

JEFFERSON ALVES DE SOUSA

**Detecção de Degeneração Macular Relacionada  
à Idade e Edema Macular Diabético em  
Imagens de Tomografia de Coerência Óptica  
Utilizando Redes Neurais Convolucionais e  
Capsule Network**

São Luís

2021

JEFFERSON ALVES DE SOUSA

**Detecção de Degeneração Macular Relacionada à Idade e  
Edema Macular Diabético em Imagens de Tomografia de  
Coerência Óptica Utilizando Redes Neurais  
Convolucionais e Capsule Network**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da UFMA como requisito para a obtenção do grau de Doutor em Engenharia de Eletricidade.

Orientador: Prof. Dr. Anselmo Cardoso de Paiva  
Coorientador: Prof. Dr. Aristófanés Corrêa Silva

São Luís  
2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Sousa, Jefferson.

Detecção de Degeneração Macular Relacionada à Idade e Edema Macular Diabético em Imagens de Tomografia de Coerência Óptica Utilizando Redes Neurais Convolucionais e Capsule Network / Jefferson Sousa. - 2021.

82 p.

Coorientador(a): Aristófanés Silva.

Orientador(a): Anselmo Paiva.

Tese (Doutorado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, São Luís, 2021.

1. Computer-Aided Diagnosis. 2. Degeneração Macular. 3. Edema Macular Diabético. 4. Imagens médicas. 5. Redes Neurais Convolucionais. I. Paiva, Anselmo. II. Silva, Aristófanés. III. Título.

Tese de autoria de Jefferson Alves de Sousa, sob o título “**Detecção de Degeneração Macular Relacionada à Idade e Edema Macular Diabético em Imagens de Tomografia de Coerência Óptica Utilizando Redes Neurais Convolucionais e Capsule Network**”, apresentado ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Doutor em Engenharia de Eletricidade, aprovado em **20 de Agosto de 2021** pela comissão examinadora constituída pelos doutores:

**Prof. Dr. Anselmo Cardoso de Paiva**  
Orientador

**Prof. Dr. Aristófanês Corrêa Silva**  
Coorientador

**Prof. Dr. Geraldo Braz Júnior**  
Examinador Interno

**Prof. Dr. João Dallyson Sousa de Almeida**  
Examinador Interno

**Prof. Dr. Kelson Rômulo Teixeira Aires**  
Examinador Externo

**Prof. Dr. Rodrigo de Melo Souza Veras**  
Examinador Externo

# Agradecimentos

Primeiramente agradeço a Deus, que me permitiu todas as alegrias, conquistas e que me deu forças para lutar sem desistir ou recuar.

A minha família, pelo apoio durante todos os anos da minha vida. Cito meus pais, pela criação e por sempre enfatizarem a importância dos estudos, minhas irmãs que sempre caminharam comigo em momentos felizes, tristes ou de estresse.

Ao amigo, professor e orientador Anselmo, pela orientação durante esses quase dez anos de parceria, por apostar em mim durante minha graduação, mestrado e doutorado, e principalmente pelos conselhos que sempre foram além da vida acadêmica.

Ao amigo, co-orientador e professor Aristófanês, pelos conselhos, ajuda e puxões de orelha quando foi preciso.

Meus professores, pelo conhecimento que me proporcionaram. Em especial aos professores: João Dallyson, Geraldo e Simara por sempre se manterem dispostos a auxiliar de alguma forma.

Aos amigos que acompanham minha caminhada desde os primeiros períodos da graduação: João, Caio Eduardo, Giovanni, Ricardo, Johnatan e Kelson.

Todos os amigos do laboratório, pelas ideias, sugestões, brincadeiras, lanches, por tudo. Obrigado, me sinto em casa com vocês.

A todos que compõem a PPGEE-UFMA: professores, alunos e funcionários.

A CAPES, por conceder a bolsa de estudos, permitindo a viabilidade desta pesquisa.

Agradeço a todos que contribuíram direta ou indiretamente para a conclusão deste trabalho.

*“O único lugar onde sucesso vem antes do trabalho é no dicionário.”*

*(Albert Einstein)*

# Resumo

Esta tese apresenta dois métodos completamente automáticos para diagnóstico de degeneração macular relacionada à idade (DMRI) e edema macular diabético (EMD), por meio da utilização de técnicas de processamento de imagens e reconhecimento de padrões aplicadas a exames de tomografia de coerência óptica (OCT). A degeneração macular relacionada a idade e o edema macular diabético são doenças oculares que podem causar deficiência visual. A degeneração macular acomete principalmente idosos com mais de 50 anos, já o edema macular é uma consequência da retinopatia diabética que por sua vez acomete pessoas portadoras do diabetes. A tomografia de coerência óptica é um dos principais exames para a detecção e acompanhamento de doenças oculares. A avaliação das alterações causadas por DMRI e EMD a partir de imagens de OCT se dá através da avaliação dos sucessivos cortes seccionais na busca por alterações morfológicas. A utilização de sistemas CAD (*Computer-Aided Detection*) e CADx (*Computer-Aided Diagnosis*) tem contribuído para aumentar as chances de detecção e diagnósticos corretos, auxiliando os especialistas na tomada de decisões sobre o tratamento dessas doenças. Assim, desenvolveu-se um método que utiliza a segmentação das bordas das camadas da retina em imagens de OCT feita com duas redes neurais profundas, U-Net e DexiNed, para delimitar as bordas, e a classificação utiliza uma rede neural profunda residual. O segundo método desenvolvido para a detecção automática da DMRI e do EMD utiliza uma arquitetura *Capsule Network* com camadas baseadas no *Local Binary Pattern*. O método proposto de segmentação mostrou-se promissor, alcançando um erro médio absoluto geral de 0,49 *pixels* para a membrana limitante interna (MLI), 0,57 para o epitélio pigmentar retiniano (EPR) e 0,66 para a membrana de Bruch (MB). Os métodos de classificação atingiram respectivamente uma acurácia de 99% e 98% entre múltiplas classes (Drusas, Neovascularização Coroidal, EMD e Normal).

Palavras-chaves: Computer-Aided Diagnosis. Redes Neurais Convolucionais. Degeneração Macular. Edema Macular Diabético. Imagens médicas.

# Abstract

This thesis presents two fully automatic methods for diagnosing age-related macular degeneration (AMD) and diabetic macular edema (DME) through the use of image processing and pattern recognition techniques applied to optical coherence tomography (OCT) examinations. Age-related macular degeneration and diabetic macular edema are eye diseases that can cause visual impairment. Macular degeneration mainly affects older adults over 50 years of age, whereas macular edema is a consequence of diabetic retinopathy, affecting people with diabetes. Optical coherence tomography is one of the main tests for detecting and monitoring eye diseases. The evaluation of alterations caused by AMD and DME from OCT images is done by evaluating successive sectional cuts in search of morphological alterations. The use of CAD (Computer-Aided Detection) and CADx (Computer-Aided Diagnosis) systems has contributed to increasing the chances of correct detection and diagnosis, helping specialists make decisions about the treatment of these illnesses. Thus, a method was developed that uses segmentation of the edges of the retinal layers in OCT images made with two deep neural networks, U-Net and DexiNed, to delimit the edges. The classification uses a deep residual neural network. The second method for the automatic detection of AMD and DME uses a Capsule Network architecture with layers based on the Local Binary Pattern. The proposed segmentation method proved to be promising, reaching an overall mean absolute error of 0.49 pixels for the inner limiting membrane (ILM), 0.57 for the retinal pigment epithelium (RPE), and 0.66 for Bruch's membrane (BM). The classification methods reached, respectively, an accuracy of 99% and 98% among multiple classes (Drusen, Choroidal Neovascularization, DME, and Normal).

Keywords: Computer-Aided Diagnosis. Convolutional Neural Networks. Macular Degeneration. Diabetic Macular Edema. Medical images.

# Lista de ilustrações

|   |    |
|---|----|
| Figura 1 – Estruturas do Olho. . . . .  | 19 |
| Figura 2 – Corte histológico da região macular da retina humana com identificação das camadas. . . . .  | 21 |
| Figura 3 – Ilustração da consequência da DMRI na visão. . . . .   | 22 |
| Figura 4 – Comparação entre uma retina saudável (A), retina acometida por DMRI não-exsudativa com formação de drusas (B) e retina com neovascularização (C). . . . .                                    | 23 |
| Figura 5 – Efeitos do edema macular diabético. . . . .  | 24 |
| Figura 6 – Diagrama básico de um equipamento de OCT. . . . .  | 25 |
| Figura 7 – Esquema de aquisição de um tomógrafo de coerência óptica no domínio espectral. . . . .   | 26 |
| Figura 8 – Aquisição de SD-OCT. À esquerda, aquisição unidimensional ( <i>A-scan</i> ). Na imagem central, a imagem bidimensional ( <i>B-scan</i> ). Na imagem à direita, temos a aquisição 3D. . . . . | 26 |
| Figura 9 – <i>B-scan</i> da região central da mácula. . . . .   | 28 |
| Figura 10 – Imagem de OCT com destaque para o ruído característico inserido pelo processo de aquisição (ampliação no canto inferior direito). . . . .   | 28 |
| Figura 11 – Resultado do LBP . . . . .  | 30 |
| Figura 12 – Bloco residual. . . . .   | 32 |
| Figura 13 – Arquitetura da ResNet. . . . .  | 33 |
| Figura 14 – Fluxograma da arquitetura da <i>capsule network</i> . . . . .   | 34 |
| Figura 15 – Roteamento dinâmico para envio de informações da cápsula de nível inferior para a cápsula de nível superior. . . . .  | 36 |
| Figura 16 – Arquitetura da rede U-Net. . . . .  | 37 |
| Figura 17 – Arquitetura da rede DexiNed. . . . .  | 39 |
| Figura 18 – Fluxograma do método. . . . .   | 49 |
| Figura 19 – Exemplos de imagens da base. . . . .  | 50 |
| Figura 20 – Fluxograma do método de segmentação. . . . .  | 52 |
| Figura 21 – Etapa de Pré-processamento. . . . .   | 53 |

|  |    |
|--|----|
| Figura 22 – Processo para a área de interesse. . . . .   | 54 |
| Figura 23 – Imagem de entrada para a U-Net. . . . .  | 55 |
| Figura 24 – Arquitetura da U-Net. . . . .  | 55 |
| Figura 25 – Arquitetura da DexiNed. . . . .  | 56 |
| Figura 26 – Fluxograma da metodologia proposta. . . . .  | 59 |
| Figura 27 – Fluxograma da arquitetura <i>capsule network</i> com camadas LBC. . . . .  | 60 |
| Figura 28 – Resultados da U-Net. Na esquerda a imagem original da marcação e na<br>direita o resultado da rede. . . . .  | 64 |
| Figura 29 – Exemplos de casos de acerto. Imagens do exame “AMD_1057”, na coluna<br>da esquerda estão os resultados do método, na coluna da direita estão<br>as marcações da base. . . . .  | 66 |
| Figura 30 – Exemplos de casos de acerto. Imagens do exame “AMD _1090”, na<br>coluna da esquerda estão os resultados do método, na coluna da direita<br>estão as marcações da base. . . . . | 67 |
| Figura 31 – Exemplos de erros na segmentação. . . . .  | 67 |
| Figura 32 – Curvas de aprendizagem. . . . .  | 71 |

# Lista de tabelas

|  |    |
|--|----|
| Tabela 1 – Arquitetura da rede residual. . . . .   | 57 |
| Tabela 2 – Arquitetura da rede residual. . . . .   | 62 |
| Tabela 3 – Resultados da segmentação das bordas das camadas da retina. . . . .                       | 65 |
| Tabela 4 – Comparação com trabalhos relacionados. . . . .  | 68 |
| Tabela 5 – Resultados do experimento 1. . . . .  | 69 |
| Tabela 6 – Resultados do experimento 2 - base ZhangLab - LBCapsNet. . . . .                          | 69 |
| Tabela 7 – Resultados do experimento 2 - base Farsiu - LBCapsNet. . . . .                            | 69 |
| Tabela 8 – Resultados do experimento 2 - base ZhangLab - Método proposto com<br>segmentação. . . . . | 70 |
| Tabela 9 – Resultados do experimento 2 - base Farsiu - Método proposto com<br>segmentação. . . . .   | 70 |
| Tabela 10 – Tempo de execução do treinamento. . . . .  | 70 |
| Tabela 11 – Comparação com trabalhos relacionados. . . . .   | 73 |
| Tabela 12 – Comparação com trabalhos relacionados. . . . .   | 74 |

# Lista de abreviaturas e siglas

|         |   |
|---------|---|
| ACC     | Acurácia  |
| CAD     | <i>Computer-Aided Detection</i>                 |
| CADx    | <i>Computer-Aided Diagnostic</i>                |
| CNN     | <i>Convolutional Neural Networks</i>            |
| DH      | Distância de Hausdorff                          |
| DMRI    | Degeneração Macular Relacionada à Idade         |
| DexiNed | <i>Dense Extreme Inception Network</i>          |
| EMA     | Erro Médio Absoluto                             |
| EMD     | Edema Macular Diabético                         |
| EPR     | Epitélio Pigmentar Retiniano                    |
| EPRCD   | Epitélio Pigmentar Retiniano Complexo de Drusas |
| ESP     | Especificidade                                  |
| FN      | Falso Negativo                                  |
| FP      | Falso Positivo                                  |
| IBGE    | Instituto Brasileiro de Geografia e Estatística |
| LBP     | <i>Local Binary Pattern</i>                     |
| LBC     | <i>Local Binary Convolution Module</i>          |
| MB      | Membrana de Bruch                               |
| OCT     | <i>Optical Coherence Tomography</i>             |
| P       | Precisão  |
| ReLU    | <i>Rectified Linear Units</i>                   |

|        |   |
|--------|---|
| ResNet | <i>Residual Network</i>                             |
| ROI    | Região de Interesse                                 |
| SD-OCT | <i>Spectral Domain Optical Coherence Tomography</i> |
| SS-OCT | <i>Swept Source Optical Coherence Tomography</i>    |
| SEN    | Sensibilidade                                       |
| VN     | Verdadeiro Negativo                                 |
| VP     | Verdadeiro Positivo                                 |

# Sumário

|            |   |           |
|------------|---|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b>   | <b>15</b> |
| <b>1.1</b> | <b>Objetivos</b>  | <b>17</b> |
| 1.1.1      | Objetivos específicos   | 17        |
| <b>1.2</b> | <b>Contribuição científica</b>  | <b>18</b> |
| <b>1.3</b> | <b>Estrutura do trabalho</b>  | <b>18</b> |
| <b>2</b>   | <b>FUNDAMENTAÇÃO TEÓRICA</b>  | <b>19</b> |
| <b>2.1</b> | <b>Olho Humano e a Retina</b>   | <b>19</b> |
| <b>2.2</b> | <b>Degeneração Macular Relacionada à Idade - DMRI</b>                               | <b>20</b> |
| <b>2.3</b> | <b>Edema Macular Diabético - EMD</b>  | <b>23</b> |
| <b>2.4</b> | <b>Tomografia de Coerência Óptica - OCT</b>   | <b>24</b> |
| <b>2.5</b> | <b>Eliminação de Ruído em Imagens</b>   | <b>27</b> |
| 2.5.1      | Filtro Bilateral  | 28        |
| <b>2.6</b> | <b>Aprendizagem Profunda</b>  | <b>29</b> |
| 2.6.1      | <i>Local Binary Convolution Module</i> - LBC  | 30        |
| 2.6.2      | Rede Neural Residual - ResNet   | 31        |
| 2.6.3      | <i>Capsule Network</i> - CapsNet  | 33        |
| 2.6.4      | U-Net   | 37        |
| 2.6.5      | Dense Extreme Inception Network - DexiNed   | 38        |
| <b>2.7</b> | <b>Métricas de desempenho</b>   | <b>39</b> |
| 2.7.1      | Coeficiente Dice  | 40        |
| 2.7.2      | Avaliação de Erro Médio   | 40        |
| 2.7.3      | Distância de Hausdorff  | 41        |
| 2.7.4      | Validação do Classificador  | 41        |
| <b>3</b>   | <b>TRABALHOS RELACIONADOS</b>   | <b>43</b> |
| <b>3.1</b> | <b>Segmentação das Camadas da Retina em Imagens de OCT</b>                          | <b>43</b> |
| <b>3.2</b> | <b>Deteccção de Degeneração Macular e Edema Macular Diabético em Imagens de OCT</b> | <b>46</b> |

|            |   |           |
|------------|---|-----------|
| <b>4</b>   | <b>MÉTODO PARA CLASSIFICAÇÃO DE IMAGENS DE OCT UTILIZANDO A SEGMENTAÇÃO DAS CAMADAS DA RETINA .</b>                         | <b>49</b> |
| <b>4.1</b> | <b>Bases de Imagens . . . . .</b>   | <b>50</b> |
| 4.1.1      | Base ZhangLab . . . . .   | 50        |
| 4.1.2      | Base Farsiu . . . . .   | 51        |
| <b>4.2</b> | <b>Segmentação das Bordas das Camadas ILM, EPR e MB . . . . .</b>   | <b>52</b> |
| 4.2.1      | Pré-processamento . . . . .   | 53        |
| 4.2.2      | Definição da Região de Interesse . . . . .  | 53        |
| 4.2.3      | Segmentação Inicial . . . . .   | 53        |
| 4.2.4      | Segmentação Final . . . . .   | 56        |
| 4.2.5      | Validação . . . . .   | 57        |
| <b>4.3</b> | <b>Classificação . . . . .</b>  | <b>57</b> |
| 4.3.1      | Treinamento . . . . .   | 58        |
| <b>5</b>   | <b>MÉTODO PARA CLASSIFICAÇÃO DE IMAGENS DE OCT UTILIZANDO UMA NOVA ARQUITETURA DE CAPSULE NETWORK - LBCAPSNET . . . . .</b> | <b>59</b> |
| <b>5.1</b> | <b>Pré-processamento . . . . .</b>  | <b>59</b> |
| <b>5.2</b> | <b>Capsule Network com Local Binary Convolution Module - LBCapsNet</b>  | <b>60</b> |
| 5.2.1      | Treinamento . . . . .   | 61        |
| <b>6</b>   | <b>RESULTADOS E DISCUSSÃO . . . . .</b>   | <b>63</b> |
| <b>6.1</b> | <b>Hardware utilizado . . . . .</b>   | <b>63</b> |
| <b>6.2</b> | <b>Resultados da Segmentação das Camadas da Retina . . . . .</b>  | <b>63</b> |
| 6.2.1      | Comparação com Trabalhos Relacionados . . . . .   | 67        |
| <b>6.3</b> | <b>Resultados dos Métodos para Detecção de DMRI e EMD . . . . .</b>   | <b>68</b> |
| 6.3.1      | Comparação com Trabalhos Relacionados . . . . .   | 73        |
| <b>7</b>   | <b>CONCLUSÃO . . . . .</b>  | <b>75</b> |
| <b>7.1</b> | <b>Trabalhos Futuros . . . . .</b>  | <b>76</b> |
| <b>7.2</b> | <b>Produção científica . . . . .</b>  | <b>76</b> |
|            | <b>REFERÊNCIAS . . . . .</b>  | <b>77</b> |

# 1 Introdução

Segundo dados da *Organização Mundial da Saúde* (2019) estima-se que pelo menos 2,2 bilhões de pessoas tenham deficiência visual ou cegueira, das quais pelo menos 1 bilhão tenham alguma deficiência visual que poderia ter sido evitada ou ainda não foi tratada. As cinco doenças que mais causam cegueira são: erros de refração não corrigidos, catarata, degeneração macular relacionada à idade, glaucoma e retinopatia diabética.

O Instituto Brasileiro de Geografia e Estatística (IBGE) estima que em 2060, um quarto da população do Brasil será de pessoas com mais de 60 anos. A degeneração macular relacionada à idade (DMRI) é uma doença ocular que acomete principalmente pessoas com mais de 50 anos, sendo uma das principais causas de deficiência visual em pacientes idosos (CHOU et al., 2013). É uma condição degenerativa e progressiva, decorrente do envelhecimento, relacionada a alterações na estrutura da retina, causando sintomas de perda da visão central de um ou ambos os olhos por disfunção dos fotorreceptores (RIM et al., 2012). Pode se apresentar em duas formas, sendo mais comum a forma seca ou não exsudativa, identificada em 85 a 90% dos casos. A forma seca é caracterizada pela presença de drusas <sup>1</sup> e/ou alteração de pigmentação, determinando perda visual mais lenta. Já a forma neovascular ou exsudativa ocorre em 10 a 15% dos casos, no entanto é responsável por 90% dos casos de cegueira por DRMI (RIM et al., 2012). A DMRI é a principal causa de perda visual na terceira idade e a terceira maior causa de cegueira no Brasil, acometendo cerca de 3 milhões de idosos (OFTALMOLOGIA, 2019).

Já o edema macular diabético (EMD), é uma doença que decorre da retinopatia diabética, que por sua vez acomete pacientes portadores do diabetes. É um tipo de distúrbio macular onde os vasos sanguíneos fracos começam a vazar líquido ou depósitos de proteínas nas camadas da retina. O edema macular é também causado devido a cirurgias oculares, especialmente cirurgias de catarata onde os vasos sanguíneos sob a mácula ficam irritados e vazam fluido formando cístico<sup>2</sup> nas camadas da retina. Edema macular pode ser causado por outros problemas, como, DMRI, uveíte<sup>3</sup>, oclusão da veia retiniana, telangiectasia

---

<sup>1</sup> Drusas são depósitos extracelulares patológicos, de formato redondo e amarelado de material degenerativo (BUSCHINI et al., 2011).

<sup>2</sup> Cístico refere-se a cisto, que vem a ser um tumor benigno revestido por uma membrana de tecido.

<sup>3</sup> Uveíte é a inflamação da úvea, a camada média do globo ocular.

macular (bloqueio nas veias pequenas da retina devido à radiação) e distúrbios genéticos (como retinose pigmentar<sup>4</sup>) (HASSAN et al., 2015).

O edema macular diabético é a maior causa de perda de visão central nos pacientes com retinopatia diabética. A porcentagem de pacientes diabéticos com algum grau de retinopatia diabética aumenta em função do tempo de instalação da doença: estima-se que, após 15 anos, a prevalência do edema macular é cerca de 20% nos portadores de diabetes do tipo 1 e 18% nos do tipo 2, após 25 anos, 80% dos pacientes apresentarão algum grau de retinopatia diabética (MOTTA, 2019). No Brasil, estima-se que a cegueira entre diabéticos possa alcançar a prevalência de 4,8% (ÁVILA; ALVES; NISHI, 2016).

O diagnóstico clínico da DMRI e do EMD é comumente feito pela retinografia (fotografia do fundo do olho), por ser mais barato, mas outro exame, a tomografia de coerência óptica (OCT, do inglês *optical coherence tomography*), por ter maior precisão, têm sido bastante utilizada em conjunto com a retinografia.

A OCT surgiu como uma opção viável para o diagnóstico de alterações da retina em 1991 (HUANG et al., 1991a). Trata-se de uma tecnologia não-invasiva de captação de imagens que permite a visualização de tecidos biológicos *in vivo* em cortes seccionais com alta resolução. Assim, logo ganhou grande importância no diagnóstico de doenças maculares como a DMRI e o EMD. Drusas, neovascularizações, cicatrizes fibrovasculares e edema macular intraretinal ou sub-retiniano são alterações típicas da retina que são detectáveis na tomografia de coerência óptica.

