition</i>
SVM	<i>Support Vector Machine</i>
TC	Tomografia Computadorizada
TCAR	Tomografia Computadorizada de Alta Resolução
TIG	Textura-intensidade-gradiente
VOI	<i>Volume-of-interest</i>

SUMÁRIO

1 INTRODUÇÃO	13
1.1 MOTIVAÇÃO	14
1.2 DESCRIÇÃO DO PROBLEMA	14
1.2.1 Doenças Pulmonares Intersticiais	14
1.2.2 Sistemas CAD	16
1.3 PROPOSTA	17
1.4 OBJETIVO GERAL	17
1.5 OBJETIVOS ESPECÍFICOS	17
1.6 DESAFIOS	18
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 REDES NEURAIS CONVOLUCIONAIS	19
2.2 APRENDIZADO PROFUNDO	20
2.3 PROCESSO DE APRENDIZAGEM	21
2.3.1 Aprendizado Supervisionado	21
2.4 REDES NEURAIS CONVOLUCIONAIS	21
2.4.1 Convoluções	23
2.4.2 Agregação	25
2.5 TRANSFERÊNCIA DE APRENDIZAGEM	26
2.5.1 ImageNet	26
2.5.2 AlexNet	27
2.5.3 GoogleNet	27
2.6 PRODUÇÃO DE DADOS SINTÉTICOS	27
2.7 DETECÇÃO E CLASSIFICAÇÃO COM CNN	28
2.7.1 R-CNN	28
2.7.2 Fast R-CNN	29
2.7.3 Faster R-CNN	30
2.7.3.1 ZF	32
2.7.3.2 VGG16	32
2.8 MÉTRICAS	32
3 REVISÃO BIBLIOGRÁFICA	35
3.1 ESTADO DA ARTE	35
3.1.1 Métodos de Extração e Classificação	35
3.1.1.1 Métodos de Aprendizado Automático	39
3.1.2 Detecção e Classificação de Objetos	42
3.2 RESUMO DO ESTADO DA ARTE	43
4 METODOLOGIA E RESULTADOS	47

4.1 VISÃO GERAL	47
4.2 BASES DE IMAGENS	48
4.3 CLASSIFICAÇÃO COM REDES NEURAIIS CONVOLUCIONAIS	49
4.3.1 Pré-Processamento	49
4.3.2 Produção de Dados Sintéticos	50
4.3.3 Criação de uma Arquitetura Própria de CNN	51
4.3.4 Transferência de Aprendizagem	52
4.3.4.1 Arquiteturas de Redes Neurais	52
4.4 DETECÇÃO E CLASSIFICAÇÃO COM A FASTER R-CNN	55
4.4.1 Pré-processamento de Imagens	56
4.4.2 Inicialização da Faster R-CNN	58
4.4.3 Arquiteturas de Redes Neurais	60
4.4.4 Treinamento e Teste	60
4.4.5 Informações de Processamento	65
5 DISCUSSÃO	67
5.1 CONSIDERAÇÕES FINAIS	68
REFERÊNCIAS	70

1 INTRODUÇÃO

O câncer de pulmão é um dos tipos de câncer mais agressivos no mundo, e a razão entre mortalidade e incidência é de aproximadamente 90%. No Brasil, para o ano de 2016, foram estimados 17.330 novos casos de câncer de traqueia, brônquios e pulmões entre homens e 10.890 entre mulheres. Com essa hipótese calculou-se um risco estimado de 17,49 novos casos a cada 100 mil homens e 10,54 para cada 100 mil mulheres (SILVA, 2016). Além do Brasil, o câncer de pulmão é a principal causa de morte em vários países tornando esse tumor maligno um problema global com projeção de aumento significativo para os próximos anos (MALTA et al., 2016).

Nesse contexto, vários estudos evidenciam a possível relação entre Doenças Pulmonares Intersticiais (DPIs) e a incidência de câncer de pulmão. As DPIs envolvem vários padrões anormais observados em imagens de Tomografia Computadorizada de Alta Resolução (TCAR). Para a tomada de decisão clínica, o diagnóstico é baseado na reflexão sobre as evidências que são fortemente correlacionadas com a classificação precisa desses padrões radiológicos. Porém, a avaliação radiológica sofre restrições como a aparência similar dos padrões anormais, a experiência do especialista, a dificuldade de identificação a olho nu de alguns padrões, entre outros fatores. Por esses motivos, é importante o desenvolvimento automatizado de sistemas de detecção assistida por computador para complementar o diagnóstico de doenças pulmonares (ARCHONTOGEOGRIS et al., 2012).

Na análise geral dos dados, as redes mais profundas demonstraram ser capazes de representar de maneira eficiente determinadas aplicações (SCHMIDHUBER, 2015). Além disso, houve um aumento considerável no número de trabalhos que utilizam modelos de aprendizado de máquina como as Redes Neurais Convolucionais (CNN) (LECUN et al., 2015). Tal avanço no reconhecimento de imagem, é devido principalmente a disponibilidade de conjuntos de dados anotados em grande escala (ImageNet (KRIZHEVSKY et al., 2012)) que possuem características representativas de distribuição dos dados possibilitando o aprendizado de modelos mais precisos ou mais generalizáveis (LU et al., 2016).

Recentemente, a CNN também foi incorporada na análise de imagens médicas, como exemplo, pode-se citar o trabalho de Lu et al. publicado em 2016, que analisa a utilização da aprendizagem de uma rede pré-formada para a identificação de padrões de DPIs, o qual atingiu 91,1% de acurácia na classificação por blocos de imagens de TCAR. Ainda em imagens médicas, no trabalho de (GAO et al., 2015) foi obtida uma taxa de acerto de 89,6% para a classificação holística de seis padrões por meio de CNNs profundas. A partir de tais observações, o presente trabalho propõe explorar os métodos de classificação para a predição dos padrões radiológicos utilizando um modelo de arquitetura profunda de rede neural convolucional projetada para reconhecer padrões diretamente a

partir dos *pixels* das imagens, incorporando as tarefas de aprendizado de representação do conhecimento e classificação.

1.1 MOTIVAÇÃO

Com base no World Cancer Report, é inquestionável que o câncer é um problema global (MALTA et al., 2016). Atualmente, o câncer é um tema em destaque devido à sua grande responsabilidade por ser a principal causa de morte nos países desenvolvidos e em desenvolvimento. Vários estudos e as estimativas da Organização Mundial da Saúde apontam que o número de óbitos devido ao câncer aumentou 1,6 milhões no período de 2000 a 2007. De acordo com (MALTA et al., 2016) e (BARBOSA et al., 2015) as projeções para o ano de 2030 revelam a ocorrência de 26 milhões de novos casos e 17 milhões de mortes por câncer e mostram que para os próximos 20 a 40 anos o número atual de casos e óbitos duplicará (SILVA, 2016; MALTA et al., 2016).

No Brasil, depois das doenças cardíacas e cerebrovasculares, o câncer é a segunda causa mais comum de morte (BARBOSA et al., 2015). Dentre os tipos mais incidentes de câncer destacam-se pulmão, mama, próstata e colorretal (MALTA et al., 2016; BARBOSA et al., 2015). O câncer de pulmão é uma das doenças mais agressivas em humanos, com uma sobrevida global na faixa de 5 a 10 anos após o início do desenvolvimento da doença (MALTA et al., 2016). Estudos mostram também que a sobrevida dos pacientes com câncer de pulmão é baixa na maioria das populações do mundo, cerca de 10% a 15%, isso deve-se, pelo menos em parte, por esse tipo de câncer ser detectado em estágios avançados, visto que não são observados sintomas em seus estágios iniciais (SILVA, 2016).

1.2 DESCRIÇÃO DO PROBLEMA

1.2.1 Doenças Pulmonares Intersticiais

As doenças pulmonares infiltrativas parenquimatosas agudas e crônicas que possuem graus variáveis de inflamação e fibrose tecidual são coletivamente chamadas de doenças pulmonares intersticiais e estão possivelmente correlacionadas com a incidência de câncer de pulmão. Em geral, as DPIs evoluem ao longo de meses a anos e incluem distúrbios de causa conhecida e desconhecida (WELLS; HIRANI, 2008). Essas doenças podem ser caracterizadas pela alteração gradual do parênquima pulmonar. No entanto, para o reconhecimento de DPIs, existem diversas formas de diagnóstico que variam de acordo com o paciente, seu quadro clínico e de acordo com a aparência dos padrões anormais de DPIs observados nos exames médicos (BAGCI et al., 2011).

Na caracterização de Doenças Pulmonares Intersticiais, a tomografia computadorizada de alta resolução, Figura 1, ajuda a identificar de maneira não invasiva padrões de imagem específicos (Figura 2), que podem fornecer um diagnóstico ou apoiar fortemente a

decisão do especialista na diferenciação de DPIs. A TCAR do pulmão permite a visualização detalhada do parênquima pulmonar possibilitando uma avaliação da aparência visual do tecido pulmonar saudável e do patológico que auxilia na identificação da natureza das lesões do parênquima e fornece indicativos de inflamação (DEPEURSINGE et al., 2011).



FIGURA 1 – IMAGEM DE TCAR. FONTE: BAGESTEIRO et al. (2015).

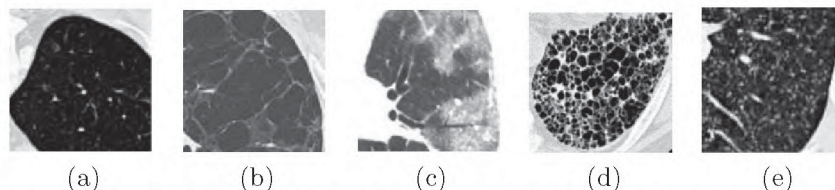


FIGURA 2 – PADRÕES RADIOLÓGICOS MAIS COMUNS: (A) NORMAL; (B) ENFISEMA; (C) VIDRO-FOSCO; (D) FIBROSE; (E) MICRONÓDULOS. FONTE: DEPEURSINGE et al. (2010).

A TCAR produz para cada imagem uma espessura da fatia entre 1-2mm, com tamanho total de 512×512 pixels sendo o pixel variado de 0.4 - 1mm (DEPEURSINGE et al., 2011). O valor numérico de cada pixel está relacionado com a atenuação de raios X, e é expresso em Unidades Hounsfield (HU) (DEPEURSINGE et al., 2011). A tabela 1 apresenta os valores de unidades de Hounsfield para as diferentes estruturas orgânicas do corpo humano. É interessante notar que os valores gerais estão na faixa de $[-1000, 1000]$ e os valores de tecido pulmonar normal estão entre $[-400, -600]$ diferenciando-se dos outros materiais orgânicos.

Com o uso da TCAR na última década, houve um avanço muito grande na habilidade da análise clínica, diminuindo na maioria dos casos o diagnóstico diferencial, aquele baseado apenas nos sinais e sintomas, para um diagnóstico mais confiável. Também, com o uso generalizado da TCAR para avaliar pacientes com DPIs, houve uma grande redução na necessidade de procedimentos diagnósticos invasivos (WELLS; HIRANI, 2008).

Entretanto, assim como em todos os outros testes diagnósticos, a análise clínica sofre influência de alguns fatores como, por exemplo, a variação inter-observador que é igualmente problemática entre radiologistas experientes de TCAR, particularmente

TABELA 1 – VALORES DAS UNIDADES DE HOUNSFIELD. FONTE: DEPEURSINGE et al. (2011).

Material Orgânico	Unidades Hounsfield (HU)
Ar	-1000
Pulmão Normal	-400 a -600
Gordura	-60 a -100
Água	0
Partes Moles	40 a 80
Osso	acima de 400

nos casos mais difíceis que levam ao encaminhamento para a biópsia cirúrgica. Ainda assim, mesmo com a biópsia pulmonar, pode-se chegar em outro fator de confusão na avaliação histológica que é o problema do “erro de amostragem”: a possibilidade de uma amostra de biópsia ter sido retirada de uma área não representativa do processo de doença predominante (WELLS; HIRANI, 2008).

Diante dos fatores acima apontados, busca-se uma alternativa para reduzir as variantes que interferem o processo de diagnóstico clínico. Nesse contexto, o diagnóstico auxiliado por computador vem sendo explorando e os sistemas de auxílio ao diagnóstico conduzem a uma nova alternativa para apoiar a decisão clínica.

1.2.2 Sistemas CAD

A abordagem multidisciplinar é considerada o “padrão-ouro” para o diagnóstico de doença pulmonar difusa (WELLS; HIRANI, 2008). Diante disso, recentemente vem sendo explorado o potencial dos sistemas CAD que usam a resposta do computador como referência para o diagnóstico clínico (BAGCI et al., 2011). Estes sistemas têm um grande potencial para melhorar as habilidades de diagnóstico dos médicos ajudando a superar as limitações da memória humana, distração e fadiga, além de melhorar a capacidade de fazer cálculos e interpretar dados, impedindo assim, que decisões sejam afetadas por viés ou pré-conceitos (EADIE et al., 2012).

Os sistemas CAD possuem dois tipos de aplicações: auxílio a detecção de lesões e auxílio ao diagnóstico (AZEVEDO-MARQUES, 2001). Neste trabalho, será abordado apenas um dos tipos de sistema CAD que é o sistema para auxílio a detecção que envolve a quantificação de características de uma imagem e a classificação dessas características.

De forma simplificada, um sistema CAD utiliza o resultado de análises quantitativas automatizadas de imagens radiográficas como uma “segunda opinião” para auxiliar o diagnóstico de um especialista da área médica. É de grande relevância ressaltar que no presente momento o computador apenas fornece informações adicionais para o radiologista, e que o profissional qualificado é que irá fornecer o diagnóstico final (AZEVEDO-MARQUES, 2001).

1.3 PROPOSTA

Uma revisão publicada por (LITJENS et al., 2017), mostrou que as aplicações da aprendizagem profunda na análise de imagens médicas são um tópico dominante nas principais conferências atuais. Além disso, algoritmos que aprendem os dados de forma automática, em particular as redes convolucionais, tornaram-se rapidamente uma metodologia de escolha para a análise de imagens médicas, justificando o interesse por sua aplicação.

Nesse cenário, outra vantagem de escolher uma rede neural como metodologia é que o algoritmo pode ser generalizado para resolver diferentes tipos de problemas utilizando arquiteturas semelhantes. Outro ponto que deve ser considerado é que as CNNs demonstram um desempenho superior na resolução de muitos problemas de classificação de imagens (LITJENS et al., 2017).

1.4 OBJETIVO GERAL

O objetivo geral do presente trabalho é classificar os padrões radiológicos normal, enfisema, vidro-fosco, fibrose e micronódulos observados em imagens de TCAR. Buscando automatizar o processo de diagnóstico de doenças pulmonares, tanto na detecção de regiões de interesse quanto na quantificação e classificação dessas regiões em uma imagem, é explorada a utilização de metodologias de aprendizado profundo.

1.5 OBJETIVOS ESPECÍFICOS

Os objetivos específicos são listados abaixo:

- Analisar o desempenho da técnica de criação de dados sintéticos;
- Observar o desempenho de uma arquitetura de Rede Neural própria criada para o problema de classificação de DPIs.
- Testar diferentes arquiteturas de Redes Neurais Convolucionais existentes;
- Analisar o comportamento da técnica de transferência de aprendizagem em CNNs para classificar blocos de imagens de TCAR;
- Descrever o desempenho da transferência de aprendizagem de outros domínios para o domínio de classificação de imagens médicas;
- Analisar o desempenho da *Faster Region-based Convolutional Network* (Faster R-CNN) aplicada ao contexto de detecção e classificação de imagens médicas;

- Testar a Faster R-CNN para detectar e classificar automaticamente os padrões radiológicos de DPIS em imagens inteiras de TCAR;
- Processar as imagens de TCAR e testar diferentes entradas para a Faster R-CNN;
- Treinar a Faster R-CNN com diferentes configurações de anotações de dados;
- Testar a Faster R-CNN com três variações de arquiteturas de CNNs para a detecção e classificação de DPIS;

1.6 DESAFIOS

De maneira geral, dentre os desafios observados pode-se citar a pequena quantidade de amostras na base de imagens original (DEPEURSINGE et al., 2011), a aparência similar dos diferentes padrões radiológicos e a presença de mais de uma doença em algumas imagens de TCAR, problemas levados em consideração na escolha da metodologia.

Em relação a aplicação da metodologia usando CNN's pode ser destacada a necessidade de uma grande quantidade de dados de treinamento rotulados. Também, deve ser observada a necessidade de grandes recursos computacionais e de memória para treinar uma CNN profunda. Outro ponto a ser considerado, é que no treinamento de uma CNN profunda podem surgir problemas de *overfitting* e convergência, os quais, exigem ajustes repetitivos na arquitetura ou nos parâmetros de aprendizado da rede para garantir que todas as camadas estejam aprendendo com velocidade comparável.

De todas as dificuldades acima mencionadas, a tarefa mais desafiadora é detectar e localizar os diferentes padrões radiológicos com precisão, especialmente para as classes não têm uma aparência bem definida. Também, deve ser observado o fato de que nessa etapa serão trabalhadas com fatias inteiras de TCAR e a base de dados (DEPEURSINGE et al., 2011) contém imagens parcialmente rotuladas. Isso ocorre quando há divergência de opinião na marcações feitas pelos radiologistas, sendo assim, a região questionada da imagem é mantida sem rotulação (DEPEURSINGE et al., 2011). Além de tudo, a própria detecção de objetos já é uma tarefa desafiadora em diferentes contextos, e portanto, é uma das principais áreas de pesquisa no campo de visão computacional (SHAFII et al., 2017).

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são descritos os conceitos usados para o desenvolvimento deste trabalho. Inicialmente, na Seção 2.1, é apresentado o conceito de Redes Neurais Artificiais seguido do conceito de Aprendizado Profundo (Seção 2.2) e de Redes Neurais Convolucionais (Seção 2.3). Na Seção 2.4 é apresentada a técnica de Transferência de Aprendizagem, e na sequência a Geração de Dados Sintéticos (Seção 2.5), Detecção e Classificação com CNNs (Seção 2.6), e por fim na Seção 2.7 são discutidas as métricas para cálculo dos resultados.

2.1 REDES NEURAIAS CONVOLUCIONAIS

As abordagens atuais para o reconhecimento de objetos fazem uso essencial de métodos de aprendizagem de máquinas baseados em Redes Neurais Artificiais (RNA) (KRIZHEVSKY et al., 2012; LECUN et al., 2015; LITJENS et al., 2017). Uma RNA é um algoritmo de inteligência artificial inspirado na estrutura neural de organismos inteligentes que adquirem conhecimento por meio da experiência (SCHMIDHUBER, 2015).

Essa rede consiste em uma sequência de neurônios artificiais conectados e organizados em camadas dentro de uma arquitetura que produzem, cada um, uma sequência de ativações de valor real. Os dados vindos das ativações dos neurônios são repassados, ponderados e transformados a outros neurônios de acordo com a função utilizada pela rede. Este processo se repete até que a RNA promova um sinal de saída e um neurônio de saída é ativado, funcionando como um classificador binário linear (SCHMIDHUBER, 2015).

O neurônio artificial é constituído basicamente pelo seus pesos, *bias* e a função de ativação (ROSENBLATT, 1958). Os pesos são utilizados para amplificar ou atenuar o valor passado de um neurônio para outro e os *bias* são ajustáveis e determinam a direção da classificação. Além disso, uma função de ativação determina se o neurônio está ativo ou não por meio de ligações ponderadas para um dado conjunto de características recebidas como entrada. Um modelo simplificado dos neurônios utilizados em uma Rede Neural Artificial pode ser observado na Figura 3.

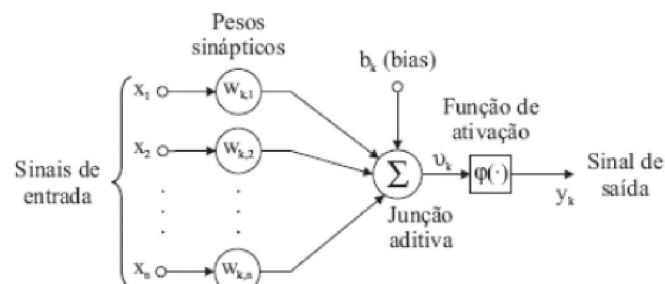


FIGURA 3 – MODELO DOS NEURÔNIOS UTILIZADOS NA RNA. FONTE: HAYKIN (2001).

O treinamento com o *backpropagation* compõe o modelo de Redes Neurais Artificiais mais utilizado atualmente, as redes “Perceptron Multi-Camada” (MLP). Para atingir o aprendizado de uma rede neural é necessário passar por um processo de atribuição de valores, que é uma tentativa de encontrar pesos que exibem o comportamento desejado da rede (SCHMIDHUBER, 2015). Esse treinamento é feito nas camadas intermediárias da rede por meio da atualização dos parâmetros de cada neurônio (pesos, *bias* e ativação) em função do erro calculado na camada de saída.

Basicamente, após calcular o avanço na rede a partir de um padrão apresentado à camada de entrada, a saída obtida é comparada com a saída desejada para um padrão particular. Assim, quando a saída não é correta, o erro é calculado e propagado da camada de saída até a camada de entrada, e os pesos das conexões das camadas internas vão sendo modificados conforme o erro é retropropagado (LECUN et al., 2015; SCHMIDHUBER, 2015; LITJENS et al., 2017). Na Figura 4 é possível visualizar a estrutura de uma RNA.

