

UNIVERSIDADE DO VALE DO RIO DOS SINOS – UNISINOS
UNIDADE ACADÊMICA DE GRADUAÇÃO
CURSO DE ENGENHARIA ELÉTRICA

FERNANDA SCHÄFER TESCH DA SILVA

**ESTUDO DA APLICAÇÃO DE DEEP LEARNING COMO AUXÍLIO AO
DIAGNÓSTICO DE GLAUCOMA**

São Leopoldo, RS

2022

FERNANDA SCHÄFER TESCH DA SILVA

**ESTUDO DA APLICAÇÃO DE DEEP LEARNING COMO
AUXÍLIO AO DIAGNÓSTICO DE GLAUCOMA**

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica, pelo curso de Engenharia Elétrica da Universidade do Vale do Rio dos Sinos (UNISINOS).

Orientador: Prof. Dr. Rodrigo Marques de Figueiredo

São Leopoldo, RS

2022

RESUMO

Atualmente, aproximadamente 80 milhões de pessoas possuem glaucoma em todo o mundo. Estima-se que metade dos portadores da doença não saibam de sua condição. Sabe-se que nos países de renda baixa e/ou média, mais de 90% das pessoas com glaucoma ainda não foram diagnosticadas. Dentre os diagnosticados, 35% já estão cegos. São necessárias, portanto, abordagens de triagem acessíveis e eficazes para permitir a identificação de indivíduos em risco de perda de visão. O uso generalizado de triagem usando imagens de fundo de olho, com classificação assistida por inteligência artificial, pode permitir que o glaucoma seja diagnosticado juntamente com as outras principais causas de cegueira a baixo custo. Estudos de implementação são necessários para determinar como e onde aplicar essas novas ferramentas. Muitas questões importantes de pesquisa permanecem sem solução e exigem investimentos substanciais e um esforço global concentrado para serem respondidas. Indo de encontro a atual necessidade, e a fim de facilitar o atendimento em áreas com acesso restrito a serviços médicos, a presente pesquisa se propõe a avaliar a aplicação de métodos de *deep learning* no auxílio ao diagnóstico glaucoma. Para isso, aplicou-se quatro arquiteturas de CNN existentes (*InceptionV3*, *SqueezeNet*, VGG16 e VGG19) à aproximadamente 1550 exames de fundo de olho, avaliando-se diferentes tratamentos de imagem. Observou-se, no cenário que melhor performou, uma acurácia de 0,71014, precisão de 0,71019, *recall* de 0,71014 e *F1 score* de 0,71012. Como próximos passos, planeja-se vincular os resultados obtidos com o treinamento à outros dados médicos, bem como outros exames de imagem a fim de obter-se melhores resultados.

Palavras-chave: Glaucoma; CNN; Fundoscopia; Predição; *Deep Learning*

ABSTRACT

Approximately 80 million people have glaucoma worldwide. It is estimated that half of them are unaware of their condition. In low and/or middle-income countries, more than 90% of people with glaucoma are still undiagnosed. Among those diagnosed, 35% are already blind. Affordable and effective screening approaches are therefore needed to identify the individuals at risk for vision loss. Widespread use of screening using fundus imaging, with AI-assisted classification, could allow glaucoma to be diagnosed alongside the other leading causes of blindness at low cost. Implementation studies are needed to determine how and where to apply these new tools. Many important research questions remain unresolved and require substantial investment and a concerted global effort to answer. Meeting the current need, the present research proposes to evaluate the application of deep learning methods in the diagnosis of glaucoma. To accomplish it, four existing CNN architectures (InceptionV3, SqueezeNet, VGG16 and VGG19) were applied to approximately 1550 fundus exams, evaluating different image treatments. In the best performance scenario, an accuracy of 0.71014, precision of 0.71019, recall of 0.71014 and F1 score of 0.71012 were observed. As next steps, it is planned to link the training results to other medical data, as well as other imaging tests in order to improve the method.

Palavras-chave: Glaucoma; CNN; Fundoscopic Examination; Prediction; Deep Learning

LISTA DE ILUSTRAÇÕES

Figura 1 – Partes do olho presentes em uma fundoscopia	14
Figura 2 – Diferença entre um olho saudável e com glaucoma	15
Figura 3 – Efeitos do glaucoma no olho	16
Figura 4 – Fluxograma geral	24
Figura 5 – Diagrama de blocos do desenvolvimento da pesquisa	25
Figura 6 – Exemplo de imagens removidas	27
Figura 7 – Imagens originais com e sem glaucoma, respectivamente	28
Figura 8 – Fluxograma do tratamento de imagens	28
Figura 9 – Imagem em escala de cinza	29
Figura 10 – Imagem cortada e binarizada	29
Figura 11 – Figura cortada na ROI e demarcada	30
Figura 12 – Filtros aplicados	31
Figura 13 – Arquitetura da <i>InceptionV3</i>	32
Figura 14 – Arquitetura da <i>SqueezeNet</i>	34
Figura 15 – Arquiteturas VGG16 e VGG19	37
Figura 16 – Diagrama de pré-testes	39
Figura 17 – Ilustração representativa das alterações do n° de neurônios	40
Figura 18 – Matrizes de Confusão para rede <i>InceptionV3</i>	46
Figura 19 – Matrizes de Confusão para rede <i>SqueezeNet</i>	46
Figura 20 – Matrizes de Confusão para rede VGG 16	47
Figura 21 – Matrizes de Confusão para rede VGG 19	48
Figura 22 – Matrizes de Confusão com os resultados finais	50

LISTA DE QUADROS

Quadro 1 – Exemplo de uma matriz confusão	41
---	----

LISTA DE TABELAS

Tabela 1 – Trabalhos correlatos	22
Tabela 2 – Acurácia obtida com os 3 métodos aplicados	43
Tabela 3 – Métricas da rede <i>InceptionV3</i>	46
Tabela 4 – Métricas da rede <i>SqueezeNet</i>	47
Tabela 5 – Métricas da rede VGG 16	47
Tabela 6 – Métricas da rede VGG 19	48
Tabela 7 – Métricas de acordo com o n° de neurônios	49
Tabela 8 – Métricas de acordo com o n° de iterações	49
Tabela 9 – Métricas finais	50

LISTA DE ABREVIATURAS E SIGLAS

AGCNN	Adaptive Gabor Convolutional Neural Networks with Receptive Fields
AUC	Área Under the Curve
CA	Classification Accuracy
CLAHE	Contrast Limited Adaptive Histogram
CNN	Convolutional Neural Network
DENet	Disc-aware Ensemble Network
EUA	Estados Unidos da América
GARDnet	Glaucoma Automated Retinal Detection Network
IA	Inteligencia Artificial
ReLU	Rectified Linear Unit
RGB	Red Green Blue
MC-CNN	CNN multiclass
ML	Machine Learning
NASNet	Neural Search Architecture Networ
OMS	Organização Mundial da Saúde
Rk-CNN	Ranking-CNN
RNN	Region Based Convolutional Neural Networks
ROC	Receiver Operating Characteristic
ROI	Region of Interest Equalization
UNISINOS	Universidade do Vale do Rio dos Sinos
TRk-CNN	Rede Neural Convolutacional de Ranking Transferível
VGG	Visual Geometry Group

SUMÁRIO

1	INTRODUÇÃO	10
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Revisão conceitual	13
2.1.1	Fundoscopia	13
2.1.2	Emprego de IA em exames de diagnóstico	16
2.1.3	Redes Neurais Convolucionais	17
2.2	Estado da arte	19
3	METODOLOGIA	24
3.1	Datasets selecionados	25
3.1.1	Seleção das Imagens	26
3.1.2	Remoção de imagens impróprias para aplicação	27
3.2	Condicionamento das imagens	27
3.2.1	Conversão para escala de cinza	28
3.2.2	Pré-redimensionamento e binarização	29
3.2.3	Seleção da área de interesse	30
3.2.4	Delimitação e redimensionamento	30
3.2.5	Ajuste de contraste e aplicação de filtro	30
3.2.6	CNN's avaliadas	31
3.2.7	<i>InceptionV3</i>	31
3.2.8	<i>SqueezeNet</i>	33
3.2.9	VGG	35
3.2.10	Testes Propostos	38
3.3	Métricas de avaliação dos resultados	40
3.3.1	AUC e ROC	41
3.3.2	Acurácia ou Acurácia de Classificação	41
3.3.3	Precisão	42
3.3.4	<i>Recall</i>	42
3.3.5	<i>F1 Score</i>	42
4	RESULTADOS	43
4.1	Análise dos testes preliminares	43
4.2	Análise do cenário de melhor desempenho selecionado	45
4.2.1	<i>InceptionV3</i>	45
4.2.2	<i>SqueezeNet</i>	46

4.2.3	VGG 16	47
4.2.4	VGG 19	48
4.3	Aprofundamento de testes com a Rede VGG16	48
4.3.1	Variação do número de neurônios nas camadas ocultas	48
4.3.2	Variação do número de iterações	49
4.3.3	Resultados finais	49
4.4	Discussão	50
5	CONCLUSÃO	52
	REFERÊNCIAS	53

1 INTRODUÇÃO

Atualmente, aproximadamente 80 milhões de pessoas são portadoras de glaucoma em todo o mundo. Estima-se que metade dos portadores da doença não saibam de sua condição. Apenas nos Estados Unidos, mais de 3 milhões de norte americanos recebem o diagnóstico positivo. Ao todo mais de 120.000 casos de cegueira são atribuídos ao glaucoma, representando de 9 a 12% do do total da população de cegos (GLAUCOMA RESEARCH FOUNDATION, 2022).

De acordo com o Community Eye Health Journal (2022), nos países de renda baixa e/ou média, mais de 90% das pessoas com glaucoma ainda não foram diagnosticadas. Dentre os diagnosticados, 35% já estão cegos. Essa alta porcentagem ocorre, pois o glaucoma é assintomático até que a doença já esteja em estado avançado. Até o ano de 2040, estima-se que haverá 22 milhões de indivíduos em todo o mundo cegos por glaucoma. Ainda, sabe-se que o distúrbio visual é responsável por mais de 10 milhões de visitas a médicos a cada ano (UNITED NATIONS, 2021).

Em termos de benefícios da Previdência Social, receitas perdidas de imposto de renda e despesas com saúde, o custo para o governo dos Estados Unidos da América (EUA) é estimado em mais de US\$ 1,5 bilhão anualmente (GLAUCOMA RESEARCH FOUNDATION, 2022). Esse fenômeno não é restrito aos EUA, a deficiência visual representa um enorme ônus financeiro global, afetando severamente a qualidade de vida. Adultos com deficiência visual geralmente têm taxas mais baixas de participação na força de trabalho e produtividade e taxas mais altas de depressão e ansiedade. No caso de idosos, a deficiência visual pode contribuir para o isolamento social, dificuldade para caminhar, maior risco de quedas e fraturas e maior probabilidade de entrada precoce em lares de idosos (WORDL HEALTH ORGANIZATION, 2022).

Visando atentar a importância da saúde oftalmológica global, a Organização Mundial da Saúde (OMS) se comprometeu com as seguintes propostas:

- Trabalhar com os Estados-Membros e outros parceiros para fornecer recomendações sobre metas globais viáveis para 2030 sobre cuidados oftalmológicos integrados centrados nas pessoas;
- O desenvolvimento contínuo de ferramentas técnicas para apoiar a implementação das recomendações do Relatório Mundial sobre visão;
- Guia de ação para cuidados oftalmológicos integrados centrados nas pessoas;
- As ferramentas de desenvolvimento e implementação para apoiar os países

na avaliação da prestação de serviços de cuidados oftalmológicos, incluindo ferramentas para avaliação de Glaucoma (UNITED NATIONS, 2021).

Ciente da gravidade do problema, a septuagésimo quinta Assembleia Geral também adotou projetos de resolução destinados a garantir o acesso global a cuidados oftalmológicos. Apresentando o texto intitulado “Visão para Todos: acelerando a ação para alcançar os Objetivos de Desenvolvimento Sustentável” (documento A/75/L.108), o representante de Bangladesh destacou o trabalho realizado pelas partes interessadas, desde a OMS até a parceiros da sociedade. Foi enfatizada a importância de garantir cuidados oftalmológicos adequados para todos, reconhecendo-se o impacto dos problemas oculares na saúde e na vida das pessoas afetadas, desde o aprendizado até os ganhos. Dado que há uma perda global anual de produtividade de US\$ 411 bilhões, ações urgentes devem aumentar o potencial de emprego e a educação. A Assembleia ainda instou os Estados Membros a garantir o acesso aos serviços de cuidados oftalmológicos para suas populações e a mobilizar os recursos e apoio necessários a esse respeito, a fim de contribuir para os esforços globais para alcançar, até 2030, pelo menos 1,1 bilhão de pessoas com deficiência visual e que atualmente não têm acesso aos serviços de atendimento oftalmológico de que precisam (UNITED NATIONS, 2021).

São necessárias, portanto, abordagens de triagem acessíveis e eficazes para permitir a identificação de indivíduos em risco de perda de visão. O uso generalizado de triagem usando imagens de fundo de olho, com classificação assistida por inteligência artificial, pode permitir que o glaucoma seja diagnosticado juntamente com as outras principais causas de cegueira a baixo custo. Estudos de implementação são necessários para determinar como e onde aplicar essas novas ferramentas. Muitas questões importantes de pesquisa permanecem sem solução e exigem investimentos substanciais e um esforço global concentrado para serem respondidas (HU, 2021).