Especificamente para o diagnóstico da DMRI, a imagem da tomografia de coerência óptica permite a avaliação de alterações no epitélio pigmentar retiniano, de drusas moles e de deslocamentos da retina (HEE et al., 1996). Em relação aos outros exames recomendados, foi evidenciado que a OCT tem maior sensibilidade à detecção das alterações patológicas relacionadas à DMRI. Foi verificado que 57% dos olhos avaliados apresentavam alterações percebidas somente através da OCT e não visualizadas através de outros exames, como a angiofluoresceinografia (ETER; SPAIDE, 2005). Outra vantagem em relação à retinografia é a rápida aquisição dispensando a utilização de contrastes e assim os riscos de reações alérgicas.

Para o edema macular, as imagens de retinografia e de OCT podem fazer a detecção a partir da medição da espessura da retina. A imagem do fundo de olho mostra apresenta

---

<sup>4</sup> Retinose pigmentar é um conjunto de doenças hereditárias que causam degeneração na retina, região do fundo do olho.

ao especialista a espessura e exsudatos duros <sup>5</sup> na região macular, a imagem de OCT mostra as mesmas alterações, fornecendo um corte transversal, com vista das camadas sub-retinianas. É muito difícil detectar edema macular precoce a partir de imagens de fundo de olho, em imagem de OCT podemos observar facilmente algumas alterações provocadas pela doença, como a cística, enquanto eles não são proeminentes na retinografia (HASSAN et al., 2015).

O diagnóstico precoce da DMRI e do EMD é essencial para o controle destas doenças sem resultar em sequelas aos pacientes, e esse diagnóstico quando feito a partir da imagem do exame de OCT, se dá com a análise visual das várias imagens geradas, e se levarmos em conta que um especialista atende e acompanha vários pacientes por dia, é necessário um apoio tecnológico para a prática clínica. Assim vários métodos computacionais foram e estão sendo pesquisados para a prevenção e detecção de tais doenças. Neste sentido, estamos propondo dois métodos para classificação imagens 2D de exames OCT normais, com DMRI e EMD.

## 1.1 Objetivos

O objetivo geral desta tese é desenvolver métodos computacionais baseados em técnicas de processamento de imagens e aprendizagem profunda para a detecção automática da degeneração macular e do edema macular diabético em imagens de tomografia de coerência óptica.

### 1.1.1 Objetivos específicos

Desta maneira pretende-se:

- propor uma técnica de remoção de ruídos que viabilizem melhores características para segmentação de imagens de OCT e proveja melhores resultados na classificação de imagens de OCT;
- Propor arquiteturas de aprendizado profundo adequadas à tarefa de segmentação das camadas da retina e classificação de imagens de OCT;

---

<sup>5</sup> Exsudatos duros são lesões amareladas causadas por edemas retinianos, são um dos primeiros sinais que podem ser observados da retinopatia diabética (SOUZA, 2019).

- validar os métodos através de métricas comumente usadas em trabalhos de imagens médicas;
- comparar o método proposto com metodologias já consolidadas na literatura.

## 1.2 Contribuição científica

A partir dos métodos propostos podemos destacar as seguintes contribuições:

- aplicação inovadora da arquitetura DexiNed (SORIA; RIBA; SAPP, 2020) em imagens médicas com resultados promissores;
- método que utiliza o resultado da segmentação de bordas para fazer a classificação de imagens de OCT;
- método que utiliza uma nova arquitetura baseada em *capsule network* (XI; BING; JIN, 2017) que utiliza uma camada que descreve o *Local Binary Pattern* para classificação de imagens de OCT;

## 1.3 Estrutura do trabalho

Além deste capítulo de introdução, esta tese ainda está organizado em mais 6 capítulos.

O Capítulo 2 apresentará conceitos fundamentais pertinentes para o entendimento do método proposto.

No Capítulo 3 será apresentado o estado da arte e como os métodos propostos preenchem lacunas nos trabalhos já existentes.

Nos Capítulos 4 e 5 serão apresentados os materiais e os métodos utilizados para segmentar as camadas da retina e classificar imagens de OCT com a degeneração macular e o edema macular.

Os resultados alcançados na aplicação de cada método e discussão dos mesmos serão mostrados no Capítulo 6.

Por fim, no Capítulo 7 serão apresentadas as conclusões do trabalho, onde é feita uma avaliação do trabalho exposto e planos futuros para melhoramento do mesmo.

## 2 Fundamentação teórica

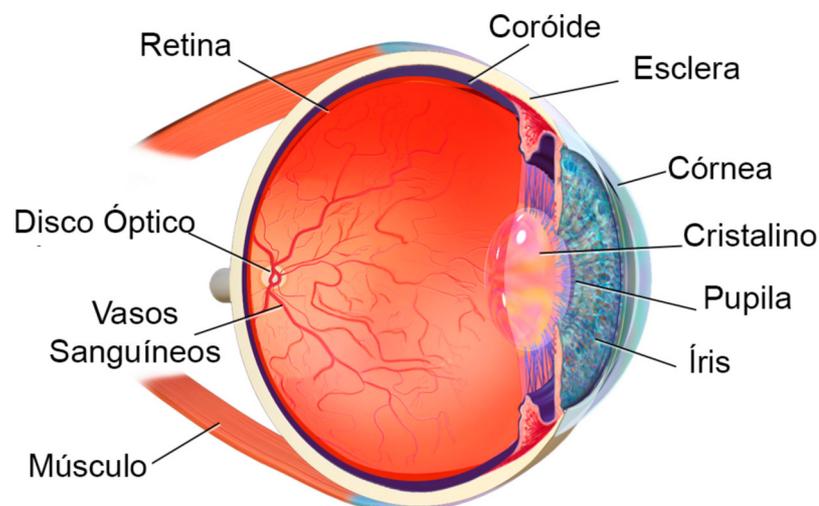
Este capítulo apresenta conceitos importantes para compreensão dos métodos propostos. Serão abordados conceitos como: anatomia do olho humano e da retina, degeneração macular e edema macular, eliminação de ruídos, aprendizagem profunda e métricas de desempenho.

### 2.1 Olho Humano e a Retina

O olho humano é responsável pela capacidade do homem de interagir visualmente com o meio ambiente que o rodeia. Suas estruturas são responsáveis pela refração e focalização dos raios luminosos e, por fim, sua conversão em impulsos elétricos que são transmitidos ao cérebro através do nervo óptico.

O globo ocular (Figura 1) tem formato esférico e um diâmetro de cerca de  $20mm$ . Ele é envolvido por três membranas: a camada externa formada pela córnea e pela esclerótica, a coróide e a retina, que é a camada interna. A córnea é uma película transparente que cobre a parte anterior do olho. Dando continuidade à córnea, a esclera é uma membrana opaca que reveste o globo ocular (MARQUES FILHO; NETO, 1999).

Figura 1 – Estruturas do Olho.



Fonte: (STAFF, 2014).

A coróide está situada abaixo da esclerótica. Essa membrana contém uma rede de vasos sanguíneos que servem como a principal fonte de nutrição do olho. O revestimento da coróide é fortemente pigmentado, o que ajuda a reduzir a quantidade de luz que entra no olho. Ela é dividida em corpo ciliar e diafragma da íris, sendo este último responsável por controlar a quantidade de luz que deve penetrar no olho. A parte frontal da íris contém o pigmento visível do olho, enquanto sua porção posterior possui um pigmento negro (MARQUES FILHO; NETO, 1999).

A membrana mais interna do olho é a retina, situada na sua parede posterior. A retina é uma estrutura multicamadas composta por células nervosas que recebem os estímulos luminosos e os convertem em estímulos elétricos. Essa estrutura contém duas partes: a parte óptica da retina, situada nas regiões posterior e lateral, e a parte cega, situada na região frontal da retina. Na parte óptica temos duas estruturas de maior destaque, o disco óptico e a mácula.

O disco óptico é o ponto onde o nervo óptico sai da retina. Esse local não possui células fotorreceptoras, sendo conhecido como o ponto cego da retina. Esta região tem a coloração mais clara e do seu centro partem ramos da artéria central da retina. A região da retina que apresenta maior espessura é denominada mácula (SANTOS, 2019).

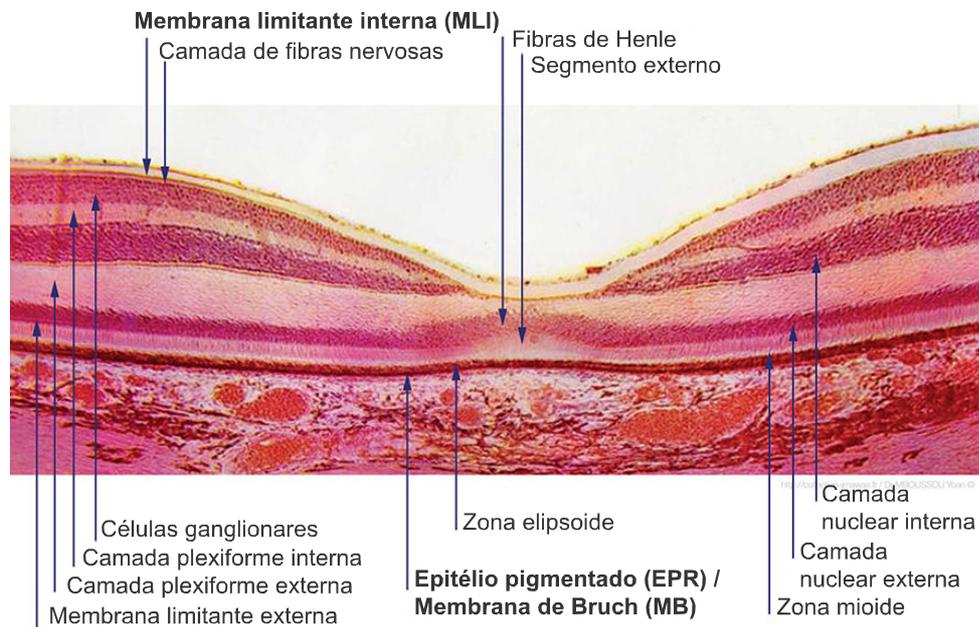
A mácula situa-se temporalmente ao disco óptico e possui formato ovalado com aproximadamente 5,5 *mm* de diâmetro. No centro da mácula encontra-se a fóvea, uma depressão da retina que constitui o local de maior acuidade visual. Alterações morfológicas nas camadas retinianas desta área estão diretamente relacionados à diminuição ou perda da capacidade visual. Na Figura 2 é apresentado um corte histológico da região de uma mácula saudável com a identificação das camadas (SANTOS, 2019).

As camadas de interesse para esta tese são a membrana limitante interna (MLI), o epitélio pigmentado retiniano (EPR) e a membrana de Bruch (MB). A MLI delimita a retina internamente e a MB externamente. O EPR é a camada que sofre alterações diretamente quando há o acometimento por DMRI e EMD.

## 2.2 Degeneração Macular Relacionada à Idade - DMRI

A Degeneração Macular Relacionada à Idade é uma das principais causas de cegueira. A idade avançada é o principal fator de risco para DMRI. A incidência e gravidade aumentam conforme aumenta a idade do paciente. Outros fatores de risco são:

Figura 2 – Corte histológico da região macular da retina humana com identificação das camadas.



Fonte: (SANTOS, 2019).

ascendência caucasiana, tabagismo, história familiar de DMRI, sedentarismo, obesidade e doenças cardiovasculares, como hipertensão e elevação do nível de colesterol no sangue (SOON-CHUN, 2015).

A DMRI atinge a região macular, área central da retina, provocando a degeneração progressiva de fotorreceptores, epitélio pigmentar retiniano (EPR) e coróide. Frequentemente, a DMRI provoca atrofia do EPR e/ou neovascularização na coróide, levando à perda de visão central (YANOFF; DUKER; AUGSBURGER, 2014). O comprometimento da visão é ilustrado pela Figura 3, onde se percebe perda de visão que começa na região central e se estende na direção das regiões periféricas.

O processo de desenvolvimento da DMRI se inicia com o depósito de lipídios na MB. Estes depósitos são conhecidos como depósitos basais laminares que, posteriormente serão classificados como drusas (YANOFF; DUKER; AUGSBURGER, 2014).

As drusas são formadas entre a lâmina basal do EPR e a camada colágena interna da membrana de Bruch. Podem ser divididas em duras ou moles. As primeiras são redondas e pequenas, de cor branco-amarelada e geralmente inócuas. Já as drusas moles são maiores e sem margens distintas, podendo aumentar de tamanho e agregar-se no chamado complexo

Figura 3 – Ilustração da consequência da DMRI na visão.



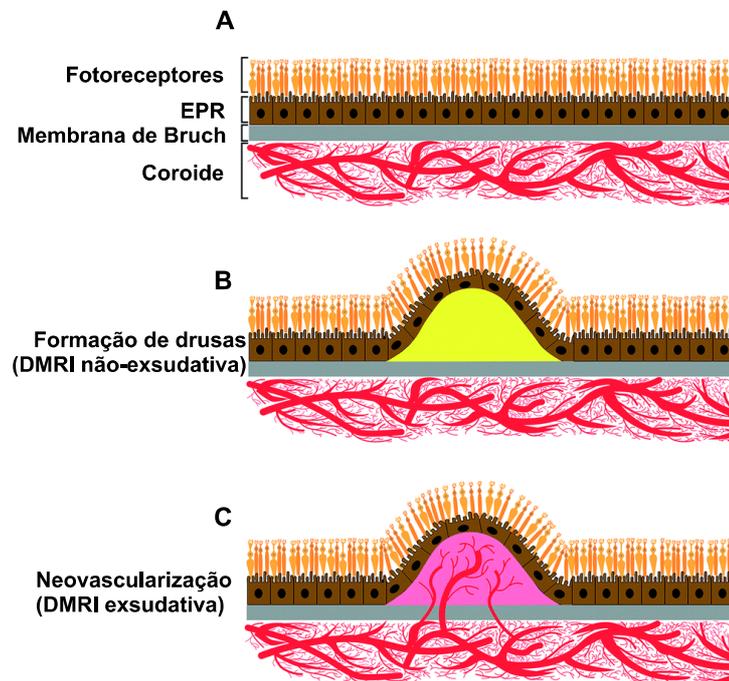
Fonte: (CBO. Conselho Brasileiro de Oftalmologia, 2017).

de drusas (drusenóide). Além disso, podem passar por processo de calcificação (KANSKI, 2012).

A DMRI pode ser classificada em não-exsudativa (ou seca) e exsudativa (ou úmida). A manifestação não-exsudativa é caracterizada pela presença de drusas geralmente duras, e atrofia do EPR, com perda de fotorreceptores e progressiva diminuição da visão central.

Já na DMRI exsudativa, ocorre o aparecimento de uma membrana neovascular sub-retiniana, com extravasamento de sangue e/ou soro, alterando a anatomia macular, com perda devastadora de fotorreceptores locais. Neste tipo da doença, a consequente baixa da visão é muito mais rápida e acentuada quando comparada à forma não-exsudativa. A presença de drusas moles é considerada fator predisponente para o desenvolvimento de DMRI exsudativa (NEHEMY, 2006; KANSKI, 2012). A Figura 4 ilustra a manifestação de drusas nas formas exsudativa e não exsudativa da DMRI. Podemos observar o atrofiamento dos fotorreceptores provocado pela presença das drusas (B e C) em comparação com uma retina saudável (A).

Figura 4 – Comparação entre uma retina saudável (A), retina acometida por DMRI não-exsudativa com formação de drusas (B) e retina com neovascularização (C).



Fonte: (FARJOOD; VARGIS, 2018).

## 2.3 Edema Macular Diabético - EMD

A diabetes é considerada pela OMS a 3<sup>a</sup> maior causa de morte no Brasil. De acordo com o Vigitel 2007 (Sistema de Monitoramento de Fatores de Risco e Proteção para Doenças Crônicas Não Transmissíveis), a ocorrência média de diabetes na população adulta (acima de 18 anos) é de 5,2%, o que representa quase sete milhões de pessoas que confirmaram ser portadoras da doença. Entretanto, este número atinge proporções muito maiores ao considerarmos as crianças (cerca de 5 milhões), portadoras de *diabetes mellitus* tipo 1 (RD1), mais o contingente de portadores que não sabem que têm a doença. A prevalência aumenta com a idade: o diabetes atinge 18,6% da população com idade superior a 65 anos (ÁVILA; ALVES; NISHI, 2016).

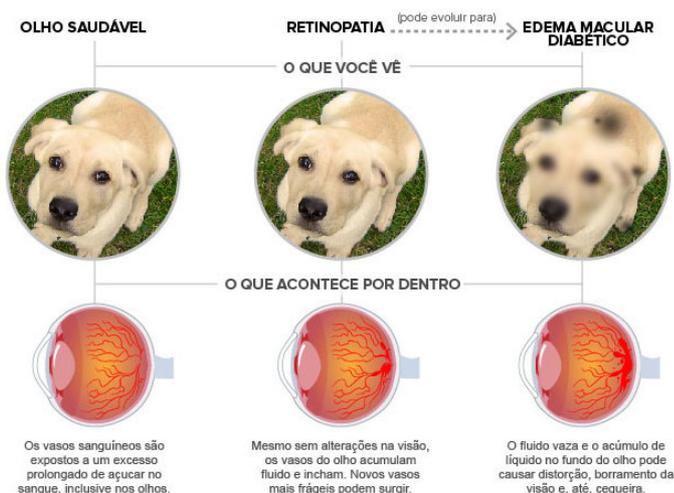
O edema macular diabético (EMD), também conhecido como maculopatia diabética, é caracterizado pelo acúmulo de líquido na mácula, que é a área da retina responsável pela visão central nítida, usada para ler, reconhecer rostos, cores e dirigir (NOVARTIS, 2017).

O edema macular diabético é uma consequência da retinopatia diabética. O EMD pode ocorrer em qualquer estágio da retinopatia diabética, embora o risco aumente de

acordo com a gravidade da retinopatia diabética. Cerca de metade das pessoas com retinopatia diabética desenvolve também EMD (NOVARTIS, 2017).

A principal causa do desenvolvimento do EMD são alterações estruturais nos vasos da retina causadas pela hiperglicemia (altos níveis de açúcar no sangue). Quando o diabetes não está controlado, a hiperglicemia desencadeia várias alterações no nosso corpo, como aumento do fator de crescimento endotelial vascular (VEGF), danos na estrutura do DNA e de outros componentes celulares (danos oxidativos) e da inflamação crônica. Essas alterações lesam os vasos sanguíneos e aumentam a permeabilidade vascular da retina. À medida que se tornam mais permeáveis, líquidos, proteínas e outras moléculas vazam de dentro dos vasos sanguíneos e se acumulam próximos da retina e da mácula formando o edema macular diabético (NOVARTIS, 2017). A Figura 5 ilustra os efeitos da doença nos vasos sanguíneos e na visão.

Figura 5 – Efeitos do edema macular diabético.



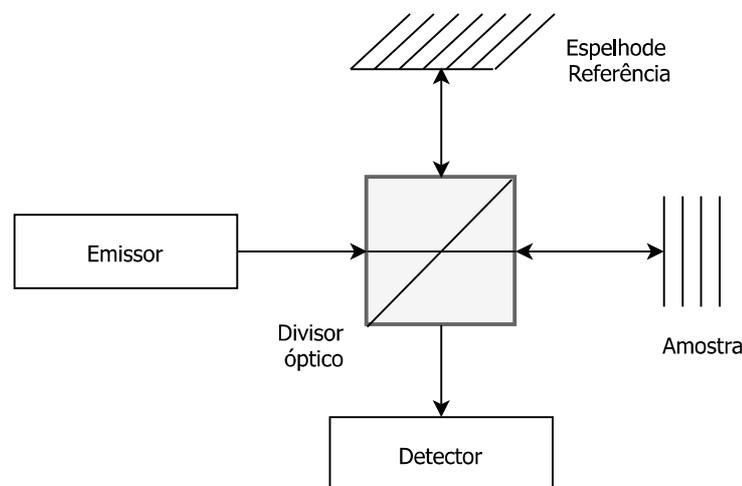
Fonte: (MOTTA; COBLENTZ; MELO, 2008)

## 2.4 Tomografia de Coerência Óptica - OCT

A tomografia de coerência óptica (OCT) é um método de aquisição de imagens, não-invasiva, que detalha estruturas da retina. As imagens são capturadas em resolução micrométrica em duas ou três dimensões a partir da sondagem por um sinal luminoso de baixa coerência. Essa tecnologia possui aplicações na indústria e na área médica e boa aceitação na área oftalmológica (SANTOS, 2019).

O funcionamento de um tomógrafo de OCT se dá a partir de uma fonte de luz de baixa coerência, que emite um feixe que passa por um divisor óptico onde é dividido em dois. Uma parte é direcionada a um ponto da amostra e a outra a um espelho de referência. A luz refletida pela amostra é recombinada à que foi refletida pelo espelho de referência dando origem a um padrão de interferência que é captado pelo detector. O emissor aponta para um ponto imediatamente vizinho e o processo é repetido por toda a região de interesse. A Figura 6 ilustra o princípio por trás do exame de OCT.

Figura 6 – Diagrama básico de um equipamento de OCT.

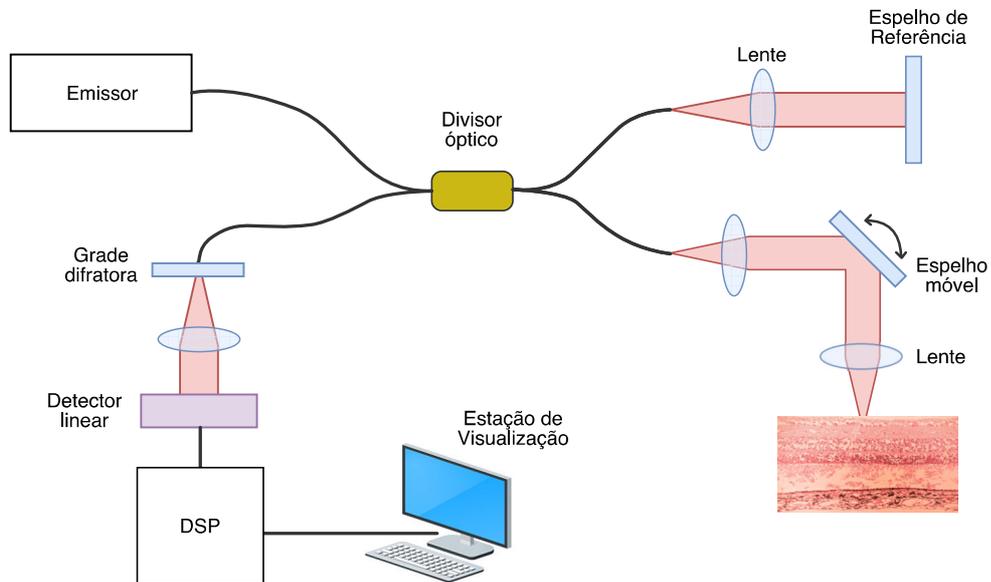


Fonte: (SANTOS, 2019).

Os tomógrafos ópticos atuais se baseiam na captação no domínio da frequência e se dividem em duas categorias: OCT de domínio espectral (SD-OCT, do inglês *spectral domain optical coherence tomography*) e OCT de fonte de varredura (SS-OCT, do inglês *swept source optical coherence tomography*). A SD-OCT é a tecnologia mais utilizada atualmente (SANTOS, 2019).