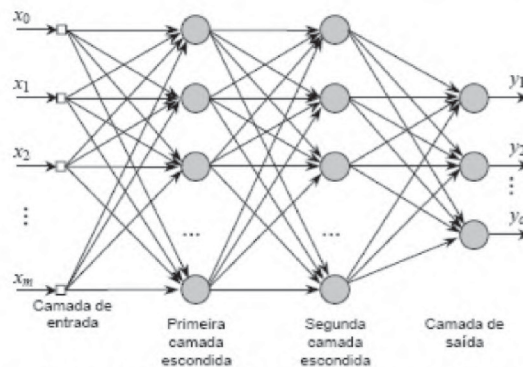


FIGURA 4 – REDE NEURAL ARTIFICIAL. FONTE: HAYKIN (2001).

O meio de encontrar uma atribuição precisa de valores no estágio de treinamento da rede é chamado de Aprendizado Profundo ou *Deep Learning*, conceito o qual é explorado na próxima Seção.

2.2 APRENDIZADO PROFUNDO

Técnicas de aprendizagem profunda vêm sendo aplicadas em diferentes casos, tais como, identificação de objetos em imagens, transcrição de fala em texto, combinação de itens de notícias, postagens ou produtos com os interesses dos usuários e seleção de resultados relevantes de uma pesquisa (LECUN et al., 2015). Como consequência, a aprendizagem profunda está trazendo grandes avanços na resolução de problemas que a inteligência artificial não alcançou por muitos anos.

Diferente dos métodos tradicionais que pré-definiam descritores de características para representar dados de um problema, a aprendizagem profunda busca descobrir uma estrutura intrínseca em grandes conjuntos de dados. Essa técnica utiliza algoritmos de representação da aprendizagem que possuem múltiplos níveis obtidos pela composição

de módulos simples, os quais transformam a representação de um nível (começando com uma entrada bruta) em outro nível mais alto (nível abstrato). Dessa forma, é possível representar os dados em múltiplos níveis de abstração, visto que os modelos são capazes de extrair as características automaticamente, ou seja, tornam-se modelos de descritores treináveis (SCHMIDHUBER, 2015).

2.3 PROCESSO DE APRENDIZAGEM

A forma mais comum de aprendizado de máquina é a aprendizagem supervisionada, seja no aprendizado profundo ou não (LECUN et al., 2015). A partir do final dos anos 90, as técnicas supervisionadas usadas para desenvolver sistemas por meio de dados de treinamento tornaram-se cada vez mais populares na análise de imagens médicas (LITJENS et al., 2017).

2.3.1 Aprendizado Supervisionado

O aprendizado supervisionado é utilizado para seguir um modelo e a partir dele possibilitar a generalização de modo que se consiga aplicar a qualquer dado de entrada. Essa generalização é feita durante o treinamento, onde a rede adapta os seus pesos gerando na saída um estado de ativação compatível com o esperado. Esse método de treinamento da rede, utiliza conhecimentos empíricos comumente representados por um conjunto de exemplos rotulados, ou seja, exemplos de dados de entrada associados com a respectiva saída esperada (LECUN et al., 2015).

Neste trabalho, utilizou-se Redes Neurais Convolucionais (CNNs) sob o ponto de vista de aprendizado supervisionado e profundo para o problema de classificação de padrões radiológicos. Na próxima seção, é aprofundado o conceito de CNNs para contextualização da metodologia proposta.

2.4 REDES NEURAI CONVOLUCIONAIS

Nos últimos anos, houve um aumento no número de trabalhos publicados que incluem em sua metodologia modelos de aprendizado de máquina. Um modelo que vem sendo bastante disseminado são as Redes Neurais Convolucionais (CNNs ou ConvNets) que trabalham com o conceito de auto-aprendizagem criando extratores de características para tratar problemas que possam ser compreendidos pela máquina. As CNNs usam dados brutos para aprender as representações em vez de dependerem de extratores de características manuais (LECUN et al., 1998; LECUN et al., 2015).

As CNNs são projetadas para processar dados que vêm na forma de múltiplas matrizes, e são tipicamente treinadas de ponta a ponta de maneira supervisionada. Na Figura 5, estão representados de forma simplificada os dois grandes estágios de uma rede

convolucional, a extração de características e a classificação. A camada de extração de características aprende características que são genéricas até um determinado ponto e que são úteis para uma tarefa destino. Além disso, as camadas de extração de características são capazes de extrair características invariantes a rotação de formas bidimensionais por meio de uma arquitetura composta de camadas convolucionais que trabalham com filtros, ativações e agregações (*pooling*). Tais conceitos, são abordados de maneira mais aprofundada nas próximas seções.

Em um segundo estágio, as camadas de classificação usam as características locais extraídas e armazenadas em vetores de características para executar a classificação da entrada da rede após a aplicação de um algoritmo de aprendizagem de máquina. O objetivo do procedimento de aprendizado é encontrar conjuntos de matrizes que extraiam boas características discriminativas para serem usadas na classificação de imagens (LITJENS et al., 2017; ROSENBLATT, 1958; LECUN et al., 1998; LECUN et al., 2015). Dessa forma, a CNN é usada para extrair uma nova representação do conjunto de dados de destino, semelhante ao uso de um extrator de características para uma determinada entrada, o qual, obtém uma representação vetorial para cada amostra.

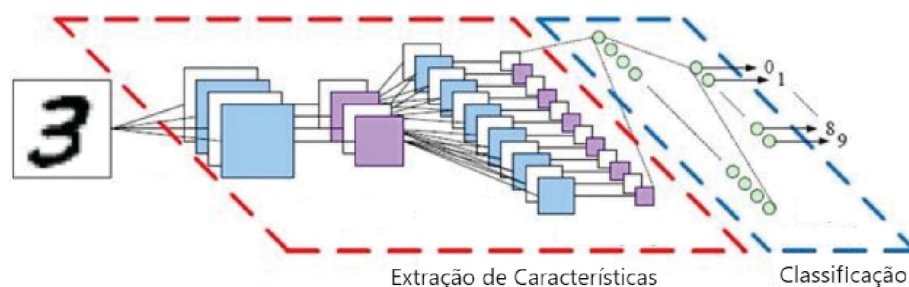


FIGURA 5 – ESTÁGIOS DE UMA REDE CONVOLUCIONAL SIMPLIFICADA. FONTE: PEEMEN et al. (2011).

De maneira simplificada, as quatro ideias-chave por trás de ConvNets são: conexões locais, pesos compartilhados, agregação e o uso de múltiplas camadas. A principal força das CNNs está nas camadas de neurônios convolucionais e na sua partilha de peso, explorando a ideia de que estruturas semelhantes ocorrem em diferentes locais de uma mesma imagem (LECUN et al., 1998; LECUN et al., 2015).

A arquitetura dentro de uma ConvNet típica aplicada a imagem de um cão Samoieda é apresentada na Figura 6. Como visto, a ConvNet é estruturada em uma série de estágios e na ilustração podem ser observados os primeiros estágios compostos pelas camadas convolucionais e camadas de agregação. No canto inferior direito estão as entradas RGB (vermelho, verde e azul) da imagem. Já, as saídas de cada camada convolucional são representadas horizontalmente, onde cada imagem retangular é um mapa de características para uma das características aprendidas em cada uma das posições da

imagem. As informações seguem um fluxo de baixo para cima, onde as características de nível inferior atuam como detectores de borda orientados e na saída é calculada uma pontuação para cada uma das possíveis classes da imagem (LECUN et al., 2015; PEEMEN et al., 2011).

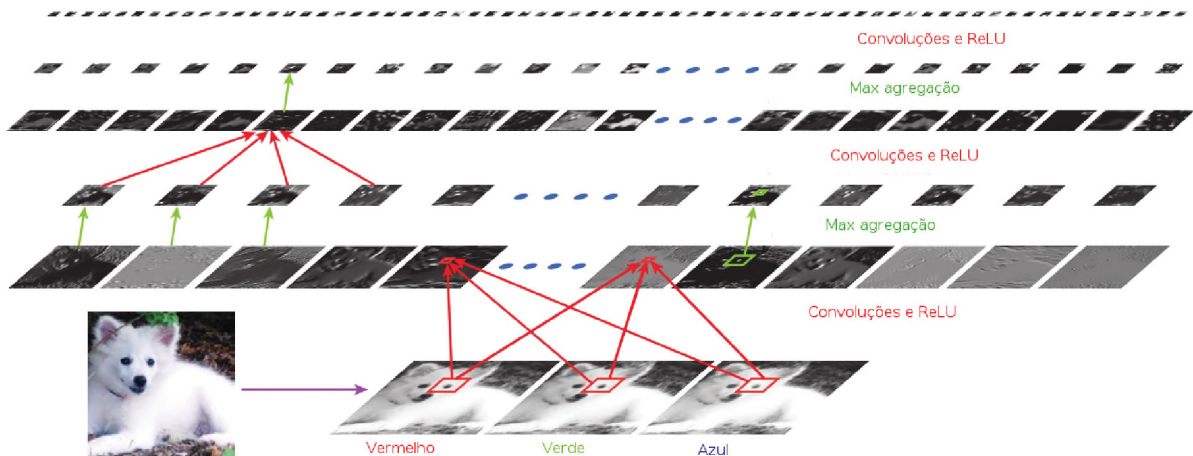


FIGURA 6 – DENTRO DE UMA REDE CONVOLUCIONAL. FONTE: LECUN et al. (2015).

2.4.1 Convoluções

A CNN é composta pelas convoluções e o processo de convolução consiste em um conjunto de N filtros treinados para reconhecer automaticamente as características da camada de entrada. O principal papel da convolução é extrair características preservando a relação espacial entre os *pixels* e aprendendo características da imagem. A convolução envolve a aplicação de um filtro sobre uma imagem, gerando uma nova imagem como resultado. O filtro é uma matriz que “desliza” sobre toda a imagem, modificando-a. Os filtros definem uma pequena área, por exemplo, uma área de 3×3 ou 5×5 *pixels* e a aplicação desses N filtros resulta em N mapas característicos (FIGURA 7), gerando as camadas de convoluções (LECUN et al., 2015).

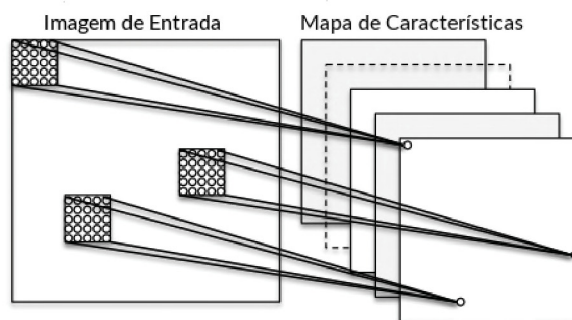


FIGURA 7 – CONVOLUÇÃO EM UMA IMAGEM DE ENTRADA COM UM FILTRO 5×5 . FONTE: LI (2015).

Os pesos das camadas convolucionais são compartilhados entre os neurônios, levando os filtros a aprenderem padrões frequentes que ocorrem em qualquer parte da

imagem. Dessa forma, as convoluções fazem com que a resposta de um neurônio individual aos estímulos dentro de seu campo receptivo possa ser aproximada matematicamente (LECUN et al., 2015). Na Figura 8 é apresentado um exemplo de filtros que foram aprendidos por uma CNN no contexto de imagens médicas.

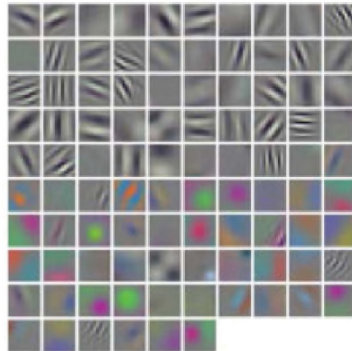


FIGURA 8 – FILTROS APRENDIDOS NO TREINAMENTO COM A CNN. FONTE: ANTHIMOPOULOS et al. (2016).

De forma mais detalhada, (LITJENS et al., 2017) explica que, por exemplo, ao considerar \mathbf{X} como uma imagem vetorizada, os pesos podem ser compartilhados resultando na operação de convolução e conseqüentemente reduzindo a quantidade de parâmetros que a rede deve aprender. Em cada camada, a imagem de entrada é convoluída em um conjunto de k *kernels* (Equação 2.1) e subsequentemente *bias* (Equação 2.2) são adicionados, cada um gerando um novo mapa de características \mathbf{X}_k . O neurônio *bias* aumenta os graus de liberdade permitindo uma melhor adaptação da rede neural ao conhecimento que está sendo aprendido.

$$W = \{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_k\} \quad (2.1)$$

$$B = \{b_1, \dots, b_k\} \quad (2.2)$$

As novas características são submetidas a uma função não-linear de elementos $\sigma(\cdot)$ e o mesmo processo é repetido para cada camada convolucional l (Equação 2.3).

$$\mathbf{X}_k^l = \sigma(\mathbf{W}_k^{l-1} \otimes \mathbf{X}^{l-1} + b_k^{l-1}) \quad (2.3)$$

Essa função não-linear ou função de ativação determina se um neurônio está ativo ou não. A função ReLU é uma das mais utilizadas e é uma operação aplicada a cada *pixel* que substitui todos os valores de *pixels* negativos do mapa de características por zero, determinando um limiar em zero. Essa operação tem por objetivo introduzir a não-linearidade na CNN pois a maioria dos problemas reais tratam-se de dados não-lineares (LITJENS et al., 2017). Apesar do grande uso da função ReLU e em geral do seu

melhor desempenho, outras funções não lineares, como *tanh* ou *sigmoid* também podem ser utilizadas para trabalhar com a não-linearidade (LECUN et al., 2015; LI et al., 2014).

2.4.2 Agregação

A inspiração para as camadas convolucionais e camadas de agregação em ConvNets está diretamente relacionada as noções clássicas de células simples e células complexas da neurociência visual, e a arquitetura geral faz referência à hierarquia do córtex visual (LECUN et al., 2015). As camadas de agregação adicionam robustez ao modelo, gerando um modelo invariante que aprende características da imagem independente da sua localização na imagem.

O processo de agregação reduz a dimensionalidade de uma mapa de características mantendo as informações mais importantes. De forma simplificada, este processo consiste em deslizar uma janela quadrada de determinada dimensão sobre a imagem de forma que não haja sobreposição das regiões. O tamanho da janela, é definido pelo parâmetro passo (*stride*), quando o passo é igual a 1 é equivalente a usar todas as janelas de *pooling* possíveis, e o passo igual ao tamanho da área de *pooling* é equivalente a usar janelas que não se sobrepõem.

Os tipos de agregação podem ser Max, Média, Mediana, entre outros. O *Max Pooling* é o mais comum e consiste em pegar o maior valor do mapa de características dentro da janela definida. A Figura 9 mostra um exemplo de operação *Max Pooling* em um mapa de características.

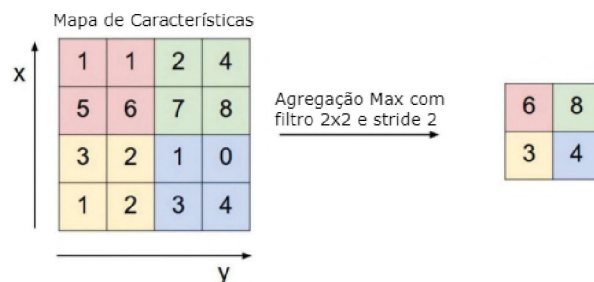


FIGURA 9 – OPERAÇÃO MAX POOLING. FONTE: LI (2015).

2.5 TRANSFERÊNCIA DE APRENDIZAGEM

Uma Rede Neural Convolutiva em geral contém uma grande quantidade de parâmetros ajustáveis podendo atingir a ordem de milhões de parâmetros. Logo, aprender um grande número de parâmetros torna-se uma tarefa difícil quando o conjunto de amostras é razoavelmente pequeno. Como alternativa, a transferência de aprendizagem tenta contornar o requisito de grandes conjuntos de dados para treinamento em uma rede profunda com o uso de redes pré-treinadas (LITJENS et al., 2017).

Os métodos transferência de aprendizagem exploram o conhecimento acumulado a partir de dados do domínio de origem para facilitar a modelagem preditiva de padrões em um novo domínio. Ou seja, as camadas ocultas compartilhadas treinadas por uma tarefa de origem podem ser reutilizadas em uma tarefa de destino diferente. Para isso, apenas a última camada de classificação precisa ser treinada novamente, e, se necessário, pode ser ajustada para o novo modelo. Os parâmetros das camadas ocultas da tarefa de origem atuam como parâmetros de inicialização da tarefa-alvo, tornando essa estratégia promissora para um modelo em que a boa inicialização é muito importante (LU et al., 2015).

Na sequência são descritas brevemente, a arquitetura de algumas CNNs conhecidas na área de visão computacional e que foram utilizadas para avaliar a técnica de transferência de aprendizagem.

2.5.1 ImageNet

A técnica de transferência de aprendizagem juntamente com a ImageNet serve como suporte para melhorar significativamente muitos problemas de detecção de objetos e segmentação de imagens usando outros conjuntos de dados (ANTHIMOPOULOS et al., 2016; LU et al., 2016). A rede ImageNet tem mais de 1,2 milhões de imagens de 256×256 pixels anotadas em 1000 categorias de classes de objetos. A arquitetura da ImageNet é composta por cinco camadas de convolução (*conv*), duas camadas de agregação (*pool*) e duas camadas totalmente conectadas (*fc*) (Figura 10).

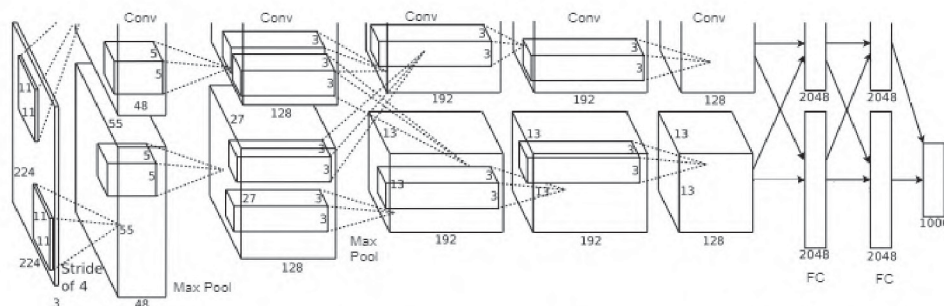


FIGURA 10 – ARQUITETURA IMAGENET. FONTE: KRIZHEVSKY et al. (2012).

2.5.2 AlexNet

A arquitetura da AlexNet tem cinco camadas convolucionais com ativações ReLU, três camadas de agregação e três camadas totalmente conectadas com aproximadamente 60 milhões de parâmetros livres (Figura 11). A arquitetura da AlexNet alcançou um desempenho significativamente melhor aos outros métodos de aprendizado não profundos em um Desafio de Reconhecimento Visual de Grande Escala. Esta rede foi projetada para classificar imagens RGB e com o tamanho de entrada de 224×224 (ANTHIMOPOULOS et al., 2016; LU et al., 2016).

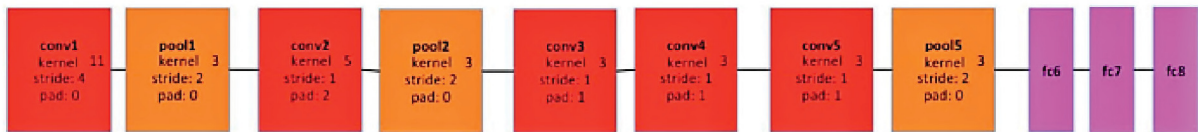


FIGURA 11 – ARQUITETURA ALEXNET. FONTE: LU et al. (2016).

2.5.3 GoogleNet

O modelo GoogleNet é uma das redes mais recentes e complexas proposta na área de Redes Neurais Convolucionais. A arquitetura GoogleNet (Figura 12) tem duas camadas de convolução, duas camadas de agregação e nove camadas do módulo *Inception* que é um novo conceito que concatena filtros de diferentes tamanhos e dimensões em um único novo filtro. As camadas *inceptions*, consistem em seis camadas de convolução com diferentes tamanhos de *Kernel* e uma camada de agregação (LU et al., 2016).

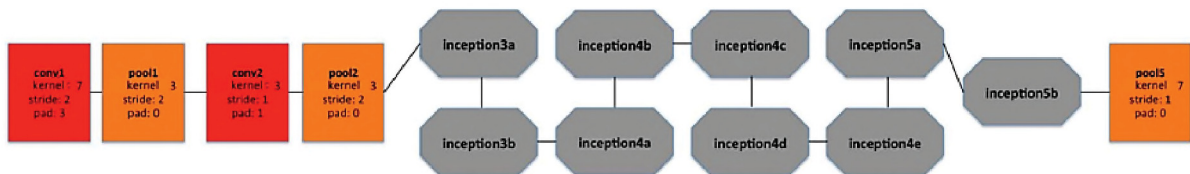


FIGURA 12 – ARQUITETURA GOOGLNET. FONTE: LU et al. (2016).

2.6 PRODUÇÃO DE DADOS SINTÉTICOS

Visto que para usar a CNN é necessário uma grande quantidade de amostras, nesta seção é apresentada uma técnica utilizada para aumentar artificialmente a quantidade de dados usando transformações que preservam o rótulo da imagem. O método de produção de dados sintéticos é considerado uma técnica fácil e comum para reduzir o *overfitting* em dados de imagens. O *overfitting* é o termo usado quando um classificador gera superfícies de decisão mais complexas do que o requisitado para resolver um determinado problema (KRIZHEVSKY et al., 2012).

