

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Programa de Pós-graduação em Engenharia Elétrica

Douglas Abreu da Rocha

**CLASSIFICAÇÃO DO GRAU DE RETINOPATIA DIABÉTICA E
SEGMENTAÇÃO DE LESÕES EM IMAGENS DE RETINA USANDO
REDES NEURAS CONVOLUCIONAIS**

Belo Horizonte
2021

Douglas Abreu da Rocha

**CLASSIFICAÇÃO DO GRAU DE RETINOPATIA DIABÉTICA E
SEGMENTAÇÃO DE LESÕES EM IMAGENS DE RETINA USANDO
REDES NEURAIIS CONVOLUCIONAIS**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Orientadora: Prof^a. Dr^a. Zélia Myriam Assis Peixoto

Linha de pesquisa: Processamento e Análise de Sinais

Belo Horizonte
2021

FICHA CATALOGRÁFICA

Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

R672c	<p>Rocha, Douglas Abreu da</p> <p>Classificação do grau de retinopatia diabética e segmentação de lesões em imagens de retina usando redes neurais convolucionais / Douglas Abreu da Rocha. Belo Horizonte, 2021.</p> <p>61 f. : il.</p> <p>Orientadora: Zélia Myriam Assis Peixoto</p> <p>Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Engenharia Elétrica</p> <p>1. Retinopatia Diabética - Classificação. 2. Diabetes Mellitus. 3. Processamento de Imagem Assistida por Computador. 4. Redes neurais (Computação). 5. Inteligência artificial. 6. Aprendizado Profundo. 7. Algoritmos computacionais. I. Peixoto, Zélia Myriam Assis. II. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.</p> <p>CDU: 681.3.093</p>
-------	---

Douglas Abreu da Rocha

**CLASSIFICAÇÃO DO GRAU DE RETINOPATIA DIABÉTICA E
SEGMENTAÇÃO DE LESÕES EM IMAGENS DE RETINA USANDO
REDES NEURAIS CONVOLUCIONAIS**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Linha de pesquisa: Processamento e Análise de Sinais

Prof^a. Dr^a. Zélia Myriam Assis Peixoto (Orientadora) – PUC Minas

Prof. Dr. Hugo Neves de Oliveira – IME/USP

Prof^a. Dr^a. Flávia Magalhães Freitas Ferreira – PUC Minas

Prof. Dr. Maury Meirelles Gouvêa Júnior (Suplente) – PUC Minas

Belo Horizonte, 30 de agosto de 2021

Aos meus pais, Ezenilda e Carlos, meus irmãos, Marcelo, Renato e Gabriel.

AGRADECIMENTOS

A Deus, onipresente em minha vida, agradeço por guiar meus passos e me dar força para seguir em frente, mesmo com tantos obstáculos.

Aos meus pais, Ezenilda e Carlos, meus irmãos, Marcelo, Renato e Gabriel, por todo amor, apoio e incentivo, por serem exemplo de que a educação é verdadeiramente transformadora e capaz de mudar nossa realidade. Minhas cunhadas Jéssicas e sobrinhos Lucas e Leonardo, por serem felicidade em nossas vidas.

À minha amiga, orientadora e principal incentivadora para a pesquisa, Zélia, por todo apoio, força, confiança e parceria durante estes anos.

Aos meus professores da graduação, em especial ao Mário Buratto, que também é um querido amigo e uma das minhas inspirações na área da eletrônica.

Aos amigos de longa data e os que fiz durante esse tempo.

Ao PPGEE PUC Minas e demais colegas pesquisadores, que em meio às dificuldades e limitações enfrentadas em nosso país, continuam a acreditar, estimular e contribuir exaustivamente para a evolução da pesquisa, ciência e tecnologia.

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela bolsa de estudos concedida.

"Foi o melhor dos tempos, foi o pior dos tempos; a era da sabedoria, a era da tolice; a época da crença, a época da incredulidade; a estação da Luz, a estação das Trevas; a primavera da esperança, o inverno do desespero; tínhamos tudo diante de nós, tínhamos nada diante de nós (...)" Charles Dickens

RESUMO

A retinopatia diabética é uma das principais complicações do diabetes mellitus. Para avaliar a evolução da doença, comumente são obtidas imagens da retina através do exame de fundo de olho. Estas imagens apresentam problemas como baixo contraste, iluminação não adequada, nível de ruído insatisfatório e outros aspectos que podem dificultar a análise e intervenção médica. Neste contexto, este trabalho visa aplicar a rede neural VGG16 para classificar o grau de retinopatia diabética em 5 categorias e uma classe adicional para informar baixa qualidade da imagem, além da segmentação de 4 tipos de lesões com a rede U-Net, em imagens digitais da retina publicamente disponíveis nas bases de dados DDR, EyePACS/Kaggle e IDRiD. A metodologia utilizada abrange o pré-processamento das imagens de retina, composto por adequação de tamanho, limpeza (remoção das imagens de baixa qualidade das demais classes e inclusão na classe 5), data augmentation e balanceamento das classes, até a fase de treinamento, ajuste de hiperparâmetros, classificação e segmentação das imagens através das redes neurais VGG16 e U-Net, respectivamente. As contribuições deste trabalho estão na melhoria dos resultados alcançados em comparação com trabalhos recentes para as bases de dados IDRiD e DDR e inclusão das métricas de desempenho sensibilidade, especificidade, precisão e F1-score.

Palavras-chave: Classificação. Imagens de Retina. Redes Neurais Convolucionais. Retinopatia Diabética. Segmentação.

ABSTRACT

Diabetic retinopathy is one of the main complications of diabetes mellitus. To assess the progression of the disease, retinal images are commonly obtained through fundus examination. However, those images may present problems such as low contrast, inadequate lighting, unsatisfactory noise level, among other aspects that can compromise medical analysis and intervention. In this context, this work aims to apply the neural network VGG16 to classify the diabetic retinopathy in 5 categories and an additional class (named class 5) to report low quality of digital retinal images and segmentation of 4 lesion types using U-Net, in publicly data available in the DDR, EyePACS/Kaggle and IDRiD databases. The proposed methodology includes pre-processing of retinal images, consisting of size adequacy, data cleaning (the removal of low-quality images from other classes and inclusion of them in class 5), data augmentation and class balancing during the training phase and, hyperparameter adjustment, image classification using the VGG16 neural network and image segmentation using U-Net neural network. Among the tests carried out on the three databases, this proposal presented the best performance for DDR database regarding accuracy, precision, sensitivity, specificity, and F1-score. This work contributes towards the improvement of results comparing with recent works for the IDRiD and DDR databases, and the inclusion of performance metrics sensitivity, specificity, precision, and F1-score.

Keywords: Classification. Retinal Images. Convolutional Neural Networks. Diabetic Retinopathy. Segmentation.

LISTA DE FIGURAS

FIGURA 1 – Exame de fundoscopia com câmera digital	12
FIGURA 2 – Resultado da busca por " <i>Diabetic retinopathy detection</i> " no <i>Web of Science</i>	13
FIGURA 3 – Manifestações de RD em uma imagem de retina, em corte. a) MA formado na parede de um vaso sanguíneo; b) HMs, que são vazamentos de sangue dos vasos em camadas da retina; c) EXs moles, formado pelo depósito de detritos axoplasmáticos em regiões isquêmicas, gerando manchas brancas superficiais; d) EXs duros, formados por depósitos de lipídeos e proteínas que vazam através de vasos sanguíneos anormais.	18
FIGURA 4 – Manifestações de RD em uma imagem de retina	19
FIGURA 5 – Camadas básicas de uma rede neural convolucional.	20
FIGURA 6 – Estrutura da rede VGG16.	24
FIGURA 7 – Estrutura da rede U-Net, destacando-se as etapas do codificador e do decodificador.	25
FIGURA 8 – Diagrama em blocos do sistema proposto, com destaque para os processos de classificação do grau de RD e de segmentação das imagens das lesões	35
FIGURA 9 – Modificação do tamanho da imagem original (a), tornando-a quadrada e reduzindo a área da máscara (b). Imagem 007-0473-000 do banco de dados DDR.	39
FIGURA 10 – Distribuição desbalanceada entre as classes das bases de dados. As classes de 0 a 5 representam, respectivamente, os níveis de RD ausente, leve, moderado, severo e proliferativo.	39
FIGURA 11 – Artefatos em imagens da base de dados EyePACS/Kaggle.	40
FIGURA 12 – Aumento de dados aplicada à imagem original mostrada na Figura 9. Da esquerda para a direita e de cima para baixo: rotação em 90°, 180° e 270°, espelhamento vertical e horizontal, e modificação de contraste com CLAHE.	40

FIGURA 13 – Distribuição balanceada entre as classes das bases de dados.	41
FIGURA 14 – Exemplo da geração de sub-imagens com dimensões 256×256 pixels. Imagem IDRiD_01 da base IDRiD.	43
FIGURA 15 – Exemplos de sub-imagens RGB removidas do conjunto de dados. Imagem original 007-1774-100-0-3 da base DDR.	44
FIGURA 16 – Exemplo da segmentação obtida para 4 tipos de lesões de RDs. Coluna da esquerda: sub-imagens originais, de cima para baixo, EXs duros, HMs, MAs e EXs moles. Coluna central: respectivas máscaras (<i>ground truths</i>). Coluna da direita: imagens segmentadas pela U-Net.	49

LISTA DE TABELAS

TABELA 1 – Critérios baseados nas diretrizes do ICDR (<i>International Clinical Diabetic Retinopathy</i> , 2002)	19
TABELA 2 – Visão geral dos trabalhos relacionados à classificação do grau de Retinopatia Diabética	32
TABELA 3 – Visão geral dos trabalhos relacionados à segmentação de lesões	33
TABELA 4 – Comparação entre as bases de dados	36
TABELA 5 – Hiperparâmetros e dados de treinamento da rede VGG16	42
TABELA 6 – Hiperparâmetros e dados de treinamento da rede U-Net	44
TABELA 7 – Matriz de confusão genérica para 6 classes	45
TABELA 8 – Análise comparativa de resultados sobre a classificação de RDs	48
TABELA 9 – Análise comparativa de resultados sobre a segmentação de MAs	50
TABELA 10 – Análise comparativa de resultados sobre a segmentação de HMs	50
TABELA 11 – Análise comparativa de resultados sobre a segmentação de EXs moles	50
TABELA 12 – Análise comparativa de resultados sobre a segmentação de EXs duros	50

LISTA DE SIGLAS

- AUC – Área Abaixo da Curva (*Area Under Curve*)
AUPR – *Area Under Precision Recall*
CLAHE – *Contrast Limited Adaptive Histogram Equalization*
CNN – Rede Neural Convolutacional (*Convolutional Neural Network*)
CPU – Unidade Central de Processamento (*Central Processing Units*)
DColor-SIFT – *Dense Color Scale-Invariant Feature Transform*
DIARETDB1 – *Standard Diabetic Retinopathy Database Calibration level 1*
DLNN – Rede Neural de Aprendizagem Profunda (*Deep Learning Neural Network*)
DVF – *Deep Visual Features*
EMD – Edema Macular Diabético
ETDRS – *Early Treatment Diabetic Retinopathy Study*
EX – Exsudato
EyePACS – *Eye Picture Archive Communication System*
FAZ – *Foveal Avascular Zone*
FC – Camada Densamente Conectada (*Fully Connected Layer*)
FN – Falso-Negativo (*False Negative*)
FOV – Campo de Visão (*Field Of View*)
FP – Falso-Positivo (*False Positive*)
GB – *Gigabyte*
GLOH – *Gradient Location-Orientation Histogram*
GPUs – Unidade de Processamento Gráfico (*Graphic Processing Units*)
HE – Equalização de Histograma (*Histogram Equalization*)
HM – Hemorragia
ICDRS – Escala Internacional de Retinopatia Diabética Clínica (*International Clinical Diabetic Retinopathy Scale*)
IDRiD – *Indian Diabetic Retinopathy Image Dataset*
IEEE – *Institute of Electrical and Electronics Engineers*
ILSVRC – *ImageNet Large Scale Visual Recognition Challenge*
IoU – Interseção sobre União (*Intersection over Union*)
ISBI – *International Symposium on Biomedical Imaging*
MA – Microaneurisma
MESSIDOR – *Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology*
MISP – *Medical Image and Signal Processing research center*
NLMD – *NonLocal Means Denoising*
NV – Neovascularização
OMS – Organização Mundial da Saúde
PCA – Análise de Componente Principal (*Principal Component Analysis*)
PPV – Precisão (*Positive Predictive Value*)
RAM – *Random Memory Access*
RBE – *Research on Biomedical Engineering*
RGB – Vermelho, Verde e Azul (*Red, Green and Blue*)
RD – Retinopatia Diabética
ReLU – *Rectified Linear Unit*
RDNP – Retinopatia Diabética Não-Proliferativa

RDP – Retinopatia Diabética Proliferativa
Se – Sensitividade (*Sensitivity*)
SGD – Gradiente Descendente Estocático (*Stochastic Gradient Descent*)
Sp – Especificidade (*Specificity*)
SVM – Máquina de Vetores de Suporte(*Support Vector Machine*)
TN – Verdadeiro-Negativo (*True Negative*)
TP – Verdadeiro-Positivo (*True Positive*)
VGG – *Visual Geometry Group*

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Contexto	11
1.2 Problema motivador	11
1.3 Justificativa	12
1.4 Objetivo, Escopo e Metas	13
1.4.1 <i>Escopo</i>	14
1.4.2 <i>Metas</i>	14
1.5 Contribuições da Pesquisa	14
1.6 Estrutura do Trabalho	15
2 REFERENCIAL TEÓRICO	17
2.1 Caracterização da Retinopatia Diabética e Diretrizes Internacionais	17
2.2 Rede Neural Convolutacional	18
2.2.1 <i>Camada de Entrada</i>	20
2.2.2 <i>Camada de Convolação</i>	20
2.2.3 <i>Função de Ativação</i>	21
2.2.4 <i>Camada de Pooling</i>	21
2.2.5 <i>Camada Densamente Conectada</i>	21
2.2.6 <i>Transferência de Aprendizagem</i>	21
2.3 Aumentação de Dados	22
2.4 Rede Neural VGG16 para classificação	23
2.5 Rede Neural U-Net para segmentação	24
2.6 Conclusão	24
3 TRABALHOS RELACIONADOS	27
3.1 Conclusão	33
4 METODOLOGIA E PROPOSTA DE DESENVOLVIMENTO	35
4.1 Bases de dados	36
4.1.1 <i>DDR</i>	36
4.1.2 <i>EyePACS/Kaggle</i>	37
4.1.3 <i>IDRiD</i>	37
4.2 Ambiente de desenvolvimento: <i>Google Colaboratory</i>	37
4.3 Pré-processamento para classificação do grau de RD	38
4.3.1 <i>Adequação das dimensões das imagens</i>	38
4.3.2 <i>Limpeza (Data Cleaning) e Aumentação de Dados (Data Augmentation)</i>	38
4.3.3 <i>Normalização dos dados</i>	42
4.4 Treinamento da rede VGG16	42
4.5 Pré-processamento para segmentação das lesões	42
4.6 Segmentação das lesões com U-Net	44
4.7 Métricas de Desempenho	44
4.8 Conclusão	46
5 RESULTADOS E DISCUSSÕES	47
5.1 Classificação do grau de RD	47
5.2 Segmentação das lesões	47
5.3 Conclusão	49
6 CONSIDERAÇÕES FINAIS E PROPOSTAS DE CONTINUIDADE	53
6.1 Considerações Finais	53
6.2 Propostas de Continuidade	55
REFERÊNCIAS	57

1 INTRODUÇÃO

Neste capítulo serão apresentados o contexto, o problema motivador, justificativa, objetivo e metas que nortearam esta pesquisa, além das contribuições já agregadas ao trabalho proposto.

1.1 Contexto

O aumento da expectativa de vida e envelhecimento da população mundial é um dos principais fatores que contribuem para o aumento do número de casos de diabetes, dentre os quais, os adultos mais idosos constituem o segmento onde pode-se observar o aparecimento e evolução mais rápidos da população com diabetes (KALYANI; GOLDEN; CEFALU, 2017). Sendo assim, ações que possam contribuir para os tratamentos e cuidados médicos nesse contexto devem se tornar mais relevantes nos próximos anos.

De acordo com dados divulgados pela Organização Mundial da Saúde (OMS) em 2014, estima-se que 422 milhões de adultos são acometidos pelo diabetes *mellitus* e é esperado que a população diabética entre 18 e 99 anos aumente para 693 milhões até 2045 (CHO et al., 2017), sendo a retinopatia diabética (RD) uma das principais complicações dessa doença (WORLD HEALTH ORGANIZATION, 2014).

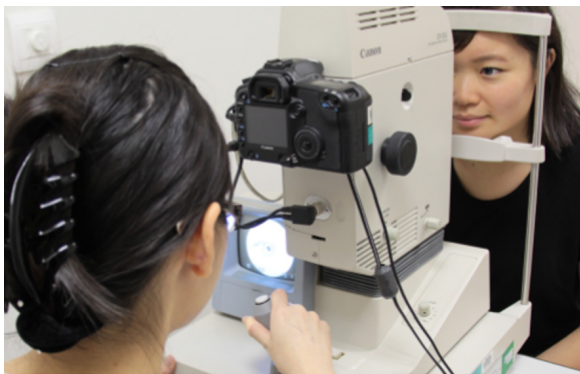
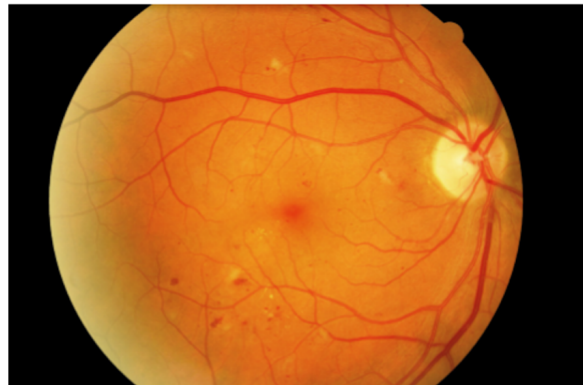
As fases ou graus de desenvolvimento, da retinopatia diabética não-proliferativa (RDNP) para a retinopatia diabética proliferativa (RDP), são caracterizadas pela presença de microaneurismas (MAs), hemorragias (HMs), exsudatos (EXs) e neovascularização (NVs), sendo essa última presente apenas em casos de RDP. Os casos de RDNP são classificados em 3 estágios: leve, moderado e severo. Portanto, é necessário que o paciente faça um acompanhamento regular da evolução do quadro clínico para manter a saúde visual e evitar danos irreversíveis que possam levar à cegueira (INTERNATIONAL COUNCIL OF OPHTHALMOLOGY, 2017).