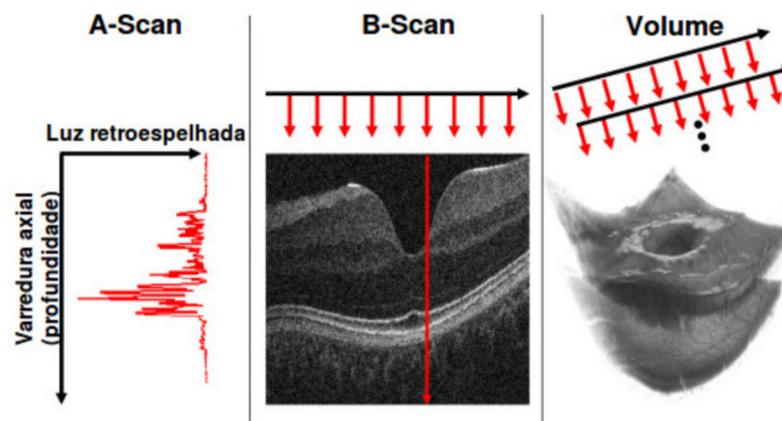
Na SD-OCT, a aquisição da imagem também se baseia na divisão de um feixe de luz em dois segmentos e na captura do padrão de interferência capturado. Entretanto, conforme apresentado na Figura 7, o sinal de interferência é disperso por uma grade difratora e analisado por um espectrômetro de detecção linear. Cada linha de profundidade da imagem (*A-scan*) é reconstruída pelo cálculo da transformada inversa de Fourier (DUKER; WAHEED; GOLDMAN, 2015) obtida pelo detector. A partir dos *A-scans* adquiridos sucessivamente, é realizada a reconstrução da imagem tomográfica (*B-scan*) ou até mesmo uma reconstrução volumétrica da região escaneada (Figura 8).

Figura 7 – Esquema de aquisição de um tomógrafo de coerência óptica no domínio espectral.



Fonte: (SANTOS, 2019).

Figura 8 – Aquisição de SD-OCT. À esquerda, aquisição unidimensional (*A-scan*). Na imagem central, a imagem bidimensional (*B-scan*). Na imagem à direita, temos a aquisição 3D.



Fonte: (SANTOS, 2019).

O sinal  $I(\omega)$  adquirido no domínio da frequência pode ser representado por (ALI; PARLAPALLI, 2010):

$$I(\omega) \approx K_1 S(\omega) + K_2 Re\{S(\omega)H(\omega)\} \quad (2.1)$$

O termo  $K_1S(\omega)$  representa um valor DC (valor médio) que é comumente subtraído do sinal capturado para eliminação de ruído e sinal de fundo, resultando em:

$$I'(\omega) = K_2Re\{S(\omega)H(\omega)\} \quad (2.2)$$

onde  $K_2Re\{S(\omega)H(\omega)\}$  representa a parte real da interferência capturada entre o sinal de referência  $S(\omega)$  e o sinal refletido pela amostra  $H(\omega)$ .

A reconstrução de uma (*A-scan*) é realizada pela aplicação da transformada inversa de Fourier:

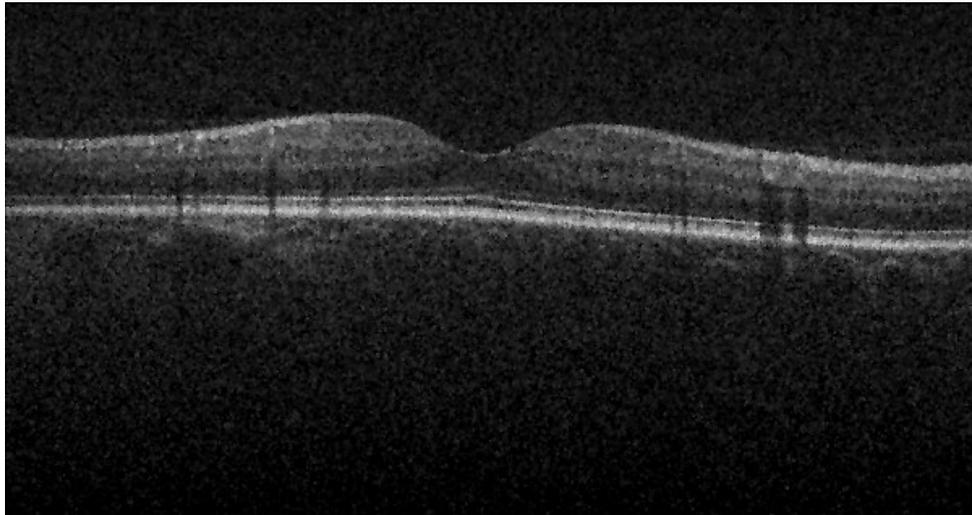
$$i(t) = FT^{-1}(I'(\omega)) \quad (2.3)$$

A principal vantagem da SD-OCT é que a aquisição de uma *A-scan* individual é realizada inteiramente a partir da análise do sinal em um único espectro e sem a necessidade da movimentação mecânica do conjunto de detecção. O sinal medido a partir da amostra representa a combinação das propriedades de refletividade, absorção e espalhamento do feixe de luz pelas camadas do tecido estudado.

A imagem de SD-OCT resultante é um mapa da refletividade da amostra representado em escala de cinza. Áreas escuras representam material homogêneo com baixa refletividade (ar ou fluidos) e áreas mais claras representam as áreas de maior refletividade como fibras colágenas, paredes e núcleos celulares (DEBUC, 2011). O sinal luminoso é atenuado nos tecidos de forma exponencial, reduzindo a intensidade da imagem com a profundidade. Vasos sanguíneos apresentam maior atenuação que tecidos com colágeno e gordura e fluídos são os que possuem menor atenuação. Uma imagem típica de OCT da região macular adquirida pela tecnologia espectral é apresentada na Figura 9. Nessa imagem podemos observar a variação de intensidade entre as camadas da estrutura da retina e áreas de sombra devido à presença de estruturas de alta refletividade (SANTOS, 2019).

## 2.5 Eliminação de Ruído em Imagens

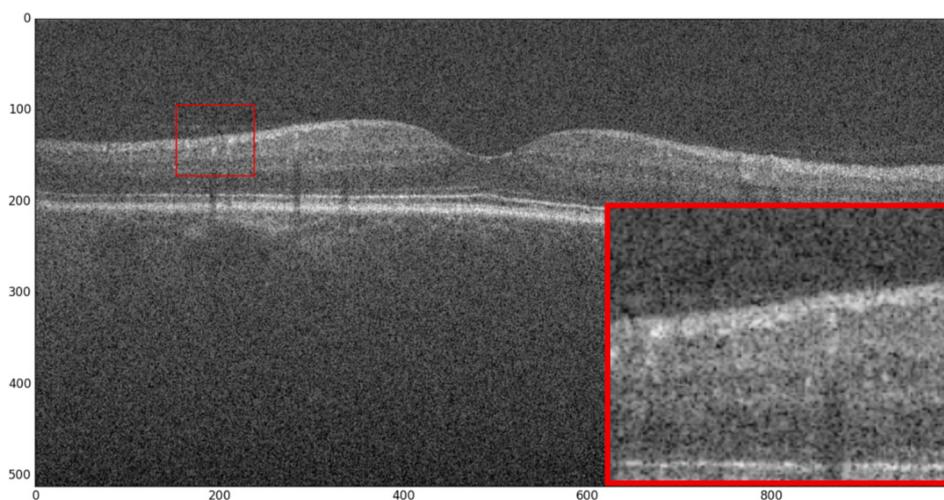
A presença de ruído em um sinal é uma característica indesejada. Ruídos podem ser introduzidos através do processo de aquisição, armazenamento, transmissão, etc (RUDIN; OSHER; FATEMI, 1992). São anomalias que dificultam o processamento das imagens,

Figura 9 – *B-scan* da região central da mácula.

Fonte: (FARSIU et al., 2014)

podendo diminuir a eficiência de um método ou, dependendo da intensidade, até mesmo inviabilizar o processamento.

Figura 10 – Imagem de OCT com destaque para o ruído característico inserido pelo processo de aquisição (ampliação no canto inferior direito).



Fonte: (SANTOS, 2019).

### 2.5.1 Filtro Bilateral

O filtro bilateral atua no domínio do espaço e foi proposto por Tomasi e Manduchi (1998). Ele é um filtro de suavização (*smoothing*) não linear, que preserva as bordas e borra (*blur*) as demais regiões. Para isto, cada *pixel* da imagem é substituído por uma média

ponderada dos *pixels* vizinhos, considerando a proximidade geométrica e sua similaridade fotométrica.

O filtro bilateral (BF) é definido pela Equação 2.4,

$$BF[I] = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(\|I_p - I_q\|) I_q \quad (2.4)$$

onde  $W_p$  é um fator de normalização dado pela Equação 2.5,

$$W_p = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(\|I_p - I_q\|) \quad (2.5)$$

A equação 2.5 é uma média normalizada dos pesos onde  $G_{\sigma_s}$  é uma gaussiana espacial que decresce de acordo com a distância do *pixel*,  $G_{\sigma_r}$  é uma gaussiana no range.

O resultado do filtro bilateral depende basicamente da escolha de três parâmetros:  $d$ , que é o diâmetro no qual são considerados os *pixels* vizinhos;  $\sigma_c$ , que é o fator de propagação fotométrica, ou seja, é o fator que determina o peso da cor no cálculo e  $\sigma_s$ , que é o fator de propagação geométrica, isto é, o fator que determina o quanto a proximidade do *pixel* interfere no cálculo. A escolha destes parâmetros depende da aplicação do filtro.

## 2.6 Aprendizagem Profunda

Durante o aprendizado, os humanos e os animais são levados primeiro a interpretar e entender conceitos mais simples, para mais tarde, no decorrer da sua vida, serem capazes de aprender abstrações de mais alto nível com base nos conceitos previamente observados (FERNANDES, 2013). É chamada de aprendizagem profunda o tipo de aprendizado onde sugere-se uma divisão em camadas hierárquicas com diferentes responsabilidades no cérebro (HUBEL, 1963; HUBEL; WIESEL, 1998).

Os métodos de aprendizagem profunda apresentam várias camadas de processamento não-linear para reconhecimento de padrões de forma análoga às hipóteses sobre o cérebro (FERNANDES, 2013). Tais métodos permitem a extração automática de características sobre os dados de treinamento, dessa forma, os modelos de arquitetura profunda podem aprender com menor envolvimento humano na construção do modelo antes do treinamento, pois integram as etapas de extração de características e classificação em um mesmo modelo (BENGIO; LECUN et al., 2007).

Atualmente, existem várias técnicas de aprendizagem profunda, entre elas as redes neurais convolucionais, as redes neurais recorrentes, as *Generative Adversarial Networks* (GANs), as redes neurais baseadas em cápsulas, entre outras.

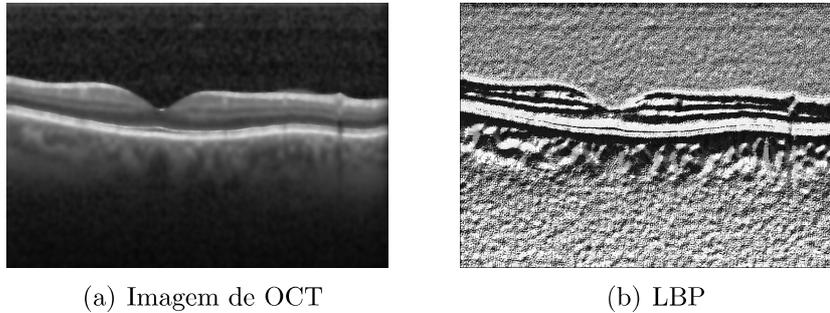
### 2.6.1 Local Binary Convolution Module - LBC

*Local Binary Pattern* (LBP) é um operador não paramétrico que pode ser usado para descrever a estrutura espacial local da imagem mostrando a alta capacidade de distinguir características de textura (OJALA; PIETIKÄINEN; HARWOOD, 1996). É calculado de acordo com a Equação 2.6.

$$LBP(x_c, y_c) = \sum_{n=0}^{n-1} S(i_n - i_c) 2^n \quad (2.6)$$

onde  $n$  é o número de vizinhos do *pixel* central  $x_c, y_c$  considerado no cálculo,  $i_c$  é o valor do nível de cinza do *pixel* central  $(x_c, y_c)$ ,  $i_n$  é o valor do nível de cinza de cada *pixel* vizinho e  $S(x)$  uma função que retorna 1 se  $x \geq 0$  e 0 caso contrário. Na Figura 11 é apresentado um exemplo da aplicação do LBP em imagem de OCT.

Figura 11 – Resultado do LBP



Em (JUEFEI-XU; BODDETI; SAVVIDES, 2017) foi desenvolvido um módulo convolucional que utiliza a codificação do LBP com filtros convolucionais. Na codificação uma convolução simples de toda a imagem com oito filtros convolucionais  $3 \times 3$ , seguidos de binarização simples, podem atingir o mesmo objetivo que o LBP original.

Cada filtro de convolução é um filtro *2-sparse difference*. A formulação padrão de LBP é uma soma de todos os mapas de bits usando um vetor de pesos predefinidos  $v = [2^7, 2^6, 2^5, 2^4, 2^3, 2^2, 2^1, 2^0]$ . Portanto, a extração padrões utilizando o LBP pode ser reformulada como  $y = \sum_{i=1}^8 \sigma(b_i * x) \cdot v_i$ , onde  $x \in \mathbb{R}^d$  é a versão vetorizada da imagem original,  $b^i$  são os filtros convolucionais esparsos,  $\sigma$  é o operador de binarização não linear,

a *heaviside step function* neste caso, e  $y \in \mathbb{R}^d$  é a imagem LBP resultante. Alterando adequadamente os pesos lineares  $v$ , a base e a ordem da codificação pode variar. Da mesma forma, alterar adequadamente os valores diferentes de zero (+1 e -1) nos filtros convolucionais nos permitem mudar o pivô. A reformulação do LBP, conforme descrito acima, forma a base da camada LBC (JUEFEI-XU; BODDETI; SAVVIDES, 2017).

Para permitir o *back propagation*, foi substituída a função não diferenciável, *Heaviside step function*, do LBP por uma função de ativação diferenciável (sigmoid ou ReLU). Os mapas de  $m$  bits são linearmente combinados por  $m$  pesos aprendíveis  $\nu_{l,i}$ ,  $i \in [m]$  para gerar um canal com o resultado final da camada LBC. A camada LBC corresponde a uma entrada  $x_l$  multicanal, que pode ser expresso como:

$$x_{l+1}^t = \sum_{i=1}^m \sigma \left( \sum_s b_i^{st} * x_l^s \right) \cdot \nu_{l,i}^t \quad (2.7)$$

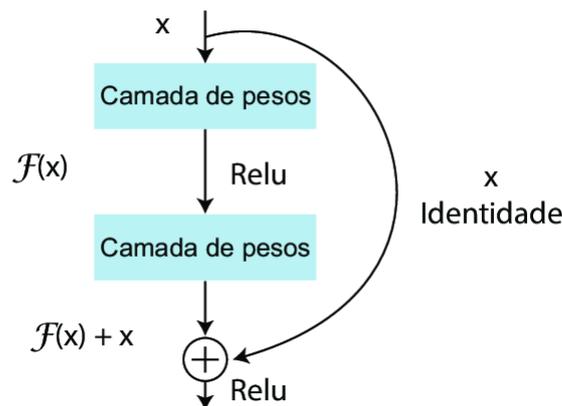
onde  $t$  é o canal de saída e  $s$  é o canal de entrada. Cada camada LBC consiste em duas camadas convolucionais, onde os pesos na primeira camada convolucional são fixos e não aprendem, enquanto os pesos na segunda camada convolucional são aprendíveis. O número de parâmetros aprendíveis na camada LBC são significativamente menores que os de uma camada convolucional padrão para o mesmo tamanho de núcleo convolucional e número de canais de entradas e saídas. Sendo os canais de entrada e saída,  $p$  e  $q$ , respectivamente. Com um núcleo convolucional de tamanho de  $h \times w$ , uma camada convolucional padrão consiste em  $p \cdot h \cdot w \cdot q$  parâmetros aprendidos. A camada LBC correspondente consiste em pesos fixos  $p \cdot h \cdot w \cdot m$  e  $m \cdot q$  que são parâmetros aprendíveis (correspondente à convolução  $1 \times 1$ ), em que  $m$  é o número de canais intermediários da camada LBC, que é essencialmente o número de filtros LBC. As convoluções  $1 \times 1$  agem nos  $m$  mapas de ativação dos filtros fixos para gerar a saída do canal  $q$  (JUEFEI-XU; BODDETI; SAVVIDES, 2017).

## 2.6.2 Rede Neural Residual - ResNet

A ResNet é uma rede convolucional clássica muito utilizada em tarefas de visão computacional. Permite treinar redes neurais extremamente profundas com um grande número de camadas sem sofrer com o problema dos gradientes de desaparecimento, que após uma dada quantidade de camadas o incremento de mais camadas não conseguia aumentar a qualidade do modelo (HE et al., 2016).

Essa arquitetura trabalha na proposta em que as camadas continuem a receber os valores resultantes da função de ativação *Rectified Linear Unit* (ReLU - uma função simples não linear em que se o valor de entrada é negativo retorna zero e caso seja positivo ou zero mantém o valor de entrada) -  $F(x)$ , da camada anterior, mas também recebam os valores de entrada  $x$  dessas funções. Como pode ser observado na Figura 12.

Figura 12 – Bloco residual.



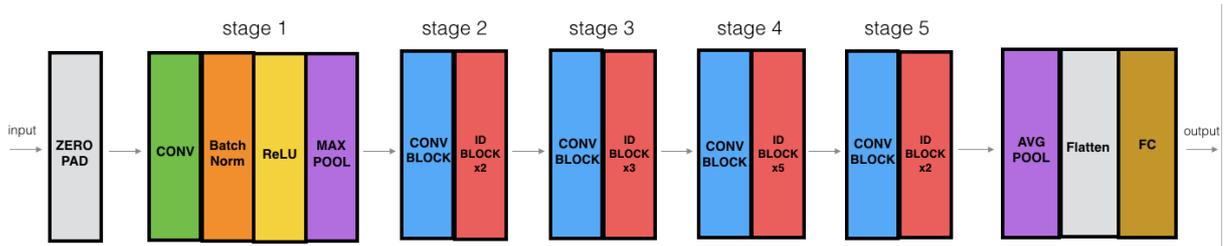
Fonte: Adaptado de (HE et al., 2016)

Essa arquitetura possui duas características principais:

1. “Conexões de atalho de identidade”: uma estratégia de “atalhos” ou “conexões de salto”, que pulam pares de grupos de camadas convolucionais. São também chamadas unidades bloqueadas ou unidades recorrentes fechadas, apesar de não apresentarem uma recorrência no sentido tradicional dos modelos de redes neurais recorrentes;
2. Grande foco em normalização de lotes, o que possibilita tornar diferentes entradas de um modelo de aprendizado de máquina mais semelhantes entre si, para que o modelo possa aprender e generalizar bem novos dados;

Os atalhos na ResNet buscam evitar que a rede, muito profunda, “morra” por esvanecimento de gradientes através do empilhamento de mapeamentos de identidades que, do ponto de vista matemático, segundo os autores (HE et al., 2016), estão simplesmente empilhando camadas que não fazem nada. Com isso, como mostra a Figura 12, a ResNet utiliza, num determinado ponto, um sinal que é a soma do sinal produzido pelas duas camadas convolucionais anteriores somado ao sinal transmitido diretamente do ponto anterior a estas camadas, juntando um sinal processado com um sinal de uma etapa anterior no processamento (SOUZA et al., 2020).

Figura 13 – Arquitetura da ResNet.



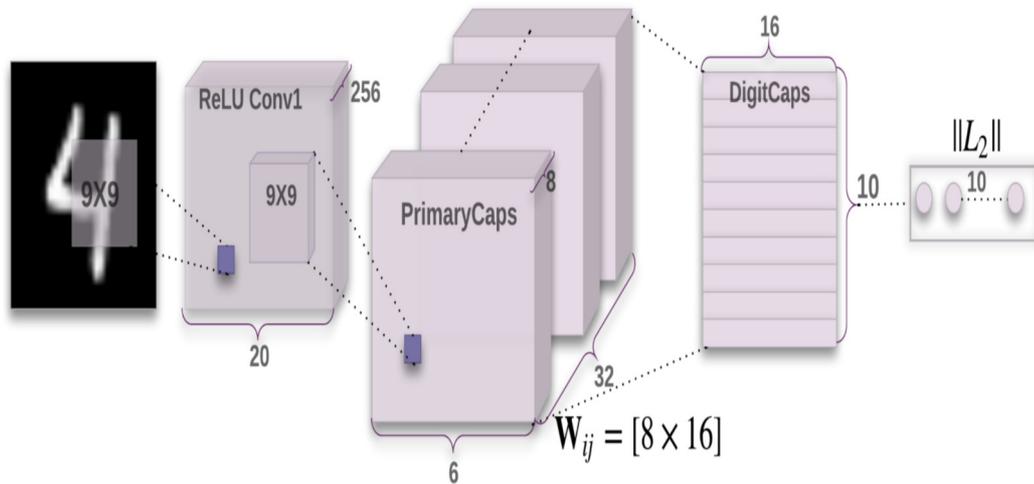
Fonte: Adaptado de (SOUZA et al., 2020)

A arquitetura básica de uma ResNet é descrita na Figura 13. É aplicada na entrada um bloco de preenchimento zero (adição de linhas e colunas com o valor zero em cada lado do filtro de convolução) de (3,3). No Estágio 1, a Convolução 2D tem 64 filtros da forma (7,7) e usa um stride de (2,2). A normalização de lotes é aplicado ao eixo dos canais da entrada. O max-pooling usa uma matriz (3,3) e uma stride (2,2). Nos estágios 2, 3, 4 e 5 o bloco convolucional usa três conjuntos de filtros. No estágio 2 usa 2 blocos de identidade que utilizam três conjuntos de filtros. No estágio 3 os 3 blocos de identidade usam três conjuntos de filtros. No estágio 4, os 5 blocos de identidade usam três conjuntos de filtros. Estágio 5, os 2 blocos de identidade usam três conjuntos de filtros. O *average-pooling* usa uma matriz de (2,2). Por fim, a camada totalmente conectada (densa) reduz sua entrada para o número de classes usando uma ativação Softmax (SOUZA et al., 2020).

### 2.6.3 Capsule Network - CapsNet

As redes neurais convolucionais (CNN) tem sido muito utilizadas nos últimos anos para detecção/diagnóstico de patologias baseados em imagens, entretanto tais métodos tem a característica de depender de uma grande quantidade de dados para a realização do treinamento. Outra característica da CNN é que elas realizam o roteamento entre camadas por meio de operações de *pooling* que descartam informações como localização e posicionamento dos objetos que podem ser valiosos para fins de classificação ou detecção. Para contornar tais problemas foi desenvolvida por Sabour, Frosst e Hinton (2017) uma rede neural profunda baseada no conceito de cápsulas.

A ideia por trás da CapsNet é que todo objeto é formado por porções menores localizados em determinada posição, com isso nosso cérebro combina essa informação hierárquica para fazer o reconhecimento, assim, não é preciso processar todos os detalhes.

Figura 14 – Fluxograma da arquitetura da *capsule network*.

Fonte: (SABOUR; FROSST; HINTON, 2017)

O responsável por fazer tal processamento são as chamadas cápsulas. Uma cápsula é um grupo de neurônios cujo vetor de atividade representa vários parâmetros de pose e o comprimento desse vetor mostra a probabilidade de que uma entidade específica exista.

Como em outras redes, a CapsNet é organizada em várias camadas. As cápsulas de nível mais baixo são chamadas cápsulas primárias. Elas são responsáveis por receber uma região pequena da imagem e tentar detectar a presença e pose de determinado objeto (GÉRON, 2018). Essa detecção é invariante quanto à mudança de posicionamento do objeto.

