



UNIVERSIDADE FEDERAL DO MARANHÃO
Curso de Ciência da Computação

Gabriel Phelipe Costa Freitas

Diagnóstico de Glaucoma via Deep Learning

São Luís

2022

Gabriel Phelipe Costa Freitas

Diagnóstico de Glaucoma via Deep Learning

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Geraldo Braz Júnior

São Luís

2022

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

Freitas, Gabriel Phelipe Costa.

Diagnóstico de Glaucoma via Deep Learning / Gabriel
Phelipe Costa Freitas. - 2022.

38 f.

Orientador(a): Geraldo Braz Júnior.

Curso de Ciência da Computação, Universidade Federal do
Maranhão, SÃO LUIS - MA, 2022.

1. Glaucoma. 2. Processamento de Imagens. 3. Rede
Neural. I. Braz Júnior, Geraldo. II. Título.

Gabriel Phelipe Costa Freitas

Diagnóstico de Glaucoma via Deep Learning

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho _____ em São Luís, 04 de agosto de 2022:

Prof. Dr. Geraldo Braz Júnior
Orientador

Prof. Dra. Simara Vieira Rocha
Examinador

Profa. Dra. João Dallyson Sousa de Almeida
Examinador

São Luís
2022

Agradecimentos

Nesta caminhada agradecimentos não poderiam faltar.

Primeiramente agradeço a Deus, por ter sido a minha força nos momentos mais difíceis e o meu apoio em todas as fases da minha vida.

A minha mãe, por ter sido minha fortaleza e minha fonte de amor incondicional durante toda a minha vida.

Ao meu pai, por também não medir esforços para que meu sonho se tornasse realidade.

A minha namorada, por todo carinho, apoio e incentivo a continuar essa caminhada mesmo nos momentos mais difíceis.

Ao meu orientador, o Prof. Dr. Geraldo Braz Jr. por toda atenção, paciência, dedicação e por todos os ensinamentos.

A minha família, por sempre proferirem palavras de conforto e de carinho. Aos meus amigos, por todo o carinho e suporte em momentos de fraquezas e alegrias.

A todo corpo docente do curso de Ciência da Computação, por tantos ensinamentos passados a cada período de aprendizado, nos motivando a buscar sempre o melhor. Por fim, agradeço a todos que me ajudaram ao longo dessa jornada. Em memória a Ednara.

*"Não fiques em terreno plano.
Não subas muito alto.
O mais belo olhar sobre o mundo
Está a meia encosta."*

Friedrich Nietzsche, em *"A Gaia Ciência"*

Resumo

Glaucoma é uma doença que promove perda adquirida de células da retina e axônios que resulta em uma lesão gradual da visão, sendo a segunda principal causa de cegueira permanente no mundo. Conforme a Organização Mundial da Saúde (OMS), aproximadamente 285 milhões de pessoas são afetadas pela doença, sendo que 60% à 80% dos casos podem ser prevenidos e remediados. Têm sido propostos diversos sistemas automatizados de diagnóstico de glaucoma, porém, é complexo com a variedade de imagens retinográficas, o que torna tais métodos inviáveis ou pouco eficazes para a aplicação em programas de imagens. O presente trabalho tem por objetivo desenvolver um método computacional para diagnóstico de glaucoma via redes neurais convolucionais. O método ainda faz uso de técnicas como segmentação e *Data Augmentation*. O emprego das técnicas de pré-processamento com *Data Augmentation* feitos nesta pesquisa, mostram-se eficazes para o uso em bases desbalanceadas, por permitir maior desempenho no aprendizado e diminuir o overfitting no emprego de CNNs. Trabalhou-se com 2488 imagens, sendo 2101 imagens classificadas como normais e 387 afetadas pelo glaucoma. Os resultados obtidos pelo método proposto atingiram, para os parâmetros de Sensibilidade, Precisão e F1-Score respectivamente, valores de 96%, com uma variação aproximada de 0,8% para a classe Normal e para o Glaucoma, 95% em média, com desvios de 1,4%, 1,3% e 0,5%, respectivamente, para as métricas citadas.

Palavras-chave: Glaucoma, Rede Neural, Processamento de Imagens.

Abstract

Glaucoma is a disease that promotes acquired loss of retinal cells and axons which results in a gradual damage to vision, being the second leading cause of permanent blindness in the world. According to the World Health Organization (WHO), approximately 285 million people are affected by the disease and about 60% to 80% of cases can be prevented and remedied. Several have been proposed Automated glaucoma diagnostic systems, however, to deal with the variety of retinographic images it's a complex issue, which makes such methods unfeasible or little efficient for it's application in imaging programs. The present work aims to develop a computational method to execute pre-processing techniques, such as Image segmentation and Data Augmentation, on fundus images of normal and pathological eyes, and apply the dataset into a Convolutional Neural Network, for automatically classification and results evaluation. The pre-processing techniques with Data Augmentation made in this research has been proofed to be a efficient when dealing with unbalanced bases, for greater performance in learning and decrease or overfitting when using the CNN model. We worked with 2488 images, with 2101 images classified as normal and 387 classified as glaucoma. The results obtained were promissory for both, normal and pathological classes. For the Recall, Precision and F1-Score data parameter that were considered "Normal", the hit rate was about 96%, with a variation of only 0.8%. Meanwhile for the affected by the disorder class, there was a positive rate of 95% on average, with standard deviation of 1.4%, 1.3% and 0.5%, respectively, for the aforementioned reference measures.

Keywords: Glaucoma, Neural Network, Image Processing.

Lista de ilustrações

Figura 1 – Olho - Corte Transversal	15
Figura 2 – Comparação - Olho Normal / Olho com Glaucomal	17
Figura 3 – Imagem Digital no plano cartesiano	18
Figura 4 – Fluxo do processamento de imagens	18
Figura 5 – Convolutional Neural Network (CNN)	20
Figura 6 – Exemplo de Convolução	21
Figura 7 – Exemplo de Convolução - Passo 1	21
Figura 8 – AVG Pooling	22
Figura 9 – Max Pooling	22
Figura 10 – Esquema com Dropout	23
Figura 11 – Metodologia	25
Figura 12 – Cubo RGB	26
Figura 13 – a: colorida, b: grayscale, c: green channel	26
Figura 14 – Oversampling	27
Figura 15 – canal verde	28
Figura 16 – Grayscale	28
Figura 17 – Threshold	28
Figura 18 – Contorno	28
Figura 19 – Área capturada	28
Figura 20 – Disco óptico	28
Figura 21 – CNN	30
Figura 22 – Gráfico de Loss e Acurácia	33
Figura 23 – Gráfico de Loss e Acurácia com Smoothing	34

Lista de tabelas

Tabela 1 – Classificação sem a segmentação e com menos layers na CNN	32
Tabela 2 – Resultados finais da classificação	33
Tabela 3 – Comparação de Resultados	34

Lista de abreviaturas e siglas

PNS	<i>Pesquisa Nacional de Saúde</i>
OMS	<i>Organização Mundial da Saúde</i>
CNN	<i>Rede Neural Convolutiva</i>
ODIR	<i>Reconhecimento Inteligente de Doenças Oculares</i>
RGB	<i>Red-Green-Blue</i>
CCIR	<i>União Internacional de Telecomunicação</i>
VP	<i>Verdadeiros Positivos</i>
FP	<i>Falsos Positivos</i>
VN	<i>Verdadeiros Negativos</i>
FN	<i>Falsos Negativos</i>

Sumário

1	INTRODUÇÃO	12
1.1	Objetivo	13
1.1.1	Objetivos Específicos	13
1.2	Contribuições	14
1.3	Organização do Trabalho	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Anatomia Ocular	15
2.2	Glaucoma	16
2.3	Processamento de Imagens	17
2.4	Deep Learning	18
2.4.1	Perceptron e Multilayer Perceptron (MLP)	19
2.4.2	Redes Neurais Convolucionais (CNN)	20
2.4.2.1	Camadas Convolucionais	20
2.4.2.2	Camadas de Pooling	21
2.4.2.3	Camadas Totalmente Conectadas	22
2.5	Dropout	23
2.6	Aumento de Dados	23
3	MATERIAIS E MÉTODO	25
3.1	Aquisição de Imagens	25
3.2	Pré-Processamento	26
3.3	Segmentação do Disco Ótico	27
3.4	Arquitetura do Modelo CNN	29
3.5	Métricas de Avaliação	30
4	RESULTADOS	32
4.1	Comparação dos Resultados	34
5	CONCLUSÃO	35
	REFERÊNCIAS	36

1 Introdução

A Pesquisa Nacional de Saúde (PNS) de 2013 revelou que dentre as deficiências físicas pesquisadas no Brasil, a com mais representatividade na população foi a visual, registrando maior proporção estimada para as pessoas que a adquiriram por doença ou acidente (SILVA et al., 2018). Segundo pesquisas de 2021 da Organização Mundial da Saúde (OMS), ao tratar sobre distúrbios oculares à nível global, estima-se que mais de 800 milhões de pessoas possuem deficiências ópticas relacionadas à distância, tais como hipermetropia ou miopia (que podem ser tratadas com um par de óculos apropriados), e mais de 100 milhões possuem deficiências que são consideradas moderadas ou graves, que podem também causar a cegueira. Ainda segundo a OMS, espera-se um aumento desses números, principalmente levando em conta os fatores de risco que acompanham o envelhecimento da população e outros agentes, tais como a mudanças no estilo de vida da população mais jovem.