Nesse contexto, uma abordagem que possa ajudar na determinação do grau de RD e na segmentação das imagens das lesões da retina mostra-se uma possibilidade relevante, visto que o grupo populacional acometido por essa doença está em expansão.

1.2 Problema motivador

Tradicionalmente, as imagens obtidas por meio dos exames de fundoscopia são obtidas através de uma câmera óptica, um processo não invasivo que permite visualizar a superfície interna traseira do globo ocular através da pupila do olho, conforme mostra a Figura 1. Dessa forma, as fotografias são coletadas e disponibilizadas para análise e marcação manual por especialistas, de forma a localizar e destacar estruturas como a mácula, fóvea, disco óptico, vasos sanguíneos e a retina, bem como as lesões presentes.

Como essa coleta é um processo manual e a análise é realizada por médicos oftalmologistas especializados em retina, a automatização da análise preliminar das imagens

Figura 1 – Exame de fundoscopia com câmera digital**(a) Exame****(b) Imagem obtida**

Fonte: INTERNATIONAL COUNCIL OF OPHTHALMOLOGY (2017).

de retina com foco em RD pode auxiliar na redução do tempo gasto para triagem dos pacientes de acordo com o respectivo grau, permitindo redução de custos (SCOTLAND et al., 2010), ampliando a cobertura e acompanhamento da saúde visual da população diabética, visto que 79% desta população vive em países de renda baixa ou média (CHO et al., 2017).

1.3 Justificativa

A retinopatia diabética é objeto de diversas pesquisas devido ao impacto que representa quanto à saúde visual dos pacientes. Diversos métodos de detecção automáticos para as anomalias típicas da retina vêm sendo propostos, bem como a divulgação de bases de dados públicas, que são de fundamental importância para desenvolvimento de novos métodos e avaliações de parâmetros de desempenho.

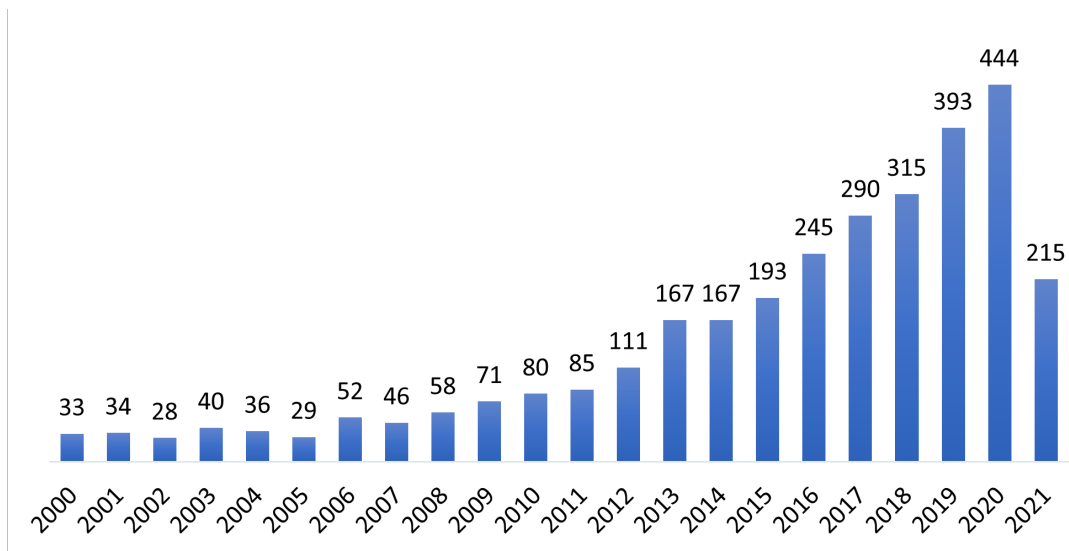
Além disso, a qualidade das imagens obtidas nos exames de fundoscopia, a habilidade dos operadores das câmeras ópticas e as manifestações diversas de retinopatias podem fazer com que a análise das imagens de fundoscopia e sua marcação manual constituam uma tarefa complexa. As principais limitações ocorrem devido ao baixo contraste, iluminação limitada e não-uniforme, nível de ruído presente nas imagens, e a presença de lesões e anomalias em tamanhos e formatos irregulares, o que dificulta o diagnóstico rápido e limita a capacidade de intervenção médica (ROCHA et al., 2020; WANG et al., 2019; KAR; MAITY, 2018; ABRÀMOFF; GARVIN; SONKA, 2010; MOOKIAH et al., 2013).

Os métodos automáticos se mostram interessantes à medida em que é possível ampliar a capacidade de processamento e análise de imagens em comparação à análise manual. Adicionalmente, há a possibilidade de se trabalhar com sistemas computadorizados que podem ser redimensionados com base na demanda de processamento. Desse modo, recursos finitos, sobretudo o tempo dos especialistas de retina, podem ser priorizados aos casos mais urgentes, a fim de impedir quadros irreversíveis na saúde visual dos pacientes.

Na área de imagens digitais médicas, a construção deste conjunto de amostras, ou

base de dados, é particularmente difícil devido ao aparecimento atípico de doenças, privacidade do paciente, expertise médica para rotular os dados, dentre outros (SHORTEN; KHOSHGOFTAAR, 2019). Alguns trabalhos propõem métodos de detecção e divulgam para a comunidade científica bases de dados públicas, que são de fundamental importância para desenvolvimento de novos métodos e avaliações de parâmetros de desempenho. A Figura 2 mostra o resultado da busca no *Web of Science* pelo termo "*Diabetic retinopathy detection*" entre o período de 2000 a julho de 2021, sendo os trabalhos não limitados a imagens médicas. Verifica-se o crescente interesse pela área, demonstrando a relevância do tema e o desenvolvimento de novas pesquisas.

Figura 2 – Resultado da busca por "*Diabetic retinopathy detection*" no *Web of Science*.



Fonte: Elaborado pelo autor.

1.4 Objetivo, Escopo e Metas

O objetivo principal deste trabalho é apresentar uma metodologia para classificar e segmentar as imagens de retina quanto ao nível de retinopatia diabética e lesões encontradas. Para isso, foram usadas imagens de retinas disponibilizadas nos bancos de dados públicos DDR (LI et al., 2019), EyePACS/Kaggle (*Eye Picture Archive Communication System*) (CUADROS; BRESNICK, 2009) e IDRiD (*Indian Diabetic Retinopathy Image Dataset*) (PORWAL et al., 2020), visando à identificação automática de anormalidades intrarretinianas decorrentes do diabetes *mellitus*, independentemente de intervenções manuais, para apoio ao diagnóstico médico.

Na metodologia proposta, tanto a abordagem de classificação quanto de segmentação foram baseadas em aplicações com redes neurais convolucionais, respectivamente, a rede VGG16 (*Visual Geometry Group*) e a U-Net. As etapas de pré-processamento das imagens de entrada incluem limpeza dos dados, aumento de dados e balanceamento das classes. Ações como a detecção da qualidade visual e segmentação em imagens particionadas

também foram desenvolvidas neste trabalho.

1.4.1 *Escopo*

Não faz parte do escopo deste trabalho a utilização de testes clínicos bem como testes em humanos. Dentre as retinopatias analisadas não será realizada a detecção de Edema Macular Diabético (EMD), pela indisponibilidade de informação nas bases de dados consultadas. As imagens utilizadas foram obtidas a partir de bancos de dados disponíveis publicamente. Os métodos investigados deverão atender à capacidade de processamento computacional necessário para a execução do treinamento e testes dos modelos.

O treinamento e testes das redes neurais foram realizados em *Python*, com a aplicação do *Jupyter Notebook*, no ambiente *Google Colab*, disponível em <https://colab.research.google.com/>.

1.4.2 *Metas*

Para alcançar o objetivo final, as seguintes metas foram contempladas:

- a) Capacidade de detecção da qualidade visual da imagem, informando aquelas que são inadequadas, considerando os artefatos apresentados por Riaz et al. (2020);
- b) Classificação das imagens de retina, considerando as classes apresentadas segundo as diretrizes do ICDR (*International Clinical Diabetic Retinopathy Scale*);
- c) Utilização de métricas de desempenho para a comparação dos resultados obtidos e apresentados em trabalhos, dentre elas, sensibilidade, especificidade, precisão e *F1-score*;
- d) Segmentação de 4 tipos de lesões (microaneurismas, hemorragias, exsudatos moles e exsudatos duros), para as bases de dados DDR e IDRiD, que apresentam as *ground truths*, ou máscaras, relativas a essas lesões;
- e) Robustez quanto às mudanças de escala, rotação e iluminação, considerando que a imagem não possui artefatos.

1.5 Contribuições da Pesquisa

As principais contribuições já alcançadas durante a pesquisa realizada são destacadas a seguir:

- a) Para a tarefa de classificação, melhoria das métricas de desempenho para a base DDR e inclusão das métricas *F1-score* para a base EyePACS/Kaggle, bem como as métricas precisão, sensibilidade e especificidade para as bases de dados DDR e IDRiD;

- b) Separação das imagens com artefatos nos bancos de dados empregados, incluindo-as em uma nova classe para a detecção de qualidade da imagem na tarefa de classificação;
- c) Para a tarefa de segmentação, melhoria das métricas de desempenho considerando as bases de dados DDR e IDRiD;
- d) Segmentação das lesões baseada em partição (*tiles*) de imagens;
- e) Submissão do artigo "*Diabetic Retinopathy Classification using VGG16 Neural Network*", na revista RBE - *Research on Biomedical Engineering*, publicada pela Springer (Fevereiro/2021).

1.6 Estrutura do Trabalho

Além do presente capítulo, este trabalho apresenta outros 5 capítulos. O capítulo 2 apresenta uma revisão das principais técnicas utilizadas. O capítulo 3 apresenta os principais trabalhos relacionados a esta pesquisa. O capítulo 4 descreve a metodologia adotada e o desenvolvimento do trabalho realizado para o processamento das imagens quanto à classificação do grau de retinopatia diabética e segmentação de lesões. O capítulo 5 mostra os resultados obtidos e discussões acerca da metodologia proposta. Por fim, o capítulo 6 apresenta as considerações finais e propostas de continuidade para a pesquisa desenvolvida.

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados alguns conceitos que foram considerados fundamentais à compreensão da caracterização das lesões de retinas, padronização do grau de retinopatia diabética, além das estruturas das redes neurais utilizadas para a classificação e segmentação de imagens de retinas obtidas a partir de bases de dados específicas.

2.1 Caracterização da Retinopatia Diabética e Diretrizes Internacionais

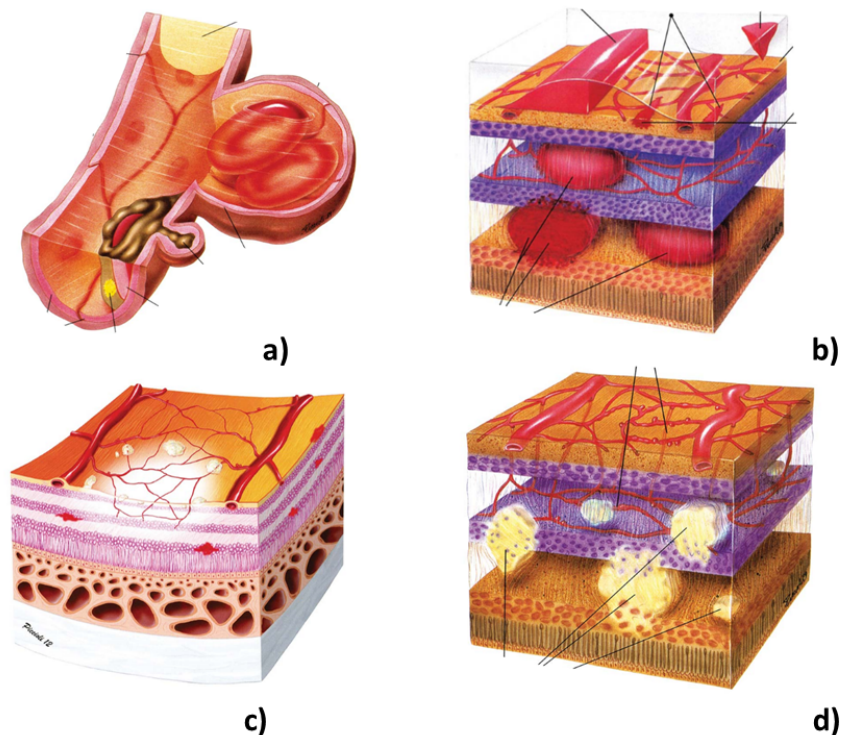
Os sinais clássicos da retinopatia diabética incluem microaneurismas, hemorragias, exsudatos duros e exsudatos moles, dilatação e anormalidades microvasculares intrarretinianas, dentre elas a neovascularização, presente em estágios avançados de RD (SALAMAT; MISSEN; RASHID, 2019; INTERNATIONAL COUNCIL OF OPHTHALMOLOGY, 2017; ALGHADYAN, 2011).

Microaneurismas (MAs) estão dentre os primeiros sinais clinicamente visíveis de retinopatia diabética, ocorrendo devido a oclusões capilares de vasos sanguíneos da retina. Eles aparecem como pequenos pontos (balões) vermelhos e, comumente, em formato circular. No entanto, não são características permanentes e podem desaparecer. Hemorragias (HMs) ocorrem devido ao vazamento de sangue quando as paredes dos vasos sanguíneos ou MAs são rompidas. Elas também podem aparecer como pequenos pontos vermelhos brilhantes (*dot*), podendo nessas circunstâncias serem confundidos com MAs, além de manchas (*blot*) avermelhadas irregulares e em formato de chama de vela (*flame*). Exsudatos (EXs) moles ou manchas algodinosas estão relacionados a acúmulos de detritos axoplasmáticos em regiões isquêmicas na camada de fibras nervosas da retina, aparecendo como manchas brancas superficiais. Exsudatos (EXs) duros são depósitos de lipídeos e proteínas que vazam através de vasos sanguíneos anormais, sendo observados como pontos amarelados. As lesões e vazamentos de fluidos podem causar cegueira total se localizados sobre a mácula (SALAMAT; MISSEN; RASHID, 2019; ALGHADYAN, 2011). MAs e HMs são frequentemente referidas como lesões avermelhadas ou escuras e EXs como lesões brilhantes.

Para ilustrar, mais detalhadamente, como cada uma das 4 lesões decorrentes da RD manifestam-se nas camadas da retina, a Figura 3 apresenta algumas vistas em corte. A Figura 3a) ilustra a presença de um microaneurisma na parede de um vaso sanguíneo, 3b) ilustra hemorragias em camadas da retina, caracterizadas pelas bolsas de sangue, 3c) mostra um exsudato mole na região superficial com aspecto branco e, em 3d), exsudatos duros em níveis abaixo da camada superficial.

Há diretrizes específicas que sugerem critérios de avaliação do grau de RD associado às manifestações presentes no exame de fundo de olho, sendo possível citar as escalas ETDRS (*Early Treatment Diabetic Retinopathy Study*) (ETDRS, 1991) e ICDRS (*International Clinical Diabetic Retinopathy Scale*) (INTERNATIONAL COUNCIL OF OPHTHALMO-

Figura 3 – Manifestações de RD em uma imagem de retina, em corte. a) MA formado na parede de um vaso sanguíneo; b) HMs, que são vazamentos de sangue dos vasos em camadas da retina; c) EXs moles, formado pelo depósito de detritos axoplasmáticos em regiões isquêmicas, gerando manchas brancas superficiais; d) EXs duros, formados por depósitos de lipídeos e proteínas que vazam através de vasos sanguíneos anormais.



Fonte: Adaptado de Lumbroso, Rispoli e Savastano (2015).

LOGY, 2017; ABRÀMOFF et al., 2013). Para evidenciar os procedimentos de tratamento para cada caso de retinopatia diabética, dentre outras ações que envolvem os trabalhos dos médicos oftalmologistas, o International Council of Ophtalmology (2017) disponibiliza um documento de diretrizes de cuidados com os olhos para diabéticos. A Tabela 1 exibe um resumo da padronização dos níveis de retinopatia diabética, considerando quadros sem RD aparente, a evolução da RDNP (Retinopatia Diabética Não-Proliferativa) e a RDP (Retinopatia Diabética Proliferativa), de acordo com resultados observáveis nas imagens de fundo. Para um detalhamento sobre cada caso, também são disponibilizados diversos anexos que descrevem e ilustram com imagens reais as manifestações associadas, conforme mostra a Figura 4.