As camadas de nível mais elevado detectam objetos mais complexos levando em consideração a detecção das camadas inferiores. A Figura 14 apresenta o arranjo da rede proposta por Sabour, Frosst e Hinton (2017) para o reconhecimento de dígitos. Nessa configuração as cápsulas primárias responsáveis pelo processamento de regiões de  $9 \times 9$  *pixels* interagem com cápsulas de alto nível (“DigitCaps”) para a identificação de algarismos manuscritos.

O fato de a saída de uma cápsula ser um vetor torna possível usar um poderoso mecanismo de roteamento dinâmico para garantir que a saída da cápsula seja enviada para uma “mãe” apropriada na camada superior. Inicialmente, a saída é roteada para todos os possíveis pais, mas é reduzida pelos coeficientes de acoplamento que somam 1 (SANTOS, 2019).

Para cada possível “mãe”, a cápsula calcula um “vetor de previsão” multiplicando sua própria saída por uma matriz de ponderação. Se esse vetor de previsão tiver um produto escalar grande para a saída de uma possível “mãe”, haverá uma realimentação de cima para baixo que aumenta o coeficiente de acoplamento para essa “mãe” e o diminui para outras. Isso aumenta a contribuição que a cápsula faz para a cápsula superior, aumentando ainda mais o produto escalar da previsão da cápsula com a saída da mãe (SANTOS, 2019).

Esse tipo de “roteamento por acordo” é mais eficaz do que a forma muito simples de roteamento implementada pelo *pool* máximo (CNN), que permite que os neurônios em uma camada ignorem todos, exceto o detector mais ativo em um conjunto local na camada abaixo.

A relação entre a  $i$ -ésima cápsula em uma camada inferior e a  $j$ -ésima cápsula na próxima camada superior é codificada usando uma matriz de transformação linear  $W_{ij}$ . A informação é propagada como:  $\hat{u}_{j|i} = W_{ij}u_i$ . O vetor  $\hat{u}_{ij}$  representa a semelhança da  $i$ -ésima cápsula em uma camada inferior sobre a  $j$ -ésima cápsula na camada superior.

Tendo computado os vetores de previsão, as cápsulas de nível inferior encaminham suas informações para as cápsulas-mães que mais concordam com suas previsões. O mecanismo que garante que as saídas das cápsulas filhas sejam enviadas para as cápsulas “mães” apropriadas é denominado *Roteamento dinâmico*.  $c_{ij}$  denota o coeficiente de roteamento da  $i$ -ésima cápsula na camada inferior à  $j$ -ésima cápsula na camada superior, onde  $\sum_j c_{ij} = 1$  e  $c_{ij} \geq 0, \forall j$ . Quando  $c_{ij} = 1$ , todas as informações da  $i$ -cápsula serão enviadas para a  $j$ -cápsula, enquanto que para  $c_{ij} = 0$ , não há informações fluindo entre as duas cápsulas.

O método de roteamento dinâmico ajusta iterativamente os coeficientes  $c_{ij}$  e direciona as saídas das cápsulas filhas para a cápsula apropriada na próxima camada, determinando, assim, a pose dos objetos com mais precisão.

A CapsNet calcula os coeficientes de roteamento da seguinte forma (MOBINY; NGUYEN, 2018):

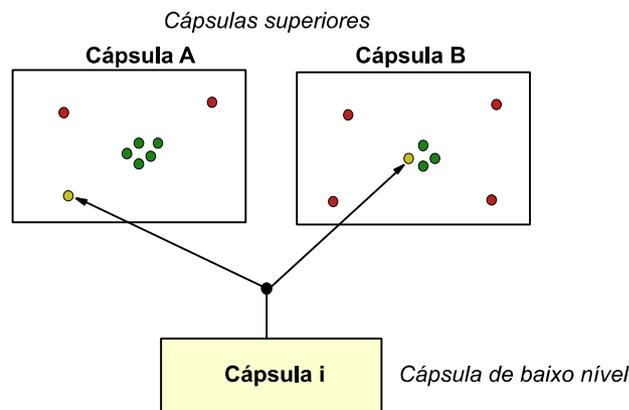
$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}, s_j = \sum_i c_{ij} \hat{u}_{j|i}, \quad (2.8)$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, b_{ij} \leftarrow b_{ij} + v_j \cdot \hat{u}_{j|i}. \quad (2.9)$$

A saída de cada cápsula pai  $v_j$  é computada como a soma ponderada de todas as previsões de cápsulas-filho, passando então por uma função *squash* (Equação 2.8) não-linear. A compactação garante que o vetor de saída tenha comprimento não maior que 1 (de modo que o comprimento pode ser interpretado como a probabilidade de que uma determinada característica seja detectada pela cápsula) sem alterar sua direção. As cápsulas pais recebem previsões de todas as cápsulas filhas. Estes vetores são representados por pontos na Figura 15. O mecanismo de roteamento dinâmico aumentará o coeficiente de roteamento para a cápsula  $j$  “mãe” por um fator de  $\hat{u}_{j|i} \cdot v_j$ . Assim, uma cápsula filha enviará mais informações para a cápsula “mãe” cuja saída  $v_j$  é mais semelhante à sua previsão  $\hat{u}_{j|i}$  (MOBINY; NGUYEN, 2018).

Para todos, exceto a primeira camada de cápsulas, a entrada total para uma cápsula  $s_j$  é uma soma ponderada sobre todos os “vetores de previsão”  $\hat{u}_{j|i}$  das cápsulas na camada abaixo e é produzida pela multiplicação da saída  $u_i$  de uma cápsula na camada abaixo por uma matriz de peso  $W_{ij}$  onde o  $c_{ij}$  são coeficientes de acoplamento determinados pelo processo de roteamento dinâmico iterativo.

Figura 15 – Roteamento dinâmico para envio de informações da cápsula de nível inferior para a cápsula de nível superior.



Fonte: (SANTOS, 2019).

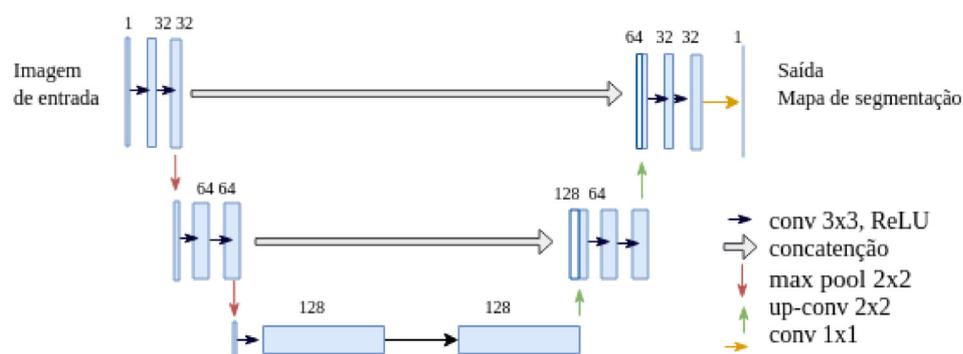
A Figura 15 mostra uma cápsula de nível inferior tomando a decisão de enviar a sua saída para as cápsulas “mãe”. Essa decisão é feita ajustando os coeficientes de roteamento,  $c_{ij} = 0$ , que serão multiplicados pelos vetores de previsão antes de enviá-los para cápsulas de alto nível. O algoritmo que realiza o processo de roteamento dinâmico é apresentado no artigo de Sabour, Frosst e Hinton (2017).

## 2.6.4 U-Net

A U-Net é uma rede neural artificial completamente convolucional proposta para segmentação de imagens biomédicas. A arquitetura da U-net consiste em um caminho de contração para capturar o contexto, e um caminho de ampliação simétrico que permite uma segmentação precisa. A principal estratégia que diferencia a U-net das outras arquiteturas de rede totalmente conectadas é a combinação entre os mapas de características do estágio de contração e seus correspondentes simétricos no estágio de expansão, permitindo a propagação das informações de contexto para os mapas de características de alta resolução. Esse modelo vem superando os métodos existentes em diversos desafios biomédicos de segmentação de imagens (FERREIRA et al., 2019).

A arquitetura consiste em um caminho de contração (descida do “U”) para capturar o contexto e um caminho de ampliação (subida do “U”) simétrico que permite uma localização precisa. A etapa de contração é composta por blocos convolucionais com duas camadas com filtros de dimensões  $3 \times 3$  e função de ativação Unidade Linear Retificada (ReLU). Cada bloco é sucedido de uma camada de subamostragem (*Max Pooling*) com janelamento  $2 \times 2$ . A partir do segundo bloco, são gerados o dobro de mapas de característica do bloco anterior. Após o último bloco de convolução, no nível mais profundo da rede, não haverá camada completamente conectada. Logo, os mapas gerados nesse nível serão a entrada da etapa de ampliação (SILVA et al., 2019).

Figura 16 – Arquitetura da rede U-Net.



Fonte: (NOH; HONG; HAN, 2015).

A fase de ampliação (subida do “U”) contém blocos convolucionais similares aos da fase de contração. Eles são precedidos da operação *upsampling*, a qual amplia o tamanho das saídas produzidas pelo bloco que o antecede. Nessa fase, concatena-se os mapas

de características dos blocos simétricos. O último bloco se diferencia dos demais pois é composto por três camadas convolucionais. As duas primeiras, contendo filtros de dimensão  $3 \times 3$  e função de ativação ReLU. A última, entretanto, possui filtros de tamanho  $1 \times 1$  e função de ativação sigmoide. Essa função atua como um classificador pixel-a-pixel, gerando valores entre 0 e 1 os quais indicam, respectivamente, fundo e região de interesse (SILVA et al., 2019).

Essa rede pode ser treinada de ponta a ponta a partir de poucas imagens, onde a U-net simplesmente concatena os mapas de características do codificador para mapear mapas de características do decodificador em todas as etapas para formar uma estrutura como escala. Essa arquitetura por suas conexões de concatenação permite que o decodificador em cada etapa aprenda as características relevantes que são perdidas quando agrupadas no codificador (FERREIRA et al., 2019).

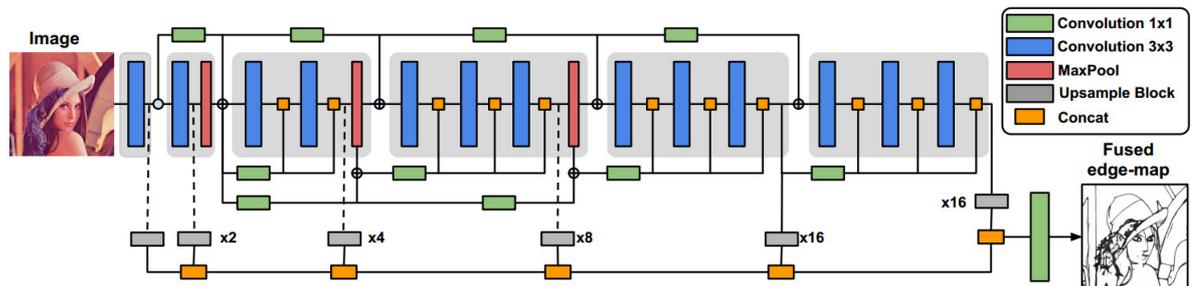
### 2.6.5 Dense Extreme Inception Network - DexiNed

A *DexiNed* é uma arquitetura de rede neural profunda, proposta para a detecção de bordas. Consiste em uma pilha de filtros que recebem como entrada uma imagem e preveem um mapa de bordas. A DexiNed pode ser vista como duas sub-redes: uma *Dense extreme inception network* (Dexi) e o bloco de *up-sampling* (UB). Enquanto a Dexi é alimentada com a imagem, o UB é alimentado com mapas de recursos de cada bloco do Dexi. A rede resultante (DexiNed) gera mapas de borda finos, evitando bordas perdidas nas camadas profundas (SORIA; RIBA; SAPPA, 2020).

A Figura 17 apresenta a estrutura da rede, que é composta por um codificador com 6 blocos principais. As saídas de rede apresentam mapas em cada um dos blocos principais para produzir mapas de borda intermediários usando um bloco de *upsampling*. Todos os mapas de borda resultantes dos blocos de amostragem são concatenados para alimentar a pilha de filtros aprendidos no final da rede e produza um mapa de borda fundido. Todos os seis blocos de *upsampling* não compartilham pesos.

Os blocos em azul consistem em uma pilha de duas camadas convolucionais com tamanho de núcleo  $3 \times 3$ , seguidas de *batch normalization* e ReLU como função de ativação (apenas as últimas convoluções nos últimos sub-blocos não possuem essa ativação). O *max pooling* é definido pelo kernel  $3 \times 3$  e pelo passo 2. Conforme a arquitetura segue o

Figura 17 – Arquitetura da rede DexiNed.



Fonte: (SORIA; RIBA; SAPPÀ, 2020).

aprendizado em várias escalas, adiciona-se um processo de *upsampling* (blocos horizontais em cinza, Figura 17).

Como muitas convoluções são realizadas, todos os blocos profundos perdem recursos importantes da aresta e apenas uma conexão principal não é suficiente, a partir da quarta camada convolucional, a perda dos recursos da aresta é mais acentuada. Portanto, desde o bloco 3, a saída de cada sub-bloco é o resultado da média com a conexão da borda (quadrados laranja na Figura 17). Após a operação de *max pooling* e antes da soma com a conexão principal, a conexão de borda é definida como a média de cada sub-bloco de saída (veja retângulos em verde, lado inferior). A saída de cada um dos blocos principais alimenta um bloco de sobre amostragem que produz um mapa de borda intermediário, e a fusão de cada mapa intermediário gera o resultado final (SORIA; RIBA; SAPPÀ, 2020).

## 2.7 Métricas de desempenho

Para alcançar o objetivo deste trabalho, que aborda os problemas de segmentação de camadas da retina e classificação de degeneração macular e edema macular, utiliza-se métricas de desempenho que avaliam os resultados obtidos, a validação da segmentação é feita a partir da avaliação de erro médio, distância de Hausdorff e coeficiente Dice, o desempenho da classificação para o diagnóstico automático utiliza a acurácia, sensibilidade, especificidade e precisão. Essas medidas são apresentadas a seguir.

### 2.7.1 Coeficiente Dice

O coeficiente de similaridade Dice é o mais utilizado para a validação de segmentação em imagens médicas (TAHA; HANBURY, 2015). Ele é definido conforme a Equação 2.10, onde  $X$  e  $Y$  representam as máscaras comparadas para cada imagem de entrada, no caso, a anotação manual e o mapa de segmentação gerado pelo processo de segmentação automática.

$$DICE = \frac{2 * |X \cap Y|}{|X| + |Y|} \quad (2.10)$$

### 2.7.2 Avaliação de Erro Médio

A comparação entre dois conjuntos de dados observados sobre o mesmo fenômeno pode ser feita a partir do cálculo do erro entre os pares correspondentes (avaliação de erro médio) ou a partir de uma métrica de distância entre os próprios conjuntos (SANTOS, 2019). A seguir falaremos de duas estratégias: a avaliação do erro médio (sumarização de erros entre pares correspondentes) e a medida de distância de Hausdorff (distância entre conjuntos).

Dadas duas variáveis contínuas  $A$  e  $B$ , o erro entre pares correspondentes é dado por  $e_i = a_i - b_i$ . A sumarização dos valores  $e_i (i = 1, 2, 3, \dots, n)$  permite estabelecer uma medida do erro médio entre as variáveis. Segundo Willmott e Matsuura (2005) essa medida é representada de forma genérica por:

$$\bar{e}_\gamma = \left[ \frac{\sum_{i=1}^n w_i |e_i|^\gamma}{\sum_{i=1}^n w_i} \right]^{1/\gamma} \quad (2.11)$$

onde  $\gamma \geq 1$  e  $w_i$  é um peso designado para cada  $|e_i|^\gamma$  referente à influência daquele elemento no conjunto.

A partir da representação genérica oferecida pela Equação 2.11, obtemos expressões comumente usadas, como o erro médio (EM) e o erro médio absoluto (EMA).

O EM é uma medida mais simples obtida quando  $\gamma = 1$  e  $w_i = 1, \forall i \in \{1, 2, \dots, n\}$  e os sinais dos erros são preservados. O cálculo de EM é dado por:

$$EM = n^{-1} \sum_{i=1}^n e_i = \bar{A} - \bar{B} \quad (2.12)$$

Já o EMA mantém os  $\gamma = 2$  e  $w_i = 1, \forall i \in \{1, 2, \dots, n\}$  e despreza os sinais dos erros, sendo representado por:

$$EMA = n^{-1} \sum_{i=1}^n |e_i| \quad (2.13)$$

A medida de EM é útil para revelar o viés médio dos erros entre os pares correspondentes. O cálculo de EMA permite uma interpretação intuitiva na mesma escala dos valores a serem comparados (WILLMOTT; MATSUURA, 2005).

### 2.7.3 Distância de Hausdorff

Diferentemente de outros métodos de comparação de similaridades entre conjuntos de dados, para o cálculo da medida de distância de Hausdorff não se estabelece qualquer pareamento de pontos de  $A$  com pontos de  $B$ .

A medida de distância de Hausdorff (CONCI et al., 2015; CONCI; KUBRUSLY, 2017) quantifica o quanto dois conjuntos de pontos estão distantes um do outro podendo ser utilizada para expressar a semelhança entre dois objetos sobrepostos, sendo definida por  $h: \wp(X) \setminus \emptyset \times \wp(X) \setminus \emptyset \rightarrow \overline{\mathbb{R}}$  é definida para  $A, B \in \wp(X)$ :

$$h(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\} \quad (2.14)$$

$$= \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}. \quad (2.15)$$

onde,  $d$  é uma medida de distância entre dois pontos (*e.g.* a medida de distância euclidiana).

A função  $h(A, B)$  determina a distância máxima entre os dois conjuntos de pontos através da medida da distância de um ponto de  $A$  que está mais afastado de qualquer ponto de  $B$ . Se  $h(A, B) = d$ , então cada ponto  $a \in A$  deve estar no máximo a uma distância  $d$  de qualquer ponto  $b \in B$  e *vice-versa*.

### 2.7.4 Validação do Classificador

A avaliação dos resultados fornecidos pelo classificador para o método proposto consiste em validar a discriminação entre os casos nos quais as imagens da retina estão saudáveis ou possuem alguma patologia. As medidas mencionadas são as mais utilizadas na área de processamento de imagens médicas.

- Sensibilidade é a proporção de verdadeiros positivos: a capacidade do sistema em prever corretamente a condição para casos que realmente a tem. O cálculo da sensibilidade é feito de acordo com a Equação 2.16.

$$SEN = \frac{VP}{VP + FN} \quad (2.16)$$

- Especificidade é a proporção de verdadeiros negativos: a capacidade do sistema em prever corretamente a ausência da condição para casos que realmente não a tem. O cálculo da especificidade é feito com na Equação 2.17.

$$ESP = \frac{VN}{VN + FP} \quad (2.17)$$

- Acurácia é a proporção de predições sem levar em consideração o que é positivo e o que é negativo. O cálculo da acurácia é feito com a Equação 2.18.

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.18)$$

- A precisão mede a capacidade de um classificador de identificar apenas as instâncias corretas para cada classe. O cálculo da precisão é feito com a Equação 2.19

$$P = \frac{VP}{VP + FP} \quad (2.19)$$

## 3 Trabalhos Relacionados

A análise automática da retina humana em exames de OCT tem sido alvo de inúmeros estudos e desenvolvimento de métodos nos últimos anos. Os métodos apresentados se aplicam principalmente aos problemas da segmentação da retina e diagnóstico de doenças oculares.

O primeiro problema enfrentado no processo de análise das imagens de OCT foi a segmentação da retina (HUANG et al., 1991b). A retina é uma estrutura multicamadas, e a identificação e separação dessas camadas é importante para uma análise anatômica detalhada. A segmentação também é importante para o diagnóstico automatizado de DMRI e EMD. Para isso, pode ser feita simplesmente uma separação da retina do fundo da imagem (para métodos que analisam globalmente a *B-scan*) ou o enfrentamento mais complexo da segmentação do epitélio pigmentar retiniano. Essa é a camada que sofre as maiores alterações provocadas pelas doenças, sendo importante para o diagnóstico automatizado.

Para diagnosticar a DMRI e a EMD de forma automática, tem sido utilizados majoritariamente métodos baseados em redes neurais de aprendizado profundo. A seguir apresentaremos os principais trabalhos desenvolvidos relacionados à esta tese que visa desenvolver e avaliar métodos para segmentação da retina e diagnóstico automático de DMRI e EMD em imagens de OCT.

### 3.1 Segmentação das Camadas da Retina em Imagens de OCT

A segmentação das camadas da retina é uma tarefa complicada, por se tratar de uma imagem complexa e com grande quantidade de ruídos inerentes ao processo de aquisição. Uma etapa padrão para a segmentação em imagens de OCT é o pré-processamento, objetivando a redução do ruído e realce das bordas para facilitar as tarefas de segmentação. Nos últimos anos vários trabalhos têm sido publicados demonstrando metodologias para a segmentação das camadas da retina. Alguns desses artigos são listados a seguir.

Em (CHIU et al., 2012), é apresentado um método para segmentação da retina total, que fica entre a membrana limitante interna (MLI) e o epitélio pigmentar da retina (EPR), e o epitélio pigmentar da retina e complexo de drusas (EPRCD), que fica entre o

EPR e a membrana de Bruch. Para validar o método, são usados 220 *B-scans* extraídos de 20 volumes. Todos os pacientes apresentavam degeneração macular relacionada à idade (DMRI) em estágio intermediário e com drusas, e alguns com atrofia geográfica. O método é baseado no algoritmo *Graph Cut*. Como etapa de pré-processamento, a redução de ruído é realizada por meio de um filtro médio retangular. A partir da imagem resultante, um grafo é construído com os valores dos *pixels* como pesos. A busca por cada borda é realizada sequencialmente. Considerando todas as camadas, o algoritmo apresentou uma diferença média de 0,95 *pixels*, valor ainda menor que a diferença de segmentação entre dois especialistas. Junto com o desenvolvido por (FARSIU et al., 2008), esses trabalhos referem-se ao desenvolvimento do software DOCTRAP (Duke OCT Retinal Analysis Program) projetado para uso em pesquisas de segmentação OCT.

Em (FANG et al., 2017), é apresentado um framework que combina redes neurais convolucionais com algoritmos de busca em grafos para segmentar nove bordas retiniais em imagens de OCT. O método foi validado com 60 volumes (2915 *B-scans*) de vinte olhos de pessoas com DMRI não exsudativa. A CNN é treinada com as características das bordas das camadas para estimar suas posições. Em seguida, esses valores são submetidos a um algoritmo de busca em grafo para a definição final dos limites. Os resultados encontrados foram comparados aos resultados obtidos pelos softwares de segmentação DOCTRAP e OCTExplorer. Esses resultados foram superiores aos obtidos pelo software OCTExplorer, mas ainda inferiores aos obtidos pelo software DOCTRAP.

O trabalho (GOPINATH; RANGREJ; SIVASWAMY, 2017) apresenta um método para segmentar as bordas das camadas da retina em imagens de OCT. Três bancos de dados públicos (CHIU et al., 2010; CHIU et al., 2012; TIAN et al., 2015) foram usados, dois com um total de 210 exames *B-Scans* de pacientes sem patologia e o outro com 220 exames *B-scans* de pacientes com DMRI intermediária. Dois modelos baseados na combinação de rede neural convolucional e memória de longo prazo (LSTM) são desenvolvidos, um para segmentar oito camadas de retina e outro para segmentar três camadas. A CNN é usada para extrair uma região de interesse e detectar as bordas das camadas, enquanto LSTM é utilizado para rastrear o limite da camada. Este modelo é treinado com uma mistura de casos normais e DMRI. Os resultados mostram um erro absoluto médio (EMA) em *pixels* de  $1,30 \pm 0,48\mu m$ , menor que o erro de marcação das bases de  $1,79 \pm 0,76\mu m$ .