No cenário supracitado, tem-se que o Glaucoma, que consiste em uma neuropatia óptica com repercussão característica no campo visual, é classificado como a principal causa de cegueira irreversível e a segunda maior causa de cegueira no mundo, sendo superado somente pela catarata (LOUREIRO; FÉLIX, 2020). Cada tipo de glaucoma, seja de ângulo aberto, fechado ou congênito, possui sua própria peculiaridade, mas independente disso, suas origens são provenientes do surgimento de uma lesão no nervo óptico, além de causar uma elevação na pressão ocular. Durante o “percurso evolutivo” da doença, ela permanece assintomática, o que torna o diagnóstico precoce algo ainda mais desafiador, e os sintomas mais graves, como a perda visual gradativa, ocorre em estágios mais avançados, comprometendo, logo no seu início, principalmente a visão periférica. De forma generalizada, a doença costuma aparecer a partir dos 40 anos, mas isso não a impede de ocorrer e se manifestar em qualquer faixa etária, já que pode existir uma dependência direta com o motivo causador do aumento da pressão intraocular. Entre os principais fatores de risco, tem-se a diabetes, fundos hereditários, a idade avançada e até traumas oculares.

Tendo em vista as intensas consequências da doença e a dificuldade de um diagnóstico precoce, tanto pelos seus sintomas quanto pelo desconhecimento em si, que por certo poderia favorecer o seu tratamento, a tecnologia surgiu como um aliado eficiente. De fato, há tempos os temas tecnologia e inovação tecnológica estão presentes em todos os meios de comunicação e evidentemente existe a difusão de que se vive em uma era histórica de desenvolvimento tecnológico. Já é um fato comum que clínicas médicas e hospitais em geral apostem no fator "recurso tecnológico" como uma forma de otimizar tarefas cotidianas (pequenas ou até complexas). O uso frequente da tecnologia no âmbito hospitalar, também abriu inúmeras portas para o avanço no tratamento, diagnóstico e até na cura de

diversas patologias. Nesse mesmo domínio, também pode-se encontrar trabalhos voltados à inteligência artificial e aprendizado de máquina que podem contribuir significativamente no possível diagnóstico de doenças, o que conseqüentemente traz melhores resultados durante o tratamento dessas patologias, visto que o diagnóstico precoce também permite uma intervenção prévia.

Levando em conta a importância na otimização e automação no diagnóstico de doenças, esse trabalho visa, por meio de técnicas de Processamento de Imagem e Aprendizado de Máquina, realizar o diagnóstico por imagem da presença de glaucoma na mesma com a finalidade de disponibilizar um método automático para este ponto.

Com finalidade semelhante, Carvalho et al. (2019) utilizaram descritores de textura (GLCM e LBP) e para a classificação o método *XGBoost*, valendo-se de uma base de dados com 200 imagens de glaucoma e 255 imagens saudáveis. Os resultados foram em torno de 81,71% de *acurácia* 77,61% para a *sensibilidade* e 85,24% de *especificidade*.

Na abordagem de Raghavendra et al. (2018) foi elaborado um modelo que contou com o uso de uma base com 1.426 imagens de fundo de olho e uma CNN com 80 camadas, e seu método atingiu, avaliado pelas métricas de *acurácia*, *sensibilidade* e *especificidade*, valores de 98,13%, 98% e 98,30%, respectivamente.

Moreira et al. (2021) aplicaram técnicas de pré-processamento, tais como a especificação de histograma e o método de CLAHE e outras, para a remoção de ruídos e para o destaque na textura das imagens e por fim usou a Rede Neural de Capsulas (CAPSNET) para o aprendizado. Contando com uma base com 2.086 imagens de retinografias, 848 de pacientes com glaucoma e 1.227 com imagens saudáveis, obtiveram valores de 85,29% de *acurácia*, 88,52 % de *especificidade* e 81,77% de *sensibilidade*.

Já Li et al. (2018), optaram pelo uso de uma Rede Neural Convolutiva utilizando a arquitetura Inception-V3 e uma base de dados que foi separada com 31.745 imagens para treino e 8.000 imagens para teste, obtendo uma *acurácia* de 98,6% *sensibilidade* de 95,6% e *especificidade* de 92,2%.

1.1 Objetivo

O objetivo geral desse trabalho é o desenvolvimento de um método computacional para realizar o diagnóstico de glaucoma em imagens de fundo de olho e redes neurais convolucionais com a finalidade de auxiliar no processo de detecção antecipada da patologia.

1.1.1 Objetivos Específicos

Destaca-se como objetivos específicos deste trabalho:

- Realizar o pré-processamento nas imagens de fundo de olho através de técnicas como a Segmentação e Data Augmentation;
- Realizar o aprendizado de máquina através de uma Rede Neural Convolutacional ajustada ao problema;
- Avaliar o método proposto por meio de experimentos e comparações com outros métodos já existentes na literatura.

1.2 Contribuições

Esse trabalho tem como contribuição:

- A proposta de um método automatizado de classificação do Glaucoma através de técnicas de processamento de imagens e aprendizado de máquina;
- Utilização de uma arquitetura de CNN com uma quantidade rasa de camadas, para ser aplicada em uma base pequena e desbalanceada de dados.
- Utilização de Data Augmentation e Pré-Processamento de imagens de uma base de dados desbalanceada no processo de classificação de patologias oculares;
- Utilização da técnica de Dropout na rede neural convolutacional, para diminuir Overfitting no treino de uma base desbalanceada.

1.3 Organização do Trabalho

Este trabalho está organizado em cinco capítulos, de forma a apresentar o conteúdo mais claramente, conforme os parágrafos a seguir. No capítulo 2 se encontra a fundamentação teórica utilizada como base para a construção do estudo e para melhor compreensão do assunto. No capítulo 3 se encontra a metodologia proposta para a elaboração desse estudo. No capítulo 4 é apresentado e discutido os resultados obtidos nesse trabalho. No capítulo 5 é apresentada a conclusão e os trabalhos futuros.

2 Fundamentação Teórica

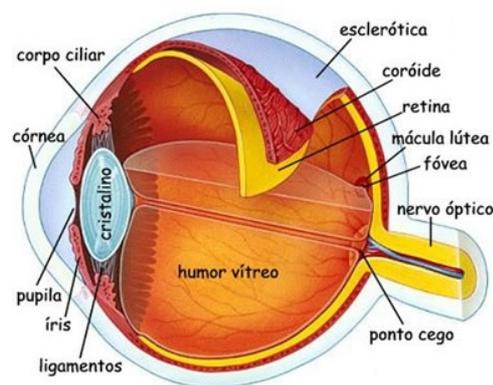
Neste capítulo é apresentado a fundamentação teórica utilizada como base para a construção da metodologia. As seções a seguir abordam os principais temas relacionados ao trabalho, tais como a anatomia ocular e o glaucoma, além de atuarem como um prefácio à abordagem escolhida durante o processo, falando sobre métodos de extração de características e de processamento de imagens.

2.1 Anatomia Ocular

Em essência, o olho humano consiste em um órgão complexo, responsável por captar a luz, projetando-a sobre a retina, transformar a informação luminosa em um impulso elétrico (através de reações químicas) e transmiti-lo, por meio do nervo óptico, para o cérebro, onde a informação capturada será finalmente processada (RAMOS, 2006).

Quanto a sua estrutura, o olho possui membranas e algumas camadas. De forma simplificada, a camada mais externa é composta pela esclera e a córnea, a camada mais inferior é composta pelo Coroide, íris e Corpo Ciliar e na camada subsequente podemos encontrar a retina (RAMOS, 2006).

Figura 1 – Olho - Corte Transversal



Fonte: <<https://www.todoestudo.com.br/biologia/olhos>>

Dentre a estrutura do olho, a Esclera é o nome do envoltório externo e branco que envolve o olho. É a estrutura que dá forma ao globo ocular (RAMOS, 2006). A córnea é a primeira estrutura do olho que a luz atinge. Sua camada mais externa, chamada de Epitélio, possui uma elevada taxa regenerativa para lesões superficiais e as outras quatro camadas mais internas, além de proporcionar mais rigidez, protegem o olho de infecções (RAMOS, 2006). A coróide é a camada intermediária do globo ocular. Consiste numa

região bem vascularizada, e exerce a função de fornecer oxigênio e nutrientes para as células da retina (CAPELINI et al., 2019). Para a porção visível, que se encontra logo atrás da córnea, dá-se o nome de íris. Constituída por músculos que estão dispostos de forma que se permite aumentar ou diminuir o tamanho da pupila, regulando a quantidade de luz que o olho pode receber, conforme as condições ambientes de luminosidade (RAMOS, 2006).

Já atrás da íris, podemos encontrar o corpo ciliar, fração do globo ocular que possui a função de produzir a substância aquosa que é responsável por manter a pressão ocular. Além disso, o corpo ciliar também é responsável por controlar a alternância da convergência do cristalino (outra estrutura ocular), o que permite que exista uma acomodação das imagens de corpos à curta e longa distância em relação ao olho (CAPELINI et al., 2019).