Indo de encontro a atual necessidade de avanços tecnológicos aplicados à área da saúde, e a fim de facilitar o atendimento em áreas com acesso restrito a serviços médicos, a presente pesquisa se propõe a avaliar a aplicação de métodos de ML no auxílio ao diagnóstico de glaucoma. As principais contribuições da pesquisa são:

- Estudar a aplicação de algoritmos de deep learning capazes de auxiliar no diagnóstico de glaucoma;
- Procurar pontuar um método acessível e aplicável a áreas remotas e de pouco acesso aos sistemas de saúde;
- Avaliar a eficácia de um modelo para predição, considerando ainda, o uso de recursos computacionais.

Por fim, o trabalho foi estruturado em seis seções. A primeira seção apresenta a introdução, contendo a justificativa, o problema e os principais objetivos da pesquisa. Na segunda seção há a revisão bibliográfica, a fim de fundamentar a solução apresentada, além de traçar um paralelo com estudos no mesmo setor ou que utilizam soluções aplicáveis ao problema em questão. Na terceira seção, a metodologia, são especificados os passos e técnicas empregadas, sendo elas: especificação de um dataset para análise, pré-processamento de dados, avaliação de modelos de redes neurais. Já a quarta seção, expõe os resultados encontrados durante o treinamento e validação das redes. Por fim, é apresentada uma conclusão, com a compilação dos resultados e *inputs* gerais sobre a pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo contém duas seções principais: os conceitos relevantes aplicados na pesquisa e o estado da arte. Primeiramente, são abordados tópicos concernentes a exames de glaucoma, aplicação de Inteligência Artificial (IA) na área de diagnósticos, bem como um *overview* sobre os conceitos de *deep learning*. Quanto ao estado da arte, são expostas pesquisas relacionadas ao auxílio de diagnóstico de glaucoma a partir de aplicação de IA, além de trabalhos voltados à algoritmos de classificação.

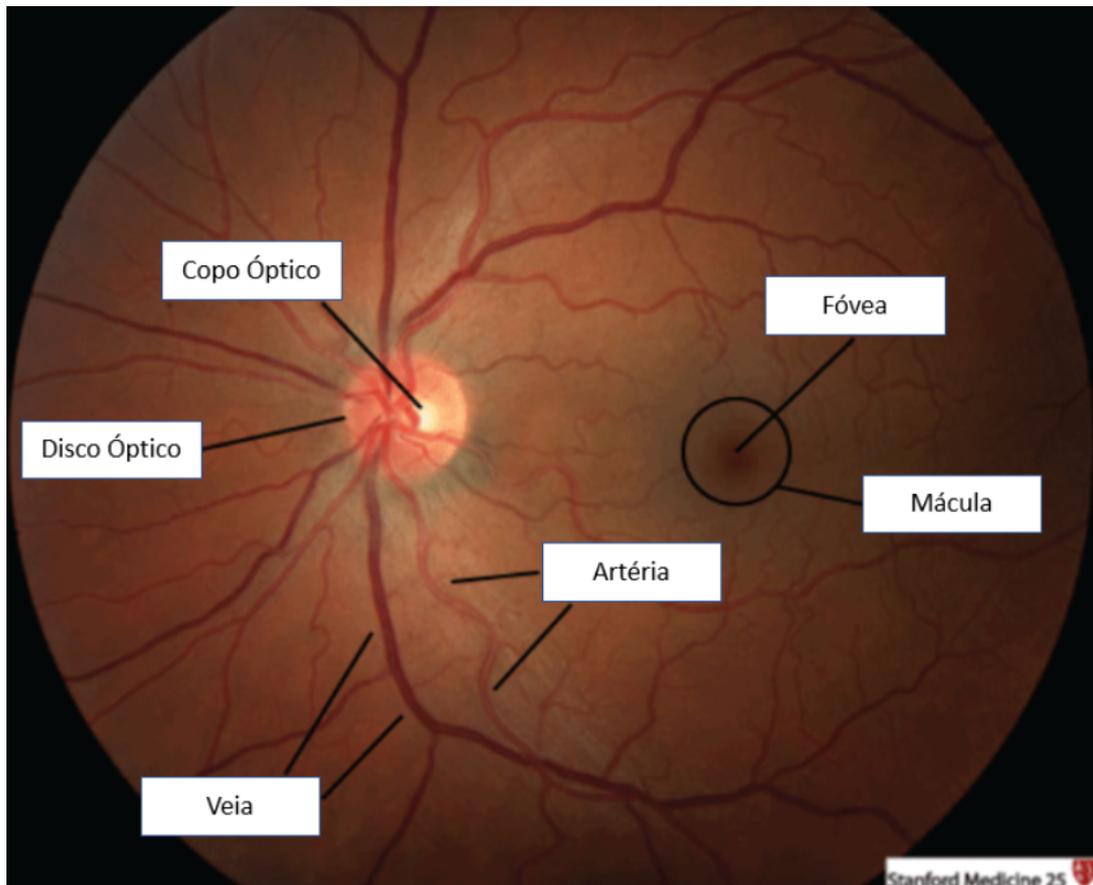
2.1 Revisão conceitual

Na seção, são expostos conceitos básicos fundamentais para a conceitualização e entendimento do trabalho desenvolvido.

2.1.1 Fundoscopia

A retina é a única porção do sistema nervoso central visível de fora. Da mesma forma, o fundo é o único local onde a vasculatura pode ser visualizada. Muito do que vemos na medicina interna é relacionado à vascularização e, portanto, visualizar o fundo é uma ótima maneira de ter uma noção da vasculatura geral do paciente, mas o exame fundoscópico pode descobrir processos patológicos de outra forma invisíveis. Ele é uma parte rotineira do exame oftalmológico de todos os médicos, não apenas do oftalmologista. (STANFORD MEDICINE, 2022) Consiste exclusivamente em inspeção. Olha-se através do oftalmoscópio, que é simplesmente uma luz com várias modificações ópticas, incluindo lentes. O oftalmoscópio ilumina a retina através do defeito normal da íris que é a pupila. Os raios de luz que formam a imagem da retina ressurtem através da pupila. A abertura de visualização (janela) do oftalmoscópio contém uma lente que modifica os raios de luz para auxiliar o usuário. No procedimento, como exposto na Figura 1 observam-se estruturas situadas no aspecto mais interno do globo, conhecidas coletivamente como fundos oculares: retina, vasos sanguíneos da retina, cabeça do nervo óptico (disco) e, em grau limitado, coróide subjacente (SCHNEIDERMAN, 1990).

Figura 1 – Partes do olho presentes em uma fundoscopia



Fonte: Adaptado de (STANFORD MEDICINE, 2022)

A pupila é frequentemente dilatada farmacologicamente para facilitar a inspeção da retina e para o exame da mácula. Um paralisa o músculo pupiloconstritor da íris com drogas parassimpáticas tópicas não absorvíveis e de curta ação, resultando em uma abertura pupilar maior. A cabeça ou disco do nervo óptico é visto quando se olha através da pupila de um ângulo de cerca de 15 graus temporal ao eixo óptico (a linha de visão do paciente, "direto"). O disco tem uma cor amarelo-rosa que se destaca da retina mais vermelha, mais marrom ou mais alaranjada. O disco é nitidamente demarcado temporalmente e em menor grau nasalmente da retina de fundo, que é toda a retina que não é disco, vasos ou mácula. Frequentemente, um crescente estreito de pigmento pontilhado se junta aos lados do disco, especialmente o lado temporal. O disco é ligeiramente mais alto do que largo. A parte central do disco é mais pálida e é chamada de taça óptica ou escavação fisiológica, normalmente ocupando menos de um terço do diâmetro do disco (SCHNEIDERMAN, 1990).

No glaucoma e na alta miopia, o copo é aumentado, como demonstrado na Figura 2.

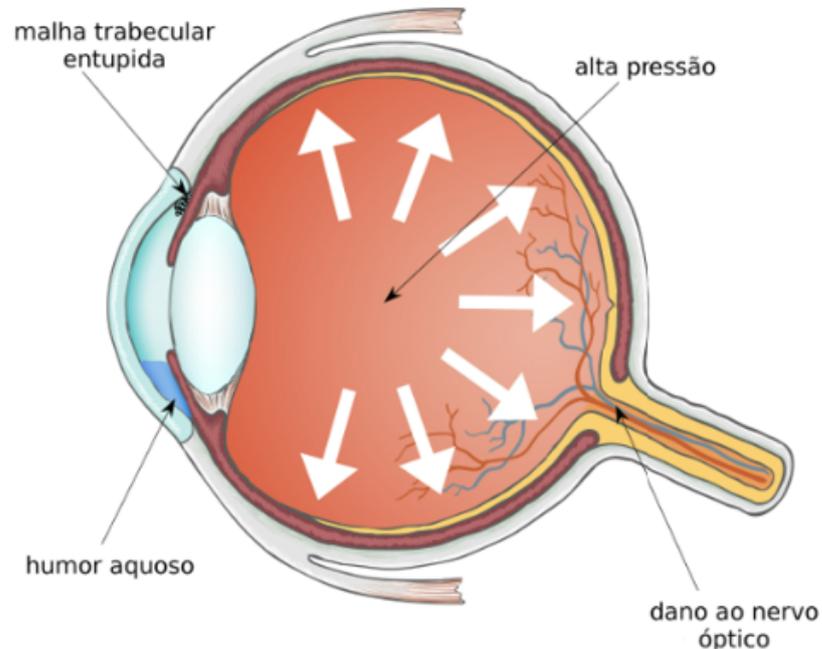
Figura 2 – Diferença entre um olho saudável e com glaucoma



Fonte: Adaptado de (JOHN LANDERS ADVANCED GLAUCOMA AND CATARACT SURGERY, 2022)

O diâmetro transversal do disco é um parâmetro padrão na descrição do fundo, de modo que, por exemplo, uma lesão pode ser caracterizada como "meio diâmetro do disco às duas horas e estendendo-se dois diâmetros do disco superiormente a partir dele". Perto do meio do disco, a artéria e a veia central da retina emergem do nervo óptico. Cada uma se bifurca em ramos superior e inferior, que correm "planos", ou seja, paralelos à superfície da retina. Além de um diâmetro de disco, eles são chamados de arteríolas e vênulas. Com todos os vasos da retina, a artéria/arteríola parece um pouco menor e distintamente mais leve, mais vermelho-alaranjado e menos roxo do que a veia/vénula. A diferença de cor reflete a coluna de sangue contida que é visualizada: as paredes vasculares são transparentes e o sangue venoso desoxigenado é mais escuro que o sangue arterial. Pontua-se, ainda, que as principais artérias, veias e quadrantes de qualquer retina são o temporal superior, o temporal inferior, o nasal inferior e o nasal superior. Especificamente no glaucoma, quando há entupimento das malhas tabeculares, como na Figura 3, a pressão interna no globo ocular é elevada, o que pode provocar a distorção do nervo óptico, que é empurrado para lateral.

Figura 3 – Efeitos do glaucoma no olho



Fonte: (SILVA, 2022)

A área avascular e escura é a mácula. Esta é a área de maior acuidade visual. Além dessa zona, a cor da retina de fundo será paralela à pigmentação da pele e do cabelo do paciente, de pálida em loiros de pele clara a um tom mais escuro nas pessoas pretas (SCHNEIDERMAN, 1990).

Normalmente, as veias maiores pulsam levemente e as artérias não, o inverso da situação em outras partes do corpo. Nenhuma hemorragia é observada em fundos normais. Quaisquer interrupções amarelas, amarelo-esbranquiçadas, cinzas ou pretas do padrão de cor da retina de fundo sugerem exsudato patológico, edema ou cicatriz. Nenhum enrugamento da retina deve ser visto. A brancura calcária ou eritema do disco é anormal, assim como as margens indistintas do disco. Qualquer mudança acentuada na elevação que torne uma área fora de foco com o oftalmoscópio, enquanto o restante da retina permanece em foco, é anormal. Vasos sanguíneos tortuosos geralmente indicam patologia (SCHNEIDERMAN, 1990).

2.1.2 Emprego de IA em exames de diagnóstico

A aplicação de IA oferece vantagens no que diz respeito ao diagnóstico de doenças. O sistema de saúde é um ambiente dinâmico e em mudança, além disso, os

médicos enfrentam continuamente novos desafios com a troca de responsabilidades e interrupções frequentes. Essa rotina faz com que o diagnóstico de doenças se torne um problema secundário para especialistas em saúde. Ainda, a interpretação clínica das informações médicas é uma tarefa cognitivamente desafiadora. Isso não se aplica apenas a médicos experientes, mas também a profissionais com diferentes ou poucas especialidades, como jovens médicos assistentes. O tempo disponível dos especialistas é geralmente limitado, as doenças podem evoluir e a dinâmica do paciente mudar ao longo do tempo, tornando o diagnóstico um processo altamente complexo. No entanto, um processo de diagnóstico preciso é de importância central para garantir o tratamento oportuno e, portanto, para obter atendimento seguro e eficaz ao paciente (MIRBABAIE; STIEGLITZ; FRICK, 2021).

A importância da IA como um componente do processo de diagnóstico tem aumentado constantemente com o desenvolvimento tecnológico. Algoritmos exploram dados médicos para gerar previsões e continuamente aprendem e se desenvolvem ao longo do tempo pelo processamento constante de dados novos e atualizados. Os algoritmos adquirem informações por meio de diferentes tipos de conhecimento e entrada ou ao longo de vários anos de experiência. Portanto, os sistemas habilitados para IA são capazes de processar mais conhecimento em comparação com os humanos, possivelmente superando-os em certas tarefas médicas. A aplicação de IA no processo de diagnóstico de especialistas médicos pode ser de grande valor para o setor de saúde e o bem-estar geral dos pacientes. A integração da IA na infraestrutura técnica existente acelera a identificação de dados médicos relevantes de várias fontes, que são adaptadas às necessidades do paciente e ao processo de tratamento. Simultaneamente, a IA desencadeia o compartilhamento de conhecimento através das fronteiras departamentais, conforme as informações de todas as áreas envolvidas são levadas em consideração. Além disso, ela gera resultados com base em uma população maior, em vez de em experiências pessoais subjetivas e atinge resultados iguais ao usar dados médicos idênticos e não está sujeita à erros humanos (MIRBABAIE; STIEGLITZ; FRICK, 2021).