2.2 Rede Neural Convolutacional

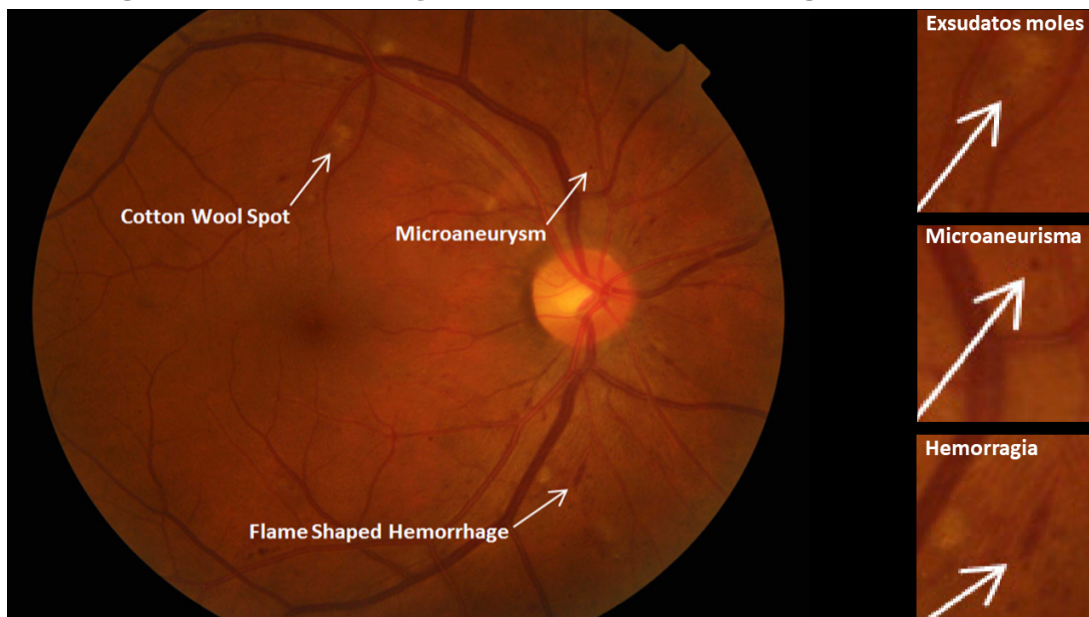
Nos anos recentes, algoritmos de *Deep Learning* estão ganhando destaque em aplicações nos mais diversos ramos da sociedade, alavancado pela evolução de processamento e barateamento de dispositivos computadorizados, crescimento na disponibilidade de dados (*Big Data*) e pela capacidade de encontrar relações para resolver problemas complexos

Tabela 1 – Critérios baseados nas diretrizes do ICDR (*International Clinical Diabetic Retinopathy, 2002*)

Retinopatia Diabética	Resultados Observáveis
Sem RD aparente	Sem anormalidades.
RDNP leve	Somente microaneurismas.
RDNP moderada	Presença de microaneurismas em grau intermediário
RDNP severa	Sintomas indicativos: <ul style="list-style-type: none"> • Hemorragia intrarretiniana (≥ 20 em cada quadrante); • Perolização venosa definitiva (em 2 quadrantes); • Anormalidade microvasculares intrarretinianas (em 1 quadrante); • Sem sinais de retinopatia proliferativa.
RD proliferativa	Sintomas indicativos: <ul style="list-style-type: none"> • Neovascularização • Hemorragia vítrea /pré-retiniana

Fonte: Adaptado de INTERNATIONAL COUNCIL OF OPHTHALMOLOGY (2017).

Figura 4 – Manifestações de RD em uma imagem de retina

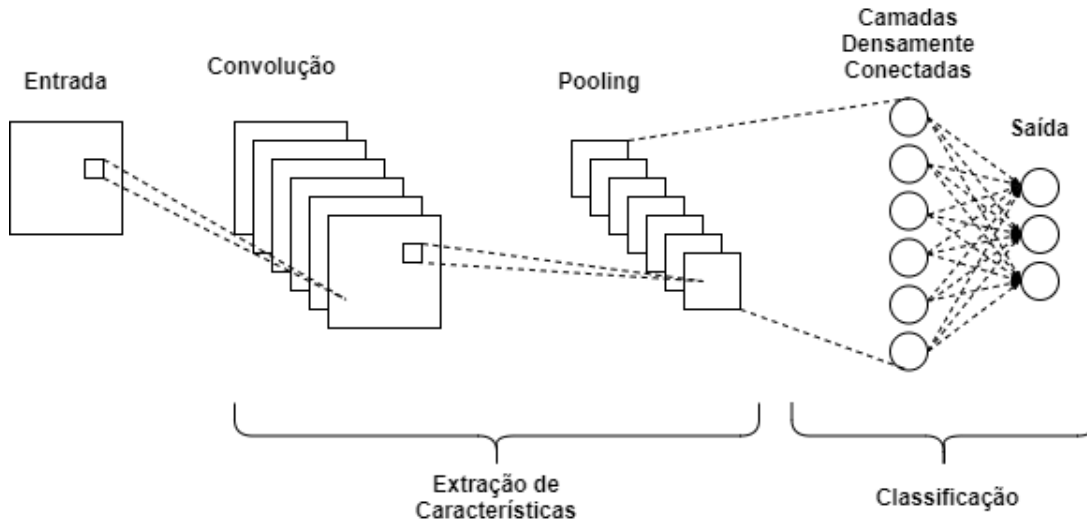


Fonte: Adaptado de INTERNATIONAL COUNCIL OF OPHTHALMOLOGY (2017).

(DARGAN et al., 2020; GUO et al., 2016). Dentre as arquiteturas disponíveis, as redes neurais convolucionais (CNN - *Convolutional Neural Network*) são as mais utilizadas no contexto de análise de imagens médicas (ALYOUBI; SHALASH; ABULKHAIR, 2020).

As camadas básicas que compõem uma rede neural convolucional podem ser separadas em duas tarefas bem definidas, a extração de características e a classificação, conforme ilustra a Figura 5. Nas subseções 2.4 e 2.5 são apresentadas as estruturas das redes neurais VGG16 e U-Net a serem aplicadas neste trabalho.

Figura 5 – Camadas básicas de uma rede neural convolucional.



Fonte: Elaborado pelo autor.

2.2.1 Camada de Entrada

A camada de entrada recebe as informações a serem utilizadas no treinamento e teste da rede neural. Em se tratando de processamento digital de imagens, ela é definida segundo altura e largura em número de pixels e a quantidade de canais, como por exemplo, no caso de imagens RGB, 3 dimensões).

2.2.2 Camada de Convolução

A camada de convolução é responsável por extrair as características de interesse de uma imagem de entrada utilizando filtros ou *kernels* que deslizam sobre ela. As características resultantes da operação de convolução são chamadas mapas de ativação ou mapas de características. Como a camada de convolução é capaz de aprender padrões locais, os padrões aprendidos são invariantes translacionais, ou seja, podem ser determinados em qualquer posição da imagem (portanto, com maior poder de generalização) e também podem ser aprendidos de maneira hierárquica, por exemplo, a primeira camada poderá aprender a detectar as bordas enquanto as próximas camadas podem aprender outros padrões baseados nas camadas anteriores) (CHOLLET, 2017).

A operação de convolução 2D pode ser descrita matematicamente (GOODFELLOW; BENGIO; COURVILLE, 2016) como:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (1)$$

onde S é o mapa de características, i e j são as coordenadas da imagem, I é a imagem de entrada, K é o *kernel* bidimensional que desliza sobre a imagem de entrada e, m e n são as coordenadas do *kernel*.

2.2.3 Função de Ativação

Após computar a soma das entradas já submetidas aos seus respectivos pesos e *bias* de um neurônio, o sinal resultante é submetido a uma função de ativação que, basicamente, definirá o valor da saída desse neurônio segundo uma função matemática pré-escolhida. Uma rede que utiliza funções de ativação lineares pode se tornar limitada em relação à resolução de problemas complexos, pois a saída será uma combinação linear das entradas dos neurônios. Para melhorar a abrangência de problemas, funções de ativação não lineares são utilizadas pois elas permitem aos neurônios aprender estruturas mais complexas dos dados que serão submetidos à rede neural.

O tipo de função de ativação a ser escolhido depende da posição na estrutura da rede neural em que ela será inserida e do tipo de ação que a transformação deverá executar. Quando colocada após as camadas escondidas ou após a etapa de convolução, a função de ativação deverá converter os mapas de características lineares em formas não-lineares (por exemplo, ReLU - *Rectified Linear Unit*), enquanto, se colocada na camada de saída, poderá produzir previsões (por exemplo, sigmoide ou softmax) (NWANKPA et al., 2018).

2.2.4 Camada de Pooling

Esta camada é responsável por agrupar os mapas de características, gerando uma representação compacta aproximada da vizinhança dos pixels sob análise, tornando-a invariante a pequenas translações da entrada (GOODFELLOW; BENGIO; COURVILLE, 2016), reduzindo o custo computacional. O parâmetro que controla a operação de subamostragem é chamado de *stride* e se refere ao deslocamento feito pelo *kernel* nas duas direções da imagem. As funções de *pooling* mais populares incluem *pool* máximo (*max-pooling*), *pool* mínimo (*min-pooling*) e *pool* de média (*average pooling*).

2.2.5 Camada Densamente Conectada

Conhecida como FC (*Fully Connected Layer*), essa camada é comumente responsável por efetuar a etapa de classificação da rede através dos vetores de características obtidas pelas camadas de convolução e *pooling*. Em uma camada densamente conectada, a saída de um neurônio, obtida a partir da soma ponderada de todas as saídas dos neurônios da camada anterior, é conectada à entrada de todos os neurônios da camada seguinte (SZE et al., 2017).

2.2.6 Transferência de Aprendizagem

A transferência de aprendizagem busca reutilizar um conhecimento já obtido em uma tarefa anterior para resolver ou acelerar soluções em outras tarefas similares. É um item importante no contexto de *deep learning* para a obtenção de modelos que resolvam problemas complexos que demandem um grande volume de dados para o treinamento e a

utilização de *labels* (TAJBAKHSI et al., 2016). Este procedimento de transferência é normalmente executado reaproveitando uma rede neural pré-treinada em um domínio similar, substituindo a(s) última(s) camada(s) conforme a dimensão de saída dos novos dados do problema a ser solucionado, e retreinando a rede (*tuning*) tipicamente com uma taxa de aprendizado mais baixa que a normal.

Neste trabalho, foi utilizada a transferência de aprendizagem para aproveitar o conhecimento obtido de um modelo com 1000 classes na base de dados *ImageNet* (DENG et al., 2009) visando acelerar o processo de treinamento da rede neural VGG16 (SIMONYAN; ZISSERMAN, 2014) responsável por classificar as imagens de retina em 6 classes distintas.

2.3 Aumentação de Dados

Um dos requisitos para treinamento de redes neurais profundas é a disponibilidade de um conjunto de amostras em grande quantidade e presença de rótulos, ou *labels* (quando utilizado abordagem supervisionada), para que o modelo obtido seja capaz de extrair as informações necessárias das classes do problema para solucioná-lo. Na área de imagens digitais médicas, a construção deste conjunto de amostras, ou base de dados, é particularmente difícil devido ao aparecimento atípico de doenças, privacidade do paciente, expertise médica para rotular os dados, dentre outros (SHORTEN; KHOSHGOFTAAR, 2019). Para expandir o conjunto de amostras a serem utilizadas no treinamento, utiliza-se a aumento de dados.

Segundo Shorten e Khoshgoftaar (2019), algumas das aplicações de aumento de dados baseadas em manipulações básicas de imagem são:

- Espelhamento (*flipping*), que pode ser horizontal (mais comum) e vertical;
- Alterações no espaço de cores, seja em um dos canais de cor, brilho ou manipulação do histograma;
- Corte, que extrai fragmentos (*patches*) da imagem, reduzindo seu tamanho;
- Rotação, que pode ser feita para a esquerda ou direita, no intervalo de 1° a 359° ;
- Translação, que desloca a imagem para quaisquer direções escolhidas;
- Injeção de ruído, usualmente através de uma distribuição Gaussiana.

Considerando a possibilidade de promover alterações no espaço de cores, pode-se citar a aplicação do CLAHE (*Contrast Limited Adaptive Histogram Equalization*) (ZUIDERVELD, 1994). Nesse método, o número de pixels em cada nível de cinza que ultrapassa um limite arbitrado de intensidade é retirado e redistribuído recursivamente. Na primeira etapa, a imagem é dividida em regiões e, posteriormente, é realizado o cálculo do histograma de cada região. Após avaliação dos níveis que

ultrapassam um limiar dos histogramas, é realizada a distribuição recursiva das intensidades excedidas. Ao término da distribuição, é realizado o cálculo da função de mapeamento e, finalmente, a interpolação e o mapeamento (PIZER et al., 1990).

2.4 Rede Neural VGG16 para classificação

VGGNet (*Visual Geometry Group*), ou simplesmente VGG, é uma rede neural convolucional desenvolvida por pesquisadores do *Visual Geometry Group* da Universidade de Oxford para a competição *ImageNet Large-Scale Visual Recognition Challenge* (ILSVR), em 2014 (SIMONYAN; ZISSERMAN, 2014), alcançando o segundo lugar (WAN; LIANG; ZHANG, 2018).

Esta CNN profunda foi construída considerando como entrada imagens RGB de 224×224 pixels e classificando-as em 1000 categorias. Sua estrutura é constituída por 16 a 19 camadas, associadas a pesos, utilizando filtros (ou *kernels*) de convolução sequenciais com dimensões 3×3 e filtros 2×2 de *max-pooling*, aplicados para melhorar o desempenho, intercalados entre as camadas da estrutura até alcançar as camadas densamente conectadas, seguidas de uma camada com ativação *Softmax*, que exhibe a distribuição de probabilidades para N classes.

A exemplo de outros artigos (LI et al., 2019; WAN; LIANG; ZHANG, 2018) que avaliaram o desempenho da rede VGG para a tarefa de classificação do grau de RD, a estrutura VGG16 foi escolhida em razão do desempenho de classificação reportado e pela característica de transferência de aprendizagem, utilizando os pesos definidos para a competição *ImageNet* (DENG et al., 2009) e redimensionando a quantidade de classes para 6.

A Figura 6 mostra a estrutura da rede VGG16 utilizada neste trabalho. Ela é composta por 13 camadas de convolução, 5 camadas de *max-pooling* e 3 camadas densamente conectadas.

O bloco conv1 possui 2 camadas de convolução em sequência, sendo que os mapas de características extraídos possuem o mesmo tamanho das imagens de entrada e dimensão 64, que corresponde ao número de *kernels* utilizados em cada camada de convolução. A camada seguinte, que é de *max-pooling*, reduz o tamanho do mapa de características na proporção de $1/4$.

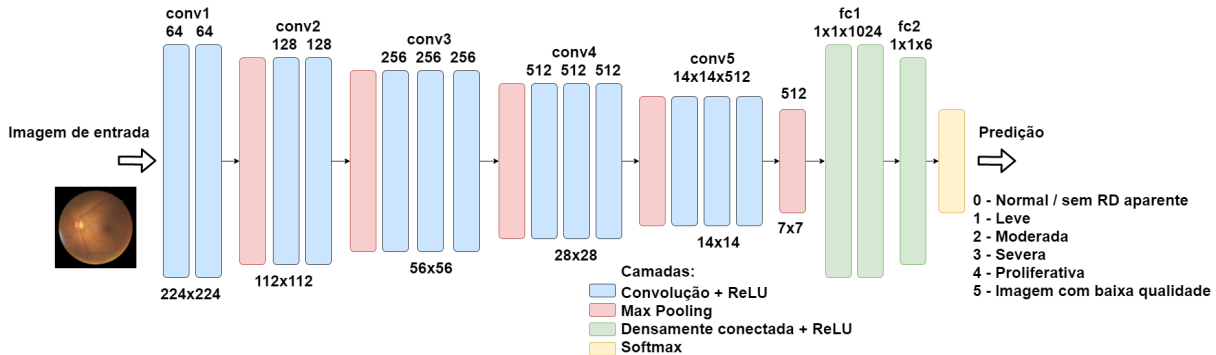
O bloco conv2 também possui 2 camadas de convolução em sequência, sendo que os mapas de características extraídos possuem o mesmo tamanho da camada de *max-pooling* anterior e dimensão 128. Novamente é utilizada uma camada de *max-pooling* para nova redução de $1/4$ dos mapas de características.

Os blocos conv3, conv4 e conv5 possuem 3 camadas de convolução em sequência para cada bloco, com 256, 512 e 512 filtros, respectivamente. Intercalando esses blocos, são utilizadas camadas de *max-pooling*.

Ao final, 3 camadas densamente conectadas são responsáveis por efetuar a classificação

da rede a partir dos vetores de características obtidos pelas camadas anteriores em 1000 categorias.

Figura 6 – Estrutura da rede VGG16.



Fonte: Elaborado pelo autor.

2.5 Rede Neural U-Net para segmentação

A rede neural U-Net desenvolvida por Ronneberger, Fischer e Brox (2015) é uma rede com estrutura composta por camadas convolucionais, *pooling* e convoluções transpostas. Em 2015, essa rede venceu a competição de rastreamento de células do ISBI (*International Symposium on Biomedical Imaging*).

Conforme apresentado na Figura 7, a arquitetura da rede pode ser dividida em duas partes principais, o caminho de contração (ou codificador) e o caminho de expansão (ou decodificador).

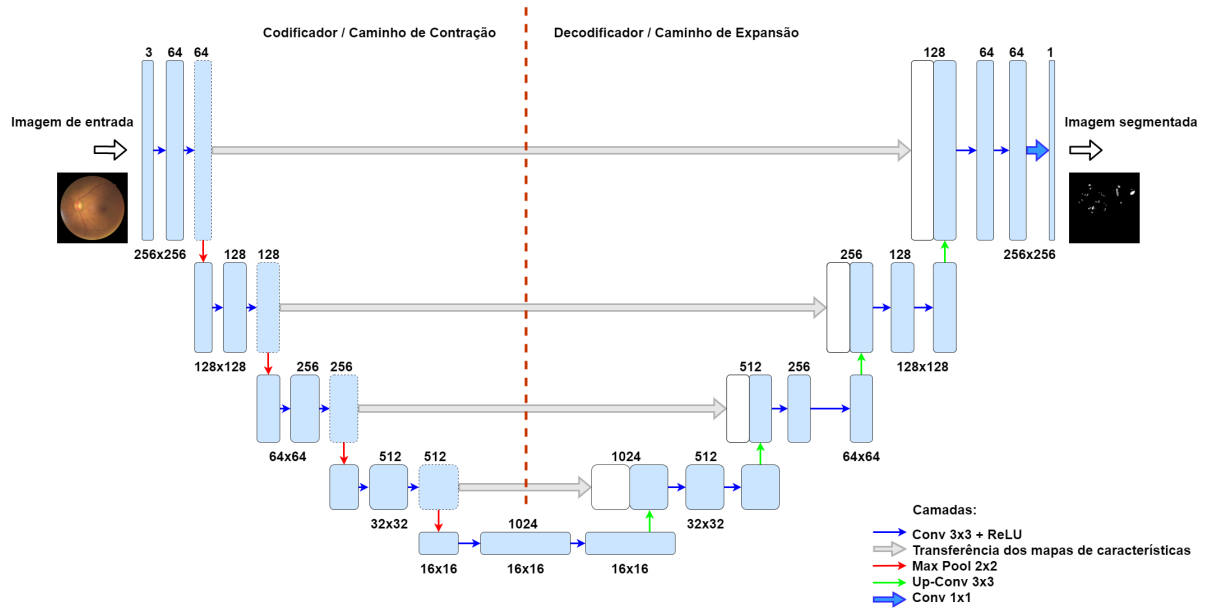
- O caminho de contração consiste de aplicações sequenciais de duas camadas de convolução com $kernel\ 3 \times 3$, uma função de ativação ReLU e uma operação de *max-pooling* para subamostragem *downsampling*. Os mapas de características extraídos utilizados em *skip connections*, que saltam algumas das camadas da rede neural, sendo agrupado na saída de uma camada como entrada para as próximas camadas;
- O caminho de expansão consiste na superamostragem (*upsampling*) dos mapas de características, através de uma convolução transposta (DUMOULIN; VISIN, 2018), seguida de convolução 2×2 que reduz o número dos mapas de características pela metade e a concatenação dos mapas extraídos do caminho de contração e duas convoluções 3×3 seguidas de ReLU.