A outra estrutura que compõe o olho é a retina, que consiste na membrana interna em volta dos olhos que recebe a luminosidade que foi focalizada pelo cristalino. Esta estrutura contém fotorreceptores que têm como responsabilidade transformar a luz em impulsos elétricos, para que o cérebro possa “decodificar” como imagens (RAMOS, 2006). Algumas fibras nervosas da retina se unem e então formam o que é denominado de nervo óptico, Este nervo é responsável por enviar as informações que foram captadas pela retina até o cérebro (CAPELINI et al., 2019)

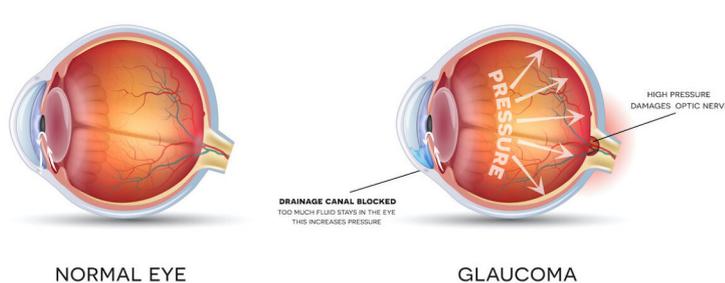
2.2 Glaucoma

O glaucoma é uma neuropatia óptica com múltiplos fatores que funcionam como agentes e é caracterizada pela lesão progressiva do nervo óptico, com um consequente impacto no campo visual do indivíduo afetado. Como já citado anteriormente, existem diferentes tipos de glaucoma. O mais comum é do tipo primário de ângulo aberto, também conhecido como glaucoma crônico simples. Nesse caso, os canais gradativamente se estreitam, dificultando a drenagem adequada do líquido, que por sua vez, ao se acumular, causa a elevação da pressão intraocular. O aumento da pressão no globo ocular, é um de seus principais fatores de risco (URBANO et al., 2003).

Apesar de todos esses fatores, raramente pacientes com glaucoma apresentam seus sintomas desde o início da patologia. Geralmente, somente após a perda de mais de 40% das fibras neurais, os afligidos pela doença poderão sentir uma perda gradativa da visão periférica (RODRIGUES, 2018).

Quanto ao diagnóstico da doença, tem-se que quanto mais cedo for detectada, melhores serão as chances do seu tratamento. Entre algumas exemplificações de exames que podem ser feitos para a obtenção desse diagnóstico, podem ser citados o de acuidade visual, que identifica alterações na visão; o exame de pupila, que percebe lesões nas vias ópticas (o que inclui o nervo óptico) e o exame com lâmpada de fenda, que examina o interior e o exterior dos olhos.

Figura 2 – Comparação - Olho Normal / Olho com Glaucomal



Fonte: <<https://coastaleyassociates.com/glaucoma/>>

Como já expressei na seção de introdução deste trabalho e frisado em parágrafos anteriores, o diagnóstico do glaucoma pode não ser uma tarefa simples. É ainda mais desafiador obter o diagnóstico precoce da doença. Por mais que seja uma doença que não possua cura, existem diversas maneiras de tratá-la para que o grau de perda da visão possa ser controlado e também para manter a melhor qualidade de vida possível para o paciente acometido pela doença. Para isso, os tratamentos costumam ser individualizados e levam sempre em conta fatores de risco, tais como a evolução da doença, a idade do indivíduo, o seu histórico familiar e a espessura da córnea.

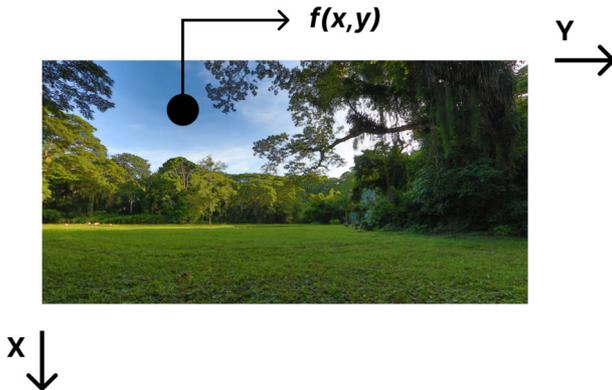
Embora o dano estrutural glaucomatoso possa ser avaliado subjetivamente examinando o nervo óptico e a fibra nervosa da retina peripapilar, a introdução de imagens oculares no manejo clínico tem contribuído significativamente a avaliação objetiva e quantitativa suplementar da estrutura ocular (BUSSEL; WOLLSTEIN; SCHUMAN, 2014).

2.3 Processamento de Imagens

Técnicas voltadas para o processamento de imagens digitais e suas aplicações, que vão desde programas espaciais até o auxílio em diversas áreas da saúde (como mostra o trabalho de MIRANDA) são elementos já diluídos no presente cenário. Refere-se à imagem, como uma função bidimensional de intensidade da luz, denotada por $f(x, y)$, onde "x" e "y" são as coordenadas espaciais de um ponto qualquer na imagem, podendo os componentes em questão assumir quaisquer valores, sendo eles discretos e finitos, e a função f em qualquer ponto dentro desse par, é o valor de equivalência do brilho ou dos níveis de cinza que aquele ponto específico carrega dentro da imagem (GONZALEZ; WOODS, 2000), conforme é apresentado na Figura 3.

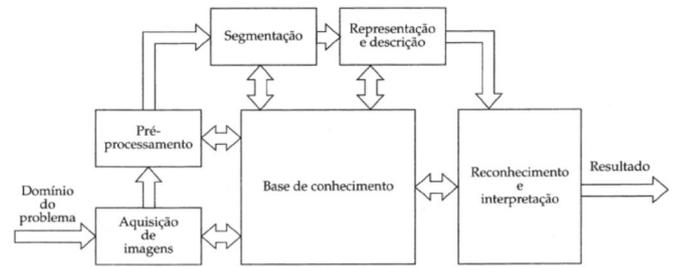
Gonzalez e Woods (2000) apresenta uma sequência de passos que são considerados como fundamentais para o autor, conforme mostra a ilustração na Figura 4. Dessa forma, o autor divide as técnicas, de forma generalizada em algumas etapas.

Figura 3 – Imagem Digital no plano cartesiano



Fonte: <<https://pxhere.com/pt/photo/1629865>>

Figura 4 – Fluxo do processamento de imagens



Fonte: GONZALEZ; WOODS

A primeira etapa seria a *aquisição da imagem*, ou seja, a etapa que tem como responsabilidade adquirir uma ou mais imagens digitais utilizando algum sensor, como câmeras de captura. A segunda etapa definida pelo autor, é a de *processamento de imagens*. A etapa propriamente dita, visa como interesse principal melhorar a imagem e melhorar as chances de sucesso de encontrar o objeto de interesse em etapas futuras. Geralmente esse passo também envolve o uso de outras técnicas, para que se possa passar para o próximo estágio, chamado de *Segmentação*. Nesta fase a imagem de entrada é dividida em subprodutos que contém objetos de interesse. Se por um lado a etapa de segmentação feita de forma assertiva contribui significativamente com os resultados, por outro lado, a mesma etapa, se mal elaborada, pode atrapalhar significativamente e levar o processo à falha. Seguindo adiante com o fluxo do processamento, pode-se então encontrar a etapa de *Representação e Descrição*, onde é selecionado características e atributos que irão gerar uma informação com valores de interesse, auxiliando na discriminação de uma classe para outra. Chegando assim na etapa final, de *Reconhecimento e Interpretação*, onde cada objeto será reconhecido por suas particularidades e características extraídas para que assim possam ser reunidos em algum dos grupos de classes ou rótulos obtidos na classe anterior.

2.4 Deep Learning

As Redes Neurais Artificiais nasceram com a premissa de que existe de fato uma semelhança entre a forma em que um computador processa a informação e a forma em que um humano realiza essa ação, (OLIVEIRA, 2019) permitindo a existência, seguindo a mesma lógica, de uma analogia entre o cérebro, um sistema complexo, não-linear e com bilhões de neurônios e um computador.

Neste mesmo contexto, nasce o Deep Learning, modelo arquitetural de aprendizagem profunda inspirada na organização do cérebro humano, que é superior ao computador em variadas atividades. O modelo segue uma estrutura hierárquica, dessa forma o aprendizado de atributos de alto nível é constituído pelo aprendizado das propriedades de níveis mais baixos (AGUIAR, 2017). Modelos de Deep Learning usam o conceito de retropropagação do erro, permitindo certa autonomia para a rede, que “visualiza” estruturas complexas em grandes volumes de dados e muda seus hiperparâmetros no decorrer das camadas, para poder haver uma representação da camada anterior (LECUN; BENGIO; HINTON, 2015).

Outra característica deste tipo de arquitetura, que foi utilizada também como base neste estudo, é a aprendizagem por experiência. O conjunto de treino possui inúmeras iterações com o modelo, iterações essas chamadas de Épocas. Ou seja, chama-se de “Época” todas as vezes que um conjunto de treino percorrer o modelo diversas vezes. Ainda segundo Aguiar (2017), esses dois fatores (hierárquico e de experiência), fazem diminuir a necessidade de intervenção humana no aprendizado (AGUIAR, 2017)

2.4.1 Perceptron e Multilayer Perceptron (MLP)

O neurônio artificial é a unidade básica das redes neurais artificiais, funciona com uma função matemática que, de acordo com a somatório ponderado dos valores de entrada, fornece uma saída (OLIVEIRA, 2019). O *Perceptron* é um algoritmo feito para treinar um neurônio, através da aprendizagem pela correção de erros, possui muitas limitações quando em comparação às redes mais modernas. Um exemplo claro dessa barreira existente é o caso dos operadores lógicos. Para os operadores AND e OR, é possível achar uma solução linear, e o perceptron é capaz de lidar com tais problemas. Mas para o caso do operador XOR, por não ser linearmente separável, faria com que o algoritmo do Perceptron entrasse em um loop infinito, por nunca ter suas condições de parada satisfeitas (AGUIAR, 2017).