A produção de dados sintéticos utiliza diferentes filtros sobre os dados originais afim de aumentar a quantidade e proporcionar dados bem variados (FIGURA 13). Alguns

exemplos que podem ser utilizados são filtros de transformações geométricas (dimensionamento, rotação, recorte), variações da imagem (borramento, nitidez) e ajustes de iluminação (KRIZHEVSKY et al., 2012). Com o emprego dessa técnica, contorna-se a necessidade de ter uma numerosa quantidade de dados para o correto treinamento das arquiteturas densas em aprendizado profundo.

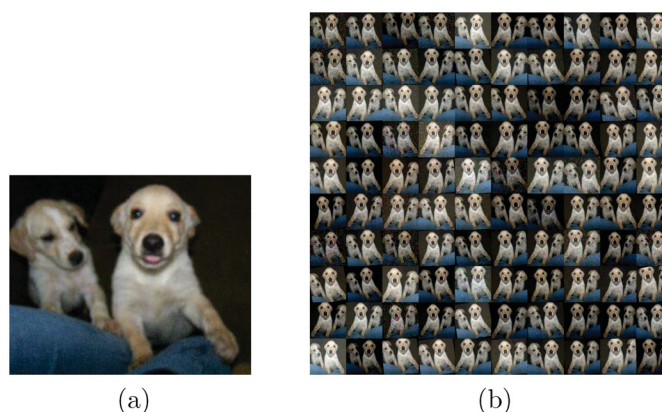


FIGURA 13 – EXEMPLO DA PRODUÇÃO DE DADOS SINTÉTICOS:
 (A) AMOSTRA ORIGINAL; (B) AMOSTRAS PRODUZIDAS ARTIFICIALMENTE.
 FONTE: LAURENT (2015).

2.7 DETECÇÃO E CLASSIFICAÇÃO COM CNN

A tendência atual é criar redes mais profundas e com o maior número de camadas possíveis, visto que atualmente existem mais dados disponíveis para treinamento, dispõem-se de um maior poder computacional e de novos algoritmos desenvolvidos. Por consequência, as ConvNets profundas trazem resultados significativamente melhores para a classificação de imagens e também para a detecção de objetos (GIRSHICK, 2015; TEICHMANN et al., 2016). Nesse cenário, são desenvolvidas novas técnicas estimuladas pelo sucesso dos métodos de detecção de objetos baseados em CNNs, tais como, R-CNN, Fast R-CNN e Faster R-CNN, apresentadas na sequência.

2.7.1 R-CNN

A Rede Neural Convolutacional baseada em Regiões (R-CNN) foi inicialmente desenvolvida utilizando a Convnet profunda para encontrar regiões propostas em uma imagem de entrada fazendo a detecção e segmentação de objetos simultaneamente, conforme pode ser observado na Figura 14.

A R-CNN é um sistema de detecção de objetos visuais que combina propostas de região de baixo para cima com características previamente calculadas pela CNN. A rede neural realiza essa tarefa por meio de regressão e classificação. Na regressão da rede são preditas as coordenadas dos objetos (*bounding boxes*) e na classificação é feita a predição da classe (GIRSHICK, 2015; SHAFII et al., 2017).

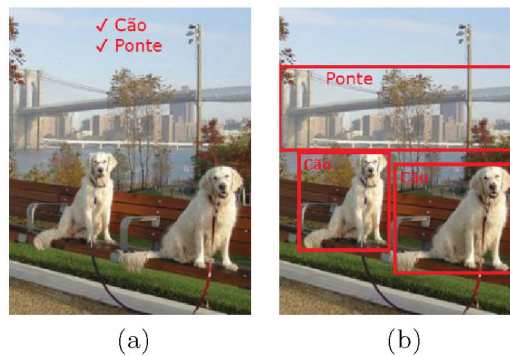


FIGURA 14 – (A) CLASSIFICAÇÃO DE OBJETOS DA IMAGEM (B) CLASSIFICAÇÃO E DETECÇÃO DE OBJETOS NA IMAGEM. FONTE: GIRSHICK (2015).

A arquitetura de uma R-CNN é composta por três módulos, no qual, o primeiro gera regiões propostas independentes da categoria da imagem com a técnica de busca seletiva, o segundo é uma grande CNN que extrai as características dessas regiões independentes, e o terceiro classifica as características extraídas por meio de um classificador SVM. A Figura 15 ilustra a R-CNN, o modelo original do qual a Fast R-CNN (Seção 2.6.2) é inspirada (GIRSHICK, 2015; SHAFII et al., 2017).

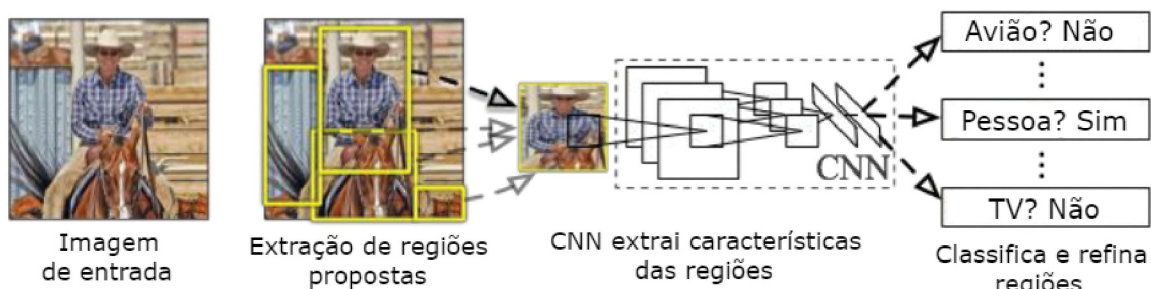


FIGURA 15 – REDE NEURAL CONVOLUCIONAL BASEADA EM REGIÕES (R-CNN). FONTE: GIRSHICK (2015).

2.7.2 Fast R-CNN

Apesar da R-CNN ter um desempenho significativamente bom na detecção de objetos e a classificação de imagens, essa metodologia apresenta algumas desvantagens, como a detecção ser lenta e o custo computacional para treinamento ser alto. Diante disso, em 2015, foi proposta a metodologia Fast R-CNN com o intuito de classificar eficientemente objetos aumentando a velocidade de redes convolucionais profundas (GIRSHICK, 2015).

A Fast R-CNN é um algoritmo de treinamento com uma única fase que aprende a classificar objetos refinando suas localizações espaciais. O treinamento em um único estágio utilizando uma multi-tarefa de perda permite a atualização de todas as camadas da rede durante o treinamento o que não requisita armazenamento em cache das características. A rede recebe como entrada uma imagem inteira e um conjunto de propostas de objetos. A imagem é então processada com *max pooling* e camadas convolucionais e é agrupada em

um mapa de características de tamanho fixo que, em seguida, é mapeado para um vetor de características por meio de camadas totalmente conectadas (FCs).

A rede tem dois vetores de saída por ROIs que são as probabilidades do *softmax* e os deslocamentos de regressão do *bounding-box* feito por classe (Figura 16). O uso da regressão com *bounding-box* reduz erros na detecção de objetos fazendo o ajuste fino das janelas de detecção para corresponder ao tamanho ideal do objeto real de interesse. Esse ajuste traz um ganho significativo na detecção correta de objetos.

A saída da camada FC de classificação é correspondente ao número de classes com adição da classe “*Background*” e a saída da camada FC de regressão possui para cada classe as 4 coordenadas dos *bounding boxes*. A classe “*Background*” é utilizada para representar as regiões das imagens que não contém nenhuma das classes de objetos do conjunto de dados original.

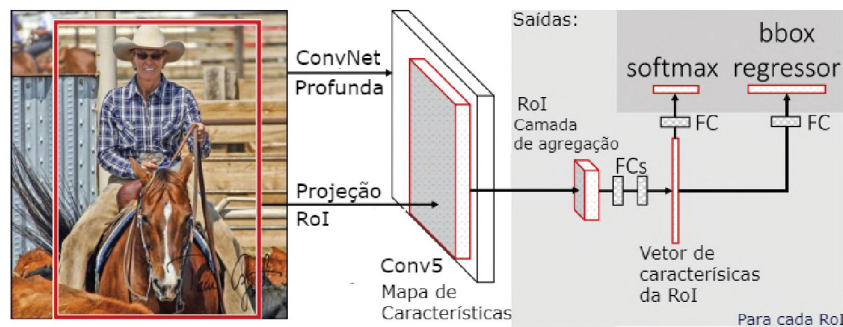


FIGURA 16 – ARQUITETURA FAST R-CNN. FONTE: GIRSHICK (2015).

Uma das vantagens da Fast R-CNN é que a rede aproveita o uso da GPU para acelerar o cálculo da proposta de objetos (SHAFII et al., 2017). Segundo (GIRSHICK, 2015), por meio da Fast R-CNN a velocidade de treino e teste é melhorada, ao mesmo tempo que a precisão de detecção é aumentada. O autor concluiu isso após um comparativo com a R-CNN, onde a Fast R-CNN treina a rede VGG nove vezes mais rápido, e é 213 vezes mais rápida em tempo de teste, além de atingir um resultado mais alto de detecção na base de dados PASCAL VOC 2012.

2.7.3 Faster R-CNN

Pouco tempo depois do desenvolvimento da rede apresentada na seção anterior, foi desenvolvido por (REN et al., 2016) o algoritmo de detecção Faster R-CNN que tem como diferencial a adição de uma Rede de Proposta de Regiões (*Region Proposal Network* - RPN) permitindo que a rede compartilhe camadas convolucionais, enquanto que, anteriormente utilizava-se um procedimento separado para extração de regiões.

A RPN compõem o primeiro módulo da Faster R-CNN que identifica a região do objeto de interesse na imagem de entrada e o segundo módulo é uma rede que classifica os objetos que foram propostos. A Faster R-CNN inicialmente processa a imagem de

entrada usando uma CNN para extrair as características da imagem e obter os mapas de características para serem usados pela RPN. Em seguida, a RPN varre os mapas de características obtidos pela janela deslizante e calcula dois tipos de pontuações. O primeiro cálculo indica se a janela em análise contém um objeto de interesse de acordo com o último mapa de características da camada de convolução e o segundo cálculo verifica se o objeto é de interesse ou não (REN et al., 2016).

No processo de janela deslizante, é difícil detectar objetos de várias formas e tamanhos se for usada apenas uma janela de tamanho fixo. Logo, para que a detecção de objetos de tamanhos diferentes seja possível, são usadas janelas deslizantes de tamanho variáveis. Ou seja, múltiplas K janelas, chamadas de caixas de âncoras (*anchors boxes*) são geradas em diferentes escalas e proporções. Uma âncora é uma região contida em uma janela deslizante que captura os objetos e prediz as suas coordenadas por meio do ajuste das dimensões. Na configuração original da Faster R-CNN são geradas 9 caixas de âncoras ($K=9$) que consiste em 3 tipos de escalas e 3 tipos de proporções definidas (Figura 17).

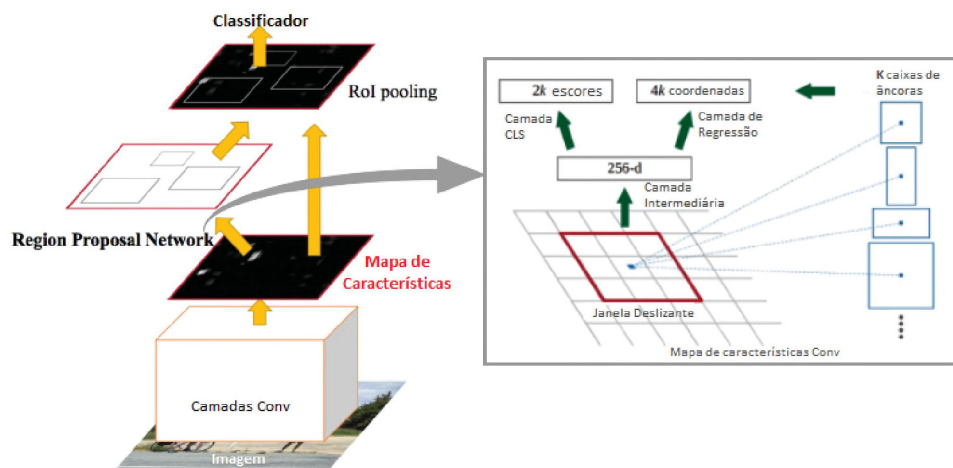


FIGURA 17 – FASTER R-CNN. À DIREITA: REDE DE PROPOSTA DE REGIÕES. FONTE: REN (2016).

A rede classifica como contendo objeto na âncora se a Intersecção sobre a União (IoU) da região for maior que um determinado valor. O IoU é um limiar no qual, um *bounding-box* é avaliado como correto se a proporção da área de sobreposição entre o *bounding-box* predito e o *bounding-box* verdadeiro, e a área abrangida por ambos os *bounding boxes* é maior que o IoU positivo pré-definido (Figura 18). Já, um *bounding-box* é avaliado como não sendo objeto se é menor que o IoU negativo parametrizado. Nesse caso, as regiões que estão entre os dois limiares não são consideradas para o treinamento (GIRSHICK, 2015).

O segundo módulo da Faster R-CNN é a própria Fast R-CNN vista na Seção 2.6.2 que agrupa as regiões e encaminha para as camadas totalmente conectadas (FC) para fazer a classificação e a regressão. O modelo proposto de treinamento na Faster R-CNN é o *End-to-End* que treina a RPN e a Fast R-CNN em conjunto.

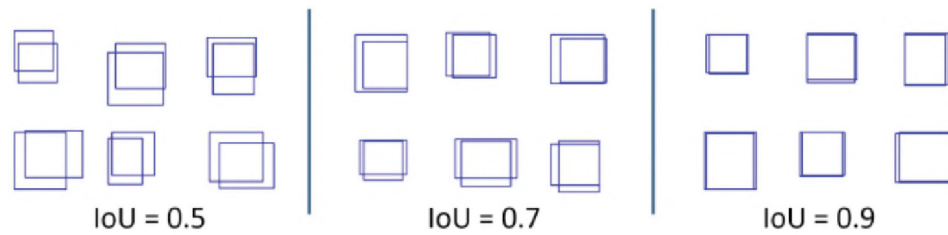


FIGURA 18 – DIFERENTES LIMIARES PARA IoU. FONTE: O autor (2018).

O autor da Faster R-CNN disponibiliza as duas arquiteturas de redes utilizadas, ZF e VGG. O projeto arquitetônico dessas duas redes é discutido brevemente nas próximas duas subseções. A versão do modelo pré-treinado da rede VGG e da rede ZF contém pesos aprendidos utilizando a base de dados PASCAL VOC 2007 (EVERINGHAM; WINN, 2007) e é disponibilizada juntamente com o algoritmo Faster R-CNN.

2.7.3.1 ZF

A Rede ZF é inspirada na AlexNet (Seção 2.4.2) e recebe originalmente imagens de entrada de 224×224 . A rede é composta por uma camada de entrada, cinco camadas de convolução, duas camadas totalmente conectadas e uma camada *softmax*. Na primeira camada de convolução, a ZF possui 96 filtros de tamanho 7×7 que são menores em relação ao tamanho de filtro 11×11 original da AlexNet. O número dos filtros usados nas camadas de convolução subsequentes são de 256, 384, 384 e 256 e o tamanho dos filtros são de 5×5 , 3×3 , 3×3 , 3×3 , respectivamente. A rede ZF utiliza a função de ativação ReLU e para função de erro utiliza a entropia cruzada.

2.7.3.2 VGG16

A rede VGG possui 16 camadas, 13 camadas de convolução e 3 camadas totalmente conectadas (FC). Essa quantidade de camadas foi definida após os autores obterem o melhor resultado criando diferentes redes que variavam o número de camadas entre 11 e 19. As camadas de convolução da VGG são agrupadas primeiramente duas a duas e depois três a três, usando uma camada de ativação ReLU após cada camada de convolução e uma camada de *pooling* após cada agrupamento de camadas. O tamanho da camada de *max pooling* usada é de 2×2 com *stride* 2, o que reduz pela metade as dimensões do mapa de características. A rede tem como entrada uma imagem RGB de dimensão 224×244 .

2.8 MÉTRICAS

Em modelos de aprendizagem e predição, o desempenho do sistema está relacionado com sua capacidade de recuperar e classificar as informações corretamente, assim como abstrair as informações erradas. Esta relação em um método de reconhecimento de padrões

é claramente representada pela matriz de confusão. A matriz de confusão apresenta as predições realizadas por um classificador, onde em sua diagonal principal estão armazenados os acertos do classificador, enquanto que nas demais posições estão as confusões ocorridas (VISA B. RAMSAY, 2011). A tabela 2 ilustra a matriz de confusão de duas classes.

TABELA 2 – MATRIZ DE CONFUSÃO GENÉRICA DE DUAS CLASSES. FONTE: VISA B. RAMSAY (2011).

	Predito Positivo	Predito Negativo
Valor Verdadeiro	VP	FN
Valor Falso	FP	VN

Na matriz de confusão de n classes, é possível classificar os dados em quatro indicadores:

- Verdadeiro Positivo (VP): representa a proporção de casos positivos que foram corretamente identificados;
- Falso Negativo (FN): A taxa FN é a proporção de positivos os casos que foram incorretamente classificados como negativos;
- Falso Positivo (FP): A taxa de falsos positivos (FP) é a proporção de casos negativos que foram incorretamente classificados como positivo;
- Verdadeiro Negativo (VN): A taxa VN é definida como a proporção de casos negativos que foram classificadas corretamente;

A partir da matriz de confusão e seus indicadores, o modelo pode ser classificado pela sua acurácia, precisão e sensibilidade/*recall*. A acurácia consiste na habilidade que o modelo possui de prever corretamente a classe de uma instância (VISA B. RAMSAY, 2011). Esta métrica é definida pela proporção de resultados corretamente classificados como positivos e negativos frente ao total de classes (VISA B. RAMSAY, 2011). A métrica acurácia é definida pela Equação 2.4.

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (2.4)$$

Precisão é o cálculo dos acertos das predições de uma classe, no caso para as predições positivas, esta relaciona a porcentagem de instâncias classificadas corretamente como positivas (VP) dentre o total de instâncias classificadas como positivas (TP+FP), apresentada pela Equação 2.5.

$$Precisão = \frac{VP}{VP + FP} \quad (2.5)$$

Recall relaciona a porcentagem de instâncias classificadas corretamente como positivas (VP), frente todas as instâncias realmente positivas (VP+FN). A Equação 2.6 define a relação da métrica *recall*.

$$Recall = \frac{VP}{VP + FN} \quad (2.6)$$

No contexto de localização de objetos, a média da Precisão Média de Detecção (mAP) (do inglês, *mean Average Precision*) é a métrica que vem sendo utilizada para medir a qualidade da predição na detecção de objetos e demonstra ter uma boa discriminação (REDMON et al., 2016).

O cálculo do AP é dado por:

$$AP = \sum_{i=1}^n Precisão_i \cdot \Delta Recall_i \quad (2.7)$$

Se todas as propostas estão corretas a Precisão e o *Recall* são iguais a 1 para cada i , de outro modo, se nenhuma das propostas é correta o AP é igual a 0. O mAP é a média do AP para todas as classes de um sistema.

3 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentados os trabalhos recentes relacionados com a classificação de padrões pulmonares. Os artigos dessa área foram escolhidos pelo critério ano de publicação, compreendendo o intervalo de 2012 a 2018 e extraídos do portal de periódicos da CAPES que compreende textos completos internacionais e nacionais de diversas bases de dados. No total foram selecionados 17 artigos que aplicam extratores de características e/ou classificadores no problema em questão e 1 artigo que faz a detecção e classificação de padrões pulmonares. Além de uma breve descrição de cada trabalho, no final deste capítulo, estão sintetizados os dados relacionados com a patologia, os padrões analisados, a base de imagens, a licença dos dados, a metodologia e os resultados individuais de cada trabalho.

3.1 ESTADO DA ARTE

3.1.1 Métodos de Extração e Classificação

A classificação de imagens normalmente é realizada em duas etapas: extração de características para codificar os *pixels* da imagem em descritores de características e a rotulagem das categorias de imagens. Vários tipos de técnicas foram propostas para caracterizar os vários padrões de doenças pulmonares extraídos de Regiões de Interesse (ROIs) de imagens médicas. Foi examinado principalmente, os trabalhos recentes que usam imagens de TCAR pulmonar em sua metodologia. Para classificar os padrões pulmonares dessas imagens, alguns autores usaram bases privadas e outros usaram a base de dados pública fornecida por Depeursinge et al. (2011), que também é adotada neste trabalho.

No estudo realizado por Chang et al. (2012), o autor relatou que a redução do número de características poderia ser mais eficaz na redução da complexidade computacional do que buscar melhorar algoritmos de classificação. Nesse contexto, foi proposto uma seleção de características com relação ao custo e uma Máquina de vetor de suporte (SVM) hierárquica com a abordagem *One-Against-All*, a fim de economizar tempo computacional usando primeiro classificação com características simples e, em seguida, classificação com características complexas. Esta abordagem garante uma classificação eficiente visto que, nos experimentos o tempo foi reduzido 57% e a exatidão global foi significativamente melhorada em comparação ao SVM *One-Against-All* (OAA) e SVM *One-Against-One* com seleção sequencial (SFS). Em relação a classificação, atingiu-se uma acurácia de 92,63% para as seis classes em questão. Como contribuição, o método desenvolvido por Chang et al. (2012) pode ser aplicado a outras tarefas de classificação em que o custo de aquisição de características é cara.

