

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

GUSTAVO DE AQUINO E AQUINO

AVALIANDO O DESEMPENHO DE REDES NEURAS CONVOLUCIONAIS COM
ARQUITETURAS DE GRAFOS ACÍCLICOS DIRETOS E SEQUENCIAIS NA
SEGMENTAÇÃO AUTOMÁTICA DE LESÕES MAMÁRIAS EM IMAGENS DE
ULTRASSOM

MANAUS

2020

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

GUSTAVO DE AQUINO E AQUINO

AVALIANDO O DESEMPENHO DE REDES NEURAIAS CONVOLUCIONAIS COM
ARQUITETURAS DE GRAFOS ACÍCLICOS DIRETOS E SEQUENCIAIS NA
SEGMENTAÇÃO AUTOMÁTICA DE LESÕES MAMÁRIAS EM IMAGENS DE
ULTRASSOM

Dissertação apresentada ao Curso de Mestrado em Engenharia Elétrica, área de concentração Controle e Automação de Sistemas e linha de pesquisa Reconhecimento de Padrões e Otimização do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas.

Orientadora: Prof^a. Dr^a. Marly Guimarães Fernandes Costa
Coorientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

MANAUS

2020

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

E999a Aquino, Gustavo de Aquino e
Avaliando o desempenho de redes neurais convolucionais com
arquitecturas de grafos acíclicos diretos e sequenciais na
segmentação automática de lesões mamárias em imagens de
ultrassom / Gustavo de Aquino e Aquino. 2020
88 f.: il. color; 31 cm.

Orientadora: Marly Guimarães Fernandes Costa
Coorientador: Cícero Ferreira Fernandes Costa Filho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade
Federal do Amazonas.

1. Redes Neurais Convolucionais. 2. Segmentação Automática de
Lesões. 3. Câncer de Mama. 4. Redes DAG. I. Costa, Marly
Guimarães Fernandes II. Universidade Federal do Amazonas III.
Título

GUSTAVO DE AQUINO E AQUINO

**AVALIANDO O DESEMPENHO DE REDES NEURAIS CONVOLUCIONAIS
COM ARQUITETURAS DE GRAFOS ACÍCLICOS DIRETOS E SEQUENCIAIS
NA SEGMENTAÇÃO AUTOMÁTICA DE LESÕES MAMÁRIAS EM IMAGENS
DE ULTRASSOM**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 03 de março de 2020.

BANCA EXAMINADORA


Prof^a. Dra. Marly Guimarães Fernandes Costa, Presidente

Universidade Federal do Amazonas


Prof. Dr. José Raimundo Gomes Pereira, Membro

Universidade Federal do Amazonas


Prof. Dr. Jozias Parente de Oliveira, Membro

Universidade do Estado do Amazonas

AGRADECIMENTOS

Agradeço aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e Prof^a. Dr^a. Marly Guimarães Fernandes Costa, pela orientação dada, pelo aprendizado adquirido e, principalmente, pela confiança em mim depositada.

A minha família, pelo apoio desde o início da minha vida estudantil e em todas as fases da minha vida.

A Kimberly, por todo o apoio mesmo nos momentos difíceis e por estar sempre ao meu lado.

A Universidade Federal do Amazonas e, em especial, ao Centro de Tecnologia Eletrônica e da Informação – CETELI – pelo provimento de toda infraestrutura para realização deste trabalho.

A Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), pela concessão de bolsa que permitiu dedicação exclusiva aos estudos.

Esta pesquisa, conforme previsto no Art. 48 do decreto nº 6.008/2006, foi financiada pela Samsung Eletrônica da Amazônia Ltda, nos termos da Lei Federal nº 8.387/1991, através de convênio nº 004, firmado com o Centro de P&D em Eletrônica e Tecnologia da Informação da Universidade Federal do Amazonas - CETELI / UFAM.

RESUMO

O câncer de mama é um problema mundial responsável por um número substancial de mortes. O diagnóstico suportado pela análise de imagem da lesão tem se mostrado eficiente, notadamente, através do uso de técnicas de aprendizado de máquinas. O sucesso alcançado em anos recentes cresceu devido ao uso das redes convolucionais, capazes de fazer tarefas de visão computacional com sucesso, como a segmentação automática de lesões nas mais variadas modalidades de imagens biomédicas. A tarefa de segmentação suporta etapas posteriores de um sistema computadorizado de auxílio a diagnóstico por imagem. Nessa dissertação, avalia-se o desempenho de redes neurais convolucionais com arquiteturas sequenciais e de grafos acíclicos diretos na segmentação automática de lesões mamárias em imagens de ultrassonografia. Foram implementadas e testadas quatro arquiteturas de redes convolucionais, três delas com grafos acíclicos diretos (DAG) e uma sequencial. Para o desenvolvimento e avaliação das propostas, dois bancos de imagens de ultrassom de mama (banco A e banco B) foram utilizados. Algumas diferenças marcantes entre esses referidos bancos são a dimensão, resolução e qualidade das imagens. Assim as imagens foram previamente processadas e adaptadas (*cropping* e *resize*). Os treinamentos dessas redes com parada pelo número de épocas se mostraram instáveis. Problema este que foi contornado com a proposição de uma função de auxílio ao treinamento, que permitiu obter o ponto de melhor desempenho do modelo. A melhor arquitetura foi escolhida com base em métricas já consagradas na literatura, como Acurácia global e coeficiente Dice. As quatro arquiteturas testadas alcançaram resultados similares, com acurácias globais de mais de 94% cada. O teste de significância estatística *t-student* apontou que para ambos os bancos de dados a melhor arquitetura de rede na validação foi a U-net, alcançando mais de 99% de acurácia global para o banco de dados B e mais de 96% para o banco de dados A. A rede apontada com melhor desempenho pôde ser testada, com outros dados de entrada, e seu desempenho se manteve o mesmo. Foi possível concluir que o procedimento de *cropping* não se mostrou crucial para uma boa precisão na segmentação. Além disso, a análise estatística das métricas de desempenho mostrou que o uso de imagens de melhor resolução (banco B) não acarretou diferenças, estatisticamente significantes ($p < 0.05$), de desempenho.

Palavras chave: Redes Neurais Convolucionais, Segmentação Automática de Lesões, Câncer de Mama, Redes DAG.

ABSTRACT

Breast cancer can be seen as a worldwide problem, which is responsible for a substantial number of deaths. Diagnosis through image analysis of the lesion is efficient, notably through the use of machine learning techniques. The success achieved in recent years has grown due to the use of convolutional networks. This method is capable of successfully performing computer vision tasks, such as the automatic segmentation of lesions in the most varied modalities of biomedical images. This stage, segmentation, supports later stages of a computer-aided diagnostic imaging system. In this dissertation, the performance of convolutional neural networks with direct acyclic graph architectures in the automatic segmentation of breast lesions in ultrasound images is evaluated. Four convolutional network architectures were implemented and tested, three of them with direct acyclic graphs (DAG) and one sequential. For the development and evaluation of the proposals, two banks of breast ultrasound images (bank A and bank B) were used. Some striking differences between these banks are the size, resolution, and quality of the images. Thus, the images were previously processed and adapted (cropping and resize). The training of these networks with a stop for the number of seasons proved to be unstable. This problem was overcome with the proposal of a training aid function, which allowed us to obtain the best performance point of the model. The best architecture was chosen based on metrics already established in the literature, such as global accuracy and the Dice coefficient. The four architectures tested achieved similar results, with a global accuracy of more than 94% each. The t-student statistical significance test showed that for both databases the best network architecture in the validation was U-net, reaching over 99% of global accuracy for database B and over 96% for the database A. The network with the best performance could be tested, with other input data, and its performance remained the same. It was possible to conclude that the cropping procedure was not crucial for good segmentation accuracy. Besides, the statistical analysis of the performance metrics showed that the use of better resolution images (bank B) did not cause statistically significant performance differences ($p < 0.05$).

Keywords: Convolutional Neural Networks, Automatic Lesions Segmentation, Breast Cancer, DAG Networks.

LISTA DE FIGURAS

Figura 1: Imagens de ultrassom com lesões mamárias. Observe as variações na forma da lesão, a baixa relação sinal ruído, o baixo contraste, entre outros. Fonte: (YAP et al., 2018)	16
Figura 2: Topologia de uma rede neural com múltiplas camadas. Fonte: (PATTERSON; GIBSON, 2008)	36
Figura 3: Neurônio biológico e neurônio artificial. Fonte: (MWANDAU; NYANCHAMA, 2018)	37
Figura 4: Funções de treinamento das Redes Neurais Profundas (RNPs). Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)	38
Figura 5: Exemplo de uma RNP. Fonte: Adaptado de Awad e Khana (2015)	39
Figura 6: Arquitetura LeNet-5. Fonte: (LeCun et al., 1998b)	41
Figura 7: Representação das camadas de uma RNC. Fonte: Adaptado de Goodfellow e contribuintes (2016)	42
Figura 8: Comparativo do comportamento do cérebro humano e as RNCs. Fonte: (ZAHARCHUK et al., 2018)	43
Figura 9: Exemplo de convolução bidimensional, em que a saída é obtida sem padding. Observe que as dimensões da entrada são de 3x4, enquanto que a saída tem dimensões 2x3. Fonte: Adaptado de Goodfellow et al. (2016)	44
Figura 10: Resultado da aplicação max pooling. Observe que foi utilizado um passo (S) de 2 e um filtro (2,2), obtendo uma imagem de metade do tamanho da original.	45
Figura 11: Exemplo da ação da camada de dropout. A rede da direita é o resultado da aplicação do dropout na rede da esquerda. Fonte: (SRIVASTAVA et al., 2014)	47
Figura 12: Convolução transposta da imagem I2 pelo kernel K.	48
Figura 13: Exemplo de uma RNC para segmentação semântica. Fonte: (LONG; SHELHAMER; DARRELL, 2015)	49
Figura 14: Processo de segmentação semântica. Fonte: (MATHWORKS, 2018)	49
Figura 15: Efeito do momento na suavização de oscilações: (a) método SGD (b) método SGDM. Fonte: (AGGARWAL, 2018)	52
Figura 16: Comparação do ADAM com outros métodos de otimização. Fonte: (KINGMA; BA, 2014)	54
Figura 17: Diagrama em blocos da Metodologia utilizada	55
Figura 18: Imagens de US de mama de BD-A	56

Figura 19: Máscaras (padrões ouro) das imagens apresentadas na Figura 18.....	56
Figura 20: (a) Exemplo de uma imagem cropped do Dataset A e (b) Exemplo de uma imagem original do Dataset B.....	57
Figura 21: Adaptações para obtenção dos subconjuntos de dados do Dataset A e Dataset B..	57
Figura 22: Divisão 60-20-20 para treino, validação e testes.	58
Figura 23: Arquitetura 1. Fonte: (MENDES et al., 2016).....	60
Figura 24: Arquitetura 2. Fonte: (MENDES et al., 2016).....	61
Figura 25: Arquitetura 3. Fonte: (MENDES et al., 2016).....	62
Figura 26: Arquitetura 4 proposta	63
Figura 27: Diagrama dos experimentos realizados com cada arquitetura, de verde o que foi adotado no modelo proposto	63
Figura 28. Descrição da Função de auxílio ao treinamento proposta.....	64
Figura 29: Diagrama de Veen ilustrando os conceitos de True Positive (TP), True Negative (TN), False Positive (FP) e False Negative (FN).	65
Figura 30: Gráfico de convergência da rede CNN1 na validação, usando o conjunto de dados BD-A	67
Figura 31: Gráfico de convergência da rede CNN2 na validação, usando o conjunto de dados BD-A	68
Figura 32: Gráfico de convergência da rede CNN3 na validação, usando o conjunto de dados BD-A	68
Figura 33: Gráfico de convergência da rede CNN4 na validação, usando o conjunto de dados BD-A	68
Figura 34: Acurácia de treinamento das 4 arquiteturas com BD-A	69
Figura 35: Acurácia de validação das 4 arquiteturas com BD-A	69
Figura 36: Gráfico de convergência da rede CNN1 na validação, usando o conjunto de dados BD-B cropped.....	71
Figura 37: Gráfico de convergência da rede CNN2 na validação, usando o conjunto de dados BD-B cropped.....	71
Figura 38: Gráfico de convergência da rede CNN3 na validação, usando o conjunto de dados BD-B cropped.....	72
Figura 39: Gráfico de convergência da rede CNN4 na validação, usando o conjunto de dados BD-B cropped.....	72
Figura 40: Acurácia de treinamento das 4 arquiteturas na validação com BD-B cropped.....	73
Figura 41: Acurácia de validação das 4 arquiteturas na validação com BD-B cropped	73

Figura 42: Acurácia de treinamento das 4 arquiteturas em BD-B sem cropping	75
Figura 43: Acurácia de validação das 4 arquiteturas em BD-B sem cropping.....	75
Figura 44: Acurácia de treinamento das 4 arquiteturas com BD-B sem cropping	77
Figura 45: Acurácia de validação das 4 arquiteturas com BD-B sem cropping.....	77

LISTA DE TABELAS

Tabela 1: Desempenho na validação sem parada antecipada com BD-A. Observe que os valores apontam um melhor desempenho de CNN2.....	70
Tabela 2: Desempenho das métricas com as arquiteturas testadas na etapa de validação com BD-A, utilizando a função de auxílio ao treinamento. Observe que agora a arquitetura 4 é a que tem melhores desempenhos.	70
Tabela 3: Desempenho das métricas com as arquiteturas testadas na validação com BD-B, sem a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN3.	74
Tabela 4: Desempenho na validação com BD-B, com a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN1.	74
Tabela 5: Desempenho na validação com BD-B sem cropping, sem a função de auxílio.	75
Tabela 6: Desempenho na validação com BD-B sem cropping, com a função de auxílio.....	76
Tabela 7: Desempenho na validação com BD-B sem cropping, sem a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN3.....	78
Tabela 8:Desempenho na validação com BD-B sem cropping, com a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN4.....	78
Tabela 9. Resultado do teste t-student entre as demais arquiteturas e CNN4 para o dataset A	79
Tabela 10: Resultado do teste t-student entre as demais arquiteturas e CNN4 para o dataset B	79
Tabela 11: Resultado da validação cruzada com o Dataset A.....	80
Tabela 12: Resultado da validação cruzada com o Dataset B	81

LISTA DE QUADROS

Quadro 1 – Resumo da análise realizada nos trabalhos relacionados.	26
-------------------------------------------------------------------------	----

LISTA DE SIGLAS

AM	Aprendizado de Máquina
ARD	<i>Average Radial Derivative</i>
RS	Razão de sobreposição
INCa	Instituto Nacional do Cancer
ASF	<i>Alternationing Sequential Filter</i>
CD	<i>Coincidende Percentage</i>
CNN	<i>Convolutional Neural Network</i>
DAG	<i>Direct Acyclic Graph</i>
AP	Aprendizado Profundo
IA	Inteligência Artificial
BN	<i>Batch Normalization</i>
OMS	Organização Muldial de Saúde
US	<i>Ultra-sound</i>
H&E	Vírus da Imunodeficiência Humana
RNA	Rede Neural Artificial
RNP	Rede Neural Profunda
RNC	Rede Neural Convolutcional
ROI	<i>Region of Interest</i>
PCNN	<i>Pulse-Coupled Neural Network</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
H&E	Hematoxilina eosina
PEP	Prontuário Eletrônico do Paciente
CAD	<i>Computer Aided Diagnosis</i>
ReLU	<i>Retifield Linear Unit</i>
FP	<i>False Positive</i>
ROC	<i>Receiver Operating Characteristic</i>
FN	<i>False Negative</i>

FCN	<i>Fully Connected Network</i>
SGDM	<i>Stochastic Gradient Descent with Momentum</i>
GPU	<i>Graphics Processing Unit</i>
SVM	<i>Support Vector Machine</i>
RMSProp	<i>Root Mean Square Prop</i>
SGD	<i>Stochastic Gradient Descent</i>
ADAM	<i>Adaptive Moment</i>
MNIST	<i>Modified National Institute of Standards and Technology</i>
RAM	<i>Random Access Memory</i>
CUDA	<i>Compute Unified Device Architecture</i>
MATLAB	<i>Matrix Laboratory</i>
BD-A	Banco de dados A
BD-B	Banco de dados B
GT	<i>Ground True</i>
CT	<i>Computed Tomography</i>

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	Objetivo Geral.....	17
1.2	Objetivos Específicos	17
1.3	Organização do Trabalho	18
2	REVISÃO DA LITERATURA	19
2.2	Discussão dos trabalhos.....	32
3	REFERENCIAL TEÓRICO.....	35
3.1	Aprendizagem profunda	35
3.2	Redes neurais convolutivas.....	40
3.2.1	Camada de convolução	43
3.2.2	Camada de unidades Retificadoras Lineares (ReLU).....	44
3.2.3	Camada de <i>pooling</i>	45
3.2.4	Camada de <i>Batch Normalization</i>	46
3.2.5	Camada de <i>Dropout</i>	47
3.2.6	Camada de convolução transposta.....	48
3.2.7	Segmentação semântica.....	48
3.2.8	Treinamento de uma RNC.....	50
3.3	Métodos de Otimização	51
3.3.1	<i>Stochastic Gradient Descent with Momentum</i> (SGDM)	51
3.3.2	<i>Root Mean Square Propagation</i> (RMSProp)	52
3.3.3	<i>Adaptive Moment Estimation</i> (ADAM).....	53
4	MATERIAIS E MÉTODOS.....	55
4.1	Definição do ambiente de trabalho	55
4.2	Modelagem das bases de dados	55
4.3	Escolha e implementação das arquiteturas	59
4.3.1	Arquitetura 1 (CNN1).....	59

4.3.2	Arquitetura 2 (CNN2).....	60
4.3.3	Arquitetura 3 (CNN3).....	61
4.3.4	Arquitetura 4 (CNN4).....	62
4.4	Experimentos	63
4.5	Métricas de desempenho.....	64
5	RESULTADOS E DISCUSSÕES.....	67
5.1	Resultados de validação do <i>Dataset A</i>	67
5.2	Resultados de validação do <i>Dataset B cropped</i>	71
5.3	Resultados de validação do <i>Dataset B</i> sem <i>cropt</i> 160x160 pixels	74
5.4	Resultados de validação do <i>Dataset B</i> sem <i>cropping</i> 320x320 pixels	76
5.5	Teste de significância estatística.....	78
5.6	Resultados de teste do <i>Dataset A</i> com validação cruzada.....	79
5.7	Resultados de teste do <i>Dataset B</i> com validação cruzada	80
5.8	Discussão dos resultados	81
6	CONCLUSÕES	84
	REFERÊNCIAS	85

1 INTRODUÇÃO

As estimativas globais da OMS sobre mortalidade e perda de saúde da população mundial, devido a doenças e lesões, apontam o câncer como a segunda maior causa morte, representando cerca de 9,6 milhões de mortes em 2018 (*Global Health Estimates*, 2018). A cada 2 anos a *American Cancer Society* descreve as últimas tendências de câncer de mama, fornecendo uma visão geral da doença nos Estados Unidos. Segundo tais tendências, o câncer de mama continua sendo o mais comuns em mulheres, independentemente da idade, e perde apenas para o câncer de pulmão como causa morte (DESANTIS et al., 2017). No Brasil, esses dados se repetem, sendo o câncer de mama o mais comum entre as mulheres, depois do câncer de pele. Ele é responsável por cerca de 28%, dos novos casos de câncer. Para o ano de 2018 foram estimados 59.700 novos casos de câncer de mama no território brasileiro (Ministério da Saúde, 2018).

De acordo com Smith e colaboradores (2006), o diagnóstico precoce traz mais eficácia ao tratamento e pode reduzir significativamente as taxas de mortalidade. Ele se dá através da análise das imagens médicas, que são uma parte integrante do prontuário eletrônico do paciente (PEP) e são fundamentais no diagnóstico do mesmo. Para além do uso na fase diagnóstica, as imagens médicas se tornaram essenciais, também, no seguimento do tratamento e no acompanhamento pós-tratamento.

A Mamografia digital e a Ultrassonografia (US) são duas técnicas comuns para o auxílio da detecção de lesão de mama. A primeira é utilizada como triagem de pacientes assintomáticos pois possui uma alta sensibilidade, identificando assim a maior parte dos verdadeiros positivos. Porém, essa modalidade de imagem possui uma desvantagem, a sua baixa especificidade, o que significa que são detectadas anormalidades que não são de fato lesões malignas. A baixa especificidade da mamografia leva o paciente a um diagnóstico diferencial, como a biópsia, para a confirmação se a lesão detectada é maligna ou benigna (AKIN et al., 2012). A segunda técnica, a Ultrassonografia não é aplicada em pacientes assintomáticos, mas tem alta precisão na diferenciação de cistos versus anormalidades sólidas detectadas através da mamografia.

O diagnóstico por imagem é realizado por profissional médico especialista, o radiologista (KER et al., 2017). A formação desses médicos requer minimamente três anos após a graduação, implicando em alto custo financeiro. Alguns sistemas de saúde, notadamente de países subdesenvolvidos e com número insuficiente de profissionais especializados, como a Índia, terceirizam os serviços de diagnóstico por imagem através da tele radiologia. Além disso, em uma era em que o uso de dados médicos está em constante crescimento, a pressão sobre os especialistas só tende a aumentar.

Identificam-se na literatura vários trabalhos visando dar suporte ao diagnóstico por imagem realizado por médicos, buscando melhorar o processo diagnóstico, algumas vezes afetado pela fadiga visual e pelas características que toda atividade visual humana está sujeita, qual seja, a variação intra e interobservador. Esses métodos analisam automaticamente as imagens buscando detectar e/ou evidenciar regiões de interesse - CADe (do inglês *Computer Aided Detection*) e/ou classificá-las como de características benignas ou malignas - CADx (do inglês *Computer Aided Diagnosis*) e estão presentes nos mais variados tipos de aplicações médicas, bem como o diagnóstico de lesões mamárias. A evidenciação da região de interesse torna mais fácil a análise de malignidade da lesão.

Os CADs (*detection* ou *diagnosis*) atuais de aplicação no câncer de mama utilizam Redes Neurais Artificiais (RNA) rasas e Redes neurais Artificiais profundas (RNP), que é a evolução da primeira (DAOUD et al., 2016; DHUNGEL; CARNEIRO; BRADLEY, 2015; HUANG et al., 2015; JIAO; WANG, 2011; PANIGRAHI; VERMA; SINGH, 2019). A maioria dos trabalhos publicados focam em um tipo de RNP específico, as Redes Neurais Convolucionais (RNC), na de segmentação e classificação de lesões em imagens de mamografia (AHN et al., 2017; BYRA et al., 2017; HAN et al., 2017; QI et al., 2019; XI; SHU; GOUBRAN, 2018; YAP et al., 2018). Ressalta-se que como o próprio nome CAD sugere, esses métodos visam dar suporte ao diagnóstico realizado pelo especialista e não o substituir.

Essa dissertação visa contribuir com o avanço do estado da arte do tema de diagnóstico de mama assistido por computador, desenvolvendo um método automático para a tarefa de detecção do contorno de lesões em imagens de ultrassom de mama, como as mostradas na Figura 1. Foram desenvolvidos quatro modelos utilizando a abordagem de aprendizagem de máquina profunda, mais especificamente através de redes neurais convolucionais e sua validação com dois bancos de imagens de mama.

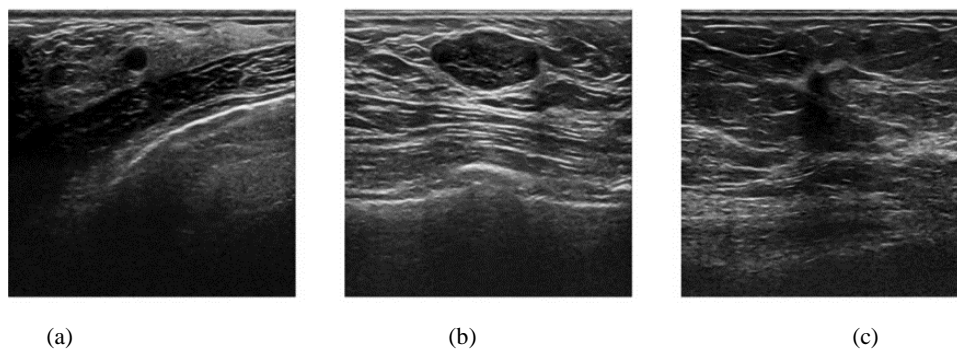


Figura 1: Imagens de ultrassom com lesões mamárias. Observe as variações na forma da lesão, a baixa relação sinal ruído, o baixo contraste, entre outros. Fonte: (YAP et al., 2018)

1.1 Objetivo Geral

Implementar um método fundamentado em redes neurais convolucionais para a detecção de contornos e irregularidades de lesões mamárias em imagens de ultrassonografia (segmentação automática de lesões de mama).

1.2 Objetivos Específicos

1. Verificar o estado da arte na tarefa de segmentação de imagens de ultrassonografia de mama;
2. Avaliar se diferentes tamanhos de imagem, ou o uso de *cropping*, influenciam no desempenho das RNCs.
3. Testar quatro arquiteturas de RNCs com estruturas acíclicas e sequenciais e determinar qual delas obtém melhor desempenho.
4. Propor a partir da melhor arquitetura um método para segmentação automática de lesões em imagens de ultrassom.

1.3 Organização do Trabalho

Este trabalho está organizado conforme a divisão descrita a seguir:

- Capítulo 1: Introdução;
- Capítulo 2: Revisão Bibliográfica;
- Capítulo 3: Referencial Teórico;
- Capítulo 4: Materiais e Métodos;
- Capítulo 5: Resultados e discussão;
- Capítulo 6: Conclusões

O Capítulo 1 contextualizou o problema do câncer de mama, a importância dos CADs no auxílio ao diagnóstico, mostrou os objetivos gerais e específicos.

O Capítulo 2 apresenta trabalhos que foram publicados na literatura na área de segmentação de lesões de mamas, por meio de variadas abordagens e técnicas e qual técnica tem se mostrado mais relevante.

O Capítulo 3 apresenta a base teórica necessária para entendimento deste trabalho: Redes neurais artificiais, redes neurais convolucionais e seu desenvolvimento na segmentação de imagens os parâmetros internos de uma rede convolucional, os métodos de otimização e as métricas de desempenho aplicadas.

No Capítulo 4 são apresentados os materiais utilizados nessa proposta de projeto, as características do banco de dados, as arquiteturas propostas e o fluxo do treinamento, validação e teste.

No Capítulo 5 são apresentados os resultados e discussões feitas através da análise dos dados e testes realizados.

Por fim, no Capítulo 6, são apresentadas as conclusões.

2 REVISÃO DA LITERATURA

Foi feita uma revisão da literatura sobre “segmentação e diagnóstico de lesões de mama” com o intuito de obter o estado da arte sobre o tema. As pesquisas foram realizadas nas bases literárias IEEE *Xplore* e *Engineering Village*. Os artigos foram obtidos utilizando a mesma estratégia de busca avançada nas duas bases de dados, fazendo-se as devidas conversões e adaptações. Além disso, alguns trabalhos foram obtidos também através da referência dos autores, nos trabalhos já encontrados. Foi também feito o uso de mapas de citação, que apresenta os artigos que citaram uma determinada publicação. Buscou-se principalmente por artigos de revista, por apresentarem maior credibilidade e confiabilidade, e deu-se preferência pelos mais recentes.

Através da análise dos trabalhos obtidos verificou-se que haviam três tipos de abordagens, os artigos que aplicavam técnicas de Processamento Digital de Imagens (PDI), os que utilizavam técnicas de Reconhecimento de Padrões (RP) e os híbridos, combinando as duas abordagens.

Foram analisados trabalhos desde 2001, o Quadro 1 permite analisar a evolução dos sistemas de diagnóstico assistido por computação (CAD), das técnicas de PDI e do uso de RP, na tarefa de segmentação automática de imagens de ultrassonografia mamária. A tendência dos trabalhos, conforme disposto no quadro, aponta uma técnica consagrada nos dias de hoje, para trabalhos relacionados à área.