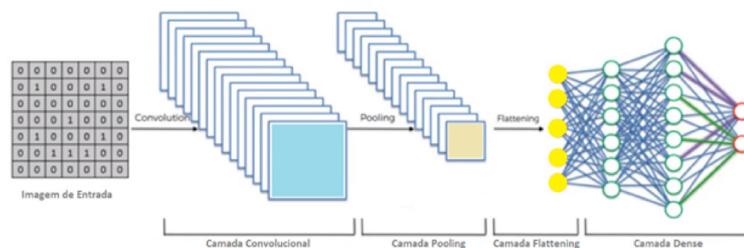
Não obstante, as redes neurais multicamadas (*Multilayer Perceptron* ou MLP) são perfeitamente capazes de aprender casos não-lineares como o do XOR. Estas MLPs funcionam como aproximadores universais que usam a minimização de erro como forma de aprendizado (AGUIAR, 2017). Esta rede é fundada por várias camadas e cada camada contém múltiplos neurônios. A primeira camada recebe a entrada da rede, as mais internas são chamadas de *hidden Layers*, ou camadas ocultas, enquanto a última é a camada de *output* ou camada de saída. Os neurônios na camada de saída desta rede, são associados às distintas classes do problema em estudo e também aos neurônios da camada anterior. Dessa forma, o neurônio da camada de entrada aprende uma função que traça um hiperplano que separa os dados e os neurônios das camadas seguintes estabelecem um grupo dos hiperplanos feitos por camadas anteriores, que ao serem passados pra camadas posteriores vão se formar novos grupos também com estes novos hiperplanos, formando regiões

convexas que adiante também serão agrupadas e permitirão formar novas regiões em formatos arbitrários (FACELI et al., 2011).

2.4.2 Redes Neurais Convolucionais (CNN)

As Redes Neurais Convolucionais, que foram introduzidas em (LECUN et al., 1998) (LECUN et al., 1998), são inspiradas no processo biológico de processamento de dados virtuais. Ao longo de seu processamento, essas redes aplicam filtros em seus dados sem perder a relação de vizinhança dos pixels da imagem (OLIVEIRA, 2019). Além disso, essas redes possuem partes multifuncionais, sendo composta tanto pela fase da extração de característica, quanto pela classificação. De forma geral, a etapa de extração de características é constituída por layers de convolução e de *Pooling*, e na fase de classificação é formada por uma camada de neurônios completamente conectados, que consiste no modelo MLP citado na Seção 2.4.1, conforme mostra a Figura 5.

Figura 5 – Convolutional Neural Network (CNN)



Fonte: <<https://medium.com/>>

2.4.2.1 Camadas Convolucionais

Nesta camada, através de pequenos blocos, definidos como filtros, é feita a extração e o mapeamento do conteúdo da imagem, que é representada por uma matriz de pixels, sem perder a informação de ordem da imagem (DUMOULIN; VISIN, 2016).

Destacam-se algumas nomenclaturas importantes durante esse processo. Para a matriz correspondente à imagem de entrada, dá-se o nome de *input feature maps* já a sua correspondente no final do processo de convolução, denomina-se *output feature maps*. Além disso, também temos a matriz de filtro ou *Kernel*, relacionada aos blocos citados no parágrafo anterior(DUMOULIN; VISIN, 2016).

Ainda neste contexto, durante o processo a matriz de filtros e a *input feature maps* são apostas uma sobre a outra (passo denominado de *stride*) de forma que o produto dessa interação é somado e o resultado obtido é posicionado na matriz de *output feature maps* em uma localização correspondente. Para cada filtro aplicado,tais como os detectores de borda ou o Blur, pode-se ter um valor diferente de tamanho do *kernel*.(DUMOULIN; VISIN, 2016).

Para exemplificações, tomemos como base a matriz 5×5 de pixels utilizados por Oliveira (2019), que recebem valores entre 0 e 1 e o kernel 3×3 , que após a sequência do primeiro passo, geram uma matriz de *output*, também de tamanho 3×3 , como mostra a Figura 6

Figura 6 – Exemplo de Convolução

$$\begin{array}{ccc}
 \text{Matriz de Pixels} & \text{Kernel} & \text{Matriz de Saída} \\
 \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} & = \begin{bmatrix} - & - & - \\ - & - & - \\ - & - & - \end{bmatrix}
 \end{array}$$

Fonte: Gerada pelo autor

Assim, como ilustrado na Figura 7 e explicado em parágrafos anteriores, na primeira execução de passos é calculado o valor da soma dos produtos das matrizes que foram sobrepostas, gerando uma simples equação que se constitui em $(0 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 1 + 1 \cdot 0)$, resultando no valor 2, que será o valor equivalente à posição correspondente na matriz de saída

Figura 7 – Exemplo de Convolução - Passo 1

$$\begin{array}{ccc}
 \text{Matriz de Pixels} & \text{Kernel} & \text{Matriz de Saída} \\
 \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} & = \begin{bmatrix} 2 & - & - \\ - & - & - \\ - & - & - \end{bmatrix}
 \end{array}$$

Fonte: Gerada pelo autor

2.4.2.2 Camadas de Pooling

A operação de Pooling é responsável por receber a saída da camada convolucional, ou seja, o mapa de características, e condensar esses valores (SAHA, 2018), reduzindo o tamanho do mapa utilizando alguma métrica como base. Algumas dessas medidas de uso comum são o *AVG Pooling* e o *Max Pooling*. Para este estudo, foi considerado somente o segundo método na *CNN* proposta.

- *AVG Pooling*: para esse método, o valor de saída na matriz reduzida corresponde à média dos valores dos pixels vizinhos definidos pela grade de *Pooling* (DUMOULIN; VISIN, 2016).

Tomando como exemplo novamente uma matriz 5×5 e uma camada de pooling 3×3 , $\text{stride} = 1$, seria gerado uma matriz de *AVG Pooling* 3×3 como saída. Conforme

Figura 8 – AVG Pooling

Matriz de Pixels	Matriz de AVG Pooling
$\begin{bmatrix} 2 & 3 & 1 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.8 & - & - \\ - & - & - \\ - & - & - \end{bmatrix}$

Fonte: Gerada pelo autor

mostra na Figura 8, na primeira interação, teria-se que o valor do primeiro elemento da matriz de saída é obtido pela média dos elementos vizinhos que são definidos pela grade de Pooling escolhida, no caso 3 x 3. Nesse caso, tem-se que a média dos valores dos pixels (2, 3, 1, 0, 4, 1, 2, 3, 1) seria **1,8**.

- *Max Pooling*: Neste tipo de *pooling*, o valor de saída na matriz reduzida corresponde simplesmente ao maior valor entre os pixels vizinhos definidos pela grade de *Pooling* (DUMOULIN; VISIN, 2016).

Figura 9 – Max Pooling

Matriz de Pixels	Matriz de Max Pooling
$\begin{bmatrix} 2 & 3 & 1 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 4 & - & - \\ - & - & - \\ - & - & - \end{bmatrix}$

Fonte: Gerada pelo autor

Por exemplo, seja também uma matriz de entrada com os mesmos valores citados na matriz do item anterior e com os mesmos tamanhos referenciados, 5 x 5 e uma operação de pooling 3 x 3 sendo o $stride = 1$, para a primeira interação, teria-se o valor do primeiro elemento na matriz de saída no primeiro *stride* como o maior valor entre o conjunto de amostras na matriz de entrada, como mostra na Figura 9, ou seja: $Max(2,3,1,0,4,1,2,3) = 4$

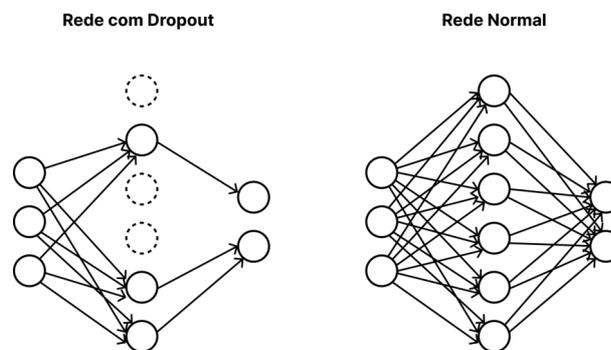
2.4.2.3 Camadas Totalmente Conectadas

A camada em questão, dentro de uma CNN, posiciona-se na etapa final do processo de convoluções e *poolings*, sendo responsável pela classificação dos resultados. Geralmente, usa-se a MLP nesta camada, que recebe as *features* de saída da CNN, depois da execução das convoluções e de *poolings*, como valores de entrada para então realizar a classificação.