Segundo Depeursinge et al. (2012), a *affine invariance* é desejável para permitir a aprendizagem de texturas não determinísticas sem localizações, orientações ou tamanhos a priori. De acordo com essa afirmação, o autor apresentou descritores de textura *near-affine-invariant* derivados de estruturas de *wavelets* isotrópicas para a caracterização de padrões de tecido pulmonar em imagens de TCAR. A quantificação dos coeficientes de *wavelet* utilizando parâmetros de uma mistura de gaussianos combinados com o histograma de níveis de cinza em HUs (*Hounsfield Unit*) caracterizou de maneira eficiente os cinco padrões de tecidos estudados com uma acurácia global de 76,9% mostrando-se mais eficiente quando comparado com características baseadas na densidade e no padrão da estrutura *wavelet* discreta.

Seguindo a mesma linha de raciocínio de Depeursinge et al. (2012), Rodriguez et al. (2012) também utilizou um sistema baseado na transformada *wavelet* incluindo a Diferença de Gaussianos (DoG) trabalhando com a recuperação de imagem semelhantes e assim, permitindo a interpretação de imagens de TCAR. As palavras visuais foram utilizadas para a agregação de características e para descrever padrões de tecido. Diferente dos trabalhos anteriormente apresentados, a classificação baseou-se na imagem original de TCAR e não em blocos. O resultado de classificação de 95,5% de acurácia mostra um bom desempenho do sistema quando os parâmetros corretos são escolhidos e que o sistema é capaz de aprender as grandes variações intra-classe de tecido saudável e as características de tecido pulmonar anormal para fornecer assistência confiável aos médicos.

Conforme os estudos sobre classificação de padrões pulmonares foram avançando, outros desafios surgiram e nesse cenário a publicação de Song et al. (2013) propôs enfrentar a baixa variação extra-classes e alta variação intra-classe projetando um método de aproximação de *patch* de imagem baseado na característica. Inicialmente, um bloco da imagem foi representado como um vetor de características, baseado na textura proposta pelo descritor Gabor-LBP invariante a rotação (RGLBP) e pelos descritores *Multi-coordinate histogram of oriented gradients* (MCHOG). Em seguida, o bloco de imagem foi classificado em uma das cinco possíveis categorias de tecidos, usando o classificador *Patch-Adaptive Aparse Approximation* (PASA) proposto com base em blocos de imagens de referência.

O novo método PASA foi feito com os seguintes componentes principais: critérios de discrepância mínimos para a classificação escassa, adaptação específica do bloco para aproximação discriminativa e ponderação do espaço de características. Os rótulos dos blocos foram então acumulados em estimativas probabilísticas para a classificação. Por fim, foi atribuída uma única rotulagem para cada ROI adaptativa com base na estimativa probabilística coletiva. O método proposto obteve mais de 80% de acerto na classificação de cada uma das classes.

Ainda no ano de 2013, o artigo de Zhao et al. (2013) descreveu um método de diagnóstico assistido por computador (CAD) para classificar padrões de Doenças

Pulmonares Difusas (DLD) em imagens de TCAR. O método utilizado foi o *Sparse Representation*. Este método é composto por três etapas: aprendizagem de dicionário, codificação esparsa e agrupamento espacial. A extração de características foi feita por meio de valores de Tomografia Computadorizada (TC) e autovalores de matriz Hessiana para calcular características locais. Após a extração de características, o dicionário foi aprendido executando os algoritmos K - *Singular Value Decomposition* (K-SVD) e *Orthogonal Matching Pursuit* (OMP) iterativamente até que o erro de reconstrução fosse menor que a condição de parada. Em seguida, as características locais originais foram representadas por OMP usando o dicionário aprendido. Finalmente, o agrupamento espacial foi adotado para combinar os coeficientes de representação do mesmo VOI (*Volume-of-interest*) em um descritor servindo como entrada de um classificador SVM. O método proposto tem uma precisão global de 95,4%, e as taxas de reconhecimento dos padrões são superiores a 90%.

Assim como Zhao et al. (2013), o estudo de Chang et al. (2013) também usou o classificador SVM mas com o intuito de comparar o mesmo com o classificador Bayesiano em três condições diferentes para investigar a precisão de classificação de padrões regionais nas imagens de TCAR. Na etapa de extração de características utilizou-se 22 características quantitativas obtidas de diferentes descritores. Primeiro, na etapa de classificação, optou-se por estimar a precisão dos classificadores usando imagens de *scanners* separadamente. Após isso, os dados dos *scanners* GE e Siemens foram utilizados para treinamento e teste. Em seguida, os dados de ROIs das diferentes fontes foram integrados, treinados e testados juntos. Como resultado, foi alcançada uma precisão de 92% e 82% nas imagens do *scanner* GE com o SVM e o classificador Bayesiano, respectivamente. Já, na classificação das imagens Siemens foi alcançado 92% e 86%, também com o SVM e o classificador Bayesiano, respectivamente, ou seja, notou-se que foram obtidas melhores precisões de classificação com o classificador SVM do que o classificador Bayesiano.

Ainda no propósito de auxiliar o diagnóstico clínico, um novo método utilizando *bag-of-features* para a classificação de seis tipos de padrões de DLD foi apresentado por Xu et al. (2013). A proposta considerou as texturas pulmonares por valores de TC e pela forma, assim, foram considerados um conjunto de medidas estatísticas baseadas em características locais calculadas a partir de valores de TC e de valores próprios da matriz Hessiana. As características estatísticas utilizadas foram média, desvio padrão, *skewness* e curtose. A classificação foi feita por meio de um SVM e a precisão global do método desenvolvido foi de 93,18%.

Observando os estudos anteriores, Anthimopoulos et al. (2014) afirmou que tipicamente as patologias existentes de Doenças Pulmonares Intersticiais (DPIs) são identificadas e quantificadas pela classificação de blocos da imagem local. Diante disso, o autor propôs um método para a descrição e classificação de blocos de imagem de TCAR com patologias de DPI com base na transformada local 2D de cosseno discreto (DCT) e na classificação por

Floresta Randômica (RF). Para a análise espectral, utilizou um banco de filtros baseados em DCT. A distribuição das imagens filtradas foi descrita por *q-quantiles* juntamente com os valores de máximo e mínimo, em adição a um histograma de nível de cinza da imagem original. O conjunto de características altamente discriminativas combinado com o classificador de Floresta Randômica (RF) apresentou resultado de F-score médio igual a 89% e acurácia igual a 85,61%.

Na sequência dos trabalhos publicados, os autores de Bagesteiro et al. (2015) propuseram a análise e quantificação de cinco padrões de doenças pulmonares em imagens de TCAR usando o descritor *Complete Local Binary Pattern* com um classificador SVM. Além da classificação dos padrões radiológicos o método proposto realiza uma segmentação pulmonar grosseira com a técnica de dilatação para classificar áreas corporais fora dos pulmões. Com isso, a principal contribuição do método proposto é a não dependência de um pulmão previamente segmentado, pois a segmentação grosseira é realizada por meio da classificação de áreas corporais fora dos pulmões. Após extrair características com o *Complete Local Binary Pattern* (CLBP) e classificar com SVM, atingiu-se 98,91% de acurácia na classificação de blocos não pulmonares e 91,81% de precisão na classificação dos padrões pulmonares normais e patológicos, apresentando resultados semelhante aos trabalhos da literatura utilizam a pré-segmentação do pulmão.

Após 2 anos da primeira publicação nessa área, em 2015, Song et al. (2015) apresenta uma nova proposta ainda com o objetivo de melhorar a precisão de classificação tratando a grande variação intra-classe e a ambiguidade inter-classe no espaço de características. Neste contexto, foi utilizado um novo modelo de classificação chamado de *Large Margin Local Estimate* (LMLE). Para aplicar o método LMLE, cada imagem foi primeiramente representada por um vetor de características de textura-intensidade-gradiente (TIG). Na sub-categorização dos conjuntos de referência, as subcategorias de referência derivadas apresentaram uma menor variação intra-classe em comparação com os conjuntos de referência geral. Depois disso, a representação esparsa foi usada para gerar as estimativas locais em nível de subcategoria. O experimento feito mostrou uma acurácia de 86,10% ao tratar o problema pelo qual foi proposto.

Mais recentemente, Revathi et al. (2016) desenvolveu um sistema de identificação de pulmão patológico, o qual, visa fornecer uma solução genérica para a identificação de patologias pulmonares em TC por meio de uma abordagem com duas etapas. Na primeira etapa, *Fuzzy-Connectedness* (FC) são usados para segmentar automaticamente e obter o volume das imagens de pulmão que tem uma grande quantidade diversificada de anormalidades pulmonares. O primeiro volume utilizado foi o resultado da segmentação FC e o segundo foi o resultado do método *Morphological Rib Cage* que remove regiões usando operações morfológicas. Com o resultado do teste de patologia foi possível inferir a presença de patologia ou não pois a imagem foi classificada como patológica se existisse uma

diferença de volume. Na segunda fase, para refinar os resultados usou-se um classificador de floresta randômica, que é uma coleção de árvores de decisão, em que uma grande gama de características de textura foram utilizadas para treinar o classificador. Esse classificador refina as imagens e detecta a existência de patologia. A saída do classificador teve uma acurácia de 99,89% apresentando como resposta se a classe era patológica ou não patológica.

Por último, Yang et al. (2016) utilizou *textons* afirmando que estes são mais adaptáveis para tarefa de classificação com conjuntos de treinamento pequenos em comparação com métodos baseados em *deep-learning*. Nesse contexto, foi explorada a caracterização de texturas pulmonares com decomposição esparsa de dicionários de *textons*. Métodos baseados em *textons* clássicos constroem dicionários dos blocos de imagens, caracterizando as texturas e rotulando os blocos dentro de ROIs de acordo com o *texton* mais semelhante e, então, geram histogramas de frequência de *textons*. Neste trabalho, primeiro foram exploradas características de textura do pulmão com decomposição esparsa de dicionários de *textons*, utilizando diferentes estratégias de regularização e, em seguida, foi feita a redução de restrição de indução de dispersão para a construção de dicionários. Duas estratégias de construção de dicionário foram utilizadas, primeiro a construção baseada em *textons* e segundo a alternativa de aprendizado não supervisionado de dicionário com regularização da dispersão. Em todos os modelos, exceto para o modelo de *multiple-dictionary*, foi alcançada uma acurácia de classificação superior a 90%. O dicionário concatenado apresentou a melhor taxa de acerto, chegando a 91,7%.

3.1.1.1 Métodos de Aprendizado Automático

Nessa seção, estão descritos os trabalhos relacionados com redes neurais na etapa de aprendizado de representação de conhecimento e/ou na classificação. No relato de cada trabalho são mostradas as arquiteturas de redes neurais propostas por cada autor.

Para diferenciar vários padrões de tecido pulmonar comumente vistos nas DPIS, Li et al. (2013) propõe criar extratores de características em escala múltipla com base em um algoritmo de aprendizado não supervisionado (Figura 19). Nesse caso, foram obtidos os vetores de características da imagem por um processo de convolução. A rede escolhida foi uma máquina de Boltzmann Restrita (RBM) que é uma rede neural estocástica e generativa capaz de capturar e reproduzir a estrutura estatística de um determinado conjunto de dados. O método Gaussiano RBM (GRBM) foi utilizado por ser capaz de modelar o valor real de entrada como imagens naturais.

O GRBM consiste em duas camadas de neurônios, uma camada visível e uma camada oculta. Os nós em camada visível e camada escondida foram conectados por uma matriz de pesos não direcionada, sendo que cada nó tem um valor de polarização associado (Figura 19). As amostras foram utilizadas em várias escalas para capturar

características de diferentes tamanhos e três redes foram treinadas, todas, configuradas com 16 neurônios ocultos. O vetor de características de cada bloco foi classificado em uma das cinco categorias de tecido usando um classificador SVM, chegando a 84% de acurácia na classificação da classe micronódulos.

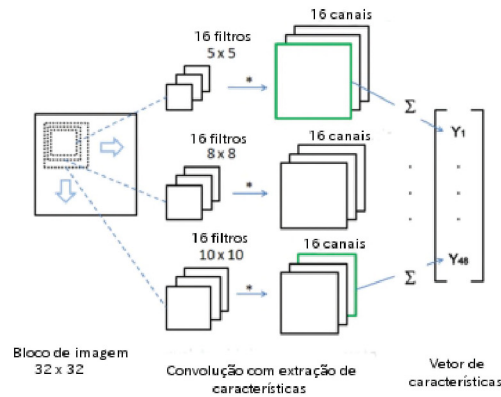


FIGURA 19 – ARQUITETURA PROPOSTA POR LI ET AL. (2013) PARA EXTRAÇÃO DE CARACTERÍSTICAS. FONTE: LI et al. (2013).

O mesmo autor do artigo anterior, em 2014, apresentou outro trabalho para propor uma Rede Neural Convolutiva (CNN) auto supervisionada (Figura 20) que detecta os diferentes padrões de doenças pulmonares intersticiais usando remoção randômica de nodos e uma única camada convolutiva. A proposta foi usar poucas camadas para simplificar a rede melhorando sua eficiência e evitando o problema de *overfitting*. O algoritmo de remoção randômica de um nodo funciona isolando ele na fase de treinamento. Sua saída e sua propagação de erro são ignoradas. A seguir, uma matriz preenchida aleatoriamente a cada iteração do treinamento indica os estados *on* e *off* de cada neurônio. A proposta apresentada mostrou uma taxa de acerto de 74%. Os resultados apresentados são valores aproximados pois não estão contidos no texto e foram inferidos por meio de uma análise da tabela de resultados do autor (LI et al., 2014).

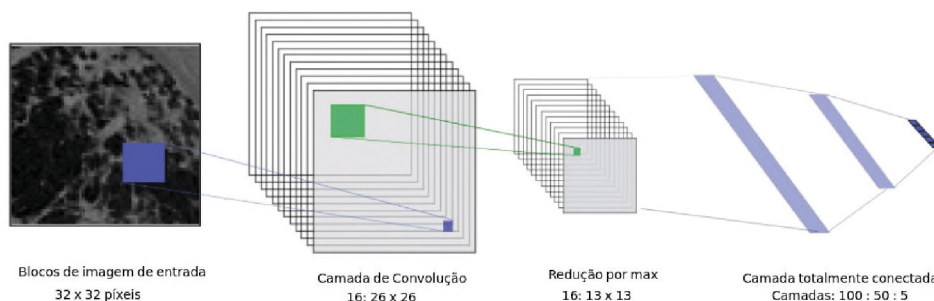


FIGURA 20 – ARQUITETURA PROPOSTA POR LI ET AL. (2014). PARA EXTRAÇÃO E CLASSIFICAÇÃO. FONTE: LI et al. (2014).

Aprimorando mais ainda as técnicas de classificação de padrões pulmonares, Gao et al. (2015) faz uma classificação holística de seis padrões em imagens de TC por meio

de redes neurais convolucionais profundas. A principal diferença com os trabalhos vistos anteriormente é que o algoritmo proposto além de classificar blocos, também utiliza a imagem inteira como uma entrada holística evitando a necessidade das ROIs como pré-requisito de entrada. O método foi proposto com três escalas de atenuação com relação aos padrões de anormalidade pulmonar que foram capturadas por reescalonamento da imagem de TC original em Unidades Hounsfield (HU), as quais, sofreram redimensionamento para $[0, 255]$ e serviram como entrada no treinamento e no teste. As faixas selecionadas foram as de padrões com menor atenuação, com maior atenuação e de atenuação pulmonar normal.

Segundo Gao et al. (2015), o uso de três faixas de atenuação oferece melhor visibilidade ou separação visual entre as seis categorias de DPis. Na etapa final, para cada fatia de entrada 2-D, foi usada a técnica *data augmentation* que aplica transformações e gera amostras aleatórias a partir das imagens originais, em seguida, essas imagens foram redimensionadas para 224×224 pixels usando uma interpolação linear. A taxa de acerto obtida foi de aproximadamente 89,6% para a classificação em blocos e de 68,6% para a classificação a nível de fatia, sendo no segundo caso, a classe enfisema foi corretamente classificada em todos as vezes apareceu.

Também utilizando a técnica *data augmentation* que trouxe resultados promissores no trabalho de Gao et al. (2015), Anthimopoulos et al. (2016), indica e avalia uma CNN desenhada para a classificação de padrões DPI (Figura 21). A rede proposta possui 5 camadas convolucionais com núcleo de 2×2 e ativações *LeakyReLU*, seguidas por *pooling* médio com tamanho igual ao tamanho dos mapas de características finais e com três camadas densas. A última camada densa possui sete saídas, equivalente ao mesmo número das classes de DPI em estudo. A técnica *data augmentation* foi aplicada usando 15 transformações, para cada classe, o número necessário de amostras aumentadas foi selecionado aleatoriamente, de modo que todas as classes atingiriam o tamanho do conjunto de treinamento da classe menos rara. O desempenho da classificação foi de aproximadamente 85,5% demonstrando novamente o potencial das CNNs para análise de padrões pulmonares.

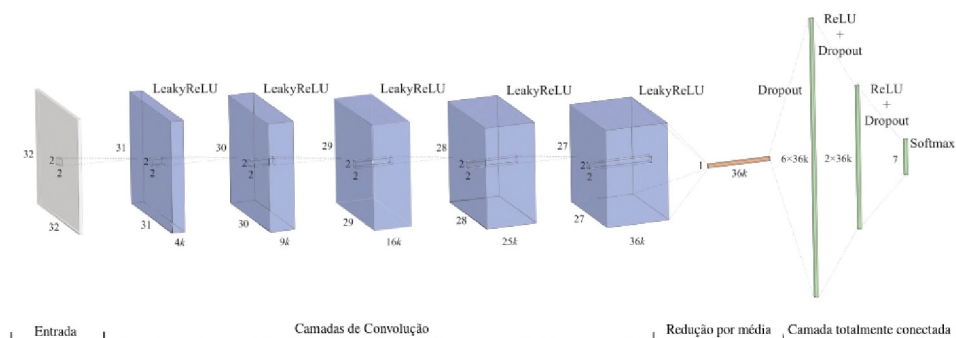


FIGURA 21 – ARQUITETURA PROPOSTA POR ANTHIMOPOULOS ET AL. (2016) PARA EXTRAÇÃO E CLASSIFICAÇÃO. FONTE: ANTHIMOPOULOS et al. (2016).

Na pesquisa de Lu et al. (2016), além de ser estudada a classificação de doença pulmonar intersticial (DPI) foi estudada a detecção de linfonodos (LN) tóraco-abdominal. Este trabalho difere-se da maioria dos outros trabalhos por utilizar CNNs Profundas para Detecção Assistida por Computador com o objetivo final de classificar a presença ou ausência da doença. Inicialmente foram exploradas e avaliadas diferentes arquiteturas de CNNs. Os modelos estudados possuem entre 5 mil e 160 milhões de parâmetros e variam também no número de camadas. Foram então avaliadas as características dos conjuntos de dados e das imagens inteiras de TCAR. Como última análise foi examinado quando e por que a transferência de aprendizagem de uma rede pré-formada pode ser útil. Pelos motivos apontados em Gao et al. (2015) e por não depender da segmentação pulmonar, Lu et al. (2016) também optou por transformar as imagens em canais de RGB e usar diferentes canais de atenuação de TC para definir regiões diretamente no espaço HU. Quanto aos resultados, no contexto de interesse desta revisão de literatura, o melhor caso de classificação por fatia de DPI foi 90,2% de acurácia (GoogleNet-TL) e a melhor classificação por blocos atingiu 91,1% de acurácia.

Por fim, concluindo a revisão dos trabalhos baseados em aprendizado automático, foi selecionado o trabalho de VanTulder et al. (2016), no qual foram realizados experimentos para a classificação de textura pulmonar e detecção de vias aéreas em imagens de TC. Nesse estudo, é usada uma RBM que é um tipo de rede neural recorrente estocástica que combina a aprendizagem de representação discriminativa supervisionada e generativa. A RBM convolucional usa o padrão de compartilhamento de peso de redes convolucionais para aprender pequenos filtros, os quais, são aplicados para todas as posições de uma imagem maior. Esse compartilhamento de peso torna a aprendizagem mais eficiente e permite que a RBM modele pequenas características que aparecem em várias áreas de uma imagem facilitando a descrição de texturas. A capacidade de usar os dois tipos de aprendizagem, generativa e discriminativa, distingue a classificação da RBM de muitos outros métodos de aprendizagem representação. Foram utilizadas as representações aprendidas pela RBM e treinados e classificados cada bloco em uma das cinco classes de tecido. No conjunto de dados de tecido pulmonar, foram estudados os efeitos da combinação de aprendizagem com um objetivo generativa e discriminativa. O classificador utilizado foi uma floresta randômica e o resultado alcançou 74,2% para a classificação dos blocos de TC.

3.1.2 Detecção e Classificação de Objetos

No contexto de detecção e classificação de padrões pulmonares sem realizar segmentação da região pulmonar, o trabalho de Agarwala et al. (2018) utilizou o algoritmo Faster R-CNN e a base de dados (DEPEURSINGE et al., 2011). Os modelos pré-treinados das redes ZF e VGG foram ajustados para localizar padrões de DPI usando a estrutura da Faster R-CNN. Nesse estudo, foram utilizadas marcações de blocos não sobrepostos das fatias de TCAR com tamanho 16×16 para uma imagem de 256×256 e marcações de 32

$\times 32$ para imagens de 512×512 , as quais, contém mais de 75% de um padrão de uma ROI. Essas marcações foram utilizadas para entrada da Faster R-CNN. O autor utilizou os três padrões consolidação, enfisema e fibrose justificando que estes são os padrões mais difíceis de classificar. Como resultado, foi apresentado um valor de mAP de 75% e 83% com a configuração das redes ZF e VGG, respectivamente.