2.1 Análise dos trabalhos

O primeiro trabalho a ser abordado apresenta um algoritmo de segmentação automática de massas em imagens de ultrassonografia mamárias (HORSCH et al., 2001). Como a maioria dos trabalhos que virão a seguir, a segmentação das lesões para o treinamento do modelo foi feita de forma manual, com um radiologista ou físico médico delineando a margem da lesão, utilizando para este fim um *software*. O algoritmo do autor envolve as etapas de pré-processamento por corte e filtragem mediana, multiplicação com uma função de restrição gaussiana, determinação de margens potenciais de lesão através do limiar de valor de cinza e maximização de uma função de utilidade nas margens de lesão em potencial. Uma das vantagens do método é que ele tende a produzir margens semelhantes às lesões. A métrica utilizada foi a Razão de Sobreposição (RS) que na época era uma forma robusta para expor irregularidades nos contornos. Este trabalho merece destaque por ser um dos pioneiros e trabalhar com um banco de dados de 400 casos (124 cistos complexos, 182 lesões benignas e

94 lesões malignas) e 757 imagens, considerado grande para a época. Para o banco utilizado, 94% das imagens apresentaram RS acima de 0,4.

Azevedo (2003) explicou de forma clara a importância da segmentação de lesões em imagens de mama. De acordo com o autor, tumores malignos geralmente são infiltrantes e invadem os tecidos adjacentes, gerando, na imagem por ultrassom, um contorno irregular ou indefinido. Logo, a segmentação do tumor pode gerar uma hipótese sobre a malignidade. Este artigo apresenta um método, baseado em operadores morfológicos (OM), para segmentar e determinar este contorno das imagens de US e faz uso de *Watershed*. Este operador determina as linhas que delimitam a transição entre os vales de uma imagem representada na forma topográfica. Porém, este conceito também pode ser utilizado para separar regiões modais de um histograma. A métrica utilizada foi a mesma de Horsch e contribuintes (2001), a RS, que para $RS = 0,6$ cerca de 90% dos contornos dos tumores foram determinados corretamente, que é um valor compatível ao da literatura da época porém com um banco de dados menos representativo de apenas 22 imagens.

Os trabalhos seguintes (GÓMEZ et al., 2009, 2010; INFANTOSI et al., 2008; MIYAGAWA et al., 2018; XU; ZHIMIN; JIWU, 2005) utilizam uma metodologia similar, aplicando abordagens para realce de contraste, combinados com formas de filtragem de ruídos, bem como diferentes métodos de segmentação para obter o contorno das lesões.

O método de Xu e colaboradores (2005) consistia em filtrar imagens com o algoritmo de difusão anisotrópica, uma técnica de filtragem não linear que servia para remover o ruído de manchas. No artigo, foi combinado o corte normalizado Ncut (um gráfico teórico que admite a combinação de diferentes recursos para a segmentação de imagens), considerada uma técnica não supervisionada, com o método de fusão de região para a segmentação. Os critérios de fusão foram derivados das regras empíricas usadas pelos radiologistas ao interpretar imagens de mama. Na avaliação de desempenho, comparou-se os limites das lesões detectadas por computador com as bordas delineadas manualmente, através do erro médio e a RS.

O artigo de Gómez et al. (2009) aplica técnicas de morfologia matemática para filtragem, Transformada de *Watershed*, o mesmo utilizado por Azevedo (2003), para a segmentação, que pode ser classificada como uma abordagem de segmentação baseada em região. A função de utilidade utilizada no algoritmo de segmentação foi a função derivativa radial média (ARD, do inglês *Average Radial Derivative*). O trabalho de Infantosi colaboradores (2008) também faz uso de operadores morfológicos com uma restrição de função gaussiana para delimitar o domínio de pesquisa associado com a técnica de *Watershed* para encontrar o contorno do tumor. Gomez no ano seguinte publicou um trabalho similar, que se diferia do trabalho anterior apenas

uso do filtro de Gabor, como técnica de filtragem (GÓMEZ et al., 2010). Gómez et al. (2010) e Infantosi et al. (2008) utilizaram a métrica RS. Os bancos diferiam pela quantidade, o primeiro com 36 imagens e o segundo com 50. Os autores observaram ainda em ambos os trabalhos a dificuldade de diagnosticar tumores malignos ou benignos apenas pela inspeção visual do especialista.

O próximo artigo, de Jiao e Wang (2011) merece destaque por ser o primeiro a fazer uso de Reconhecimento de Padrões (RP), com o uso de uma rede neural denominada pelo autor *Pulse-coupled Neural Network* (PCNN). O autor propõe usar uma estratégia de dois estágios: primeiro a geração e depois a segmentação da região de interesse (ROI). No qual, inicialmente, o modelo de PCNN é usado para categorizar a imagem em diferentes classes nas quais a ROI é selecionada, obtendo o contorno aproximado. Então, o contorno aproximado é usado como condição inicial para os contornos ativos modificados sem arestas (ACWE), um modelo baseado em análise de regiões, com o intuito de obter o limite final do tumor da mama. Este trabalho usou métricas menos subjetivas e que são utilizadas ainda atualmente, como as taxas de Verdadeiro positivo (TP), Falso positivo (FP) e Falso negativo (FN). Os autores concluem que não há necessidade de pré-processamento no seu método e que o uso da PCNN para gerar ROI reduz o custo computacional em relação ao método de contornos ativos (ACWE).

Os trabalhos de Cai e Wang (2013), Daoud e colaboradores (2012) e Xian, Cheng e Zhang (2014) continuaram utilizando técnicas de PDI e propondo seus próprios algoritmos de segmentação. Entretanto, os avanços no estado da arte alegados não podem ser de fato verificados, conforme será comentado mais à frente. O trabalho de Daoud e colaboradores (2012) além de segmentar fez a classificação de lesões em malignas e benignas, porém com um banco de dados limitado de apenas 10 imagens. Já as publicações de Cai e Wang (2013) e Xian, Cheng e Zhang (2014) utilizaram dentre as suas métricas a razão de sobreposição e Falso positivo. O primeiro trabalho alcançou resultados do estado da arte com um banco de dados de 168 imagens. Estes trabalhos são os últimos do Quadro 1 que fizeram uso exclusivo de técnicas de PDI, os trabalhos seguintes implementam técnicas de PDI e RP.

Huang (2015) fez uso de um esquema de interação automática baseado em um método de reconhecimento de objetos para segmentar as lesões nas imagens de ultrassom de mama. Nesse esquema, uma imagem de ultrassom 2D é primeiramente filtrada com um modelo de variação total para reduzir o ruído de manchas. Um método de segmentação baseado em gráfico é então usado para segmentar a imagem em várias sub-regiões. Após, um método de reconhecimento de padrões, as Máquinas de vetores de Suporte (SVM) que incorpora os procedimentos de extração, seleção e classificação de características para identificar automaticamente as regiões

associadas aos tumores de mama. Além disso, um modelo de contorno ativo é usado para refinar os contornos das regiões que são reconhecidas como tumores. Neste estudo, cinco formas de extração de características das imagens, incluindo o histograma de nível de cinza (HIS), a matriz de coocorrência de nível de cinza (GLCM), o histograma de gradientes orientados (HOG), a forma e localização são consideradas. As métricas taxas de TP, FP e FN foram utilizadas neste trabalho.

O próximo artigo propõe um algoritmo que decompõe a imagem em um conjunto de superpixels usando o método *Normalized Cuts* (NCuts), juntamente com a análise de textura (DAOUD et al., 2016). Um classificador SVM foi empregado para estimar a probabilidade do superpixel pertencer a um tumor com base em cinco características de textura. Um super pixel de semente é identificado com base nas probabilidades de pertencer a um tumor e na sua localização espacial. O super pixel de semente é estendido para destacar a região do tumor usando uma abordagem de crescimento de região que combina as probabilidades de pertencer a um tumor dos superpixels e a análise baseada na borda. O autor comparou o desempenho do seu algoritmo com dois algoritmos de segmentação já existentes o algoritmo Buts NCuts (*Biased NCuts*) e o algoritmo GVF (*Gradient Vector Flow*), no qual foi obtido superioridade para toda as métricas avaliadas.

Os artigos até então utilizavam técnicas de PDI de forma a obter a extração de características, e métodos de RP classificavam as características já extraídas, de uma forma separada, sem quaisquer relações entre as duas abordagens. Nesse contexto, buscava-se criar um método que obtivesse as melhores representações paramétricas das imagens e então apresentar essas representações a um classificador, como o SVM ou RNA para classificar as lesões. Com a criação das redes convolucionais isso passou a ser feito de forma integrada. Em uma mesma arquitetura de rede neural é possível fazer a extração de características e a classificação, utilizando ainda o mesmo algoritmo para filtrar, obter as características e os pesos para classifica-las.

Os trabalhos de Azevedo (2003), Daoud e colaboradores (2012), Gómez e contribuintes (2009), Huang e colaboradores (2015), Jiao e Wang (2011) e Xu, Zhimin e Jiwu (2005) utilizaram, conforme mencionado anteriormente, bancos de dados pequenos, com menos de 50 imagens. A implementação de técnicas de segmentação e classificação de imagens com as redes neurais convolucionais (RNCs) necessitam de uma grande quantidade de dados para serem treinadas e poder generalizar os dados de forma correta. Com essa abordagem, a classificação é feita diretamente pela análise da imagem, eliminando a dependência da rede de uma boa parametrização das características para obter um bom desempenho. Além disso, bancos de

dados maiores aumentam a confiabilidade do trabalho, e dão uma ideia mais fidedigna do real desempenho do sistema em testes reais.

O trabalho de Long e colaboradores (2015) não está no Quadro 1, porém ele é a base para o avanço dos sistemas CADs, e, portanto, merece destaque, ele é o primeiro a usar de forma eficiente uma rede convolucional para segmentar imagens. Antes do uso para segmentação, as redes convolucionais vinha obtendo sucesso nas mais variadas tarefas. Isso inclui avanços na detecção de objetos (AHN et al., 2017; DHUNGEL; CARNEIRO; BRADLEY, 2015), predição (PANIGRAHI; VERMA; SINGH, 2019; YAP et al., 2018) e outras tarefas. Long, Shelhamer e Darrell (2015) propuseram a ideia de rede totalmente convolucional (FCN) treinada em todas as partes, pixel a pixel sem a necessidade de outras abordagens de aprendizado de máquina (segmentação semântica). Versões totalmente convolucionais das redes existentes preveem saídas densas de entradas de tamanho arbitrário. De forma resumida, os autores estenderam as redes convolucionais de classificação para a tarefa de segmentação, melhorando a arquitetura com combinações de camadas com várias resoluções, ao mesmo tempo em que simplificaram e aceleraram o aprendizado e a inferência, propiciando um relevante avanço no estado da arte.

Com o sucesso das RNCs, vários trabalhos publicados a partir desse momento começaram a fazer uso dessa técnica para o problema de segmentação e classificação de lesões mamárias. Ahn e colaboradores (2017) faz uso das redes para estimativa automatizada da densidade mamográfica. A sua rede foi capaz de aprender as características de imagens e classificá-las em tecidos densos e gordurosos. Sua RNC continha as 16 camadas escondidas, dentre estas haviam camadas de convolução, ReLU, *pooling*, *drop-out*, totalmente conectada e regressão. O método proposto pelo autor exibiu uma correlação de 0.96 com as mensurações dos especialistas.

A abordagem de Byra e colaboradores (BYRA et al., 2017) é baseada em ultrassom quantitativo (imagem de Nakagami) e aprendizado profundo para a classificação. Esse tipo de imagem revela propriedades do tecido que não são visíveis em imagens padrão, criando mapas paramétricos de lesões mamárias que ilustram as propriedades de dispersão de tecidos. Para esta tarefa, foi aplicada a técnica da janela deslizante. A janela deslizante de 2 mm indicou uma região de interesse contendo amostras de amplitude que foram usadas para calcular o parâmetro Nakagami com o estimador de máxima verossimilhança. A janela foi movida por toda a imagem em etapas de um pixel. A cada vez, o valor do parâmetro Nakagami estimado foi atribuído ao pixel central da janela. Em seguida, usou-se os mapas de parâmetros de Nakagami para treinar uma rede neural convolucional. A RNC empregada tinha 14 camadas ocultas. O desempenho da classificação foi avaliado por validação cruzada com 5 pastas, obtendo 83% de acurácia.

Seokmin Han e colaboradores (2017), exploram a estrutura de aprendizado profundo para diferenciar os tipos distintos de lesões e nódulos na mama adquiridos com imagens de ultrassom. O método desenvolvido incluía equalização de histograma, corte de imagem e aumento de margem. A estrutura de RNC da GoogLeNet foi treinada no banco de dados para diferenciar tumores benignos e malignos. No treinamento as imagens foram adaptadas de nível de cinza para RGB, e a RNC foi treinada com *fine-tuning*. Sua rede RNC tem mais 90 camadas, entre elas camadas de convolução, ReLU, *pooling*, *Depth concatenation*, totalmente conectada e *softmax*. As redes foram treinadas com aumento de dados, e obtiveram 90% de acurácia.

Moi Hoon Yap e colaboradores (2018) propõe o uso de abordagens de aprendizagem profunda para detecção de lesões por ultrassom da mama e aplica para isto três abordagens diferentes: um LeNet baseado em *Patch*, um U-Net e uma abordagem de aprendizado de transferência com um FCN-AlexNet pré-treinado. Abordagem de RNCs baseadas em patches, treina as redes neurais convolucionais com amostras de imagens para treinamento e uma janela deslizante para testes. A abordagem totalmente convolucional, produz segmentação por previsão em pixels, em vez de distribuição de probabilidade única na tarefa de classificação para cada imagem. A abordagem com transferência de aprendizado utiliza um modelo pré-treinado de imagens não médicas para superar a limitação da deficiência de dados na pesquisa de imagens médicas. Os desempenhos são comparados com quatro algoritmos de detecção de lesões de última geração (ou seja, Índice de radial, Filtragem multifractal, Classificação de região com base em regras e Modelos de peças deformáveis). Além disso, este artigo compara e contrasta dois conjuntos de dados de imagens de ultrassom convencionais adquiridos de dois sistemas de ultrassom diferentes. O conjunto de dados A compreende 306 (60 malignas e 246 benignas) e o conjunto de dados B compreende 163 (53 malignas e 110 benignas). Os resultados demonstram que os três métodos obtiveram resultados diferentes, sendo o FCN-AlexNet o melhor, alcançando 99% de taxa de verdadeiro positivo no *Dataset A* e 93% no *Dataset B*.

No trabalho de Xi e colaboradores (2018) foi proposta uma abordagem para classificar e localizar calcificações e massas em imagens de mamografia. Para melhorar as abordagens convencionais, aplicou-se redes neurais convolucionais profundas para o aprendizado automático de recursos e a criação de classificadores. As redes neurais convencionais profundas e avançadas foram comparadas com o desempenho de classificar as anormalidades. Sendo necessário duas redes, uma para identificar se na imagem havia algum tumor e a segunda para identificar regiões de tumores nas imagens ditas com tumor. Resultados experimentais indicaram que a rede VGGNet recebeu a melhor precisão geral em 92,53% nas classificações. Para anormalidades na localização, o ResNet foi selecionado para calcular mapas de ativação

de classe. O autor adaptou o classificador de *patches* para localizar anormalidades em imagens de mamografia completa por meio do mapeamento de ativação de classe.

O trabalho de Qi e outros (2019), propõe um método de diagnóstico de imagens de ultrassom, utilizando duas redes convolucionais agindo em cascata. A primeira é construída para classificar as imagens de acordo com a presença de tumores malignos, uma vez que esses tumores são os mais graves entre todas as doenças da mama. A segunda rede analisa imagens com nódulos sólidos, pois os nódulos sólidos estão intimamente relacionados ao câncer e devem ser tratados com cuidado. Os autores afirmam que o uso das duas redes de forma colaborativa melhora o desempenho de ambas de forma individual. Para permitir que as duas redes trabalhem de forma colaborativa, é proposto um mecanismo de aprimoramento da região com base nos mapas de ativação de classe. O seu trabalho é comparado com duas outras abordagens, alcançando superioridade em ambas.

Quadro 1 - Resumo da análise realizada nos trabalhos relacionados.

<i>Autores</i>	<i>Artigo</i>	<i>Conjunto de dados</i>	<i>Técnica</i>	<i>Métricas/Desempenho</i>
(HORSCH et al., 2001)	<i>Automatic segmentation of Breast Lesions on Ultrassound</i>	<i>Dataset: 757</i> imagens de ultrassonografia	Filtragem: Mediana e função gaussiana Segmentação: ARD (<i>Average Radial Derivative</i>)	Razão de sobreposição (RS): 94% das imagens alcançaram RS acima de 0,4
(AZEVEDO, 2003)	<i>Application of Morphological Operators on the Segmentation and Countourn Detection of Ultrassound Breast Images</i>	<i>Dataset: 22</i> imagens de ultrassonografia	Contraste: Normalização Filtragem: Abertura por reconstrução Segmentação: <i>Watershed-based cluster</i> e Imposição de mínimo	Razão de sobreposição (RS): 90% das imagens alcançaram RS acima de 0,6
(XU; ZHIMIN; JIWU, 2005)	<i>Automated Segmentation of Breast lesions in Ultrasound Images</i>	<i>Dataset: 40</i> imagens de ultrassonografia	Contraste: Máscara <i>Unsharp</i> Filtragem: Difusão anisotrópica Segmentação: Corte normalizado (Ncut), <i>Region Mergin</i>	Erro médio (e): 5,271 Razão de Sobreposição (RS): 0,92

<i>Autores</i>	<i>Artigo</i>	<i>Conjunto de dados</i>	<i>Técnica</i>	<i>Métricas/Desempenho</i>
(INFANTOSI et al., 2008)	<i>Breast Ultrasound Segmentation Using Morphologic Operators and a Gaussian Function Constraint</i>	<i>Dataset (INCa):</i> 200 imagens de ultrassonografia	Filtragem: Abertura por reconstrução seguida de fechamento por reconstrução Segmentação: Um híbrido entre Horsch et al (2001) e Alvarenga et al. (2003) respectivamente, seguido da aplicação do operador watershed para obtenção do contorno final.	Razão de Sobreposição (RS): 91% das imagens obtiveram um $RS > 0,50$. valor médio residual quadrático (nrv):0,70 obtiveram um $nrv < 0,40$
(GÓMEZ et al., 2009)	<i>Morphological Operators on the Segmentation of Breast Ultrasound Images.</i>	<i>Dataset:</i> 36 imagens de ultrassonografia	Filtragem: <i>Alternationg sequential filter</i> (ASF) Segmentação: <i>Watershed-based cluster</i> , imposição de mínimo e função ARD (<i>Average Radial Derivative</i>)	<i>Coincidence Percentage</i> (CP):67% dos contornos obtidos com o método quando comparados ao do R1 e 81% dos contornos em relação ao R2 apresentaram $CP \geq 80\%$ <i>Proportional Distance</i> (PD): 81% dos contornos em relação ao R1 e 83% em relação ao R2 apresentaram $PD \leq 10\%$.
(GÓMEZ et al., 2010)	<i>Segmentation of Breast Nodules on Ultrasonographic Images Based on Marke dControlled Watershed Transform</i>	<i>Dataset (INCa):</i> 50 imagens de ultrassonografia	Filtragem: Filtro de Gabor, filtro de difusão anistrópica e função gaussiana Segmentação: Imposição de mínimo, <i>watershed</i> e ARD	Razão de Superposição (RS): 94% das imagens apresentaram $RS \geq 80\%$ Valor normalizado residual quadrático (nrv):86% das imagens apresentaram um $nrv \leq 0,2$.

<i>Autores</i>	<i>Artigo</i>	<i>Conjunto de dados</i>	<i>Técnica</i>	<i>Métricas/Desempenho</i>
(JIAO; WANG, 2011)	<i>Automatic Boundary Detection in Breast Ultrasound Images Based on Improved Pulse Coupled Neural Network and Active Contour Model</i>	<i>Dataset: 18</i> imagens de ultrassonografia	Contraste: <i>Pulse-coupled Neural Network</i> (PCNN) Segmentação: <i>Level-step</i> baseado em contornos-ativos	Taxa de Verdadeiro Positivo (TPR): 96,7% Taxa de Falso Positivo (FPR): 5,5% Taxa de Falso Negativo (FNR): 3,12%
(DAOUD et al., 2012)	<i>Accurate Segmentation of Breast Tumor in Ultrasound Images using a CustomMade Active Contour Model and Signal-toNoise Ratio Variations</i>	<i>Dataset: 10</i> imagens de ultrassonografia	Filtragem: Função de transformação logarítmica não linear de descompressão Segmentação: Contornos-ativos semiautomático	Taxa de Verdadeiro Positivo (TPR): 91,13% Taxa de Falso Positivo (FPR): 8,87% Taxa de Falso Negativo (FNR): 15,58%
(CAI; WANG, 2013)	<i>A Phase-Based Active contour Model for Segmentation of Breast Ultrasound Images</i>	<i>Dataset: 168</i> imagens de ultrassonografia	Segmentação: <i>Phase-Based Active Contour</i>	Taxa de Verdadeiro Positivo (TPR): 93,9% Taxa de Falso Positivo (FPR): 3,5% Razão de Sobreposição (RS): 0,8959

<i>Autores</i>	<i>Artigo</i>	<i>Conjunto de dados</i>	<i>Técnica</i>	<i>Métricas/Desempenho</i>	
(XIAN; CHENG; ZHANG, 2014)	<i>A Fully Automatic Breast Ultrasound Image Segmentation Approach Based on Neutro-Connectedness</i>	<i>Dataset: 131</i> imagens de ultrassonografia	Contraste: Normalização Filtragem: <i>Smoothing</i> Segmentação: Geração de um ponto semente; <i>Neutro-Connectedness</i>	Razão de Sobreposição (RS): 0,7965 Taxa de Falso Positivo (FPR): 9,85% Erro Médio de Hausdorff (AHE): 0,209	
(HUANG et al., 2015).	<i>Automatic Segmentation of Breast Lesions for Interaction in Ultrasonic Computer-aided Diagnosis</i>	<i>Dataset A: 47</i> imagens de ultrassonografia	Filtragem: <i>Total Variation</i> (TV) Segmentação: <i>Robust Graph Based</i> (RGB) Classificação: SVM	Taxa de Verdadeiro Positivo (TPR): 85,01% Taxa de Falso Positivo (FPR): 1,78% Taxa de Falso Negativo (FNR): 14,99%	
(DHUNGEL; CARNEIRO; BRADLEY, 2015)	<i>Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms</i>	<i>Inbreast: 112</i> imagens DDSMBCRP: 158 imagens de mamografia	Campo Aleatório Condicional (CRF) Máquinas de Vetor de Suporte estruturada	INbreast Dice: 0,90 Tempo de teste: 0,1s Tempo de treino: 360s	DDSMBCRP Dice: 0,90 Tempo de teste: 0,1s Tempo de treino: 383s

<i>Autores</i>	<i>Artigo</i>	<i>Conjunto de dados</i>	<i>Técnica</i>	<i>Métricas/Desempenho</i>
(DAOUD et al., 2016)	<i>Accurate and Fully Automatic Segmentation of Breast Ultrasound Images by Combining Image Boundary and Region Information</i>	<i>Dataset: 50 imagens de ultrassonografia</i>	<i>Filtragem: Speckle Reducing Anisotropic Diffusion Segmentação: Extração de características baseadas em textura, Decomposição da imagem em Superpixels usando Ncuts, SVM e Region Growing</i>	<i>Taxa de Verdadeiro Positivo (TPR): 90,7% Taxa de Falso Positivo (FPR): 9,3% Taxa de Falso Negativo (FNR): 7,6%</i>
(AHN et al., 2017)	<i>A Novel Deep Learning-based Approach to High Accuracy Breast Density Estimation in Digital Mammography</i>	<i>Dataset: 397 imagens de mamografia</i>	<i>RNC</i>	<i>Correlação de marcação: 0,96</i>
(BYRA et al., 2017)	<i>Combining Nakagami Imaging and Convolutional Neural Network for Breast Lesion Classification</i>	<i>Dataset: 458 imagens de ultrassonografia</i>	<i>RNC</i>	<i>AUC médio: 0,984 Acurácia (Acc): 0,8300 Sensibilidade: 0,824 Especificidade: 0,833</i>

<i>Autores</i>	<i>Artigo</i>	<i>Conjunto de dados</i>	<i>Técnica</i>	<i>Métricas/Desempenho</i>	
(HAN et al., 2017).	<i>A deep learning framework for supporting the classification of breast lesions in ultrasound images</i>	<i>Dataset: 7408</i> imagens de ultrassonografia	RNC (GoogLeNet)	AUC médio: 0,900 Acurácia (Acc): 0,900 Sensibilidade: 0,860 Especificidade: 0,960	
(YAP et al., 2018)	<i>Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks</i>	<i>Dataset A: 306</i> imagens de ultrassonografia <i>Dataset B: 163</i> imagens de ultrassonografia	RNC: U-net RNC: FCN-AlexNet Baseada em <i>Patch</i>	<i>Dataset A (FCN-AlexNet)</i> TPF: 99% FPF: 0,16 <i>F-Score: 0,92</i>	<i>Dataset B (LeNet)</i> TPF: 91,1% FPF: 0,09 F-Score: 0,91
(XI; SHU; GOUBRAN, 2018)	<i>Abnormality Detection in Mammography using Deep Convolutional Neural Networks</i>	MIAS: 322 imagens de mamografia DDSM: 5000 imagens de mamografia	RNC: AlexNet RNC: VGGNet RNC: GoogLeNet RNC: ResNet	Calcificação: 0,9242 Massa: 0,9264 Precisão geral: 0,9253	
(QI et al., 2019).	<i>Automated Diagnosis of breast ultrasonography images using deep neural networks</i>	<i>Dataset: 8145</i> imagens de ultrassonografia	2 RNCs conjuntamente (Mt-Net e Sn-Net)	F-score: 0,932 Acurácia (Acc): 0,9013 Sensibilidade: 0,9354 Especificidade 0,8318	

2.2 Discussão dos trabalhos

Os artigos analisados reforçam a importância das pesquisas na área, pela questão de o câncer de mama ser um problema mundial e da importância do diagnóstico precoce, para maior eficácia do tratamento. Além disso, há o fato de que o diagnóstico médico deve ser rápido e preciso, mostrando assim a utilidade de sistemas CAD para apoiar os especialistas.

Os primeiros trabalhos do Quadro 1, baseados em técnicas de PDI, representam um marco por serem os primeiros a tentar automatizar a tarefa da segmentação automática das lesões de mama. É possível verificar que os métodos de PDI apresentam abordagens variadas, diferentemente do que ocorre as propostas que fazem uso de redes convolucionais, que apresentam uma certa similaridade. Com respeito a esses primeiros trabalhos, observa-se que cada autor propôs um método que associa técnicas diversas de PDI. Normalmente as propostas consistiam de uma técnica de melhoramento de contraste, em conjunto com filtragem de ruídos e um método de segmentação. Não se identificava nestas abordagens, grandes incrementos no estado da arte que justificasse a adoção das mesmas por trabalhos posteriores. Era comum, também, o uso de bancos de dados proprietários, com poucas imagens. Adicionalmente, observa-se que a maioria desses trabalhos utilizaram a Razão de Sobreposição como métrica. Essa métrica, conforme mencionado, quantifica o quanto a segmentação automática se sobrepõe a segmentação realizada pelo especialista. Entretanto, a forma de apresentar os resultados não proporcionava uma ideia do desempenho global do método frente ao conjunto total de imagens de teste. Por exemplo, “o método apresentou $RS = 0,6$ para cerca de 90% dos contornos das lesões segmentadas”.

Em outras circunstâncias os autores apresentaram seus resultados com outras métricas, não usuais. O uso de métricas diferentes é um dos pontos que faz com que os desempenhos dos métodos não possam ser de fato comparados. O outro é o uso de conjunto de dados proprietários. Nos primeiros trabalhos do Quadro 1 foi possível constatar também que a maioria dos autores não apresentaram a forma como os dados foram utilizados: houve divisão de dados para a parametrização do método e para o teste deste? Ou o conjunto de dados inteiro foi utilizado tanto para o ajuste quanto para o teste? Se houve divisão, de qual forma foi realizada? Foi feito de forma randômica?