2.1.3 Redes Neurais Convolucionais

As redes neurais podem ser definidas como um subconjunto do aprendizado de máquina e estão no centro dos algoritmos de aprendizado profundo. Eles são formados por camadas de nós, contendo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada nó se conecta a outro, tendo um peso e limite associados. Se a saída de algum nó individual estiver acima do valor limite especificado, esse nó será ativado, enviando dados para a próxima camada da rede. Caso contrário, nenhum dado é transmitido (IBM CLOUD LEARN HUB, 2021).

Existem diversas redes neurais, que são aplicadas em diferentes casos de uso e tipos de dados. Por exemplo, as redes neurais recorrentes são comumente usadas para processamento de linguagem natural e reconhecimento de fala, enquanto as *Convolutional Neural Networks* (CNN's) são mais frequentemente usadas para tarefas de classificação e visão computacional. Elas fornecem uma abordagem mais escalonável para tarefas de classificação de imagens e reconhecimento de objetos, aproveitando os princípios da álgebra linear, especificamente a multiplicação de matrizes, para identificar padrões em uma imagem. Dito isso, eles podem ser computacionalmente exigentes, exigindo unidades de processamento gráfico para treinar modelos (IBM CLOUD LEARN HUB, 2021).

As CNN's se distinguem de outras redes neurais por seu desempenho superior com entradas de sinal de imagem, voz ou áudio. Eles têm três tipos principais de camadas, que são: camada convolucional, camada de *pooling* e camada *full connected*. A camada convolucional é a primeira camada de uma rede convolucional. Embora as camadas convolucionais possam ser seguidas por camadas convolucionais adicionais ou camadas de *pool*, a camada totalmente conectada é a camada final. A cada camada, a CNN aumenta em sua complexidade, identificando porções maiores da imagem. As camadas anteriores se concentram em recursos simples, como cores e bordas. À medida que os dados da imagem avançam pelas camadas da CNN, ela começa a reconhecer elementos ou formas maiores do objeto até que finalmente termine o objeto pretendido (IBM CLOUD LEARN HUB, 2021).

A camada convolucional é o bloco de construção central de uma CNN e é onde ocorre a maior parte da computação. Requer alguns componentes, que são dados de entrada, um filtro e um mapa de recursos. Ela também apresenta um recurso detector, conhecido como *kernel* ou filtro, que se moverá pelos campos receptivos da imagem, verificando se o recurso está presente. Este processo é conhecido como convolução. Camadas de *pooling*, também conhecidas como *downsampling*, conduzem a redução de dimensionalidade, diminuindo o número de parâmetros na entrada. Semelhante à camada convolucional, a operação de agrupamento varre um filtro por toda a entrada, mas a diferença é que esse filtro não tem pesos. Em vez disso, o *kernel* aplica uma função de agregação aos valores dentro do campo receptivo, preenchendo a matriz de saída. Existem dois tipos principais de *pool*:

- *Pooling* máximo: conforme o filtro se move pela entrada, ele seleciona o pixel com o valor máximo para enviar para a matriz de saída. Como um aparte, essa abordagem tende a ser usada com mais frequência em comparação com o *pool* médio.
- *Pooling* médio: conforme o filtro se move pela entrada, ele calcula o valor médio

dentro do campo receptivo para enviar para a matriz de saída.

Embora muitas informações sejam perdidas na camada de *pool*, elas também trazem vários benefícios para a CNN. Eles ajudam a reduzir a complexidade, melhorar a eficiência e limitar o risco de sobreajuste (IBM CLOUD LEARN HUB, 2021).

Por sua vez, na camada totalmente conectada, cada nó na camada de saída se conecta diretamente a um nó na camada anterior. Ela realiza a tarefa de classificação com base nas características extraídas através das camadas anteriores e seus diferentes filtros. Embora as camadas convolucionais e de *pooling* tendam a usar funções *Rectified Linear Unit* (ReLU), as camadas *full connected* geralmente aproveitam uma função de ativação softmax para classificar as entradas de forma adequada, produzindo uma probabilidade de 0 a 1 (IBM CLOUD LEARN HUB, 2021).

2.2 Estado da arte

Cientes da atual necessidade de melhorar o diagnóstico de Glaucoma, Fu et al. (2018), propuseram um novo *Disc-aware Ensemble Network* (DENet) para rastreamento automático de glaucoma. O algoritmo contém quatro *deep streams* correspondentes a vários níveis e módulos da imagem do fundo de olho. O primeiro é um *stream* de imagem global, que representa a estrutura global de fundo em nível de imagem, e funciona como uma classificação. O segundo é uma rede guiada por segmentação, que detecta a localização do disco, analisando toda a imagem. O terceiro *stream* é baseado na região do disco local, que produz a probabilidade de triagem a partir do nível da região do disco. O *stream* final se concentra na região do disco com transformação polar, que amplia a estrutura do disco e do copo com a operação de geometria e melhora o desempenho de triagem. Finalmente, as saídas dos *deep streams* são combinadas para obter o resultado da triagem. Em resumo, a DENet é um sistema automático para toda a imagem do fundo que inclui detecção de disco e triagem de glaucoma, e usa as representações hierárquicas em diferentes níveis (imagem global e região do disco) e módulos (coordenadas cartesianas e polares). Para avaliação, testou-se o método em dois conjuntos de dados clínicos, analisando-se as curvas *Receiver Operating Characteristic* (ROC) e *Area Under the Curve* (AUC). Finalmente, empregou-se três critérios para medir desempenho, incluindo: Sensibilidade, Especificidade, e Precisão Equilibrada. Os experimentos em dois conjuntos de dados diferentes mostram que o método supera outros algoritmos existentes para detecção de glaucoma.

Considerando que a metodologia básica para detectar o glaucoma é analisar imagens de fundo de olho, os algoritmos de aprendizado de máquina podem oferecer métodos sofisticados para processamento e classificação de imagens. Pensando nisso, Al-Mahrooqi et al. (2022), propuseram uma técnica avançada de pré-processamento

de imagem combinada com uma rede *multi-view* de modelos de classificação profunda para categorizar o glaucoma, o *Glaucoma Automated Retinal Detection Network* (GARDNet). Ele é composto por três diferentes redes neurais convolucionais, treinadas em diferentes vistas de imagens coloridas do fundo de olho. A primeira rede é treinada nas imagens redimensionadas originais, enquanto a segunda rede é treinada na área do disco recortado gerada a partir do step de pré-processamento e, finalmente, a terceira rede é treinada nas imagens com coordenadas polares. O treinamento de cada modelo é feito de forma independente. A escolha do modelo *multi-view* final da rede é baseada em estudos de ablação usando arquiteturas selecionadas. Por fim, no modelo final, aplica-se a transformação polar, que converte a representação da imagem de coordenadas cartesianas para coordenadas polares. A transformação converte a relação radial entre o disco óptico, copo e pano de fundo para uma estrutura hierárquica espacial, que pode fornecer uma visão alternativa ao modelo de classificação e ajudar a capturar recursos mais complexos. A previsão de classificação final é obtida tomando uma média ponderada das três previsões de *softmax*, atribuindo o rótulo de previsão à classe que obteve a maior probabilidade. O modelo foi testado com sucesso com uma AUC de 0,92 e, em seguida, ajustado e testado no conjunto de dados com uma AUC de 0,9308.

Bajwa et al. (2020), também buscando auxiliar no diagnóstico de glaucoma, avaliaram algoritmos de segmentação de última geração e redes de classificação a partir do dataset G1020. Para isso, aplicou-se segmentação automatizada, além das redes *Region Based Convolutional Neural Networks* (RCNN) e ResNet-50 como *backbone* convolucional pré-treinado em ImageNet. Primeiramente, treinou-se 80% das imagens aleatórias do G1020 e testou-se nos 20% restantes. Em seguida, treinou-se o Mask-RCNN usando de um dataset distinto, o ORIGA, e avaliou seu desempenho em todas as imagens do G1020. Para refinar os resultados de segmentação, empregou-se o método *NonMaximum Suppression*. Após localizar e extrair a região dos discos ópticos, eles foram empregados para treinar a rede *InceptionV3* para classificação de imagens saudáveis e glaucomatosas. Por fim, realizou-se a validação cruzada e aplicou-se a *Principal Component Analysis* para obter dois dos principais componentes e visualizá-los no plano 2D. A rede conseguiu atingir níveis competitivos AUC em comparação com os resultados de AUC de última geração no ORIGA, mas apresentava problemas de desempenho no *dataset* G1020.

Focando em *Deep Learning*, Singh, Saini e Lakshminarayanan (2021) propuseram um novo método para treinar redes neurais convolucionais que integra o aprendizado por reforço junto com o aprendizado supervisionado, aplicando-o para transferência de aprendizagem. O método de treinamento usa *hill climbing techniques* através de dois tipos diferentes de *climbs*: “movimento aleatório” e “detecção aleatória”. Ainda, há a integração com modelo de aprendizado supervisionado através de gradi-

ente estocástico do modelo de *stochastic gradient descent with momentum*. O método foi treinado e testado usando os conjuntos de dados, com imagens de fundo de olho normal e glaucomatoso. O desempenho para previsão foi testado por transferência de aprendizado em cinco arquiteturas CNN, a saber, *GoogLeNet*, *DenseNet-201*, *NASNet*, *VGG-19* e *Inception-Resnet-v2*. A técnica de *5-fold classification* foi usada para avaliar o desempenho e sensibilidades. Dos modelos testados, a arquitetura DenseNet-201 teve o melhor desempenho em termos de sensibilidade e área sob a curva. Este método de o treinamento permite a transferência de aprendizado em pequenos conjuntos de dados e pode ser aplicado para aplicativos de tele-oftalmologia, incluindo treinamento com conjuntos de dados locais.

Joon et al. (2021), por sua vez, propõem uma *Rede Neural Convolutacional de Ranking Transferível* (TRk-CNN) que pode ser efetivamente aplicada quando as classes de imagens para ser classificados apresentam uma alta correlação entre si. TRk-CNN, combina os pesos do modelo de classificação primitivo para refletir as informações interclasses para a fase de classificação final. Avaliou-se o TRkCNN em um conjunto de dados de imagem de glaucoma que foi rotulado em três classes: normal, suspeita de glaucoma e olhos com glaucoma. Comparou-se os resultados da avaliação do TRk-CNN com *Ranking-CNN* (Rk-CNN) e *CNN multiclass* (MC-CNN), usando o DenseNet como o modelo principal da CNN. Como resultado, o TRk-CNN alcançou uma precisão média de 92,96%, especificidade de 93,33%, sensibilidade para suspeita de glaucoma de 95,12% e sensibilidade para glaucoma de 93,98%. Com base na precisão média, TRk-CNN é 8,04% e 9,54% maior que Rk-CNN e MC-CNN e 26,83% maior para sensibilidade para suspeitos do que MC-CNN.

Através da análise dos trabalhos correlatos, observa-se, com auxílio da Tabela 1, a possibilidade de empregar diversos métodos de *Machine Learning* (ML) na detecção de glaucoma e análise de exames. O trabalho de Li et al. (2020) propôs um novo modelo de aprendizado profundo, denominado *Adaptive Gabor Convolutional Neural Networks with Receptive Fields* (AG-CNN), para detecção automática de glaucoma e localização de área patológica em imagens de fundo de olho. O algoritmo é composto pelas sub-redes de previsão de atenção, localização da área patológica e é treinado de forma fracamente supervisionada. Assim, o glaucoma pode ser detectado usando os recursos profundos destacados por mapas de visualização de áreas patológicas, com base na previsão de mapas de atenção. Para treinar o modelo AG-CNN, empregou-se o banco de dados com 11.760 imagens de fundo rotuladas como glaucoma positivo ou negativo. Apesar do estudo poder ser empregado como referência, quanto ao uso de *deep learning*, a aplicação de mapas de atenção exige um denso processamento. De forma semelhante, a pesquisa de Tina, Ing-Chou e Yi-Zhu (2021) propuseram uma estrutura para detecção de glaucoma que integra tecnologias de aprendizado profundo e metodologias estatísticas. Essas tecnologias incluem (a) Retinex, um algoritmo de

Tabela 1 – Trabalhos correlatos

Título	Autor	Ano	Relação com o trabalho
<i>A Large-Scale Database and a CNN Model for Attention-Based Glaucoma Detection</i>	Li et al. (2020)	2020	O artigo propõe uma attention-based CNN para detecção de glaucoma. A arquitetura inclui uma sub-rede de previsão de atenção, uma sub-rede de localização de área patológica e uma sub-rede de classificação de glaucoma.
<i>A Statistical Robust Glaucoma Detection Framework Combining Retinex, CNN, and DOE Using Fundus Images</i>	Tina, Ing-Chou e Yi-Zhu (2021)	2021	O estudo propõe uma estrutura para detecção de glaucoma que integra tecnologias de aprendizado profundo e metodologias estatísticas. Essas tecnologias incluem um algoritmo de aprimoramento de cores; a extração de áreas críticas para análise e uma CNN básica com 4 camadas convolucionais.
<i>An Exploratory Study for Glaucoma Detection using Densely Connected Neural Networks</i>	Ovreiu et al. (2020)	2020	Os autores propuseram um método que usa redes neurais densamente conectadas para detecção precoce de glaucoma. Aplicou-se uma rede <i>DenseNet</i> com 121 camadas, que foi inicialmente pré-treinado em um conjunto de dados de imagem natural, <i>ImageNet</i> .
<i>RGB Channel Analysis for Glaucoma Detection in Retinal Fundus Image</i>	Nugraha, Riyandari e Sutoyo (2020)	2020	A pesquisa analisou imagens com base nos 3 canais RGB, a fim de encontrar o melhor componente para detecção de glaucoma. Além disso, aplicou-se métodos como <i>Mean Square Error</i> e <i>Peak Signal to Noise Ratio</i> para análise dos dados.
<i>CNNs for automatic glaucoma assessment using fundus images: an extensive validation</i>	Diaz-Pinto et al. (2019)	2019	Empregou-se cinco modelos diferentes treinados em <i>ImageNet</i> (<i>VGG16</i> , <i>VGG19</i> , <i>InceptionV3</i> , <i>ResNet50</i> e <i>Xception</i>) para avaliação automática de glaucoma usando imagens de fundo. Resultados de extensa validação usando validação cruzada e estratégias de cross-testing foram comparadas com trabalhos anteriores na literatura.