Em (CHAKRAVARTY; SIVASWAMY, 2017) é proposto um novo modelo de campo aleatório condicional linearmente parametrizado (LP-CRF) cuja energia é aprendida de

um conjunto de imagens de treino. O LP-CRF proposto compreende dois bancos de filtros de convolução para capturar a aparência de cada região e limite das camadas, os pesos relativos da forma anterior e um termo adicional com base na similaridade de aparência dos pontos de fronteira adjacentes. Todos os tipos de energia são aprendidos em conjunto com uma Máquina de vetores de suporte estruturado. O método proposto segmenta todos os limites da retina em uma única etapa. O método foi avaliado em 107 imagens da classe Normal e 220 DMRI. O erro médio absoluto alcançado foi de  $1,52 \pm 0,29 \mu m \text{ pixels}$  para a segmentação de oito bordas no conjunto de dados Normal e  $1,9 \pm 0,65 \mu m \text{ pixels}$  para três bordas no conjunto de dados DMRI e Normal.

Um método para segmentar as bordas das camadas da retina em imagens de OCT foi proposto em (KUGELMAN et al., 2018). Foram utilizados dois bancos de dados de imagens, um para pacientes pediátricos sem histórico de patologias e pacientes e patologias intermediárias de DMRI. O método usou uma rede neural recorrente (RNN) como um classificador baseado em fragmentos para segmentar sete limites de camadas na primeira base e três bordas na segunda base. Os resultados indicaram que a arquitetura RNN é uma alternativa viável à CNN para tarefas de classificação de imagens que apresentam uma estrutura sequencial clara. Em comparação com a CNN, a RNN apresentou um erro médio absoluto ligeiramente menor.

Um método para segmentar as bordas das camadas MLI, EPR e MB é apresentado em (SANTOS, 2019). O método foi testado em duas bases de dados públicas com pacientes normais e com DMRI intermediária, uma fornecida por (FARSIU et al., 2014) com 384 volumes, a outra conta com 19 volumes disponibilizados por (CHIU et al., 2012). Uma arquitetura de rede neural profunda, *capsule network*, foi utilizada para fazer a segmentação inicial das bordas de três camadas da retina, onde classifica fragmentos extraídos de imagens de OCT nas quatro classes (MLI, EPR, MB e fundo). Para refinar a segmentação, o autor fez uso da técnica de corte de grafo, que é aplicada aos mapas de probabilidade da rede neural. Como resultado, o método atingiu, respectivamente, para as duas bases, 0,75 e 2,04 de média do erro absoluto médio para o MLI, 0,93 e 1,97 para o EPR e 1,09 e 2,05 para o MB.

Em (CHEN et al., 2019), um novo método é apresentado para segmentar MLI, EPR e MB em imagens de OCT. Três bases de imagens foram utilizadas para treinar e validar o método, com imagens de pacientes saudáveis e DMRI intermediária. Os autores usam um método de ensemble chamado *deepForest*, que usa várias *random forests* para

gerar a classificação de *patches* extraídas de imagens de OCT. Como pré-processamento, a normalização das intensidades de *pixel* é aplicada. Os *patches* são divididos em quatro classes, membrana limitante interna, EPRCD, membrana de Bruch e fundo. Após a classificação, uma técnica baseada em grafos desenvolvida pelo autor é aplicada para refinar a segmentação. Como resultado, o método obteve erros médios de 0,81, 1,35 e 1,23 *pixels* para as três bases, respectivamente.

O artigo (LOU et al., 2020) apresenta um método para segmentação de camadas retiniais (MLI, EPR e MB) em imagens OCT baseado no algoritmo de onda, um modelo matemático da equação da energia potencial do fluido em mecânica dos fluidos. O método usa a base fornecida por (FARSIU et al., 2014) e usa o erro médio absoluto para avaliar os resultados. O método obteve EMA inferior a 1,5 *pixels* em todas as avaliações.

Em (HU et al., 2021) é apresentado um método para segmentação das camadas MLI, EPR e MB em imagens de OCT utilizando uma nova arquitetura de rede neural recorrente chamada ERR-Net que utiliza uma rede residual como *backbone* e blocos recorrentes em suas etapas. Como a saída da rede neural pode deixar uns espaços sem preenchimento na segmentação, o autor utiliza uma busca em grafo para refinar o resultado final. Foram utilizadas 3 bases diferentes para testar o método (TIAN et al., 2016; CHIU et al., 2012; FARSIU et al., 2014), onde alcançou erro médio absoluto menor que 1,2 *pixels* para as 3 bordas na base de Farsiu et al. (2014).

Nos trabalhos supracitados, podemos destacar que todos usam alguma técnica de busca em grafo para gerar a segmentação final das bordas das camadas da retina. O método apresentado neste trabalho utiliza uma rede neural profunda especializada em detecção de bordas para fazer a etapa final de segmentação.

## 3.2 Detecção de Degeneração Macular e Edema Macular Diabético em Imagens de OCT

A eficiência do uso de processamento de imagens e reconhecimento de padrões para auxiliar no diagnóstico de diversas doenças tem impulsionado o desenvolvimento de métodos computacionais para o diagnóstico de doenças oculares.

Dentre as formas de se analisar imagens médicas digitais têm se destacado o uso de técnicas de reconhecimento de padrões tradicionais e técnicas de aprendizagem profunda. Na sequência, serão apresentados alguns trabalhos que usam tais formas de reconhecimento.

Nos últimos anos um número significativo de métodos para análise automática de retina a partir de exames de OCT tem sido desenvolvidos. As técnicas apresentadas se aplicam principalmente ao problema de diagnóstico de doenças oculares, entre elas a degeneração macular relacionada a idade e o edema macular diabético. Dentre as técnicas utilizadas para classificação de imagens *B-scans* de OCT, tem se destacado as redes neurais profundas. Em (KERMANY et al., 2018) é desenvolvido um método para diagnóstico da degeneração macular e edema macular diabético utilizando imagens *B-scans* de OCT e uma rede neural convolucional com uma arquitetura Inception-V3 para predição. Foi utilizada uma base de 207.130 imagens de OCT, divididas em quatro classes, neovascularização coroidal, edema macular diabético, drusa e normal.

Em (KAYMAK; SERENER, 2018) é utilizada uma rede neural convolucional (CNN do inglês *Convolutional Neural network*) com arquitetura AlexNet padrão para fazer o diagnóstico do edema macular diabético e da degeneração macular relacionada a idade,

Li et al. (2019) usam uma CNN com arquitetura VGG-16 e transferência de aprendizagem para diagnosticar a degeneração macular relacionada a idade e o edema macular diabético. As imagens da base foram redimensionadas para  $224 \times 224$ , e a base foi reduzida a partir de uma nova avaliação das imagens por outros especialistas.

No trabalho de (HUANG et al., 2019) é apresentada uma nova arquitetura de CNN chamada ‘rede neural convolucional guiada por camada’, onde a mesma é utilizada para segmentar camadas da retina e classificar as imagens. Neste trabalho são utilizadas duas bases de imagens para validar o método proposto, a base de (KERMANY et al., 2018) e uma base privada HUCM (*Hospital of Hunan University of Chinese Medicine*) que conta com 8904 imagens. As imagens são redimensionadas para  $224 \times 224$ .

Kamran et al. (2019) desenvolveram uma nova arquitetura CNN, denominada Optic-Net, especializada na identificação de doenças da retina. Nesta arquitetura é proposta uma nova unidade residual que inclui a Convolução Atrosa Separável, um novo bloco de construção e um mecanismo para impedir a degradação do gradiente. A rede é treinada a partir do zero em dois conjuntos de dados publicamente disponíveis: conjuntos de dados OCT2017 (KERMANY et al., 2018) e Srinivasan2014 (SRINIVASAN et al., 2014). A

abordagem atingiu uma acurácia de 99,8% e 100% para dois conjuntos de dados disponíveis separadamente, respectivamente.

Em (RASTOGI; PADHY; SA, 2019) é feita a detecção de distúrbios da retina em imagens de OCT da retina. Para isso foi utilizada uma Rede Neural de Convolução Densamente Conectada (DenseNet). O método foi avaliado utilizando a base OCT2017 (KERMANY et al., 2018) e alcançou resultados de acurácia de 96,5%, sensibilidade de 95,57% e especificidade de 99,15%.

Em (ALQUDAH, 2020) é apresentada uma metodologia para diagnóstico de drusas, neovascularização coroidal, DMRI, EMD e normal em imagens de OCT. Para tal é utilizado um filtro *double density wavelet transform-based adaptive thresholding* para remoção de ruídos e o autor propõe uma arquitetura CNN composta por quatro camadas convolucionais, *batch normalization*, *ReLU* e *max pooling*. O método é treinado e testado utilizando duas bases de imagens, Zanglab (KERMANY et al., 2018) e da Universidade de Duke (FARSIU et al., 2014), alcançando como melhor resultado 97% de acurácia na classificação multiclasse.

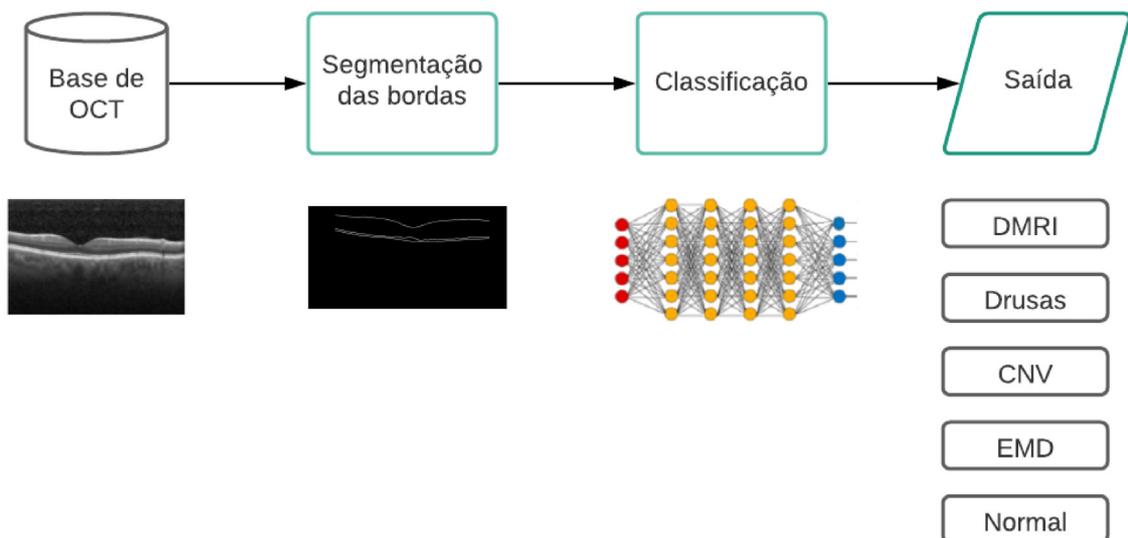
Os trabalhos relacionados acima indicam que pesquisas sobre o diagnósticos de doenças oculares como a DMRI e EMD são promissoras. Podemos verificar que, à exceção de (KERMANY et al., 2018), a maioria dos trabalhos utiliza bases de imagens muito pequenas, podendo assim haver uma baixa heterogeneidade nos casos presentes na base.

Os métodos apresentados na tese destacam-se em relação aos citados por utilizarem, no primeiro, o resultado da segmentação das camadas, e no segundo, uma nova arquitetura, que além de manter as características de rápida convergência e a necessidade de um número menor de amostras para treinamento, adiciona a utilização de uma camada convolucional que extrai características de textura.

## 4 Método para Classificação de Imagens de OCT Utilizando a Segmentação das Camadas da Retina

Este capítulo apresenta um método automático para classificação de imagens de OCT que utiliza o resultado da segmentação das bordas das camadas membrana limitante inferior, epitélio pigmentar retiniano e membrana de Bruch. A degeneração macular relacionada à idade e o edema macular diabético são doenças que podem apresentar lesões na retina e assim causando deformações nas camadas da retina. O processo de segmentação de segmentação busca expor tais deformações e auxiliar na etapa de classificação. A Figura 18 apresenta o fluxo do método, que é composto por duas etapas: segmentação das bordas de imagens de OCT e classificação dessas imagens utilizando as imagens binárias com o resultado da segmentação.

Figura 18 – Fluxograma do método.



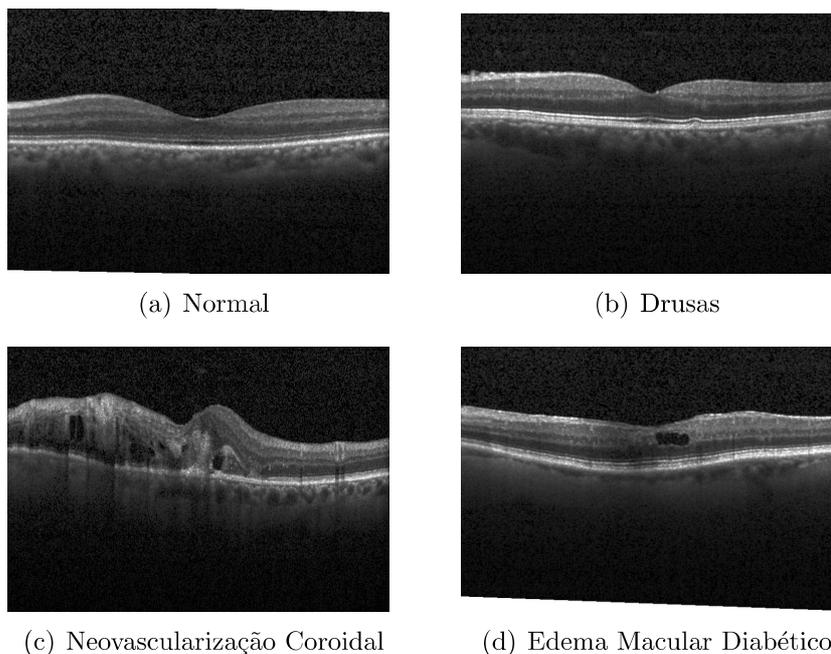
## 4.1 Bases de Imagens

Duas bases de exames de SD-OCT foram utilizadas para a avaliação dos métodos propostos. A primeira base é dividida em dois grupos (DMRI e Controle), já a segunda contém *B-scans* com quatro classificações (Normal, Drusas, Neovascularização Coroidal e Edema Macular Diabético). Somente a primeira possui marcação das bordas da RT e do EPR. As bases foram utilizadas para avaliar os dois métodos. A seguir, serão apresentados mais detalhes sobre esses dois conjuntos de imagens.

### 4.1.1 Base ZhangLab

Conjunto de dados de OCTs criado pelo ZhangLab na University of California at San Diego (UCSD), que está disponível gratuitamente online em (<https://data.mendeley.com/datasets/rscbjbr9sj/3>) (KERMANY et al., 2018). O conjunto de dados consiste em 108.312 imagens OCT de retina (37.206 com neovascularização coroidal, 11.349 com edema macular diabético, 8.617 com drusas e 51.140 normais). As imagens possuem diferentes resoluções, portanto foi aplicado um redimensionamento das mesmas para resolução de  $128 \times 128$  *pixels*.

Figura 19 – Exemplos de imagens da base.



### 4.1.2 Base Farsiu

A base de imagens de OCT utilizada por Farsiu et al. (2014) é composta por 269 volumes de olhos com DMRI do tipo seca e 115 volumes de controle (olhos considerados normais). Trata-se de um subconjunto da AREDS2 (CHEW et al., 2012) que foi coletado com o propósito de promover um estudo sobre indicadores quantitativos da presença de DMRI em adultos.

As principais características da população do estudo AREDS2 são:

- idade média de 74 anos;
- 57% são mulheres;
- 97% brancos;
- 7% fumantes;
- 19% com patologia cardiovascular;
- 44% utilizam medicamentos para regulação dos níveis de colesterol;
- 50% utilizam aspirina;
- 59% são acometidos de drusas bilaterais grandes;
- 32% como DMRI em estágio avançado em pelo menos um olho.

As imagens foram adquiridas em quatro clínicas por tomógrafos SD-OCT fabricados pela empresa Bioptigen, Inc (*Research Triangle Park, NC*). Os volumes foram adquiridos em uma região retangular de  $6,7\text{mm} \times 6,7\text{mm}$  centrada na fóvea com protocolo de *scan* rápido, resultando em volumes de tamanho  $1.000 \times 512 \times 100$ .

O subconjunto utilizado possui imagens de pacientes entre 50 e 85 anos exibindo DMRI intermediária com drusas maiores que  $125 \mu\text{m}$  nos dois olhos ou em apenas um, quando o outro olho está acometido por DMRI em estágio avançado. Os pacientes arrolados não possuem histórico de cirurgia vitreoretiniana ou outra condição patológica que afete a acuidade visual dos olhos. O grupo de controle utiliza os mesmos critérios de inclusão do estudo AREDS2 com a exceção dos que não apresentam evidência da presença de drusas.

Os exames SD-OCT fornecidos também disponibilizam a marcação das bordas que limitam a retina total (entre a membrana limitante interna e a membrana de Bruch) assim

como delimitam o epitélio pigmentado retiniano associado com o complexo de drusas (quando existente). A segmentação foi inicialmente executada de forma automatizada pelo software DOCTRAP, desenvolvido pela Universidade de Duke, e submetida à revisão e ajustes por especialistas certificados pelo *Duke Advanced Research in Spectral Domain OCT Imaging Laboratory*.

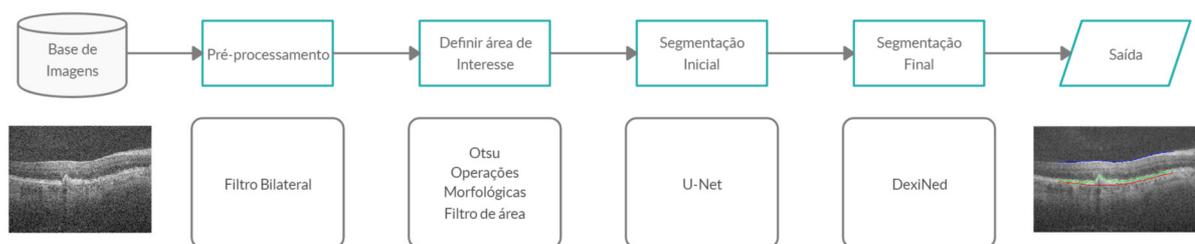
Após a segmentação, os especialistas marcaram manualmente o ponto central da fóvea e definiram como região de interesse em cada volume, um cilindro centrado neste ponto com 5mm de diâmetro. Esta revisão criteriosa e correção dos possíveis erros cometidos pelo método automatizado permitem a criação de uma representação da retina em mapas topográficos com precisão adequada para a validação desta metodologia.

## 4.2 Segmentação das Bordas das Camadas ILM, EPR e MB

Esta Seção apresenta o método que faz a segmentação das bordas que delimitam a retina total e o epitélio pigmentar retiniano. Para isso, são identificadas as bordas internas da MLI e do EPR e borda externa da MB. Com a obtenção das bordas do EPR e da MB é possível delimitar o EPR-CD (FARSIU et al., 2014). A delimitação dessa camada pode ser utilizada para avaliação de variações locais de espessura e alterações morfológicas causadas pela DMRI e EMD.

O método proposto apresenta quatro etapas, a primeira realiza um pré-processamento para eliminação de ruídos, a segunda define a região de interesse, a terceira faz uma segmentação inicial para diminuir mais a área de busca pelas bordas das camadas, e a quarta e última etapa finaliza o processo realizando a segmentação das 3 bordas. A Figura 20 apresenta um fluxograma dessas etapas e nas Seções seguintes são apresentados os detalhes sobre o uso das técnicas que compõem o método.

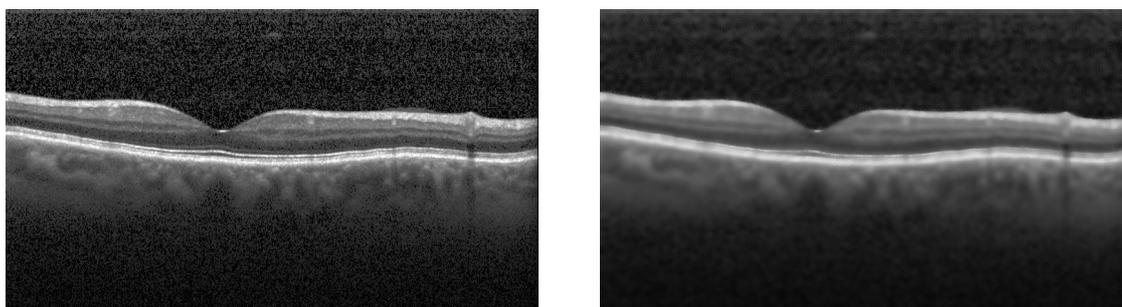
Figura 20 – Fluxograma do método de segmentação.



### 4.2.1 Pré-processamento

Para atenuar os ruídos presentes das imagens do exame de OCT, utiliza-se o filtro bilateral. Outra característica que fez tal filtro ser utilizado, é a capacidade de destacar bordas. Como foi explicado na Seção 2.5, o filtro bilateral depende de alguns parâmetros ( $d = 15$ ,  $\sigma_c = 100$ ,  $\sigma_s = 100$ ), que foram escolhidos com base em testes realizados. A Figura 21 apresenta um exemplo do resultado da aplicação desse filtro.

Figura 21 – Etapa de Pré-processamento.



(a) Sem Pré-processamento.

(b) Filtro Bilateral.

### 4.2.2 Definição da Região de Interesse

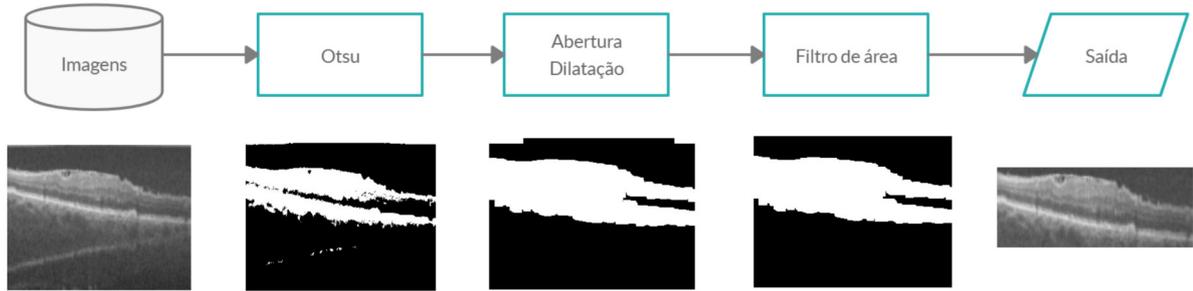
Esta etapa tem o objetivo de delimitar na imagem de cada fatia do volume em qual região será aplicado o processamento para identificação das bordas. Isso evita processamento desnecessário de regiões muito distantes da retina que representam o “fundo” da imagem.

O processo para a delimitação da região de interesse é feito em três etapas. A primeira consiste na aplicação do método de segmentação de Otsu (OTSU, 1979) para obter uma área inicial. A segunda etapa é composta pela aplicação de operações morfológicas de abertura (com elemento  $5 \times 5$  e duas iterações) e dilatação (com elemento  $5 \times 5$  e sete iterações). Por fim, para chegar à região final foi aplicado um filtro de eliminação por área, onde elimina-se objetos muito pequenos ( $1000 \text{ pixels}$ ). A Figura 22 mostra o resultado final do processo.

### 4.2.3 Segmentação Inicial

Uma arquitetura de rede neural profunda totalmente conectada baseada na U-Net proposta por Ronneberger, Fischer e Brox (2015) foi utilizada para gerar uma segmentação inicial das bordas das camadas da retina. A segmentação das camadas é feita em imagens

Figura 22 – Processo para a área de interesse.