Ainda, com relação ao banco de dados, era comum que as imagens fossem provenientes de único aparelho de ultrassom (AZEVEDO, 2003; GÓMEZ et al., 2009, 2010; HORSCH et al., 2001; INFANTOSI et al., 2008), podendo representar particularidades deste aparelho e não dando uma visão ampla do universo do problema. Autores recentes mostraram que bancos de

dados globais, obtidos de diferentes localidades e diferentes aparelhos representam melhor o problema (DHUNGEL; CARNEIRO; BRADLEY, 2015; XI; SHU; GOUBRAN, 2018; YAP et al., 2018).

A segmentação baseada em região era bastante utilizada (AZEVEDO, 2003; GÓMEZ et al., 2009; XU; ZHIMIN; JIWU, 2005), principalmente a transformada de *Watershed*, porém essa segmentação dependia de um conjunto de regras empíricas ajustadas pelos autores.

Abordagens híbridas de PDI e RP se mostraram mais eficientes e, conforme Quadro 1, são as mais utilizadas. Nos primeiros trabalhos dessa abordagem, entretanto, costumava-se utilizar uma técnica para extração de características e outra para classificação das lesões (DHUNGEL; CARNEIRO; BRADLEY, 2015; HUANG et al., 2015; JIAO; WANG, 2011). Era comum também a utilização de um método de segmentação e outro de método de classificação (SVM, por exemplo). Os trabalhos citados que utilizam essa abordagem ainda utilizavam bancos de dados proprietários pequenos.

A etapa de pré-processamento dos dados/imagens se manteve presente em todas as abordagens, e até mesmo nas redes convolutivas.

As redes convolutivas revolucionaram a forma como os sistemas CADs são desenvolvidos, a análise do Quadro 1 revela uma clara tendência das publicações a partir de 2017 de utilizarem essa técnica. Isso se deve ao fato de essas redes trabalharem bem com problemas complexos, como é caso da identificação de lesões em imagens de mama em imagens de ultrassonografia, que possuem baixa qualidade, alta relação sinal-ruído, baixo contraste e variações significativas no tecido adiposo (AHN et al., 2017). Além disso, com as RNCs é mais fácil ter noção do grau de complexidade do sistema, já que redes com mais camadas ou mais neurônios são mais complexas. Analisando o desempenho do treinamento da rede é possível verificar se o sistema está complexo demais ou muito simples para a tarefa proposta. Utilizar uma complexidade adequada é fundamental para o sucesso do experimento. Concomitantemente, as RNCs forçam os autores a trabalharem com conjunto de dados maiores, propiciando maior representatividade do universo real de possibilidades das lesões mamárias. As redes convolutivas também eliminam a necessidade de métodos distintos para as tarefas de segmentação e classificação, pela adoção de uma arquitetura apropriada. Existe também o fato dessas redes não necessitarem de métodos prévios de extração de características. Embora o tempo de treinamento dessas redes seja superior a quaisquer das abordagens analisadas, uma vez treinado, o modelo consegue analisar imagens de forma rápida, alguns milissegundos, na maioria dos casos.

Os trabalhos mostram uma carência de grandes bancos de dados públicos robustos e balanceados que viabilizem o *benchmark* dos modelos. A criação de um banco público global, com um número grande de imagens, dispensaria o uso da técnica de aumento de dados, regularmente empregada pela maioria dos trabalhos na área.

Alguns trabalhos optaram pela separação aleatória dos dados de treinamento, validação e teste em detrimento do uso da técnica de validação cruzada. Considerando que as RNCs se adaptam aos conjuntos de dados apresentados, corre-se o risco de um bom desempenho obtido para um certo subconjunto de dados, não significa melhor desempenho na prática, pois as imagens deste subconjunto particular podem ter se adaptado melhor ao modelo.

Sistemas CADs propostos com o intuito de auxiliar o diagnóstico médico têm os papéis de aperfeiçoar o diagnóstico por imagem e reduzir a carga de trabalho dos radiologistas. Em uma era que uma grande quantidade de dados está disponível, a era do “*big data*”, as redes convolucionais apontam por ser a melhor forma de algoritmo para um sistema CAD.

3 REFERENCIAL TEÓRICO

Neste capítulo serão abordados os principais conceitos necessários ao entendimento deste trabalho. A Seção 3.1 apresenta os conceitos sobre aprendizagem profunda. Mais à frente, na Seção 3.2, são introduzidos os conceitos básicos de redes neurais convolucionais. Por fim, na Seção 3.3, são apresentados os métodos de otimização empregados.

3.1 Aprendizagem profunda

Desde o início dos tempos, o homem vem construindo ferramentas cada vez melhores para entender e controlar o ambiente ao redor. A inteligência artificial (IA) tem como base o aprendizado de máquina (AM), no qual o aprendizado profundo (AP) é um subconjunto de métodos. O AM é um campo de estudo que visa dar aos computadores a capacidade de aprender sem serem explicitamente programados (TRASK, 2019). O aprendizado profundo é mais um capítulo na história da evolução humana. O aprendizado profundo se baseia em conceitos inspirados no cérebro humano (redes neurais) para alcançar aplicações nos mais variados campos da ciência, medicina, tecnologia e outros.

O termo *Deep learning* (Aprendizado profundo) é bastante amplo, sendo importante começar pela sua definição. Segundo Patterson e Gibson (2008) uma definição comum para o termo é de “uma rede neural com mais de duas camadas”(vide a Figura 2). O problema dessa definição é que ela faz com que a aprendizagem profunda pareça estar presente desde a década de 80, o que não é uma verdade. Os autores apontam que as redes neurais tiveram que transcender arquiteturalmente os estilos de rede anteriores (em conjunto com muito mais poder de processamento) para serem consideradas profundas, alcançando enorme sucesso e progresso, excedendo resultados de artigos anteriores nas tarefas de visão computacional, reconhecimento de fala, dentre outras. A seguir, estão algumas das facetas dessa evolução das redes neurais citadas em Patterson e Gibson (2008):

- Mais neurônios que as redes anteriores;
- Maneiras mais complexas de conectar camadas;
- Explosão na quantidade de energia de computação disponível para treinar;
- Extração automática de características;

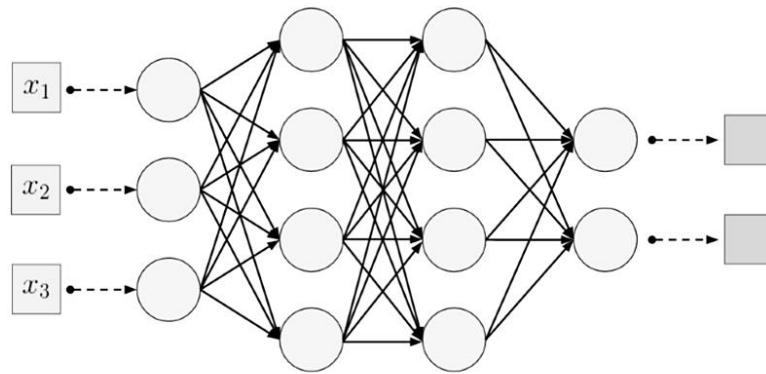


Figura 2: Topologia de uma rede neural com múltiplas camadas. Fonte: (PATTERSON; GIBSON, 2008)

A extração automática de características é outra das grandes vantagens que o aprendizado profundo tem sobre os algoritmos tradicionais de aprendizado de máquina. Extração de características é o processo de decidir quais características de um conjunto de dados podem ser usadas como indicador para rotulação desses dados de maneira confiável. Com outras palavras, extrair as características é obter modelos matemáticos que façam uma representação paramétrica precisa para resolver um determinado problema. Historicamente, os profissionais de aprendizado de máquina e das mais variadas áreas da ciência passaram meses, dedicavam um tempo substancial criando manualmente métodos de extração exaustivos para a classificação de dados. Como, por exemplo, no início da evolução dos CADs, em que era preciso criar um método para extração e outro para classificação dos dados (DHUNGEL; CARNEIRO; BRADLEY, 2015; HUANG et al., 2015; JIAO; WANG, 2011).

Na época da “explosão” do aprendizado profundo, iniciada em 2006, o desenvolvimento de algoritmos de aprendizado de máquina de última geração demandou décadas de esforço humano, que redundaram na disponibilização de técnicas e métodos relevantes para classificar informações. O aprendizado profundo superou esses algoritmos convencionais com precisão em quase todos os tipos de dados, notadamente com imagens (Quadro 1).

De maneira simples, as redes profundas têm por objetivo simular o comportamento humano em tarefas que envolvam detecção, classificação ou reconhecimento. Esta simulação dá-se através da aplicação de métodos ou algoritmos envolvendo modelos matemáticos em conjunto com redes neurais artificiais que permitam à máquina (computador) aprender por experiência a partir dos dados de entrada. Neste processo de aprendizado, a máquina cria representações hierárquicas que permitem dividir tarefas complexas em instâncias de menor complexidade.

A unidade básica das redes neurais artificiais e das redes profundas é o neurônio (em analogia com cérebro humano) podendo ser visto como um dispositivo computacional básico capaz de realizar cálculos simples de somas e multiplicações. O poder dessa unidade se dá quando ela é utilizada em rede (vários neurônios conectados) através da qual o cérebro (ou arquitetura de rede) é capaz de desempenhar cálculos complexos e tomar decisões (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Contextualizando, a biologia, em específico a neurociência, serviu de base para a criação do neurônio artificial. Um neurônio do sistema nervoso do corpo humano é composto por três estruturas: corpo da célula, que pode ter várias formas e tamanhos; dendritos, que são extensões do corpo da célula e atuam na recepção dos estímulos provenientes de outros neurônios; e axônio, que é uma estrutura fina e longa que transmite sinais elétricos para outros neurônios. É fácil notar que esse modelo biológico pode ter um equivalente matemático sendo ilustrado na Figura 3. As entradas (X_i) são conectadas ao neurônio através dos pesos (W_i) que simulam a estrutura do dendrito. O somatório, a polarização (b) e a função de ativação (f) funcionam como o corpo da célula. A propagação da saída (Y) é análoga ao axônio em um neurônio biológico (AWAD; KHANNA, 2015).

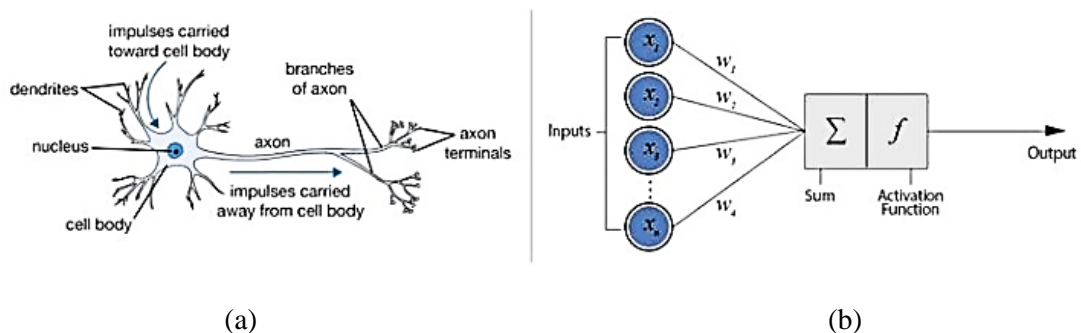


Figura 3: Neurônio biológico e neurônio artificial. Fonte: (MWANDAU; NYANCHAMA, 2018)

A representação matemática de um neurônio artificial é dada por:

$$Y = f\left(\sum_{i=1}^n W_i X_i + b\right) \quad (1)$$

e pode ser modelado usando a forma matricial:

$$Y = f(W \cdot X + b) \quad (2)$$

em que, $W = [W_1 \quad W_2 \quad \dots \quad W_n]$ e $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$.

A entrada X , ou vetor de entrada, representa um conjunto de n características que são fornecidas à entrada do neurônio artificial. Cada uma dessas características tem uma influência sobre a saída, até mesmo uma entrada nula. Essa influência é mediada pela matriz de pesos W . Para que a rede neural realize uma determinada função, essa matriz de pesos W é alterada no processo de aprendizagem do algoritmo. A polarização b tem por finalidade permitir um melhor ajuste da função de ativação f , em conformidade com a tarefa a ser desempenhada pela rede neural. A função de ativação prediz o valor da saída e pode ter várias formas. O desempenho do modelo de rede neural depende da escolha da função de ativação, bem como a escolha da complexidade do modelo adequado para o problema. As funções de ativação mais utilizadas são:

- Sigmoide: $f(x) = \frac{1}{1+e^{-x}}$
- Tangente hiperbólica: $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \tanh(x)$
- ReLU: $f(x) = \max(0, x)$
- Identidade: $f(x) = x$

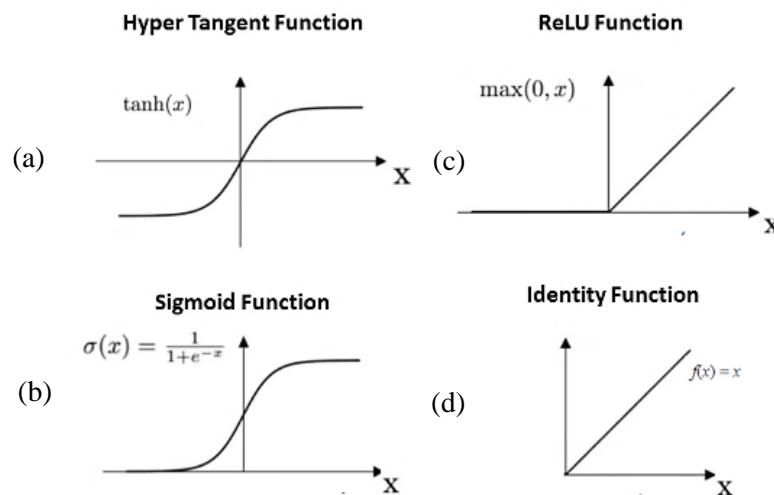


Figura 4: Funções de treinamento das Redes Neurais Profundas (RNPs). Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)

Uma camada de neurônios é formada por vários neurônios artificiais conectados a um mesmo conjunto de entrada, operando em paralelo. A saída de uma camada é um vetor, em que cada componente corresponde a saída de um neurônio artificial dessa camada. A rede neural é

simplesmente uma associação em cascata de camadas de neurônios, cada uma com sua própria matriz de peso, vetor de polarização e vetor de saída (AWAD; KHANNA, 2015). Na rede neural, as camadas adjacentes são completamente conectadas, ou seja, todos os neurônios de uma camada da rede estão conectados a todos os neurônios da camada seguinte. A quantidade de camadas e de neurônios necessários para o desempenho ótimo de um modelo depende da complexidade da aplicação para a qual o mesmo é projetado. A Figura 5 ilustra uma rede neural com duas camadas ocultas.

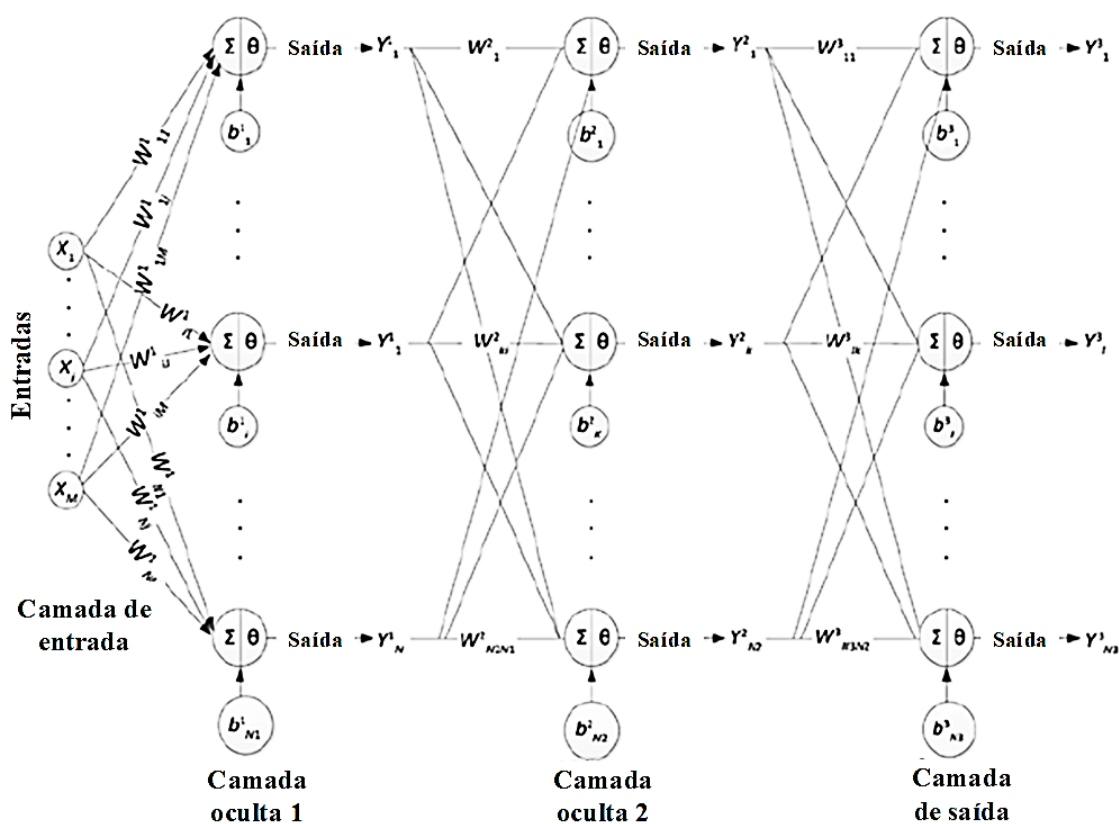


Figura 5: Exemplo de uma RNP. Fonte: Adaptado de Awad e Khana (2015)

Então as redes profundas seguem a mesma estrutura, diferenciando-se apenas pelo número de camadas, quantidades de neurônios e funções de ativação associadas as camadas. Essa estrutura de rede profunda geralmente tem um número grande de neurônios e camadas. Elas utilizam funções de ativação não lineares, pois já foi provado que o uso de múltiplas camadas lineares só é capaz de resolver problemas lineares (HAGAN et al., 2014). Além disso, conforme podemos perceber cada neurônio de uma camada é conectado a todos os outros da

camada seguinte, significando que arquiteturas complexas requerem um grande custo computacional, devido aos cálculos de parâmetros internos.

3.2 Redes neurais convolutivas

Conforme foi visto na Figura 5, as redes neurais profundas seguem uma estrutura em que cada neurônio de uma camada é conectado a todos os neurônios de uma camada anterior. Por isso, algumas abordagens chamam essas de redes densas. Além disso, elas não são ideais para se trabalhar com imagens, devido ao fato que a entrada dessas redes são vetores (unidimensionais), enquanto que uma imagem RGB pode ter 3 dimensões. Antes das redes convolutivas, quando desejava-se fazer uso de redes neurais profundas com imagens, estas sofriam um processo de “vetorização”, no qual todos os pixels de uma linha eram concatenados com a próxima linha, obtendo um vetor unidimensional. Porém nesse processo a imagem perdia suas características próprias, acarretando resultados nem sempre satisfatórios em tarefas de classificação.

Pensando nisso, LeCun e colaboradores (1998) propuseram as redes neurais convolutivas (RNC), denominada LeNet (vide Figura 6), um método que se diferencia das redes tradicionais por combinar três abordagens: campo receptivo local, pesos compartilhados e subamostragem. As RNC recebem esse nome devido ao emprego de uma operação muito conhecida na engenharia elétrica, a convolução. A convolução aplica uma máscara ou *kernel* em pequenas regiões da imagem de forma repetida (GOODFELLOW; BENGIO; COURVILLE, 2016), ou seja, o *kernel* ao ser aplicado em uma pequena região da imagem faz o papel do campo receptivo. Isso é uma ideia revolucionária. Imagine uma imagem, com milhares de pixels (por exemplo, uma imagem em níveis de cinza 1024x1024 pixels) como entrada de uma rede convolucional densa, no qual as unidades de cada camada são totalmente conectadas com as próximas. Cada unidade computacional da camada de entrada, por exemplo, estaria conectada a cada uma dessas variáveis, totalizando centenas de milhares de pesos a serem computados somente nesta camada.

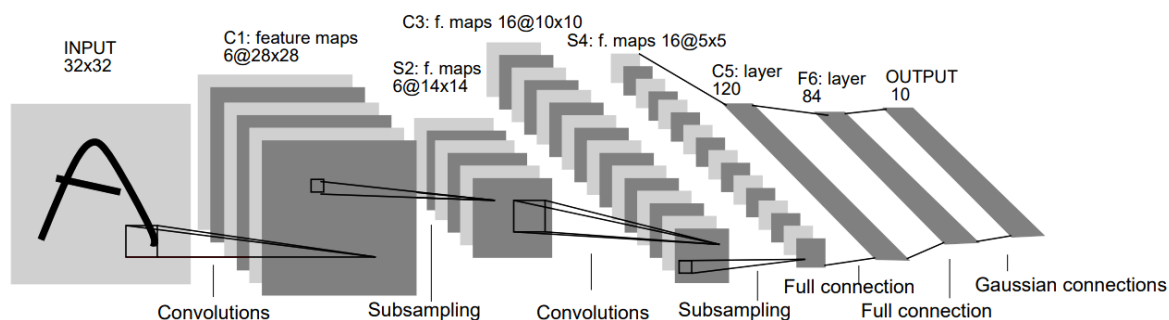


Figura 6: Arquitetura LeNet-5. Fonte: (LeCun et al., 1998b)

As RNC possuem diversas arquiteturas que variam de acordo com a aplicação que diferem entre si pelo tipo e número de camadas, bem como pela disposição destas na rede. As principais camadas, tratando-se de RNCs, são: Convolução, *Pooling*, Unidade Linear Retificadora (ReLU), Classificação, *Dropout* e *Batch Normalization*.

Essas redes apresentam uma estrutura 3D, no qual cada camada pode ter um determinado número de planos. Esses planos são responsáveis por extrair as características da imagem. Em camadas iniciais, os planos podem conter características simples como cantos e bordas em diversas orientações. Conforme a rede vai acrescentando planos e camadas as abstrações obtidas nos planos são mais profundas, a ponto de obter características mais complexas, tal como partes de um objeto ou pessoa, como orelha, boca e nariz, como pode ser observado na Figura 7.

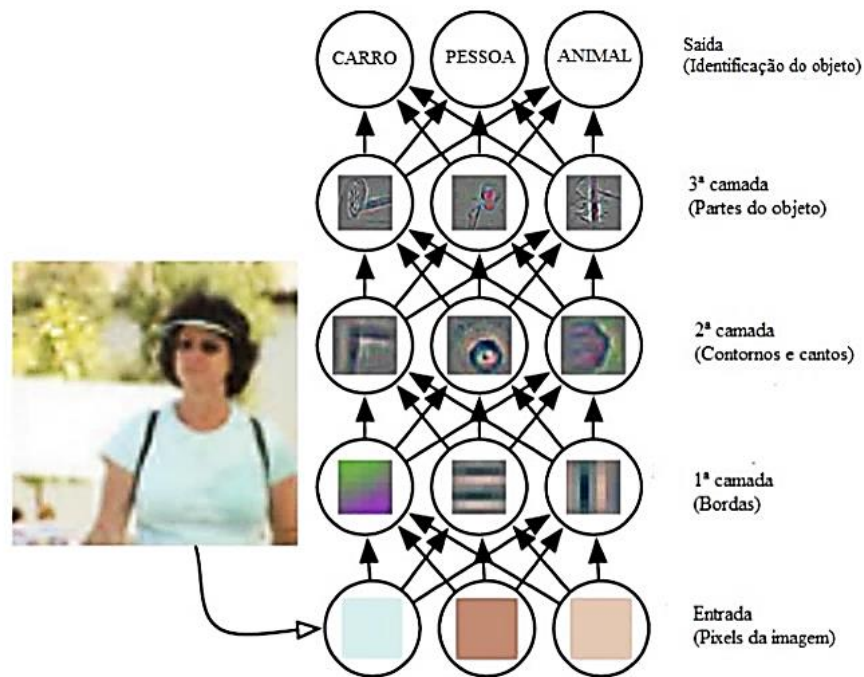


Figura 7: Representação das camadas de uma RNC. Fonte: Adaptado de Goodfellow e contribuintes (2016)

Essas redes se baseiam na forma como as informações são processadas no córtex visual dos mamíferos. No qual um conjunto de neurônios são compostos por pequenos campos receptivos locais, em que cada campo reage somente a um estímulo localizado em região limitada do campo visual. Assim, quando um estímulo visual chega à retina, cada neurônio é responsável por identificar características específicas de uma sub-região da imagem correspondentes a este estímulo. Em particular, neurônios das camadas iniciais são responsáveis por detectar formas geométricas simples, como linhas horizontais, linhas verticais, bordas e contornos, enquanto que neurônios nas camadas finais têm a atribuição de detectar formas gráficas mais complexas, compostas das formas gráficas simples detectadas por camadas anteriores. Desta forma, cada conjunto de neurônios combina características detectadas pela camada imediatamente anterior para formar outras mais complexas até, finalmente, chegar em um nível de abstração alto, como reconhecimento ou classificação de objetos (GÉRON, 2017a). A Figura 8 ilustra o que foi dito.

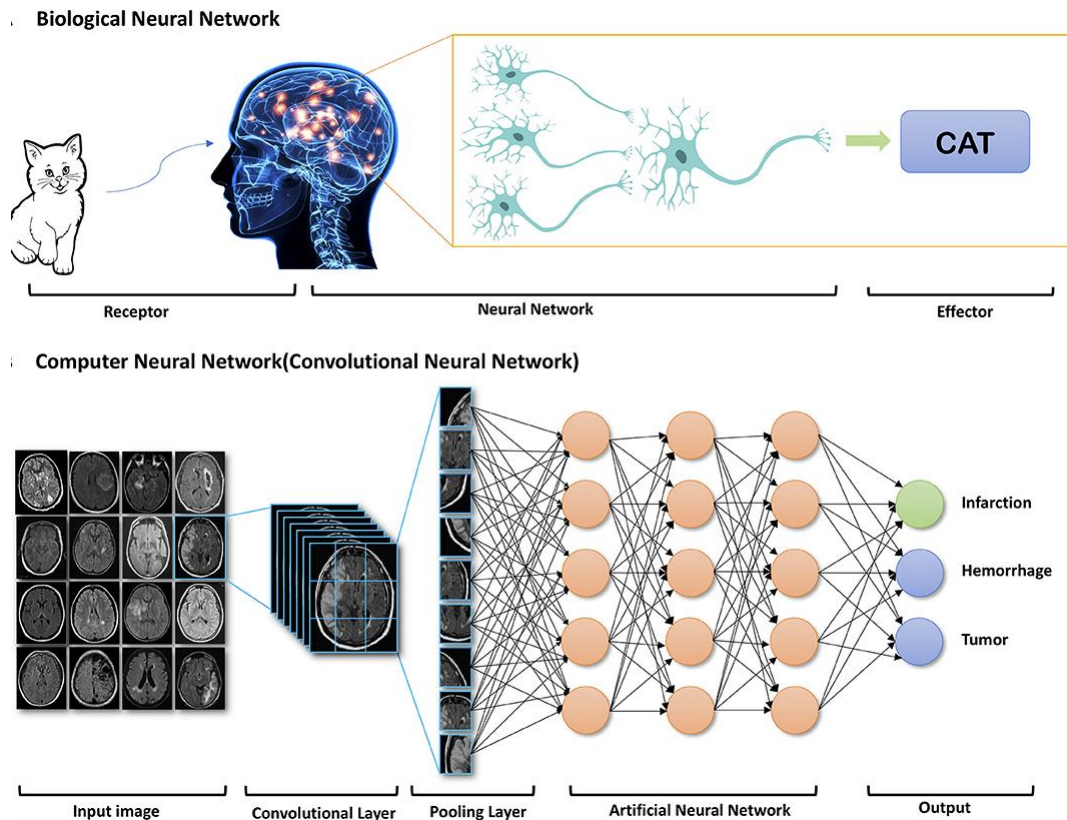


Figura 8: Comparativo do comportamento do cérebro humano e as RNCs. Fonte: (ZAHARCHUK et al., 2018)

Um estudo sobre as camadas das RNCs que foram aplicadas nas arquiteturas propostas neste trabalho é feito a seguir.