3.2 RESUMO DO ESTADO DA ARTE

Na Figura 22 é possível visualizar um comparativo geral do estado da arte para classificação de padrões radiológicos. Por questões cronológicas, os trabalhos estão listados por ano de publicação. Em uma síntese geral podemos destacar que a melhor precisão na classificação de blocos de VOIs (3D) de seis classes, foi de 95,5% obtida no trabalho de Zhao et al. (2013) por meio da classificação com um SVM. Em outra análise (Figura 22, dos quatro trabalhos que utilizam CNNs a melhor acurácia alcançada na classificação de ROIs de cinco padrões foi 91,1% no trabalho desenvolvido pelo autor Lu et al. (2016).

Trabalho	Patologia	Padrões Analisados	Licença DataSet	ROIs	Método de extração de Características	Classificação	Resultado
Chang (2012)	Doença Pulmonar Intersticial Difusa (DILD)	Normal, Enfisema, Opacidade em vidro-fosco, Consolidação, Opacidade reticular e Faveolamento	Privada	600 ROIs de 32x32 pixels	13 extratores de Textura: descritores estatísticos, descriptor de gradiente, codificação de comprimento de execução, entropia e co-ocorrência de nível de escala de cinza 11 extratores de Forma: Transformação tophat e análise de agrupamentos	SVM com arquitetura de árvore binária (hierárquico)	Classificação de ROIs: Acurácia: 92,63%
Depeursinge (2012a)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose e Micronódulos	Pública	17848 ROIs de 31x31 pixels	Valores de coeficiente de wavelet combinado com histograma de níveis de cinza	SVM com Kernel Gaussiano	Classificação de ROIs: Acurácia: 76,9%
Foncubierta-Rodriguez (2012)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose, Micronódulos e Consolidação	Pública	-	Análise Wavelet	Multiscale Visual Words	Classificação em tecido saudável e não saudável: Precisão: 94,5% Classificação de fatias de TC: Precisão: 60,96% (dado calculado)
Song (2013)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose e Micronódulos	Pública	1458 AROIS de 31x31 pixels	Gabor-LBP invariante a rotação e Multi-coordinate histogram of oriented gradients	Dicionários de referência usando o algoritmo patch-adaptive sparse approximation (PASA) seguido de estimativas probabilísticas da classificação.	Classificação de ROIs: Precisão: 82,66% (dado calculado)
Zhao (2013)	Doenças pulmonares difusas (DLD)	Normal, Enfisema, Consolidação, Opacidade em Vidro-fosco, Faveolamento e Nodular	Privada	2360 VOIs de 32x32x32	Valores de TC (Textura) e forma 3D (autovalores da matriz Hessiana)	SVM com Kernel Linear	Classificação de ROIs: Precisão: 95,4%
Xu (2013)	Doenças pulmonares difusas (DLD)	Normal, Consolidação, Opacidade em Vidro-fosco, Faveolamento, Enfisema e Nodular	Privada	3009 VOIs de 32x32x32	Conjunto de medidas estatísticas calculadas a partir de valores de TC e valores da matriz Hessiana.	SVM com Kernel Gaussiano	Classificação de ROIs: Precisão: 93,18%
Chang (2013)	Doença Pulmonar Difusa	Normal, Enfisema, Opacidade em vidro-fosco, Opacidade reticular, Faveolamento e Consolidação	Privada	600 ROIs de 20x20 pixels	Histograma, Gradiente, Comprimento de corrida, Matriz de co-ocorrência de nível de cinza, Cluster de área de baixa atenuação e Transformação Tophat	SVM e Classificador Bayesiano	Classificação de fatias de TC (Acurácia): SVM: 92% Bayesiano: 82% Classificação de ROIs (Acurácia): SVM: 92% Bayesian: 77%

Li (2013)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose e Micronódulos	Pública	16220 ROIs de 32x32 pixels	Extração automática por convolução	SVM	Classificação de ROIs: Acurácia: 74,1% (dado calculado)
Li (2014)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose e Micronódulos	Pública	16220 ROIs de 32x32 pixels	Extração automática por convolução	CNN	Classificação de ROIs: Acurácia: 74% (dado calculado)
Bragesteiro (2015)	Doenças Pulmonares Difusas (DPDs)	Não-pulmão, Normal, Enfisema, Opacidade em vidro-fosco, Fibrose e Micronódulos	Pública	20540 ROIs de 32x32 pixels	Completed Local Binary Pattern - CLBP	SVM	Classificação de ROIs: Acurácia: 91,81%
Song (2015)	-Doença Pulmonar Intersticial (ILD). Classificação fenotípica binária e Classificação contínua Regressão na ressonância magnética cerebral (MR)	Normal, Enfisema, Vidro-fosco, Fibrose e Micronódulos	Pública	2062 ROIs de 31x31 pixels	Gabor-local, Gabor-local invariante de rotação, Histograma de intensidade e Multi-coordinate histogram of oriented gradients (MCHOG)	Classificação de large margin local estimate (LMLE)	Classificação de ROIs: Acurácia: 86,10%
Gao (2015)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose, Micronódulos e Consolidação	Pública	2048 ROIs de 31x31	Extração automática por convolução	CNN	Classificação de ROIs: Acurácia: 89,6% Classificação de fatias de TC: Acurácia: 68,6%
Revathi (2016)	Patologias pulmonares	Patológico e Não-patológico	Pública	-	Matriz de Co-ocorrência de Níveis de Cinza, Matriz de Comprimento de Corrida de Cinza e Histograma	Floresta Aleatória	Classificação de ROIs: Acurácia: 99,69%
Yang (2016)	Doença pulmonar obstrutiva crônica	Normal, Enfisema centrolobular e Enfisema paraseptal	Pública	168 ROIs de 61x61 pixels	4 modelos baseados em dicionário de textons 3 modelos baseados no dicionário de aprendizagem com regularização da dispersão	Textons	Classificação de ROIs: Acurácia: 91,7% usando o modelo de dicionário concatenado
Anthimopoulos (2016)	Doença Pulmonar Intersticial (ILD)	Normal, Opacidade em vidro fosco, Micronódulos, Consolidação, Reticulação, Faveolamento e Combinação de GGO e reticulação.	Pública e Privada	14696 ROIs de 32x32 pixels	Extração automática por convolução	CNN	Classificação de ROIs: Acurácia: 85,61%
Lu (2016)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose, Micronódulos e Consolidação	Pública	100 ROIs de 64x64 pixels	Extração automática por convolução	CNN	Classificação de fatia de TC: Acurácia: 90,2% Classificação de ROIs: Acurácia: 91,1%
VanTuider (2016)	Doença Pulmonar Intersticial (ILD)	Normal, Enfisema, Vidro-fosco, Fibrose e Micronódulos	Pública	ROIs de 32x32 voxels	RBM - diferentes combinações de aprendizado discriminativo e generativo	Floresta Aleatória	Classificação de ROIs: Acurácia: 74,2% (dado calculado)

FIGURA 22 – RESUMO COMPARATIVO ENTRE AS TÉCNICAS PARA CLASSIFICAÇÃO DE PADRÕES RADIOLOGÍCOS.

Na Tabela 3 é apresentado o resumo do estado da arte para a detecção e classificação de padrões pulmonares obtido pelo autor (AGARWALA et al., 2018).

TABELA 3 – RESULTADOS DO TRABALHO AGARWALA ET AL. (2018) PARA DETECÇÃO DE PADRÕES PULMONARES. FONTE: AGARWALA et al. (2018).

Trabalho	Patologia	Licença DataSet	Tamanho da Imagem	Tamanho da Anotação	Rede	Padrões Analisados	Resultado
Agarwala (2018)	Doença Pulmonar Intersticial	Pública	512 x 512	32 x 32	ZF	Consolidação Enfisema Fibrose	mAP: 73%
					VGG		mAP: 75%
			224 x 224	16 x 16	ZF		mAP: 75%
					VGG		mAP: 83%

Em relação aos trabalhos publicados e referenciados no estado da arte, destaca-se que existe um grande número de estudos que envolvem somente a classificação de doenças pulmonares, enquanto que trabalhos que realizam a detecção e classificação são escassos. Portanto, a fim de complementar na Tabela 4, é apresentada uma comparação entre as melhores arquiteturas para detecção de objetos feita pelo autor (REDMON et al., 2016). Este comparativo foi realizado na base de dados PASCAL VOC 2007 que possui grupos de treinamento e teste com aproximadamente cinco mil imagens cada e que estão rotuladas em 20 categorias de objetos (EVERINGHAM; WINN, 2007). Observa-se que o melhor resultado geral foi obtido com a arquitetura a YOLOv2 com 78,6% mAP, porém a Faster R-CNN é utilizada neste trabalho pois disponibiliza um modelo pré-treinado, o código é acessível e os resultados são próximos ao estado da arte.

TABELA 4 – IMAGENS COM MAIS DE UMA CLASSE. FONTE: O autor (2018).

Arquitetura	mAP
YOLOv2	78,6%
SSD500	76,8%
Faster R-CNN ResNet	76,4%
SSD300	74,3%
Faster R-CNN VGG-16	73,2%
Fast R-CNN	70,0%
YOLO	63,4%

Os resultados detalhados por classe podem ser encontrados em (REDMON et al., 2016). Em geral, os resultados de detecção de objetos são ligeiramente menores que os resultados de trabalhos que fazem apenas a classificação. Essa variação é explicada pelo fato de ter mais uma etapa envolvida no processo que aumenta notavelmente o nível de dificuldade de reconhecimento de padrões.

4 METODOLOGIA E RESULTADOS

Nos problemas de classificação de imagem, o caráter descritivo e discriminativo das características extraídas é um fator crítico para se alcançar um bom desempenho de classificação (LI et al., 2014). Portanto, neste trabalho é proposta a utilização de uma estrutura de aprendizagem automática baseada em neurônios para a classificação de imagens de DPis.

4.1 VISÃO GERAL

A partir da escolha de CNNs como principal metodologia de classificação, o presente trabalho divide-se em duas grandes etapas:

- Analisar o desempenho da técnica de transferência de aprendizagem em Redes Neurais Convolucionais para classificação de DPis;
- Detectar e classificar DPis com o uso da metodologia Faster R-CNN.

Uma visão geral da primeira etapa que consiste em classificar os blocos de DPis, pode ser visualizada na Figura 23. A ideia é analisar a precisão de classificação de patologias pulmonares por meio de redes pré-treinadas com o intuito de investigar a viabilidade da utilização da técnica de transferência de aprendizagem em conjunto com outras técnicas baseadas em redes neurais convolucionais.

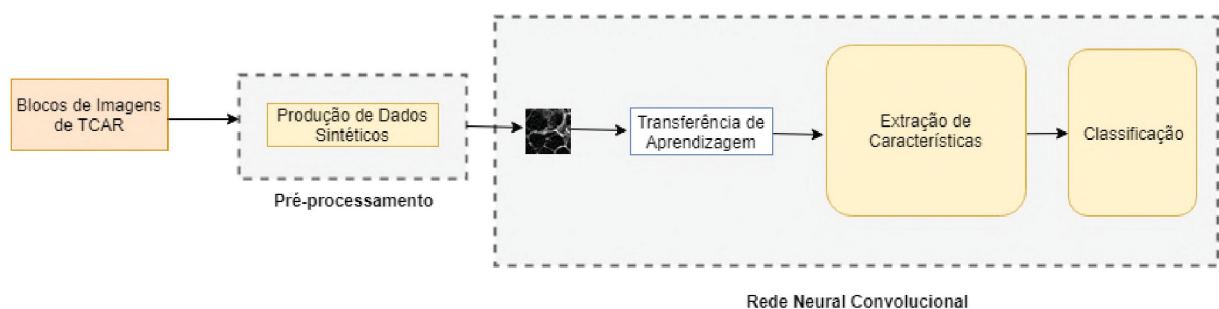


FIGURA 23 – VISÃO GERAL DA CLASSIFICAÇÃO COM A TRANSFERÊNCIA DE APRENDIZAGEM. FONTE: O autor (2018).

A segunda etapa deste trabalho propõe detectar e classificar os padrões radiológicos diretamente na imagem de TCAR buscando uma alternativa para automatizar o processo de caracterização de doenças pulmonares. Com essa técnica deseja-se suprir a necessidade de uma base pré-segmentada e trabalhar diretamente com as fatias inteiras de TCAR possibilitando a detecção de mais de uma doença por imagem. Na Figura 24, é apresentada uma visão geral dessa metodologia.

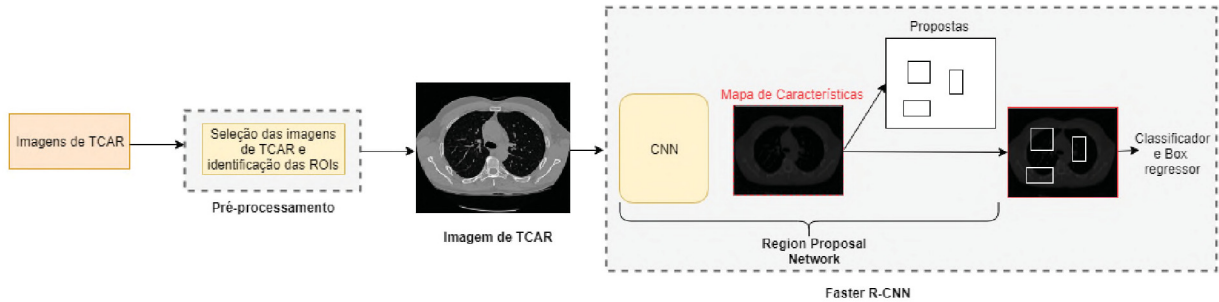


FIGURA 24 – VISÃO GERAL DA CLASSIFICAÇÃO COM A Faster R-CNN. FONTE: O autor (2018).

4.2 BASES DE IMAGENS

A realização deste trabalho tem como referência a base de imagens pública com casos de Doenças Pulmonares Intersticiais fornecida por (DEPEURSINGE et al., 2011) que contém imagens de TCAR com uma espessura de corte de 1 mm. A base é composta por conjuntos de TCAR com casos de doenças pulmonares de 113 pacientes e, cada conjunto possui de 11 a 61 cortes tomográficos.

As imagens são disponibilizadas em alta resolução no formato DICOM (*Digital Imaging and Communications in Medicine*) com tamanho de 512×512 pixels e com uma espessura de corte de 1 mm. Além disso, a base fornece máscaras dos pulmões segmentados e também, anotações de Regiões de Interesse (ROIs) de 17 padrões de tecidos realizada por especialistas da área, dos quais, cinco são utilizados no presente trabalho (Figura 25). No total são 1448 ROIs desenhadas manualmente em fatias 2-D de TCAR feitas por especialistas com 15 e 20 anos de experiência nos Hospitais da Universidade de Genebra.

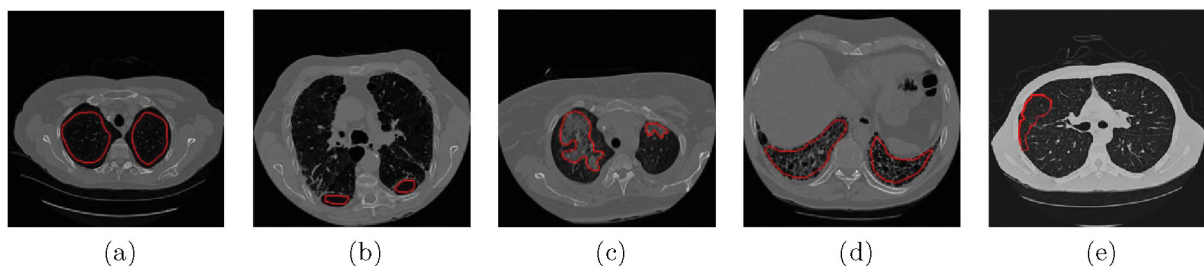


FIGURA 25 – AMOSTRA DAS PADRÕES NAS IMAGENS DA BASE DEPEURSINGE ET AL. (2010): (A) NORMAL; (B) ENFISEMA; (C) VIDRO-FOSCO; (D) FIBROSE; (E) MICRONÓDULOS. FONTE: DEPEURSINGE et al. (2010).

Para aumentar o número de amostras de tecido saudável são utilizados exames de DPIs em estágios iniciais, pois quantidade de exames sem a presença de alguma doença é baixa. O diagnóstico de cada caso foi confirmado pelo exame de patologia (biópsia e lavagem bronco-alveolar) ou por um teste de laboratório específico. Para maiores detalhes, uma descrição completa da base de dados pode ser encontrada em (DEPEURSINGE et al., 2011).

Uma segunda configuração extraída da base citada anteriormente e disponibilizada por (BAGESTEIRO et al., 2015) também foi utilizada neste trabalho. Essa base é composta por blocos de imagens com meia-sobreposição no eixo x e y de 32×32 *pixels* que contém pelo menos 75% dos *pixels* pertencente ao padrão da ROI (Figura 26).

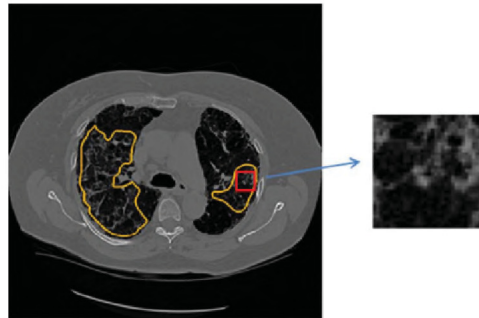


FIGURA 26 – EXTRAÇÃO DE BLOCOS DE ROIs FEITA PELO AUTOR BAGESTEIRO ET AL. (2015). FONTE: BAGESTEIRO et al. (2015).

De acordo com (BAGESTEIRO et al., 2015), o tamanho de 32×32 *pixels* foi adotado com base em estudos anteriores que mostram que esse tamanho de bloco possui o melhor equilíbrio entre o desempenho da classificação e localização, além de justificar que o tamanho mínimo para preservar as formas dos padrões pulmonares deve ser de 20×20 *pixels*. A base extraída contém 20.540 blocos de imagens de 90 pacientes rotulados em 6 classes: não pulmão (NP), normal (N), enfisema (E), vidro-fosco (GG), fibrose (F) e micronódulos (M). Todos os detalhes dessa base podem ser consultados em (BAGESTEIRO et al., 2015).

4.3 CLASSIFICAÇÃO COM REDES NEURAIAS CONVOLUCIONAIS

Para o estudo de CNNs com técnica de transferência de aprendizagem foi utilizada a base de blocos de imagens disponibilizada por (BAGESTEIRO et al., 2015). A escolha da base justifica-se pelo fato de que cada bloco contém uma única classe visto que a classificação da rede é feita por imagem.

4.3.1 Pré-Processamento

Na etapa inicial deste trabalho optou-se por remover a classe não-pulmão (3.000 blocos) da base de imagens (BAGESTEIRO et al., 2015) para que fosse possível a comparação dos resultados com a maioria dos trabalhos existentes. Na sequência, todas as imagens em formato DICOM foram convertidas para o formato PNG de modo que as amostras fossem compatíveis com a entrada do classificador utilizado (Tabela 5). Para a conversão utilizou-se os intervalos de HU no valor médio de -600 e janela de 1600 que está na faixa representativa do pulmão em Unidades de Hounsfield (WELLS; HIRANI, 2008).

4.3.2 Produção de Dados Sintéticos

Depois disso, observou-se que a distribuição das amostras era muito discrepante, variando de 9 a 5.940 amostras por paciente, valores relativos a incidência do tipo de doença. Sendo assim, para corrigir o desbalanceamento das classes nos blocos de imagens e também para suprir a necessidade de grandes conjuntos de dados para o aprendizado da rede (KRIZHEVSKY et al., 2012), utilizou-se a técnica de criação de dados sintéticos que é exemplificada na Figura 27. Essa técnica tem sido frequentemente empregada na classificação de imagem, a fim de aumentar a quantidade de dados de treinamento e evitar a sobreposição (KRIZHEVSKY et al., 2012).

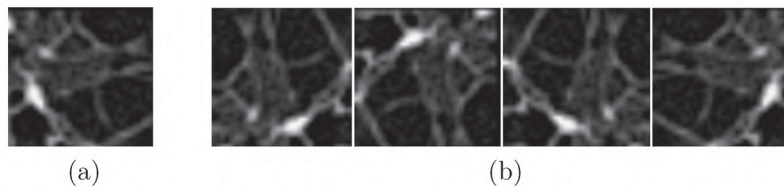


FIGURA 27 – EXEMPLO DA CRIAÇÃO DE DADOS SINTÉTICOS NA BASE BAGESTEIRO ET AL. (2015): (A) AMOSTRA ORIGINAL; (B) AMOSTRAS SINTÉTICAS. FONTE: O autor (2018).

Foram aplicados filtros de transformações geométricas nas imagens: rotação (ângulos: 90° , 180° e 270°), espelhamento (horizontal e vertical) e a combinação desses filtros, sempre com a preocupação que as transformações geradas fossem diferentes. A escolha dos filtros foi motivada no fato de não ocasionar perda de informação e não alterar a textura das imagens no momento da geração de dados sintéticos.