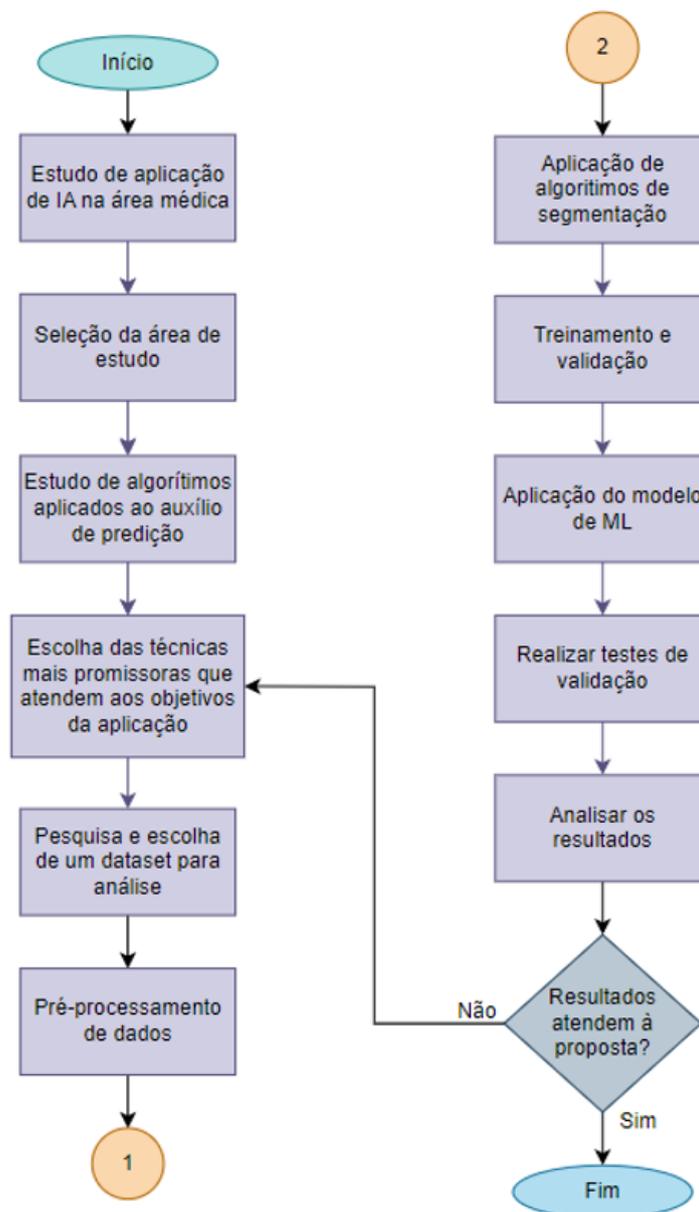
Fonte: Elaborado pela autora.

aprimoramento de cores que remove os efeitos do flash da fotografia do fundo e restaura as cores originais da imagem do fundo; (b) a extração de áreas críticas para análise (incluindo o copo do nervo óptico, disco óptico, mácula e área mínima de fibra do nervo óptico camada); e (c) uma CNN básica com 4 camadas convolucionais. A proposta se assemelha com o trabalho desenvolvido no quesito de identificação de glaucoma a partir de algoritmos de classificação, porém sem focar no auxílio ao diagnóstico. Ovreiu et al. (2020) também incorporaram um modelo à detecção precoce do glaucoma. Os autores utilizaram, primeiramente, um *DenseNet* com 121 camadas, que foi inicialmente pré-treinado em um conjunto de dados de imagem natural, o *ImageNet*. Saliencia-se como limitação o emprego de um conjunto de dados relativamente pequeno, o que pode aumentar a instabilidade e levar a uma baixa generalização dos resultados. Nugraha, Riyandari e Sutoyo (2020), por sua vez, utilizou o processamento digital de imagens com base nos canais (RGB) para facilitar a segmentação do glaucoma. Apesar de apresentar bons resultados preliminares, o método necessita de pesquisas mais profundas. Além disso, os autores focam em segmentação e não, necessariamente, na detecção da doença. Quanto ao artigo desenvolvido por Diaz-Pinto et al. (2019), empregou-se cinco modelos diferentes treinados em *ImageNet* (*VGG16*, *VGG19*, *InceptionV3*, *ResNet50* e *Xception*) para avaliação automática de glaucoma usando imagens de fundo de olho. Os resultados, usando estratégias de validação cruzada e testes cruzados, foram comparados com trabalhos anteriores na literatura. Embora tenha-se empregado um banco de dados bastante grande, ainda há uma limitação ao tentar generalizar a análise. O desempenho diminuiu ao testar a CNN em bases de dados diferentes daquelas utilizadas para treinamento. Além desse problema, concluiu-se que os distintos critérios de rotulagem são outra questão que se enfrenta ao desenvolver sistemas automáticos de avaliação de glaucoma. A maioria dos bancos de dados disponíveis publicamente diferem na maneira como são rotulados, nas informações que os especialistas clínicos usam para avaliar as imagens e na qualidade das imagens do fundo de olho. Conclui-se, por fim, que os trabalhos anteriormente citados podem ser tomados como base para análise de algoritmos para auxílio ao diagnóstico, focados em classificação com base em um dataset de imagens. Como complemento, é necessário, ainda, focar no custo computacional de cada solução, balanceado os recursos e resultados esperados.

3 METODOLOGIA

A presente seção tem como objetivo descrever os processos envolvidos no desenvolvimento da pesquisa, bem como expor os métodos aplicados na solução do problema proposto. A seguir, é possível visualizar um fluxograma geral do desenvolvimento da pesquisa, a partir da Figura 4.

Figura 4 – Fluxograma geral



Fonte: Elaborado pela autora.

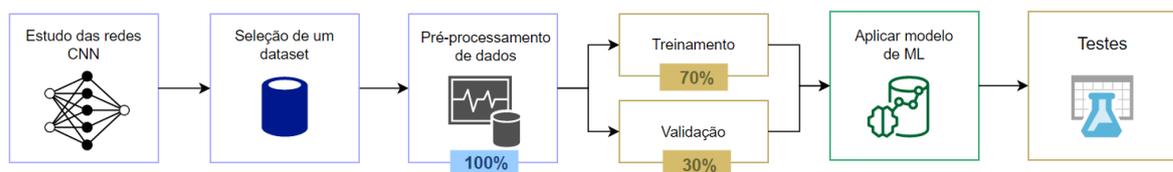
Primeiramente, investigou-se estudos voltados à utilização de IA na área médica, bem como aplicações para auxílio ao diagnóstico no setor. Após um mapeamento inicial, delimitou-se o tema e foco do estudo, ou seja, aplicação de um método como auxílio ao diagnóstico de glaucoma.

Após a escolha do tema de pesquisa, elaborou-se um levantamento, com base no estado da arte, sobre as principais aplicações existentes para classificação de imagens, bem como metodologias empregadas à investigação de glaucoma a partir de fundoscopia.

Seguindo a partir das definições iniciais, pesquisou-se e selecionou-se datasets apropriados. Em seguida, foi efetuado o pré-processamento dos dados. Tal etapa permite obter uma exatidão maior, pois analisa as informações disponíveis e realiza seu tratamento, aumentando a qualidade e facilitando a tomada de decisões das redes neurais. Nessa etapa também é efetuada a remoção de registros falhos ou de informações desconexas do restante dos dados.

Parte-se, então, para o treinamento, realizado com 70% dos dados e para validação, que utiliza 30% dos dados. Com isso, é possível aplicar o modelo que será selecionado, utilizando redes CNN. Após a execução do algoritmo, realiza-se os testes. O fluxo do processo pode ser acompanhado na Figura 5 abaixo:

Figura 5 – Diagrama de blocos do desenvolvimento da pesquisa



Fonte: Elaborado pela autora.

Para avaliação do resultado foram empregadas métricas, como: AUC, CA, precision, F1 score e recall.

3.1 Datasets selecionados

Para o desenvolvimento da pesquisa, empregou-se quatro datasets de fundoscopias: ORIGA, ACRIMA, G1020 e REFUGE. O banco de dados *Online Retinal Fundus Image Dataset for Glaucoma Analysis and Research* (ORIGA) consiste em 650 imagens adquiridas através do *Singapore Malay Eye Study*, que, por sua vez, é conduzido pelo *Singapore Eye Research Institute*. Todas as imagens foram marcadas por especialistas.

O conjunto de dados inclui 168 imagens glaucomatosas e 482 não glaucomatosas (ZHANG et al., 2010).

Quanto ao conjunto de dados ACRIMA, ele foi criado pelo *Ministerio de Economía y Competitividad* da Espanha para a classificação de imagens glaucomatosas. O conjunto de dados contém 309 imagens normais e 396 imagens glaucomatosas capturadas pela câmera de retina Topcon TRC, com um campo de visão de 35°. As imagens foram tiradas dos olhos dilatados e centrados em disco óptico. Apenas imagens de alta qualidade foram selecionadas para evitar artefatos, ruídos e baixo contraste. Dois especialistas rotularam manualmente as imagens como normal ou afetada por glaucoma (SRENG et al., 2020).

O REFUGE, por sua vez, contém 1200 imagens de retina capturadas usando as câmeras Seiss Viscucam e Canon CR-2 de pacientes chineses. As imagens foram centradas na mácula e no disco óptico visível, com o objetivo de examinar os danos no nervo da cabeça óptica e os defeitos da camada de fibras nervosas da retina. O conjunto de dados contém anotações e rótulos de classificação de normal e glaucomatoso, marcados por sete oftalmologistas. Imagens de baixa qualidade foram descartadas pelos oftalmologistas durante a rotulagem. O conjunto de dados foi predefinido em três subconjuntos: treinamento (400 imagens), validação (400 imagens) e teste (400 imagens), e cada conjunto contém uma proporção igual de glaucomatosos (10%) e não glaucomatosos (90%) (SRENG et al., 2020).

Finalmente, o G1020 é um dataset que possui 1020 imagens de fundo de olho disponíveis publicamente (724 saudáveis e 296 de glaucoma) para diagnóstico de glaucoma. A rotulagem das imagens, bem como a segmentação do disco óptico e do copo óptico, é fornecida (KOVALYK et al., 2022).

3.1.1 Seleção das Imagens

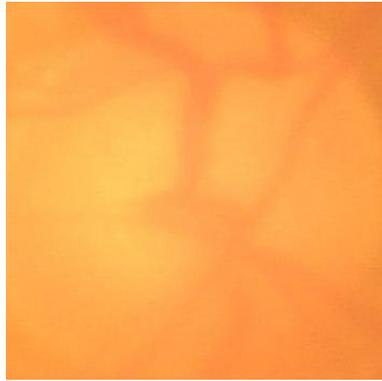
Para aplicação nas redes foi selecionado um número igual de imagens com glaucoma e sem glaucoma dos quatro datasets, considerando o número máximo de imagens com glaucoma (conjunto de menor número em todos *datasets*) de cada um. Ao final, com a remoção dos ruídos, empregou-se 1449 imagens, divididas da seguinte maneira:

- ORIGA: 131 imagens com glaucoma e 175 imagens sem glaucoma;
- ACRIMA: 297 imagens com glaucoma e 138 imagens sem glaucoma;
- G1020: 219 imagens com glaucoma e 280 imagens sem glaucoma;
- REFUGE: 77 imagens com glaucoma e 132 imagens sem glaucoma.

3.1.2 Remoção de imagens impróprias para aplicação

Previamente ao início dos testes, é necessário remover imagens com ângulos que prejudiquem a classificação ou com má qualidade, como exemplificado na Figura 6

Figura 6 – Exemplo de imagens removidas



(a) Imagem com baixa resolução



(b) Imagem não centralizada no nervo óptico

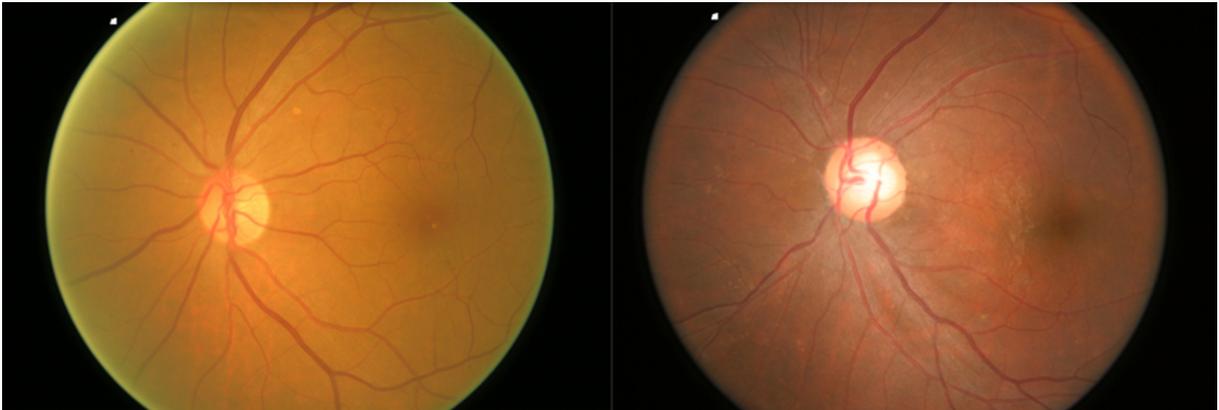
Fonte: Elaborado pela autora.