3.2.1 Camada de convolução

A camada que dá nome as RNC é de fato a mais importante, pois é responsável por criar mapas de características a partir dos dados de entrada. Esta camada aplica o processo de convolução de duas dimensões através da imagem. Na Figura 10 apresenta-se um exemplo do funcionamento da operação de convolução para duas dimensões, na qual, mostra-se que a dimensão da saída pode mudar, dependendo do tamanho do *kernel*. No entanto, é possível controlar esse parâmetro da camada convolutiva, mediante a variação do hiperparâmetro passo (*stride*). E caso deseje-se manter as dimensões de entrada, uma técnica de preenchimento como o *padding* pode ser usada. O *padding* consiste basicamente em preencher os pixels de borda da imagem com um determinado valor, como o zero.

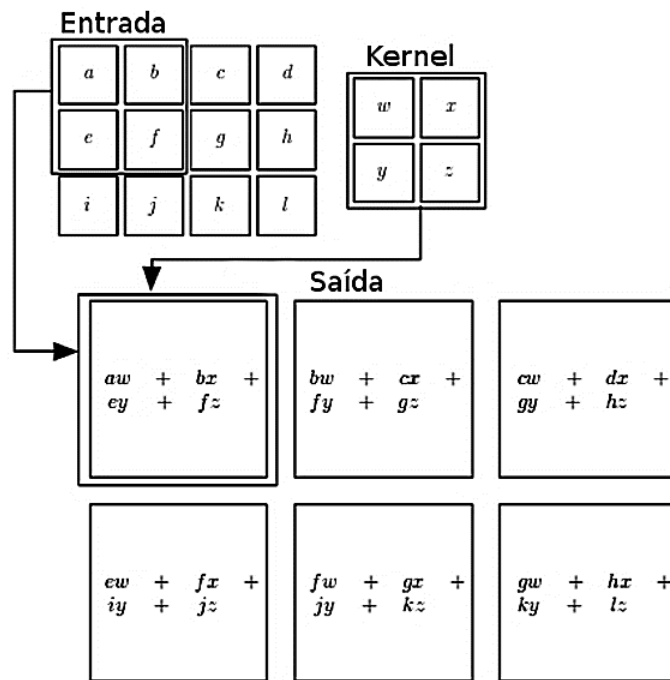


Figura 9: Exemplo de convolução bidimensional, em que a saída é obtida sem *padding*. Observe que as dimensões da entrada são de 3x4, enquanto que a saída tem dimensões 2x3. Fonte: Adaptado de Goodfellow et al. (2016)

A operação de convolução bidimensional aplicada nas RNC. No qual K representa a máscara, I a imagem, m largura da máscara e n altura da máscara, como é mostrado na equação abaixo:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (3)$$

As entradas das redes são as próprias imagens, até mesmo imagens com mais de um plano de bit (RGB por exemplo). As camadas convolutivas extraem as características. Quanto mais camadas contém as redes mais informações específicas das imagens são extraídas. Computacionalmente as redes escolhem os melhores filtros, aprendendo sozinhas como extrair as características mais relevantes para fazer uma determinada tarefa, tal como segmentação ou classificação.

3.2.2 Camada de unidades Retificadoras Lineares (ReLU)

Conforme comentado anteriormente, ao final de cada camada há uma função de ativação. Esta função assegura que o valor de ativação f de cada neurônio seja passado para a próxima camada a partir de sua entrada x . Essa função não linear pode ser de várias formas (Figura 4). O desempenho da rede depende da função de ativação escolhida.

Em arquiteturas amplamente utilizadas, como ImageNet (Krizhevsky et al., 2012) cada camada conectada é seguida da função de ativação ReLU. Isso se deve ao fato de esta função, no caso específico das camadas convolutivas, reportar melhores resultados segundo o estado da arte. A equação abaixo mostra a função ReLU:

$$f(x) = \max(0, x) \quad (4)$$

Além de melhores resultados, esta função se destaca pelo baixo custo computacional, em comparação com outras como sigmoide ou tangente hiperbólica. A derivada dessa função é 1 quando x é diferente de zero e 0 caso contrário, fazendo com que o número de cálculos necessários seja menor e, desta forma, simplificando o algoritmo utilizado para treinar a rede (*Backpropagation*). Como as RNCs são modelos com muitas conexões, o impacto da utilização desta função é considerável, tornando-as mais rápidas em tempo de treinamento. Outro fator a se destacar, a respeito, é a capacidade de evitar o desvanecimento do gradiente graças a não saturação da função para valores positivos. Uma das desvantagens dessa função é que, durante o processo de treinamento, alguns neurônios efetivamente morrem, especialmente sob circunstâncias na qual o valor empregado para a taxa de aprendizado é muito alto (GÉRON, 2017b).

3.2.3 Camada de *pooling*

Uma prática comum nas RNC é ter o uso de sucessivas associações de camadas convolutivas, seguidas de camadas ReLU e *Pooling*, nessa ordem. Basicamente a função da camada de *pooling* é reduzir as dimensões espaciais do volume de entrada, tornando a rede menos carregada de informações. Essa camada reduz a dimensionalidade aplicando um filtro com operações de média ou máximo, por exemplo, em determinadas porções da imagem. Veja o exemplo da Figura 10. O parâmetro passo e tamanho do filtro controlam o quão reduzida a imagem de saída se apresentará.

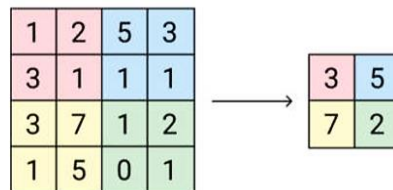


Figura 10: Resultado da aplicação *max pooling*. Observe que foi utilizado um passo (S) de 2 e um filtro (2,2), obtendo uma imagem de metade do tamanho da original.

Esta camada retorna os pesos mais representativos de uma região, tendo em vista que eles influenciarão mais na resposta final. Perceba que quanto maior o tamanho do filtro, menos

informações serão extraídas de uma sub-região, tendo maior perda de informação, o mesmo raciocínio vale para o passo. As equações abaixo dão uma noção mais ampla de como as imagens de entrada e saída se relacionam. Sendo o tamanho da dimensão (largura ou altura da imagem) de entrada (W_i), tamanho da dimensão de saída (W_o), tamanho do filtro (F) e passo (S). É importante ressaltar que a operação de *pooling* não altera em nada a profundidade da rede.

$$W_o = \frac{W_i - F}{S} + 1 \quad (5)$$

3.2.4 Camada de *Batch Normalization*

Conforme será discutido mais a frente, na Seção 3.3, durante o treinamento das redes convolutivas existe um problema conhecido como devaneio ou explosão do gradiente, que acontece pelo fato de, as vezes, o conjunto de dados ser demasiadamente grande e a atualização do próximo peso depender de um fator multiplicativo que cresce de acordo com o conjunto de dados. Essa camada também é útil para o problema de distribuição que existe quando são mudadas as entradas durante o treinamento, o que os autores chamam de “mudança de covariância interna”.

A camada de *Batch Normalization* (BN) corrige esses problemas através de uma etapa de normalização que corrige a média e a variância da saída da camada anterior. Isso é feito adicionando uma operação pouco antes da função de ativação em cada camada, simplesmente centralizando em zero e normalizando a entrada, e então escalonando-a através do uso de dois novos parâmetros. Com isso o *batch normalization* permite taxas de aprendizado maiores e contribui na normalização do modelo, minimizando a necessidade da camada de *dropout* (SZEGEDY et al., 2015). Adicionalmente, essa camada tende a minimizar o *overfitting*, que ocorre quando o modelo se adapta demais aos dados de treinamento e não consegue uma boa generalização no conjunto de dados de validação.

O BN inclui dois parâmetros γ e β a serem aprendidos durante o treinamento, adicionando então mais complexidade ao modelo. No Python, ao se utilizar camadas de *Batch normalization* é comum verificar que a rede apresenta o dobro de parâmetros de entrada que são não treináveis. A saída torna-se:

$$y = \gamma \hat{x} + \beta \quad (6)$$

sendo \hat{x} a entrada normalizada. Para calcular \hat{x} em um *minibatch* de tamanho m , $B = \{x_1, x_2, \dots, x_m\}$, calcula-se primeiramente a média, μ_B e a variância do *minibatch*, σ_B^2 .

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (7)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (8)$$

$$\hat{x}_i = \frac{(x_i - \mu_B)}{\sqrt{\sigma_B^2 + \epsilon}} \quad (9)$$

onde ϵ tem a função de evitar a divisão por zero.

3.2.5 Camada de *Dropout*

Nas redes convolucionais há uma grande preocupação com o *overfitting*. Buscando evitá-lo a camada de *dropout* é utilizada. Esta técnica foi introduzida por Srivastava e colaboradores (2014) e vem apresentando grande sucesso em trabalhos da área por se tratar de uma técnica simples e efetiva, como mostra a Figura 11. Ela consiste em desativar randomicamente a saída de um neurônio e avaliar o desempenho da rede, associando assim para cada neurônio uma probabilidade p de desativação. O termo desativar, não significa retirar o neurônio, e sim mudar o seu valor de saída para zero. Sendo assim, essa camada não reduz a dimensionalidade do modelo, a mesma também não será uma camada de neurônios e sim uma operação realizada nos neurônios já existentes. As unidades desativadas ainda estão presentes, porém deixam de participar da rede.

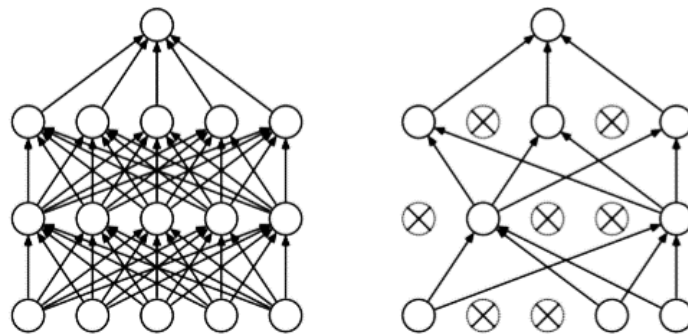


Figura 11: Exemplo da ação da camada de *dropout*. A rede da direita é o resultado da aplicação do *dropout* na rede da esquerda. Fonte: (SRIVASTAVA et al., 2014)

Os *dropouts* são definidos em porcentagem, significando que uma dada porcentagem de neurônios da camada anterior será desativada. Um valor padrão utilizado na literatura para o *dropout* é de 20%. Porém, esses valores podem mudar de arquitetura para arquitetura.

3.2.6 Camada de convolução transposta

Conhecida também como camada de deconvolução, ela aplica uma operação que funciona de forma contrária a convolução, utilizando o processo de sobreamostragem, vide a Figura 12. Podemos deconvoluir uma imagem (I_2) obtendo sua vetorização ($\text{vec}(I_2)$) e multiplicando uma matriz (A) onde cada linha representa uma convolução com o *kernel* K , resultando no vetor E , que após o processo inverso de vetorização se transforma na imagem sobreamostrada. Para ilustrar o processo um exemplo será mostrado abaixo:

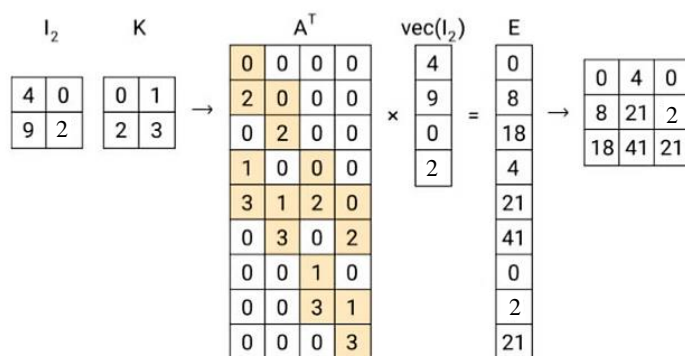


Figura 12: Convolução transposta da imagem I_2 pelo *kernel* K .

Observe que esse processo da Figura 12 obteve uma imagem de saída maior que a de entrada. Essa propriedade é muito usada em RNC na tarefa de segmentação, no qual a imagem vai perdendo dimensionalidade pelo uso de camadas de *pooling* e no final deseja-se obter uma imagem do mesmo tamanho da entrada, porém segmentada.

3.2.7 Segmentação semântica

A tarefa principal de uma RNC é fazer classificações, não é de amplo conhecimento que essas arquiteturas também podem desempenhar a tarefa de segmentação, basta encarar a segmentação como uma classificação pixel a pixel.

Inicialmente havia uma tentativa de adaptação de arquiteturas convolucionais já consagradas, como AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) ou GoogLeNet (SZEGEDY et al., 2015) para fazer a classificação, substituindo a camada final da rede por uma camada denominada “*pixelwise prediction*” como mostra a Figura 13. Porém essa abordagem

não mostrou grandes desempenhos, pois havia uma perda muito grande na dimensionalidade, e a técnica tinha por exemplo a incapacidade de classificar elementos pequenos, devido ao processo limitado de sobreamostragem (NOH; HONG; HAN, 2015).

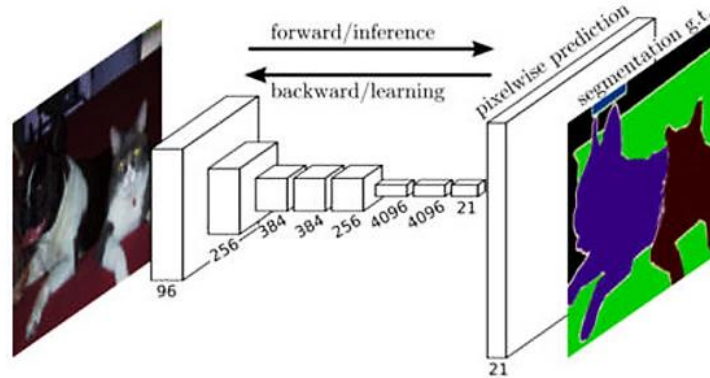


Figura 13: Exemplo de uma RNC para segmentação semântica. Fonte: (LONG; SHELHAMER; DARRELL, 2015)

Porém o processo é mais eficaz quando feito da forma apresentada na Figura 14. É possível observar, nesta figura, que na primeira metade da rede ela diminui a dimensionalidade da entrada, essa primeira parte é chamada em aprendizado de máquina de *encoder* (por obter um código reduzido da imagem de entrada). Conforme são utilizadas camadas de convolução transposta a rede ganha dimensionalidade até chegar ao tamanho da imagem de entrada. Esse processo de aumento da dimensão é conhecido como *decoder*.



Figura 14: Processo de segmentação semântica. Fonte: (MATHWORKS, 2018)

Observe que na Figura 14 diferentemente da Figura 13, o processo de sobreamostragem não é feito por apenas uma camada, e sim por um conjunto de camadas. Além disso o número de camadas utilizadas no *encoder* é o mesmo do *decoder*.

3.2.8 Treinamento de uma RNC

Assim como outras técnicas de aprendizado de máquina, o treinamento de uma RNC é supervisionado, ou seja, no treinamento são apresentados pares de entrada e saída desejada. O treinamento é baseado no algoritmo de *Backpropagation*. Nele, quando uma entrada é apresentada à rede, esta faz uma predição da saída e caso haja diferença entre o predito e o desejado, os parâmetros internos da rede sofrem modificações, na tentativa de minimizar o erro.

O *Backpropagation*, se baseia do gradiente descendente, e recebe esse nome pela sua forma de treinamento, que ocorre em sentidos opostos. No primeiro sentido *forward* (para frente), a entrada é apresentada a rede e ela prediz a saída. No segundo *backward* (para trás), se a saída não for a desejada os pesos das conexões são ajustados e então o processo se repete até que o erro seja minimizado ou outro critério de parada seja atingido.

Uma função de erro muito utilizada é o erro médio quadrático (MSE) E , onde $y_k(x_k; w)$ é o valor da saída predito através das entradas x_k e a matriz de pesos w , o termo t_{nk} é referente ao termo para rótulo no inglês (*target*) e se refere aos valores:

$$E(w) = \sum_{n=1}^N E_n(w) = \sum_{n=1}^N \sum_{k=1}^C (y_k(x_n; w) - t_{nk})^2 \quad (10)$$

Segundo Bishop (2006), de forma resumida, o algoritmo de *backpropagation* é dado seguindo os passos abaixo:

1. Encontrar as ativações de todas as unidades escondidas e de saída aplicando um vetor de entrada x_n a rede;

2. Encontrar o valor de δ_k para todas as unidades utilizando:

$$\delta_k = y_k - t_k \quad (11)$$

3. Obter δ_j de cada unidade escondida através do *backpropagation* dos δ através da equação:

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k \quad (12)$$

4. Obter as derivadas necessárias utilizando:

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i \quad (13)$$

O algoritmo distribui a contribuição do erro para cada peso na rede, levando em consideração o gradiente local da função de erro (GÉRON, 2017b). Como as funções de ativação de cada neurônio são deriváveis, o algoritmo utiliza-se da regra da cadeia para calcular

a derivada do erro com relação a cada parâmetro da rede tendo apenas o valor do erro na última camada. Os métodos de otimização são todos baseados neste algoritmo.

Observe que o método de treinamento não é tão simples, e devido ao enorme número de pesos na rede ele pode ser demorado. Normalmente essas redes tentam minimizar uma superfície de erro em uma dimensão muito grande. No qual o erro é função de um conjunto grande de imagens que precisam ser apresentadas a rede. Atualmente devido ao crescimento da área de games é possível um treinamento mais eficaz, com o uso das GPUs (*Graphics Processing Unit*). Mas esta etapa ainda representa um desafio de recursos computacionais até o ano do presente trabalho, tanto que geralmente as entradas das RNC são limitadas a imagens de dimensões pequenas. Por outro lado, embora o treinamento seja demorado, uma vez treinada a rede, a sua predição é feita de forma rápida e precisa, levando algumas dezenas ou centenas de milissegundos para obter o resultado.

3.3 Métodos de Otimização

Os métodos de otimização são os algoritmos de treinamento da rede que minimizam a função de custo (erro da rede). Conforme comentado eles são variações e melhoramentos do *backpropagation*. É comprovado que um determinado conjunto de dados pode obter melhores resultados com um método de otimização específico, não existindo assim um que seja explicitamente melhor, e sim os que se adaptaram melhor à sua problemática. Será comentado um pouco dos métodos SGDM (*Stochastic Gradient Descent With Momentum*), RMSProp (*Root Mean Square Propagation*) e ADAM (*Adaptive Moment Estimation*). Esses três métodos foram escolhidos por serem os mais utilizados no estado da arte. É importante entender este parâmetro de treinamento das RNC e como o mesmo interfere nos resultados obtidos. A escolha do algoritmo de treinamento pode significar uma diferença de bons resultados em minutos, horas e dias.

3.3.1 Stochastic Gradient Descent with Momentum (SGDM)

O Gradiente estocástico com Momentum (SGDM) é uma variação do Gradiente descendente estocástico (SGD), sendo este último um dos algoritmos mais utilizados em aprendizado de máquina e aprendizado profundo em geral. O SGD é normalmente empregado quando há uma base de dados extensa, pois, ele permite a utilização de subconjuntos (*mini batches*), já que carregar todas as imagens na memória de uma única vez seria inviável.

Uma das desvantagens do SGD é o seu tempo de convergência demorado. Então, com o intuito de acelerar este processo foi proposto o SGDM, que adiciona uma variável que controla a velocidade (V) na qual os parâmetros se movem através do espaço de parâmetros (\bar{W}), que nada mais é que uma matriz de pesos dos neurônios (GOODFELLOW; BENGIO; COURVILLE, 2016).

As equações 13 e 14 representam como a atualização do vetor de parâmetro \bar{W} é realizada com base nessas variáveis:

$$\bar{V} \leftarrow \beta \bar{V} - \alpha \frac{\partial E}{\partial \bar{w}} \quad (13)$$

$$\bar{W} \leftarrow \bar{W} + \bar{V} \quad (14)$$

em que, E é a função de custo e α é a taxa de aprendizado, V são os pesos iniciais que iniciam com o valor zero e são modificados com suavização exponencial, determinado por um parâmetro de suavização β (que pode variar entre 0 e 1, sendo 0,9 um valor padrão).

A Figura 15 representa como ocorre de forma visual esse melhoramento, ilustrando a superioridade em tempo de processamento do SGDM.

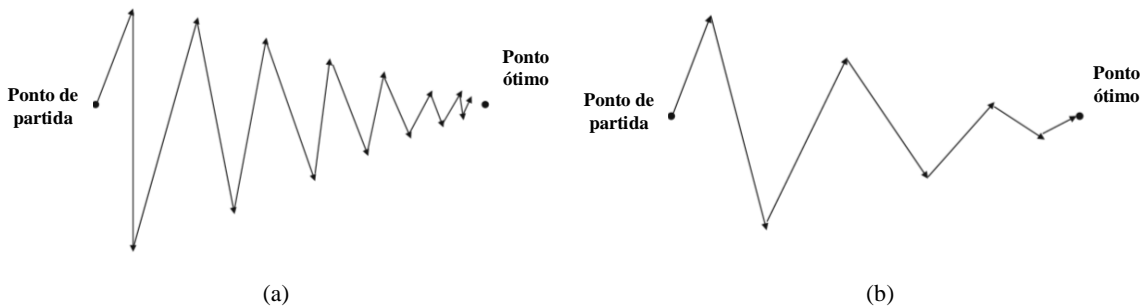


Figura 15: Efeito do momento na suavização de oscilações: (a) método SGD (b) método SGDM. Fonte: (AGGARWAL, 2018)

3.3.2 Root Mean Square Propagation (RMSProp)

O RMSProp é uma outra proposta para a convergência com menos oscilações. As equações 15 e 16 modelam o RMSProp. A ideia básica deste algoritmo é acumular somente os gradientes das iterações mais recentes, por meio da introdução de um hiperparâmetro β que representa um decaimento exponencial, o símbolo \otimes representa a multiplicação ponto a ponto. Esta forma vetorizada é equivalente ao cálculo $s_i \leftarrow s_i + (\partial/\partial\theta_j J(\theta))^2$ para cada elemento s_i do vetor s , em outras palavras, cada valor s_i acumula o quadrado da derivada parcial da função de custo em relação ao peso θ_i . A equação 16 representa a atualização dos pesos, quase idêntica ao Gradiente Descendente, a não ser pelo fato que o vetor gradiente é dividido por um fator $\sqrt{(s + \epsilon)}$ onde o símbolo \oslash representa a divisão ponto a ponto, e ϵ representa uma constante para evitar a divisão por zero. Essa forma vetorizada é

equivalente ao cálculo $\theta_i \leftarrow \theta_i - \alpha \partial/\partial\theta_i J(\theta) / \sqrt{(s_i + \epsilon)}$ para todo parâmetro θ_i , simultaneamente (GÉRON, 2017a).

$$s \leftarrow \beta s + (1 - \beta) \nabla_{\theta} J(\theta) \otimes \nabla_{\theta} J(\theta) \quad (15)$$

$$\theta \leftarrow \theta - \alpha \nabla_{\theta} J(\theta) \oslash \sqrt{s + \epsilon} \quad (16)$$

Esse método de otimização é baseado em outro, o AdaGrad. Enquanto o AdaGrad reduz a taxa de aprendizado de acordo com o acúmulo do gradiente ao quadrado, o RMSProp usa uma média exponencialmente ponderada para realizar o acúmulo do gradiente, de forma que valores históricos mais antigos da trajetória são descartados, tendo em vista que são atribuídos pesos menores aos dados mais antigos.

Na maioria dos casos o RMSProp se mostra superior ao AdaGrad, perdendo apenas quando os problemas são muito simples (GÉRON, 2017b).

3.3.3 Adaptive Moment Estimation (ADAM)

O método ADAM, é uma extensão do SGD. Ele foi apresentado por Diederick Kingma (2015) e é um dos algoritmos mais utilizados para o treinamento de redes neurais profundas. Segundo o autor o algoritmo apresenta uma série de vantagens em relação ao SGD:

- Adequado para problemas com gradientes muito ruidosos / ou esparsos;
- Computacionalmente eficiente;
- Requer pouca memória;
- Invariante a escala diagonal de gradientes;
- Hiperparâmetros têm interpretação intuitiva e geralmente requerem pouco ajuste;
- Adequado para problemas grandes em termos de dados e/ou parâmetros.

O ADAM acumula uma média de decaimento exponencial dos gradientes passados. Como RMSProp, acumula uma média de decaimento exponencial do quadrado dos gradientes de iterações mais recente. Onde β_1 e β_2 representam as taxas de decaimento para a estimativa dos momentos, m o momento, $\nabla_{\theta} J(\theta)$ é o gradiente da função de custo e θ o custo, s é a atualização da estimativa tendenciosa do primeiro momento, m_t serve para calcular a estimativa do primeiro momento corrigida pelo viés, como mostram as equações a seguir:

$$m \leftarrow \beta_1 m + (1 - \beta_1) \nabla_{\theta} J(\theta) \quad (17)$$

$$s \leftarrow \beta_2 s + (1 - \beta_2) \nabla_{\theta} J(\theta) \otimes \nabla_{\theta} J(\theta) \quad (18)$$

$$m_t \leftarrow \frac{m}{1 - \beta_1^T} \quad (19)$$

$$s \leftarrow \frac{s}{1 - \beta_2^T} \quad (20)$$

$$\theta \leftarrow \theta - m_t \oslash \sqrt{s + \epsilon} \quad (21)$$

Os autores (KINGMA; BA, 2014) demonstraram o ADAM sendo superior aos métodos de otimização mais utilizados. O ADAM foi aplicado em classificação, em análise de sentimentos e com redes convolucionais, com bases de dados bastante conhecidas. Na figura 16 é mostrado a superioridade do ADAM na tarefa de classificação, na base de dados MNIST (base de dados de números escritos a mão).

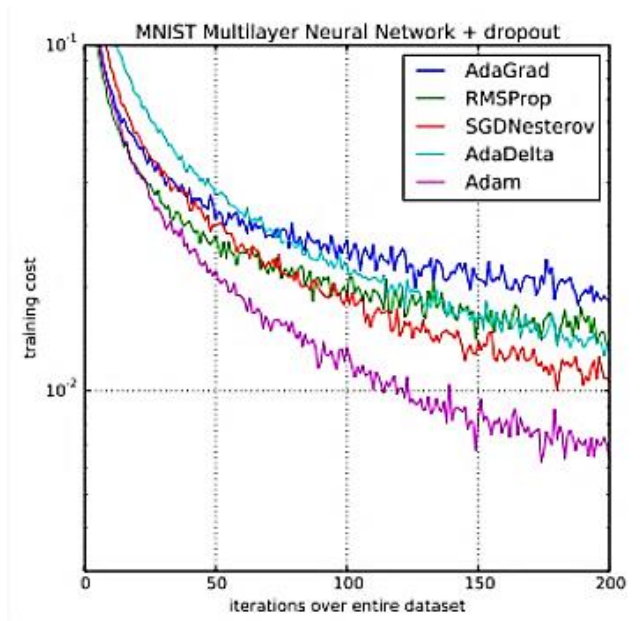


Figura 16: Comparação do ADAM com outros métodos de otimização. Fonte: (KINGMA; BA, 2014)

4 MATERIAIS E MÉTODOS

O trabalho visa o desenvolvimento de um conjunto de métodos que torne possível a segmentação automática de lesões mamárias em imagens de ultrassonografia digital. Na Figura 17, é apresentado o diagrama em blocos das etapas para o desenvolvimento do sistema proposto.

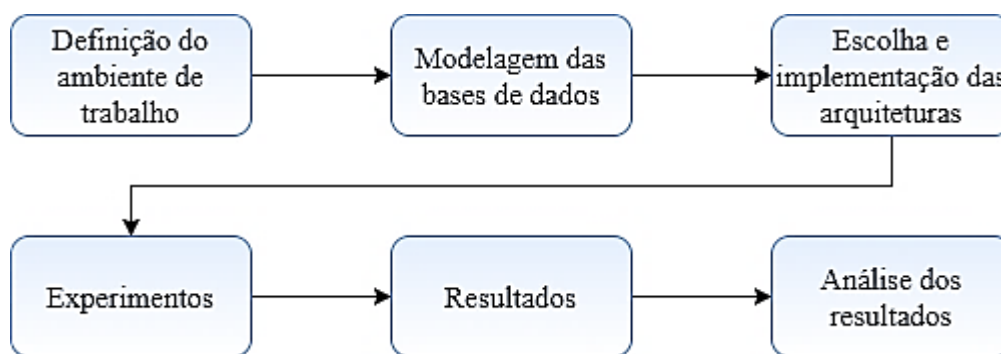


Figura 17: Diagrama em blocos da Metodologia utilizada.