A camada final possui uma convolução com kernel 1×1 que mapeia os 64 mapas de características para o número desejado de classes.

2.6 Conclusão

Os conceitos apresentados no referencial teórico trataram acerca da caracterização dos graus de RDs e das estruturas das redes neurais convolucionais, bem como sobre

Figura 7 – Estrutura da rede U-Net, destacando-se as etapas do codificador e do decodificador.



Fonte: Elaborado pelo autor.

suas vantagens e limitações. Esses conceitos constituem a base para a compreensão dos processos de classificação das RDs e segmentação das imagens das lesões, conforme a proposta deste trabalho.

3 TRABALHOS RELACIONADOS

Neste capítulo serão apresentadas as referências bibliográficas correlatas ao tema da pesquisa proposta consideradas de maior relevância, organizadas primeiramente em relação ao tipo de abordagem, técnicas e bases de dados utilizadas nos trabalhos selecionados e, em seguida, cada uma delas é individualmente discutida. Ao final do capítulo, as Tabelas 2 e 3 apresentam a visão geral dos trabalhos relacionados à classificação do grau de RD e segmentação das lesões, respectivamente.

Dentre os trabalhos relacionados, observa-se a presença significativa da utilização das redes neurais (PORWAL et al., 2020; RIAZ et al., 2020; QURESHI; MA; ABBAS, 2021; LI et al., 2019; LEEZA; FAROOQ, 2019; SHANTHI; SABEENIAN, 2019; ZHANG et al., 2019; WAN; LIANG; ZHANG, 2018; ABBAS et al., 2017; GARGEYA; LENG, 2017; QUELLEC et al., 2017; GULSHAN et al., 2016; PRATT et al., 2016) a partir do treinamento com um grande número de imagens, visando à extração de características que diferenciam cada classe de imagens (normais e com algum grau de RD). Como informado em (QUELLEC et al., 2017), como as redes neurais são estruturas caixa preta (*black boxes*), há a possibilidade de características desconhecidas ou não utilizadas por humanos serem também consideradas (GULSHAN et al., 2016). Nos trabalhos pesquisados, pôde-se observar o uso frequente da augmentação de dados no conjunto de amostras disponíveis e a utilização de estruturas de redes neurais já conhecidas, aproveitando-se de suas características de transferência de aprendizagem para acelerar o treinamento.

Em abordagens que utilizam outras técnicas e conceitos de visão computacional além das redes neurais, observa-se que as informações obtidas da retina são baseadas em descritores conhecidos como forma, coloração, contraste, formulação de modelos para imagens normais e anormais ou, ainda, características calculadas em pontos de interesse para criação de dicionários. Esses métodos possibilitam extrair informações previamente conhecidas a fim de criar vetores de características para posterior classificação (LEEZA; FAROOQ, 2019; ABBAS et al., 2017).

Em relação à classificação do grau de RD, os trabalhos utilizam 4 classes (SHANTHI; SABEENIAN, 2019; ZHANG et al., 2019) ou 5 classes (QURESHI; MA; ABBAS, 2021; RIAZ et al., 2020; LEEZA; FAROOQ, 2019; WAN; LIANG; ZHANG, 2018; ABBAS et al., 2017; GULSHAN et al., 2016; PRATT et al., 2016), sendo que um deles se destaca pela proposição de 6 classes (LI et al., 2019). Esse trabalho inclui uma categoria para informar quando as imagens possuem qualidade insuficiente e, portanto, não possibilitam a avaliação do grau de RD associado. Há diretrizes específicas que sugerem critérios de avaliação do grau de RD associado às manifestações presentes no exame de fundo de olho, e através de especialistas, as imagens são avaliadas e agrupadas de acordo com os resultados observáveis. Dentre estas diretrizes para RD, podem ser citadas as escalas ETDRS (*Early Treatment Diabetic Retinopathy Study*) (ETDRS, 1991) e ICDRS (*International Clinical*

Diabetic Retinopathy Scale) (ABRAMOFF et al., 2013; INTERNATIONAL COUNCIL OF OPHTHALMOLOGY, 2017).

Dentre as bases de dados citadas nos trabalhos de classificação do grau de RD, observou-se que a MESSIDOR e a EyePACS/Kaggle são mais frequentemente utilizadas. Ambas apresentam avaliações sobre a presença de RD em cada imagem, sendo avaliadas em escalas de 0 a 3 e de 0 a 4, respectivamente. Uma visão geral dos trabalhos de classificação é mostrada na Tabela 2.

Com foco em segmentação de lesões, são apresentados os trabalhos de Li et al. (2019) e Porwal et al. (2020), que empregam redes neurais para segmentação semântica. Os trabalhos utilizam as bases de dados públicas DDR e IDRiD (*Indian Diabetic Retinopathy Image Dataset*), respectivamente, que possuem anotações a nível de imagem (classificação do grau de RD) e a nível de pixel (segmentação das lesões). Uma visão geral desses trabalhos é apresentada na Tabela 3.

O trabalho de Li et al. (2019) apresenta a base DDR e avalia algoritmos *Deep Learning* para as tarefas de classificação de imagem em 6 categorias, segmentação de 4 tipos de lesões e detecção de objetos. No pré-processamento, as imagens foram redimensionadas para 512 x 512 pixels e uma estratégia de re-amostragem foi aplicada em todas as classes do conjunto de treinamento, balanceando todas as classes. Em seguida, para reduzir o *overfitting* e melhorar o desempenho, os autores utilizam aumento de dados composta por translação, alargamento, rotação, espelhamento e aumento de cores. Na tarefa de classificação, foram utilizadas cinco estruturas de redes profundas (VGG16, ResNet18, GoogLeNet, DenseNet121 e SE-BN-Inception) com entrada de imagens RGB de dimensões 224 x 224 pixels redimensionadas a partir das imagens de 512 x 512 pixels. O trabalho destaca a dificuldade de identificar o estágio leve de RD e também que o estágio severo é facilmente classificado como moderado, ocasionando baixa acurácia para essas classes. Para o processo de segmentação das lesões há a avaliação das redes neurais HED (*Holistically-Nested Edge Detection*) e DeepLab V3+.

No trabalho de Riaz et al. (2020), os autores utilizam uma rede neural densamente conectada para a tarefa de classificação de RD em 5 níveis, aplicadas às bases de dados da EyePACS/Kaggle e MESSIDOR. Os passos de pré-processamento envolvem limpeza dos dados, mantendo-se apenas as imagens sem presença de quaisquer artefatos que reduzam a qualidade visual das informações da retina, correção de *labels* informados pela base MESSIDOR, janela de recorte com múltiplas resoluções para eliminação da máscara, aumento de dados para balanceamento das classes e redução de *overfitting* e normalização dos dados. A estrutura da rede contém 3 blocos densamente conectados sendo cada bloco composto pela combinação de 20 camadas convolucionais de dimensões 1 x 1 e 3 x 3. Após cada bloco denso é adicionado um bloco de transição, com camadas de ativação, convolução 1 x 1, *pool* de média 2 x 2 e concatenação. O desempenho do sistema proposto foi avaliado em 1747 imagens da base MESSIDOR e 17978 imagens da base EyePACS/-

Kaggle utilizando os parâmetros precisão, sensibilidade (*recall*), especificidade, F1-score e AUC (*Area Under Curve*).

O artigo de Abbas et al. (2017) propõe um sistema capaz de determinar o grau de severidade de RD utilizando características visuais profundas (DVF - *Deep Visual Features*). As técnicas DColor-SIFT (*Dense Color Scale-Invariant Feature Transform*) e GLOH (*Gradient Location-Orientation Histogram*) foram utilizadas para extração de descritores das imagens e a técnica PCA (*Principal Component Analysis*) foi aplicada para reduzir a dimensão dos vetores de descritores. Em seguida, uma rede neural de aprendizado profundo (DLNN - *Deep Learning Neural Network*) foi utilizada para classificação em 5 níveis de RD (ausente, leve, moderado, severo e proliferativo). A base de dados para validação incluiu 750 imagens (150 imagens para cada grau de RD) selecionadas por dois oftalmologistas a partir das bases de dados públicas DIARETDB1, FAZ (*Foveal Avascular Zone*) disponibilizada pelo centro de pesquisas MISP (*Medical Image and Signal Processing research center*), MESSIDOR e de uma base privada do Hospital Puerta del Mar, localizado na Espanha.

O artigo de Gargeya e Leng (2017) desenvolve um algoritmo para a detecção de RD através de uma rede neural profunda com 5 camadas convolucionais para a análise de patologias utilizando as três componentes do espaço RGB (*Red, Green, Blue*). O diagnóstico final foi obtido a partir de um classificador de *gradient boosting* responsável por determinar, em 2 classes, a ausência de RD ou presença em qualquer grau de severidade. A avaliação foi feita a partir das bases de dados MESSIDOR e E-Ophtha.

Gulshan et al. (2016) apresentam uma CNN (*Convolutional Neural Network*, ou *Conv-Net*) profunda com arquitetura Inception-v3, aplicada à detecção de retinopatia diabética e edema macular diabético. Segundo os autores, o algoritmo classifica as imagens em 5 níveis de RD (ausente, leve, moderada, severa e proliferativa), mas não explicita as lesões detectadas (MAs, HMs) e as características exatas para fazer as previsões, sendo possível que o algoritmo utilize características ignoradas por humanos. Neste estudo, a base de dados de desenvolvimento foi constituída de 128175 imagens, marcadas por oftalmologistas licenciados ou *trainees* no último ano de residência. Já para a validação foram utilizadas as bases de dados EyePACS (privada) e MESSIDOR (pública).

Na proposta de Leeza e Farooq (2019), a detecção do grau de RD é feita a partir do modelo *Bag of Features* e a base de dados Kaggle, sem incluir as etapas de pré e pós-processamento. O modelo baseia-se na criação de um dicionário de características visuais a partir dos pontos de interesse, onde são computadas as características descritivas e, em seguida, agrupadas em uma representação mais compacta. O SVM (*Support Vector Machine*) e redes neurais foram usadas para classificar a RD em 5 níveis, como ausente, leve, moderada, severa e proliferativa.

Quellec et al. (2017) propõem a generalização do método de retropropagação para o treinamento de CNNs e a criação de *heatmaps* que mostram quais pixels são significativos

para classificação do grau de RD das imagens. A abordagem foi aplicada ao banco de dados gerado para uma competição de retinopatia diabética da Kaggle, em 2015, ao banco de dados E-Ophtha para avaliação do grau de RD e ao banco de dados público DIARETDB1, para avaliação do local da lesão e do grau de RD considerando 2 classes: ausente ou leve, e moderada ou proliferativa.

Shanthi e Sabeenian (2019) utilizam uma arquitetura AlexNet modificada, com dimensões de entrada 512×512 , a fim de executar a classificação de RD em 4 categorias (Ausente, RD estágio 1 correspondente a leve, RD estágio 2 correspondente a moderado e severo, RD estágio 3 correspondente a proliferativo). A entrada da rede é alimentada com informações do canal verde, utilizando o espaço de cores RGB. O treinamento e teste da arquitetura proposta foram executados, respectivamente, com 710 e 303 imagens de fundo de olho da base de dados MESSIDOR e as métricas utilizadas para avaliação foram a sensibilidade, especificidade, acurácia e precisão.

Zhang et al. (2019) apresentam um método para a identificação de imagens de retina em 2 classes (normais ou anormais) e avaliação do grau de RD em 4 classes (ausente, leve a moderado, severo a proliferativo sem risco elevado, proliferativo com risco elevado). A base de dados utilizada é privada, fornecida pela *Sichuan Academy of Medical Sciences e Sichuan Provincial Peoples Hospital*, China, e é composta por 13767 imagens de 1872 pacientes. Para melhorar a qualidade visual das imagens é aplicada a equalização de histograma (HE - *Histogram Equalization*). A augmentação de dados foi utilizada para ampliar o tamanho do conjunto de treinamento em tempo real, além do recorte por meio de máscaras para a redução da área de informações irrelevantes. O grau de RD foi executado em 4 classes, e com a avaliação individual e a combinação (*ensemble*) de 7 modelos de redes, sendo elas: InceptionV3, InceptionResNetV2, Xception, ResNet50, DenseNet169, DenseNet201 e DenseNet121. Os autores reportam que a combinação das arquiteturas ResNet50, DenseNet169 e DenseNet201 produziram os melhores resultados entre as estruturas avaliadas, considerando as métricas de acurácia, sensibilidade, precisão, Kappa, F1-score, F_β -score e Índice de Youden.

A estrutura proposta por Pratt et al. (2016) conta com 10 camadas convolucionais e 3 camadas completamente conectadas para avaliação do grau de RD em 5 classes (ausente, leve, moderado, severo e proliferativo). Na etapa de pré-processamento as cores das imagens foram normalizadas e as dimensões foram reduzidas para 512×512 pixels. Para a etapa de treinamento foi utilizada a base de dados da Kaggle e executada uma rotina de tempo real para augmentação dos dados, além de atribuição de pesos por classe para reduzir o risco de *overfitting*. As métricas utilizadas foram sensibilidade, especificidade e acurácia.

Wan, Liang e Zhang (2018) avaliam 4 tipos de arquiteturas de redes neurais para a classificação do grau de RD em 5 categorias (ausente, leve, moderado, severo e proliferativo), sendo elas AlexNet, VGGNet, GoogLeNet e ResNet, todas com capacidade

de transferência de aprendizado e ajuste (*tuning*) de hiperparâmetros. A base de dados utilizada para treinamento foi a Kaggle, que passa pelos processos de aumento de dados, normalização e recorte de imagem para redução da área de máscara. Um filtro NLMD (*NonLocal Means Denoising*) para a remoção de ruídos também é aplicado durante o pré-processamento. No treinamento, a taxa de aprendizado da rede é configurada com variação linear entre 0,0001 a 0,1 e o método de otimização escolhido é o Gradiente Descendente Estocástico (SGD - *Stochastic Gradient Descent*). Após o ajuste dos hiperparâmetros, os autores reportam que a melhor acurácia foi obtida para a arquitetura VGGNet-s, com valor percentual de 95,68%.

Porwal et al. (2020) apresenta uma coletânea de trabalhos sobre os modelos que obtiveram melhores resultados na competição de segmentação e classificação do grau de RD no ISBI 2018 (*IEEE International Symposium on Biomedical Imaging*). Para a tarefa de segmentação foram elencados os três trabalhos que produziram os melhores *scores*. Son et al. (equipe VRT) modificaram uma U-Net de modo que as camadas de *upsampling* tivessem o mesmo número de mapas de recursos com camadas concatenadas. Também houve ajuste na quantidade de vezes que a camada de max-pooling foi aplicada, sendo 6 vezes para EXs duros e HMs, 4 vezes para EXs moles e 2 vezes MAs. A função de custo foi escolhida para penalizar falsos-negativos e aumentar a sensibilidade. A rede foi treinada com otimizador de Adam em 2 etapas, com mudança de hiperparâmetros a partir de 250 *epochs* até 300 *epochs*.

Ainda, com base na coletânea de Porwal et al. (2020), Wang et al. (equipe IFlyTEK-MIG) propõem uma nova abordagem com uma CNN em cascata baseada na U-Net composta de 3 estágios, sendo que o primeiro produz uma segmentação inicial grossa (*coarse*) a partir de *patches* de imagens centradas nas lesões, o segundo é um classificador em cascata para reduzir o número de falso-positivos treinados com as regiões candidatas à lesão do estágio anterior, e o terceiro produz a segmentação fina a partir das regiões obtidas do segundo estágio. Cada estágio foi treinado com 100, 300 e 100 *epochs*, respectivamente, e *momentum* de 0,9 com estratégia de decaimento da taxa de aprendizagem.

Também em Porwal et al. (2020), Lihong et al. (equipe PATech) apresentam um novo método de CNN, combinando a U-Net com DenseNets e blocos de dilatação para capturar mais informação de contexto da imagem e recursos multi-escalas. O modelo possui blocos de transição *downsampling*, *upsampling* e dilatação com taxas de 1, 3 e 5, respectivamente, concatenados com um bloco de convolução. Essa característica foi avaliada pelos autores como benéfica na tratativa de lesões em diferentes tamanhos, onde as menores taxas de dilatação conseguem capturar características de lesões menores, enquanto maiores taxas de dilatação focam em lesões maiores. *Patches* de imagens foram extraídas e normalizadas e aquelas com alguma lesão sobreposta aos *ground truths* foram reservadas ao treinamento e submetidas à aumento de dados. O otimizador escolhido foi Adam, com taxa de aprendizagem inicial de 2×10^{-4} , dividido por 20 a cada 20 *epochs*.