2.5 Dropout

Muitas redes neurais são treinadas utilizando uma base de dados relativamente pequenas, o que pode causar um *Overfitting* durante o treino dos modelos o que pode os levar à uma performance baixa durante validação do aprendizado. O *Overfitting* acontece quando durante os treinos o modelo possui um desempenho surpreendente, mas o mesmo não acontece quando os dados são repassados e testados na fase de validação (HAWKINS, 2004). Em síntese, consiste em quando a rede aprende tão bem as relações dos dados durante a fase de treino, que apenas memoriza o que deve ser feito e ao receber dados diferentes na fase de teste e aplicar as regras que foram decoradas, o desempenho é completamente afetado. Os mecanismos de regularização mais utilizados para evitar que é o Data Augmentation, citado na Seção 2.6 deste estudo, e o uso do *Dropout*.

Figura 10 – Esquema com Dropout



Fonte: Gerada pelo autor

Definimos o termo *Dropout* se refere à descartagem de neurônios na rede neural. Seguindo os estudos feitos por Srivastava (2013) e conforme mostra o esquema da Figura 10 pode-se inicialmente eliminar, aleatória e temporariamente, unidades da camada oculta da rede, sem interferir nos neurônios de entrada ou saída. Dessa forma, os dados de entrada percorrem a rede modificada e os seus resultados de saída são retropropagados através da rede. Depois de fazer isso para um mini-lote de exemplos, os pesos e outros vieses são atualizados e o processo se repete por mais algumas vezes. No final de todo o processo, a rede terá aprendido um outro conjunto de pesos e vieses para então ser executada normalmente, com todos os seus neurônios novamente ativos.

2.6 Aumento de Dados

Existem diversas formas para enfrentar os problemas que estão associados à uma quantidade limitada de dados no contexto do aprendizado de máquina. O aumento de imagens é uma técnica muito útil, por permitir o aumento do tamanho da base de treinamento sem necessariamente exigir a aquisição de uma base de dados nova ou de

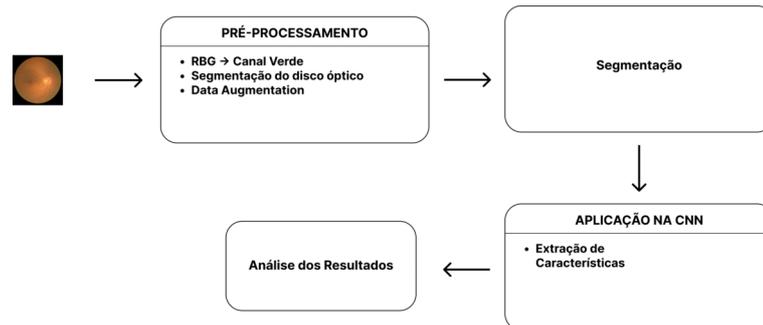
novas imagens. A ideia consiste basicamente em duplicar as imagens já existentes na base utilizada com variância o suficiente para que o modelo possa aprender com mais modelos, preservando as particularidades que são primordiais para que se possa fazer as predições, mas reorganizando os pixels de forma o suficiente para produzir algum ruído na nova imagem gerada.

Dentre as técnicas de aumento de dados, existem as chamadas transformações geométricas Claro et al. (2020). A primeira técnica a ser tratada, é a de *Flipping*. Esta é uma técnica de espelhamento da imagem original, que pode ser invertida horizontalmente ou verticalmente. A inversão do eixo horizontal é mais comum do que a inversão do eixo vertical. Este método é um dos que possuem a mais simples implementação. Outro método que convém citação é o de *Rotação*. Para esse método, são feitas rotações tanto para a direita, quanto para a esquerda, em torno de um eixo fixo entre 1° (grau) e 359° (graus). Para a técnica da *Translação*, as imagens sofrem mudanças para direita, esquerda, para cima e para baixo, sem que sofra distorções posicionais nos dados das imagens, o que traz enorme vantagem para a abordagem. Isso ocorre porque, quando imagem sofre a distorção, o espaço restante, que completa o tamanho da imagem, é preenchido com o valor constante, podendo ser ele 0 ou 255, ou até mesmo ser preenchido com ruídos.

3 Materiais e Método

O método proposto para o diagnóstico de glaucoma usando imagens de fundo de olho é descrita neste capítulo.

Figura 11 – Metodologia



Fonte: Gerada pelo autor

A base de imagens na qual o método é avaliado é apresentada primeiro, em seguida, descreve-se a sequência de passos percorridos para atingir o objetivo da metodologia. As etapas da metodologia propostas são apresentadas na figura 11. A primeira etapa é a de pré-processamento, onde foi feita a extração para o canal verde de cada imagem e em seguida aplicada a técnica de Aumento de Dados, ou *Data Augmentation*, para a geração de um maior volume de dados na base de treinos. Em seguida é feita a segmentação do disco óptico para que, na etapa seguinte, as imagens possam ser aplicadas numa Rede Neural Convolutiva (CNN), responsável pela extração de características das imagens e do reconhecimento de padrões. Por fim, é feita a análise dos resultados.

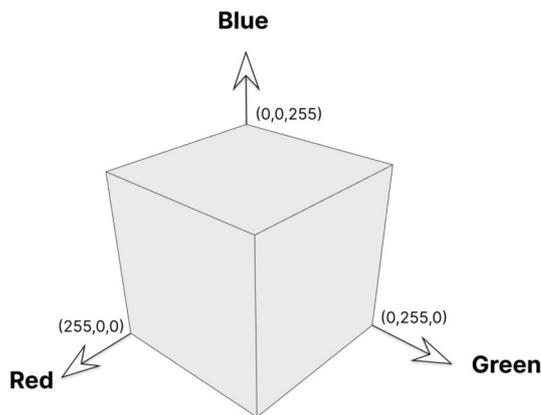
3.1 Aquisição de Imagens

A base de imagens utilizada neste trabalho é a Li et al. (2020), retirada do kaggle¹, que conta com um banco de dados oftalmológico estruturado, com imagens de fundo de olho de 5 mil pacientes, feita em diferentes hospitais e com imagens provenientes de diferentes tipos de câmeras, trazendo maior variedade nas resoluções e também uma situação mais próxima da realidade.

Desta base, foram obtidas as imagens de classe normal e glaucoma. Existe desbalanceamento no material utilizado. Tem-se 2101 imagens de fundo de olho que são classificadas como normais e somente 397 imagens classificadas como afetadas pelo glaucoma.

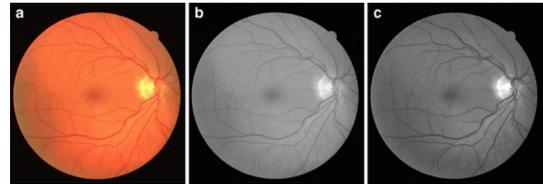
¹ <www.kaggle.com/>

Figura 12 – Cubo RGB



Fonte: Gerada pelo autor

Figura 13 – a: colorida, b: grayscale, c: green channel



Fonte: (SALEH; ESWARAN; MUEEN, 2011)

3.2 Pré-Processamento

A etapa de pré-processamento foi responsável por preparar as imagens originais e tornar evidente as características típicas da doença, além de ter sido fundamental para a etapa de classificação, pois foi também responsável por diminuir o desbalanceamento na base original desse estudo.

Como já explanado na Seção 2.3 da fundamentação deste trabalho, costuma-se descrever as imagens, dentro de um plano bidimensional correlacionado à intensidade da luz, por uma função do tipo $f(x,y)$, cujo valor em qualquer ponto das coordenadas espaciais (x,y) respeita a proporcionalidade do brilho da imagem naquela área específica (ERPEN, 2004). Esse ponto consiste na menor unidade que compõe uma imagem digital e é denominado de *pixel*.

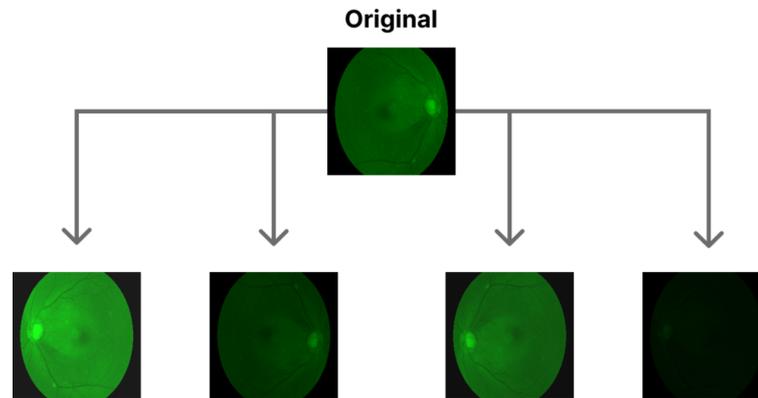
Ao se tratar de imagens coloridas, estas podem ser representadas tanto por uma tabela de cores, quanto pelo modelo RGB (*Red, Green, Blue*). Nesta segunda representação, cada pixel ocupa o valor de 3 bytes, que carregam consigo o valor da intensidade daquele pixel em cada espectro de cor primária do modelo RGB - vermelho, verde ou azul (GONZALEZ; WOODS, 2000), como mostra a representação da Figura 12.

Para a execução deste estudo, inicialmente as imagens são extraídas no canal verde. Tendo em vista o padrão RGB, o canal verde é comumente usado em imagens de fundo de olho com representação da retina por melhorar o contraste das imagens apresentadas. Além disso, estruturas fisiológicas também são melhores contrastadas quando as imagens são extraídas nesse canal (GHARABAGHI; DANESHVAR; SEDAAGHI, 2013), como mostra a Figura 13.