4.1 Definição do ambiente de trabalho

Para o desenvolvimento deste trabalho foi utilizado um computador com sistema operacional Windows 10, um processador Intel® Core™ i7- 8700 @3.20 GHz 3.19 GHz, 16 GB de memória RAM e uma GPU GeForce 1070 GTX com 8GB de memória RAM e 1920 núcleos CUDA.

O software utilizado foi o MATLAB® R2019a com as *toolboxes*: *Computer Vision System*, *Neural Network*, *Parallel Computing and Machine Learning*, *Image Processing* e *Database Processing*.

4.2 Modelagem das bases de dados

Com o intuito de ser feito um *benchmark*, foram utilizados dois bancos de dados (BD), denominados de *dataset A* (BD-A) e *dataset B* (BD-B). Os dois bancos de dados apresentam algumas diferenças, a primeira é o *cropping* (procedimento de corte feito nas imagens). Enquanto BD-A continha somente imagens *cropped* (imagens previamente cortadas), BD-B tinha imagens completas. Outra diferença primordial é que as imagens do BD-A foram obtidas por digitalização das imagens de ultrassom, enquanto que no BD-B as imagens foram obtidas diretamente do aparelho de ultrassom. Os dois bancos de dados e suas particularidades são comentados a seguir. A Figura 18 ilustra exemplos de imagens de ultrassom de mama contidas na base de dados A.

A base de dados A contém 387 imagens de ultrassom de mama (US), com seus padrões ouro (GT, de *Ground True*), disponibilizadas por (GÓMEZ et al., 2010). Sendo deste total, 208 imagens com lesões benignas e 179 lesões malignas. Todas as imagens foram adquiridas durante a rotina de diagnóstico do câncer de mama, uma para cada paciente, no Instituto Nacional do Câncer (INCa), Rio de Janeiro, por equipamentos diversos. As imagens originais estão em escala de cinza e possuem diferentes tamanhos.

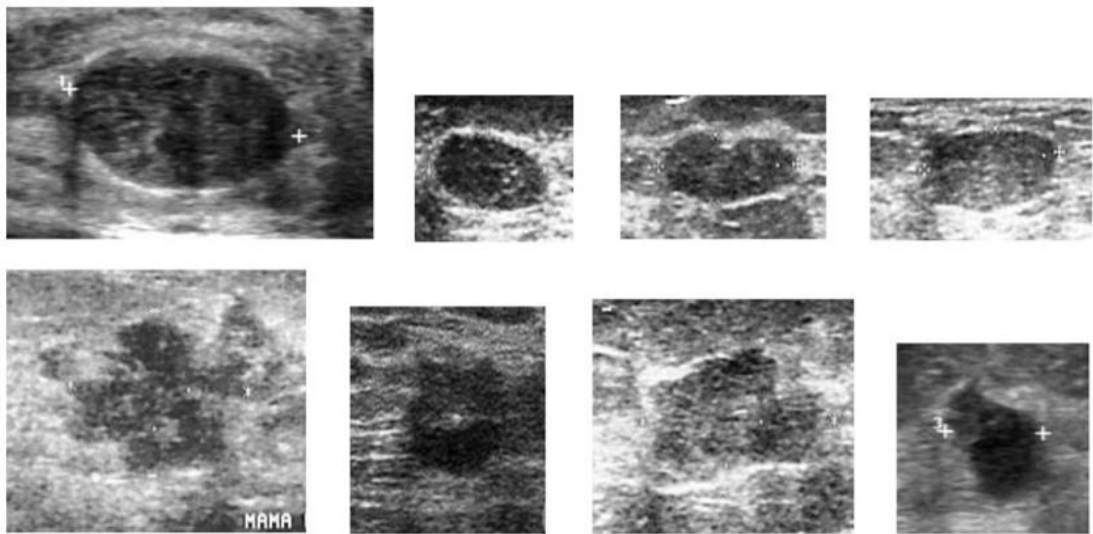


Figura 18: Imagens de US de mama de BD-A

A Figura 19 apresentam os respectivos padrões ouro ou máscaras (imagens binárias já segmentadas) das imagens de ultrassom de mama apresentadas na Figura 18.

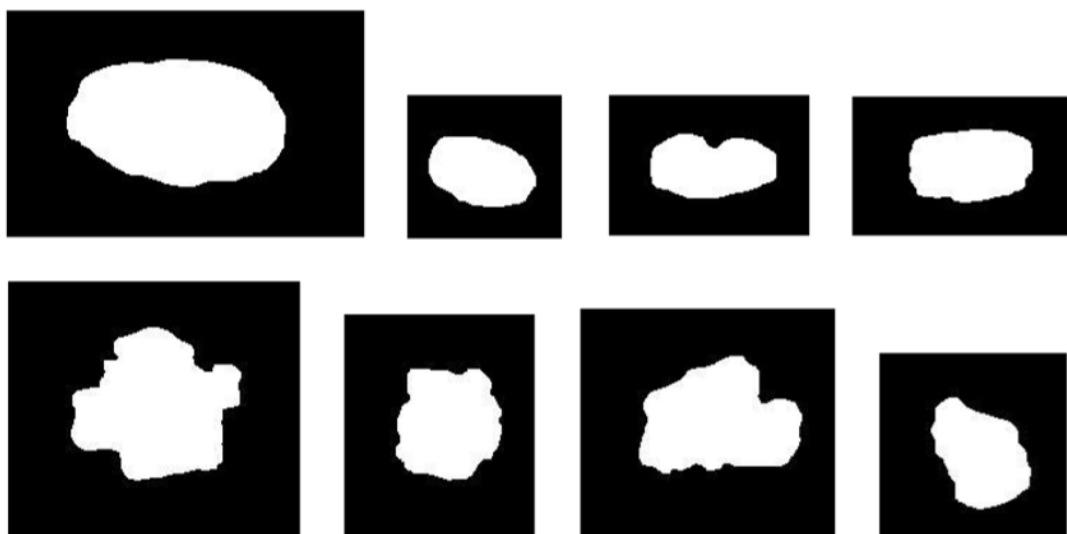


Figura 19: Máscaras (padrões ouro) das imagens apresentadas na Figura 18.

A segunda base de dados, BD-B, foi cedida por Yap et al. (2018). Foi coletada em 2012 em UDIAT (*Diagnostic Centre of the Parc Tauli Corporation*), em Sabadell (Espanha), com um aparelho de ultrassom *Siemens ACUSON Sequoia C512 system 17L5 HD* com 8.85 MHz. O BD-B consiste em 163 imagens de ultrassom das quais, 53 são imagens contendo lesões malignas e 110 com lesões benignas. As imagens também possuem diferentes tamanhos. A Figura 20 ilustra as diferenças entre as imagens do BD-A e BD-B.

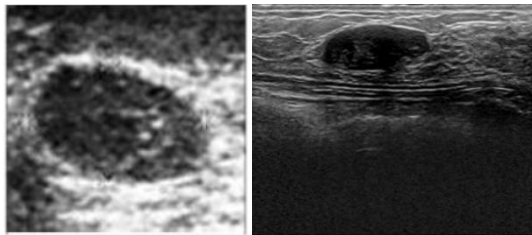


Figura 20: (a) Exemplo de uma imagem *cropped* do *Dataset A* e (b) Exemplo de uma imagem original do *Dataset B*.

Devido a limitações de hardware, todas as imagens de ambos os BDs foram redimensionadas para 160x160 ou para 320x320 pixels. Foram então construídos a partir desses dois BDs quatro outros subconjuntos (ver a Figura 21) que são: *Dataset A*: Subconjunto 1 com a imagem *cropped* reescalada para 160x160; *Dataset B*: Subconjunto 2 com a imagem *cropped* reescalada para 160x160, Subconjunto 3 com a imagem original reescalada para 160x160 e Subconjunto 4 com a imagem de 320x320 pixels sem *cropping*.

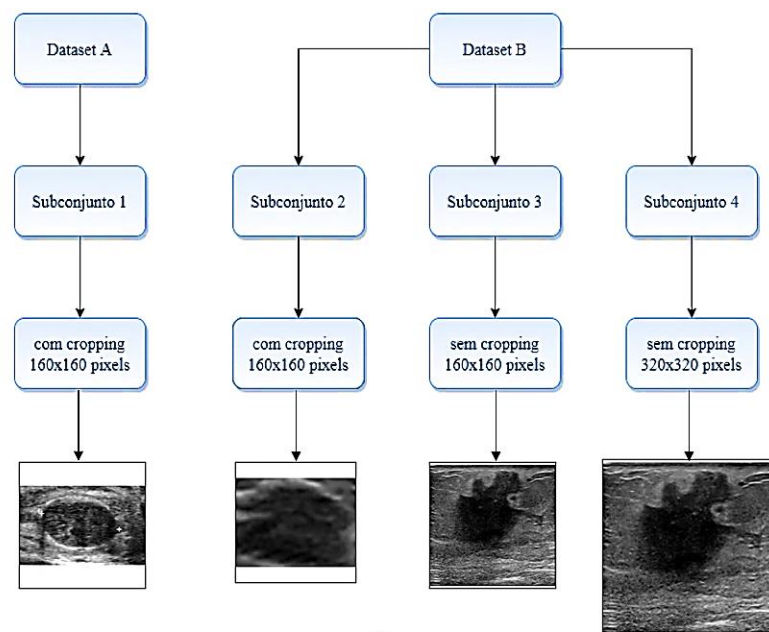


Figura 21: Adaptações para obtenção dos subconjuntos de dados do *Dataset A* e *Dataset B*.

O redimensionamento das imagens foi realizado por interpolação bilinear através da função `imresize()` e `padarray()`, disponíveis na toolbox de *Image Processing* do MATLAB®. Para manter a razão de aspecto das imagens, a função `imresize()` foi utilizada de forma que somente uma dimensão (altura ou largura) da imagem, a de maior escala, fosse redimensionada para 160. Em seguida, a dimensão de menor escala foi completada com *padding* de valor 255 (correspondente ao branco na escala de cinza) através da função `padarray()` de forma que a imagem obtivesse a dimensão final de 160x160.

O tamanho de 160x160 foi escolhido baseado no *Dataset A*, que apresentava em média imagens com essas dimensões. Porém as dimensões médias das imagens do BD-B não eram as mesmas, então fez-se esses dois testes com imagens 320x320 e 160x160. Observe que o *benchmark* só é possível quando as imagens são do mesmo tamanho.

Optou-se por ser feito o uso da técnica de validação cruzada com 5 pastas e separação de dados aleatória em ambos os BDs, buscando obter um desempenho médio mais fidedigno ao real. Foi escolhida uma subdivisão de 60-20-20. No qual, 60% dos dados foram para treinamento (232 imagens para BD-A e 97 imagens BD-B), 20% para validação e 20% para teste (77 imagens em BD-A e 33 imagens BD-A). A validação foi feita para avaliar a melhor arquitetura. O diagrama da Figura 22 deixa clara esta etapa da metodologia.

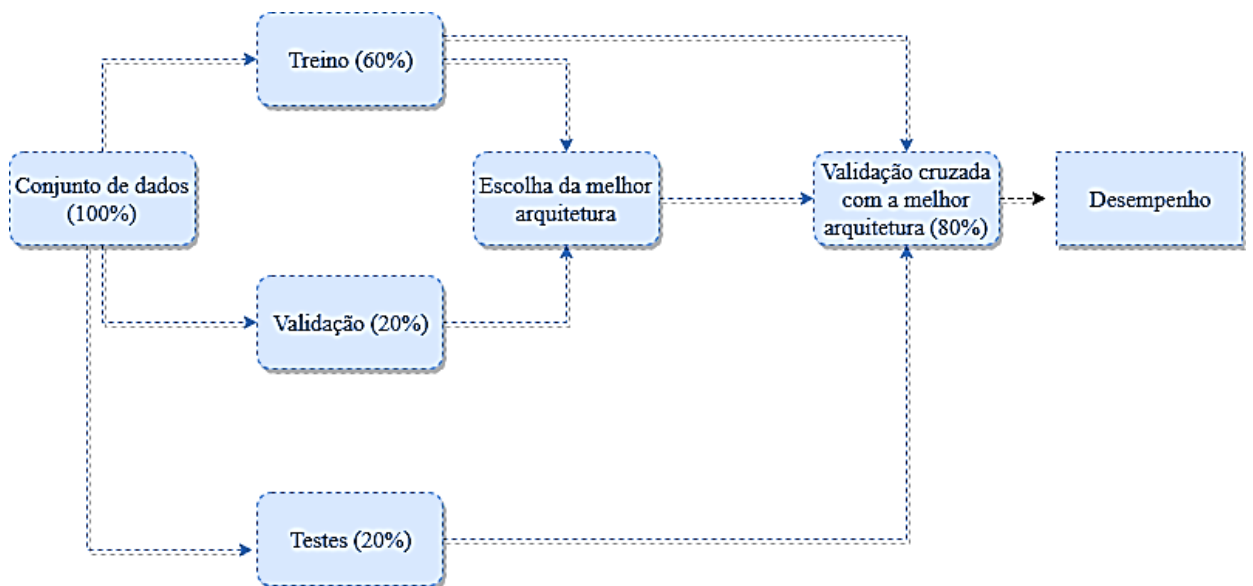


Figura 22: Divisão 60-20-20 para treino, validação e testes.

Os dois BDs já foram utilizados na literatura. Um subconjunto do BD-A, composto por 200 imagens, foi empregado por Infantosi e contribuintes (2008). Um outro subconjunto de BD-A com 50 imagens foi utilizado por Gomez (2009). Enquanto BD-B foi utilizado por Yap

e contribuintes (2018), porém não com o intuito de fazer a segmentação e sim a detecção dessas lesões.

4.3 Escolha e implementação das arquiteturas

Foram selecionadas quatro arquiteturas de redes convolucionais, sendo 1 sequencial e as outras 3 utilizando estruturas de grafos acíclicos, para a tarefa de segmentação de lesões mamárias em imagens de ultrassom.

4.3.1 Arquitetura 1 (CNN1)

A primeira rede (CNN1, do inglês *Convolutional Neural Network*) é uma arquitetura em série, o que significa dizer que a entrada de cada camada é a saída de uma camada prévia. Arquiteturas em série são utilizadas em estudos de Roth et al. (2015). A arquitetura CNN1 pura é utilizada para segmentação do pâncreas em imagens de tomografia computadorizada (CT), e Shelhamer et al. (2017) faz uso da CNN1 para segmentação semântica de objetos. A Figura 23 mostra a referida rede.

Essa arquitetura foi a mais simples com relação a construção e número de camadas. As seguintes camadas são utilizadas: Convolutiva (Conv), *Batch normalization* (BatchNorm), Unidade retificadora linear (ReLU), *Max pooling* (MaxPooling), *Deconvolution* (Deconv). No geral a arquitetura contém 11 camadas convolucionais. Os primeiros 10 campos receptivos têm 3x3 pixels e o último é de 1x1 pixel. Todas as camadas convolutivas, exceto a última tem 64 mapas de características, ou profundidade. Entre os dois primeiros módulos de sobreamostragem é introduzida a técnica de *Dropout* com a função de prevenir o *overfitting* em fase de treinamento. Por fim, após o módulo de sobreamostragem, vem o módulo de classificação que consiste na operação de convolução seguida da função *Softmax* e da classificação do pixel em si. A última camada convolutiva produz dois mapas de características, já que os pixels serão classificados em duas classes (0) fundo (*background*) e (1) lesão. Todos os pesos das camadas convolutivas foram inicializados de acordo com uma distribuição gaussiana de média zero e desvio padrão 0,01. Esta arquitetura possui 52 camadas.

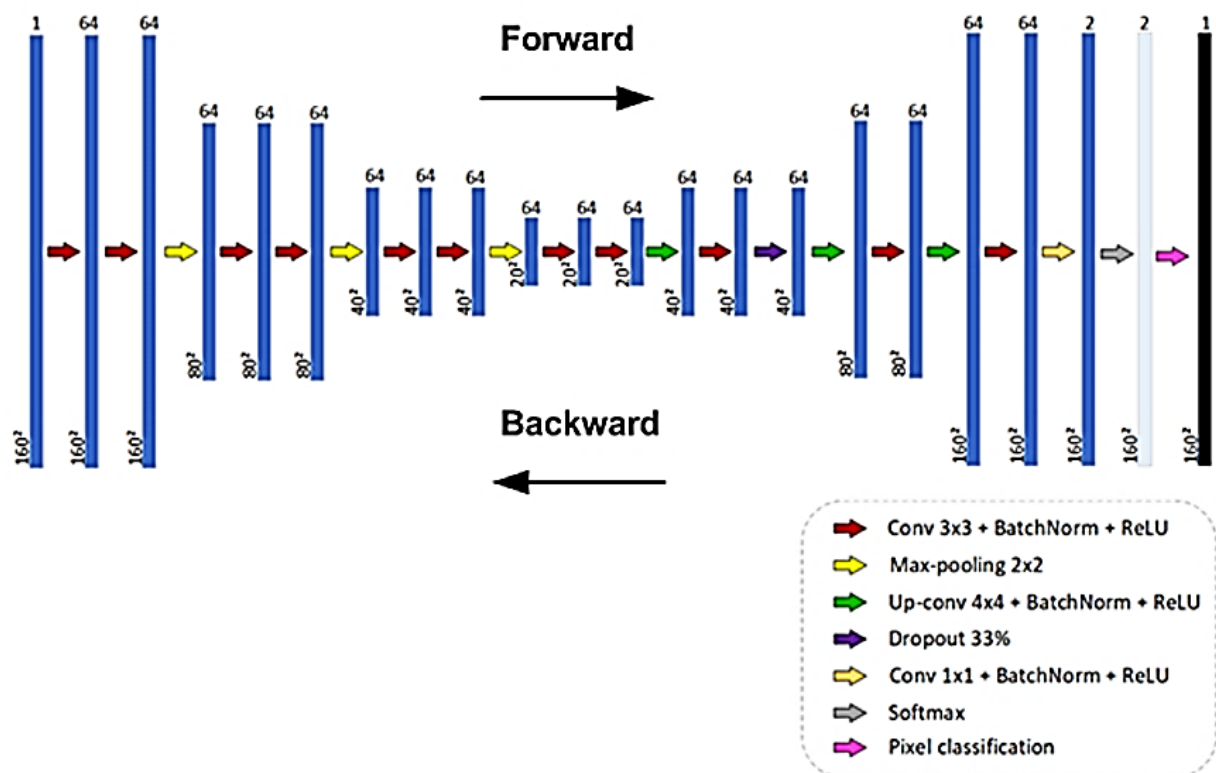


Figura 23: Arquitetura 1. Fonte: (MENDES et al., 2016)

4.3.2 Arquitetura 2 (CNN2)

Essa e as demais arquiteturas são conhecidas como arquiteturas de grafo acíclico direto (DAG). São mais complexas que as redes construídas de forma sequencial. Nas DAGs tanto as entradas quanto as saídas podem ser para camadas não subsequentes. Quando aplicadas no processamento de imagens, essas arquiteturas agregam informação da localização de pixels contidos desde as camadas iniciais até as finais. Em tarefas de segmentação semântica, é esperado que elas produzam melhoramentos refinados em detalhes da imagem. O estudo de Chen et al. (2016) usa arquiteturas DAG para segmentar estruturas neurais em Imagens de microscópio eletrônico.

Nesta segunda arquitetura, denominada CNN2 (vide Figura 24), a informação do primeiro módulo de subamostragem, depois da segunda operação de convolução, é agregada antes da operação de convolução do primeiro módulo de sobreamostragem. Para implementar a agregação, é necessário que a representação de dados em ambas as entradas seja de mesmo tamanho. Considerando que a dimensão do dado depois de duas operações de convolução da primeira camada é 160×160 e depois da deconvolução do primeiro módulo é 40×40 , uma

redução de dimensionalidade dessa primeira é necessária e isso foi feito com o uso *MaxPooling*. Esta rede possui 61 camadas.

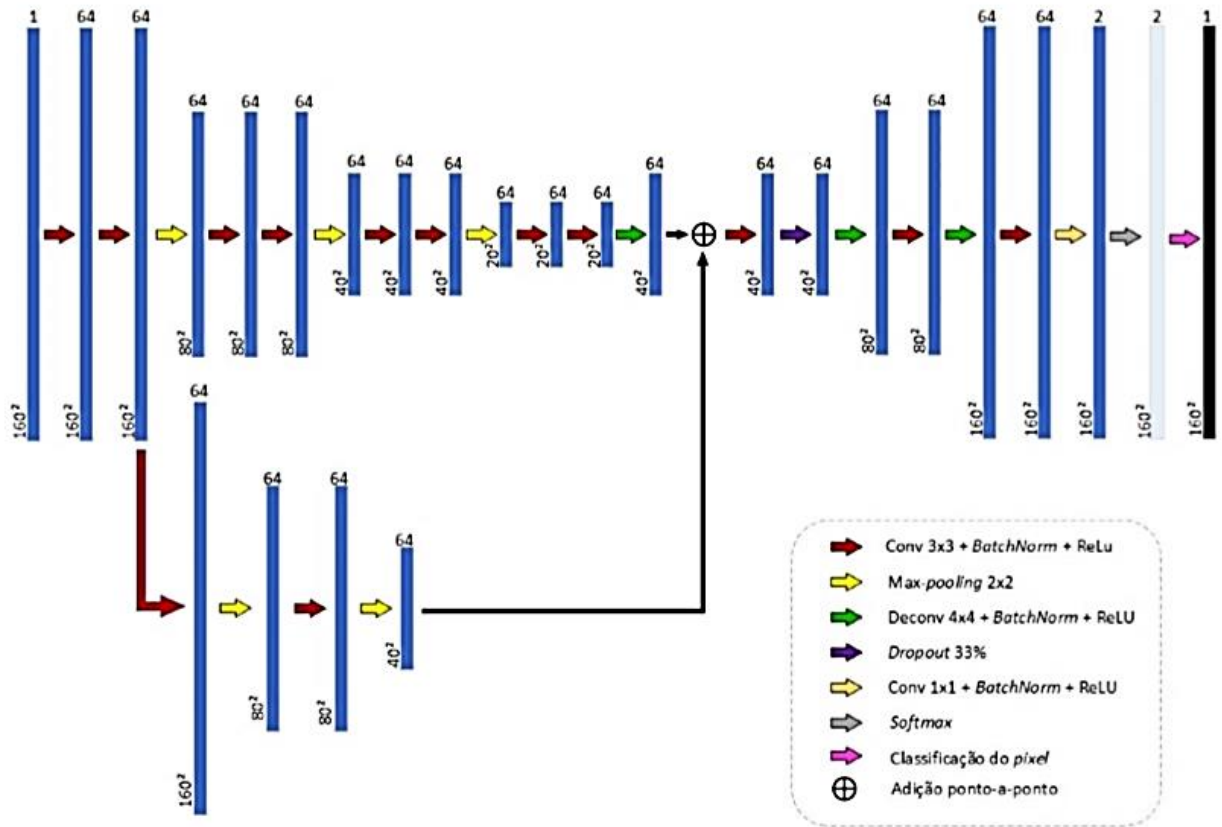


Figura 24: Arquitetura 2. Fonte: (MENDES et al., 2016)

4.3.3 Arquitetura 3 (CNN3)

Nessa arquitetura, denominada CNN3 (vide Figura 25), a informação de três módulos de subamostragem, depois da segunda camada de convolução de cada uma das Convs, é agregada depois da última camada de convolução, Conv 2. As operações de deconvolução utilizam um filtro 4x4 de profundidade 64 e S=2, enquanto que as operações de convolução utilizam um filtro 3x3 com profundidade de 64, S=1 e P=1. Esta rede possui 68 camadas.

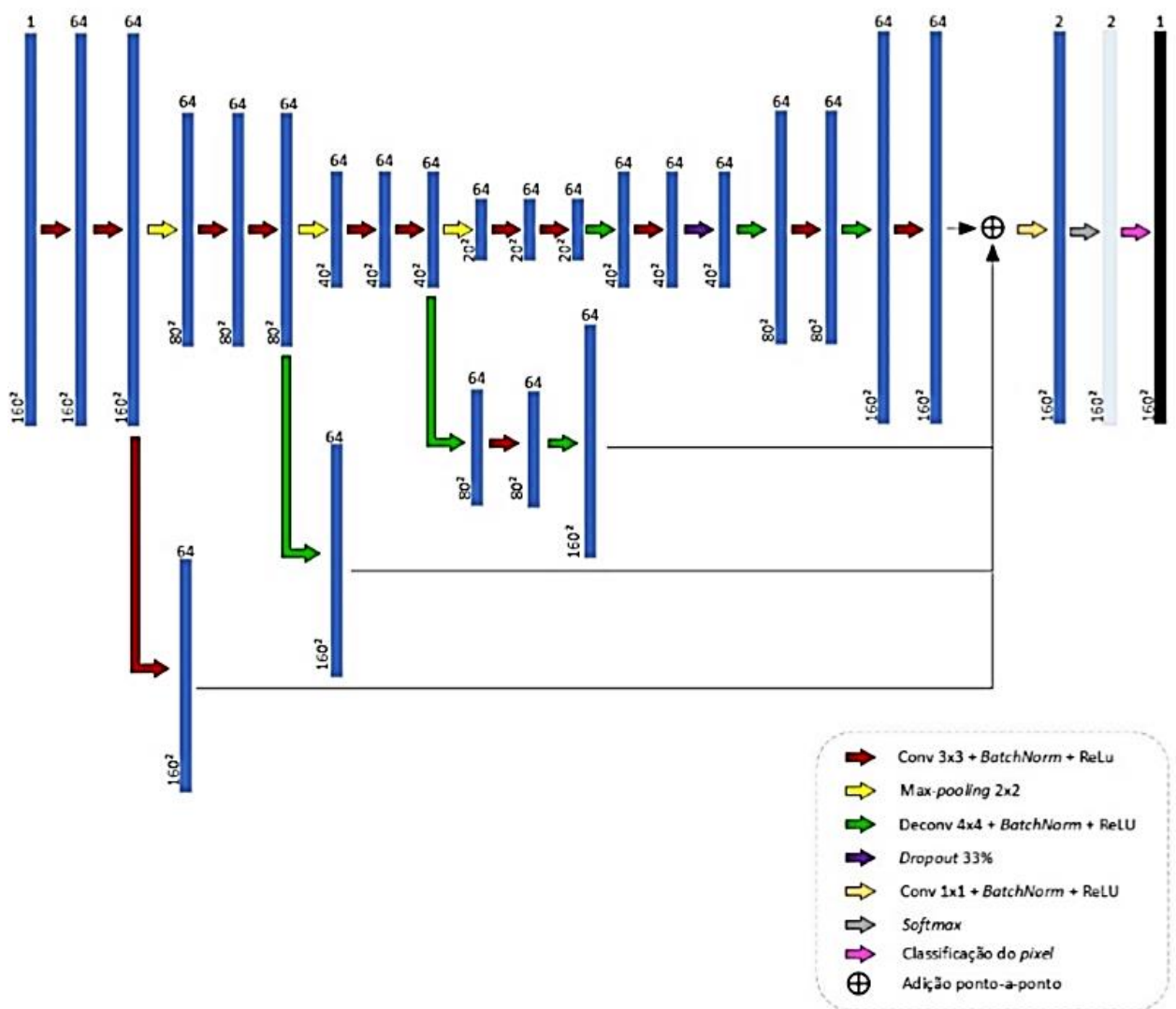


Figura 25: Arquitetura 3. Fonte: (MENDES et al., 2016)

4.3.4 Arquitetura 4 (CNN4)

Essa última arquitetura, conhecida como U-net (RONNEBERGER; FISCHER; BROX, 2015). Ela é muito utilizada para segmentação semântica de imagens médicas. Vários autores já demonstraram a utilidade dessa arquitetura com diferentes aplicações na área médica (JIN et al., 2019; RONNEBERGER; FISCHER; BROX, 2015; YAP et al., 2018). Possui cinco camadas *encoder* de profundidade, conforme mostra a Figura 26. Ela também possui estrutura DAG. Essa rede apresenta uma estrutura lógica, na qual o dado da última camada convolutiva do *encoder*, aparece na camada correspondente de deconvolução do *decoder*, antes do uso da *MaxPooling*. Isso se repete por toda a rede. Esta rede possui 70 camadas, se configurando então por ser a maior arquitetura, dentre as 4 testadas.