Tabela 2 – Visão geral dos trabalhos relacionados à classificação do grau de Retinopatia Diabética

Trabalho	Abordagem	Base	Classes
LI et al., 2019	Apresenta a base DDR e avalia as redes profundas VGG16, ResNet18, GoogLeNet, DenseNet121 e SE-BN-Inception.	DDR	6 (ausente a proliferativa e baixa qualidade)
WAN;LIANG;ZHANG, 2018	Avaliação das arquiteturas AlexNet, VGGNet, GoogleNet e ResNet.	Kaggle	5 (ausente a proliferativa)
LEEZA; FAROOQ, 2019	<i>Bag of Features</i> (dicionário de características visuais), SVM e ANN.	Kaggle	5 (ausente a proliferativa)
QUELLEC et al., 2017	CNN. Criação de <i>heat maps</i> que mostram quais pixels são significativos para predições a nível de imagem e de lesão.	Kaggle, E-Ophtha e DIARETDB1	2 (ausente ou leve e moderada ou proliferativa)
RIAZ et al., 2020	DenseNet, com 3 blocos densos e cada bloco com 20 camadas convolucionais. Único que faz limpeza do banco de dados.	Kaggle e Messidor-2	5 (ausente a proliferativa)
GULSHAN; et al., 2016	CNN Inception-V3. Combinação dos pixels próximos em características locais, agregando-os segundo características globais	EyePACS-1 e Messidor-2	5 (ausente a proliferativa)
SHANTHI; SABEENIAN, 2019	Arquitetura AlexNet modificada.	Messidor	4 (ausente, leve, moderada ou severa, proliferativa)
GARGEYA; LENG, 2017	Extração de características com uma CNN customizada e classificação por <i>Gradient Boosting</i> .	Messidor-2 e E-Ophtha	2 (ausência ou presença em qualquer grau)
ABBAS; et al., 2017	<i>Deep Visual Features</i> (DColor-SIFT + GLOH), e rede neural de aprendizado profundo.	Messidor, DIARETDB1, FAZ, e uma base privada	5 (ausente a proliferativa)
ZHANG et al., 2019	Avaliação e combinação de CNNs (InceptionV3, InceptionResNetV2, Xception, ResNet50, DenseNet169, DenseNet201 e DenseNet121).	Privada	4 (ausente, leve ou moderado, severo e proliferativo)

Fonte: Elaborado pelo autor.

Tabela 3 – Visão geral dos trabalhos relacionados à segmentação de lesões

Trabalho	Abordagem	Base	Número de lesões
PORWAL et al., 2020 Son et al. (VRT)	U-Net modificada	IDRiD	4
PORWAL et al., 2020 Wang et al. (IFLYTEK-MIG)	CNN cascata baseada em U-Net	IDRiD	4
PORWAL et al., 2020 Lihong et al. (PATech)	CNN baseado em <i>patch</i> , combinando DenseNets e U-Net	IDRiD	4
LI et al., 2019	Redes neurais HED (<i>Holistically-Nested Edge Detection</i>) e DeepLab V3+	DDR	4

Fonte: Elaborado pelo autor.

3.1 Conclusão

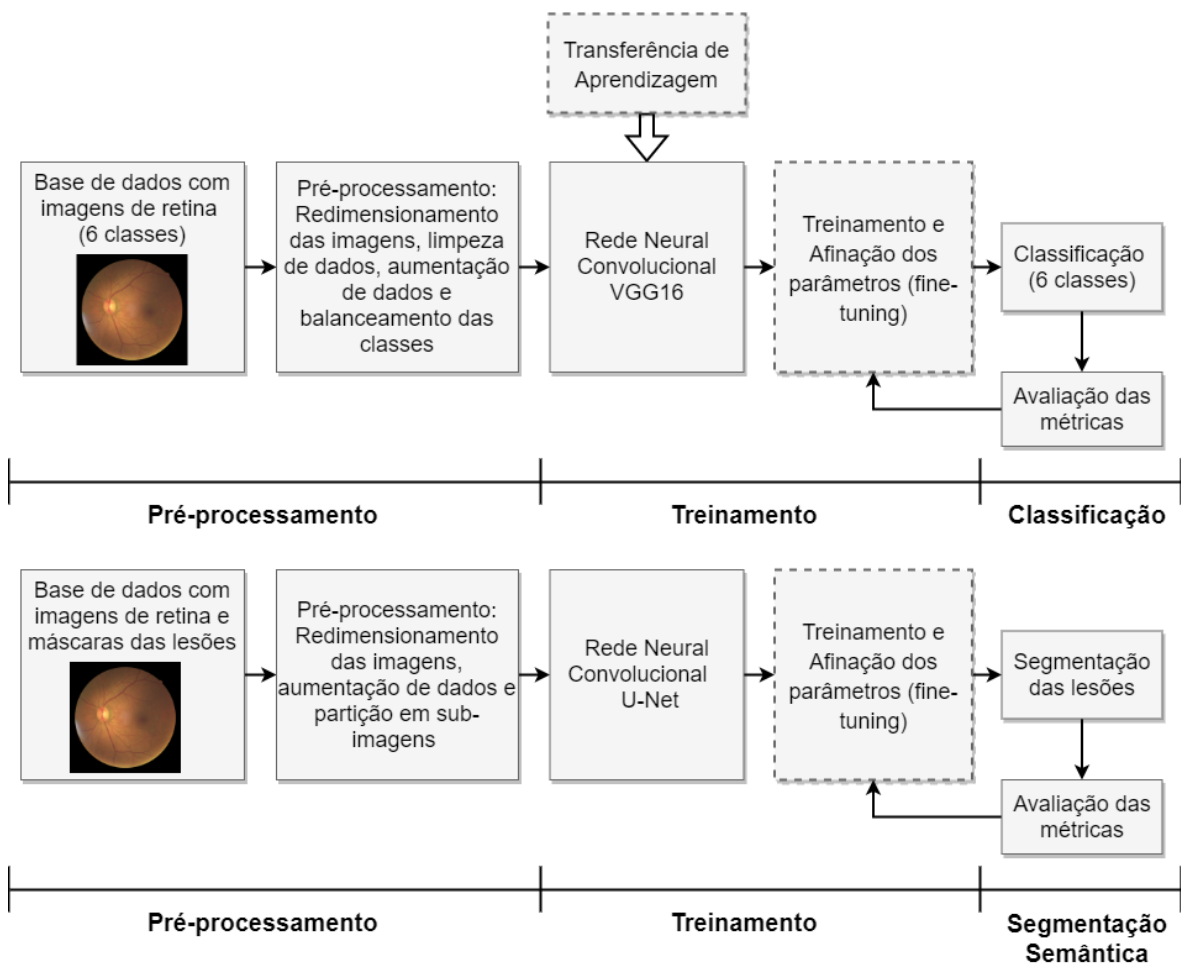
Com base na revisão bibliográfica realizada, o processo de classificação das imagens de retina com a rede neural VGG16 a ser apresentado neste trabalho busca melhorar os resultados alcançados para as métricas de desempenho em comparação com trabalhos recentes, bem como executar a classificação das imagens nos bancos de dados DDR, EyePACS/Kaggle e IDRiD em 6 classes, sendo de 0 a 4 para o grau de severidade da RD associada e a classe 5 para indicar imagens cuja qualidade que possa dificultar ou impossibilitar a análise adequada das estruturas presentes na retina.

Já o processo de segmentação com a rede neural U-Net será utilizada para detectar 4 tipos de lesões a partir da imagem original separada em sub-imagens, tendo em vista que essas lesões podem se manifestar em tamanhos, formatos e colorações distintas.

4 METODOLOGIA E PROPOSTA DE DESENVOLVIMENTO

Neste capítulo apresentam-se as etapas de desenvolvimento escolhidas conforme a metodologia adotada para a classificação do grau de retinopatia diabética e segmentação das lesões. De modo geral, esses dois processos podem ser decompostos em três etapas: pré-processamento da imagem, treinamento da rede e processo de classificação ou segmentação das imagens para a avaliação das métricas de desempenho. Um diagrama simplificado destacando as principais etapas é mostrado na Figura 8.

Figura 8 – Diagrama em blocos do sistema proposto, com destaque para os processos de classificação do grau de RD e de segmentação das imagens das lesões



Fonte: Elaborado pelo autor.

Os processos de classificação e segmentação são compostos segundo as estruturas de redes convolucionais para classificação e segmentação comumente utilizadas na literatura técnico-científica corrente, compreendendo também a etapa de limpeza das bases de dados conforme Riaz et al. (2020) e o procedimento de classificação em 6 classes com base no trabalho de Li et al. (2019). No processo de segmentação, a rede U-Net é empregada para segmentação semântica. Conforme visto na coletânea de trabalhos de Porwall et

al. (2020), esta proposta apresenta um diferencial por executar a segmentação sobre a particionada, similar ao processo de partição de imagens apresentado por Amorim et al. (2019), de forma que o redimensionamento não impacte significativamente na redução do tamanho das lesões, sobretudo aquelas menores.

A seguir, cada uma das etapas será detalhada.

4.1 Bases de dados

Nesta subseção são apresentadas as três bases de dados utilizadas neste trabalho. A Tabela 4 apresenta uma breve comparação entre elas.

Tabela 4 – Comparação entre as bases de dados

Base de dados	Total de Imagens	Classes	Segmentação de lesões	Múltiplos avaliadores
DDR	13673	0 a 5	Sim, 4 tipos (757 imagens)	Sim
EyePACS/Kaggle	88702	0 a 4	Não	Não
IDRiD	516	0 a 4	Sim, 4 tipos (81 imagens)	Sim

Fonte: Elaborado pelo autor.

4.1.1 DDR

A base de dados DDR (LI et al., 2019) está disponível publicamente e é composta por 13673 imagens de fundo coloridas em alta resolução, coletadas entre 2016 e 2018 em 147 hospitais situados em 23 províncias na China. Essas imagens foram providas por 9598 pacientes com idade entre 1 e 100 anos (idade média de 54,13), com proporções de 48,23% e 51,77% pertencentes ao sexo masculino e sexo feminino, respectivamente. Todas as imagens possuem 45° FOV (*Field of View*) e são centradas no ponto médio da linha que conecta o disco óptico à fóvea. Como proposto por essa base de dados, para pacientes saudáveis, as estruturas oculares apresentadas nas imagens devem ser claramente identificáveis, enquanto para pacientes com RD as lesões devem ser claras e legíveis. Todas as 13673 imagens possuem avaliação do grau de RD (anotação a nível de imagem em 5 níveis de severidade e 1 nível adicional para avaliar imagens com baixa qualidade visual) feita por 7 avaliadores, com classificação final determinada por maioria. Além disso 757 imagens são fornecidas com segmentação de 4 tipos de lesões (MAs, HMs, EXs moles e EXs duros) e detecção com *bounding-box*, para localização das lesões. A base de dados também fornece as imagens divididas nos conjuntos de treino, validação e teste, nas proporções de 50%, 20% e 30%, respectivamente.

4.1.2 *EyePACS/Kaggle*

Popularmente conhecida como Kaggle DR, a base de dados pública da EyePACS (*Eye Picture Archive Communication System*) (CUADROS; BRESNICK, 2009) fez parte de uma competição de detecção de RDs da Kaggle em 2015 e é composta por 88702 imagens de fundo de olho coloridas em alta resolução, sendo divididas em 35126 imagens para treinamento e 53576 imagens para teste. As imagens apresentam avaliação clínica do grau de RD em uma escala de 0 a 4, sendo 0 considerado saudável e 4 como RD proliferativa. Em contrapartida, esta base não apresenta segmentação das lesões. Atualmente, ela é a maior base de dados pública para classificação de RD, mas apresenta diversos problemas em relação à qualidade visual das imagens, com presença de ruídos, artefatos causados pelas lentes do sistema de captura, aberrações cromáticas, saturação excessiva, iluminação inadequada, dentre outros (RIAZ et al., 2020).

4.1.3 *IDRiD*

A base de dados pública IDRiD (*Indian Diabetic Retinopathy Image Dataset*) (PORWAL et al., 2020) é composta por imagens de fundo de olho em alta resolução de pacientes indianos. Ela está disponível como parte da competição *Diabetic Retinopathy: Segmentation and Grading Challenge*, organizada no ano de 2018 em conjunto com o IEEE International Symposium on Biomedical Imaging (ISBI-2018). Todas as imagens possuem 50° FOV. A base possui 516 imagens com classificação de RD na escala de 0 a 4 e Edema Macular Diabético (EMD) na escala de 0 a 2 de acordo com o ICDRS (*International Clinical Diabetic Retinopathy Scale*) e 81 imagens segmentadas considerando 4 tipos lesões (MAs, HMs, EXs moles e EXs duros) e o disco óptico. A base de dados também já fornece as imagens separadas nos subconjuntos de treinamento e teste.

4.2 Ambiente de desenvolvimento: *Google Colaboratory*

O *Google Colaboratory*, ou simplesmente *Colab*, é produto do *Google Research* que permite escrita e execução de código *Python* por meio do navegador e é especialmente adequado para aprendizado de máquina, análise de dados e educação. Mais tecnicamente, é um serviço de *notebook Jupyter* hospedado que não requer configuração para uso, ao mesmo tempo que fornece acesso gratuito a recursos de computação, incluindo GPUs (GOOGLE RESEARCH, 2021).

Com o acesso gratuito às máquinas virtuais, o *Colab* conta com bibliotecas essenciais para desenvolvimento de soluções de *machine learning* como *TensorFlow* e *Keras*, utilizadas neste trabalho, dentre outras. Já a conexão a estas máquinas virtuais é de até 12 horas na versão gratuita e a disponibilidade de recursos não é garantida, podendo variar conforme demanda e utilização dos usuários. Em se tratando de capacidade computacional, ele fornece 12 GB de RAM (*Random Memory Access*) e algumas dezenas de GB

de armazenamento em disco. Também há a possibilidade de integração com uma conta do *Google Drive* para alocação de informações, já que após encerramento de execução do *Jupyter Notebook*, todos os dados do usuário e configurações da máquina virtual são perdidos (CARNEIRO et al., 2018).

A depender da estrutura e parametrização da rede, tamanho dos dados de entrada e complexidade do problema, essas limitações terão diferentes graus de impacto no tempo de treinamento e teste dos modelos.

4.3 Pré-processamento para classificação do grau de RD

Antes de utilizar as imagens para treinamento da rede neural, é necessário adequar as dimensões das imagens, avaliar a distribuição entre as classes e agrupá-las de acordo os respectivos graus de RD.

4.3.1 Adequação das dimensões das imagens

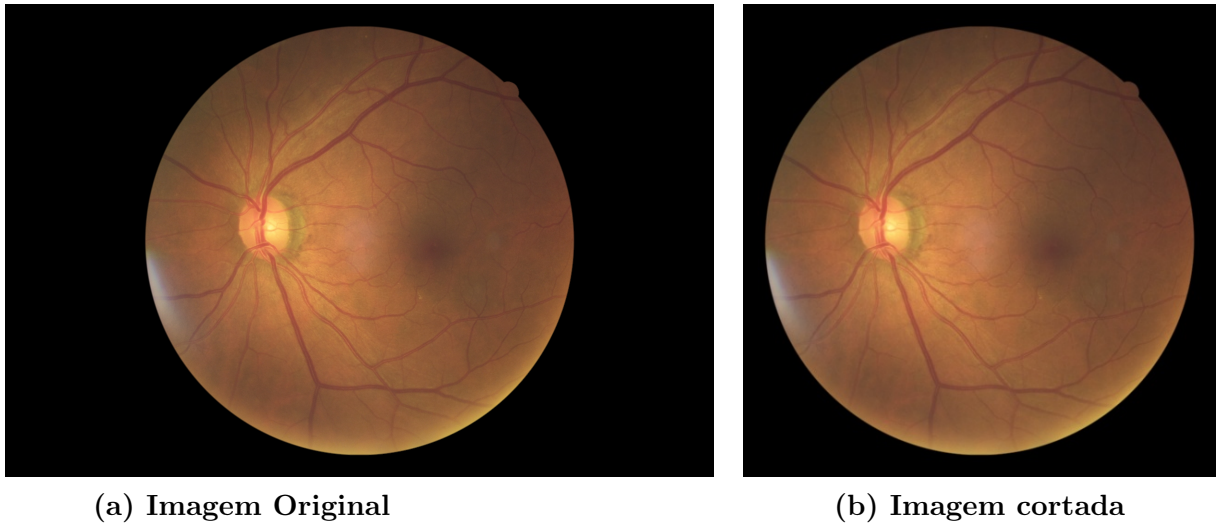
Seguindo a proposta de Li et al. (2019) e Wan, Liang e Zhang (2018), todas as imagens foram redimensionadas para 512×512 pixels, reduzindo a quantidade de memória necessária para alocar toda a base de dados (PRATT et al., 2016) e o processamento em regiões com pixels de máscara. Para isso, e de forma a evitar que a imagem fique distorcida devido às diferenças entre a altura e largura durante o redimensionamento, a menor dimensão da imagem foi selecionada e a partir daí, fez-se o redimensionamento mantendo-se as devidas proporções. A partir do centro dessa imagem, foram calculadas as coordenadas para cortá-la, reduzindo-se parte da máscara e aproximando as extremidades da imagem da área que possui informações da retina, produzindo-se uma imagem quadrada com dimensão 512×512 pixels. A Figura 9 mostra a imagem número 007-0473-000 do banco de dados DDR, antes e após a adequação das dimensões da imagem.

4.3.2 Limpeza (*Data Cleaning*) e Aumentação de Dados (*Data Augmentation*)

As três bases de dados escolhidas são desbalanceadas e com predominância da classe 0, que corresponde a pacientes sem manifestações aparentes na retina. Existem classes com pouca representatividade em relação à quantidade de amostras disponíveis, o que leva a uma contribuição reduzida durante a etapa de treinamento e, conseqüentemente, limitações quanto à obtenção do modelo para a classificação do grau de RD. A Figura 10 mostra a distribuição das classes de RD no conjunto de treinamento das bases de dados. Vale observar que a base DDR apresenta uma classe a mais em relação às outras devido à avaliação de imagens com qualidade pobre ou seja, que não possibilitam obter o grau de RD associado.