Cada filtro foi aplicado sobre as imagens originais de acordo com a classe, de modo que todas as classes atingissem um tamanho semelhante ao conjunto de imagens da classe com mais exemplares. Após gerar um número semelhante de amostras por classe criou-se novamente dados artificiais sobre o novo conjunto de dados com o objetivo de aumentar a quantidade total de blocos para a classificação. O aumento de dados na base pode ser visto na Figura 28, e o detalhe de todas as transformações na Tabela 5 resultando em um total de 82.494 blocos de imagens de 90 pacientes.

TABELA 5 – BLOCOS DE IMAGENS. FONTE: O autor (2018).

	N	E	GG	F	M	Total	Nº de Pacientes
Base Original (PNG)	5707	1017	1942	2694	5492	16852	90
Base Aumentada (PNG)	17121	15255	17478	16164	16476	82494	90

Na sequência, foi feita a divisão dos blocos de imagens em conjuntos de pacientes para compor os grupos de treinamento, teste e validação, garantindo que as imagens de um mesmo paciente não pertencessem ao mesmo tempo a grupos diferentes. Para isso, foram selecionadas aproximadamente 70% das imagens para treinamento, 20% para teste e 10%

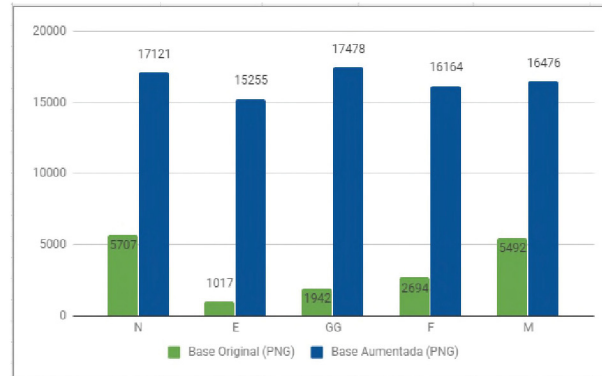


FIGURA 28 – QUANTIDADE DE BLOCOS APLICANDO A PRODUÇÃO DE DADOS SINTÉTICOS. FONTE: O autor (2018).

para validação, a distribuição foi escolhida baseada no critério que o grupo de treinamento deve ser composto pelo maior número de amostras conforme as distribuições mais utilizadas nos trabalhos relacionados (ANTHIMOPOULOS et al., 2014; GAO et al., 2015; LU et al., 2016). Na Tabela 6 é apresentada a divisão dos grupos para a CNN aplicada a base desbalanceada enquanto que na Tabela 7 é apresentada a divisão dos grupos aplicada a base balanceada. A quantidade de pacientes por grupo pode ser observada na Tabela 8.

TABELA 6 – DIVISÃO DOS GRUPOS PARA CNN NA BASE DESBALANCEADA. FONTE: O autor (2018).

	Treino	Teste	Validação	Total
Imagens	10605	2003	1144	13752

TABELA 7 – DIVISÃO DOS GRUPOS PARA CNN NA BASE BALANCEADA. FONTE: O autor (2018).

	Treino	Teste	Validação	Total
Imagens	56943	17079	8472	82494

TABELA 8 – DIVISÃO DOS PACIENTES POR GRUPO. FONTE: O autor (2018).

	Treino	Teste	Validação	Total
Pacientes	69	10	9	90

4.3.3 Criação de uma Arquitetura Própria de CNN

Para o aprendizado automático foi utilizada uma estrutura de aprendizado de máquina neural automática específica para extrair características discriminantes dos blocos de imagem de treinamento e classificá-las. Por questões comparativas, antes de testar a técnica de transferência de aprendizagem proposta, foi criada uma arquitetura própria com pesos não definidos para o problema e testado o desempenho de classificação dessa rede no contexto de imagens médicas.

Utilizou-se como modelo de arquitetura a rede AlexNet e a partir desta arquitetura foram variados os parâmetros de taxa de aprendizagem, número de neurônios, número de

camadas convolucionais, número de *pooling*, número filtros, tamanho do *kernel* e *stride*. Os parâmetros definidos foram escolhidos de acordo com estudos anteriores de CNN (HAFEMANN et al., 2014) que tiveram bons resultados para a classificação de textura em outros contextos.

A rede foi ajustada por meio de testes empíricos para a problemática apresentada e a melhor arquitetura construída é composta pelas seguintes camadas e parâmetros:

- Camada de entrada: parâmetros de acordo com a resolução da imagem de 32×32 *pixels*;
- Duas combinações de camadas convolucionais e de *pooling*: cada camada convolucional com 64 filtros, com tamanho de filtro definido para cada problema e *stride* definido igual a 1. O *pooling* possui 34 filtros de tamanho 3×3 e *stride* 2;
- Duas camadas localmente conectadas: 32 filtros de tamanho 3×3 e *stride* igual a 1;
- Camada de saída altamente conectada com 5 saídas equivalente ao número de classes do problema.

A matriz de confusão obtida com a arquitetura própria adaptada da rede AlexNet para a classificação das 5 classes em estudo é apresentada na Tabela 9.

TABELA 9 – ARQUITETURA PRÓPRIA ADAPTADA DA ALEXNET. FONTE: O autor (2018).

	N	E	GG	F	M
N	74,40%	3,18%	7,11%	5,96%	9,32%
E	21,05%	76,26%	0,00%	0,67%	2,02%
GG	0,03%	0,03%	90,39%	7,24%	0,05%
F	2,81%	5,56%	6,17%	81,17%	4,29%
M	25,53%	22,06%	0,00%	0,06%	52,34%

4.3.4 Transferência de Aprendizagem

Após a classificação inicial sem o uso de pesos aprendidos, foi testada a técnica de transferência de aprendizagem de um domínio de imagem diferente do domínio de imagens médicas. Ao contrário de uma CNN criada para um determinado problema em que o aprendizado da rede é iniciado do zero, uma representação já aprendida em uma tarefa anterior foi utilizada como entrada da rede neural para avaliar a performance de classificação.

4.3.4.1 Arquiteturas de Redes Neurais

Devido a aquisição de dados ser difícil e a anotação de qualidade ser custosa, não existe um conjunto de dados de imagens médicas rotulados em grande escala. Diante disso,

optou-se por elaborar os testes com as redes AlexNet, ImageNet e GoogleNet já montadas. Inicialmente foi testada a configuração original das redes sendo necessário redimensionar o tamanho da imagem de entrada de 32×32 *pixels* para o tamanho utilizado por cada arquitetura e considerar a imagem de tons de cinza como uma imagem de 3 canais. Além disso, foi alterado o número de camadas de saída para 5 e a taxa de aprendizagem para 0.001.

Segundo (LITJENS et al., 2017), a transferência de aprendizagem e o ajuste fino são componentes chave para o uso de CNNs profundas em aplicações de imagens médicas. Portanto, depois da parametrização inicial, foi feito o ajuste fino para o problema apresentado por meio de testes empíricos buscando diminuir o erro de classificação.

Com o intuito de gerar um comparativo mais amplo e para destacar a importância da inicialização de parâmetros e o impacto do uso da técnica de criação de dados sintéticos foi testada a configuração acima proposta nas variações com e sem transferência de aprendizagem e com e sem produção de dados sintéticos, conforme apresentado abaixo:

- (a) Rede com os pesos zerados e a base desbalanceada;
- (b) Rede com os pesos zerados e a base balanceada;
- (c) Rede com os pesos aprendidos e a base desbalanceada;
- (d) Rede com os pesos aprendidos e a base balanceada;

Abaixo estão apresentados os resultados da classificação das diferentes configurações testadas após o ajuste fino de parâmetros. Para ambas as tabelas, o resultado apresentado refere-se a melhor acurácia obtida e a época correspondente. Na Tabela 10, estão os resultados sem o uso da transferência de aprendizagem e sem o balanceamento das classes enquanto que na Tabela 11 é mostrado o resultado após a utilização da produção de dados sintéticos.

TABELA 10 – REDE COM OS PESOS ZERADOS E BASE DESBALANCEADA. FONTE: O autor (2018).

Rede	Melhor Acurácia do Teste	Época
AlexNet	72,68%	5000
ImageNet	78,24%	4000
GoogleNet	80,92%	7000

Nos testes usando a técnica de transferência de aprendizagem e mantendo a mesma parametrização inicial, a rede não estava apresentando convergência da taxa de erro. As primeiras épocas de treinamento apresentavam uma precisão maior e conforme o treinamento avançava a taxa de precisão não sofria variações relevantes. Visto que os pesos não estavam sendo aprendidos do zero, admitiu-se como hipótese que a taxa

TABELA 11 – REDE COM OS PESOS ZERADOS E BASE BALANCEADA. FONTE: O autor (2018).

Rede	Melhor Acurácia do Teste	Época
AlexNet	77,92%	8000
ImageNet	85,40%	7000
GoogleNet	86,00%	8000

de aprendizagem deveria ser menor para que a rede aprendesse do ponto de onde foi inicializada. Por este motivo, no momento do treinamento usando a técnica de transferência de aprendizagem o parâmetro taxa de aprendizagem foi alterado de 0.001 para 0.0001 e a rede começou a convergir e aprender novamente conforme os resultados das Tabelas 12, 13.

TABELA 12 – REDE COM OS PESOS APRENDIDOS E BASE DESBALANCEADA. FONTE: O autor (2018).

Rede	Melhor Acurácia do Teste	Época
AlexNet	76,75%	6000
ImageNet	84,71%	7000
GoogleNet	84,36%	6000

TABELA 13 – REDE COM OS PESOS APRENDIDOS E BASE BALANCEADA. FONTE: O autor (2018).

Rede	Melhor Acurácia do Teste	Época
AlexNet	75,8%	7000
ImageNet	87,32%	10000
GoogleNet	90,95%	9000

Um comparativo dos resultados pode ser observado na Figura 29. Os valores apresentados correspondem a média calculada após a execução do treinamento trinta vezes, garantindo que o algoritmo de treinamento é o mesmo e que as condições de inicialização são as mesmas.

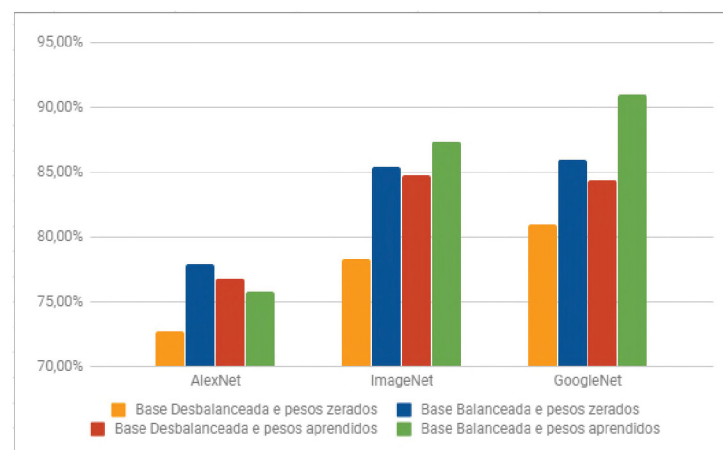


FIGURA 29 – COMPARATIVO DAS CLASSIFICAÇÕES PROPOSTAS. FONTE: O autor (2018).

Dentre as configurações testadas, a rede que apresentou melhor resultado foi a *GoogleNet* combinada com as técnicas de aumento de dados e transferência de aprendizagem. A matriz de confusão obtida é mostrada na Tabela 14.

TABELA 14 – MATRIZ DE CONFUSÃO DA MELHOR CONFIGURAÇÃO PARA CLASSIFICAÇÃO POR BLOCOS. FONTE: O autor (2018).

	N	E	GG	F	M
N	69,41%	14,20%	9,15%	1,92%	5,32%
E	2,04%	94,11%	0,00%	1,50%	2,35%
GG	5,03%	0,07%	78,87%	10,02%	6,01%
F	0,09%	1,60%	4,14%	89,87%	4,30%
M	10,52%	0,00%	3,13%	2,18%	84,17%

4.4 DETECÇÃO E CLASSIFICAÇÃO COM A FASTER R-CNN

Como segunda etapa deste trabalho foi aplicada a técnica Faster R-CNN para detectar e classificar automaticamente os padrões radiológicos em imagens de TCAR. Foi utilizada a estrutura desenvolvida pelos autores da *Faster Region-based Convolutional Networks* (REN et al., 2016) e construída usando o Framework Caffe para o aprendizado profundo.

Conforme apresentado na Seção 4.2.2, os melhores resultados são encontrados quando combinadas as técnicas de transferência de aprendizagem e produção de Dados sintéticos em CNNs. Concluindo isso, nessa seção o algoritmo Faster R-CNN foi parametrizado para utilizar ambas as técnicas. Para aumentar a diversidade dos dados durante o treinamento utilizou-se a implementação existente de filtros de transformações geométricas da Faster R-CNN. Em relação a transferência de aprendizagem, foram empregados os pesos aprendidos pelas redes ZF, VGG e GoogleNet após o treinamento na base ImageNet, a qual, tem mais de 1,2 milhões de imagens anotadas em 1000 categorias de classes de objetos. Justifica-se a utilização da arquitetura da rede GoogleNet na detecção e classificação das DPis pois, esta rede apresentou os melhores resultados no estudo inicial de classificação de DPis.

Para analisar o desempenho da Faster R-CNN foram definidos diferentes pré-processamentos de imagens, configurações de anotações de dados e arquiteturas de redes neurais, conforme listado nos tópicos abaixo e discutido nas próximas seções:

- Pré-Processamento de Imagens
 - (a) Conversão em Escala de Cinza: 1 canal em Unidades de Hounsfield
 - (b) Conversão para RGB: 3 canais em Unidades de Hounsfield
- Configuração de Anotações

- (a) *Bounding-Box* por ROI do pulmão
 - (b) *Bounding-Box* por ROI de DPI
 - (c) *Bounding-Box* por Blocos de ROI de DPI
- Arquiteturas de Redes Neurais
 - (a) VGG original da Faster R-CNN
 - (b) ZF original da Faster R-CNN
 - (c) GoogleNet adaptada para a Faster R-CNN

4.4.1 Pré-processamento de Imagens

Para que fosse possível trabalhar diretamente com os exames de TCAR da base de imagens original fornecida por (DEPEURSINGE et al., 2011), foram realizados pré-processamentos dos dados para compatibilidade das imagens com a entrada da Faster R-CNN.

O primeiro pré-processamento consistiu em converter as imagens do formato DICOM para PNG em escala de cinza e a aplicação de um filtro de escala de HU. Para a conversão, foram considerados os mesmos intervalos de HU utilizados nos blocos na Seção 4.3.1, sendo a faixa de HU médio de -600 e janela de 1600. A Figura 30 apresenta um exemplo do pré-processamento com a aplicação do filtro em escala de HU para entrada da Faster R-CNN.

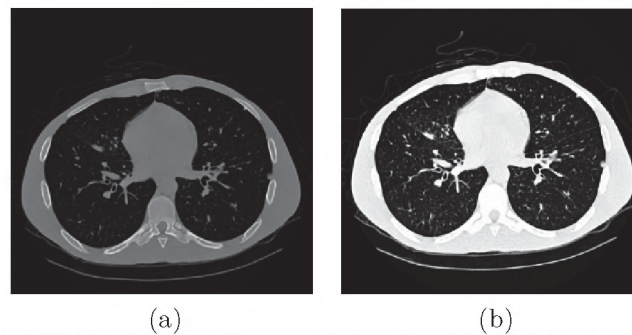


FIGURA 30 – (a) CONVERSÃO SIMPLES. (b) CONVERSÃO NA ESCALA DE HU. FONTE: O autor (2018)

O segundo pré-processamento realizado foi a conversão das imagens DICOM para PNG em escala RGB. A proposta da utilização de três canais foi para estar em concordância com as arquiteturas da CNN, que são treinadas com o conjunto de dados da ImageNet (KRIZHEVSKY et al., 2012), a qual utiliza os valores RGB das imagens.

Em adição, como citado no trabalho de (GAO et al., 2015), uma prática para otimizar os efeitos de visualização de determinados tecidos ou patologias nas tomografias computadorizadas é o ajuste de contraste da atenuação pelos radiologistas. Com isto, o

autor destaca que a utilização de três faixas de atenuação oferece melhor visibilidade ou separação visual entre as categorias de doenças pulmonares.

Assim, foram selecionadas três janelas de atenuação de TCAR com reescalamento para $[0, 255]$ e aplicadas nas imagens. As janelas de atenuação foram definidas no intervalo $HU = [-1400, -950]$ para baixa atenuação, $HU = [-1400, -200]$ para escala normal e $HU = [-160, 240]$ para alta atenuação. A partir das três imagens, uma nova imagem RGB foi gerada posicionando cada faixa de atenuação em um canal de cor. A Figura 31 apresenta a união das três faixas de atenuação selecionadas, as quais, geram a nova imagem RGB.

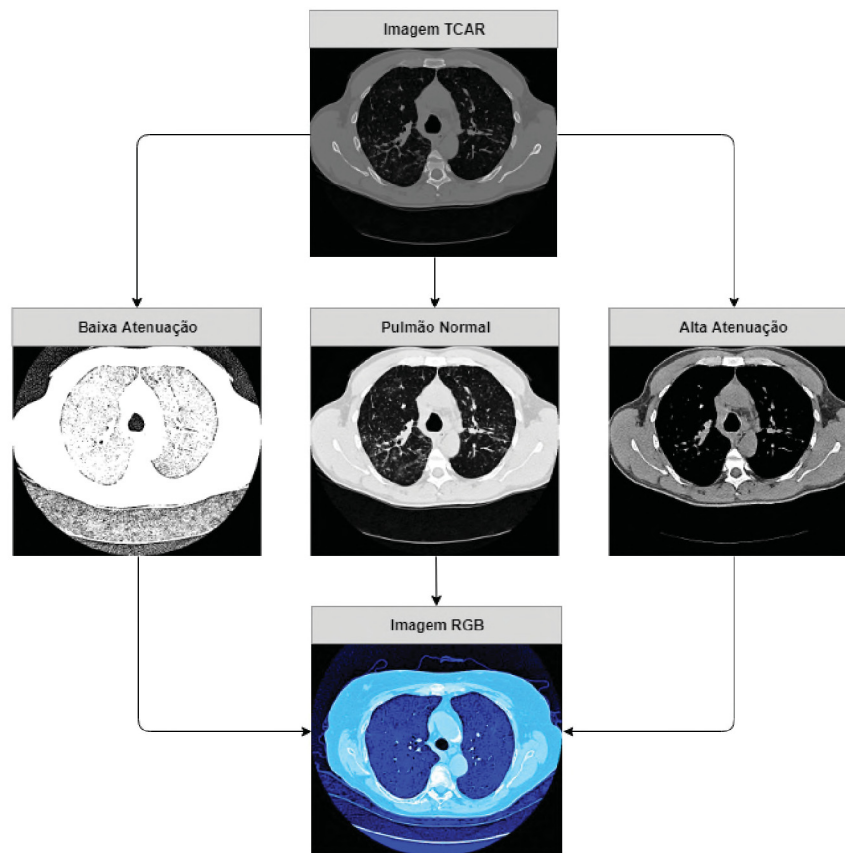


FIGURA 31 – ESCALAS DE ATENUAÇÃO DE TC. FONTE: O autor (2018).

O total de imagens de TCAR obtido foi de 632 e de ROIs foi de 1103, a quantidade de imagens por classe estão descritas na Tabela 15. Na Tabela 16 também estão listados os pacientes que apresentam mais de um padrão de patologia.

TABELA 15 – QUANTIDADE DE IMAGENS POR CLASSE. FONTE: O autor (2018).

Classe	Normal	Enfisema	Vidro-fosco	Fibrose	Micronódulos	Total
Imagens	69	41	241	278	153	632
ROIs	148	86	284	337	248	1103

TABELA 16 – PACIENTES COM MAIS DE UMA CLASSE. FONTE: O autor (2018).

Paciente	Classes	Quantidade de Imagens
017	Enfisema e Fibrose	2
142	Fibrose e Vidro-Fosco	2
160	Enfisema e Fibrose	4
172	Enfisema e Fibrose	2
205	Normal e Enfisema	1
208	Normal e Enfisema	4

4.4.2 Inicialização da Faster R-CNN

Na entrada da Faster R-CNN, além da imagem, devem ser adicionados arquivos *xml* de anotações conforme o Código 4.1 com as informações do tamanho da imagem, da classe e das coordenadas iniciais e finais de cada padrão contido na imagem.

Código 4.1 – EXEMPLO DE ARQUIVO DE ANOTAÇÃO USADO PELA Faster R-CNN

```
<annotation>
  <folder>ILD</folder>
  <filename>008_CT-INSPIRIUM-8871_CT-slice007_label_2.png</filename>
  <source>
    <database>ILD Database</database>
  </source>
  <size>
    <width>512</width>
    <height>512</height>
    <depth>2</depth>
  </size>
  <object>
    <name>ground_glass</name>
    <bndbox>
      <xmin>147</xmin>
      <ymin>330</ymin>
      <xmax>176</xmax>
      <ymin>347</ymin>
    </bndbox>
  </object>
</annotation>
```

Portanto, para trabalhar com a Faster R-CNN é necessário definir as coordenadas iniciais e finais de uma área na qual a ROI está contida. Uma vez que a base (DEPEUR-SINGE et al., 2011) fornece somente informações das coordenadas do contorno de cada ROI, para encontrar os *bounding boxes* das ROIs foi utilizado o algoritmo *Label Image Regions* da biblioteca *Scikit-image* desenvolvida em *Python* (YU, 2013) que segmenta uma imagem e encontra cada região fechada contida na mesma. O passo a passo do algoritmo é descrito sequencialmente a seguir:

- Faz um limiar com o método Otsu automático;
- Fecha os pequenos orifícios com o fechamento binário;

- Remove os artefatos das bordas que ficaram unidas;
- Mede as regiões da imagem para filtrar objetos pequenos.