2D do exame de OCT. Como entrada da rede são passados os exames e suas respectivas marcações das camadas.

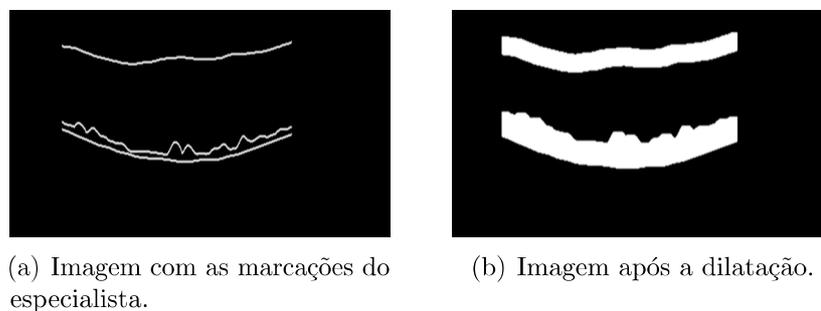
Neste trabalho utiliza-se a rede U-Net com a arquitetura proposta por FERREIRA et al. (2019), a qual mostrou-se superior à arquitetura original para o problema. Nesta arquitetura a rede possui apenas quatro camadas de subamostragem com função de ativação do tipo média, onde a média de intensidade dos *pixels* da janela é mantida. Para a etapa de ampliação é utilizada a convolução transposta. As camadas de convolução são seguidas por uma camada de normalização de *batch* e uma camada de *dropout*, com uma taxa de 20% para desativação de neurônios. Por fim, a segmentação é gerada por meio de uma convolução ( $1 \times 1$ ), com função de ativação *softmax*. A rede é treinada com a função de perda *Dice Loss(S)*, como mostra a Equação 4.1.

$$S = 1 - \left( \frac{2VP}{2VP + FP + FN} \right) \quad (4.1)$$

O treinamento da U-Net é realizado com imagens *B-scans* extraídas da base de imagens citada na Seção 4.1.2. As imagens possuem as marcações das bordas da MLI, do EPR e da MB. Operações morfológicas de dilatação são aplicadas às marcações (Figura 23), portanto, há duas áreas maiores para detecção. Essas operações foram aplicadas porque as marcações são muito finas e, no processo de *downsampling*, a rede tende a perder essas informações. Portanto, o processo de dilatação tenta aumentar o tamanho das marcações das bordas, fazendo com que ela se mantenha melhor durante as fases de *downsampling*. Os rótulos inseridos na rede são as bordas (ILM, RPE e BM) e o fundo. Para a entrada da rede, redimensionamos as imagens para uma resolução de  $128 \times 128$ . Para chegar ao tamanho de imagem usado, foram testados previamente com vários tamanhos ( $64 \times 64$ ,  $128 \times 128$ ,  $256 \times 256$ ,  $512 \times 512$ ). Verificou-se que o tamanho utilizado não

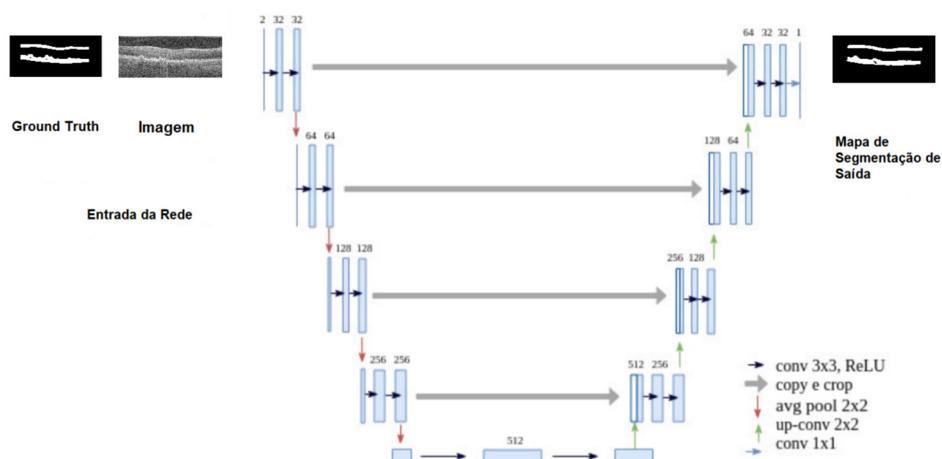
impacta significativamente nos resultados. Isso ocorre porque esta etapa visa reduzir a área de busca para a próxima, realizando uma segmentação aproximada. O que se busca é que a área resultante contenha as camadas que se pretende segmentar na etapa final. O redimensionamento é feito utilizando o menor tamanho de imagem capaz de gerar uma segmentação aproximada útil para a próxima etapa, que não gera nenhuma perda de performance no processo de segmentação. O aspecto do esforço computacional também influenciou no menor tamanho da imagem. Com o tamanho reduzido, foi possível treinar a rede com mais imagens e, assim, ter uma maior diversidade de casos no conjunto.

Figura 23 – Imagem de entrada para a U-Net.



O processo de treinamento é apresentado na Figura 24 em forma de fluxo com indicação da ordem das etapas. Após o processamento das imagens, estas são submetidas à U-Net para o treinamento, ou seja, o ajuste dos pesos de forma que a rede possa segmentar corretamente os *pixels* correspondentes às bordas.

Figura 24 – Arquitetura da U-Net.

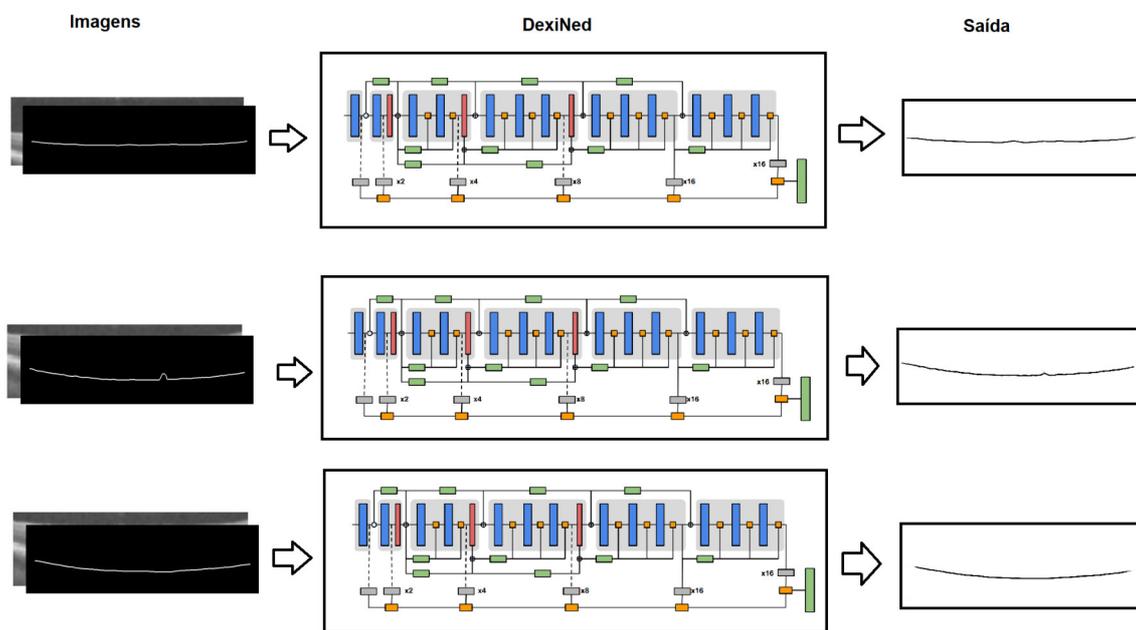


### 4.2.4 Segmentação Final

A rede neural DexiNed foi utilizada neste trabalho para gerar a segmentação final das bordas das camadas que são o objetivo do método. A arquitetura utilizada foi a mesma proposta em (SORIA; RIBA; SAPPA, 2020). Como entrada para a rede utiliza as imagens delimitadas pelo resultado da U-Net, assim diminuí-se o espaço de busca e processamento. Tanto no treino quanto no teste as imagens de entrada foram redimensionadas para  $992 \times 128$ , esta resolução leva em consideração que a rede trabalha com tamanhos de altura e largura múltiplos de 16. Foram treinados três modelos, um para cada borda que se pretende segmentar (MLI, EPR e MB), como ilustra a Figura 25.

Cada modelo foi treinado com 1000 imagens, utilizando os mesmos parâmetros da arquitetura padrão proposta por Soria, Riba e Sappa (2020). Uma função de perda especial é utilizada nesta rede, ela foi proposta por Xie e Tu (2015) e tem por objetivo se ajustar a rede a uma característica particular da segmentação de bordas, onde de 85 a 90% da imagem não são bordas.

Figura 25 – Arquitetura da DexiNed.



Fonte: adaptado de (SORIA; RIBA; SAPPA, 2020).

### 4.2.5 Validação

Após a segmentação das bordas das camadas da retina, é necessário validar os resultados. Este trabalho utilizou métricas comumente empregadas em métodos que visam tal objetivo. Essas são, o Dice, o Erro médio Absoluto (EMA) e a Distância de Hausdorf (DH), descritas na Seção 2.7. Tais métricas tem o objetivo de medir o desempenho do método proposto como satisfatório ou não, além de auxiliar na identificação de melhorias futuras.

## 4.3 Classificação

A partir do resultado da segmentação, a imagem binária é aplicada a uma rede neural para classificação, para tal foi utilizada uma Resnet50. Como explicado na Seção 2.6.2, a rede residual é uma rede neural convolucional construída por blocos residuais. Os blocos residuais da ResNet tem como valor de entrada um dado  $X$  que passa por uma série de operações em camadas de convolução-relu-convolução, e a saída  $f(X)$  é mapeada novamente para a entrada original  $X$ . A Tabela 4.3 apresenta a configuração da rede.

Tabela 1 – Arquitetura da rede residual.

| Camada | Configuração  | quantidade |
|--------|---|------------|
| Conv1  | $7 \times 7$ , 64, stride 2                                     | 1          |
| Conv2  | $1 \times 1$ , 64<br>$3 \times 3$ , 64<br>$1 \times 1$ , 264    | x 3        |
| Conv3  | $1 \times 1$ , 128<br>$3 \times 3$ , 128<br>$1 \times 1$ , 512  | x 4        |
| Conv4  | $1 \times 1$ , 256<br>$3 \times 3$ , 256<br>$1 \times 1$ , 1024 | x 6        |
| Conv5  | $1 \times 1$ , 512<br>$3 \times 3$ , 512<br>$1 \times 1$ , 2048 | x 3        |
|        | Average pool  | 1          |
|        | Flatten   | 1          |
|        | Dense, Softmax  | 1          |

### 4.3.1 Treinamento

O treinamento é realizado a partir de um conjunto de *B-scans* de OCT reservados para este fim. Foram realizados dois experimentos para verificar a eficácia das técnicas propostas.

No experimento 1 foi utilizada a divisão original da base fornecida em (KERMANY et al., 2018), onde foram separadas 250 imagens de cada classe para teste, e o restante da base foi utilizada para treino e validação.

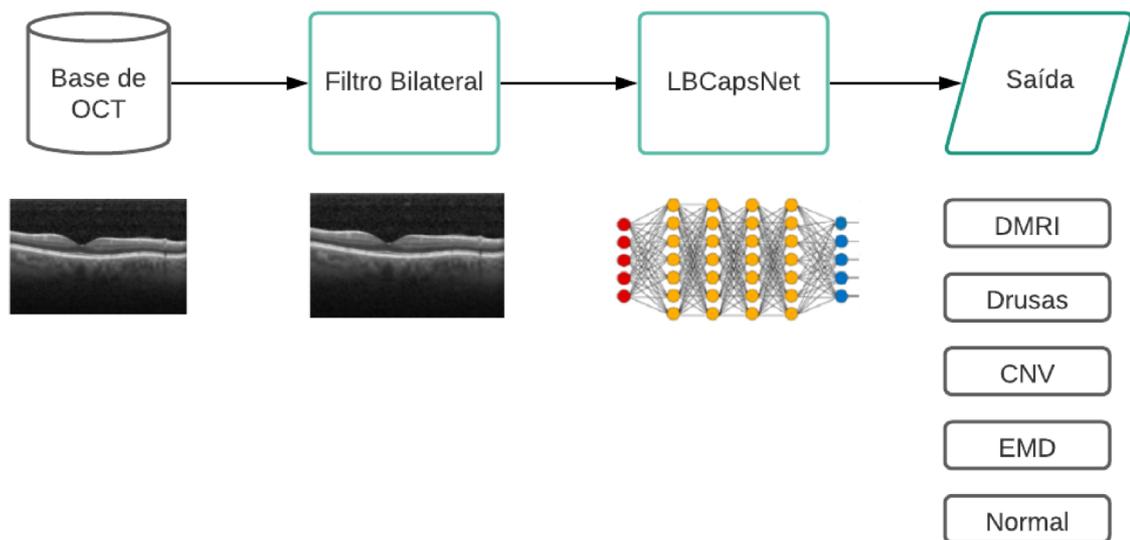
No experimento 2 as bases de imagens foram divididas novamente em treino, validação e teste, mas foram escolhidas proporções das mesmas, ou seja, 70/30, 80/20 e 90/10, que representam respectivamente a porcentagem de treino e validação e a outra a porcentagem de teste. Tal experimento tem o objetivo de verificar o comportamento da rede com a diminuição das bases de treino/validação e aumento da base de teste. Foram feitas 5 repetições de cada teste, com divisão aleatória, este processo tem o objetivo de garantir que os modelos gerados mantenham um padrão de desempenho mesmo com a mudança das imagens utilizadas em treino, validação e teste.

Todos os experimentos foram realizados utilizando 100 épocas como padrão para as redes neurais, este valor foi estipulado a partir da verificação de que as mesmas já estavam estabilizadas neste ponto do treinamento.

## 5 Método para Classificação de Imagens de OCT Utilizando uma Nova Arquitetura de *Capsule Network* - LBCapsNet

Este capítulo apresenta um método para classificação de imagens *B-scans* de OCT em cinco classes, normal, DMRI, drusas, neovascularização e edema macular diabético. O processo é dividido em duas etapas: pré-processamento e classificação. A Figura 26 mostra o fluxograma com as etapas e as técnicas utilizadas nos mesmos.

Figura 26 – Fluxograma da metodologia proposta.



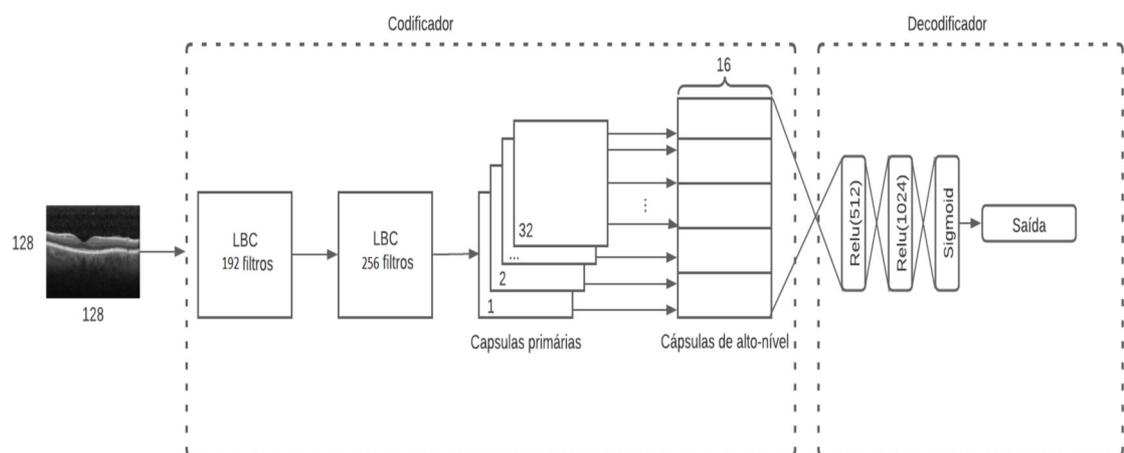
### 5.1 Pré-processamento

Depois do redimensionamento, as imagens são submetidas a etapa de pré-processamento. Neste trabalho é utilizado o filtro bilateral para diminuir a quantidade de ruídos, característico de imagens *B-scans* de OCT. Trata-se de um filtro de suavização (*smoothing*) não linear, que preserva as bordas e borra (*blur*) as demais regiões. A Figura 21 apresenta o resultado da aplicação do filtro.

## 5.2 Capsule Network com Local Binary Convolution Module - LBCapsNet

Esta etapa consiste em gerar um modelo de classificação para rotular as imagens em DMRI, drusas, neovascularização, EMD e normal utilizando uma rede neural profunda. Foi proposta uma modificação na arquitetura da *capsule network* (CapsNet), onde utilizou-se camadas convolucionais baseadas no LBP (Seção 2.6.1). A Figura 27 apresenta a organização da arquitetura.

Figura 27 – Fluxograma da arquitetura *capsule network* com camadas LBC.



As redes neurais convolucionais tem sido muito utilizadas nos últimos anos para detecção e diagnóstico de patologias baseados em imagens, entretanto tais métodos tem a característica de depender de uma grande quantidade de dados para a realização do treinamento. Outra característica da CNN é que ela realiza o roteamento entre camadas por meio de operações de *pooling* que descartam informações como localização e posicionamento dos objetos que podem ser valiosos para fins de classificação ou detecção. Para contornar tais problemas foi proposta em (SABOUR; FROSST; HINTON, 2017) uma rede neural profunda baseada no conceito de cápsulas.

A ideia por trás da *capsule network* é que todo objeto é formado por porções menores localizadas em determinada posição, com isso nosso cérebro combina essa informação hierárquica para fazer o reconhecimento, assim, não é preciso processar todos os detalhes. Os responsáveis por realizarem tal processamento são as chamadas cápsulas, que são

grupos de neurônios cujo vetor de atividade representa vários parâmetros de pose e o comprimento desse vetor mostra a probabilidade de que uma entidade específica exista.

A rede é dividida em duas partes, um codificador e um decodificador. O codificador recebe como entrada as imagens para a extração de descritores vetoriais. Já o decodificador é responsável por identificar a qual classe cada imagem corresponde e fornecer a saída da rede.

A parte do codificador consiste em quatro camadas. Sua primeira camada na arquitetura original é composta por uma camada de convolução, na arquitetura proposta utilizamos dois módulos convolucionais LBC, que diferentemente da *capsule network*, descreverá características de textura da imagem.

A segunda camada contém 32 cápsulas primárias. Sua função é gerar combinações de descritores a partir das características básicas extraídas pelas camadas anteriores. Eles são organizados em 16 canais, onde cada cápsula é composta por uma camada de LBC  $9 \times 9$ .

A última camada do codificador consiste em cinco cápsulas de alto nível. Elas são responsáveis por codificar os descritores vetoriais que representam as cinco classes a serem classificadas.

O algoritmo de roteamento dinâmico (SABOUR; FROSST; HINTON, 2017) é empregado para codificar entradas de cápsulas primárias em saídas de cápsulas de alto nível (16 dimensões). Os comprimentos desses vetores de previsão são usados no decodificador para determinar a classe prevista para cada imagem.

O decodificador consiste em três camadas, duas camadas densas e uma totalmente conectada, ele pega a saída do codificador e a decodifica em uma classe correspondente.

### 5.2.1 Treinamento

O treinamento utiliza a mesma configuração citada na Seção 4.3.1 e com as mesmas bases de imagens, entretanto neste método são utilizadas as imagens originais, sem segmentação das bordas. Na tabela abaixo são listados os parâmetros de treinamento.

Tabela 2 – Arquitetura da rede residual.

| Camada                 | Configuração                 | quantidade |
|------------------------|------------------------------|------------|
| LBC1                   | $3 \times 3$ , 192, stride 2 | 1          |
| LBC2                   | $3 \times 3$ , 264, stride 2 | 1          |
| Capsulas Primárias     | LBC, $3 \times 3$ , stride 2 | 1          |
| Cápsulas de alto nível | Roteamento dinâmico          | 1          |
| Dense                  | Relu                         | 1          |
| Dense                  | Sigmoid                      | 1          |

## 6 Resultados e Discussão

Este capítulo apresenta os resultados obtidos aplicando os métodos propostos para segmentação das camadas da retina e para a classificação de imagens de OCT com degeneração macular relacionada à idade e edema macular diabético.

Para validar o método de segmentação calcula-se o erro de segmentação em comparação à base de volumes de OCT com marcação das bordas da MLI, EPR e MB (Seção 4.1). As marcações das bordas presentes nas fatias dos volumes foram feitas de forma automática e revisadas por um especialista. Para a classificação das imagens foram utilizadas duas bases de imagens rotuladas com uma grande quantidade de *B-Scans* para cada uma das classes presentes (Seção 4.1). A validação é feita com métricas que descrevem a porcentagem de acerto em relação às patologias.

Nas seções seguintes são apresentados e discutidos todos os resultados dos experimentos realizados.

### 6.1 Hardware utilizado

Para o desenvolvimento e aplicação das técnicas apresentadas neste trabalho foram utilizadas duas máquinas com as seguintes configurações:

1. Processador Intel Core I7, 16gb de memória RAM, duas placas de vídeo com 8gb de memória cada e 1Tb de HD;
2. Processador Intel Core I7, 16gb de memória RAM, uma placa de vídeo de 6gb de memória e 1Tb de HD;

### 6.2 Resultados da Segmentação das Camadas da Retina

Nesta seção são apresentados os resultados encontrados nos testes realizados com o método de segmentação das camadas da retina apresentados na Seção 4.

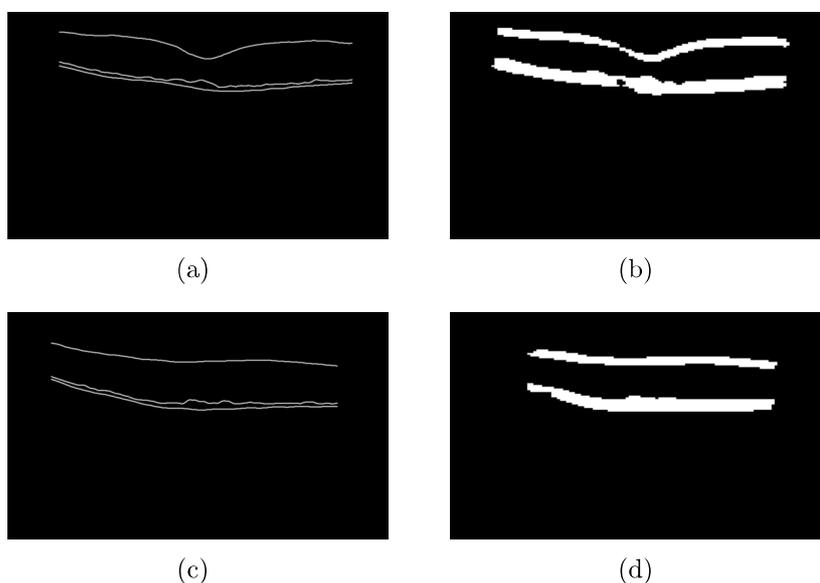
A base de imagens, Farsiu, foi dividida em dois grupos. Um grupo composto por 308 volumes de OCT foi destinado ao treinamento e outro com 76 volumes foi reservado para testes. A seleção para a composição dos grupos foi realizada de forma aleatória, porém

respeitando a proporção da distribuição de exames com DMRI e do grupo de controle existentes na base (269:115). Logo, o conjunto de imagens para o treinamento foi composto por 220 volumes com DMRI e 88 do grupo de controle, e foi utilizado para treinar as redes que compõem as etapas de segmentação inicial e final. Já o conjunto de teste foi composto por 50 volumes com DMRI e 26 do grupo de controle.