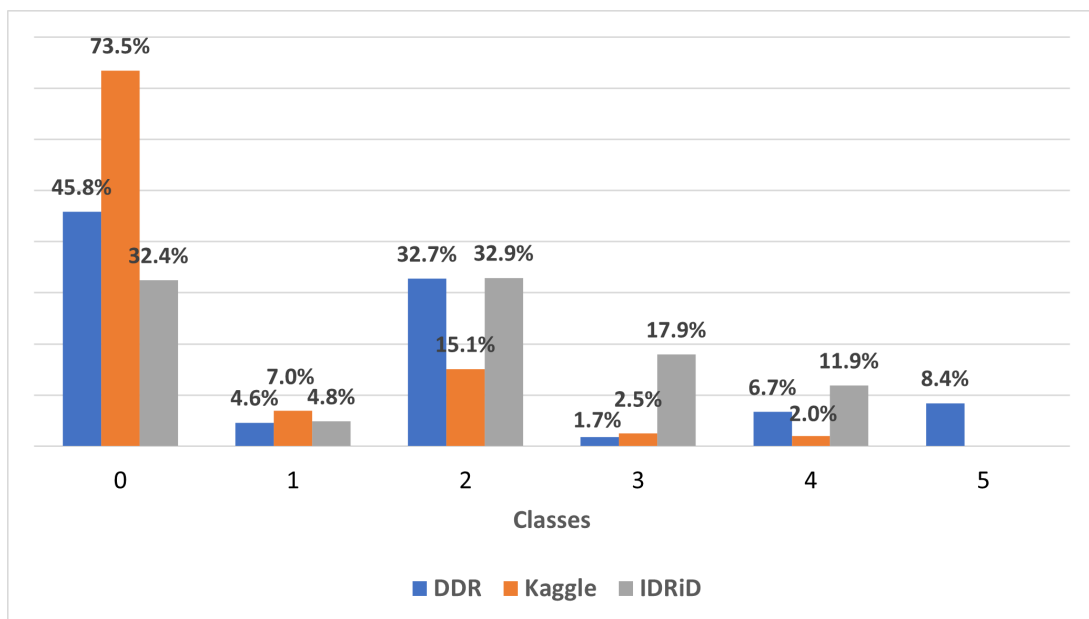
Nas bases de dados também são encontradas imagens com excesso de artefatos e de baixa qualidade nas diversas classes, principalmente na base EyePACS/Kaggle. Seguindo

Figura 9 – Modificação do tamanho da imagem original (a), tornando-a quadrada e reduzindo a área da máscara (b). Imagem 007-0473-000 do banco de dados DDR.



Fonte: Elaborado pelo autor.

Figura 10 – Distribuição desbalanceada entre as classes das bases de dados. As classes de 0 a 5 representam, respectivamente, os níveis de RD ausente, leve, moderado, severo e proliferativo.

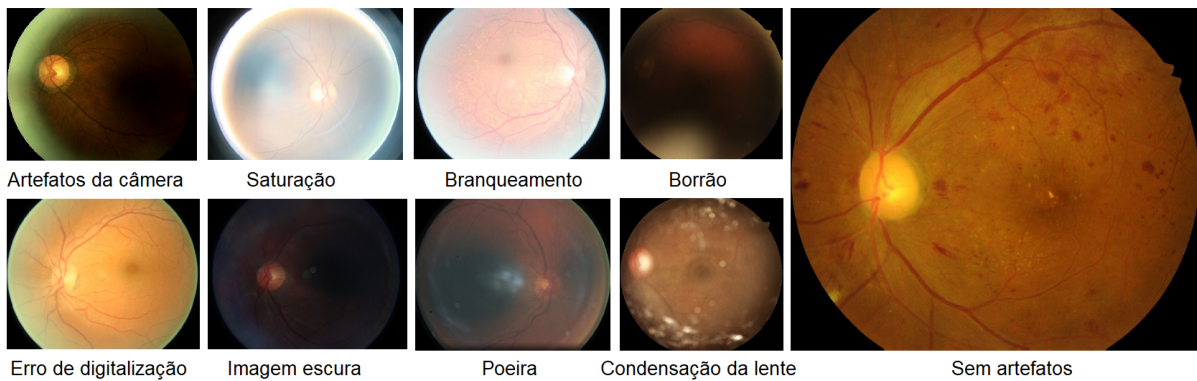


Fonte: Elaborado pelo autor.

parte da proposta de Riaz et al. (2020), que indica e caracteriza os artefatos presentes em imagens de retina, para este trabalho foi feito uma limpeza manual nas 3 bases de dados, movendo-se para a classe 5 as imagens que apresentavam artefatos e qualidade insuficiente, conforme apresenta a Figura 11.

Para minimizar a preferência da rede em relação a um determinado subconjunto de imagens devido à diferença do número de amostras pertencentes a cada classe, é usual a

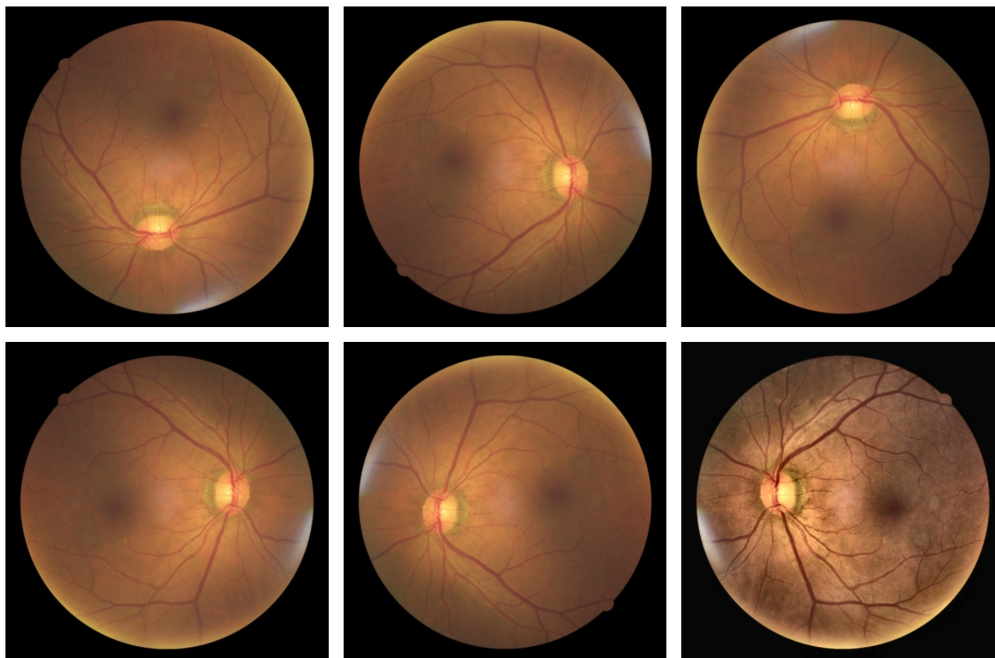
Figura 11 – Artefatos em imagens da base de dados EyePACS/Kaggle.



Fonte: Adaptado de Riaz (2020).

aplicação de técnicas de aumento de dados. Neste trabalho foram aplicados rotação (90° , 180° e 270°), espelhamento vertical e horizontal e modificação de contraste por meio da técnica CLAHE (*Contrast Limited Adaptive Histogram Equalization*) (ZUIDERVELD, 1994), conforme ilustrado na Figura 12 (LI et al., 2019; ORNEK; CEYLAN, 2019; ZHANG et al., 2019; WAN; LIANG; ZHANG, 2018; PAWARA et al., 2017; QUELLEC et al., 2017).

Figura 12 – Aumentação de dados aplicada à imagem original mostrada na Figura 9. Da esquerda para a direita e de cima para baixo: rotação em 90° , 180° e 270° , espelhamento vertical e horizontal, e modificação de contraste com CLAHE.



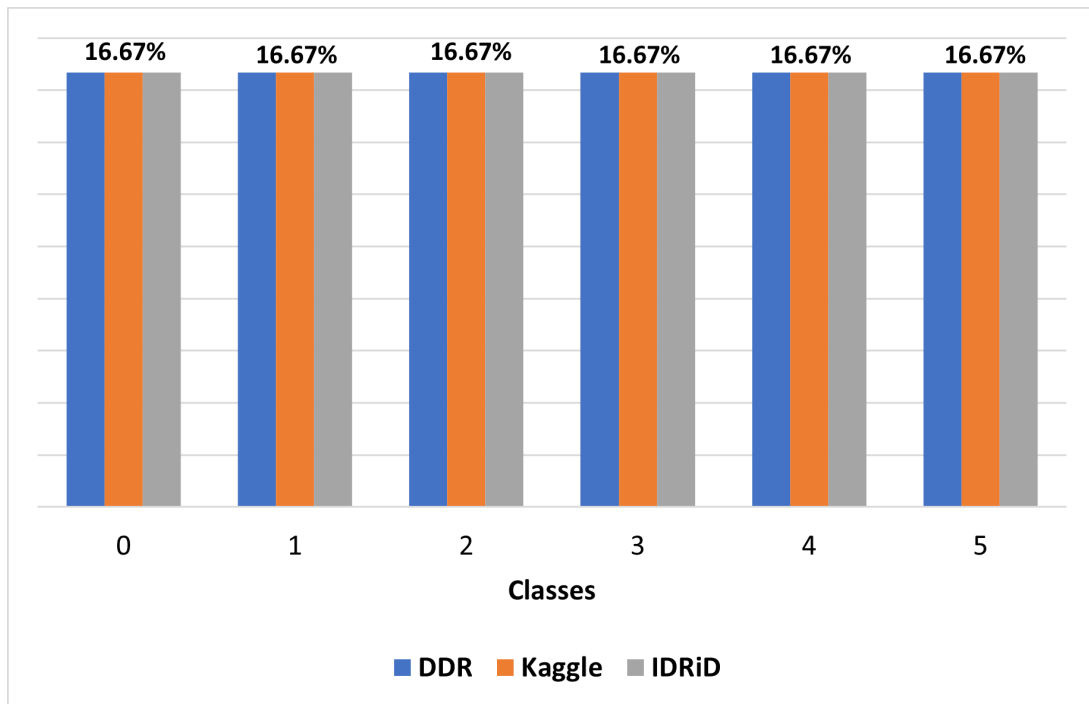
Fonte: Elaborado pelo autor.

A proposta deste trabalho inclui uma classe específica (classe 5) para imagens de baixa qualidade ou que apresentem artefatos. Considerando a inexistência dessa classe nas prin-

cipais bases de dados públicos, um filtro gaussiano com kernel 25×25 foi utilizado para borrar uma seleção aleatória das imagens disponíveis nas três bases de dados, piorando a qualidade visual. As imagens resultantes desse processo foram agrupadas com outras imagens que apresentavam artefatos durante a limpeza dos dados.

Assim, para o treinamento, cada uma das 6 classes na base DDR foram representadas com 800 imagens, totalizando 4800 imagens. A base da EyePACS/Kaggle foi representada com 3000 imagens por classe, totalizando 18000 imagens. Já na base IDRiD foram 140 imagens por classe, totalizando 840 imagens. Portanto, todas as classes foram igualmente representadas nas três bases de dados com uma proporção de 16,67% para cada classe, conforme mostra a Figura 13.

Figura 13 – Distribuição balanceada entre as classes das bases de dados.



Fonte: Elaborado pelo autor.

Após balancear as classes conforme apresentado na Figura 13, as imagens resultantes foram submetidas ao retreino da rede, previamente treinada utilizando os pesos definidos com os dados da *ImageNet*. Também foi incluído um bloco para aumento de dados gerando dados novos (*on the fly*) a partir das imagens selecionadas em cada *batch* para que a rede neural não utilizasse a mesma imagem após nova submissão de imagens do banco de dados. Essa estrutura proporciona rotação de (0° a 45°), translação (10% da imagem para esquerda ou direita e para cima ou para baixo), *zoom* (0,95 - 1,05), espelhamento e deformação (*shear*).

Tabela 5 – Hiperparâmetros e dados de treinamento da rede VGG16

Hiperparâmetros	Valor
Função de Otimização	Gradiente Descendente Estocástico (SGD - Stochastic Gradient Descent)
<i>Epochs</i>	100
Tamanho do <i>batch</i>	32
Taxa de Aprendizado	1×10^{-2}
Dropout	0,5
Número de classes	6
Amostras por classe	800 (DDR), 3000 (Kaggle), 140 (IDRiD)

Fonte: Elaborado pelo autor.

4.3.3 Normalização dos dados

Cada imagem pré-processada é individualmente padronizada de 0 a 255 para 0 a 1, respectivamente, e normalizada com média zero e desvio padrão unitário. Esse procedimento é empregado para evitar que características (ou valores) com intervalos muito diferentes sejam aplicadas à entrada da rede, impactando na convergência do modelo (CHOLLET, 2017). As operações aritméticas são realizadas com dados em ponto flutuante pelas GPUs (*Graphic Processing Units*) e CPUs (*Central Processing Units*), dispositivos já otimizados para esses cálculos (RIAZ et al., 2020).

4.4 Treinamento da rede VGG16

A rede VGG16 desenvolvida por Simonyan e Zisserman (2014), apresentada na Figura 6, foi treinada para classificação em 1000 categorias. A fim de acelerar o treinamento, foi utilizado o modelo pré-treinado da VGG16 com pesos obtidos pelos criadores dessa rede na competição *ImageNet*, em 2014. As camadas de convolução e funções de ativação foram mantidas congeladas, ou seja, durante o processo de treinamento os pesos dessas camadas não serão modificados pela técnica de retropropagação. As camadas densamente conectadas foram modificadas de $1 \times 1 \times 4096$ para $1 \times 1 \times 1024$ (PRATT et al., 2016) e a última camada foi modificada de $1 \times 1 \times 1000$ para $1 \times 1 \times 6$, utilizando-se a função de ativação *softmax*, para reduzir o número de categorias de 1000 para 6, em conformidade com a proposta deste trabalho. A Tabela 5 destaca os hiperparâmetros utilizados no treinamento.

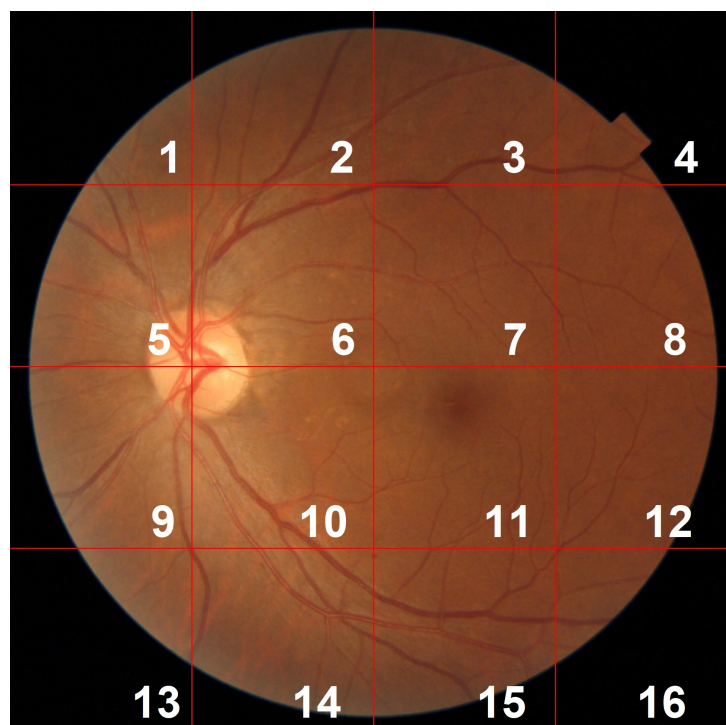
O treinamento da rede VGG16 para as bases de dados foi executado com 100 *epochs* com a função de otimização SGD. As técnicas de *dropout* e decaimento ponderado também foram aplicadas a fim de minimizar o risco de *overfitting*.

4.5 Pré-processamento para segmentação das lesões

Considerando que a representação das lesões em imagens de fundo de olho normalmente ocupam uma porção pequena da imagem, principalmente os MAs (INTERNATI-

ONAL COUNCIL OF OPHTHALMOLOGY, 2017; LI et al., 2019), a detecção dos pixels dessas lesões podem ser comprometidas durante o processo de segmentação caso toda a imagem seja redimensionada para dimensões menores. Em relação aos bancos de dados DDR e IDRiD, pôde-se observar que todas as imagens possuem dimensões superiores a 1024×1024 pixels. Assim, todas as imagens e suas respectivas máscaras foram redimensionadas para o tamanho de 1024×1024 pixels e particionadas em 16 fragmentos (ou *tiles*), gerando sub-imagens com dimensões 256×256 pixels, conforme ilustra a Figura 14.

Figura 14 – Exemplo da geração de sub-imagens com dimensões 256×256 pixels. Imagem IDRiD_01 da base IDRiD.

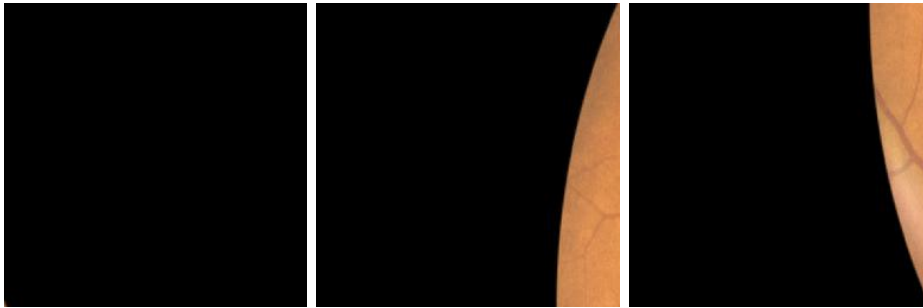


Fonte: Elaborado pelo autor.

Ao avaliar todo o novo conjunto de dados gerado a partir da base DDR, notou-se que algumas sub-imagens representavam partes da máscara e, portanto, não possuíam informação relevante sobre a retina, conforme apresentado na Figura 15. Por essa razão, optou-se por descartá-las, reduzindo em aproximadamente 35% o conjunto de dados gerado para a base DDR. Não houve remoção de dados nas sub-imagens da base IDRiD, já que todas elas apresentam porções representativas da retina, e não somente da máscara. Para cada sub-imagem resultante da etapa de seleção foi aplicado o processo de melhoria de contraste através do CLAHE (ZUIDERVELD, 1994).

Assim como ocorre no procedimento de aumento de dados para classificação do grau de RD, nesta etapa do pré-processamento das sub-imagens das lesões também utilizou-se a aumento de dados (*on the fly*) a partir das imagens selecionadas em cada *batch*, através de rotação (0° a 45°), translação (10% da imagem para esquerda ou direita e para

Figura 15 – Exemplos de sub-imagens RGB removidas do conjunto de dados. Imagem original 007-1774-100-0-3 da base DDR.



Fonte: Elaborado pelo autor.

cima ou para baixo), *zoom* (0.95 - 1.05) e espelhamento horizontal e vertical. As mesmas transformações utilizadas nas imagens RGB são aplicadas nas respectivas máscaras.

4.6 Segmentação das lesões com U-Net

A fim de identificar os 4 tipos de lesões (MAs, HMs, EXs moles e EXs duros) a partir das imagens de retina, 4 redes U-Net foram treinadas separadamente para obtenção dos modelos em cada base de dados. A camada de entrada foi padronizada considerando imagens com dimensões 256×256 pixels e espaço de cores RGB. A estrutura utilizada neste trabalho foi apresentada previamente na Figura 7.