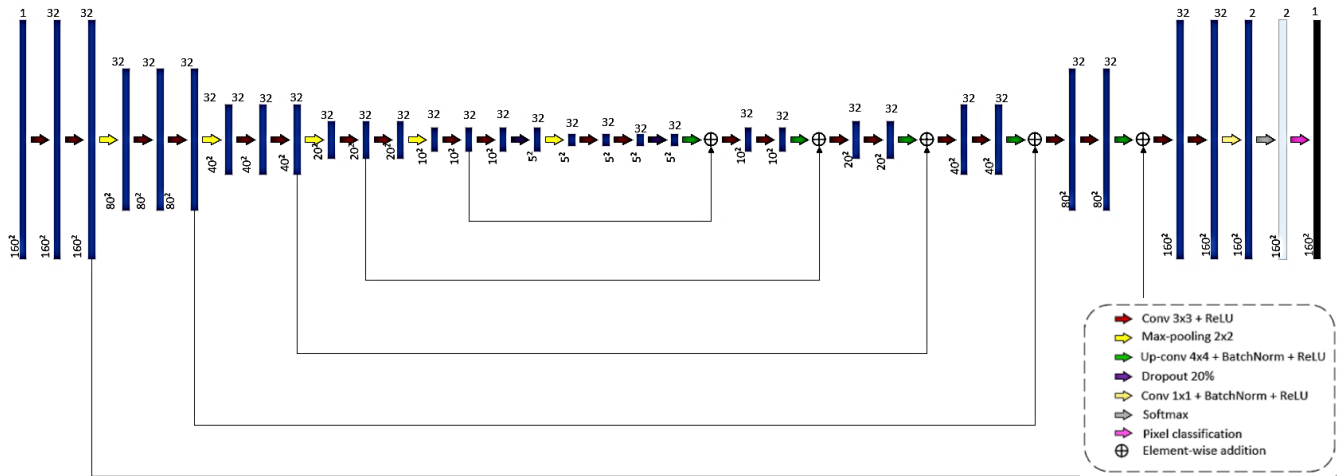


Figura 26: Arquitetura 4 proposta

4.4 Experimentos

Os experimentos tem início na modelagem e adaptação nas bases de dados, como explicado anteriormente, onde foram realizados os pré-processamentos necessários.

Nesta etapa os parâmetros internos das CNNs foram variados e avaliados de forma a obter-se os melhores resultados em cada experimento, com todos os quatro subconjuntos, na Figura 27 os dados em verde representam as medidas adotadas buscando melhor desempenho nos testes.

Com o intuito de comparar o desempenho de cada uma das arquiteturas propostas, foram utilizados os mesmos conjuntos de treinamento (as mesmas imagens de treino e validação para o *Dataset B*), bem como os mesmos hiperparâmetros internos das redes.

Na etapa de treinamento, foi utilizada a técnica de otimização ADAM com os hiperparâmetros padrões do MATLAB. Foi utilizado um *minibatch* de 8 imagens, para tornar o treinamento mais rápido em comparação com métodos tradicionais.

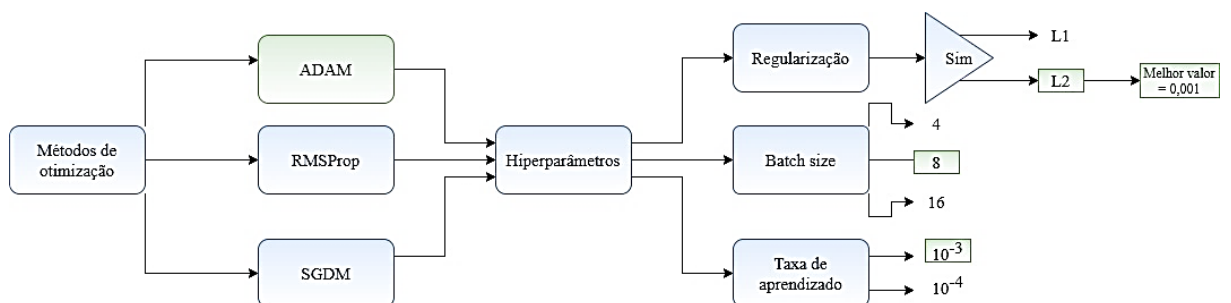


Figura 27: Diagrama dos experimentos realizados com cada arquitetura, de verde o que foi adotado no modelo proposto

Foi proposta e implementada uma “função de auxílio ao treinamento”, que treina a rede por 300 épocas e retorna o melhor desempenho obtido nas iterações realizadas, vide Figura 28. Esta função faz no final de cada iteração do treinamento um teste com os dados de validação, e então, a cada vez que o mesmo obtém um desempenho superior, um “*checkpoint*” do modelo é realizado. No final das épocas de treinamento, a função de auxílio substitui o modelo obtido no final das épocas pelo melhor modelo encontrado (ou seja, o modelo do *checkpoint*). Isso pode ser encarado como um tipo de parada antecipada, pois os pesos do modelo ficam congelados em uma região e os treinamentos seguintes não são considerados pela rede.

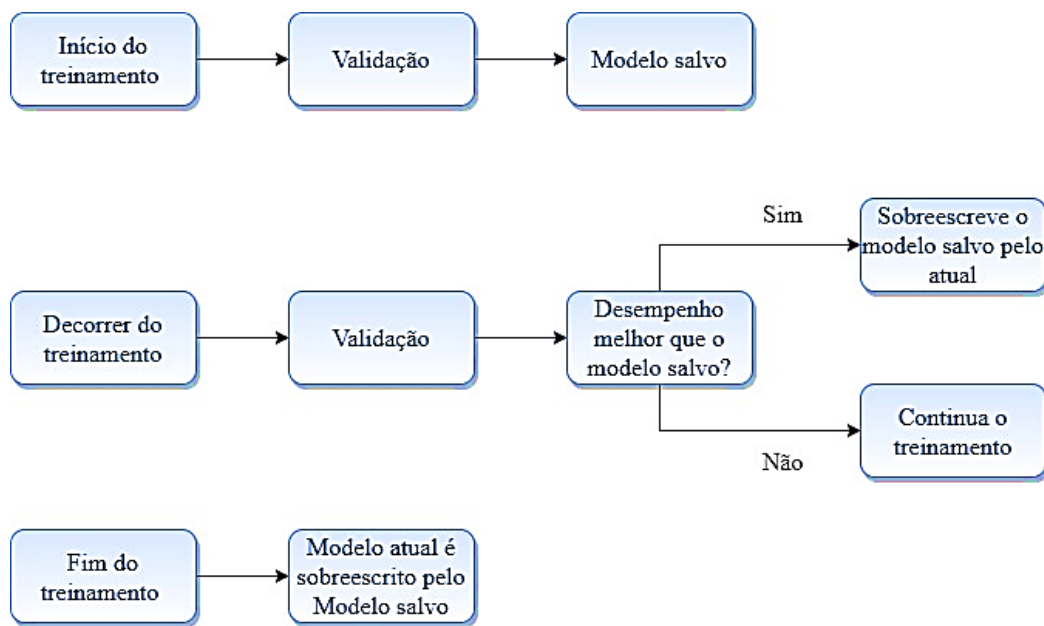


Figura 28. Descrição da Função de auxílio ao treinamento proposta

4.5 Métricas de desempenho

Para a escolha da melhor arquitetura na validação e nos testes foi feito o uso de métricas baseadas na predição das arquiteturas. Essas métricas são apresentadas como valores médios dos conjuntos de teste e de validação, como na Tabela 1, considerando que saídas com valores positivos dizem respeito à lesão, enquanto que as negativas dizem respeito ao *background*. Pela análise das métricas utilizadas na literatura, a maioria delas utiliza algo proveniente das subdivisões dos pixels da imagem predita em termos de TP (verdadeiro positivo), TN (verdadeiro negativo), FP (falso positivo) e FN (falso negativo), como ilustrado no diagrama de Veen da Figura 29. FN pode ser visto como os pixels que foram erroneamente classificados como *background*, TP pode ser visto como os pixels que foram classificados como lesão e que realmente são lesão, FP como os pixels que foram classificados como lesão mas que na

realidade são *background* e TN como os pixels que foram classificados como *background* e que de fato são *background*. As seguintes métricas foram utilizadas no presente trabalho: Acurácia, Acurácia Global, Interseção Sobre União (IOU), IOU Ponderada, Coeficiente Dice de Similaridade e Taxa BF (*Boundary F1*).

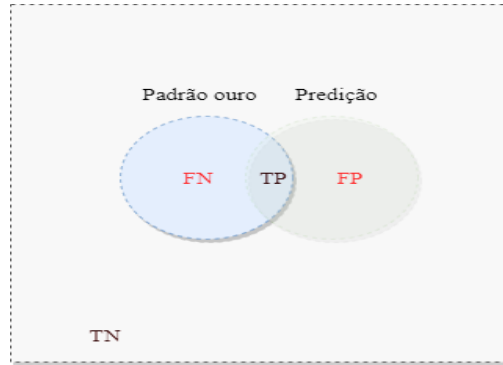


Figura 29: Diagrama de Venn ilustrando os conceitos de *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) e *False Negative* (FN).

A Acurácia média (*Mean Accuracy*), ou só acurácia indica a proporção de pixels que foram corretamente classificados em cada classe, e é dada por:

$$Acurácia = \frac{\left(\frac{TP}{TP + FN}\right) + \left(\frac{TN}{TN + FP}\right)}{2} \quad (22)$$

A Acurácia Global indica a proporção de pixels classificados corretamente, independentemente da classe, para o total de pixels, e é dada por:

$$Acurácia\ Global = \frac{TP + TN}{TP + TN + FP + FN} \quad (23)$$

A IOU média, ou Coeficiente Jaccard de similaridade, penaliza a classificação incorreta dos pixels ou como lesão (FP) ou como *background* (FN), e é dada por:

$$Lesão = \frac{TP}{TP + FN + FP} \quad (24)$$

$$Background = \frac{TN}{TN + FN + FP} \quad (25)$$

$$IOU = \frac{Lesão + Background}{2} \quad (26)$$

A IOU Ponderada normalmente é utilizada quando existe uma desproporcionalidade entre os tamanhos das classes na imagem, o que é exatamente o caso em questão,

principalmente no caso em que o *Dataset B* foi adaptado para a imagem inteira com 160x160. Essa métrica leva em consideração a proporção (número de pixels) que a classe tem na imagem, reduzindo, desta forma, a penalização alta nas classificações incorretas em classes de menor proporção e é dado por:

$$Peso_{Lesão} = (n^o \text{ de pixels pertencentes a lesão}) / (n^o \text{ total de pixels}) \quad (27)$$

$$Peso_{Bg} = (n^o \text{ de pixels pertencentes ao background}) / (n^o \text{ total de pixels}) \quad (28)$$

$$IOU \text{ Ponderada} = (Peso_{Lesão} \times Lesão) + (Peso_{Bg} \times Background) \quad (29)$$

O Coeficiente Dice de Similaridade mede a proporção de pixels corretamente classificados como lesão, penalizando a classificação incorreta (FP ou FN), e é dado por:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (30)$$

A Taxa *Boundary F1(BF)*, baseada na métrica *F1-measure*, indicará o quão bem as bordas de cada classe predita alinham-se com as bordas respectivas do seu padrão ouro. Seja B_{gt}^c a imagem contendo as bordas da lesão, gerada a partir do padrão ouro, S_{gt}^c , e B_{ps}^c a imagem contendo as bordas da lesão, gerada a partir da segmentação predita, S_{ps}^c . No qual z é o suporte a inversões, ou seja, $z=1$ se z é verdadeiro e $z=0$ se é falso. Onde θ é a distância de erro tolerada (normalmente empregado o valor 0,75%), as equações de *precision* e *recall* são dadas, respectivamente, por:

$$precision^c = \frac{1}{|B_{ps}|} \sum_{z \in B_{ps}^c} [d(z, B_{gt}^c < \theta)] \quad (31)$$

$$recall^c = \frac{1}{|B_{gt}|} \sum_{z \in B_{gt}^c} [d(z, B_{ps}^c < \theta)] \quad (32)$$

Onde $d()$ representa a distância Euclidiana. Logo, a taxa BF para a classe c é dada por:

$$Taxa \text{ BF} = \frac{2(precision^c \times recall^c)}{precision^c + recall^c} \quad (33)$$

5 RESULTADOS E DISCUSSÕES

Esta seção apresenta o desempenho das quatro arquiteturas propostas. Primeiramente são mostrados o resultado de validação e testes das arquiteturas com os bancos de dados. Depois os resultados são discutidos e contrastados.

5.1 Resultados de validação do *Dataset A*

A Figura 30 mostra os gráficos de convergência de rede, usando o conjunto de dados A, com o algoritmo de otimização ADAM. Nos eixos x e y estão as iterações e os valores de perda de *minibatch*, respectivamente. Todas as simulações, tanto do banco de dados A quanto de B, foram executadas com os mesmos parâmetros de treinamento, utilizando o método de otimização ADAM, taxa de aprendizado de 0,001, *minibatch* de 8 e regularização L2 de 10^{-3} . As Figuras 30 até 33 mostram os treinamentos das arquiteturas propostas, em azul está disposta a Acurácia global no conjunto de testes, já os dados em preto representam a acurácia global no conjunto de validação. Nas Figuras 34 e 35 são dispostos os resultados das quatro arquiteturas no mesmo gráfico, possibilitando contrastar a evolução das arquiteturas com o passar das épocas de treinamento.

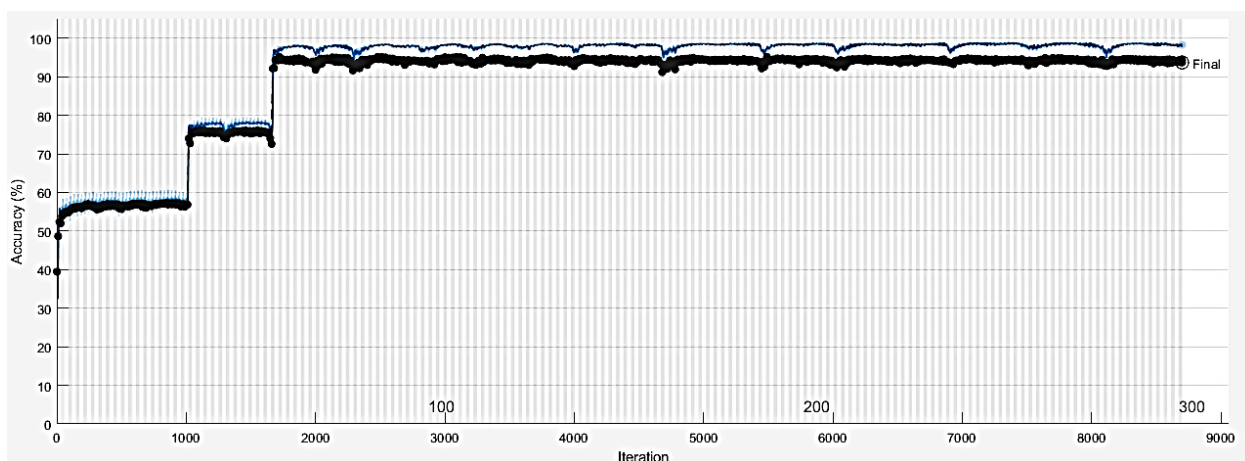


Figura 30: Gráfico de convergência da rede CNN1 na validação, usando o conjunto de dados BD-A

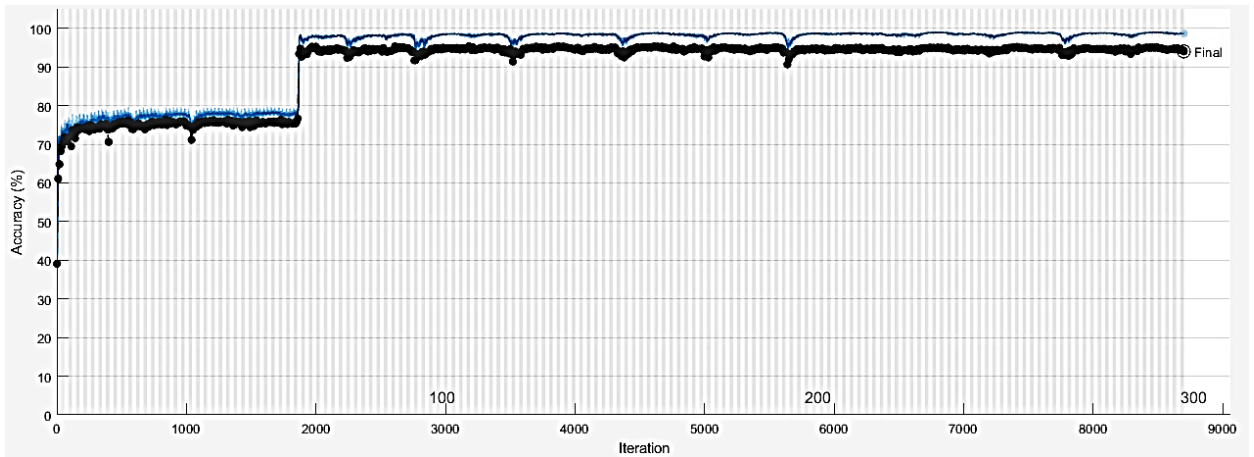


Figura 31: Gráfico de convergência da rede CNN2 na validação, usando o conjunto de dados BD-A

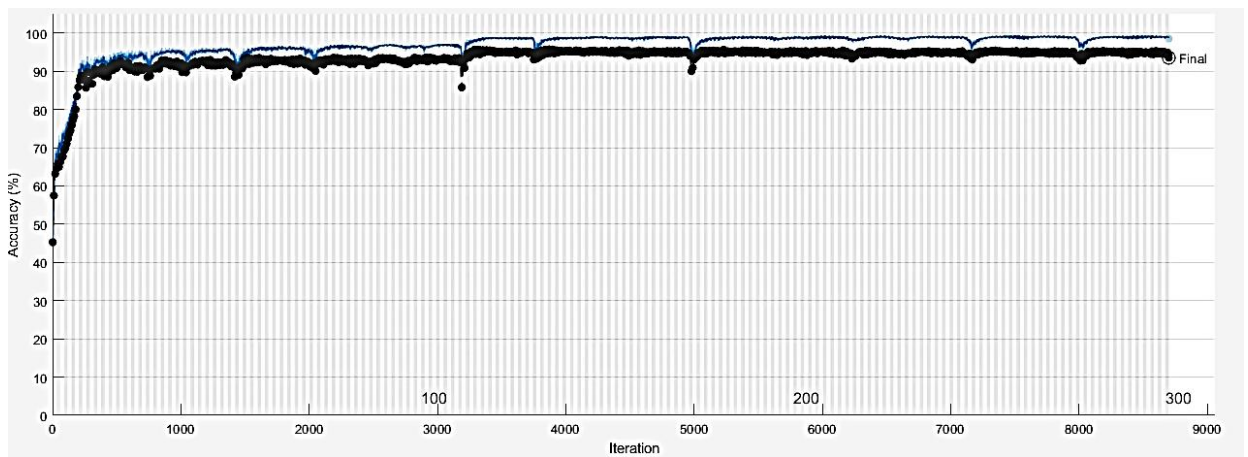


Figura 32: Gráfico de convergência da rede CNN3 na validação, usando o conjunto de dados BD-A

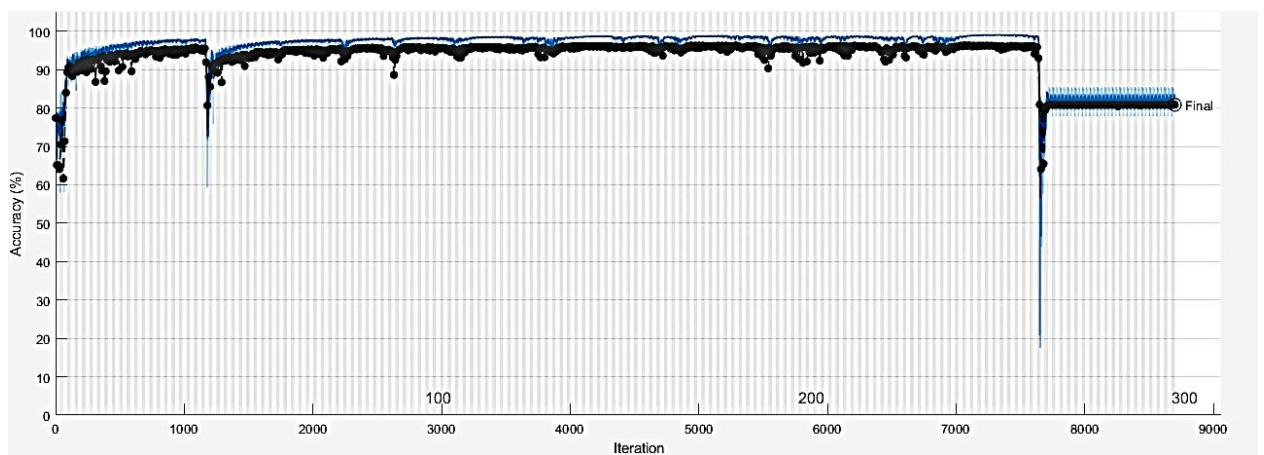


Figura 33: Gráfico de convergência da rede CNN4 na validação, usando o conjunto de dados BD-A

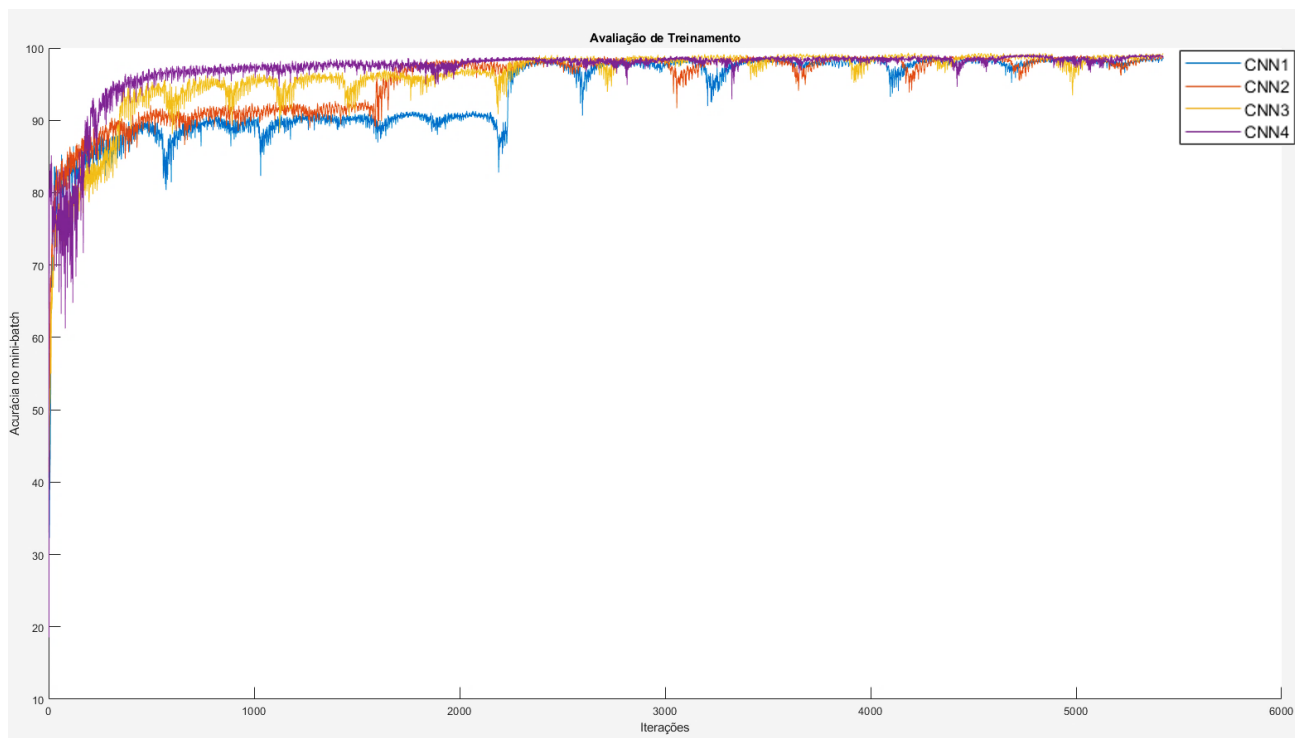


Figura 34: Acurácia de treinamento das 4 arquiteturas com BD-A

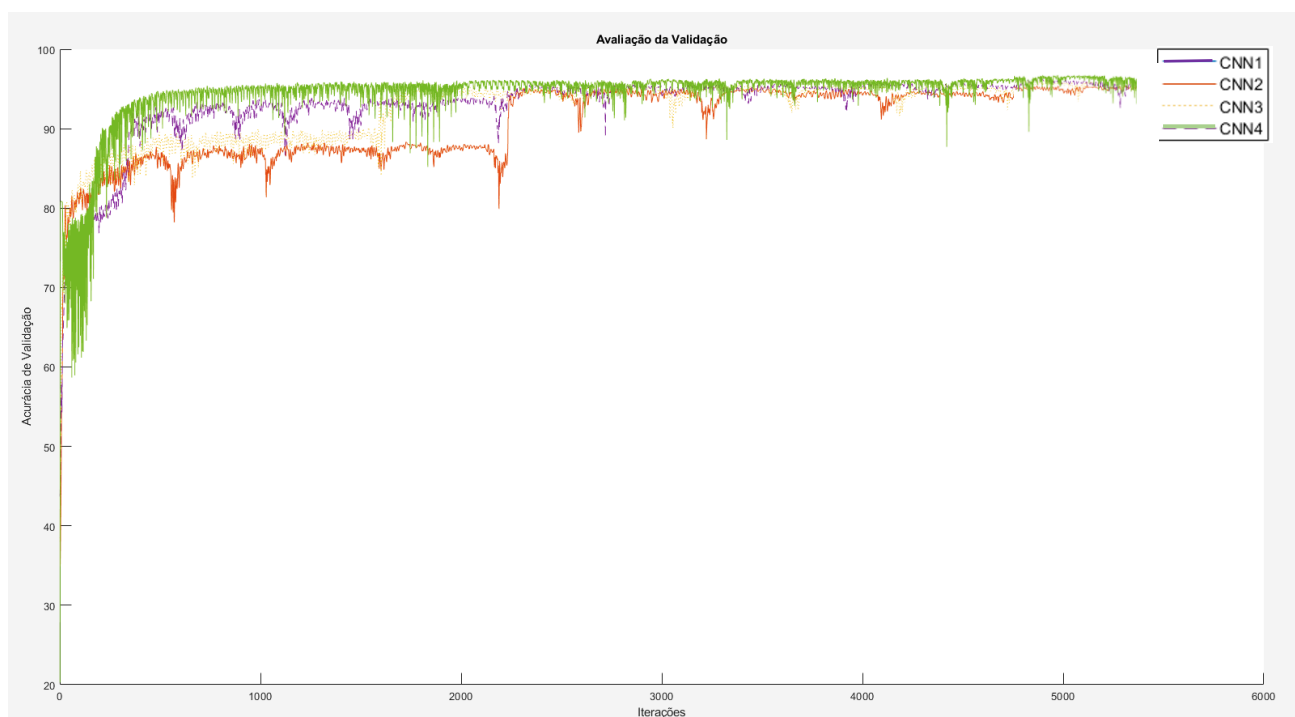


Figura 35: Acurácia de validação das 4 arquiteturas com BD-A

A partir das figuras 34 e 35 é possível verificar que o critério de paradas apenas pelo número de épocas não é o mais eficaz, vide a Tabela 1. Além disso, para fazer afirmações sobre

qual rede apresenta melhor desempenho é preciso treiná-las e avaliá-las no ponto de máximo desempenho, e não apenas um dos desempenhos possíveis.