Para cada volume do conjunto de treinamento, foram selecionadas 5 *B-scans* para treinar as redes, logo composto por 1540 imagens, sendo 540 utilizadas para validação. Para o teste utilizou-se 3.600 *B-scans*, as imagens eliminadas não continham a marcação do especialista ou eram muito pequenas.

A etapa de segmentação inicial têm o objetivo de separar duas áreas na imagem onde estão posicionadas as bordas das camadas, podemos verificar pela estrutura da retina que a MLI fica separada das outras duas (EPR, MB), como explanado na Seção 4.2.4. A U-net foi utilizada para tanto e gerou os resultados pretendidos, alcançando 97% de Dice no teste. Na Figura 28 podemos ver exemplos dos resultados gerados pela etapa.

Figura 28 – Resultados da U-Net. Na esquerda a imagem original da marcação e na direita o resultado da rede.



A Tabela 3 apresenta os resultados obtidos para a segmentação da borda da MLI, da borda interna do EPR e da borda externa da MB, com os resultados do cálculo do erro entre o método de segmentação automática e as marcações fornecidas pela base. Os valores da média do EMA são considerados baixos, apresentando valores mais elevados para as bordas do EPR e MB, pois em retinas com DMRI formam-se drusas que causam

deformações entre essas duas camadas. Os resultados apresentaram valores de desvio padrão baixos, evidenciando que o método é bem consistente, inclusive entre classes.

Nas Figuras 29 e 30 podemos ver casos de acerto, onde o método conseguiu gerar um resultado muito próximo da marcação da base, uma característica que se destaca é a capacidade de detectar as curvaturas presentes nas bordas das camadas, principalmente na EPR e MB, pois as mesmas sofrem com as deformações causadas pelas drusas. Na Figura 31 temos casos de erro, onde o método gerou algumas falhas na segmentação das bordas, isto se dá pelas características das imagens que podem mudar muito de uma para a outra, mesmo com o pré-processamento a grande quantidade de ruídos influi nestes casos, a rede treinada não reconhece aquela área como pertencente à borda.

Tabela 3 – Resultados da segmentação das bordas das camadas da retina.

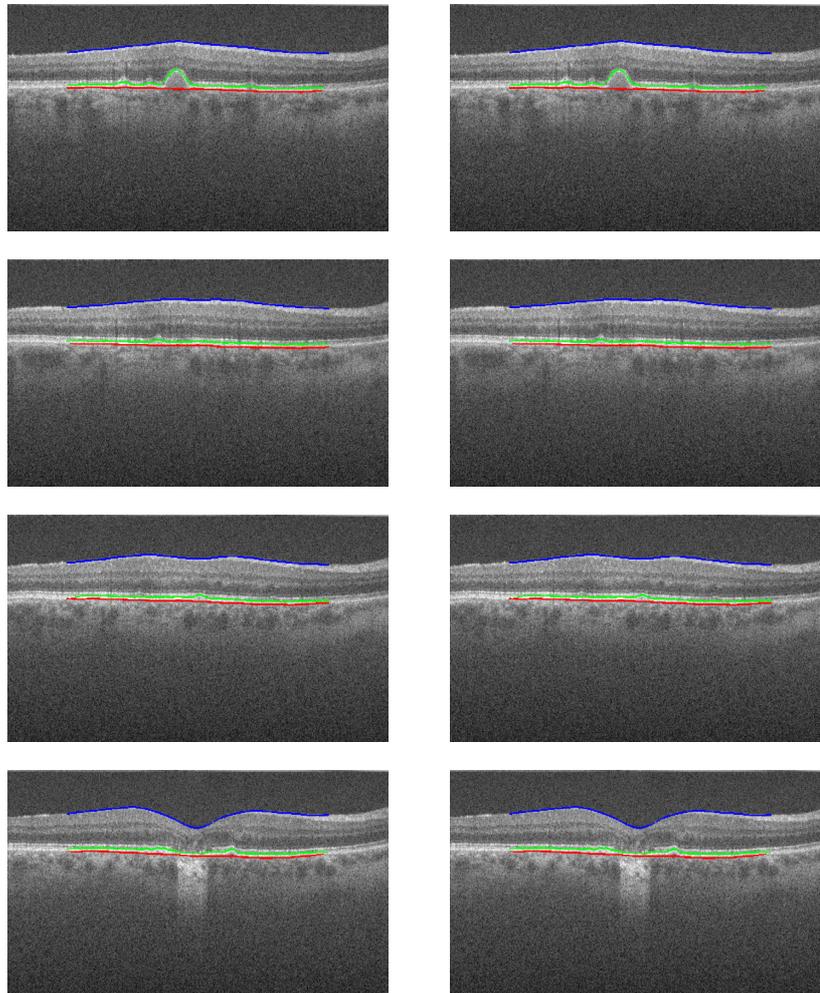
| Grupo   | Borda | EMA   |               | Hausdorff |               |
|---------|-------|-------|---------------|-----------|---------------|
|         |       | Média | Desvio Padrão | Média     | Desvio Padrão |
| DMRI    | MLI   | 0,49  | 0,10          | 4,50      | 2,62          |
|         | EPR   | 0,56  | 0,06          | 5,67      | 2,35          |
|         | MB    | 0,70  | 0,13          | 6,03      | 2,33          |
| Control | MLI   | 0,48  | 0,08          | 4,34      | 2,50          |
|         | EPR   | 0,58  | 0,09          | 5,17      | 2,32          |
|         | MB    | 0,59  | 0,08          | 5,79      | 2,39          |
| Todos   | MLI   | 0,49  | 0,09          | 4,43      | 2,57          |
|         | EPR   | 0,57  | 0,07          | 5,48      | 2,35          |
|         | MB    | 0,66  | 0,12          | 5,93      | 2,36          |

A partir do resultados apresentados, o método proposto demonstrou grande robustez na tarefa de segmentar as bordas das camadas retiniais em imagens de OCT. A dificuldade dessa tarefa é amplamente reconhecida. A maioria dos estudos obtiveram resultados inferiores usando o mesmo conjunto de dados.

A utilização de um conjunto de dados público é uma característica importante deste trabalho, pois permite que outros autores reproduzam os experimentos e estabeleçam análises comparativas. Foi utilizado o conjunto de dados de (FARSIU et al., 2014), que é composto por exames de OCT em pacientes com manifestações de DMRI intermediária e saudáveis.

A segmentação das bordas das camadas é um processo totalmente automatizado realizado inteiramente por dois modelos de CNN, que são ferramentas muito robustas que executam recursos implícitos de extração e seleção de características. Este é um

Figura 29 – Exemplos de casos de acerto. Imagens do exame “AMD\_1057”, na coluna da esquerda estão os resultados do método, na coluna da direita estão as marcações da base.



aspecto positivo porque elimina a necessidade de determinar empiricamente o conjunto de características a serem usados no processo de aprendizagem.

A etapa de segmentação final é realizada pela DexiNed que é uma rede neural convolucional criada especificamente para detecção de bordas. A combinação de todas as técnicas deste trabalho proporcionaram uma melhor segmentação como mostram os resultados apresentados na Tabela 3.

Apesar de apresentar vários fatores positivos, o método proposto também possui algumas limitações. A base de dados é um conjunto muito complexo e as imagens possuem muitos ruídos, a etapa de pré-processamento que utiliza o filtro bilateral busca amenizar isto. As imagens para pacientes com DMRI só conta com casos intermediários, assim limitando a atuação do método.

Figura 30 – Exemplos de casos de acerto. Imagens do exame “AMD \_1090”, na coluna da esquerda estão os resultados do método, na coluna da direita estão as marcações da base.

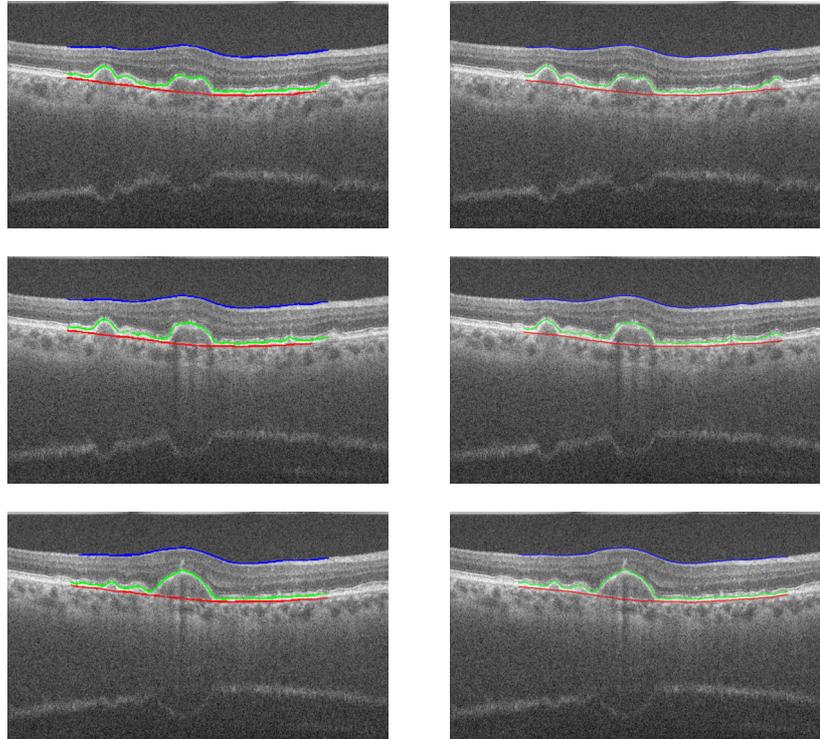
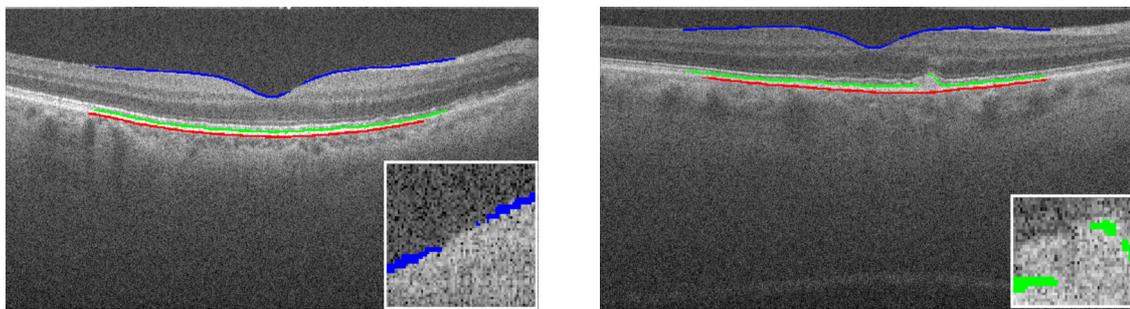


Figura 31 – Exemplos de erros na segmentação.



(a) Exam B-scan 46 “AMD\_1053”.

(b) Exam B-scan 32 “AMD\_1081”.

### 6.2.1 Comparação com Trabalhos Relacionados

A Tabela 6.2.1 apresenta uma comparação do método proposto com outros trabalhos que utilizaram a mesma base de imagens. Os resultados demonstram que o método proposto permite realizar a segmentação das bordas com boa precisão. Nossa metodologia conseguiu alcançar valores de EMA menores ou bem próximos a todos os resultados dos trabalhos relacionados, obtendo um desvio padrão bem menor que a maioria, evidenciando a robustez e generalização do método. Para o MB e o EPR o método proposto apresenta os menores valores de erro médio absoluto e desvio padrão, para o MLI, os trabalhos de (KUGELMAN

et al., 2018) e (CHEN et al., 2019) apresentam valores de EMA ligeiramente melhores. Entretanto a variação é maior, alcançando valores máximos de 0,59 e 1,3 respectivamente, eles também utilizaram uma quantidade muito pequena de imagens para testar os métodos propostos em seus trabalhos, utilizando 200 e 300 fatias respectivamente.

Tabela 4 – Comparação com trabalhos relacionados.

| Autor                   | Método     | Borda | Erro Médio Absoluto |               |
|-------------------------|------------|-------|---------------------|---------------|
|                         |            |       | Média               | Desvio Padrão |
| (KUGELMAN et al., 2018) | RNN-GS     | MLI   | 0,38                | 0,92          |
|                         |            | EPR   | 1,05                | 2,91          |
|                         |            | MB    | 2,07                | 4,31          |
| (KUGELMAN et al., 2018) | CNN-GS     | MLI   | 1,10                | 7,21          |
|                         |            | EPR   | 1,17                | 3,15          |
|                         |            | MB    | 2,31                | 4,60          |
| (KUGELMAN et al., 2018) | FCN-GS     | MLI   | 0,65                | 4,24          |
|                         |            | EPR   | 1,03                | 2,97          |
|                         |            | MB    | 1,53                | 3,50          |
| (SANTOS, 2019)          | CapsNet    | MLI   | 0,75                | 3,42          |
|                         |            | EPR   | 0,93                | 0,86          |
|                         |            | MB    | 1,09                | 2,49          |
| (CHEN et al., 2019)     | DeepForest | MLI   | 0,47                | 0,12          |
|                         |            | EPR   | 0,99                | 0,48          |
|                         |            | MB    | 1,24                | 0,52          |
| (LOU et al., 2020)      | WAVE       | MLI   | 0,68                | 4,09          |
|                         |            | EPR   | 1,22                | 2,80          |
|                         |            | MB    | 1,90                | 3,01          |
| (HU et al., 2021)       | ResNet     | MLI   | 1,23                | 4,77          |
|                         |            | EPR   | 1,93                | 5,10          |
|                         |            | MB    | 2,46                | 4,97          |
| Nosso método            | DexiNed    | MLI   | 0,49                | 0,09          |
|                         |            | EPR   | 0,57                | 0,07          |
|                         |            | MB    | 0,66                | 0,12          |

### 6.3 Resultados dos Métodos para Detecção de DMRI e EMD

Nesta tese propomos dois métodos para a classificação de imagens de OCT. O primeiro utiliza a imagem resultante da segmentação das camadas da retina e aplica em uma rede neural ResNet50. O segundo utiliza uma nova arquitetura de CapsNet com camadas convolucionais baseadas no LBP, nomeada LBCapsNet, para classificação de imagens de OCT em normal, DMRI, drusas, neovascularização coroidal e edema macular diabético. Diversos experimentos foram realizados, onde buscamos comprovar a eficiência

dos métodos. Para avaliar o desempenho dos métodos, utilizamos as métricas sensibilidade, especificidade, acurácia e precisão, descritas na Seção 2.7.4.

Foram realizados dois experimentos, como mostra a Seção 4.1. O experimento 1 utilizou as imagens com a divisão original da base ZhangLab, na qual temos 250 imagens de cada classe para teste e o restante para treino e validação, para a classe DMRI também foram separadas 250 imagens para teste. No segundo experimento utilizamos dois conjuntos de imagens, ZhangLab e Farsiu, a base de teste aumentou e passou a ter divisão proporcional de imagens para cada classe. No experimento 1 a metodologia foi aplicada com a utilização de três diferentes abordagens para a etapa de classificação de modo a propiciar uma comparação dos métodos propostos nesta tese com a arquitetura CapsNet Original proposta em (SABOUR; FROSST; HINTON, 2017) e a arquitetura LBCNN proposta em (JUEFEI-XU; BODDETI; SAVVIDES, 2017). As imagens foram redimensionadas para  $128 \times 128$ , onde esse tamanho foi definido através dos experimentos com o hardware disponível.

Tabela 5 – Resultados do experimento 1.

| Arquitetura                     | Acurácia (%) | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|---------------------------------|--------------|-------------------|--------------------|--------------|
| LBCNN                           | 50,7         | 75,3              | 20                 | 47,2         |
| CapsNet Original                | 92           | 93,3              | 88                 | 95           |
| Método proposto com segmentação | 98,1         | 97,6              | 98,4               | 97,6         |
| LBCapsNet                       | 99,2         | 99                | 99,3               | 99,7         |

Tabela 6 – Resultados do experimento 2 - base ZhangLab - LBCapsNet.

| Treino/Teste | Acurácia (%) | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|--------------|--------------|-------------------|--------------------|--------------|
| 90/10        | 91,2         | 90,3              | 91,9               | 90,3         |
| 80/20        | 90,7         | 89,5              | 91,3               | 89,6         |
| 70/30        | 90,1         | 89,3              | 90,8               | 89           |

Tabela 7 – Resultados do experimento 2 - base Farsiu - LBCapsNet.

| Treino/Teste | Acurácia (%) | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|--------------|--------------|-------------------|--------------------|--------------|
| 90/10        | 95           | 94,7              | 95,6               | 98           |
| 80/20        | 92,2         | 92,8              | 90,7               | 95           |
| 70/30        | 90,6         | 90,3              | 91,3               | 96           |

A Tabela 5 apresenta os resultados usando o experimento 1. Verifica-se que os dois métodos propostos apresentam os melhores resultados entre as arquiteturas. As Tabelas 6, 7, 8 e 9 mostram os resultados dos testes usando o experimento 2, o método

Tabela 8 – Resultados do experimento 2 - base ZhangLab - Método proposto com segmentação.

| Treino/Teste | Acurácia (%) | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|--------------|--------------|-------------------|--------------------|--------------|
| <b>90/10</b> | 88,7         | 87,3              | 89,9               | 87,3         |
| <b>80/20</b> | 87,2         | 86,8              | 88,5               | 86,6         |
| <b>70/30</b> | 87,6         | 86,9              | 88                 | 86           |

Tabela 9 – Resultados do experimento 2 - base Farsiu - Método proposto com segmentação.

| Treino/Teste | Acurácia (%) | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|--------------|--------------|-------------------|--------------------|--------------|
| <b>90/10</b> | 97,9         | 98,3              | 97,1               | 98,7         |
| <b>80/20</b> | 96,3         | 96,6              | 95,6               | 98           |
| <b>70/30</b> | 93,8         | 96,7              | 86,9               | 94,5         |

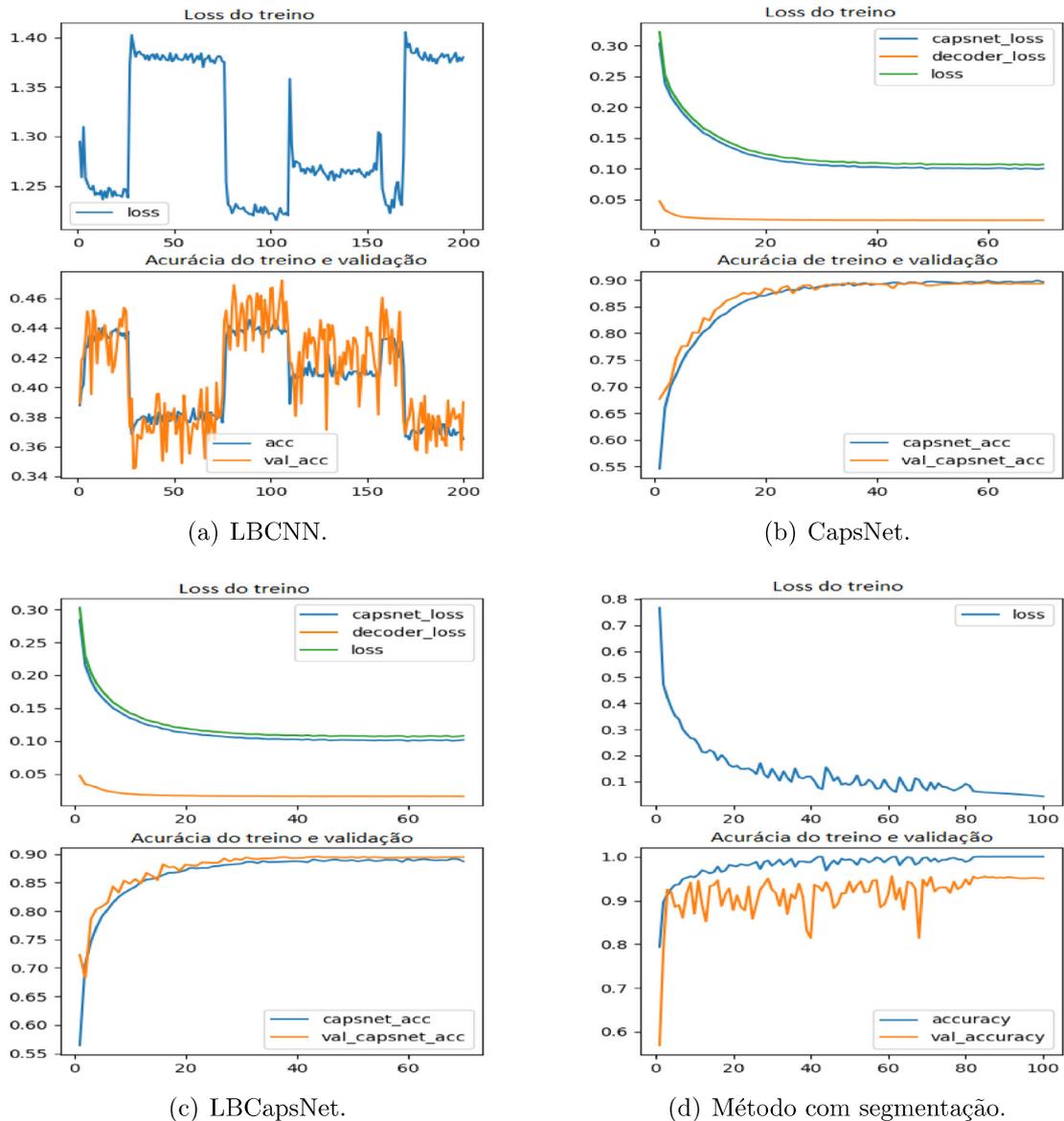
com segmentação obteve os melhores resultados na base Farsiu e a LBCapsNet na base ZhangLab. A diferença de performance na base Farsiu se dá pelas características das imagens, pois as mesmas são modificadas e apresentam recortes e rotações, dificultando a etapa de segmentação. Houve uma queda de desempenho de um experimento para o outro, acredita-se que com o aumento na quantidade de imagens testadas, há uma maior heterogeneidade das imagens.

Tabela 10 – Tempo de execução do treinamento.

| Arquitetura             | Tempo (s) |
|-------------------------|-----------|
| <b>CapsNet Original</b> | 97.869    |
| <b>LBCNN</b>            | 20.328    |
| <b>LBCapsNet</b>        | 20.988    |

Na Tabela 10 é apresentada uma comparação dos tempos de execução do treino, a CapsNet original caracteriza-se por demorar um tempo muito maior que as outras arquiteturas de CNN, e os resultados mostram isso. A arquitetura proposta resolve tal problema da CapsNet e com resultados melhores.

Figura 32 – Curvas de aprendizagem.



A Figura 32 apresenta as curvas de aprendizagem e a acurácia durante o treinamento e validação das redes para classificação, demonstrando que as redes utilizadas possuem convergência rápida, ou seja, em poucas épocas atingem bons resultados.