A função de ativação da camada de saída é do tipo sigmoide, enquanto que a função de custo é a entropia cruzada binária (*binary cross-entropy*). A Tabela 6 destaca os hiperparâmetros utilizados no treinamento.

Tabela 6 – Hiperparâmetros e dados de treinamento da rede U-Net

Hiperparâmetros	Valor
Função de Otimização	Adam
Epochs	100
Tamanho do batch	16
Número de classes	Classificação binária por lesão
Número de imagens para treinamento	383 (DDR), 54 (IDRiD)
Número de sub-imagens para treinamento	3984 (DDR), 864 (IDRiD)

Fonte: Elaborado pelo autor.

4.7 Métricas de Desempenho

Utilizando-se a matriz de confusão obtida para as 6 classes do problema é possível obter algumas métricas de desempenho conforme descrito a seguir.

Segundo Sokolova e Lapalme (2009), Zhang et al. (2019), a acurácia indica a proporção de amostras classificadas corretamente; a precisão indica a proporção de predições positivas (corretamente classificadas) dentre aquelas classificadas como positivas; a sensibilidade (*recall*) mede a fração de predições positivas (corretamente classificadas) dentre

todas as amostras positivas; a especificidade mede a proporção de amostras corretamente identificadas como negativas; o *F1-score* é a média harmônica entre precisão e sensibilidade.

A partir das equações para as métricas de desempenho acurácia, sensibilidade (*recall*), especificidade, precisão e *F1-score* para classificação binária apresentadas nas equações (2) a (6), em combinação com a matriz de confusão, mostrada Tabela 7 para 6 classes e as equações (7) a (10), é possível expandir as definições dessas métricas de desempenho para n classes. A métrica IoU (Interseção sobre União média) indicada na equação (12) permite avaliar a área de sobreposição entre a segmentação predita (A) e o seu respectivo *ground truth* (B), utilizada apenas para verificar o desempenho da segmentação binária.

Tabela 7 – Matriz de confusão genérica para 6 classes

Classe		Predito					
		0	1	2	3	4	5
Real	0	$A_{0,0}$	$A_{0,1}$	$A_{0,2}$	$A_{0,3}$	$A_{0,4}$	$A_{0,5}$
	1	$A_{1,0}$	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{1,4}$	$A_{1,5}$
	2	$A_{2,0}$	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{2,4}$	$A_{2,5}$
	3	$A_{3,0}$	$A_{3,1}$	$A_{3,2}$	$A_{3,3}$	$A_{3,4}$	$A_{3,5}$
	4	$A_{4,0}$	$A_{4,1}$	$A_{4,2}$	$A_{4,3}$	$A_{4,4}$	$A_{4,5}$
	5	$A_{5,0}$	$A_{5,1}$	$A_{5,2}$	$A_{5,3}$	$A_{5,4}$	$A_{5,5}$

$$Acurcia (Acc) = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Sensitividade (recall, Sp) = \frac{TP}{TP + FN} \quad (3)$$

$$Especificidade (Sp) = \frac{TN}{TN + FP} \quad (4)$$

$$Precisão (PPV) = \frac{TP}{TP + FP} \quad (5)$$

$$F1-score = 2 \times \frac{Precisão \times Recall}{Precisão + Recall} \quad (6)$$

Os verdadeiros-positivos (TP) referem-se às amostras corretamente classificadas como positivas enquanto os verdadeiros-negativos (TN) se referem às amostras corretamente classificadas como negativas. Os falsos-positivos (FP) e falsos-negativos (FN) são as amostras classificadas erroneamente como positivas e negativas, respectivamente.

$$TP \text{ para classe } k = A_{k,k} \quad (7)$$

$$TN \text{ para classe } k = \left(\sum_{i=0}^n \sum_{j=0}^n A_{i,j} \right) - \sum_{i=0}^n A_{i,k} - \sum_{i=0}^n A_{k,i} + A_{k,k} \quad (8)$$

$$FP \text{ para classe } k = \left(\sum_{i=0}^n A_{i,k} \right) - A_{k,k} \quad (9)$$

$$FN \text{ para classe } k = \left(\sum_{i=0}^n A_{k,i} \right) - A_{k,k} \quad (10)$$

$$Acurácia \text{ global} = \frac{\sum_{i=0}^n A_{i,i}}{\sum_{i=0}^n \sum_{j=0}^n A_{i,j}} \quad (11)$$

$$IoU(A,B) = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN} \quad (12)$$

onde i e j indicam a linha e coluna da matriz de confusão, respectivamente.

Neste trabalho, para o processo de classificação do grau de RD serão utilizadas as métricas acurácia (ACC), precisão (PPV - *Positive Predictive Value*), sensibilidade ou *recall* (Se), especificidade (Sp) e F1-score para avaliação das respectivas médias (*macro-averaging*) para avaliação global do sistema. Já para o processo de segmentação das lesões serão utilizadas as métricas IoU, sensibilidade ou *recall* (Se) e precisão (PPV).

4.8 Conclusão

Com base na revisão bibliográfica realizada, os processos de classificação do grau de RD e segmentação de lesões desenvolvidos deverão buscar a melhoria das métricas de desempenho.

No pré-processamento, as etapas de adequação das dimensões das imagens reduz a área de máscara, a qual não possui informação relevante, a limpeza manual das bases de dados excluem as amostras com artefatos nas diversas classes e permite agrupá-las em uma nova classe, a fim de possibilitar a classificação das amostras com qualidade insuficiente. O balanceamento das classes e a aumentação de dados possibilitam representar cada classe com as mesmas quantidades de amostras, minimizando a predominância de determinada classe durante o treinamento da rede VGG16.

Para a segmentação, a utilização da partição da imagem original permite submeter à U-Net amostras adequadas ao tamanho da camada de entrada utilizada, reduzindo a taxa de redimensionamento a fim de evitar exclusões de regiões com presença de lesões, sobretudo aquelas presentes em regiões muito pequenas como são os casos dos MAs.

As imagens a serem utilizadas no treinamento e teste foram obtidas a partir das bases de dados DDR, EyePACS/Kaggle e IDRiD, publicamente disponíveis. Os treinamentos e testes de ambas as redes neurais foram realizados em Python, com a aplicação de *Jupyter Notebook*, em ambiente *Google Colab*.

5 RESULTADOS E DISCUSSÕES

Neste capítulo serão apresentados os resultados obtidos e análises relativas aos processos de classificação do grau de retinopatia diabética considerando 6 classes para as bases de dados DDR, EyePACS/Kaggle e IDRiD, além da segmentação de 4 tipos de lesões considerando as bases de dados DDR e IDRiD. A base de dados da EyePACS/Kaggle não apresenta *ground truths* de lesões, portanto ela não foi utilizada na tarefa de segmentação.

5.1 Classificação do grau de RD

A Tabela 8 apresenta uma comparação entre os resultados da classificação dos graus de RDs obtidos neste trabalho e outros levantados a partir publicações recentes, destacados em negrito as melhores métricas para cada uma das base de dados. Para os trabalhos que apresentaram resultados para diferentes estruturas de redes neurais, considerou-se apenas os melhores desempenhos indicados pelos respectivos autores. Observa-se que a estratégia adotada para o desenvolvimento da proposta aqui apresentada produziu resultados competitivos, exceto em relação à base Kaggle.

A DDR é a mais recente dentre as três bases utilizadas, adotada em poucos trabalhos até o momento. Em relação à acurácia, foi obtida uma melhoria de 12,8% em relação ao trabalho de Li et al. (2019), o que mostra uma maior taxa de acerto na determinação correta do grau de RD associado à imagem sob análise.

Quanto à base IDRiD há uma ressalva, pois os resultados obtidos por Porwal et al. (2020) também incluem detecção de edema macular diabético (DME - *Diabetic Macular Edema*), e não há registro informando o comportamento do sistema de classificação em termos de variação nas métricas de desempenho a partir da inclusão ou remoção desta detecção adicional. Assim, entende-se que uma comparação direta dos resultados deve ser feita com cautela.

Quanto à base Kaggle/EyePACS, que é uma das mais populares dentre as que envolvem imagens de fundo de olho, quase todos os trabalhos (RIAZ et al., 2020; GULSHAN et al., 2016; LEEZA; FAROOQ, 2019; WAN; LIANG; ZHANG, 2018) reportaram métricas de desempenho acima das alcançadas neste trabalho, exceto em relação à sensibilidade e acurácia segundo Pratt et al. (2016).

Vale destacar que todos os trabalhos elencados na Tabela 8 utilizam 4 ou 5 classes na avaliação do grau de RD, exceto em Li et al. (2019) onde também são avaliadas 6 classes, como também proposto neste trabalho.

5.2 Segmentação das lesões

Esta seção apresenta algumas ilustrações sobre os processos de segmentação das lesões MAs, HMs, EXs moles e duros, bem como análises comparativas dos resultados obtidos em relação a publicações disponíveis na literatura técnico-científica recente.

Tabela 8 – Análise comparativa de resultados sobre a classificação de RDs

Trabalho	Base de dados/Classes	Acc	PPV	Se	Sp	F1-score
Gulshan et al., 2016	EyePACS-1 (privado) / 5	-	-	0,903	0,981	-
Wan et al., 2018	Kaggle / 5	0,9568	0,8647	0,9743	-	-
Leeza and Farooq 2019	Kaggle / 5	0,9830	0,9574	0,9592	0,989	-
Riaz et al., 2020	Kaggle / 5	-	-	0,94	0,97	-
Li et al., 2019	DDR / 6	0,7659	-	-	-	-
Porwal et al., 2020	IDRiD / 5	0,6311	-	-	-	-
Proposto	DDR / 6	0,892	0,857	0,881	0,861	0,867
Proposto	Kaggle / 6	0,835	0,858	0,856	0,832	0,857
Proposto	IDRiD / 6	0,902	0,866	0,879	0,890	0,872

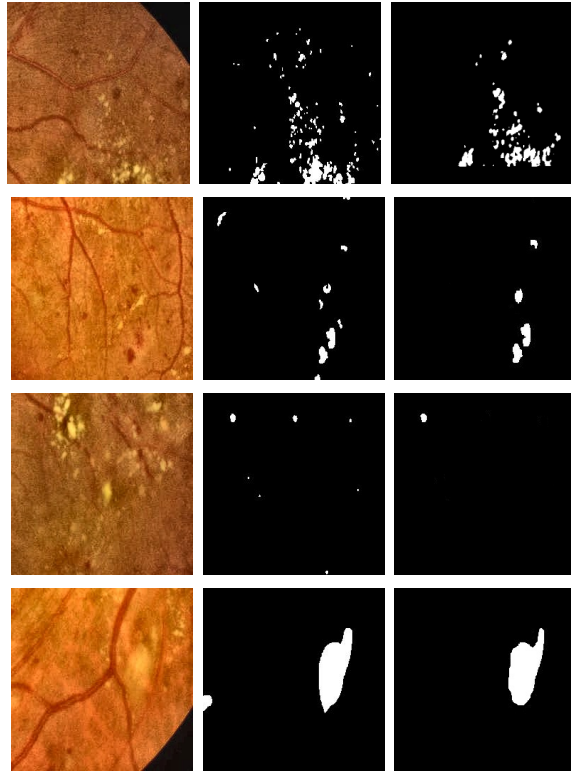
Fonte: Elaborado pelo autor.

A Figura 16 mostra algumas ilustrações da segmentação destas lesões obtidas para 4 sub-imagens do banco de dados IDRiD. Ao se comparar as imagens originais com as máscaras, dispostas sobre a segunda coluna, pode-se perceber que os modelos conseguiram produzir uma segmentação grosseira que contém as regiões das lesões identificadas, sendo que lesões em tamanhos maiores e com formas mais concentradas são melhor detectadas se comparadas às regiões de lesões espalhadas e em formatos menores. Ainda assim, pode-se notar a presença de falso-positivos e falso-negativos em determinadas partes das imagens após o processo de aplicação das máscaras.

A fim de avaliar o desempenho dos processos de segmentação para esses 4 tipos de lesões, as Tabelas 9 a 12 mostram os resultados das métricas alcançados neste trabalho em relação a outros disponíveis na literatura atual, onde os melhores índices foram destacados em negrito. Pode-se observar que todas as métricas obtidas foram superiores aos resultados apresentados por Porwal et al. (2020) e Li et al. (2019), que disponibilizaram publicamente os banco de dados IDRiD e DDR, respectivamente. Estas melhorias são atribuídas também ao processo de utilização de sub-imagens para a segmentação, evitando que lesões menores ou com representação espacial dispersa sejam reduzidas ainda mais durante o redimensionamento das imagens para entrada na rede U-Net. Assim, com a utilização das sub-imagens, as lesões ocupam uma região proporcionalmente maior do que nas imagens originais, permitindo que características sejam identificadas e extraídas mais facilmente.

Como na base DDR, a quantidade de imagens de retina e *ground truths* disponíveis é maior em relação ao banco de dados IDRiD, sua utilização permite melhorar a eficiência dos processos de desenvolvimento e avaliação de novos modelos de segmentação semântica

Figura 16 – Exemplo da segmentação obtida para 4 tipos de lesões de RDs. Coluna da esquerda: sub-imagens originais, de cima para baixo, EXs duros, HMs, MAs e EXs moles. Coluna central: respectivas máscaras (*ground truths*). Coluna da direita: imagens segmentadas pela U-Net.



Fonte: Elaborado pelo autor.

ou outros tipos de redes neurais.

Os trabalhos apresentados por Porwal et al. (2020) contemplam unicamente a métrica de AUPR (*Area Under Precision Recall*) ou área abaixo da curva de precisão-*recall*, enquanto que em (LI et al., 2019) há a utilização das métricas IoU e PPV. Adicionalmente, optou-se pela inclusão da sensibilidade de modo a reforçar a avaliação da segmentação obtida e para que futuros trabalhos possam estabelecer comparações com base em um conjunto mais amplo de métricas de desempenho.

5.3 Conclusão

Neste capítulo foram apresentados os resultados obtidos para os processos de classificação do grau de RD e segmentação das lesões, comparados aos trabalhos disponíveis na literatura atual. Também foram destacadas as limitações da metodologia proposta considerando-se as informações fornecidas pelas bases de dados Kaggle e IDRiD. Pôde-se, também, comprovar os ganhos alcançados em relação às métricas na classificação e segmentação, sobretudo na base DDR, se comparadas com o trabalho de Li et al. (2019). Para a segmentação, destaca-se os resultados alcançados com a proposta da utilização

Tabela 9 – Análise comparativa de resultados sobre a segmentação de MAs

Trabalho	Base de dados	IoU	Se	PPV	AUPR
Porwal et al. (2020)	IDRiD	-	-	-	0,5017
Porwal et al. (2020)	IDRiD	-	-	-	0,4951
Porwal et al. (2020)	IDRiD	-	-	-	0,4740
Li et al. (2019)	DDR	0,0325	-	0,0205	-
Proposto	IDRiD	0,1222	0,3935	0,0987	0,5244
Proposto	DDR	0,1074	0,3506	0,0701	0,5012

Fonte: Elaborado pelo autor.

Tabela 10 – Análise comparativa de resultados sobre a segmentação de HMs

Trabalho	Base de dados	IoU	Se	PPV	AUPR
Porwal et al. (2020)	IDRiD	-	-	-	0,6804
Porwal et al. (2020)	IDRiD	-	-	-	0,6490
Porwal et al. (2020)	IDRiD	-	-	-	0,5588
Li et al. (2019)	DDR	0,1425	-	0,1842	-
Proposto	IDRiD	0,2546	0,5167	0,2975	0,7120
Proposto	DDR	0,2267	0,5034	0,2598	0,6858

Fonte: Elaborado pelo autor.

Tabela 11 – Análise comparativa de resultados sobre a segmentação de EXs moles

Trabalho	Base de dados	IoU	Se	PPV	AUPR
Porwal et al. (2020)	IDRiD	-	-	-	0,6995
Porwal et al. (2020)	IDRiD	-	-	-	0,6607
Porwal et al. (2020)	IDRiD	-	-	-	0,6588
Li et al. (2019)	DDR	0,2295	-	0,4359	-
Proposto	IDRiD	0,3846	0,5854	0,4875	0,7491
Proposto	DDR	0,3444	0,5438	0,4750	0,7002

Fonte: Elaborado pelo autor.

Tabela 12 – Análise comparativa de resultados sobre a segmentação de EXs duros

Trabalho	Base de dados	IoU	Se	PPV	AUPR
Porwal et al. (2020)	IDRiD	-	-	-	0,8850
Porwal et al. (2020)	IDRiD	-	-	-	0,8741
Porwal et al. (2020)	IDRiD	-	-	-	0,8582
Li et al. (2019)	DDR	0,3118	-	0,5634	-
Proposto	IDRiD	0,4146	0,6993	0,6175	0,8936
Proposto	DDR	0,3864	0,6425	0,5850	0,8623

Fonte: Elaborado pelo autor.

de sub-imagens, que produziu resultados competitivos em comparação as propostas de Porwal et al. (2020). Adicionalmente, a inclusão de outras métricas de desempenho possibilitaram uma análise mais ampla que poderá ser investigada mais detalhadamente em trabalhos futuros.

6 CONSIDERAÇÕES FINAIS E PROPOSTAS DE CONTINUIDADE

Este trabalho apresentou os resultados obtidos para os processos de classificação do grau de retinopatia diabética por meio da rede neural VGG16 e segmentação de quatro tipos de lesões (MAs, HMs, EXs moles e EXs duros), utilizando-se 4 redes independentes do tipo U-Net. As imagens de fundo de olho aplicadas para o treinamento e teste das redes foram extraídas dos bancos de dados públicos DDR, IDRiD e EyePacs/Kaggle, dentre as quais, no processo de segmentação, foram consideradas as bases DDR e IDRiD, pois a base de dados da EyePACS/Kaggle não apresenta *ground truths* de lesões. A complexidade demandada para a marcação manual devido aos problemas de contraste, iluminação e ruído inerentes às imagens, principalmente com o crescimento de casos de diabetes na atualidade e, conseqüentemente, de RD ((WORLD HEALTH ORGANIZATION, 2014; INTERNATIONAL COUNCIL OF OPHTHALMOLOGY, 2017; CHO et al., 2017)), torna relevante a investigação de soluções para acelerar e auxiliar o diagnóstico médico nessa área.