A remoção foi realizada de forma manual, com base nos critérios expostos acima.

3.2 Condicionamento das imagens

Primeiramente, foi necessário separar as imagens em pastas distintas, conforme o diagnóstico. Para isso, empregou-se um *script* em python para leitura de planilhas em Excel, que continham a identificação da imagem e o diagnóstico, e separação das imagens, de acordo com as anotações fornecidas. Na Figura 7, é possível ver as imagens originais, com e sem glaucoma, respectivamente.

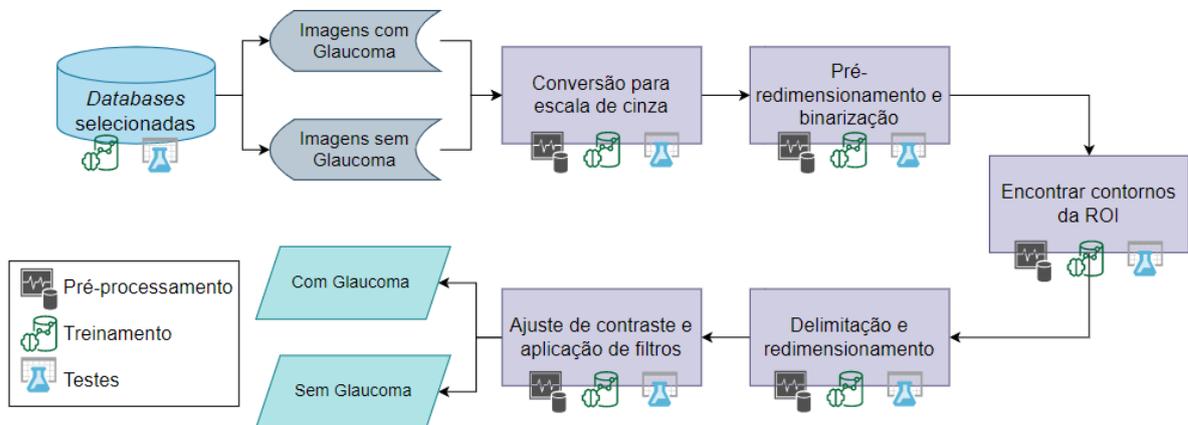
Figura 7 – Imagens originais com e sem glaucoma, respectivamente



Fonte: (KAGGLE, 2022)

Após a separação, iniciou-se o tratamento das imagens, conforme a Figura 8:

Figura 8 – Fluxograma do tratamento de imagens



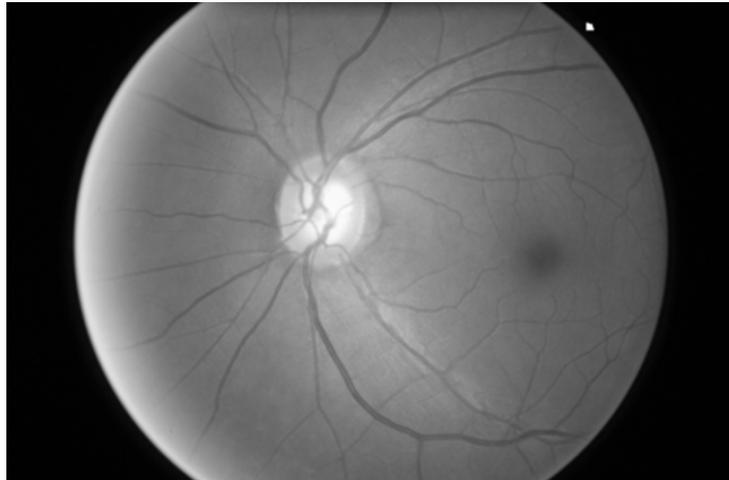
Fonte: Elaborado pela autora.

De acordo com a legenda, é possível destacar por quais etapas a imagem passou, considerando seu estado de tratamento.

3.2.1 Conversão para escala de cinza

Primeiramente, converteu-se a imagem para cinza, reduzindo as dimensões no espaço de cor e facilitando o processamento. Abaixo, observa-se a Figura 9 após a conversão:

Figura 9 – Imagem em escala de cinza



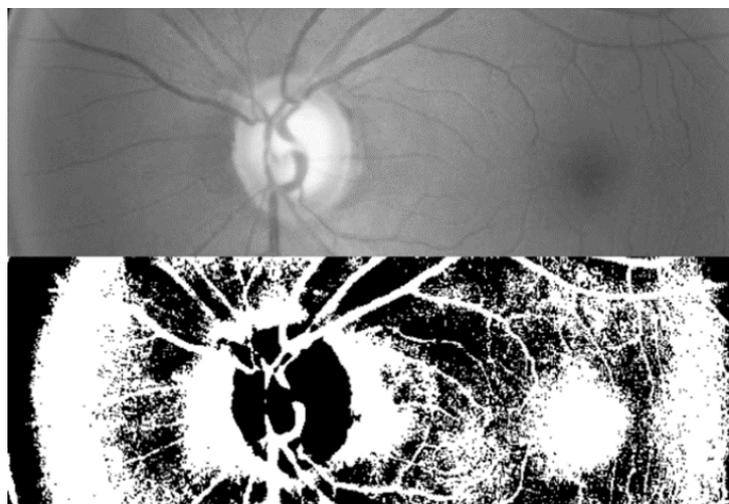
Fonte: Elaborado pela autora.

Após a conversão, partiu-se para as etapas de pré-redimensionamento e binarização.

3.2.2 Pré-redimensionamento e binarização

Para o dataset ACRIMA, desenvolveu-se um algoritmo capaz de encontrar a região do disco óptico, e cortar a imagem a fim de evitar ruídos na CNN. Primeiramente, foi necessário pré-redimensionar a imagem e binarizá-la, a fim de encontrar com maior precisão a *Region of Interest* (ROI), como demonstrado na Figura 10.

Figura 10 – Imagem cortada e binarizada

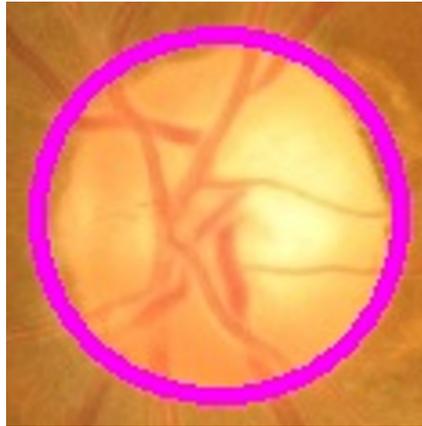


Fonte: Elaborado pela autora.

3.2.3 Seleção da área de interesse

Após a binarização, aplicou-se o método de *Rough-Circles*, capaz de detectar a região circular do disco óptico, de acordo com a Figura 11.

Figura 11 – Figura cortada na ROI e demarcada



Fonte: Elaborado pela autora.

Tendo a ROI definida, foi possível partir para o redimensionamento.

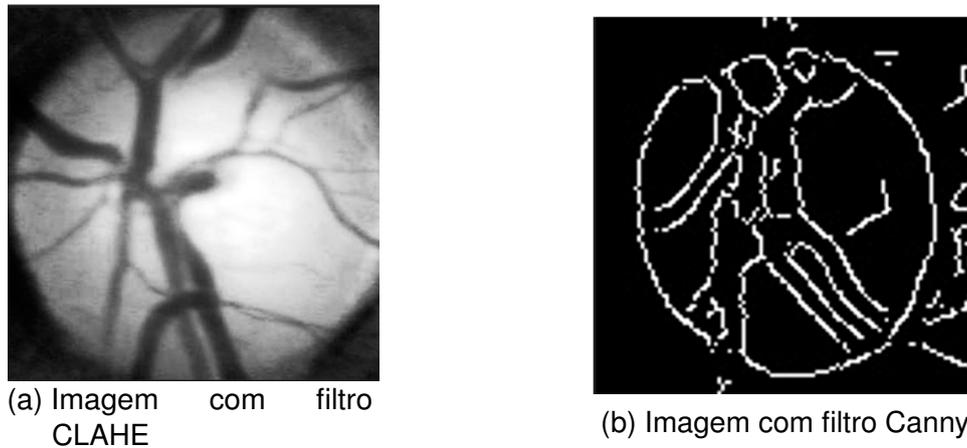
3.2.4 Delimitação e redimensionamento

Com base na posição encontrada anteriormente, cortou-se as imagens considerando o centro do círculo adicionando-se uma margem para evitar a perda de informação. Após, redimensionou-se todas as imagens para o tamanho de 150x150 e 350x350.

3.2.5 Ajuste de contraste e aplicação de filtro

Nessa etapa, aplicou-se o método de *Contrast Limited Adaptive Histogram* (CLAHE) para equalizar as imagens, além do filtro Canny, como demonstrado na Figura 12.

Figura 12 – Filtros aplicados



Fonte: Elaborado pela autora.

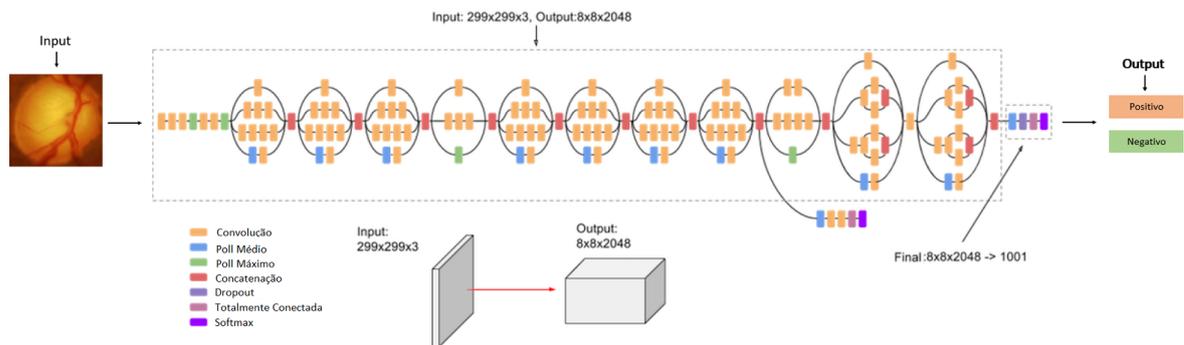
Após finalizada a aplicação dos filtros, seguiu-se para o estudo das redes de treinamento.

3.2.6 CNN's avaliadas

Aqui são descritas brevemente as CNN's avaliadas para aplicação na predição de glaucoma.

3.2.7 InceptionV3

O *InceptionV3* é um modelo composto de blocos de construção simétricos e assimétricos, incluindo convoluções, *pool* médio, *pool* máximo, concatenações, *dropouts* e camadas totalmente conectadas. A normalização em lote é usada extensivamente em todo o modelo e aplicada às entradas de ativação. A perda é calculada usando *Softmax*. O diagrama do modelo pode ser observado na Figura 13.

Figura 13 – Arquitetura da *InceptionV3*

Fonte: Adaptado de (GOOGLE CLOUD, 2022).

O *InceptionV3* se concentra principalmente em utilizar menos poder computacional modificando as arquiteturas anteriores do Inception. Em comparação com o VGGNet, as Redes *Inception* mostraram-se mais eficientes computacionalmente, tanto em número de parâmetros gerados pela rede quanto em termos de custo econômico incorrido (memória e outros recursos). Se alguma mudança for feita em uma *Inception Network*, deve-se tomar cuidado para garantir que as vantagens computacionais não sejam perdidas. Assim, a adaptação de uma rede *Inception* para diferentes casos de uso acaba sendo um problema devido à incerteza da eficiência da nova rede.

A arquitetura de uma rede *InceptionV3* é construída progressivamente, passo a passo, conforme explicado abaixo:

- Convoluções fatoradas: ajuda a reduzir a eficiência computacional, pois reduz o número de parâmetros envolvidos em uma rede. Ele também mantém uma verificação da eficiência da rede;
- Convoluções menores: substituir convoluções maiores por convoluções menores leva a um treinamento mais rápido. Por exemplo, um filtro 5×5 tem 25 parâmetros; dois filtros 3×3 , substituindo uma convolução 5×5 , têm apenas 18 ($3 \times 3 + 3 \times 3$) parâmetros;
- Convoluções assimétricas: Uma convolução 3×3 pode ser substituída por uma convolução 1×3 seguida por uma convolução 3×1 . Se uma convolução 3×3 for substituída por uma convolução 2×2 , o número de parâmetros seria um pouco maior do que a convolução assimétrica proposta;
- Classificador auxiliar: um classificador auxiliar é uma pequena CNN inserida entre as camadas durante o treinamento, e a perda incorrida é somada à perda da rede principal. No *GoogLeNet* foram usados classificadores auxiliares para uma

rede mais profunda, enquanto no *InceptionV3* um classificador auxiliar atua como regularizador.