Tabela 1: Desempenho na validação sem parada antecipada com BD-A. Observe que os valores apontam um melhor desempenho de CNN2.

Arquitetura	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice	Tempo de treino (min)
CNN 1	0,9369	0,9350	0,8310	0,8879	0,5876	0,8426	35
CNN2	0,9400	0,9426	0,8391	0,8931	0,6200	0,8495	47
CNN3	0,9351	0,9401	0,8288	0,8855	0,5889	0,8403	61
CNN4	0,9083	0,8000	0,8042	0,8534	0,5335	0,8403	99

Dessa forma foi implementada uma função de auxílio ao treinamento que força a rede parar na região de melhor desempenho, de acordo com a métrica de acurácia global. A partir da função percebeu-se que o melhor desempenho na arquitetura 1 foi de 94,23%, obtido na iteração 2480. A segunda arquitetura obteve o seu melhor desempenho de 94,81% na iteração 3370. A terceira arquitetura, 95,24% na iteração 3870. Por fim a quarta arquitetura obteve 96,00% na iteração 3140. A Tabela 2 mostra as métricas de desempenho das arquiteturas, com parada antecipada, mostrando a CNN4 com melhores resultados em 5 das 6 métricas utilizadas. Nessa segunda tabela não há o tempo de treinamento, pois levou o mesmo tempo da Tabela 1, porém os pesos no ponto de melhor desempenho foram congelados através da função de auxílio ao treinamento, o mesmo vale para as Tabelas 4, 6 e 8.

É importante ressaltar que uma vez treinada essa rede consegue fazer previsões de novas imagens de forma rápida e precisa, isso pode ser comprovado pelo fato de que em todos os casos dos bancos de dados de A e B o conjunto de validação inteiro, como no caso de A com 78 imagens e no caso de B com 33 imagens, foi predito em menos de um segundo, conforme contabilizado através do MATLAB.

Tabela 2: Desempenho das métricas com as arquiteturas testadas na etapa de validação com BD-A, utilizando a função de auxílio ao treinamento. Observe que agora a arquitetura 4 é a que tem melhores desempenhos.

Arquitetura	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice
CNN 1	0,9423	0,9455	0,8445	0,8969	0,5539	0,8535
CNN2	0,9481	0,9421	0,8559	0,9058	0,6645	0,8660
CNN3	0,9524	0,9475	0,8665	0,9130	0,6491	0,8771
CNN4	0,9600	0,9289	0,8799	0,9243	0,6990	0,8905

Conforme foi visto, os valores obtidos nas métricas das arquiteturas foram próximos, levando a dúvida se há significância estatística ou se os melhores resultados podem ser atribuídos ao acaso. O teste de significância estatística, realizado mais a diante, ajuda a demonstrar se há relevância entre os valores.

5.2 Resultados de validação do *Dataset B cropped*

As Figuras 36, 37, 38, 39, 40 e 41 mostram os gráficos de convergência de rede, para as quatro arquiteturas propostas, sendo que nas figuras 40 e 41 os dados de todas as arquiteturas estão dispostos no mesmo gráfico. Sendo esta primeira para imagens *cropped* 160x160.

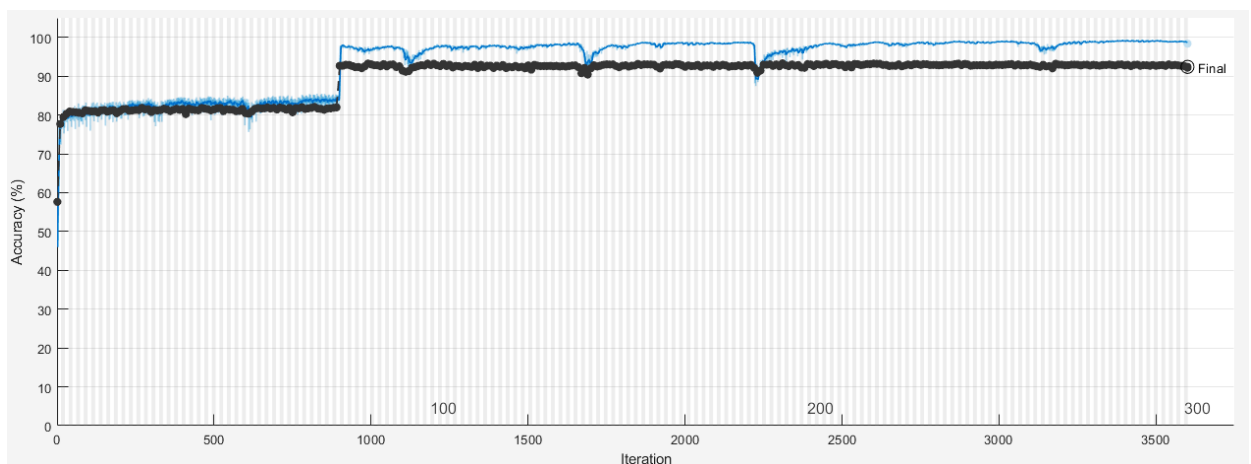


Figura 36: Gráfico de convergência da rede CNN1 na validação, usando o conjunto de dados BD-B *cropped*

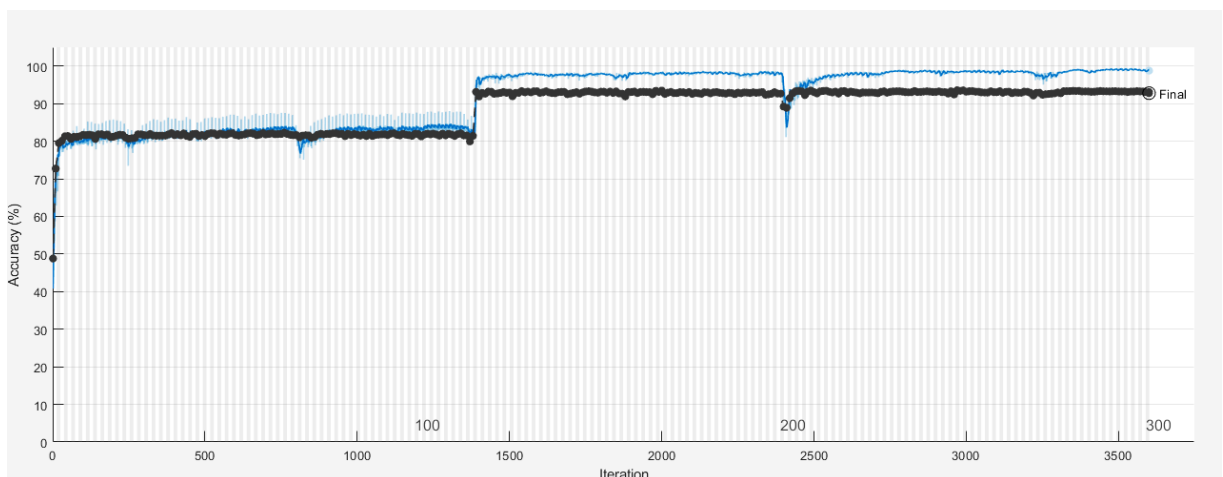


Figura 37: Gráfico de convergência da rede CNN2 na validação, usando o conjunto de dados BD-B *cropped*

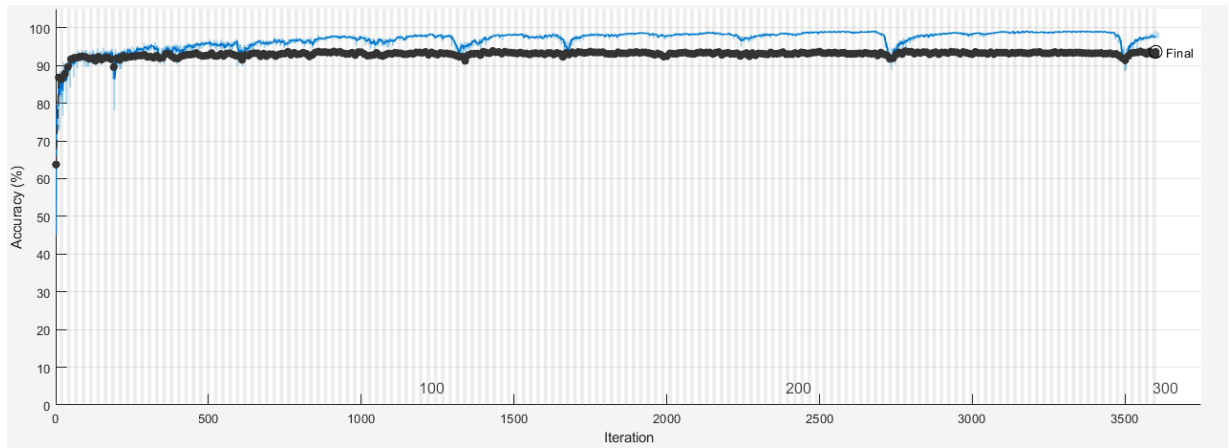


Figura 38: Gráfico de convergência da rede CNN3 na validação, usando o conjunto de dados BD-B *cropped*

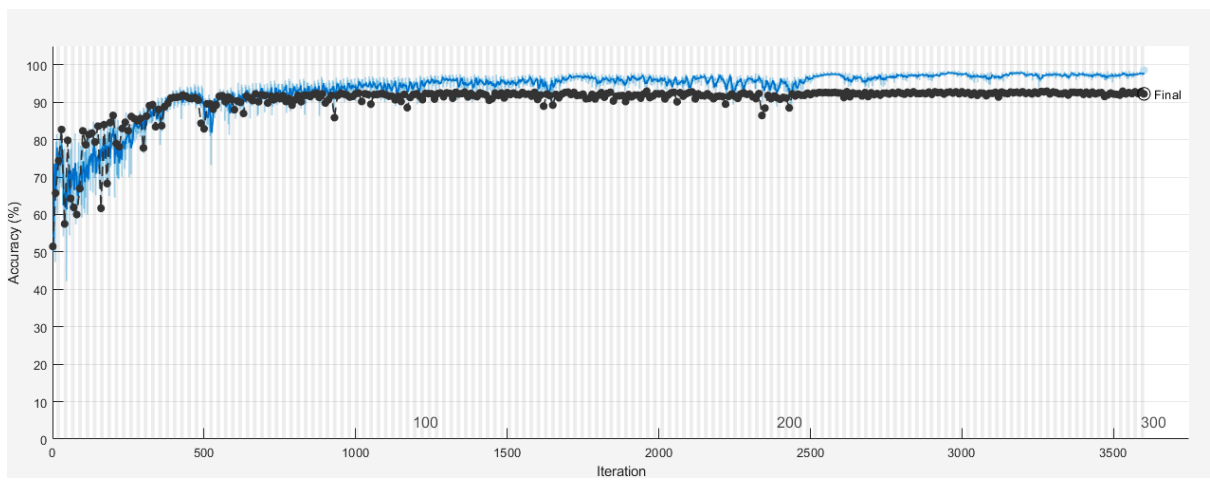


Figura 39: Gráfico de convergência da rede CNN4 na validação, usando o conjunto de dados BD-B *cropped*

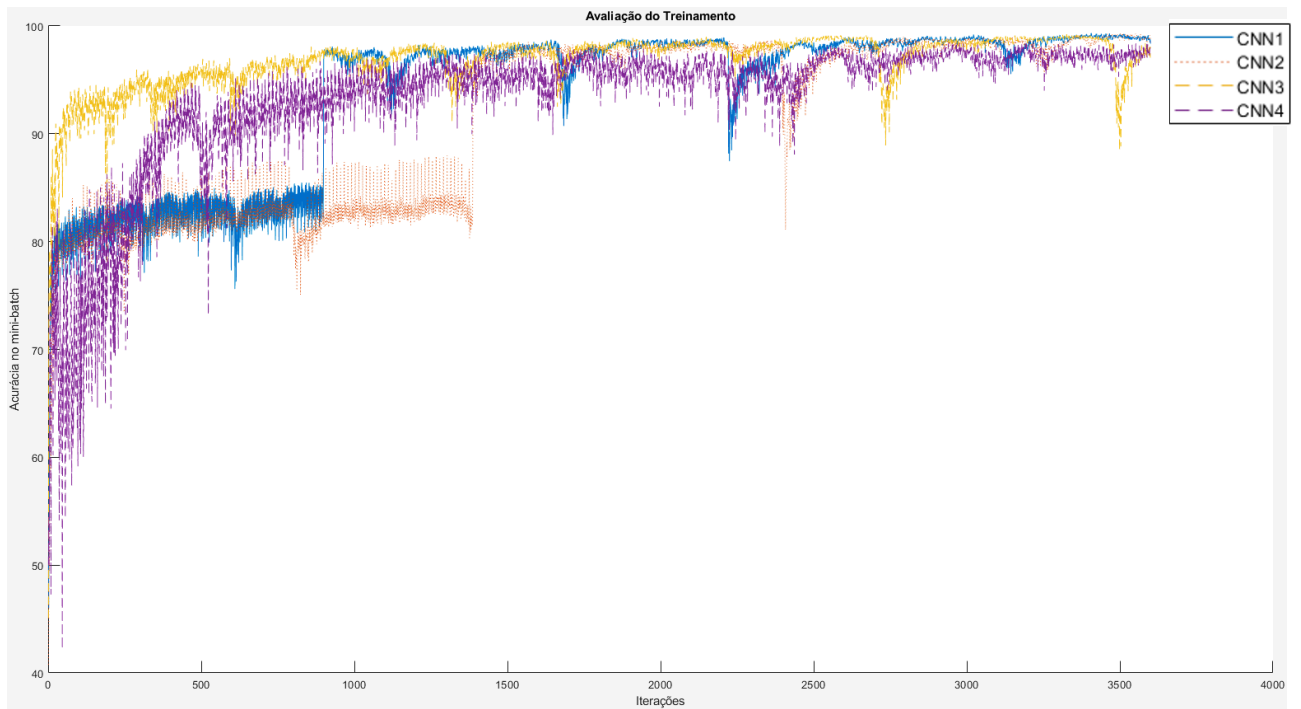


Figura 40: Acurácia de treinamento das 4 arquiteturas na validação com BD-B *cropped*

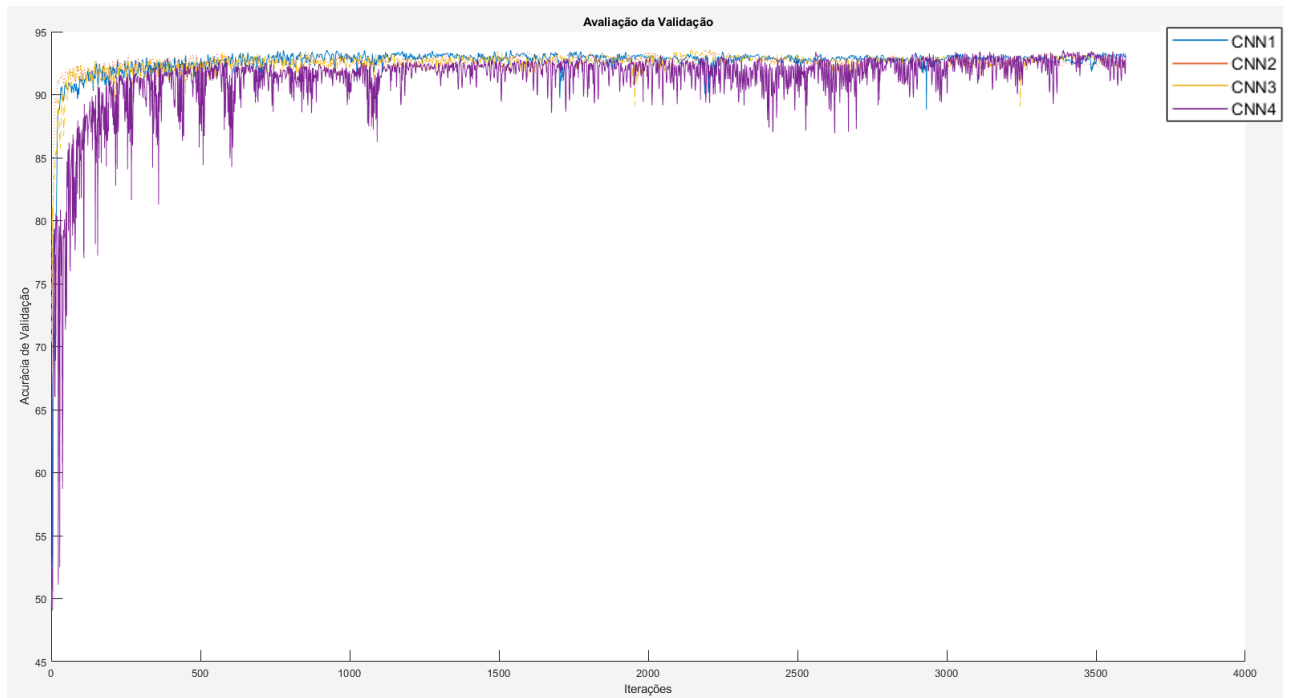


Figura 41: Acurácia de validação das 4 arquiteturas na validação com BD-B *cropped*

Com a implementação da função que força a rede parar na região de melhor desempenho, percebeu-se que o melhor desempenho na arquitetura 1 foi de 93,28%, obtido na iteração 2480. A segunda arquitetura obteve o seu melhor desempenho de 93,17% na iteração 3370. A terceira

arquitetura, 93,10% na iteração 3870. Por fim a quarta arquitetura obteve 92,96% na iteração 7070. Os resultados sem e com o uso dessa função de parada antecipada são mostrados abaixo, nas Tabelas 3 e 4.

Tabela 3: Desempenho das métricas com as arquiteturas testadas na validação com BD-B, sem a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN3.

Arquitetura	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice	Tempo de treino (min)
CNN 1	0,9219	0,9249	0,8586	0,8587	0,5431	0,9245	14
CNN2	0,9277	0,9288	0,8652	0,8653	0,5469	0,9304	17
CNN3	0,9309	0,9342	0,8695	0,8296	0,5565	0,9355	22
CNN4	0,9221	0,9234	0,8555	0,8555	0,5412	0,9255	55

A Tabela 4 mostra as métricas de desempenho das arquiteturas, mostrando a CNN1 com melhores resultados em das 5 métricas utilizadas.

Tabela 4: Desempenho na validação com BD-B, com a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN1.

Arquitetura	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice
CNN 1	0,9328	0,9337	0,8741	0,8741	0,5502	0,9350
CNN2	0,9317	0,9325	0,8721	0,8721	0,5319	0,9328
CNN3	0,9310	0,9317	0,8710	0,8710	0,5520	0,9326
CNN4	0,9296	0,9299	0,8684	0,8685	0,5339	0,9270

Como os resultados de desempenho foram muito próximos seria novamente necessário o uso de um teste de significância para determinar se há superioridade de alguma das arquiteturas.

5.3 Resultados de validação do *Dataset B* sem *cropt 160x160 pixels*

Afim de verificar se o *cropping* é realmente essencial para uma boa segmentação da lesão foram testadas as imagens do banco de dados B sem este procedimento. Os gráficos das figuras 39 e 40 mostram a forma como o treinamento evoluiu com o decorrer das épocas. Analisando estes gráficos é possível perceber que a parada pelo número de épocas ficou mais instável do que quando foi realizado o *cropping*, como na CNN1 e CNN 3.

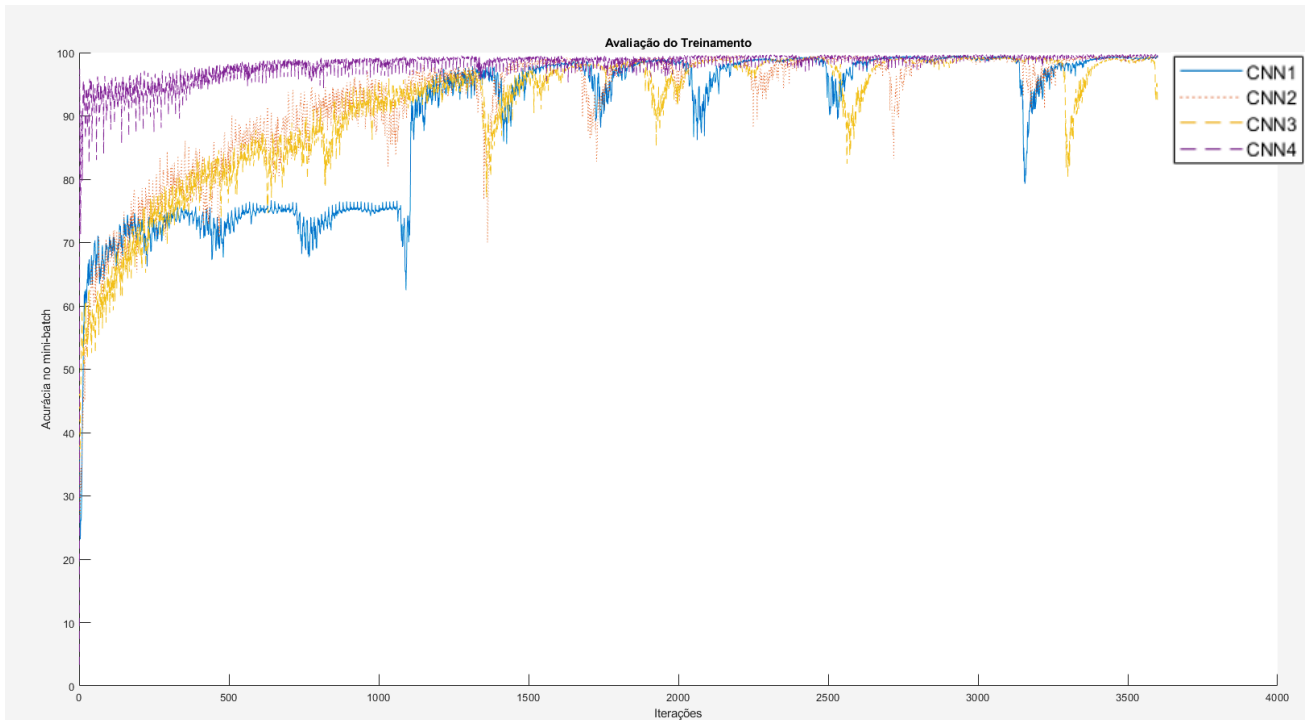


Figura 42: Acurácia de treinamento das 4 arquiteturas em BD-B sem *cropping*

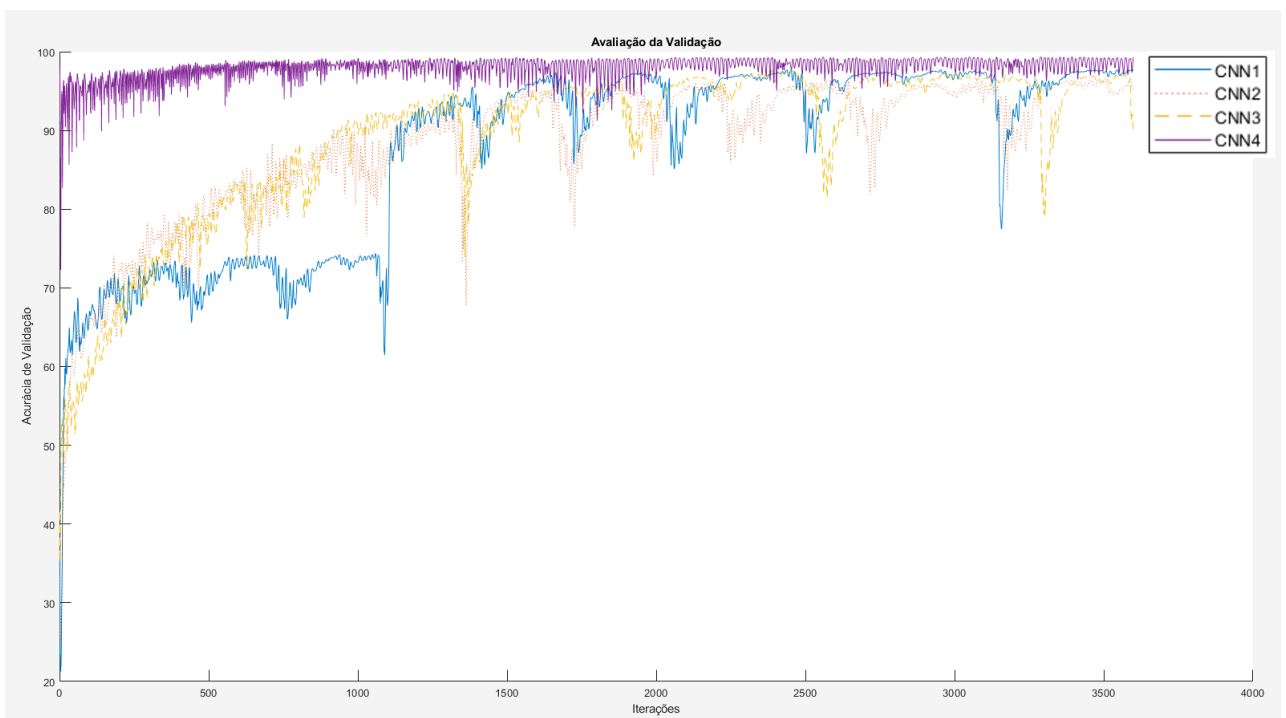


Figura 43: Acurácia de validação das 4 arquiteturas em BD-B sem *cropping*

As Tabelas 5 e 6 mostram o desempenho dessas redes de acordo com as métricas adotadas.

Tabela 5: Desempenho na validação com BD-B sem *cropping*, sem a função de auxílio.

Arquitetura	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice	Tempo de treino (min)
CNN 1	0,9845	0,9409	0,7560	0,9754	0,6433	0,5647	15
CNN2	0,9855	0,9412	0,7649	0,9767	0,6589	0,6023	18
CNN3	0,9028	0,9312	0,5307	0,8867	0,3901	0,2660	24
CNN4	0,9927	0,8991	0,8375	0,9866	0,8039	0,6176	56

Tabela 6: Desempenho na validação com BD-B sem *cropping*, com a função de auxílio.

Arquitetura	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice
CNN 1	0,9827	0,9412	0,7416	0,9731	0,6319	0,5968
CNN2	0,9781	0,9470	0,7125	0,9675	0,5273	0,5203
CNN3	0,9817	0,9478	0,7363	0,9719	0,5714	0,5640
CNN4	0,9928	0,8649	0,8281	0,9863	0,7778	0,5236

Com ou sem a função de treinamento criada, a arquitetura CNN4 foi superior. Porém, com o uso dessa função CNN4 foi superior em 3 e não 4 das métricas apresentadas. Isso pode ser explicado pelo fato de a função de treinamento levar em conta apenas a acurácia global, e não o conjunto de todas métricas. Neste caso poderia ser feito um estudo mais detalhado para avaliar quais métricas seriam mais importantes para o critério de parada.

5.4 Resultados de validação do *Dataset B* sem *cropping* 320x320 pixels

Afim de verificar o comportamento da rede com uma imagem de tamanho maior é possível verificar a real diferença entre o *cropping*. As figuras 39 e 40 mostram a forma como o treinamento evoluiu com o decorrer das épocas.