Para lidar com o desbalanceamento da base, foi aplicado o método de *Data Augmentation* para criar um *Oversampling* das imagens, que consiste em gerar novas figuras sintéticas na classe minoritária a partir dos dados iniciais. Para tal, foram aplicadas

as técnicas de translação, rotação e flippings horizontais e verticais, com probabilidade de 100% de ocorrerem, alterações aleatórias de brilho e contraste, variando de -0,9 até 0,9, com uma probabilidade de 70% de serem aplicadas, conforme mostra a Figura 14.

Figura 14 – Oversampling



Fonte: Gerada pelo autor

Para esse processo também é levado em conta a desproporcionalidade da base inicial. Considerando que existiam 2101 imagens classificadas como normais e somente 387 imagens classificadas como afetadas pelo glaucoma, para cada imagem do tipo normal, produziu-se uma cópia da imagem sintética e para cada imagem do tipo patológica, produziu-se 4 outras imagens sintéticas, alterando os dados de uma proporção de 2101 normais por 387 patológicas para 2101 normais por 1588 patológicas.

3.3 Segmentação do Disco Ótico

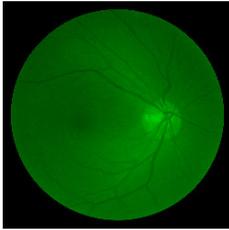
Ainda no contexto, a segmentação consiste na subdivisão da imagem em áreas/regiões (que são um conjunto de pixels) de interesse, o que facilita e simplifica a análise da base de dados. Existem diversas técnicas e algoritmos para realizar a segmentação de imagens, para a simplificação, nesta subseção, é tratada a estratégia de *Thresholding*, que foi implementada durante a execução desta pesquisa.

Primeiramente, a imagem pré-processada, extraída no canal verde (Figura 15) passou por uma conversão para *grayscale* (Figura 16). Para essa conversão, faz-se a soma linear dos componentes do RGB utilizando pesos para as correções, conforme mostra a fórmula da equação 3.1.

$$\text{RGB}[A] \text{ to Gray: } Y \leftarrow (0.299 \times R) + (0.587 \times G) + (0.114 \times B) \quad (3.1)$$

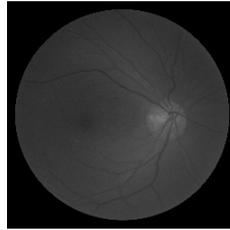
Os pesos são coeficientes baseados nas funções e padrões publicados pela CCIR (União Internacional de Telecomunicação) (ROBERTS, 1994), no modelo de recomendação "BT.601" ou "Rec.601". Observe que as cores não são ponderadas igualmente. Como o verde

Figura 15 – canal verde



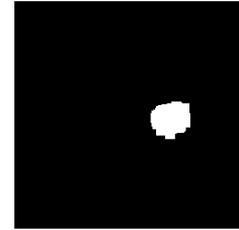
Fonte: Gerada pelo autor

Figura 16 – Grayscale



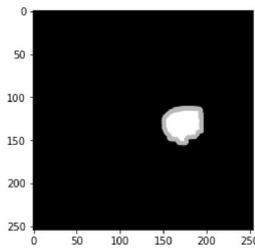
Fonte: Gerada pelo autor

Figura 17 – Threshold



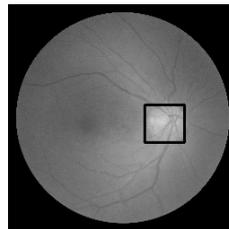
Fonte: Gerada pelo autor

Figura 18 – Contorno



Fonte: Gerada pelo autor

Figura 19 – Área capturada



Fonte: Gerada pelo autor

Figura 20 – Disco óptico



Fonte: Gerada pelo autor

primário é mais claro que o vermelho e o azul primário, ele tem um peso maior. Por outro lado, o azul primário é o mais escuro entre as três cores primárias, levando-o a receber o menor peso.

Depois da conversão, aplica-se o *threshold* na imagem. Nesta operação, uma dada imagem em tons de cinza é segmentada e os componentes gerados são desassociados em classes através do controle feito por um valor (*limiar* ou *threshold*). Dessa forma, pixels que possuem valores menores ou maiores que o limite estabelecido, são classificados de formas distintas (MEIRA et al., 2020). Posto isto, temos que os pixels $f(x,y)$ da imagem, poderão ser agrupados e gerar uma nova imagem binária $g(x,y)$, conforme a fórmula da equação 3.2:

$$f(x) = \begin{cases} 1, & \text{se } f(x,y) > L, \text{ sendo } L \text{ o limiar escolhido} \\ 0, & \text{Caso contrário} \end{cases} \quad (3.2)$$

sendo que os pixels contendo o valor 1 correspondem aos objetos de interesse na imagem (no caso desse estudo, o disco óptico), já os pixels com valor 0 constituem o fundo da imagem (SILVA; TAVARES, 2011). Para esse trabalho, foi aplicado um valor limiar L de 180.

Atingido o momento da execução dessa etapa, tem-se a imagem binária gerada a partir das imagens da base original já pré-processadas (extraídas no canal verde e transformadas em escalas de cinza) com destaque no disco óptico, como mostra a Figura 17.

Uma vez obtida a imagem binária, faz-se a marcação do contorno do objeto em evidência. Assim sendo, analisa-se pixel a pixel, a fim de unir os pontos contínuos ao longo das bordas que possuem a mesma cor ou intensidade, conforme mostra a imagem na Figura 18. A partir desse ponto, é possível criar uma região, que será correspondente à área de interesse da imagem, conforme mostra a Figura 19. E também torna-se possível criar uma única imagem que contem a informação necessária, que compreende o disco óptico da imagem de fundo, conforme a Figura 20, que será utilizada na rede neural convolucional para a realização dos treinos e testes de predição.

3.4 Arquitetura do Modelo CNN

Esta seção apresenta a arquitetura CNN proposta para esse trabalho. Por se tratar de uma quantidade de imagens relativamente pequena, optou-se por uma arquitetura rasa.

A primeira camada da rede neural é um layer convolucional da CNN 2D proposta no trabalho, que contém um filtro de tamanho 16, que precede um layer de agrupamento com MaxPooling. Pode-se entender esse processo como a geração de 16 novas imagens a partir da entrada. Foi também utilizado nas camadas de convolução o processo de padding. Dessa forma, conforme mostra a ilustração da CNN proposta nesta pesquisa na Figura 21.

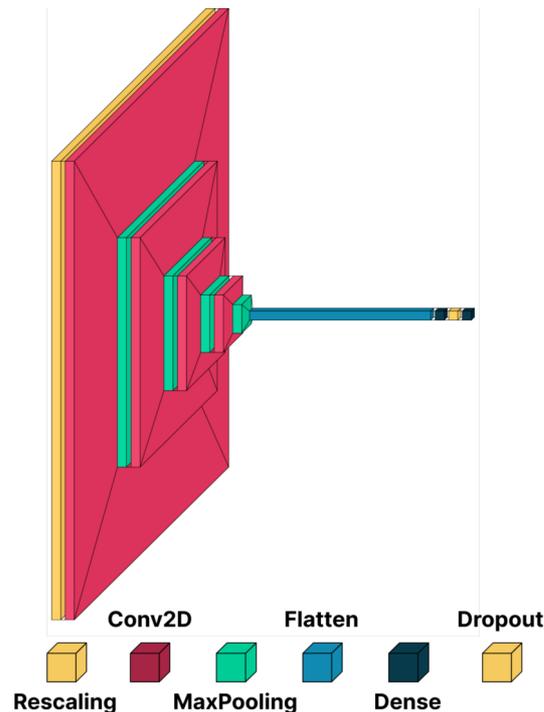
A arquitetura flui para uma estrutura constituída por mais 3 camadas de convolução, cujo filtro é 2 vezes o valor utilizado no filtro do layer de convolução anterior (12,32, ...,128), que intercalam com camadas de MaxPooling. Depois dessas 8 camadas, usa-se o Flatten ou achatamento, que, como já citado, converte a matriz resultante em um vetor linear. Em seguida, usa-se uma camada densa oculta, que antecede uma camada de Dropout para diminuir a sensibilidade da rede neural e, por fim, uma camada densa de saída.

Uma vez criado o modelo, durante sua execução a rede percorre os dados e reconhece os padrões. Para as duas classes existentes no problema em estudo (0 para imagens "normais" e 1 para imagens "patológicas").

Também foi aplicado o sistema de pesos para diminuir os efeitos negativos nas predições causados pelo desbalanceamento da base. Para cada previsão feita erroneamente pelo modelo, há uma penalidade e isso é representado no valor da perda (ou loss) para cada época.

Para a situação em foco nesta pesquisa, 80% da base foi utilizada para treino e 20% foi utilizada para realizar os testes e a rede foi executada por 150 épocas.

Figura 21 – CNN



Fonte: Gerada pelo autor

3.5 Métricas de Avaliação

Para validar a metodologia proposta, os testes e o desempenho do trabalho, faz-se a análise dos resultados que foram gerados pelo classificador, utilizando como métrica a matriz de confusão. Essa matriz, comumente empregue em modelos de classificação, tem como objetivo calcular a quantidade de casos que são rotulados como (VP) Verdadeiros Positivos, (VN) Verdadeiros Negativos, (FP) Falsos Positivos e (FN) Falsos Negativos (HARRISON, 2019). As condições de acerto encontram-se na diagonal principal da matriz de confusão e correspondem aos casos classificados como Verdadeiros Positivos e Verdadeiros Negativos.