Na triagem do tamanho dos objetos, foi definido um tamanho mínimo de *bounding box* de pelo menos 32 *pixels* de altura ou largura. Um exemplo visual da aplicação do algoritmo pode ser observado nas Figuras 32 e 33. Além das marcações na imagem de saída, o algoritmo fornece as coordenadas de cada objeto a partir da máscara contida na base de imagens, tais coordenadas foram utilizadas como entrada para o detector de objetos Faster R-CNN.

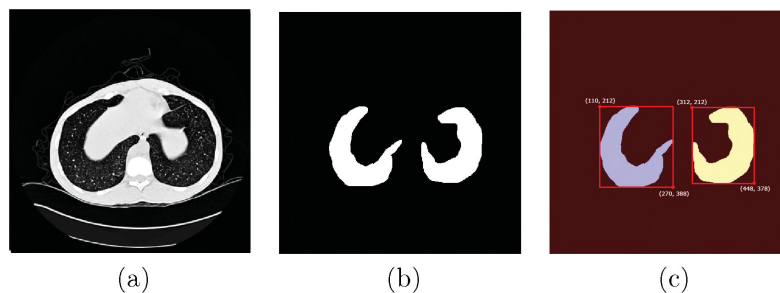


FIGURA 32 – UTILIZAÇÃO DO ALGORITMO *LABEL IMAGE REGIONS* PARA REGIÃO DO PULMÃO (A) TCAR ORIGINAL (B) MÁSCARA: ENTRADA DO ALGORITMO; (C) LOCALIZAÇÃO DAS REGIÕES FECHADAS: SAÍDA DO ALGORITMO. FONTE: O autor (2018).

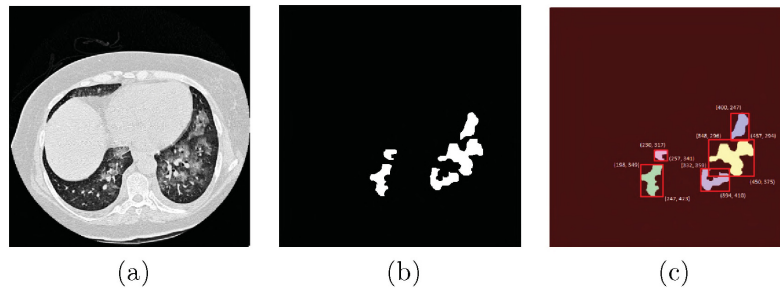


FIGURA 33 – UTILIZAÇÃO DO ALGORITMO *LABEL IMAGE REGIONS* PARA REGIÃO DE DPIs (A) TCAR ORIGINAL (B) MÁSCARA: ENTRADA DO ALGORITMO; (C) LOCALIZAÇÃO DAS REGIÕES FECHADAS: SAÍDA DO ALGORITMO. FONTE: O autor (2018).

Outra proposta para configuração das anotações da Faster R-CNN foi definir *Bounding-Box* por blocos de ROI de DPIs uma vez que a maioria dos trabalhos relacionados com DPIs estudados no Capítulo 3 utilizam essa metodologia. Além disso, refinando a região de interesse por blocos é garantido que pelo menos 75% da ROI pertence a área selecionada. Nesse caso, a entrada da rede continua sendo a imagem inteira de TCAR mas o *Bounding-Box* passa a ser blocos fixos de 32×32 *pixels* de meia sobreposição no eixo x e y. Na Figura 34 é apresentado um exemplo de como são definidos os blocos de ROIs utilizando as mesmas coordenadas da base (BAGESTEIRO et al., 2015).

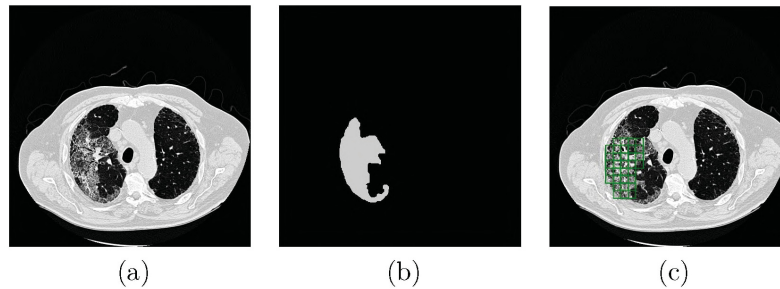


FIGURA 34 – (A) TCAR ORIGINAL (B) MÁSCARA DA ROI (C) LOCALIZAÇÃO DE BLOCOS DE ROIS DE DPIS. FONTE: BAGESTEIRO et al. (2015).

4.4.3 Arquiteturas de Redes Neurais

Por padrão, na implementação original da Faster R-CNN as redes utilizadas como extratores de características para a CNN são a VGG e ZF (REN et al., 2016). Por não existir a construção da rede GoogleNet na Faster R-CNN foi feita a adaptação dessa rede para a arquitetura da Faster R-CNN. Foram utilizadas todas as camadas antes da saída da camada *inception4e* (*conv1 - inception4e*) como “extratores de características” para a CNN. Após isso, foram incluídas as camadas da RPN e na sequência utilizou-se as camadas restantes da GoogleNet (*inception5a - pool5*) como a “rede de Classificação”. Em detalhes, a construção dessa nova arquitetura pode ser observada na Figura 35.

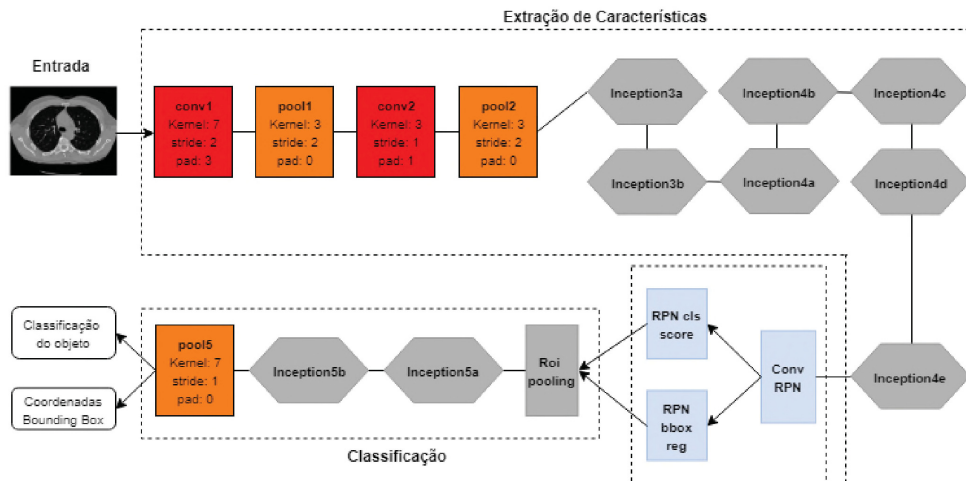


FIGURA 35 – FASTER R-CNN ADAPTADA PARA A ARQUITETURA DA GOOGLNET. FONTE: O autor (2018).

4.4.4 Treinamento e Teste

Primeiramente, foi explorada a eficiência da Faster R-CNN no contexto de imagens médicas. Para tanto, foi assumida a detecção e classificação do problema binário de segmentação de região pulmonar. O algoritmo foi adaptado do contexto original para trabalhar com duas classes, ou seja, Pulmão e Não Pulmão (*Background*).

Como a classificação do pulmão não é o foco do presente trabalho e é feita apenas para fins de avaliar o desempenho da Faster R-CNN no contexto de imagens médicas, a execução do algoritmo foi realizada na sua configuração original. Os resultados para esta aplicação estão na Tabela 17 e foram avaliadas de acordo com a Precisão Média de Detecção (mAP), métrica comumente utilizada para a detecção de objetos (REDMON et al., 2016).

TABELA 17 – DETECÇÃO E CLASSIFICAÇÃO DO PULMÃO COM A FASTER R-CNN NA CONFIGURAÇÃO ORIGINAL. FONTE: O autor (2018).

Imagem	Rede	Épocas	mAP
1 Canal	ZF	10mil	87.5%
1 Canal	VGG	10mil	91.1%
1 Canal	GoogleNet	10mil	90.6%

Na Imagem 36, é apresentado um exemplo de detecção de pulmão com a Faster R-CNN, onde o *bounding-box* verde delimita a região que contém o pulmão e o *bounding-box* azul representa a marcação gerada pelo detector na imagem de TCAR.

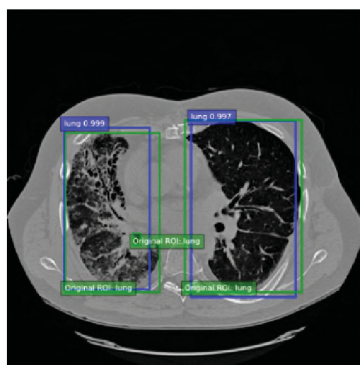


FIGURA 36 – EXEMPLO DE DETECÇÃO E CLASSIFICAÇÃO DO PULMÃO COM A FASTER R-CNN. FONTE: O autor (2018).

Uma vez que a Faster R-CNN detectou e classificou a região do pulmão com um mAP alto, o algoritmo foi adaptado novamente para trabalhar com seis classes: as cinco classes de interesse já utilizadas anteriormente na Seção 4.3 (Normal, Enfisema, Fibrose, Vidro-fosco e Micronódulos) com adição da classe “*Background*”. Foi modificada a camada de saída de classificação da Faster R-CNN para ser igual a 6 enquanto que a camada FC de regressão foi modificada para 24, ou seja, 4 coordenadas dos *bounding boxes* vezes 6 classes.

O desempenho da detecção de DPIs pulmonares foi avaliado pelo modelo proposto *End-to-End*. O conjunto de dados foi dividido em treino/validação e teste de acordo com a Tabela 18. A divisão dos grupos de imagens foi empregada para não conter imagens de TCAR derivadas do mesmo paciente no conjunto de treinamento e no conjunto de teste.

Com relação à configuração de parâmetros, muitas das configurações utilizadas nessa pesquisa foram herdadas das configurações originais da Faster R-CNN. No entanto,

TABELA 18 – DIVISÃO DOS GRUPOS PARA FASTER R-CNN. FONTE: O autor (2018).

	Treino/Validação	Teste	Total
Imagens	540	92	632

foram realizados vários testes empíricos variando os parâmetros taxa de aprendizagem, tamanho da imagem de entrada, IoU positivo, IoU negativo e NMS (*Non-Max Suppression*).

Na primeira abordagem utilizando *Bounding-Box* por ROI, a imagem de entrada foi testada em duas escalas, na escala original 512 e na escala 1024, de acordo com estudos anteriores que tiveram melhor desempenho aumentando a escala da imagem (REIAZI et al., 2018). Enquanto que na abordagem *Bounding-Box* por Blocos de ROI não foi aumentada a escala da imagem pois os blocos de ROIs têm tamanho fixo de 32×32 logo, utilizou-se a imagem original de 512×512 . O ajuste da Supressão Não Máxima (NMS) com base no escore da classe também foi empregada na abordagem de *Bounding-Box* por Blocos de ROI pois, como são consideradas as anotações de objetos com meia sobreposição o NMS deve ser ajustado para a definir a melhor janela entre as múltiplas janelas produzidas pela RPN (REN et al., 2016; SUN et al., 2018; REIAZI et al., 2018).

Em ambas as abordagens os parâmetros IoU de regiões sobrepostas foram variados, para exemplos positivos foram testadas as configurações de $\text{IoU} \geq 0.5$ e ≥ 0.75 e para exemplos negativos foram testados os valores de $\text{IoU} \leq 0.3$ e ≤ 0.35 , ambos para a RPN e a *Fast* R-CNN, em concordância com as duas configurações mais utilizadas nos trabalhos de detecção de objetos (GIRSHICK, 2015; REN et al., 2016; REDMON et al., 2016). Também, a taxa de aprendizagem variou de 0.001 (configuração original) a 0.0001 uma vez que observou-se que a variação desse parâmetro altera significativamente o resultado da classificação. Uma listagem da compilação do melhor resultado e a época correspondente para cada uma das configurações testadas na Faster R-CNN utilizando a anotação por ROI é exibida na Tabela 19 e os principais testes utilizando marcações por Blocos de ROIs são mostrados na Tabela 20.

TABELA 19 – CONFIGURAÇÕES TESTADAS NA FASTER R-CNN PARA ANOTAÇÃO POR ROI INTEIRA. FONTE: O autor (2018).

Imagem	Rede	Tamanho Imagem	Taxa de Aprendizagem	IoU Positivo	IoU Negativo	NMS	Épocas	Uso GPU	N	E	GG	F	M	mAP Teste
1 Canal	ZF	512 × 512	0,001	0,5	0,3	0,3	1mil	1,8GB	0,666	0,213	0,13	0,657	0,375	0,408
			0,001	0,75	0,35	0,3	2mil	1,8GB	0,717	0,31	0,202	0,332	0,435	0,399
			0,0001	0,5	0,3	0,3	9mil	1,8GB	0,647	0,218	0,256	0,755	0,36	0,447
		0,0001	0,75	0,35	0,3	5mil	1,8GB	0,653	0,176	0,138	0,612	0,464	0,409	
		0,001	0,5	0,3	0,3	1mil	3,1GB	0,512	0,278	0,543	0,354	0,389	0,389	
		0,001	0,75	0,35	0,3	3mil	3,1GB	0,576	0,534	0,196	0,42	0,391	0,424	
	VGG	1024 × 1024	0,0001	0,5	0,3	0,3	4mil	3,1GB	0,392	0,473	0,066	0,658	0,299	0,377
			0,0001	0,75	0,35	0,3	7mil	3,1GB	0,643	0,228	0,233	0,628	0,502	0,447
			0,001	0,5	0,3	0,3	2mil	5,1GB	0,495	0,417	0,209	0,719	0,501	0,468
		0,001	0,75	0,35	0,3	3mil	5,1GB	0,538	0,265	0,212	0,647	0,401	0,413	
		0,0001	0,5	0,3	0,3	4mil	5,1GB	0,538	0,077	0,288	0,598	0,544	0,409	
		0,0001	0,75	0,35	0,3	6mil	5,1GB	0,498	0,255	0,287	0,34	0,42	0,36	
3 Canais	ZF	512 × 512	0,001	0,5	0,3	0,3	3mil	10,4GB	0,536	0,228	0,087	0,66	0,455	0,393
			0,001	0,75	0,35	0,3	3mil	10,4GB	0,538	0,265	0,212	0,647	0,401	0,413
			0,0001	0,5	0,3	0,3	3mil	10,4GB	0,521	0,45	0,105	0,563	0,429	0,414
		0,0001	0,75	0,35	0,3	7mil	10,4GB	0,407	0,239	0,376	0,696	0,383	0,42	
		0,001	0,5	0,3	0,3	3mil	2GB	0,656	0,382	0,092	0,277	0,392	0,36	
		0,001	0,75	0,35	0,3	4mil	2GB	0,445	0,279	0,242	0,46	0,476	0,38	
	GoogleNet	512 × 512	0,0001	0,5	0,3	0,3	7mil	2GB	0,576	0,459	0,152	0,434	0,371	0,398
			0,0001	0,75	0,35	0,3	8mil	2GB	0,68	0,394	0,039	0,409	0,43	0,39
			0,001	0,5	0,3	0,3	4mil	3,6GB	0,479	0,178	0,3	0,544	0,318	0,364
		0,001	0,75	0,35	0,3	5mil	3,6GB	0,631	0,705	0,174	0,559	0,405	0,495	
		0,0001	0,5	0,3	0,3	5mil	3,6GB	0,454	0,554	0,103	0,635	0,453	0,44	
		0,0001	0,75	0,35	0,3	5mil	3,6GB	0,559	0,723	0,207	0,745	0,463	0,539	
VGG	512 × 512	0,001	0,5	0,3	0,3	3mil	1,8GB	0,436	0,058	0,113	0,503	0,285	0,279	
		0,001	0,75	0,35	0,3	1mil	1,8GB	0,637	0,032	0,165	0,554	0,344	0,346	
		0,0001	0,5	0,3	0,3	7mil	1,8GB	0,505	0,357	0,135	0,548	0,334	0,376	
		0,0001	0,75	0,35	0,3	8mil	1,8GB	0,63	0,189	0,233	0,508	0,342	0,381	
		0,001	0,5	0,3	0,3	2mil	3,1GB	0,637	0	0,052	0,41	0,251	0,27	
		0,001	0,75	0,35	0,3	4mil	3,1GB	0,675	0,308	0,087	0,41	0,37	0,37	
	GoogleNet	1024 × 1024	0,0001	0,5	0,3	0,3	6mil	3,1GB	0,464	0,378	0,232	0,441	0,399	0,383
			0,0001	0,75	0,35	0,3	9mil	3,1GB	0,484	0,421	0,214	0,461	0,349	0,3858
			0,001	0,5	0,3	0,3	3mil	5,1GB	0,446	0,033	0,039	0,631	0,255	0,281
		0,001	0,75	0,35	0,3	2mil	5,1GB	0,572	0,139	0,2	0,689	0,426	0,405	
		0,0001	0,5	0,3	0,3	4mil	5,1GB	0,568	0,103	0,197	0,57	0,182	0,324	
		0,0001	0,75	0,35	0,3	5mil	5,1GB	0,533	0,327	0,286	0,744	0,328	0,444	
VGG	512 × 512	0,001	0,5	0,3	0,3	1mil	10,4GB	0,068	0	0,057	0,065	0,207	0,08	
		0,001	0,75	0,35	0,3	2mil	10,4GB	0,231	0	0,072	0,25	0,225	0,156	
		0,0001	0,5	0,3	0,3	7mil	10,4GB	0,49	0,345	0,282	0,451	0,482	0,41	
		0,0001	0,75	0,35	0,3	6mil	10,4GB	0,475	0,428	0,287	0,685	0,417	0,458	
		0,001	0,5	0,3	0,3	1mil	2GB	0,478	0,015	0,044	0,273	0,331	0,228	
		0,001	0,75	0,35	0,3	1mil	2GB	0,257	0,001	0,12	0,497	0,195	0,214	
	GoogleNet	512 × 512	0,0001	0,5	0,3	0,3	2mil	2GB	0,478	0,015	0,044	0,273	0,331	0,228
			0,0001	0,75	0,35	0,3	3mil	2GB	0,257	0,001	0,12	0,497	0,195	0,214
			0,001	0,5	0,3	0,3	1mil	3,6GB	0,784	0,311	0,051	0,512	0,361	0,404
		0,001	0,75	0,35	0,3	5mil	3,6GB	0,508	0,546	0,136	0,843	0,546	0,516	
		0,0001	0,5	0,3	0,3	8mil	3,6GB	0,493	0,568	0,228	0,735	0,414	0,488	
		0,0001	0,75	0,35	0,3	10mil	3,6GB	0,389	0,255	0,24	0,954	0,574	0,482	

TABELA 20 – CONFIGURAÇÕES TESTADAS NA FASTER R-CNN PARA ANOTAÇÃO POR BLOCOS DE ROIS. FONTE: O autor (2018).

Imagem	Rede	Tamanho Imagem	Taxa de Aprendizagem	IoU Positivo	IoU Negativo	NMS	Uso GPU	Épocas	N	E	GG	F	M	mAP Teste
1 Canal	ZF		0,001	0.5	0.3	0.3	1,8GB	3mil	0,157	0,111	0,138	0,347	0,15	0,186
			0,001	0.75	0.35	0.3	1,8GB	5mil	0,2	0,121	0,143	0,389	0,144	0,199
			0,0001	0.5	0.3	0.3	1,8GB	1mil	0,253	0,102	0,075	0,455	0,117	0,2
			0,0001	0.75	0.35	0.3	1,8GB	3mil	0,183	0,125	0,163	0,432	0,172	0,215
			0,0001	0.75	0.35	0.5	1,8GB	5mil	0,184	0,135	0,092	0,549	0,202	0,232
			0,001	0.5	0.3	0.3	5,1GB	2mil	0,194	0,088	0,133	0,356	0,205	0,195
	VGG		0,001	0.75	0.35	0.3	5,1GB	5mil	0,232	0,014	0,045	0,518	0,208	0,203
			0,0001	0.5	0.3	0.3	5,1GB	4mil	0,157	0,168	0,187	0,242	0,214	0,193
			0,0001	0.75	0.35	0.3	5,1GB	3mil	0,214	0,154	0,113	0,535	0,227	0,249
			0,0001	0.75	0.35	0.5	5,1GB	9mil	0,229	0,087	0,154	0,572	0,362	0,281
			0,001	0.5	0.3	0.3	2GB	2mil	0,215	0,136	0,023	0,375	0,189	0,188
			0,0001	0.75	0.35	0.3	2GB	4mil	0,155	0,108	0,091	0,475	0,243	0,214
3 Canais	ZF	512	0,0001	0.5	0.3	0.3	2GB	4mil	0,113	0,043	0,149	0,504	0,308	0,223
			0,0001	0.75	0.35	0.3	2GB	8mil	0,158	0,19	0,128	0,406	0,25	0,227
			0,0001	0.75	0.35	0.5	2GB	10mil	0,201	0,082	0,124	0,555	0,256	0,243
			0,001	0.5	0.3	0.3	1,8GB	5mil	0,204	0,061	0,069	0,355	0,173	0,172
			0,0001	0.75	0.35	0.3	1,8GB	6mil	0,162	0,161	0,037	0,361	0,199	0,184
			0,0001	0.5	0.3	0.3	1,8GB	6mil	0,116	0,168	0,059	0,354	0,186	0,176
	VGG		0,0001	0.75	0.35	0.3	1,8GB	5mil	0,124	0,184	0,154	0,461	0,204	0,225
			0,0001	0.75	0.35	0.5	1,8GB	4mil	0,292	0,171	0,103	0,493	0,153	0,243
			0,001	0.5	0.3	0.3	5,1GB	4mil	0,09	0,196	0,132	0,357	0,156	0,186
			0,001	0.75	0.35	0.3	5,1GB	5mil	0,194	0,146	0,127	0,348	0,203	0,203
			0,0001	0.5	0.3	0.3	5,1GB	4mil	0,228	0,182	0,167	0,347	0,175	0,22
			0,0001	0.75	0.35	0.3	5,1GB	7mil	0,114	0,178	0,183	0,351	0,203	0,206
GoogleNet		0,0001	0.75	0.35	0.5	5,1GB	9mil	0,131	0,192	0,202	0,528	0,24	0,259	
		0,001	0.5	0.3	0.3	2GB	5mil	0,276	0,191	0,12	0,446	0,259	0,259	
		0,001	0.75	0.35	0.3	2GB	6mil	0,194	0,154	0,203	0,37	0,191	0,222	
		0,0001	0.5	0.3	0.3	2GB	9mil	0,248	0,127	0,166	0,297	0,226	0,213	
		0,0001	0.75	0.35	0.3	2GB	7mil	0,27	0,197	0,143	0,395	0,163	0,234	
		0,0001	0.75	0.35	0.5	2GB	9mil	0,262	0,173	0,196	0,421	0,257	0,262	

Para avaliação utilizou-se a curva de precisão/*recall* que destaca a precisão do detector em diferentes níveis de *recall*. Em um sistema ideal, a precisão é 1 em todos os níveis de *recall*. O maior valor de mAP indica robustez e efetividade da Faster R-CNN para localização de objetos.