6.1 Considerações Finais

Com base no levantamento de trabalhos recentemente realizados acerca do tema escolhido, percebeu-se a predominância de aplicações baseadas em redes neurais, tanto para classificação quanto para segmentação. Adicionalmente, pôde-se observar o uso frequente da aumentação de dados no conjunto de amostras disponíveis e a utilização de estruturas de redes neurais já conhecidas, aproveitando-se suas características de transferência de aprendizagem para acelerar a etapa de treinamento. Além disso, especificamente na tarefa de classificação, observou-se que os trabalhos utilizam 4 ou 5 classes, de acordo com a base de dados utilizada. Adicionalmente, algumas dessas bases de dados seguem critérios de avaliação do grau de RD, sendo comumente empregadas as escalas ETDRS (*International Clinical Diabetic Retinopathy Scale*) e ICDRS (*International Clinical Diabetic Retinopathy Scale*).

Na metodologia proposta para a tarefa de classificação, as imagens RGB da retina, disponíveis nos bancos de dados públicos da EyePACS/Kaggle, DDR e IDRiD, foram pré-processadas para a padronização de tamanho e redução da região de máscara, verificação da presença de artefatos e imagens de baixa qualidade, movendo-as para a classe 5 e a aumentação de dados para balanceamento das classes, através de rotação, espelhamento e modificação de contraste. Em seguida, foi executado o treinamento da rede VGG16.

Já para a tarefa de segmentação, as imagens dos bancos DDR e IDRiD e suas respectivas máscaras foram pré-processadas através de redimensionamento e partição em 16 sub-imagens, eliminando-se aquelas que apresentavam somente região de máscara (todos os pixels pretos). Para cada sub-imagem resultante, a partir da etapa de seleção, foi aplicado o processo de melhoria de contraste através do método CLAHE. Posteriormente,

foram treinadas 4 redes U-Net, sendo uma para cada tipo de lesão.

Quanto à classificação, os resultados obtidos apresentaram melhoria de desempenho para as bases DDR e IDRiD, conforme pode ser observado nas métricas apresentadas na Tabela 8. Para as bases EyePACS/Kaggle, os resultados obtidos neste trabalho não foram suficientes para superar os trabalhos atualmente disponíveis, correlatos ao tema. Para a base IDRiD, os resultados apresentados por Porwal et al. (2020) incluem detecção simultânea de RD e EMD (Edema Macular Diabético), o que pode ser o motivo da baixa acurácia encontrada neste trabalho. A detecção de EMD não faz parte do escopo deste trabalho, portanto uma comparação direta dos resultados deve ser feita com cautela.

Em resumo, dentre as contribuições deste trabalho para o processo de classificação, pode ser destacadas:

- Melhoria dos resultados alcançados por Li et al. 2019;
- Inclusão das métricas de desempenho sensibilidade, especificidade, precisão and F1-score para as bases de dados IDRiD e DDR;
- Separação das imagens de fundo de olho em 6 classes, com as classes 0 a 4 associadas ao grau de DR e a classe 5 para indicar imagem com baixa qualidade, passível de dificultar ou impossibilitar a análise adequada das estruturas presentes na retina;
- Adequação e balanceamento dos dados na fase de pré-processamento, juntamente com a seleção adequada dos hiperparâmetros destacados na Tabela 5 na etapa de retreino da rede VGG16, o que possibilitou melhoria da métrica de acurácia global, se comparada ao trabalho de Li et al. 2019, tornando o sistema mais robusto em relação à taxa de detecção correta do grau de DR;
- Análise e avaliação prévia da qualidade visual das imagens para a distribuição das classes, com a remoção de imagens com artefatos que pudessem comprometer o desempenho do processo. Essa tarefa auxiliou na obtenção de dados mais adequados para o treinamento da rede e, dentre os trabalhos relacionados, somente foi abordada por Riaz et al. (2020).

Quanto à segmentação das lesões, todas as métricas obtidas superaram o trabalho de Li et al. (2019), também desenvolvido a partir de imagens da base DDR, conforme mostram as métricas apresentadas nas Tabelas 9, 10, 11 e 12. Dessa forma, pode-se concluir que os objetivos e metas propostas nesta pesquisa foram cumpridos, culminando em melhorias dos parâmetros de desempenho nos processos de classificação e segmentação, sobretudo na base de dados DDR apresentada por Li et al. (2019).

O desenvolvimento desta pesquisa possibilitou a submissão do artigo *Diabetic Retinopathy Classification using VGG16 Neural Network* na revista *Research on Biomedical Engineering* (RBE), publicada pela *Springer*. Além disso, um artigo sobre o processo de segmentação encontra-se em fase de elaboração.

6.2 Propostas de Continuidade

A qualidade visual das imagens constitui um fator importante para a correta classificação do grau de RD e segmentação das lesões, visto que os modelos a serem desenvolvidos devem ser capazes de diferenciar as classes do problema com assertividade. Esses modelos devem ser preparados para lidar com imagens cuja localização das lesões podem estar presentes em diferentes tamanhos e posições da retina, além de variações na rotação e iluminação, já que as estruturas oculares e lesões apresentam tamanhos e formatos irregulares e a captura destas imagens de retina ocorre em um ambiente de baixo contraste, com iluminação não-uniforme e diferentes tipos de câmeras. Portanto, para serem capazes de executar tais tarefas com robustez e segurança e, futuramente, serem empregadas como ferramenta de apoio aos profissionais da área, novas pesquisas a respeito do tema ainda são relevantes e demonstram a importância deste trabalho.

Com base nas propostas aqui desenvolvida e as outras, apresentadas nos trabalhos relacionados, pôde-se constatar que novos trabalhos poderão ser desenvolvidos a partir de outras estruturas ou pela combinação de diferentes tipos de redes neurais, incluindo-se a detecção de pontos de interesse em regiões que permitam caracterizar mais adequadamente o grau de DR, além da extração, seja por segmentação ou informações descritivas da retina acerca dos tipos de lesões e manifestações detectadas, que poderão ser aplicadas como complemento nas entradas das redes, visando melhorar os processos de classificação do grau de RD e segmentação de lesões.

Para maior alcance e validação da metodologia proposta, o algoritmo desenvolvido pode ainda ser testado em outras bases de dados, desde que estas possuam identificação adequada das classes e as máscaras das lesões para comparação.

REFERÊNCIAS

- ABBAS, Q. et al. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. Medical & Biological Engineering & Computing, v. 55, n. 11, p. 1959–1974, Nov 2017. ISSN 1741-0444. Disponível em: <<https://doi.org/10.1007/s11517-017-1638-6>>. 27, 29, 32
- ABRAMOFF, M. D. et al. Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy. Automated Analysis of Retinal Images. JAMA Ophthalmology, v. 131, n. 3, p. 351–357, 03 2013. ISSN 2168-6165. Disponível em: <<https://doi.org/10.1001/jamaophthalmol.2013.1743>>. 18, 28
- ABRAMOFF, M. D.; GARVIN, M. K.; SONKA, M. Retinal imaging and image analysis. IEEE Reviews in Biomedical Engineering, v. 3, p. 169–208, 2010. ISSN 1937-3333. 12
- ALGHADYAN, A. A. Diabetic retinopathy – an update. Saudi Journal of Ophthalmology, v. 25, n. 2, p. 99 – 111, 2011. ISSN 1319-4534. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1319453411000105>>. 17
- ALYOUBI, W. L.; SHALASH, W. M.; ABULKHAIR, M. F. Diabetic retinopathy detection through deep learning techniques: A review. Informatics in Medicine Unlocked, v. 20, p. 100377, 2020. ISSN 2352-9148. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352914820302069>>. 19
- AMORIM, L. da S. et al. Automatic segmentation of blood vessels in retinal images using 2d gabor wavelet and sub-image thresholding resulting from image partition. Research on Biomedical Engineering, 2019. 36
- CARNEIRO, T. et al. Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. IEEE ACCESS, 6, p. 61677–61685, 2018. ISSN 2169-3536. 38
- CHO, N. et al. Idf diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Research and Clinical Practice, Elsevier, v. 138, p. 271–281, 2017. 11, 12, 53
- CHOLLET, F. Deep Learning with Python. 1st. ed. USA: Manning Publications Co., 2017. ISBN 1617294438. 20, 42
- CUADROS, J.; BRESNICK, G. Eyepacs: An adaptable telemedicine system for diabetic retinopathy screening. Journal of Diabetes Science and Technology, v. 3, n. 3, p. 509–516, 2009. PMID: 20144289. Disponível em: <<https://doi.org/10.1177/193229680900300315>>. 13, 37
- DARGAN, S. et al. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. ARCHIVES OF COMPUTATIONAL METHODS IN ENGINEERING, 27, n. 4, p. 1071–1092, SEP 2020. ISSN 1134-3060. 19

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. 2009 IEEE conference on computer vision and pattern recognition. [S.l.], 2009. p. 248–255. 22, 23

DUMOULIN, V.; VISIN, F. A guide to convolution arithmetic for deep learning. 2018. 24

ETDRS. Fundus photographic risk factors for progression of diabetic retinopathy: Etdrs report number 12. Ophthalmology, v. 98, n. 5, Supplement, p. 823 – 833, 1991. ISSN 0161-6420. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0161642013380142>>. 17, 27

GARGEYA, R.; LENG, T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology, v. 124, n. 7, p. 962 – 969, 2017. ISSN 0161-6420. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0161642016317742>>. 27, 29, 32

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. [S.l.]: The MIT Press, 2016. ISBN 0262035618. 20, 21

GOOGLE RESEARCH. Frequently Asked Questions. 2021. "<<https://research.google.com/colaboratory/faq.html>>". "[Online; Acessado em 06 de Março de 2021]". 37

GULSHAN, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 2016. Disponível em: <<http://jamanetwork.com/journals/jama/fullarticle/2588763>>. 27, 29, 32, 47

GUO, Y. et al. Deep learning for visual understanding: A review. NEUROCOMPUTING, 187, n. SI, p. 27–48, APR 26 2016. ISSN 0925-2312. 19

INTERNATIONAL COUNCIL OF OPHTHALMOLOGY. ICO Guidelines for Diabetic Eye Care. [S.l.]: International Council of Ophthalmology, 2017. "<<http://www.icoph.org/downloads/ICOGuidelinesforDiabeticEyeCare.pdf>>". "[Online; Acessado em 19 de Agosto de 2019]". 11, 12, 17, 18, 19, 28, 43, 53

KALYANI, R. R.; GOLDEN, S. H.; CEFALU, W. T. Diabetes and aging: Unique considerations and goals of care. Diabetes Care, American Diabetes Association, v. 40, n. 4, p. 440–443, 2017. ISSN 0149-5992. Disponível em: <<https://care.diabetesjournals.org/content/40/4/440>>. 11

KAR, S. S.; MAITY, S. P. Automatic detection of retinal lesions for screening of diabetic retinopathy. IEEE Transactions on Biomedical Engineering, v. 65, n. 3, p. 608–618, March 2018. ISSN 0018-9294. 12

LEEZA, M.; FAROOQ, H. Detection of severity level of diabetic retinopathy using bag of features model. IET Computer Vision, v. 13, n. 5, p. 523–530, 2019. ISSN 1751-9632. 27, 29, 32, 47

LI, T. et al. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Information Sciences, v. 501, p. 511 – 522, 2019. ISSN 0020-0255. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0020025519305377>>. 13, 23, 27, 28, 32, 33, 35, 36, 38, 40, 43, 47, 48, 49, 50, 54

LUMBROSO, B.; RISPOLI, M.; SAVASTANO, M. C. Diabetic Retinopathy. 1st. ed. India: Jaypee Brothers Medical Publishers, 2015. ISBN 978-9351528982. 18

MOOKIAH, M. R. K. et al. "computer-aided diagnosis of diabetic retinopathy: A review". "Computers in Biology and Medicine", v. 43, n. 12, p. 2136 – 2155, 2013. ISSN 0010-4825. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0010482513002862>>. 12

NWANKPA, C. et al. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. 2018. 21

ORNEK, A. H.; CEYLAN, M. Comparison of Traditional Transformations for Data Augmentation in Deep Learning of Medical Thermography. In: IEEE Reg 8; IEEE Hungary Sect; IEEE Czechoslovakia Sect & SP CAS COM Joint Chapter; Sci Assoc Infocommunicat; Brno Univ Technol, Dept Telecommunicat; Budapest Univ Technol & Econ, Dept Telecommunicat & Media Informat; Czech Tech Univ Prague, Dept Telecommunicat Engn; Isik Univ, Dept Elect & Elect Engn,; Istanbul Tech Univ, Elect & Communicat Engn Dept; Josip Juraj Strossmayer Univ Osijek, Fac Elect Engn Comp Sci & Informat Technol; Karadeniz Tech Univ, Dept Elect & Elect Engn; Natl Taiwan Univ Sci & Technol, Dept Elect & Comp Engn; Seikei Univ, Grad Sch & Fac Sci & Technol, Informat Networking Lab; Slovak Univ Technol Bratislava, Inst Multimedia Informat & Commun Technologies; Escola Univ Politecnica Mataro, Tecnocampus; Tech Univ Sofia, Fac Telecommunicat; Univ Paris 8, UFR MITSIC, Lab Informatique Avancee Saint Denis; Univ Politehnica Bucharest, Ctr Adv Res New Mat Prod & Innovat Proc; Univ Ljubljana, Lab Telecommunicat; Univ Patras, Phys Dept; VSB Tech Univ Ostrava, Dept Telecommunicat; W Pomeranian Univ Technol, Fac Elect Engn. 2019 42ND INTERNATIONAL CONFERENCE ON TELECOMMUNICATIONS AND SIGNAL PROCESSING (TSP). [S.l.], 2019. p. 191–194. ISBN 978-1-7281-1864-2. 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, HUNGARY, JUL 01-03, 2019. 40

PAWARA, P. et al. Data Augmentation for Plant Classification. In: BlancTalon, J and Penne, R and Philips, W and Popescu, D and Scheunders, P (Ed.). ADVANCED CONCEPTS FOR INTELLIGENT VISION SYSTEMS (ACIVS 2017). [S.l.], 2017. (Lecture Notes in Computer Science, 10617), p. 615–626. ISBN 978-3-319-70353-4; 978-3-319-70352-7. ISSN 0302-9743. 18th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS), Antwerp, BELGIUM, SEP 18-21, 2017. 40

PIZER, S. et al. Contrast-limited adaptive histogram equalization: speed and effectiveness. [1990] Proceedings of the First Conference on Visualization in Biomedical Computing, p. 337–345, 1990. 23

PORWAL, P. et al. Idrid: Diabetic retinopathy – segmentation and grading challenge. Medical Image Analysis, v. 59, p. 101561, 2020. ISSN 1361-8415. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1361841519301033>>. 13, 27, 28, 31, 33, 36, 37, 47, 48, 49, 50, 51, 54

PRATT, H. et al. Convolutional neural networks for diabetic retinopathy. Procedia Computer Science, v. 90, p. 200 – 205, 2016. ISSN 1877-0509. 20th Conference on Medical Image Understanding and Analysis (MIUA 2016). Disponível em:

<<http://www.sciencedirect.com/science/article/pii/S1877050916311929>>. 27, 30, 38, 42, 47

QUELLEC, G. et al. Deep image mining for diabetic retinopathy screening. Medical Image Analysis, v. 39, p. 178 – 193, 2017. ISSN 1361-8415. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S136184151730066X>>. 27, 29, 32, 40

QURESHI, I.; MA, J.; ABBAS, Q. Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. MULTIMEDIA TOOLS AND APPLICATIONS, v. 80, n. 8, p. 11691–11721, MAR 2021. ISSN 1380-7501. 27

RIAZ, H. et al. Deep and Densely Connected Networks for Classification of Diabetic Retinopathy. DIAGNOSTICS, 10, n. 1, JAN 2020. 14, 27, 28, 32, 35, 37, 39, 40, 42, 47, 54

ROCHA, D. A. da et al. An unsupervised approach to improve contrast and segmentation of blood vessels in retinal images using clahe, 2d gabor wavelet, and morphological operations. Research on Biomedical Engineering, 2020. 12

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. CoRR, abs/1505.04597, 2015. Disponível em: <<http://arxiv.org/abs/1505.04597>>. 24

SALAMAT, N.; MISSEN, M. M. S.; RASHID, A. Diabetic retinopathy techniques in retinal images: A review. Artificial Intelligence in Medicine, v. 97, p. 168 – 188, 2019. ISSN 0933-3657. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0933365718300630>>. 17

SCOTLAND, G. S. et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. British Journal of Ophthalmology, BMJ Publishing Group Ltd, v. 94, n. 6, p. 712–719, 2010. ISSN 0007-1161. Disponível em: <<https://bj.o.bmj.com/content/94/6/712>>. 12

SHANTHI, T.; SABEENIAN, R. Modified alexnet architecture for classification of diabetic retinopathy images. Computers & Electrical Engineering, v. 76, p. 56 – 64, 2019. ISSN 0045-7906. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0045790618334190>>. 27, 30, 32

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for Deep Learning. JOURNAL OF BIG DATA, 6, n. 1, JUL 6 2019. 13, 22

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014. 22, 23, 42

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. Information Processing & Management, v. 45, n. 4, p. 427 – 437, 2009. ISSN 0306-4573. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306457309000259>>. 44

SZE, V. et al. Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, v. 105, 03 2017. 21

TAJBAKHSI, N. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Transactions on Medical Imaging, v. 35, n. 5, p. 1299–1312, 2016. 22

WAN, S.; LIANG, Y.; ZHANG, Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. Computers & Electrical Engineering, v. 72, p. 274 – 282, 2018. ISSN 0045-7906. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0045790618302556>>. 23, 27, 30, 32, 38, 40, 47

WANG, R. et al. Weakly supervised lesion detection from fundus images. IEEE Transactions on Medical Imaging, v. 38, n. 6, p. 1501–1512, June 2019. ISSN 0278-0062. 12

WORLD HEALTH ORGANIZATION. Global report on diabetes. [S.l.]: World Health Organization, 2014. "<<https://apps.who.int/iris/handle/10665/204871>>". "[Online; Acessado em 20 de Junho de 2019]". 11, 53

ZHANG, W. et al. Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowledge-Based Systems, v. 175, p. 12 – 25, 2019. ISSN 0950-7051. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705119301303>>. 27, 30, 32, 40, 44

ZUIDERVELD, K. J. Contrast limited adaptive histogram equalization. In: Graphics Gems. [S.l.: s.n.], 1994. 22, 40, 43