Como exposto acima, em um modelo *InceptionV3*, várias técnicas para otimizar a rede foram sugeridas para facilitar a adaptação do modelo. As técnicas incluem convoluções fatoradas, regularização, redução de dimensão e cálculos paralelizados (KURAMA, 2020).

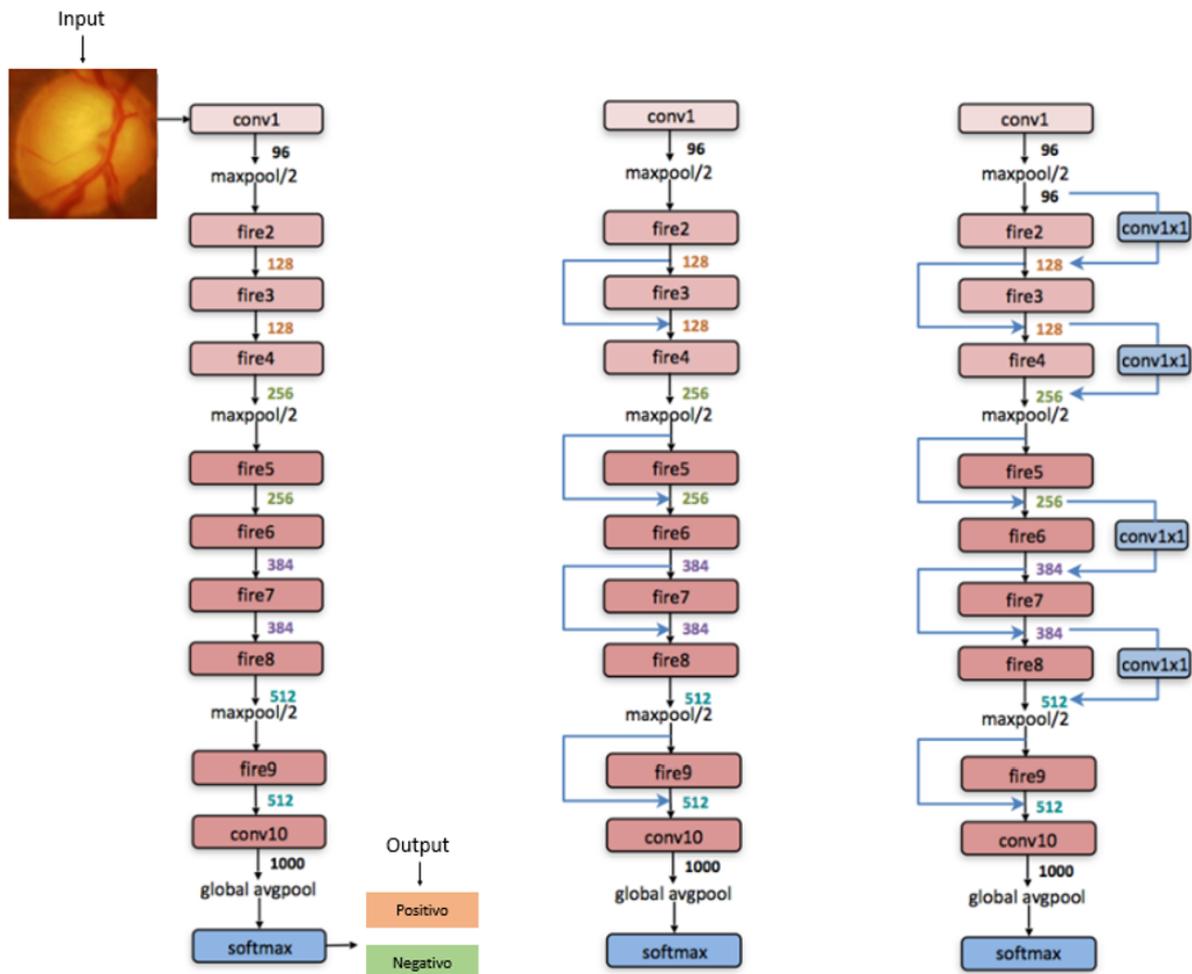
3.2.8 SqueezeNet

O *SqueezeNet* é uma rede projetada como um substituto mais compacto para o AlexNet, possuindo quase 50x menos parâmetros e funcionando 3x mais rápido.

A arquitetura *SqueezeNet* é composta por camadas "squeeze" e "expand". Uma camada convolucional de compressão tem apenas filtros 1×1 . Estes são alimentados em uma camada de expansão que possui uma mistura de filtros de convolução 1×1 e 3×3 .

Uma imagem de entrada é enviada primeiro para uma camada convolucional autônoma, seguida por 8 "módulos de incêndio" que são denominados "fire2-9", de acordo com a estratégia acima. A ilustração do *SqueezeNet* resultante é mostrada na Figura 14 (KURAMA, 2020).

Figura 14 – Arquitetura da SqueezeNet



Fonte: Adaptado de (KURAMA, 2020)

Seguindo a estratégia dois, os filtros por módulo de incêndio são aumentados com *bypass* simples. Por fim, o *SqueezeNet* executa o pool máximo com um passo de 2 após as camadas conv1, fire4, fire8 e conv10. De acordo com a estratégia três, o pooling recebe um posicionamento relativamente tardio, resultando no *SqueezeNet* com um "*bypass* complexo". Abaixo estão os detalhes de outros parâmetros usados na rede:

- A ativação ReLU é aplicada entre todas as camadas de compressão e expansão dentro do módulo de incêndio;
- Camadas de *dropout* são adicionadas para reduzir o *overfitting*, com probabilidade de 0,5 após o módulo de incêndio9;
- Não há camadas totalmente conectadas usadas na rede. Esta escolha de *design* foi inspirada na arquitetura *Network In Network*;

- O *SqueezeNet* foi treinado com uma taxa de aprendizado de 0,04, que diminui linearmente ao longo do processo de treinamento;
- O tamanho do lote para treinamento é 32 e a rede usou um otimizador Adam (KURAMA, 2020).

Finalmente, pode-se pontuar que o *SqueezeNet* facilita o processo de implantação devido ao seu pequeno tamanho. Inicialmente, essa rede foi implementada no Caffe, mas o modelo ganhou popularidade e foi adotado em muitas plataformas diferentes (KURAMA, 2020).

3.2.9 VGG

A rede VGG possui uma arquitetura de CNN profunda com várias camadas. A ferramenta, assim como indicado pelo nome, VGG-16 ou VGG-19, consiste em 16 e 19 camadas convolucionais. Ela é empregada em diferentes modelos de reconhecimento de objetos e uma das arquiteturas de reconhecimento de imagem mais populares.

O modelo VGG, ou VGGNet, que suporta 16 camadas também é conhecido como VGG16, atinge quase 92,7% de precisão de teste no *ImageNet*, um conjunto de dados composto por mais de 14 milhões de imagens pertencentes a cerca de 1.000 classes. A rede emprega filtros do tamanho do kernel 3×3 , um após o outro, fazendo melhorias significativas em relação a redes como o *AlexNet*.

Como mencionado acima, o VGGNet-16 suporta 16 camadas e pode classificar imagens em 1000 categorias de objetos, incluindo teclado, animais, lápis, mouse, etc. Além disso, o modelo possui um tamanho de entrada de imagem de 224 por 224. A arquitetura consiste em 13 camadas convolucionais e três camadas totalmente conectadas, conforme detalhado abaixo:

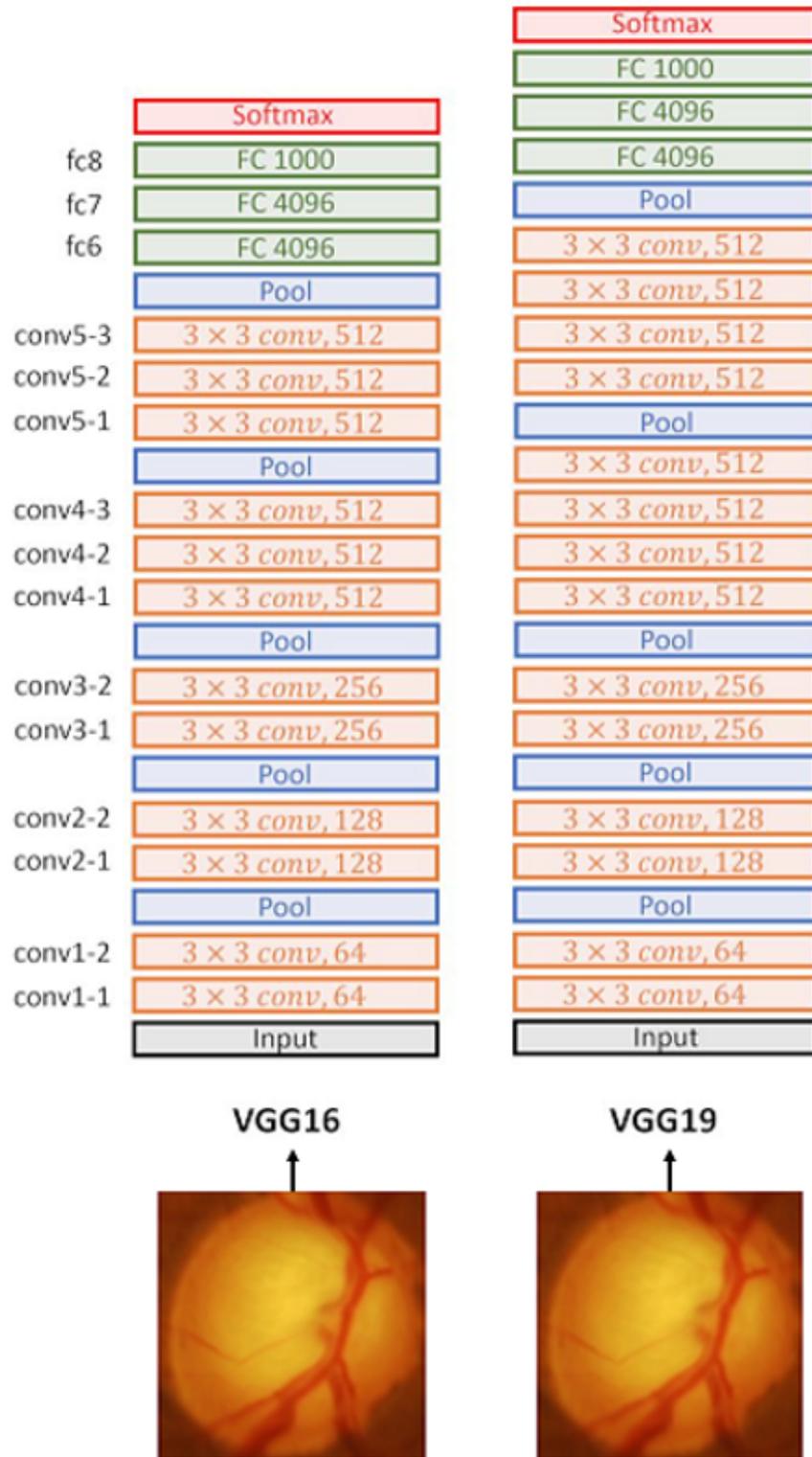
- Input: O VGGNet recebe um tamanho de entrada de imagem de 224×224 . Para a competição ImageNet, os criadores do modelo cortaram o patch central de 224×224 em cada imagem para manter o tamanho de entrada da imagem consistente.
- Camadas convolucionais: as camadas convolucionais do VGG aproveitam um campo receptivo mínimo, ou seja, 3×3 , o menor tamanho possível que ainda captura para cima/baixo e esquerda/direita. Além disso, também existem filtros de convolução 1×1 atuando como uma transformação linear da entrada. Isto é seguido por uma unidade ReLU, que reduz o tempo de treinamento. ReLU significa função de ativação de unidade linear retificada; é uma função linear por partes que produzirá a entrada se for positiva; caso contrário, a saída é zero.

O passo de convolução é fixado em 1 pixel para manter a resolução espacial preservada após a convolução (passo é o número de deslocamentos de pixel sobre a matriz de entrada).

- Camadas ocultas: Todas as camadas ocultas na rede VGG usam ReLU. O VGG geralmente não aproveita a normalização de resposta local, pois aumenta o consumo de memória e o tempo de treinamento. Além disso, não faz melhorias na precisão geral.
- Camadas Totalmente Conectadas: O VGGNet possui três camadas totalmente conectadas.

Das três camadas, as duas primeiras possuem 4.096 canais cada, e a terceira, 1.000 canais, 1 para cada classe. A VGG16 é uma rede bastante extensa e possui um total de cerca de 138 milhões de parâmetros. No entanto, a simplicidade da arquitetura VGGNet16 é o que torna a rede mais atraente. Apenas olhando para sua arquitetura, pode-se dizer que é bastante uniforme. Existem algumas camadas de convolução seguidas por uma camada de agrupamento que reduz a altura e a largura. Ainda, estão disponíveis cerca de 64 filtros que podem ser obrados para cerca de 128 e depois para 256 filtros. Nas últimas camadas, pode-se usar 512 filtros. É possível observar a arquitetura da rede na Figura 15.

Figura 15 – Arquiteturas VGG16 e VGG19



Fonte: Adaptado de (STANFORD, 2017)

Uma das desvantagens cruciais da rede VGG16 é que ela é uma rede enorme, o que significa que leva mais tempo para treinar seus parâmetros. Por causa de sua profundidade e número de camadas totalmente conectadas, o modelo VGG16 tem mais de 533 MB. Isso torna a implementação de uma rede VGG uma tarefa demorada. O modelo VGG16 é usado em vários problemas de classificação de imagem de aprendizado profundo, mas arquiteturas de rede menores, como *GoogLeNet* e *SqueezeNet*, geralmente são preferíveis.