A curva de precisão/*recall* foi gerada executando o detector treinado sobre as amostras de teste para os melhores resultados quantitativos obtidos pelo modelo. Na Figura 37 é apresentada a curva para o melhor resultado utilizando imagens em escala de cinza para a configuração de anotações por ROI inteira.

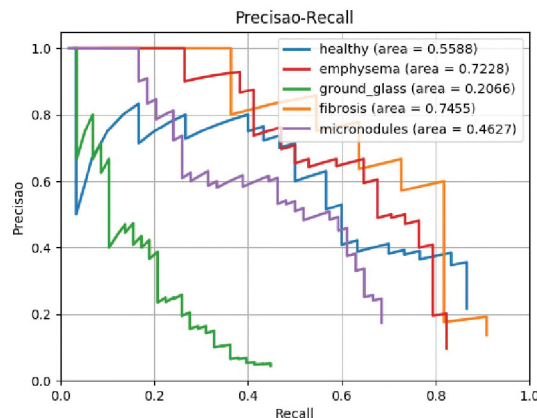


FIGURA 37 – MELHOR CURVA DE PRECISÃO-RECALL UTILIZANDO A FASTER R-CNN. FONTE: O autor (2018).

Alguns exemplos de saída da Faster R-CNN podem ser observados na Figura 38, onde, cada imagem contém o *bounding-box* que representa a ROI original e o *bounding-box* gerado pelo detector Faster R-CNN.

4.4.5 Informações de Processamento

Todos os experimentos listados neste Capítulo foram executados na GPU (Unidades de Processamento Gráfico) NVIDIA TITAN Xp (Pascal) - 12GB (3840 CUDA cores) que é eficiente para a computação paralela e que tem uma ordem de magnitude com mais *threads* de execução do que unidades de processamento central (CPUs). Por esse motivo, a aprendizagem profunda em GPUs é tipicamente 10 a 30 vezes mais rápida do que em CPUs (LITJENS et al., 2017). O software utilizado foi o Framework Caffe (BERKELEY, 2014), desenvolvido e disponibilizado pela Universidade da Califórnia, que fornece interfaces em C++ e Python além de bibliotecas com implementações eficientes de GPU para processamento de redes neurais (JIA et al., 2014).

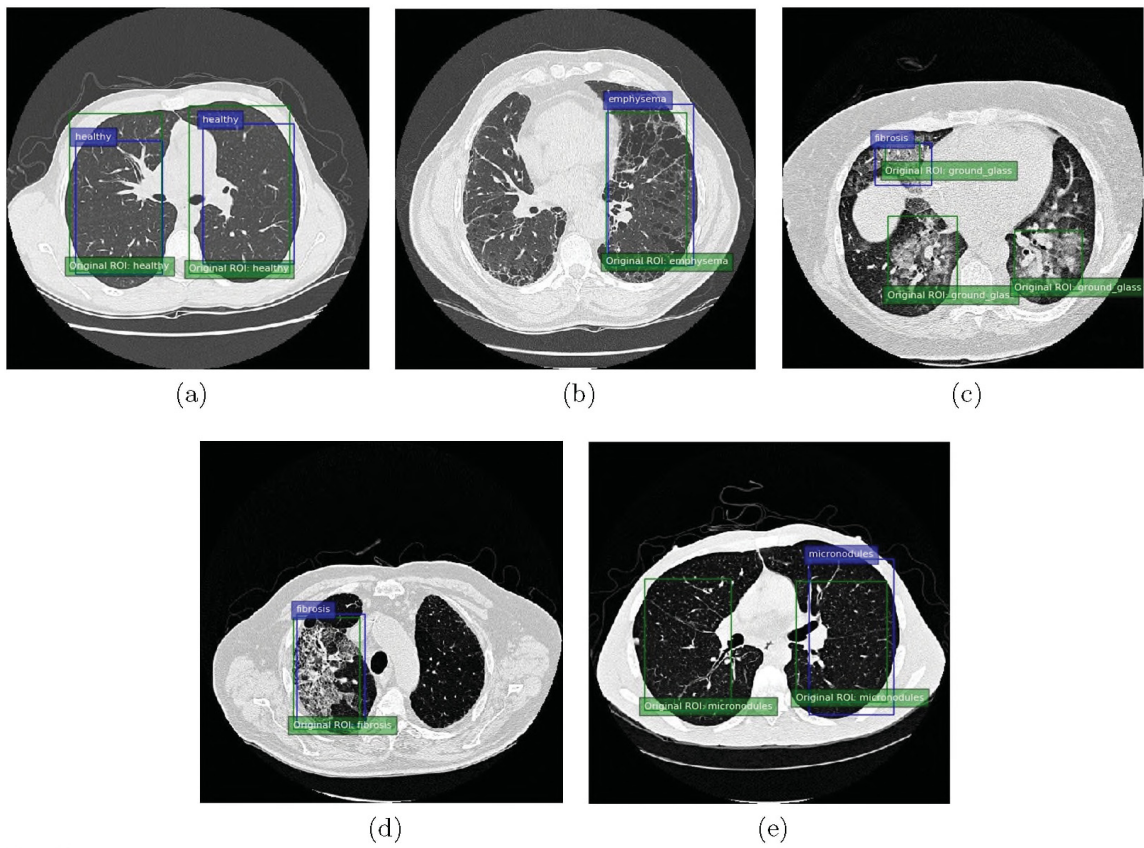


FIGURA 38 – EXEMPLOS DE DETECÇÃO COM A FASTER R-CNN: (A) NORMAL; (B) ENFISEMA; (C) VIDRO-FOSCO; (D) FIBROSE; (E) MICRONÓDULOS. FONTE: O autor (2018).

5 DISCUSSÃO

Os experimentos foram conduzidos de acordo com as metodologias apresentadas no capítulo anterior. Como previamente descrito, a primeira etapa consiste em classificar os blocos de DPIS com o intuito de avaliar a viabilidade da utilização da técnica de transferência de aprendizagem, assim como, o efeito da produção de dados sintéticos quando aplicadas em sistemas baseados em redes neurais convolucionais.

Como resultado inicial, o teste utilizando uma rede própria não treinada e a base de dados balanceada, obteve uma acurácia de 73,40% na classificação de blocos de imagens de DPIS. Na sequência nos resultados utilizando as redes existentes AlexNet, ImageNet e GoogleNet, e a base de dados balanceada, a acurácia foi de 77,92%, 85,40% e 86,00%, respectivamente. Ao aplicar a técnica de transferência de aprendizagem nas redes AlexNet, ImageNet e GoogleNet a precisão obtida foi de 75,8%, 87,32%, 90,95%, para cada rede.

No comparativo do impacto do uso da técnica de criação de dados sintéticos observou-se que para a maioria das arquiteturas testadas houve um aumento significativo de aproximadamente 6% na acurácia quando feito o balanceamento da base de dados. Em relação a técnica de transferência de aprendizagem, também houve um aumento de aproximadamente 5% para a rede GoogleNet.

Em comparação a literatura, o melhor resultado obtido na classificação de blocos de TCAR se aproximou muito do resultado de 91.2% de acurácia na classificação de seis classes da mesma base de imagens que havia sido apresentado pelo autor (LU et al., 2016), o qual, também utiliza transferência de aprendizagem mas com a rede ImageNet em vez da GoogleNet.

Em uma análise geral do desempenho da técnica de transferência de aprendizagem para o reconhecimento de DPIS observou-se que a metodologia é promissora pois, não é necessária a criação de uma CNN profunda do zero para um problema específico. A criação de uma CNN é uma tarefa não trivial, uma vez que um grande número de valores de parâmetros mutuamente dependentes e escolhas algorítmicas devem ser definidas. Além disso, pode-se destacar a importância da inicialização de parâmetros mesmo que os pesos tenham sido aprendidos de um contexto diferente do contexto de imagens médicas.

O resultado superior obtido para arquiteturas de redes já existentes e pré-treinadas, justifica a utilização da transferência de aprendizagem e do aumento de dados em conjunto com outras técnicas baseadas em redes neurais convolucionais.

Tendo em vista os resultados anteriores e dando sequência ao desenvolvimento deste trabalho, a segunda etapa, propôs detectar e classificar os padrões radiológicos diretamente na imagem de TCAR com o algoritmo Faster R-CNN juntamente com a transferência

de aprendizagem e a geração de dados sintéticos. A ideia foi suprir a necessidade de pré-segmentação de regiões de interesse e automatizando o processo de caracterização de doenças pulmonares.

Como resultado da segunda etapa, o teste inicial para verificação da eficácia da metodologia Faster R-CNN na detecção e classificação de imagens médicas, apresentou um mAP de 87.5%, 91.1% e 90.6% para as redes ZF, VGG e GoogleNet no problema binário de segmentação do pulmão. Deve-se notar que na classificação binária, o teste foi realizado utilizando somente as classes “pulmão” e “não-pulmão”, assim independente da presença de patologia nas imagens de TCAR a Faster R-CNN detectou e classificou regiões do pulmão corretamente na maioria das imagens.

Em relação ao desempenho da Faster R-CNN no contexto de classificação das cinco classes de patologias, o melhor resultado obtido foi utilizando anotações de ROIs inteiras e a rede GoogleNet, com um mAP de 53.9%, onde as classes enfisema e fibrose foram detectadas com um AP de 72.3% e 74.5%. Quando alteradas as imagens para escala RGB e ainda utilizando marcações de ROIs inteiras, o AP da classe Fibrose aumentou para 95.4%, o AP da classe pulmão melhorou para 78.4% e também houve um ligeiro aumento da classe micronódulos de 46.3% para 57.4%, porém, na média dos resultados não houve melhora. Para a configuração por blocos de ROIs, a Faster R-CNN não exibe resultados satisfatórios, com um valor de mAP de no máximo 28.1%.

No resultado individual geral por classe, os melhores valores de AP em diferentes configurações foram de 78.4%, 72.3%, 37.6%, 95.4% e 57.4% para as classes normal, enfisema, vidro-fosco, fibrose e micronódulos, respectivamente. No geral, a classe fibrose apresenta o maior resultado em diferentes configurações devido ao grande número de dados de treinamento disponíveis dessa classe em comparação aos outros padrões.

Uma fato relevante no uso da Faster R-CNN para detecção de DPIs, é que tanto para anotações de ROIs inteiras como para blocos, são selecionadas apenas regiões que pertencem ao pulmão como propostas de regiões para classificação de DPIs, ou seja, o algoritmo não faz confusões com regiões que não estejam dentro da área pertencente ao pulmão. Quanto as redes utilizadas, em geral, a detecção com as redes VGG e GoogleNet apresentaram resultados próximos, porém o uso da arquitetura da GoogleNet apresentou-se mais viável pelo fato da execução da arquitetura VGG demandar um alto custo computacional que é de aproximadamente três vezes mais que a GoogleNet.

5.1 CONSIDERAÇÕES FINAIS

Devido à falta de imagens médicas anotadas em grande escala, o treinamento de redes neurais profundas pode ser muito desafiador. Para superar esse problema, mostrou-se que, ao ajustar uma rede profunda pré-treinada em um conjunto de dados médicos, pode-se

alcançar resultados satisfatórios para a classificação de blocos de padrões radiológicos.

Na tarefa de detecção e classificação de padrões pulmonares, observou-se que o modelo Faster R-CNN apresenta bons resultados para as classes enfisema e fibrose, porém não identifica todos os padrões de doenças pulmonares em imagens inteiras de TCAR com um AP aceitável. A baixa performance do método pode ser justificada pelo fato do banco de dados (DEPEURSINGE et al., 2011) ser parcialmente rotulado conforme também mencionado pelo autor (AGARWALA et al., 2018), pois, no momento da rotulação, se houve uma contradição na opinião dos radiologistas para alguma região em particular, essa região não foi rotulada e permanece sem nenhuma marcação.

Além disso, conforme mencionado por (REN et al., 2016) a Faster R-CNN possui uma fraca capacidade de identificar textura local, supondo assim, que esta metodologia não aprende corretamente as classes de DPIS pois este algoritmo trabalha melhor com problemas de objetos que têm um contorno bem definido, como por exemplo, na detecção do pulmão.

Em ambas as etapas deste trabalho foi possível levantar a hipótese que as confusões de classificação se deram pela pouca quantidade de dados em comparação com a quantidade de amostras normalmente utilizadas nos trabalhos de redes neurais. Além disso, a aparência similar dos tipos de padrões radiológicos também pode gerar confusões na detecção e classificação da rede neural convolucional.

Ainda, como contribuição final do presente trabalho, destaca-se a adaptação da rede GoogleNet para a Faster R-CNN, arquitetura, a qual, não era implementada na literatura existente e que apresenta ganhos significativos no custo de processamento.

REFERÊNCIAS

- AGARWALA, S. et al. Convolutional neural networks for efficient localization of interstitial lung disease patterns in hrct images. v. 2, p. 1–10, 2018. Citado 2 vezes nas páginas 46 e 69.
- ANTHIMOPOULOS, M. et al. Classification of interstitial lung disease patterns using local dct features and random forest. *Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, p. 6040–6043, 2014. Citado na página 51.
- ANTHIMOPOULOS, M. et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, v. 35, n. 5, 2016. Citado 2 vezes nas páginas 26 e 27.
- ARCHONTOGEOGRIS, K. et al. Lung cancer and interstitial lung diseases: A systematic review. *Pulmonary Medicine*, v. 2012, 2012. Citado na página 13.
- AZEVEDO-MARQUES, P. M. Diagnóstico auxiliado por computador na radiologia. 2001. Citado na página 16.
- BAGCI, U. et al. Computer-assisted detection of infectious lung diseases: A review. *Computerized Medical Imaging and Graphics*, v. 36, p. 72–84, 2011. Citado 2 vezes nas páginas 14 e 16.
- BAGESTEIRO, L. D.; OLIVEIRA, L. F.; WEINGAERTNER, D. Blockwise classification of lung patterns in unsegmented ct images. *IEEE Symposium on Computer-Based Medical Systems*, v. 2015-July, p. 177–182, 2015. Citado 2 vezes nas páginas 49 e 59.
- BARBOSA, I. R. et al. Cancer mortality in brazil. *Medicine*, v. 94, n. 16, p. 1–6, 2015. Citado na página 14.
- BERKELEY, A. I. R. *Caffe - Deep learning framework by BAIR*. [S.l.], 2014. Disponível em: <<http://caffe.berkeleyvision.org/>>. Acesso em: 01/05/2017. Citado na página 65.
- DEPEURSINGE, A.; VARGAS, A.; AL., A. P. et. Building a reference multimedia database for interstitial lung diseases. 2011. Citado 7 vezes nas páginas 15, 18, 42, 48, 56, 58 e 69.
- EADIE, L. H.; TAYLOR, P.; GIBSON, A. P. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *European Journal of Radiology*, Elsevier Ireland Ltd, v. 81, n. 1, p. e70–e76, 2012. Citado na página 16.
- EVERINGHAM, M.; WINN, J. *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Development Kit*. [S.l.], 2007. v. 2007, 1–23 p. Citado 2 vezes nas páginas 32 e 46.
- GAO, M. et al. Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Workshop on Deep Learning in Medical Image Analysis (MICCAI)*, 2015. Citado 3 vezes nas páginas 13, 51 e 56.
- GIRSHICK, R. Fast r-cnn. *IEEE International Conference on Computer Vision (ICCV)*, p. 1440–1448, 2015. Citado 5 vezes nas páginas 28, 29, 30, 31 e 62.

- HAFEMANN, L. G.; OLIVEIRA, L. S.; CAVALIN, P. R. An analysis of deep neural networks for texture classification. 2014. Citado na página 52.
- JIA, Y. et al. *Caffe: Convolutional Architecture for Fast Feature Embedding*. [S.l.], 2014. 675–678 p. Disponível em: <<http://arxiv.org/abs/1408.5093>>. Acesso em: 08/02/2018. Citado na página 65.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, p. 1–9, 2012. Citado 6 vezes nas páginas 13, 19, 27, 28, 50 e 56.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 1, p. 436–444, 2015. Citado 8 vezes nas páginas 13, 19, 20, 21, 22, 23, 24 e 25.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *IEEE*, v. 86, n. 11, p. 2278–2323, 1998. Citado 2 vezes nas páginas 21 e 22.
- LI, Q. et al. Medical image classification with convolutional neural network. *13th International Conference on Control, Automation, Robotics, and Vision*, IEEE, v. 2014, n. December, p. 844–848, 2014. Citado 3 vezes nas páginas 25, 40 e 47.
- LITJENS, G. et al. A survey on deep learning in medical image analysis. n. 1995, 2017. Disponível em: <<http://arxiv.org/abs/1702.05747>>. Acesso em: 01/10/2017. Citado 9 vezes nas páginas 17, 19, 20, 21, 22, 24, 26, 53 e 65.
- LU, J. et al. Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, Elsevier B.V., v. 80, p. 14–23, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.knosys.2015.01.010>>. Acesso em: 29/01/2018. Citado na página 26.
- LU, L. et al. Deep convolutional neural networks for computer-aided detection. *IEEE Transactions on Medical Imaging*, v. 35, n. 5, p. 1285–1298, 2016. Citado 5 vezes nas páginas 13, 26, 27, 51 e 67.
- MALTA, D. C. et al. Trends in corrected lung cancer mortality rates in brazil an regions. *Revista de saúde pública*, 2016. Citado 2 vezes nas páginas 13 e 14.
- PEEMEN, M.; CORPORAAL, B.; HENK. Speed sing detection and recognition by convolutional neural networks. 2011. Citado na página 23.
- REDMON, J. et al. You only look once: Unified, real-time object detection. 2016. Citado 4 vezes nas páginas 34, 46, 61 e 62.
- REIAZI, R. et al. Mammography lesion detection using faster r-cnn detector. *Computer Science and Information Technology*, n. February, p. 111–115, 2018. Citado na página 62.
- REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. 2016. Citado 6 vezes nas páginas 30, 31, 55, 60, 62 e 69.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 521, 1958. Citado 2 vezes nas páginas 19 e 22.

- SCHMIDHUBER, J. Deep learning in neural networks: An overview. 2015. Citado 4 vezes nas páginas 13, 19, 20 e 21.
- SHAFII, N. Z. et al. Performance evaluation of faster r-cnn on gpu for object detection. 2017. Citado 4 vezes nas páginas 18, 28, 29 e 30.
- SILVA, I. N. de Cancer José Alencar Gomes da. Inca - instituto nacional de câncer - estimativa 2016. p. 124, 2016. Citado 2 vezes nas páginas 13 e 14.
- SUN, X.; WU, P.; HOI, S. C. Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing*, v. 299, p. 42–50, 2018. Citado na página 62.
- TEICHMANN, M. et al. Multinet: Real-time joint semantic reasoning for autonomous driving. 2016. Disponível em: <<http://arxiv.org/abs/1612.07695>>. Acesso em: 22/11/2017. Citado na página 28.
- VISA B. RAMSAY, A. R. e. a. S. Confusion matrix-based feature selection. 2011. Citado na página 33.
- WELLS, A. U.; HIRANI, N. Interstitial lung disease guideline. *British Thoracic Society*, v. 63, p. 1–58, 2008. Citado 4 vezes nas páginas 14, 15, 16 e 49.
- YU, J. S. T. S. Label image regions. 2013. Disponível em: <http://scikit-image.org/docs/dev/auto_examples/segmentation/plot_label.html>. Acesso em: 14/11/2017. Citado na página 58.