Além da matriz de confusão, outra métrica que também foi tida como parâmetro neste estudo foi o *Recall* ou *Sensibilidade*. Este critério é conceituado como a razão entre os verdadeiros positivos (VP) pela soma dos verdadeiros positivos (VP) com os falsos negativos (FN), conforme mostra a Equação 3.3. Assim, esse parâmetro informa a proporção dos casos que são verdadeiros positivos e que foram classificados corretamente, entre todos os casos positivos da observação (VP + FN).

$$Sensibilidade = \frac{VP}{VP + FN} \quad (3.3)$$

Já a precisão, é a razão entre a quantidade de verdadeiros positivos (VP) com a soma dos casos classificados como verdadeiros positivos (VP) e falsos positivos (FP), conforme demonstra a Equação 3.4. Este argumento indica o quão precisas são as classificações

positivas, ou seja, dos casos tidos como verdadeiros positivos, quantos realmente são verdadeiros positivos.

$$Precisão = \frac{VP}{VP + FP} \quad (3.4)$$

Tendo em vista o que já foi dito, vale mencionar um fator conhecido como o Tradeoff entre sensibilidade e Precisão, em que para algumas ocorrências, é preferível priorizar um ou outro, em virtude da relação inversamente proporcional entre ambos, onde quando a precisão aumentar, a sensibilidade tende a diminuir, funcionando de forma recíproca caso no contrário.(HARRISON, 2019).

O outro indicador aplicado na pesquisa, depende tanto da *precisão* quanto da *sensibilidade*, a fim de avaliar o desempenho. Este modelo é denominado de *F1-Score* e sua fórmula se resume na seguinte equação:

$$F1-Score = 2 \times \frac{Precisão \times Sensibilidade}{Precisão + Sensibilidade} \quad (3.5)$$

4 Resultados

Este capítulo aborda e discorre os resultados obtidos através da execução do método proposto, utilizando os parâmetros de avaliação supracitados.

Para a implementação foi utilizada a linguagem Python¹. Durante a organização dos dados fornecidos foi usada a biblioteca Pandas. Na etapa de manipulação das imagens, foram principalmente empregados os recursos das bibliotecas OpenCV², sklearn³ e albumations⁴. Para a visualização dos dados, além do OpenCv², foi empregado também a biblioteca Matplotlib⁵. A fim de gerar uma visualização das camadas (layers) da Rede Neural Convolutiva, foi feita a manipulação dos dados através da biblioteca Visualkeras⁶. Na classificação e criação do modelo da Rede Neural, optou-se pela biblioteca TensorFlow⁷ e por fim, todos os experimentos foram executados na plataforma Google Colab⁸ com aceleração de GPU para diminuir o tempo de processamento.

Como já citado brevemente na seção 3.5, para realizar a avaliação dos resultados obtidos, foi operado o uso dos modelos de *Sensibilidade*, *Precisão* e *F1-Score*. Durante a organização dos dados de treino e teste, foi utilizado o modelo Hold-out, onde os dados são separados entre treino e teste, utilizando uma proporção de 80% para dados de treino e 20% para dados de teste.

Os experimentos são executados em torno de 5 vezes cada. Para as primeiras execuções, conforme mostra a Tabela 1 não foi utilizado a etapa de segmentação e a CNN empregada possuía uma camada de convolução e Max-Pooling a menos do que a rede com os resultados finais, além de não possuir a camada de Dropout.

Tabela 1 – Classificação sem a segmentação e com menos layers na CNN

<i>Classe</i>	<i>Sensibilidade</i>	<i>Precisão</i>	<i>F1-Score</i>
<i>Normal</i>	0,82 ± 0,151	0,83 ± 0,042	0,82 ± 0,024
<i>Glaucoma</i>	0,78 ± 0,040	0,77 ± 0,032	0,78 ± 0,029

Os primeiros resultados foram relativamente insatisfatórios. Para o conjunto de dados classificados como normais a média de precisão na predição foi em torno de 83%, com um desvio de 4,2%, enquanto para os dados classificados como patológicos, a taxa de

¹ <www.python.org>

² <www.opencv.org>

³ <www.scikit-learn.org>

⁴ <www.albumations.ai>

⁵ <www.matplotlib.org>

⁶ <www.github.com/paulgavrikov/visualkeras>

⁷ <www.tensorflow.org>

⁸ <https://colab.research.google.com/>

precisão foi de 77% com um desvio de 3,2%. Para a métrica de sensibilidade, obteve-se uma taxa de 82% com um desvio de 5,2% para a primeira classe e para a segunda um percentual de somente 0,77% com um desvio de 4%. Fechando com o argumento F1-Score, para as classes 0 (Normais) e 1 (Patológicas) obteve-se valores de 82% e 78%, respectivamente.

O segundo lote de testes foi feito utilizando também a segmentação e adicionando uma camada convolucional a mais na rede neural, seguida de outra de Max-Pooling. Conjuntamente, para diminuir o Overfitting, adicionou-se uma camada de Dropout. Os resultados em relação a primeira bateria de testes foram discrepantes, em conformidade com o que é mostrado na Tabela 2.

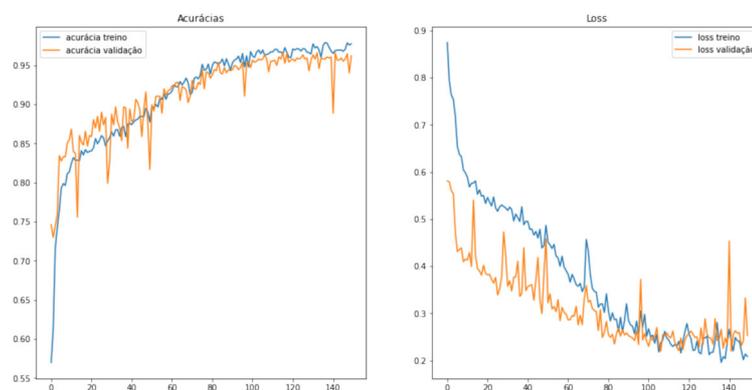
Tabela 2 – Resultados finais da classificação

<i>Classe</i>	<i>Sensibilidade</i>	<i>Precisão</i>	<i>F1-Score</i>
<i>Normal</i>	0,96 ± 0,008	0,96 ± 0,008	0,96 ± 0,008
<i>Glaucoma</i>	0,95 ± 0,014	0,95 ± 0,013	0,95 ± 0,005

Os resultados foram muito mais promissores neste segundo, tanto para a classe normal, quanto para os dados classificados com a doença. Para o parâmetro Sensibilidade, Precisão e F1-Score dos dados que foram tidos como "Normais", a taxa de acerto foi de 96%, com uma variação aproximada de somente 0,8%. Enquanto para as classificações feitas nos casos acometidos pelo distúrbio, houve uma taxa positiva de 95% em média, com desvios de 1,4%, 1,3% e 0,5%, respectivamente, para as métricas citadas.

Fora esses indicadores, foi possível fazer uma análise dos resultados valendo-se dos gráficos de Loss e Acurácia dos treinos e das validações feitas, que foram gerados no fim de cada execução. A Figura 22, mostra em média a relação entre ambos os critérios.

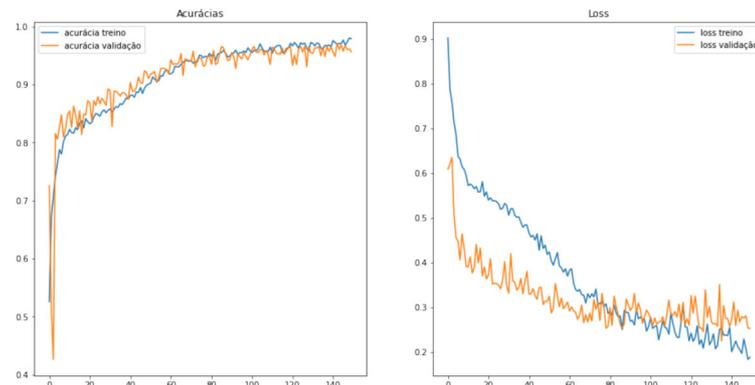
Figura 22 – Gráfico de Loss e Acurácia



Fonte: Gerada pelo autor

Após a aplicação da média na função de Loss, usando o parâmetro smoothing, os picos no gráfico durante o aprendizado se amenizaram, resultando nos dados que são expostos na Figura 23.

Figura 23 – Gráfico de Loss e Acurácia com Smoothing



Fonte: Gerada pelo autor

4.1 Comparação dos Resultados

Nesta seção, os resultados finais obtidos são comparados com outros trabalhos relacionados, utilizando a métrica de Sensibilidade (Sensibilidade), métrica em comum entre estes trabalhos apresentados e o método proposto neste estudo. Na Tabela 3, são evidenciados os resultados aproximados de Sensibilidade dos estudos feitos por cada autor.