Em termos de desempenho *single-net*, o modelo VGGNet-16 alcança o melhor resultado com cerca de 7,0% de erro de teste, superando assim um GoogLeNet único em cerca de 0,9%.

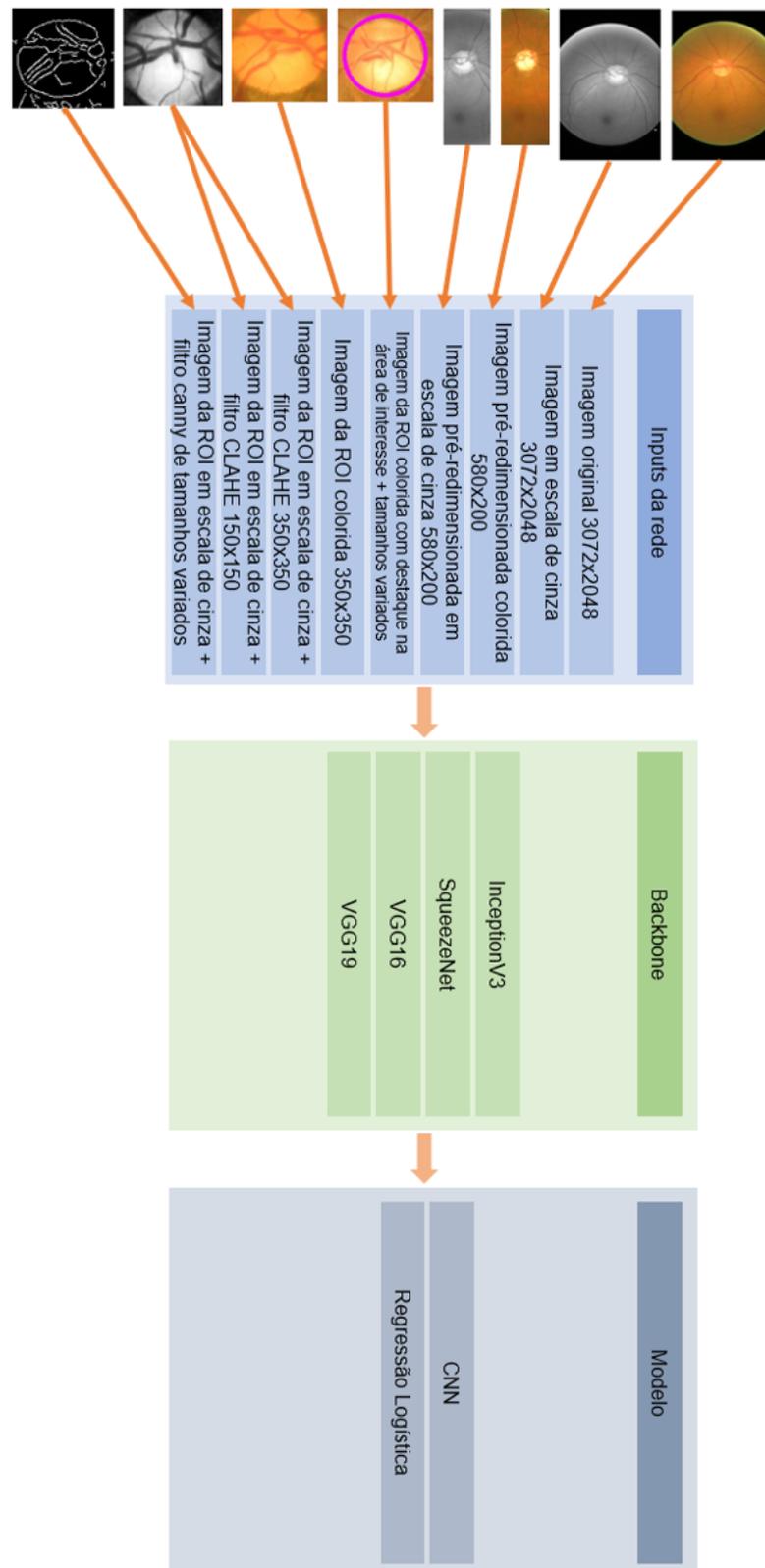
Já o modelo VGG19 (também VGGNet-19), por sua vez, é o mesmo do VGG16, exceto por suportar 19 camadas. O “16” e o “19” representam o número de camadas de peso no modelo (camadas convolucionais). Isso significa que o VGG19 possui três camadas convolucionais a mais que o VGG16 (BOESCH, 2022).

3.2.10 Testes Propostos

Para aplicação das imagens nas redes, optou-se por empregar a ferramenta de análise de dados, Orange. Primeiramente, conduziu-se testes preliminares, com uma quantidade reduzida de 160 imagens (80 positivas e 80 negativas), selecionadas de forma aleatório do conjunto de 1449 imagens, para diferentes tipos de pré-processamento. A intenção era avaliar o comportamento, para posteriormente aprofundar o estudo com os melhores resultados.

Cada pré-processamento, foi utilizado como *input* para as 4 redes mencionadas anteriormente: *InceptionV3*, *SqueezeNet*, VGG16 e VGG19. Ainda, empregou-se modelos de regressão logística e CNN, conforme a Figura 16.

Figura 16 – Diagrama de pré-testes

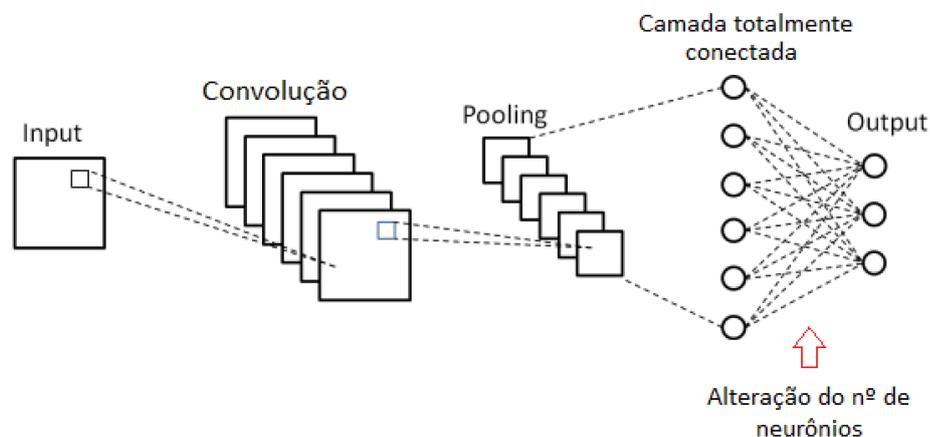


Fonte: Elaborado pela autora.

O primeiro teste consistiu em aplicar nos modelos a imagem original, colorida e em escala de cinza. Em seguida, diminuiu-se o tamanho com a tentativa de eliminar ruídos. O próximo passo, foi inserir apenas as imagens da ROI com diferentes filtros, marcações e tamanhos.

Uma vez selecionado o tratamento de imagem com melhor resposta, os testes e parâmetros foram aprofundados. Aplicou-se os métodos de ativação Relu, bem como o *solver* Adam para as redes neurais. Ainda, variou-se o número de iterações das redes e o número de neurônios na camada totalmente conectada, conforme a figura 17.

Figura 17 – Ilustração representativa da alterações do n° de neurônios



Fonte: Adaptado de (ANALYTICS VIDHYA, 2022)

Vale salientar que o número de épocas empregadas foi fixado em 100, seguindo o padrão da ferramenta Orange. Destaca-se, também, que todas as redes mencionadas são pré-treinadas.

3.3 Métricas de avaliação dos resultados

Por se tratar de uma problema de classificação binária, com ou sem glaucoma, os resultados foram avaliados com base nos acertos e erros das previsões. Com esse intuito, foi empregada como base de validação a matriz de confusão do Quadro 1, como exemplificado abaixo:

Quadro 1 – Exemplo de uma matriz confusão

Matriz de Confusão		Predito	
		Neg	Pos
Real	Neg	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Pos	Falso Negativo (FN)	Verdadeiro Positivo (VP)

Fonte: Elaborado pela autora.

Onde:

- VP: o exame classificado como glaucoma realmente apresenta a doença.
- FP: o exame classificado como glaucoma apresenta um olho saudável.
- VN: o exame classificado como sem glaucoma realmente apresenta um olho saudável.
- FN: o exame classificado como sem glaucoma apresenta a doença.

Partindo da matriz de confusão, é possível analisar métricas como AUC, acurácia, precisão, recall e F1-score, capazes de quantificar a resposta do método empregado.

3.3.1 AUC e ROC

A curva AUC é derivada da curva ROC. A curva ROC, por sua vez, mostra o quão bom o modelo criado pode distinguir entre duas classes, ou seja, positivo e negativo. Os melhores modelos conseguem distinguir com precisão o binômio. Uma curva ROC traça *True Positive Rate vs. False Positive Rate* em diferentes limiares de classificação. Assim, na tentativa de simplificar a análise da ROC, a AUC é uma maneira de resumir a curva ROC em um único valor, agregando todos os limiares da ROC, calculando a área sob a curva. O valor do AUC varia de 0,0 até 1,0 e o limiar entre a classe é 0,5. Ou seja, acima desse limite, o algoritmo classifica em uma classe e abaixo na outra classe. Quanto maior o AUC, melhor o resultado (RODRIGUES, 2018).

3.3.2 Acurácia ou Acurácia de Classificação

A acurácia da classificação é a proporção de exemplos classificados corretamente.

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.1)$$

3.3.3 Precisão

Precisão é a proporção de verdadeiros positivos entre instâncias classificadas como positivas, por exemplo: a proporção de glaucoma positivo corretamente identificada como glaucoma positivo.

$$Precisão = \frac{VP}{VP + FP} \quad (3.2)$$

3.3.4 Recall

Recall é a proporção de verdadeiros positivos entre todas as instâncias positivas nos dados, por exemplo. o número de olhos doentes entre todos os diagnosticados como olhos doentes.

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

3.3.5 F1 Score

F1 Score é uma média harmônica ponderada de precisão e recall.

$$F1Score = \frac{2 \times Precisão \times Recall}{Precisão + Recall} \quad (3.4)$$

4 RESULTADOS

Nesta seção será apresentada a análise e discussão dos resultados obtidos, de acordo com as atividades descritas na metodologia. O capítulo foi dividido entre a avaliação dos testes preliminares e o cenário final de estudo.

4.1 Análise dos testes preliminares

A partir da avaliação dos métodos de classificação, com base em amostras das imagens tratadas, elaborou-se a tabela abaixo, destacando-se a acurácia de cada modelo e o tipo de pré-processamento empregado.

Tabela 2 – Acurácia obtida com os 3 métodos aplicados

Pré-processamento	Backbone	Modelo	Acurácia
Imagem original 3072x2048	<i>InceptionV3</i>	CNN	0.63750
		Regressão Logística	0.61250
	<i>SqueezeNet</i>	CNN	0.61875
		Regressão Logística	0.63125
	VGG16	CNN	0.65625
		Regressão Logística	0.56875
	VGG19	CNN	0.66250
		Regressão Logística	0.61875
Imagem em escala de cinza 3072x2048	<i>InceptionV3</i>	CNN	0.6625
		Regressão Logística	0.6125
	<i>SqueezeNet</i>	CNN	0.6125
		Regressão Logística	0.61875
	VGG16	CNN	0.6625
		Regressão Logística	0.53125
	VGG19	CNN	0.625
		Regressão Logística	0.6375
Imagem pré-redimensionada colorida 580x200	<i>InceptionV3</i>	CNN	0.675
		Regressão Logística	0.68125
	<i>SqueezeNet</i>	CNN	0.675
		Regressão Logística	0.675
	VGG16	CNN	0.80390
		Regressão Logística	0.77390
	VGG19	CNN	0.64375

		Regressão Logística	0.65625
Imagem pré-redimensionada em escala de cinza 580x200	<i>InceptionV3</i>	CNN	0.66875
		Regressão Logística	0.63125
	<i>SqueezeNet</i>	CNN	0.6875
		Regressão Logística	0.61875
	VGG16	CNN	0.5625
		Regressão Logística	0.60625
	VGG19	CNN	0.61875
		Regressão Logística	0.66875
Imagem da ROI colorida com destaque na área de interesse + tamanhos variados	<i>InceptionV3</i>	CNN	0.7625
		Regressão Logística	0.73125
	<i>SqueezeNet</i>	CNN	0.70625
		Regressão Logística	0.68125
	VGG16	CNN	0.73125
		Regressão Logística	0.73125
	VGG19	CNN	0,7
		Regressão Logística	0,65
Imagem da ROI colorida 350x350	<i>InceptionV3</i>	CNN	0.725
		Regressão Logística	0.66875
	<i>SqueezeNet</i>	CNN	0.65625
		Regressão Logística	0.6125
	VGG16	CNN	0.74375
		Regressão Logística	0.675
	VGG19	CNN	0.6625
		Regressão Logística	0.63125
Imagem da ROI em escala de cinza + filtro CLAHE 350x350	<i>InceptionV3</i>	CNN	0.65
		Regressão Logística	0.64375
	<i>SqueezeNet</i>	CNN	0.66875
		Regressão Logística	0.64375
	VGG16	CNN	0.70625
		Regressão Logística	0.64375
	VGG19	CNN	0.61875
		Regressão Logística	0.66875
Imagem da ROI em escala de cinza + filtro CLAHE 150x150	<i>InceptionV3</i>	CNN	0.73125
		Regressão Logística	0.78125
	<i>SqueezeNet</i>	CNN	0.7375
		Regressão Logística	0.6875
	VGG16	CNN	0.80625
		Regressão Logística	0.7875

Filtro Canny	VGG19	CNN	0.71847
		Regressão Logística	0.72482
	<i>InceptionV3</i>	CNN	0.72955
		Regressão Logística	0.70440
	<i>SqueezeNet</i>	CNN	0.66037
		Regressão Logística	0.62893
	VGG16	CNN	0.62893
		Regressão Logística	0.63522
	VGG19	CNN	0.62264
		Regressão Logística	0.70440

Fonte: Elaborado pela autora.