Os métodos propostos mostraram grande robustez na tarefa de classificação de imagens de OCT. Analisando o desenvolvimento e a avaliação do trabalho, podemos destacar os seguintes aspectos:

1. apresenta-se um método que utiliza a imagem binária resultado da segmentação para realizar a classificação e como o formato das bordas revelam aspectos do estado fisiológico da retina, isso levou a maior assertividade do método;

2. LBCapsNet, uma arquitetura baseada na CapsNet que, de acordo com as métricas de avaliação utilizadas, melhora os resultados para o problema proposto. Além disso, a arquitetura proposta apresenta um menor tempo de treinamento da rede quando comparada com a arquitetura original;
3. a utilização da camada convolucional baseada no LBP que descreve características de textura em substituição a convolução tradicional, se mostrou determinante para os resultados;
4. os experimentos realizados, utilizando dois conjuntos de dados públicos permitem observar que os métodos propostos também apresentam bons resultados de generalização, e por se tratarem de dados públicos facilitam a reprodução deste trabalho e comparação com outros métodos que venham a surgir;
5. cada um dos métodos de classificação obteve melhores resultados em uma das bases, as características de cada base levou a tal resultado, e cada método se mostrou eficaz;
6. os valores de precisão e sensibilidade apresentados pela LBCapsNet demonstram que o método tem grande capacidade de classificar corretamente os casos que apresentam doença;
7. a combinação das técnicas, filtro bilateral com LBCapsNet, resultou em melhores resultados, como foi apresentado na Seção 6.3;

Apesar de apresentar vários fatores positivos, o método proposto também possui algumas limitações, entre as quais pode-se destacar:

1. a base de dados é um conjunto muito complexo e as imagens possuem muitos ruídos, a etapa de pré-processamento que utiliza o filtro bilateral busca amenizar isso, mas ainda dificulta muito o processo de classificação;
2. os modelos de aprendizado profundo, como a arquitetura proposta na tese, têm um grande número de parâmetros. Frequentemente, esse conjunto de parâmetros é empiricamente determinado pelo usuário, o que pode não fornecer ótimos resultados;

### 6.3.1 Comparação com Trabalhos Relacionados

A Tabela 12 apresenta uma comparação de desempenho com outros trabalhos que foram descritos na Seção 3.2. A maioria dos trabalhos utiliza a base OCT2017 que também foi liberada por (KERMANY et al., 2018), difere de ZhangLab na quantidade de imagens na classe Normal e nas imagens do teste, onde apesar de possuírem a mesma quantidade, são imagens diferentes. Assim, fizemos um teste com as imagens da base OCT2017 para fim de comparação. Podemos verificar que o método proposto apresenta resultados melhores que a maioria dos trabalhos.

Para ter uma melhor comparação com o trabalho de (ALQUDAH, 2020), testamos adicionando imagens da DMRI da base Farsiu no treino (26.650 imagens) e no teste (250 imagens) e ter uma saída com 5 classes (DMRI, EMD, Drusas e Neovascularização Coroidal), e os dois métodos alcançaram acurácia melhor. Kermany et al. (2018), Kaymak e Serener (2018), Li et al. (2019) e Kamran et al. (2019) utilizam somente a ZhangLab. Alcançamos resultados melhores que Kermany et al. (2018) e não utilizamos transferência de aprendizagem. Li et al. (2019) e Kamran et al. (2019) conseguiram alguns resultados melhores, entretanto, ambos os trabalhos utilizam tamanhos de imagens bem maiores, assim exigindo maior poder computacional e a classificação é feita em 4 classes, o nosso com 5 classes. Nesses dois trabalhos, o método não foi testado com uma quantidade maior de imagens. Em (LI et al., 2019) o método também foi testado na base de (SRINIVASAN et al., 2014), entretanto tal base possui apenas 324 imagens para teste.

Como a base de teste é muito pequena os resultados acabam sendo muito próximos e todos muito bons. Nesta seção mostramos que com o aumento da amostra para teste, as estatísticas dos resultados diminuem, mas isso pode ser explicado por uma provável maior heterogeneidade da base, o que é menos provável com uma quantidade pequena de imagens, e somente nosso trabalho realizou tal teste.

Tabela 11 – Comparação com trabalhos relacionados.

| Autor                  | Técnica                         | Imagens        | ACC.        | SEN.        | ESP.        |
|------------------------|---------------------------------|----------------|-------------|-------------|-------------|
| (ALQUDAH, 2020)        | AOCT-NET                        | 137.437        | 97,1        | 97,1        | 99,2        |
| <b>Método Proposto</b> | <b>LBCapsNet</b>                | <b>135.212</b> | <b>99</b>   | <b>97,6</b> | <b>99,4</b> |
| <b>Método Proposto</b> | <b>ResNet50 com Segmentação</b> | <b>135.212</b> | <b>98,5</b> | <b>96</b>   | <b>99</b>   |

Tabela 12 – Comparação com trabalhos relacionados.

| Autor                   | Técnica                           | Imagens        | ACC.        | SEN.        | ESP.        |
|-------------------------|-----------------------------------|----------------|-------------|-------------|-------------|
| (KERMANY et al., 2018)  | InceptionV3 com Transfer Learning | 108.312        | 96,6        | 97,8        | 97,4        |
| (KAYMAK; SERENER, 2018) | AlexNet                           | 84.484         | 97,1        | 99,6        | 98,4        |
| (LI et al., 2019)       | VGG-16 com Transfer Learning      | 84.484         | 98,6        | 97,8        | 99,4        |
| (KAMRAN et al., 2019)   | OpticNet                          | 84.484         | 99,8        | 99,8        | 99,9        |
| <b>Método Proposto</b>  | <b>LBCapsNet</b>                  | <b>108.312</b> | <b>99,2</b> | <b>99</b>   | <b>99,3</b> |
| <b>Método Proposto</b>  | <b>ResNet50 com Segmentação</b>   | <b>108.312</b> | <b>98,1</b> | <b>97,6</b> | <b>98,4</b> |

## 7 Conclusão

O diagnóstico precoce da degeneração macular e do edema macular é muito importante pois são doenças oculares que podem causar deficiência visual. Nesta tese, apresentamos dois métodos para a detecção automática de DMRI e EMD em imagens de OCT. A primeira utiliza o resultado da segmentação das camadas da retina, que é uma tarefa desafiadora, dificultado pelo fato de imagens de retinas, com uma das doenças, poderem apresentar deformações causadas pelas mesmas.

O método proposto para a segmentação das bordas das camadas da retina alcançou bons resultados, com erros médio absolutos melhores aos encontrados no estado da arte para as camadas do epitélio pigmentar retiniano e membrana de Bruch. Os valores de desvio padrão baixos evidenciam a consistência do método. A utilização de uma rede especializada em detecção de bordas se mostrou determinante para o sucesso do método.

Para a detecção das doenças apresentamos dois métodos. O primeiro método utiliza uma imagem binária resultante da etapa de segmentação das camadas e aplica em uma rede residual ResNet50 para realizar a classificação, os resultados comprovam que a utilização de tais imagens ajudam no processo de classificação. O segundo método tem como principal componente, uma rede baseada no conceito de cápsulas que utilizam camadas convolucionais inspiradas na técnica de extração de características de textura, *Local Binary Pattern*, para fazer a classificação das *B-Scans*. A rede mostrou um resultado promissor, levando em conta que a entrada da rede foi somente a imagem pré-processada com o filtro bilateral. Entretanto, tal método necessita de melhorias, pois ficou evidente que com o aumento da base de teste o resultado sofreu uma queda de desempenho. Na comparação com a arquitetura original da *capsule network* apresentada na Seção de resultados, podemos verificar que a adição das camadas LBC melhoraram a arquitetura em termos tanto de resultados como de tempo de treinamento.

Assim, a tese apresenta três boas contribuições comprovadas nos resultados, um método para segmentação das bordas das camadas da retina em imagens de OCT, um método para detecção DMRI e EMD em imagens de OCT que utiliza o resultado da segmentação das camadas, e uma nova arquitetura de rede neural de cápsulas.

## 7.1 Trabalhos Futuros

Apesar dos resultados obtidos serem satisfatórios para a segmentação das bordas das camadas da retina e para os métodos de classificação, melhorias ainda podem ser realizadas visando incrementar a sua eficiência. Algumas sugestões são apresentadas a seguir:

- otimizar parâmetros das redes neurais;
- modificar camada LBC para reproduzir modificações do LBP clássico;
- testar a arquitetura da CapsNet com camadas LBC em imagens com outras enfermidades;

## 7.2 Produção científica

O desenvolvimento desta tese levou a publicação do artigo “*Automatic segmentation of retinal layers in OCT images with intermediate age-related macular degeneration using U-Net and DexiNed*”, na revista PlosOne, que possui qualis B1 segundo a Plataforma Sucupira (<https://sucupira.capes.gov.br/sucupira/>) da CAPES.

# Referências

- ALI, M.; PARLAPALLI, R. Signal processing overview of optical coherence tomography systems for medical imaging. In: SPRABB9. Dallas: Texas Instruments, 2010.
- ALQUDAH, A. M. Aoct-net: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Medical & biological engineering & computing*, Springer, v. 58, n. 1, p. 41–53, 2020.
- ÁVILA, M.; ALVES, M.; NISHI, M. *As condições de Saúde Ocular no Brasil-2015. São Paulo (SP): Conselho Brasileiro de Oftalmologia; 2015.[cited 2016 Dec 6]. 145p.* 2016.
- BENGIO, Y.; LECUN, Y. et al. Scaling learning algorithms towards ai. *Large-scale kernel machines*, v. 34, n. 5, p. 1–41, 2007.
- BUSCHINI, E.; PIRAS, A.; NUZZI, R.; VERCELLI, A. Age related macular degeneration and drusen: neuroinflammation in the retina. *Progress in neurobiology*, Elsevier, v. 95, n. 1, p. 14–25, 2011.
- CBO. Conselho Brasileiro de Oftalmologia. DMRI. 2017. Disponível em: <http://www.cbo.net.br/novo/publico-geral/dmri.php>. Acesso em: 03 jun. 2017.
- CHAKRAVARTY, A.; SIVASWAMY, J. End-to-end learning of a conditional random field for intra-retinal layer segmentation in optical coherence tomography. In: SPRINGER. *Annual Conference on Medical Image Understanding and Analysis*. [S.l.], 2017. p. 3–14.
- CHEN, Z.; LI, D.; SHEN, H.; MO, Y.; WEI, H.; OUYANG, P. Automated retinal layer segmentation in oct images of age-related macular degeneration. *IET Image Processing*, IET, v. 13, n. 9, p. 1401–1411, 2019.
- CHEW, E. Y.; CLEMONS, T.; SANGIOVANNI, J. P.; DANIS, R.; DOMALPALLY, A.; MCBEE, W.; SPERDUTO, R.; FERRIS, F. L. The age-related eye disease study 2 (AREDS2): Study design and baseline characteristics (AREDS2 report number 1). *Ophthalmology*, v. 119, n. 11, p. 2282 – 2289, 2012. ISSN 0161-6420.
- CHIU, S. J.; IZATT, J. A.; O’CONNELL, R. V.; WINTER, K. P.; TOTH, C. A.; FARSIU, S. Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Investigative Ophthalmology & Visual Science*, v. 53, n. 1, p. 53, 2012.
- CHIU, S. J.; LI, X. T.; NICHOLAS, P.; TOTH, C. A.; IZATT, J. A.; FARSIU, S. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Biomedical Optics Express*, OSA, v. 18, n. 18, p. 19413–19428, aug 2010.
- CHOU, C.-F.; COTCH, M. F.; VITALE, S.; ZHANG, X.; KLEIN, R.; FRIEDMAN, D. S.; KLEIN, B. E.; SAADDINE, J. B. Age-related eye diseases and visual impairment among us adults. *American journal of preventive medicine*, Elsevier, v. 45, n. 1, p. 29–35, 2013.
- CONCI, A.; GALVÃO, S. S.; SEQUEIROS, G. O.; SAADE, D. C.; MACHENRY, T. A new measure for comparing biomedical regions of interest in segmentation of digital images. *Discrete Applied Mathematics*, v. 197, p. 103 – 113, 2015. ISSN 0166-218X.

- CONCI, A.; KUBRUSLY, C. Distance between sets—a survey. *Advances in Mathematical Sciences and Applications*, v. 20, n. 1, p. 1 – 18, 2017.
- DEBUC, D. C. *A review of algorithms for segmentation of retinal image data using optical coherence tomography*. INTECH Open Access, 2011. Disponível em: <http://cdn.intechopen.com/pdfs-wm/15371.pdf>. Acesso em: 01 jan. 2016.
- DUKER, J.; WAHEED, N.; GOLDMAN, D. *Manual de OCT em Retina*. Rio de Janeiro: Elsevier, 2015. ISBN 9788535283334.
- ETER, N.; SPAIDE, R. F. Comparison of fluorescein angiography and optical coherence tomography for patients with choroidal neovascularization after photodynamic therapy. *Retina*, LWW, v. 25, n. 6, p. 691–696, 2005.
- FANG, L.; CUNEFARE, D.; WANG, C.; GUYMER, R. H.; LI, S.; FARSIU, S. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomedical Optics Express*, OSA, v. 8, n. 5, p. 2732–2744, 2017.
- FARJOOD, F.; VARGIS, E. Novel devices for studying acute and chronic mechanical stress in retinal pigment epithelial cells. *Lab on a Chip*, Royal Society of Chemistry, v. 18, n. 22, p. 3413–3424, 2018.
- FARSIU, S.; CHIU, S. J.; IZATT, J. A.; TOTH, C. A. Fast detection and segmentation of drusen in retinal optical coherence tomography images. In: SPIE BIOS – OPHTHALMIC TECHNOLOGIES, 18., 2008, San Jose, Califórnia. *Proceedings...* Bellingham, Washington: SPIE, 2008. v. 6844, p. 1–12.
- FARSIU, S.; CHIU, S. J.; O’CONNELL, R. V.; FOLGAR, F. A.; YUAN, E.; IZATT, J. A.; TOTH, C. A. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology*, v. 121, n. 1, p. 162 – 172, 2014. ISSN 0161-6420.
- FERNANDES, B. J. T. *Redes neurais com extração implícita de características para reconhecimento de padrões visuais*. Tese (Doutorado) — Tese de Doutorado. Centro de Informática, Universidade Federal de Pernambuco . . . , 2013.
- FERREIRA, J. L. et al. Segmentação automática da próstata em imagens de ressonância magnética utilizando redes neurais convolucionais, mapa probabilístico e treinamento adversário. Universidade Federal do Maranhão, 2019.
- GÉRON, A. Introducing capsule networks. 2018. Disponível em: <https://www.oreilly.com/ideas/introducing-capsule-networks>. Acesso em: 02 jan. 2019.
- GOPINATH, K.; RANGREJ, S. B.; SIVASWAMY, J. A deep learning framework for segmentation of retinal layers from oct images. In: IEEE. *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. [S.l.], 2017. p. 888–893.
- HASSAN, T.; AKRAM, M. U.; HASSAN, B.; NASIM, A.; BAZAZ, S. A. Review of oct and fundus images for detection of macular edema. In: IEEE. *2015 IEEE International Conference on Imaging Systems and Techniques (IST)*. [S.l.], 2015. p. 1–4.

- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- HEE, M. R.; BAUMAL, C. R.; PULIAFITO, C. A.; DUKER, J. S.; REICHEL, E.; WILKINS, J. R.; COKER, J. G.; SCHUMAN, J. S.; SWANSON, E. A.; FUJIMOTO, J. G. Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. *Ophthalmology*, Elsevier, v. 103, n. 8, p. 1260–1270, 1996.
- HU, K.; LIU, D.; CHEN, Z.; LI, X.; ZHANG, Y.; GAO, X. Embedded residual recurrent network and graph search for the segmentation of retinal layer boundaries in optical coherence tomography. *IEEE Transactions on Instrumentation and Measurement*, IEEE, v. 70, p. 1–17, 2021.
- HUANG, D.; SWANSON, E. A.; LIN, C. P.; SCHUMAN, J. S.; STINSON, W. G.; CHANG, W.; HEE, M. R.; FLOTTE, T.; GREGORY, K.; PULIAFITO, C. A. et al. Optical coherence tomography. *Science*, American Association for the Advancement of Science, v. 254, n. 5035, p. 1178–1181, 1991.
- HUANG, D.; SWANSON, E. A.; LIN, C. P.; SCHUMAN, J. S.; STINSON, W. G.; CHANG, W.; HEE, M. R.; FLOTTE, T.; GREGORY, K.; PULIAFITO, C. A. et al. Optical coherence tomography. *Science*, NIH Public Access, v. 254, n. 5035, p. 1178, 1991.
- HUANG, L.; HE, X.; FANG, L.; RABBANI, H.; CHEN, X. Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network. *IEEE Signal Processing Letters*, IEEE, v. 26, n. 7, p. 1026–1030, 2019.
- HUBEL, D. H. The visual cortex of the brain. *Scientific American*, JSTOR, v. 209, n. 5, p. 54–63, 1963.
- HUBEL, D. H.; WIESEL, T. N. Early exploration of the visual cortex. *Neuron*, Elsevier, v. 20, n. 3, p. 401–412, 1998.
- JUEFEI-XU, F.; BODDETI, V. N.; SAVVIDES, M. Local binary convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 19–28.
- KAMRAN, S. A.; SAHA, S.; SABBIR, A. S.; TAVAKKOLI, A. Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images. *arXiv preprint arXiv:1910.05672*, 2019.
- KANSKI, J. J. *Oftalmologia Clínica*. Rio de Janeiro: Elsevier, 2012.
- KAYMAK, S.; SERENER, A. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. In: IEEE. *2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP)*. [S.l.], 2018. p. 265–269.
- KERMANY, D. S.; GOLDBAUM, M.; CAI, W.; VALENTIM, C. C.; LIANG, H.; BAXTER, S. L.; MCKEOWN, A.; YANG, G.; WU, X.; YAN, F. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, Elsevier, v. 172, n. 5, p. 1122–1131, 2018.

- KUGELMAN, J.; ALONSO-CANEIRO, D.; READ, S. A.; VINCENT, S. J.; COLLINS. Automatic segmentation of oct retinal boundaries using recurrent neural networks and graph search. *Biomedical Optics Express*, OSA, v. 9, n. 11, p. 5759–5777, 2018.
- LI, F.; CHEN, H.; LIU, Z.; ZHANG, X.; WU, Z. Fully automated detection of retinal disorders by image-based deep learning. *Graefe's Archive for Clinical and Experimental Ophthalmology*, Springer, v. 257, n. 3, p. 495–505, 2019.
- LOU, S.; CHEN, X.; HAN, X.; LIU, J.; WANG, Y.; CAI, H. Fast retinal segmentation based on the wave algorithm. *IEEE Access*, IEEE, v. 8, p. 53678–53686, 2020.
- MARQUES FILHO, O.; NETO, H. V. *Processamento digital de imagens*. [S.l.]: Brasport, 1999.
- MOBINY, A.; NGUYEN, H. V. Fast CapsNet for lung cancer screening. In: INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER ASSISTED & INTERVENTION – MICCAI, 21., 2018, Granada, Espanha. *Proceedings...* Cham, Switzerland: Springer, 2018. p. 741–749.
- MOTTA, A. A. L. d. *Efeito do tratamento com e sem bevacizumabe intravítreo associado ao controle glicêmico otimizado na função visual e morfologia macular de pacientes com edema macular diabético*. Tese (Doutorado) — Universidade de São Paulo, 2019.
- MOTTA, M. M. d. S.; COBLENTZ, J.; MELO, L. G. N. d. Aspectos atuais na fisiopatologia do edema macular diabético. *Revista Brasileira de Oftalmologia*, SciELO Brasil, v. 67, p. 45–49, 2008.
- NEHEMY, M. B. Degeneração macular relacionada à idade: novas perspectivas. *Arquivos Brasileiros de Oftalmologia*, v. 69, n. 6, p. 955–958, 2006.
- NOH, H.; HONG, S.; HAN, B. Learning deconvolution network for semantic segmentation. In: *Computer Vision (ICCV), 2015 IEEE International Conference on*. [S.l.: s.n.], 2015.
- NOVARTIS. *O que é edema macular diabético (EMD)?* 2017. Disponível em: <https://saude.novartis.com.br/doencas-da-visao/o-que-e-edema-macular-diabetico-emd/>.
- OFTALMOLOGIA, C. B. de. Censo oftalmológico: As condições de saúde ocular no brasil [internet]. *São Paulo: CBO*, 2019.
- OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, Elsevier, v. 29, n. 1, p. 51–59, 1996.
- ORGANIZATION, W. H. *Blindness and vision impairment*. 2019. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>.
- OTSU, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, IEEE, v. 9, n. 1, p. 62–66, 1979.
- RASTOGI, D.; PADHY, R. P.; SA, P. K. Detection of retinal disorders in optical coherence tomography using deep learning. In: IEEE. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. [S.l.], 2019. p. 1–7.

- RIM, P. H. H. et al. Degeneração macular relacionada à idade - estudo dos fatores de risco em uma população brasileira. [sn], 2012.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.
- RUDIN, L. I.; OSHER, S.; FATEMI, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, v. 60, n. 1, p. 259 – 268, 1992.
- SABOUR, S.; FROSST, N.; HINTON, G. E. Dynamic routing between capsules. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 30., 2017. *Proceedings...* New York: Curran, 2017. p. 3856–3866.
- SANTOS, A. M. Diagnóstico de degeneração macular relacionada à idade a partir de tomografia de coerência óptica usando geoestatística e capsule network. Universidade Federal do Maranhão, 2019.
- SILVA, I. F. S. d. et al. Detecção automática da presença de patologia na visão baseada em imagens do teste de brückner. Universidade Federal do Maranhão, 2019.
- SOON-CHUN, P. *Handbook Of Vitreo-retinal Disorder Management: A Practical Reference Guide*. Singapura: World Scientific, 2015. ISBN 9789814663311.
- SORIA, X.; RIBA, E.; SAPPÀ, A. Dense extreme inception network: Towards a robust cnn model for edge detection. In: *The IEEE Winter Conference on Applications of Computer Vision (WACV '20)*. [S.l.: s.n.], 2020.
- SOUZA, L. C. d. *Segmentação de exsudatos duros em em imagens de retinografia para auxílio do diagnóstico de retinopatia diabética*. Dissertação (Mestrado) — Universidade do Rio Grande do Norte, 2019.
- SOUZA, V.; SILVA, L.; SANTOS, A.; ARAÚJO, L. Análise comparativa de redes neurais convolucionais no reconhecimento de cenas. *Anais do Computer on the Beach*, v. 11, n. 1, p. 419–426, 2020.
- SRINIVASAN, P. P.; KIM, L. A.; METTU, P. S.; COUSINS, S. W.; COMER, G. M.; IZATT, J. A.; FARSIU, S. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical Optics Express*, OSA, v. 5, n. 10, p. 3568–3577, 2014.
- STAFF, B. Blausen gallery 2014. *Wikiversity Journal of Medicine*, v. 1, 2014.
- TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, BioMed Central, v. 15, n. 1, p. 29, 2015.
- TIAN, J.; VARGA, B.; SOMFAI, G. M.; LEE, W.-H.; SMIDDY, W. E.; DEBUC, D. C. Real-time automatic segmentation of optical coherence tomography volume data of the macular region. *PloS one*, Public Library of Science, v. 10, n. 8, p. e0133908, 2015.
- TIAN, J.; VARGA, B.; TATRAI, E.; FANNI, P.; SOMFAI, G. M.; SMIDDY, W. E.; DEBUC, D. C. Performance evaluation of automated segmentation software on optical coherence tomography volume data. *Journal of biophotonics*, Wiley Online Library, v. 9, n. 5, p. 478–489, 2016.

TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: IEEE. *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. [S.l.], 1998. p. 839–846.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, Luhe, v. 30, n. 1, p. 79–82, 2005.

XI, E.; BING, S.; JIN, Y. Capsule network performance on complex data. *arXiv preprint arXiv:1712.03480*, 2017.

XIE, S.; TU, Z. Holistically-nested edge detection. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2015. p. 1395–1403.

YANOFF, M.; DUKER, J.; AUGSBURGER, J. *Ophthalmology*. 4. ed. Philadelphia: Saunders Elsevier, 2014.