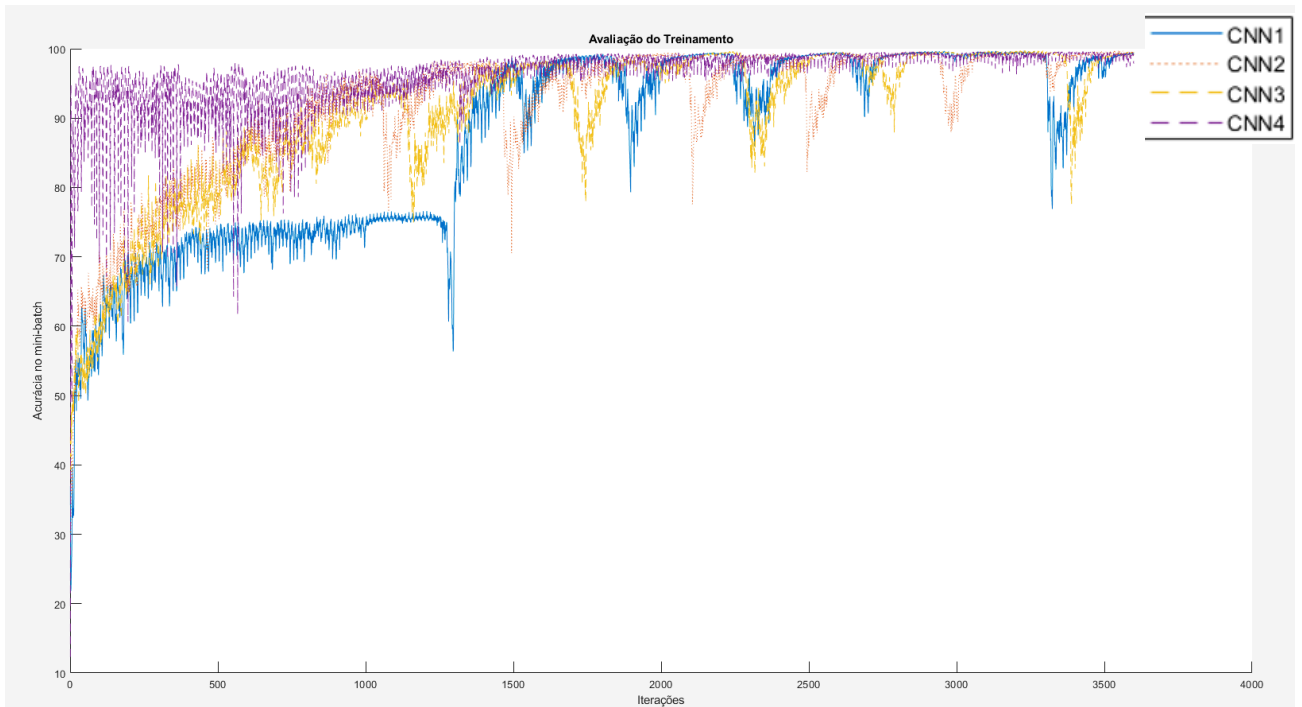


Figura 44: Acurácia de treinamento das 4 arquiteturas com BD-B sem *cropping*

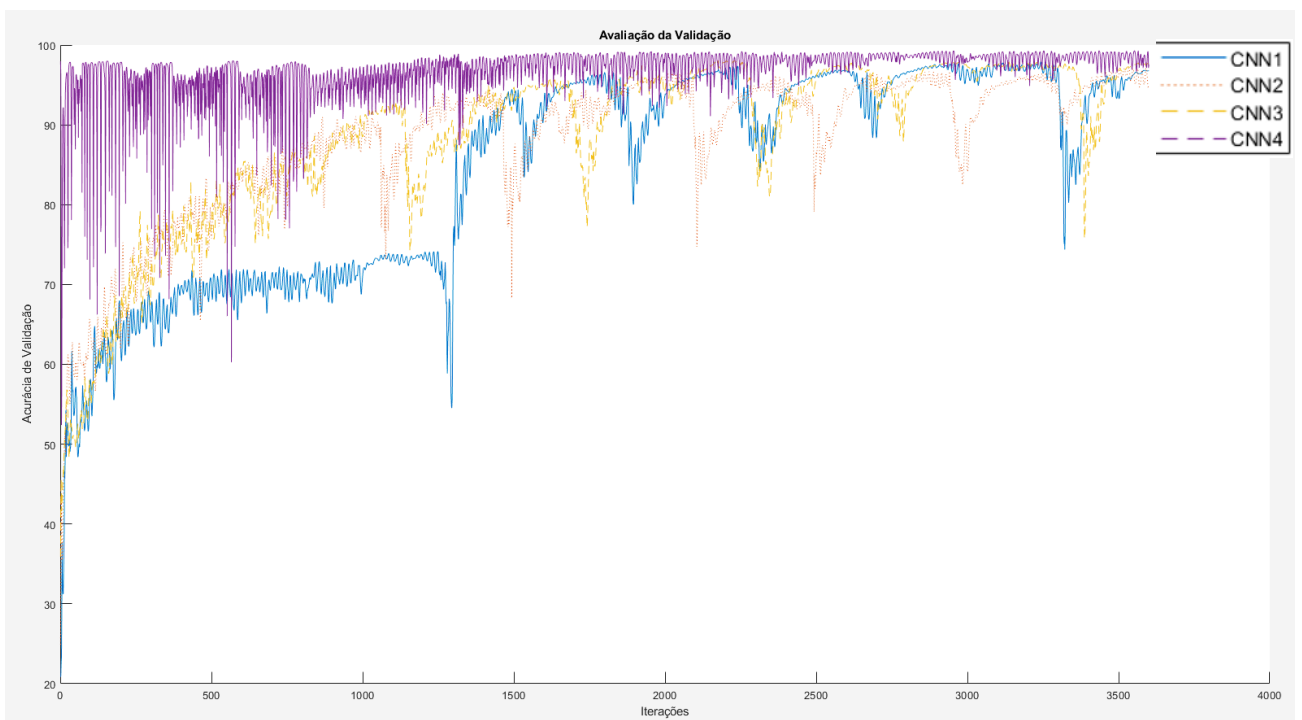


Figura 45: Acurácia de validação das 4 arquiteturas com BD-B sem *cropping*

As Tabelas 7 e 8 mostram os resultados do treinamento com e sem a função de auxílio ao treinamento criada.

Tabela 7: Desempenho na validação com BD-B sem *cropping*, sem a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN3.

	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice	Tempo de treino (min)
CNN 1	0,9814	0,9425	0,7330	0,9715	0,6303	0,5987	39
CNN2	0,9733	0,9447	0,6858	0,9617	0,6059	0,5336	46
CNN3	0,9820	0,9404	0,7368	0,9722	0,6424	0,6098	54
CNN4	0,9756	0,9383	0,6958	0,9643	0,5019	0,4875	157

Tabela 8: Desempenho na validação com BD-B sem *cropping*, com a função de auxílio. Observe que os valores apontam um melhor desempenho de CNN4.

	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice
CNN 1	0,9857	0,9099	0,7574	0,9766	0,6884	0,6281
CNN2	0,9802	0,9331	0,7227	0,9698	0,6125	0,5656
CNN3	0,9833	0,9220	0,7414	0,9737	0,6799	0,2986
CNN4	0,9911	0,8793	0,8073	0,9838	0,7048	0,5819

5.5 Teste de significância estatística

Os testes foram realizados com os subconjuntos de validação, para a escolha da melhor arquitetura e identificação de superioridade de um determinado subconjunto (como por exemplo para provar que o subconjunto 3 era preferível). Para realizar os testes de igualdade de variâncias e os testes de médias, precisamos que as populações testadas sejam independentes. Porém, no caso deste trabalho essas populações não são independentes. Neste caso, utilizamos o teste t pareado. A hipótese nula testada que representa uma incapacidade de provar uma significância estatística também é testada.

Para o conjunto de dados A. Utilizando-se o teste *t-student* para métrica de acurácia global, com graus de liberdade de 154, para 78 amostras e com 95% de certeza, a arquitetura CNN4 foi superior a todas as outras, na etapa de validação. A coluna “Estratégias” representa as hipóteses testadas, se uma arquitetura é superior a outra. Os resultados do valor do teste para as arquiteturas são mostrados na Tabela 9. O valor crítico é 1,64, quando o resultado fica acima deste valor, a superioridade de desempenho da arquitetura é estatisticamente significativa.

Tabela 9. Resultado do teste *t-student* entre as demais arquiteturas e CNN4 para o *dataset A*

Estratégias	Média	Variância	Média CNN4	Variância CNN4	<i>t-student</i>
CNN1 x CNN4	0,9423	0,000878	0,9600	0,000576	5,23. $p < 0,05$
CNN2 x CNN4	0,9481	0,001107	0,9600	0,000576	3,95. $p < 0,05$
CNN3 x CNN4	0,9524	0,000754	0,9600	0,000576	2,82. $p < 0,05$

Para o conjunto de dados BD-B sem *cropping* (imagens de 160x160 pixels). Utilizando-se o teste *t-student* com grau de liberdade de 62, para 32 amostras, e com 99% de certeza a arquitetura CNN4 foi superior a todas as outras, na etapa de validação, os resultados do valor do teste para as arquiteturas é mostrado na Tabela 10. O valor crítico é 2,39 quando o resultado fica acima deste valor, a superioridade de desempenho da arquitetura é estatisticamente significativa.

Tabela 10: Resultado do teste *t-student* entre as demais arquiteturas e CNN4 para o *dataset B*

Estratégias	Média	Variância	Média CNN4	Variância CNN4	<i>t-student</i>
CNN1 x CNN4	0,9827	0,000468	0,992831	0,0000027	3,3. $p < 0,01$
CNN2 x CNN4	0,9781	0,000271	0,992831	0,0000027	4,96. $p < 0,01$
CNN3 x CNN4	0,9817	0,000275	0,992831	0,0000027	5,58. $p < 0,01$

Analisando-se os conjuntos de dados BD-B 320x320 pixels e 160x160 pixels verificou-se que não há significância estatística. Considerando-se um grau de liberdade de 62, para 32 amostras, e com 95% de certeza. O valor de *t* obtido é de 0,75. O valor crítico é 1,67 quando o resultado fica acima deste valor, a superioridade de desempenho da arquitetura é estatisticamente significativa. Então o uso de imagens maiores não acarretou em melhores métricas de desempenho, e sim em mais custo computacional (como tempo de treinamento 3 vezes maior).

5.6 Resultados de teste do *Dataset A* com validação cruzada

A Tabela 11 mostra o desempenho de cada uma das 5 pastas do *Dataset A*, bem como o valor médio do desempenho obtido. Os valores obtidos já são após o uso da função de auxílio ao treinamento. Esta etapa foi realizada apenas com a arquitetura CNN4, já que a mesma foi a que alcançou melhores resultados na etapa de validação. Os dados utilizados na etapa de

validação foram desconsiderados na etapa de testes, para que a rede não se tornasse tendenciosa a apresentar resultados superiores aos verdadeiros.

Tabela 11: Resultado da validação cruzada com o *Dataset A*

Pasta	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice	Tempo de treino (min)
1	0,9720	0,9463	0,9110	0,9462	0,7635	0,9236	235
2	0,9722	0,9498	0,9129	0,9467	0,7708	0,9253	263
3	0,9697	0,9411	0,9054	0,9417	0,7388	0,9171	242
4	0,9727	0,9532	0,9161	0,9476	0,7602	0,9253	243
5	0,9725	0,9528	0,9156	0,9473	0,7752	0,9275	244
Média	0,9718	0,9486	0,9122	0,9459	0,7617	0,9238	

5.7 Resultados de teste do *Dataset B* com validação cruzada

A etapa de validação permitiu algumas conclusões sobre os testes realizados com o conjunto de dados B. Uma delas foi o fato de que, para a tarefa de segmentação o procedimento de *cropt* não é necessário. Avaliando as métricas, o uso desse corte na região do tumor acabou dificultando o processamento da rede, levando a resultados levemente inferiores, como por exemplo o decréscimo na acurácia global de CNN4 de 99% para 93%. Outra questão a ser levada em conta é que o aumento espacial de 160x160 pixels para 320x320 pixels não é justificado analisando-se as métricas de desempenho, tendo em vista um aumento substancial no tempo de processamento sem o devido acréscimo às métricas de desempenho, conforme pode ser analisado comparando as Tabelas 5 e 7. Com base nisso, o único dos subconjuntos de dados utilizado na etapa de testes foi o *Dataset B* 160x160 pixels sem *cropping*.

A Tabela 12 mostra o desempenho de cada uma das 5 pastas do *Dataset B* 160x160 pixels sem *cropping*, bem como o valor médio do desempenho obtido. Os valores obtidos já são após o uso da função de auxílio ao treinamento. Esta etapa foi realizada apenas com a arquitetura CNN4, já que a mesma foi a que alcançou melhores resultados na etapa de validação, conforme a Tabela 5.

Tabela 12: Resultado da validação cruzada com o *Dataset B*

Pasta	Acurácia Global	Acurácia média	IoU médio	IoU Ponderado	BF Score médio	Dice	Tempo de treino (min)
1	0,9751	0,8325	0,7884	0,9532	0,6000	0,9751	75
2	0,9769	0,8079	0,7414	0,9586	0,6617	0,9769	73
3	0,9705	0,7832	0,7436	0,9446	0,6704	0,9705	82
4	0,9805	0,8357	0,7811	0,9640	0,5564	0,9805	70
5	0,9807	0,7818	0,7524	0,9632	0,6550	0,9807	78
Média	0,9767	0,8082	0,7614	0,9567	0,6287	0,9767	

5.8 Discussão dos resultados

Os dados apresentados buscam contrastar as quatro arquiteturas de redes convolucionais para identificar qual seria mais recomendada para a uma segmentação precisa de leões de mama em imagens de ultrassonografia. Durante o treinamento, os pesos da rede são ajustados para diminuir o valor da perda em cada *minibatch*, forçando a convergência do algoritmo para um ponto de mínimo. Como esse é um processo estocástico e os pesos são inicializados aleatoriamente, treinamentos sucessivos no mesmo conjunto de dados não resultam em pesos iguais na iteração final. Isso se deve ao fato de, como o processo de inicialização é aleatório, o estado inicial da rede pode ser muito variável, podendo está localizado próximo a um mínimo global, ou próximo a um dos muitos mínimos locais, com diferentes características cada um. Isso faz com que em alguns casos a rede pare em regiões com um desempenho bastante inferior ao seu potencial máximo, como o caso extremo da Figura 40, em que CNN3 finalizou seu treinamento em um valor de desempenho inferior aos demais. O uso da “função de auxílio ao treinamento”, definida na seção 4.4, que foi criada para identificar, dentro de um *minibatch*, a época com o melhor resultado, permitiu obter a configuração de pesos da CNN que resultou em um valor ótimo para a acurácia. O conceito de valor ótimo, nesse caso, diz respeito ao melhor valor obtido no treinamento.

A rede CNN4 apontou ser a melhor, apresentando melhores desempenhos com os bancos de dados A e B sem *cropping*. Esta rede é uma rede de estrutura acíclica denominada U-net. Em trabalhos anteriores já havia alcançado melhores resultados em tarefas de classificação com a base de dados B (YAP et al., 2018). Porém, como constatado, o seu treinamento é instável e um bom critério de parada é fundamental. A CNN1, que também já havia sido testada em outras modalidades de imagens médicas, chega a alcançar os melhores resultados quando é utilizado BD-B *cropped* (ROTH et al., 2015). Esta é uma rede sequencial, com 54 camadas, 16 camadas

a menos que a CNN4. Se estivéssemos preocupados com baixo custo computacional esta arquitetura poderia ser escolhida sem grandes perdas no desempenho.

Para a escolha da melhor rede, verificou-se qual rede apresentava melhores desempenhos em um maior número de métricas, tendo em vista que não é possível saber se uma métrica é superior a outra, pois envolvem abordagens diferentes, como o caso da acurácia global, que identifica o total de pixels que foram atribuídos a classe certa, ou o coeficiente Dice que é uma medida de sobreposição espacial. O dado apresentado na Tabela 4, em que com parada antecipada o valor do coeficiente Dice diminuiu, embora a acurácia global tenha aumentado, pode ser explicado pela variabilidade deste coeficiente de acordo com as imagens, conforme comentado por um outro trabalho (ZOU et al., 2004) que utilizou esse coeficiente como métrica de desempenho. Além disso, outros autores apontam que essas diferenças entre as métricas são devido ao uso de banco de dados pequenos, como é caso do estudo. No caso do BD-B, temos apenas 163 imagens (AHN et al., 2017; DHUNGEL; CARNEIRO; BRADLEY, 2015; YAP et al., 2018).

A realização do *cropping* se mostrou dispensável com relação as métricas, exceto com respeito ao coeficiente Dice, que ficou com níveis inferiores sem este procedimento. Há uma diminuição de qualidade da imagem quando é utilizada a imagem sem *cropping*. Essa diminuição levou a um desempenho inferior em algumas das métricas adotadas, como pode ser constatado comparando-se as Tabelas 4 e 8. Na literatura é conhecido o fato das imagens com mais qualidade apresentarem métricas piores do que as com qualidade inferior (YAP et al., 2018). Isso se deve ao fato de imagens com maior qualidade conterem informações indesejáveis, como ossos, músculo peitoral ou ar no pulmão.

A etapa de validação permitiu que somente uma parte das arquiteturas passasse para a fase subsequente. O banco de dados B com o *cropping* não foi levado a próxima etapa, devido a diminuição observada das métricas. O banco de dados B com imagem de 320x320 pixels também não alcançou a fase de teste, tendo em vista a pouca diferença entre as métricas, quando se compara os resultados com aqueles obtidos com imagens de 160x160 pixels, sem contar o alto custo computacional de se trabalhar com imagens maiores. O teste de significância estatística ajudou a eleger a CNN4 como melhor rede.

Os resultados na fase de testes com o uso da técnica de validação cruzada alcançaram valores de métricas próximos àqueles obtidos na etapa de validação. Supõe-se que o desempenho de CNN4 em uma imagem diferente, nunca vista pela rede, deve alcançar valores de métricas próximos a média das 5 pastas, obtidos na fase de testes.

Por fim, ressaltamos que a principal contribuição desse estudo foi avaliar se as arquiteturas de RNCs poderiam delinear contornos irregulares, com alto nível de ruídos e margens anguladas, como as encontradas nas imagens de ultrassonografia de lesões mamárias. Além das arquiteturas sequencias, deu-se preferência pela utilização de arquiteturas DAGs, tendo em vista um conhecimento prévio do desempenho da arquitetura DAG em aplicações de segmentação de imagens, sendo as mesmas capazes de preservar detalhes finos. No presente estudo, constatou-se que o desempenho das arquiteturas DAG, quando comparado ao desempenho de uma arquitetura em série, é superior, tanto qualitativamente quanto quantitativamente, quando se leva em conta a acurácia da segmentação. Porém, o resultado da Tabela 4 mostra que a arquitetura sequencial obtém uma boa performance em imagens de maior resolução. Em particular, o valor da acurácia obtida pela CNN4 superior a 99% leva a concluir que um método preciso de segmentação foi desenvolvido.

6 CONCLUSÕES

Neste trabalho explorou-se um método de segmentação automática de lesões em imagens de ultrassonografia digital utilizando-se redes convolutivas. Os bons resultados alcançados, mormente em imagens de Ultrassonografia com características que tornam difíceis o processamento e a segmentação, tais como baixo contraste, baixa resolução, baixa relação sinal ruído, entre outros, mostraram a relevância do trabalho. Os conjuntos de dados contrastados, *dataset A* e *dataset B*, com propriedades e características diferentes, revelaram uma arquitetura de rede convolucional que sobressaiu em relação às demais, a CNN4, uma rede com 70 camadas de estrutura de grafo acíclico direto. Os testes de significância estatística mostraram que a arquitetura CNN4 é superior às demais. Porém, analisando os resultados após a aplicação da “função de auxílio ao treinamento”, verificou-se que os valores das métricas para as quatro arquiteturas testadas foram muito próximos, mostrando que para um trabalho em que fosse necessário uma rede menor, a arquitetura CNN1 poderia ser escolhida sem grandes sacrifícios no desempenho.

A criação da “função de auxílio ao treinamento” foi uma estratégia que permitiu alcançar melhores resultados, levando a uma acurácia média (obtida através de validação cruzada) das segmentações realizadas nos conjuntos de teste A e B de mais de 96%. Verificou-se que, para a etapa de segmentação, nem o uso de *cropping* nem tão pouco o aumento da imagem para 320x320 pixels levou a um melhor resultado. De certa forma, o redimensionamento da imagem para o tamanho 160x160 pixels mostrou-se ideal para a obtenção de uma melhor segmentação. As redes com estruturas de grafos acíclicos não obtiveram resultados com diferenças estatisticamente significante em relação as redes sequenciais na tarefa segmentação.

As observações sobre a forma instável com que as redes param o treinamento em função do número de épocas pode ser estendida para outras aplicações, sendo necessário para algumas plataformas como o MATLAB® uma “função de auxílio ao treinamento”, que possibilita um congelamento da matriz de pesos na região de melhor desempenho. Uma continuação desse trabalho poderia aproveitar parte da melhor arquitetura (CNN4), ou então as imagens já segmentadas, para a tarefa de classificação das lesões em malignas ou benignas. Poderia também ser realizado uma junção de mais conjuntos de dados para ter uma representação mais global do problema, obtendo assim um método mais preciso.

Portanto, o método de segmentação automático de lesões de mama pode constituir um bom ponto de partida para pesquisadores que desejem criar um sistema computadorizado para segmentação automática de lesões mamárias em imagens de ultrassonografia.

REFERÊNCIAS

- AHN, C. K. et al. A novel deep learning-based approach to high accuracy breast density estimation in digital mammography. **Medical Imaging 2017: Computer-Aided Diagnosis**, v. 10134, n. Cc, p. 101342O, 2017.
- AKIN, O. et al. Advances in oncologic imaging: Update on 5 common cancers. **CA Cancer Journal for Clinicians**, v. 62, n. 6, p. 364–393, 2012.
- AWAD, M.; KHANNA, R. **Efficient learning machines: Theories, concepts, and applications for engineers and system designers**. New York: Apress Open, 2015.
- AZEVEDO, C. M. DE. Aplicação de operadores morfológicos na segmentação e determinação do contorno de tumores de mama em imagens por ultra-som Application of morphological. **Sociedade Brasileira de Engenharia Biomédica**, p. 91–101, 2003.
- BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg: Springer-Verlag, 2006.
- BYRA, M. et al. Combining Nakagami imaging and convolutional neural network for breast lesion classification. **IEEE International Ultrasonics Symposium, IUS**, p. 5–8, 2017.
- CAI, L.; WANG, Y. A phase-based active contour model for segmentation of breast ultrasound images. **Proceedings of the 2013 6th International Conference on Biomedical Engineering and Informatics, BMEI 2013**, n. 2012, p. 91–95, 2013.
- CHEN, H. et al. **Deep Contextual Networks for Neuronal Structure Segmentation**. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. **Anais...: AAAI'16.AAAI** Press, 2016Disponível em: <<http://dl.acm.org/citation.cfm?id=3015812.3015985>>
- DAOUD, M. I. et al. Accurate segmentation of breast tumors in ultrasound images using a custom-made active contour model and signal-to-noise ratio variations. **8th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2012r**, p. 137–141, 2012.
- DAOUD, M. I. et al. Accurate and fully automatic segmentation of breast ultrasound images by combining image boundary and region information. **Proceedings - International Symposium on Biomedical Imaging**, v. 2016- June, p. 718–721, 2016.
- DESANTIS, C. E. et al. Breast cancer statistics, 2017, racial disparity in mortality by state. **CA: A Cancer Journal for Clinicians**, v. 67, n. 6, p. 439–448, 2017.
- DHUNGEL, N.; CARNEIRO, G.; BRADLEY, A. P. Deep learning and structured prediction for the segmentation of mass in mammograms. **Lecture Notes in Computer Science**

(including subseries **Lecture Notes in Artificial Intelligence** and **Lecture Notes in Bioinformatics**), v. 9349, p. 605–612, 2015.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. [s.l.] O'Reilly Media, 2017a.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. [s.l.] O'Reilly Media, 2017b.

GÓMEZ, W. et al. Morphological operators on the segmentation of breast ultrasound images. **2009 Pan American Health Care Exchanges - PAHCE 2009**, p. 67–71, 2009.

GÓMEZ, W. et al. Segmentation of Breast Nodules on Ultrasonographic Images Based on Marked-Controlled Watershed Transform. **Computación y Sistemas**, v. 14, n. 2, p. 165–174, 2010.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [s.l.] The MIT Press, 2016.

HAGAN, M. T. et al. **Neural Network Design**. [s.l.] Martin Hagan, 2014.

HAN, S. et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. **Physics in Medicine and Biology**, v. 62, n. 19, p. 7714–7728, 2017.

HORSCH, K. et al. Automatic segmentation of breast lesions on ultrasound. **Medical Physics**, v. 28, n. 8, p. 1652–1659, 2001.

HUANG, Q. et al. Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis. **Information Sciences**, v. 314, p. 293–310, 2015.

INFANTOSI, A. F. C. et al. Breast ultrasound segmentation using morphologic operators and a Gaussian function constraint. **IFMBE Proceedings**, v. 20 IFMBE, n. January, p. 520–523, 2008.

JIAO, J.; WANG, Y. Automatic boundary detection in breast ultrasound images based on improved pulse coupled neural network and active contour model. **5th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2011**, n. 1, p. 1–4, 2011.

JIN, Q. et al. DUNet: A deformable network for retinal vessel segmentation. **Knowledge-Based Systems**, v. 178, n. 8, p. 149–162, 2019.

KER, J. et al. Deep Learning Applications in Medical Image Analysis. **IEEE Access**, v. 6, p. 9375–9379, 2017.

KINGMA, D.; BA, J. Adam: A Method for Stochastic Optimization. **International**

Conference on Learning Representations, 2014.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **ImageNet Classification with Deep Convolutional Neural Networks**. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. **Anais...**: NIPS'12.USA: Curran Associates Inc., 2012Disponível em: <<http://dl.acm.org/citation.cfm?id=2999134.2999257>>

LECUN, Y. et al. A B7CEDGF HIB7PRQTSUDGQICWVYX HIB edCdSISIXvg5r ` CdQTW XvefCdS. **proc. OF THE IEEE**, 1998.

LONG, J.; SHELHAMER, E.; DARRELL, T. FCN ppt. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, p. 3431–3440, 2015.

MENDES, J. P. C. et al. SEGMENTAÇÃO AUTOMÁTICA DE LESÕES DE MAMA EM IMAGENS DE ULTRASSOM. 2016.

MIYAGAWA, M. et al. Lumen Segmentation in Optical Coherence Tomography Images using Convolutional Neural Network. **Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS**, v. 2018-July, n. c, p. 600–603, 2018.

NOH, H.; HONG, S.; HAN, B. Learning Deconvolution Network for Semantic Segmentation. 17 maio 2015.

PANIGRAHI, L.; VERMA, K.; SINGH, B. K. Ultrasound image segmentation using a novel multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution. **Expert Systems with Applications**, v. 115, p. 486–498, 2019.

PATTERSON, J.; GIBSON, A. **Deep Learning A practitioner's approach**. [s.l: s.n.]. v. 14

QI, X. et al. Automated diagnosis of breast ultrasonography images using deep neural networks. **Medical Image Analysis**, v. 52, p. 185–198, 2019.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9351, p. 234–241, 2015.

ROTH, H. et al. Deep convolutional networks for pancreas segmentation in CT imaging. v. 9413, 2015.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. New York: Cambridge University Press, 2014. v. 9781107057

SHELHAMER, E.; LONG, J.; DARRELL, T. Fully Convolutional Networks for Semantic

Segmentation. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 39, n. 4, p. 640–651, 2017.

SMITH, R. A.; COKKINIDES, V.; EYRE, H. J. American Cancer Society Guidelines for the Early Detection of Cancer, 2006. **CA: A Cancer Journal for Clinicians**, v. 56, n. 1, p. 11–25, 2006.

SRIVASTAVA, N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, v. 15, p. 1929–1958, 2014.

SZEGEDY, C. et al. Going deeper with convolutions. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 1–9, 2015.

TRASK, A. W. **Grokking Deep Learning**. [s.l.] Manning, 2019.

XI, P.; SHU, C.; GOUBRAN, R. Abnormality Detection in Mammography using Deep Convolutional Neural Networks. **MeMeA 2018 - 2018 IEEE International Symposium on Medical Measurements and Applications, Proceedings**, p. 1–6, 2018.

XIAN, M.; CHENG, H. D.; ZHANG, Y. A fully automatic breast ultrasound image segmentation approach based on neutro-connectedness. **Proceedings - International Conference on Pattern Recognition**, p. 2495–2500, 2014.

XU, L.; ZHIMIN, H.; JIWU, Z. Automated segmentation of breast lesions in ultrasound images. **Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings**, v. 7 VOLS, p. 7433–7435, 2005.

YAP, M. H. et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 4, p. 1218–1226, 2018.

YE, J. et al. Breast Cancer Image Classification on WSI with Spatial Correlations. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**, v. 2019- May, p. 1219–1223, 2019.

ZOU, K. H. et al. Statistical validation of image segmentation quality based on a spatial overlap index. **Academic radiology**, v. 11, n. 2, p. 178–189, fev. 2004.