Tabela 3 – Comparação de Resultados

<i>Autor</i>	<i>Sensibilidade</i>
RAGHAVENDRA et al.	98%
LI et al.	95%
MOREIRA et al.	88%
CARVALHO et al.	77%
Método Proposto	95%

É importante destacar que cada autor mencionado na tabela de resultados acima utilizaram tanto a base de dados quanto metodologias diferentes. Os autores Moreira et al. (2021), Raghavendra et al. (2018) e LI et al., assim como na metodologia da pesquisa proposta neste trabalho, utilizaram redes neurais convolucionais. Mesmo prevalendo-se de arquiteturas distintas e considerando suas respectivas bases, as pesquisas suprarreferidas obtiveram valores que possuem proximidade com os resultados do trabalho aqui proposto, levando em consideração a métrica avaliada. Os autores da primeira citação utilizaram, além de uma arquitetura diferente, estratégias distintas na etapa de pré-processamento, como a normalização, remoção de ruído e a técnica de Clahe. Em seguida, foi subtraído o valor RGB médio das imagens, para padronizar as imagens, para em seguida fazer o uso da normalização *min-max*. Enquanto Carvalho et al. (2019), utilizaram transformações de intensidade logarítmicas na fase de pré-processamento, para então aplicar os resultados em descritores de textura (LBP e GLCM), para que a classificação seja feita utilizando o método XGBoost, obtendo valores mais distantes das demais obras.

5 Conclusão

Este estudo apresentou uma proposta de método computacional para a classificação de Glaucoma a partir de imagens de fundo de olho.

A metodologia proposta consiste em basicamente 4 etapas, iniciando-se com o pré-processamento, etapa esta que foi responsável por aumentar drástica e positivamente os resultados das operações. Em seguida a imagem é segmentada, para que no passo seguinte possa ser repassada para uma CNN, responsável por fazer a classificação dos indivíduos da base, finalizando na etapa de análise dos resultados.

O uso da segmentação e das técnicas utilizadas no pré-processamento, das camadas extras de Convolução e pooling, acompanhada do uso da camada de Dropout para evitar overfitting e a técnica de Data Augmentation, demonstraram-se fatores essenciais para a conclusão e para os bons resultados desta pesquisa.

Em síntese, espera-se que este trabalho possa contribuir para futuras pesquisas, principalmente relacionada ao tema aqui tratado. Tendo em vista o que foi dito neste trabalho, seguem algumas sugestões para o aprimoramento da pesquisa feita e dentro do tema abordado neste trabalho

- Adquirir uma nova base ou novos itens que possam preencher a base na classe minoritária sem a utilização de imagens sintéticas e comparar resultados.
- Realizar testes utilizando a metodologia proposta em outras patologias oculares.
- Aplicar a sequência logística na metodologia , atribuída na seção de pré-processamento, à outras Redes Neurais e comparar resultados.
- Utilizar um otimizador e especificar extratores de características diferentes.
- Aplicar e avaliar diferentes combinações de hiperparâmetros aplicadas à CNN utilizada

Referências

- AGUIAR, R. d. L. *Classificação de gêneros musicais utilizando convolutional neural network e data augmentation*. Dissertação (Mestrado) — Universidade Estadual de Maringá, 2017. Citado na página 19.
- BUSSEL, I. I.; WOLLSTEIN, G.; SCHUMAN, J. S. Oct for glaucoma diagnosis, screening and detection of glaucoma progression. *British Journal of Ophthalmology*, BMJ Publishing Group Ltd, v. 98, n. Suppl 2, p. ii15–ii19, 2014. Citado na página 17.
- CAPELINI, L. F. et al. *O funcionamento do olho humano: uma sequência didática para o ensino de óptica geométrica no ensino médio*. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, 2019. Citado na página 16.
- CARVALHO, N. R. d. S.; FILHO, A. O. de C.; JÚNIOR, F. d. C. dos A.; SOUSA, D. D. de; ARAÚJO, R. L.; RODRIGUES, M. da C.; LIMA, T. J.; VIEIRA, P. d. A. Método automático para classificação do glaucoma usando análise de textura, xgboost e grid search. In: SBC. *Anais da VII Escola Regional de Computação Aplicada à Saúde*. [S.l.], 2019. p. 300–305. Citado 2 vezes nas páginas 13 e 34.
- CLARO, M.; VOGADO, L.; SANTOS, J.; VERAS, R. Utilização de técnicas de data augmentation em imagens: Teoria e prática. *Sociedade Brasileira de Computação*, 2020. Citado na página 24.
- DUMOULIN, V.; VISIN, F. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016. Citado 3 vezes nas páginas 20, 21 e 22.
- ERPEN, L. R. C. Reconhecimento de padrões em imagens por descritores de forma. 2004. Citado na página 26.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. F. d. Inteligência artificial: uma abordagem de aprendizado de máquina. 2011. Citado na página 20.
- GHARABAGHI, S.; DANESHVAR, S.; SEDAAGHI, M. H. Retinal image registration using geometrical features. *Journal of digital imaging*, Springer, v. 26, n. 2, p. 248–258, 2013. Citado na página 26.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. [S.l.]: Editora Blucher, 2000. Citado 3 vezes nas páginas 17, 18 e 26.
- HARRISON, M. *Machine Learning—Guia de Referência Rápida: Trabalhando com dados estruturados em Python*. [S.l.]: Novatec Editora, 2019. Citado 2 vezes nas páginas 30 e 31.
- HAWKINS, D. M. The problem of overfitting. *Journal of chemical information and computer sciences*, ACS Publications, v. 44, n. 1, p. 1–12, 2004. Citado na página 23.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado na página 19.

- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 20.
- LI, N.; LI, T.; HU, C.; WANG, K.; KANG, H. A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In: SPRINGER. *International Symposium on Benchmarking, Measuring and Optimization*. [S.l.], 2020. p. 177–193. Citado na página 25.
- LI, Z.; HE, Y.; KEEL, S.; MENG, W.; CHANG, R. T.; HE, M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*, Elsevier, v. 125, n. 8, p. 1199–1206, 2018. Citado 2 vezes nas páginas 13 e 34.
- LOUREIRO, F. L. B.; FÉLIX, K. A. d. C. Perfil clínico e epidemiológico dos pacientes com glaucoma atendidos em um ambulatório no interior da amazônia. *Revista Brasileira de Oftalmologia*, SciELO Brasil, v. 79, p. 12–20, 2020. Citado na página 12.
- MEIRA, S. A.; SACRAMENTO, L. A.; LIMA, M. P.; PESSÔA, J. R. de C.; NASCIMENTO, F. L. de A.; ASSIS, J. T. de. Análise da porosidade de concreto por processamento de imagem: uma visão da sensibilidade do threshold na binarização. *Brazilian Journal of Development*, v. 6, n. 3, p. 16449–16459, 2020. Citado na página 28.
- MIRANDA, V. Processamento de imagens médicas digitais: restauração de imagens degradadas e codificação para compressão sem perdas. Citado na página 17.
- MOREIRA, J. M. M.; ALMEIDA, J. D. S. de; JUNIOR, G. B.; PAIVA, A. C. de. Detecção de glaucoma usando redes em cápsula. In: SBC. *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2021. p. 188–199. Citado 2 vezes nas páginas 13 e 34.
- OLIVEIRA, W. D. G. d. Data augmentation via generative adversarial networks aplicado em classificação de imagens. Universidade Federal de São Paulo, 2019. Citado 4 vezes nas páginas 18, 19, 20 e 21.
- RAGHAVENDRA, U.; FUJITA, H.; BHANDARY, S. V.; GUDIGAR, A.; TAN, J. H.; ACHARYA, U. R. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Information Sciences*, Elsevier, v. 441, p. 41–49, 2018. Citado 2 vezes nas páginas 13 e 34.
- RAMOS, A. Fisiologia da visão. 2006. Citado 2 vezes nas páginas 15 e 16.
- ROBERTS, A. Constant-luminance coding for hdtv. In: *Signal Processing of HDTV*. [S.l.]: Elsevier, 1994. p. 643–652. Citado na página 27.
- RODRIGUES, V. A. M. *Edema macular cistóide pseudofáquico em doente com glaucoma crônico simples*. Tese (Doutorado), 2018. Citado na página 16.
- SAHA, S. *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way*. 2018. <<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>>. Acessado em 21/07/2022. Citado na página 21.

- SALEH, M. D.; ESWARAN, C.; MUEEN, A. An automated blood vessel segmentation algorithm using histogram equalization and automatic threshold selection. *Journal of digital imaging*, Springer, v. 24, n. 4, p. 564–572, 2011. Citado na página 26.
- SILVA, M. G.; PESSOA, A. C.; ALMEIDA, J. D. de; JUNIOR, G. B.; PAIVA, A. C. de. Diagnóstico do glaucoma em imagens de retinografia usando variantes de padrões locais binários. In: SBC. *Anais do XVIII Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2018. Citado na página 12.
- SILVA, T. D.; TAVARES, J. M. R. Algoritmos de segmentação de imagem e sua aplicação em imagens do sistema cardiovascular. In: *Actas do 10º Congresso Iberoamericano de Engenharia Mecânica (CIBEM 10)*. [S.l.: s.n.], 2011. Citado na página 28.
- SRIVASTAVA, N. Improving neural networks with dropout. *University of Toronto*, v. 182, n. 566, p. 7, 2013. Citado na página 23.
- URBANO, A. P.; FREITAS, T. G.; ARCIERI, E. S.; URBANO, A. P.; COSTA, V. P. Avaliação dos tipos de glaucoma no serviço de oftalmologia da unicamp. *Arquivos Brasileiros de Oftalmologia*, SciELO Brasil, v. 66, p. 61–65, 2003. Citado na página 16.