Em um primeiro momento, tomou-se a métrica de acurácia como uma aproximação qualitativa, com o intuito de explorar o comportamento das redes para posterior aprofundamento no cenário de melhor resposta.

Observa-se ao longo dos testes que, quanto mais delimitada a área de análise, maior a acurácia dos resultados. Ainda, notou-se que imagens de menor tamanho apresentaram melhores performance e acurácia. Quanto a escala de cores, não foi possível observar uma discrepância significativa entre os resultados. Em relação aos filtros, o CLAHE se mostrou mais eficaz. Optou-se, portanto, em seguir os testes com imagens em escala de cinza, com filtro CLAHE e dimensão de 150x150.

No que concerne aos modelos de CNN, observou-se que o VGG16, em geral, resultou em melhor acurácia, seguido pelo *InceptionV3* e VGG19. Comparando as redes neurais com a regressão logística, notou-se que o primeiro método se destacou.

4.2 Análise do cenário de melhor desempenho selecionado

A fim de avaliar o cenário escolhido, inicialmente, aplicou-se um total de 1449 imagens nas diferentes redes. Para abordagem CNN, o modelo foi configurado com ativador ReLu e *solver* Adam. Ainda, empregou-se uma iteração máxima de 1000 e 500 neurônios na camada oculta. Na Regressão Logística, empregou-se como regularização o Laço(L1).

4.2.1 InceptionV3

Para a rede *InceptionV3*, observam-se as matrizes de confusão obtidas e os dados extraídos na Figura 18 e na Tabela 3:

Figura 18 – Matrizes de Confusão para rede *InceptionV3*

		Predito					Predito		
		neg	pos	Σ			neg	pos	Σ
Atual	neg	63.5 %	36.4 %	725	Atual	neg	69.0 %	30.7 %	725
	pos	36.5 %	63.6 %	724		pos	31.0 %	69.3 %	724
Σ		731	718	1449	Σ		732	717	1449

(a) Regressão Logística

(b) CNN

Fonte: Elaborado pela autora.

Tabela 3 – Métricas da rede *InceptionV3*

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.74427	0.69151	0.69150	0.69152	0.69151
Logistic Regression	0.70446	0.63561	0.63560	0.63561	0.63561

Fonte: Elaborado pela autora.

No tocante a regressão logística, nota-se 63,5% de verdadeiros negativos, já os verdadeiros positivos seguem com valores aproximados, com 63,6% de acertos. Na matriz de confusão da rede neural, observa-se uma porcentagem de verdadeiros negativos de 69%. Os verdadeiros positivos, por sua vez, ficam com a porcentagem 69,3% de acertos. Comparando-se as matrizes e os valores tabelados, nota-se uma melhor resposta com a aplicação da rede neural.

4.2.2 SqueezeNet

Para a rede *SqueezeNet*, observam-se as matrizes de confusão obtidas e os dados extraídos na Figura 19 e na Tabela 4.

Figura 19 – Matrizes de Confusão para rede *SqueezeNet*

		Predito					Predito		
		neg	pos	Σ			neg	pos	Σ
Atual	neg	65.5 %	33.9 %	725	Atual	neg	68.4 %	30.4 %	725
	pos	34.5 %	66.1 %	724		pos	31.6 %	69.6 %	724
Σ		739	710	1449	Σ		749	700	1449

(a) Regressão Logística

(b) CNN

Fonte: Elaborado pela autora.

Tabela 4 – Métricas da rede *SqueezeNet*

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.75761	0.68944	0.68935	0.68964	0.68944
Logistic Regression	0.72820	0.65769	0.65766	0.65774	0.65769

Fonte: Elaborado pela autora.

Avaliando a matriz de confusão da regressão logística, percebe-se 65,5% de verdadeiros negativos, já os verdadeiros positivos resultam em 66,1%. Quanto a rede neural, observa-se uma porcentagem de verdadeiros negativos de 68,4%. Para os verdadeiros positivos, obteve-se um acerto de 69,6%. Nota-se, no teste, um desempenho melhor com a aplicação da rede neural.

4.2.3 VGG 16

Para a rede VGG16, observam-se as matrizes de confusão obtidas e os dados extraídos na Figura 20 e na Tabela 5:

Figura 20 – Matrizes de Confusão para rede VGG 16

		Predito					Predito		
		neg	pos	Σ			neg	pos	Σ
Atual	neg	69.2 %	30.3 %	725	Atual	neg	70.3 %	29.0 %	725
	pos	30.8 %	69.7 %	724		pos	29.7 %	71.0 %	724
Σ		734	715	1449	Σ		738	711	1449

(a) Regressão Logística

(b) CNN

Fonte: Elaborado pela autora.

Tabela 5 – Métricas da rede VGG 16

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.74427	0.69151	0.69150	0.69152	0.69151
Logistic Regression	0.70446	0.63561	0.63560	0.63561	0.63561

Fonte: Elaborado pela autora.

Concernente a regressão logística, nota-se 69,2% de verdadeiros negativos e 69,7% de verdadeiros positivos. Quanto a rede neural, observa-se uma porcentagem de verdadeiros negativos de 70,3% e verdadeiros positivos com 71% de acertos. Comparando-se as matrizes e os valores tabelados, nota-se mais uma vez uma resposta melhor com a aplicação da rede neural. Quanto aos *backbones*, o VGG16 se destaca com melhores métricas.

4.2.4 VGG 19

Para a rede VGG 19, observam-se as matrizes de confusão obtidas e os dados extraídos na Figura 21 e na Tabela 6:

Figura 21 – Matrizes de Confusão para rede VGG 19

		Predito					Predito		
		neg	pos	Σ			neg	pos	Σ
Atual	neg	68.4 %	31.4 %	725	Atual	neg	66.1 %	33.0 %	725
	pos	31.6 %	68.6 %	724		pos	33.9 %	67.0 %	724
Σ		730	719	1449	Σ		746	703	1449

(a) Regressão Logística (b) CNN

Fonte: Elaborado pela autora.

Tabela 6 – Métricas da rede VGG 19

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.72307	0.66528	0.66521	0.66541	0.66528
Logistic Regression	0.75689	0.68461	0.68460	0.68461	0.68461

Fonte: Elaborado pela autora.

Por fim, tomando a matriz de confusão da regressão logística, percebe-se 68,4% de verdadeiros negativos. Os verdadeiros positivos, por sua vez, possuem 68,6% de acerto. Na matriz de confusão da rede neural, percebe-se uma porcentagem de verdadeiros negativos de 66,1%. Já nos verdadeiros positivos, obteve-se um acerto de 67%. Nota-se, nesse caso, que a regressão logística possui um melhor desempenho.

4.3 Aprofundamento de testes com a Rede VGG16

A fim de explorar o cenário mais eficaz, desenvolveu-se testes a partir da variação dos parâmetros do método de redes neurais, que apresentou melhor desempenho.

4.3.1 Variação do número de neurônios nas camadas ocultas

Testou-se a variação do número de neurônios nas camadas ocultas, segundo a Tabela 7 com as principais métricas:

Observa-se que quanto maior o número de neurônios, melhores são as métricas de acurácia, F1, precisão e *recall*. Sienta-se que a discrepância é pequena e deve-se considerar o desempenho de processamento, uma vez que, quanto maior o número de neurônios, mais densa fica a rede.

Tabela 7 – Métricas de acordo com o n° de neurônios

N° de neurônios	AUC	CA	F1	Precision	Recall
250	0.76219	0.69979	0.69978	0.69980	0.69979
500	0.76563	0.70669	0.70666	0.70675	0.70669
750	0.76432	0.70392	0.70393	0.70392	0.70393
1000	0.76110	0.70048	0.70047	0.70050	0.70048
1250	0.76647	0.70738	0.70737	0.70739	0.70738
1500	0.76505	0.71014	0.71012	0.71019	0.71014

Fonte: Elaborado pela autora.

4.3.2 Variação do número de iterações

Por fim, avaliou-se a variação do número de iterações, conforma a Tabela 8.

Tabela 8 – Métricas de acordo com o n° de iterações

N° de iterações	AUC	CA	F1	Precision	Recall
250	0.76432	0.70392	0.70393	0.70396	0.70393
500	0.76432	0.70393	0.70393	0.70396	0.70393
750	0.76432	0.70392	0.70392	0.70393	0.70393
1000	0.76432	0.70392	0.70393	0.70392	0.70393
1250	0.76432	0.70392	0.70393	0.70392	0.70393

Fonte: Elaborado pela autora.

Nota-se, após a execução dos testes, que a variação é mínima, não interferindo no resultado final. Deve-se, portanto, selecionar o menor número de iterações, visando melhorar o desempenho da rede.

4.3.3 Resultados finais

Finalmente, os resultados obtidos no cenário de melhor desempenho podem ser observados na Tabela 9 e na matriz de confusão da Figura 22.

Figura 22 – Matrizes de Confusão com os resultados finais

		Predito		Σ
		neg	pos	
Atual	neg	70.7 %	28.7 %	725
	pos	29.3 %	71.3 %	724
Σ		737	712	1449

Fonte: Elaborado pela autora.

Tabela 9 – Métricas finais

AUC	CA	F1	Precision	Recall
0.76505	0.71014	0.71012	0.71019	0.71014

Fonte: Elaborado pela autora.

Nota-se que os verdadeiros negativos atingiram 70,7%. Já os verdadeiros positivos, alcançaram 71,3% de acertos.

4.4 Discussão

A partir dos testes realizados foi possível analisar o desempenho de quatro redes existentes aplicadas ao diagnóstico de glaucoma, submetidas à imagens com diferentes tratamentos: *InceptionV3*, *SqueezeNet*, VGG16 e VGG19.

No que concerne ao pré-processamento, observou-se um melhor desempenho das redes com imagens focadas na ROI, o que era esperado, uma vez que, quando mais regiões envolvidas, mais ruído é inserido na rede. Ainda, notou-se que imagens em escala de cinza e com filtro CLAHE geraram respostas mais eficientes. A conversão para escala de cinza pode ajudar no processamento, uma vez que, para determinadas aplicações, as cores não são úteis. Além disso, ter apenas um canal melhora o desempenho de processamento. Quanto ao tamanho das imagens, notou-se que imagens menores não causarem perda de informação, mas, pelo contrário, são favoráveis à performance da rede.

No tocante aos *backbones* empregados, notou-se um melhor desempenho da rede VGG16, em grande partes dos testes. Já em relação aos métodos aplicados, redes

neurais e regressão logística, observou-se melhor desempenho das redes neurais, com uma maior quantidade de neurônios nas camadas ocultas. Observou-se, no cenário que melhor performou, uma acurácia de 0,71014, precisão de 0,71019, *recall* de 0,71014 e *F1-score* de 0,71012. Pontua-se que a discrepância entre o resultado encontrado e os expostos na literatura pode ter sido ocasionada devido a metodologia estudada: empregou-se *backbones* individuais pré-treinados em uma vista específica do globo ocular. Alguns trabalhos, por sua vez, utilizaram abordagens *multi-view* com diferentes camadas de treinamento, redes mais complexas para detecção de áreas patológicas e alteração dos pesos dos modelos de classificação baseados em correlação. Ainda, é possível empregar diferentes *backbones* em uma única rede para atingir melhores resultados. O tamanho do conjunto de dados também é um fator de influência, uma vez que um conjunto pequeno de exames pode gerar *overfitting*.

De forma geral, as métricas obtidas apresentaram bons resultados e boa performance. Pontua-se que os métodos estudados podem ser aplicados no auxílio ao diagnóstico de glaucoma de forma simples e prática. Como próximos passos, vê-se a necessidade de vincular os atuais resultados à outros dados do paciente, bem com outros exames de imagem oculares.

5 CONCLUSÃO

A presente pesquisa tem como objetivo estudar a aplicação de *deep learning* como auxílio no diagnóstico de Glaucoma, pontuando um método acessível e aplicável a áreas remotas e de pouco acesso aos sistemas de saúde. Além disso, procura-se avaliar a eficácia de modelos existentes para predição da doença. Focando nisso, estudou-se quatro arquiteturas de redes existentes: *InceptionV3*, *SqueezeNet*, VGG16 e VGG19. Ainda, explorou-se diversos tratamentos de imagem e seu comportamento quando aplicados à exames de fundo de olho.

Notou-se que a rede VGG16, empregada como *backbone*, obteve a melhor performance, atingindo uma acurácia de 71%. Salienta-se que entre os métodos estudados, regressão logística e redes neurais, o segundo apresentou melhor desempenho. No tocante ao tratamento das imagens, os exames com dimensões de 150x150, focados na ROI, com escala de cinza e filtro CLAHE apresentaram resultados mais satisfatórios. Observa-se que a discrepância entre o resultado encontrado e os expostos no estado da arte pode ser relacionada à metodologia estudada, que empregou *backbones* individuais pré-treinados em uma vista específica do globo ocular. Algumas pesquisas, no entanto, utilizaram abordagens *multi-view* com diferentes camadas de treinamento, redes mais complexas para detecção de áreas patológicas, além da alteração dos pesos dos modelos de classificação com base em correlação. Vale pontuar que o tamanho do conjunto de dados também é um fator de influência, uma vez que um conjunto pequeno de exames pode gerar *overfitting*.

A partir do estudo, portanto, foi possível extrair dados que comprovam a eficácia das redes estudadas no auxílio ao diagnóstico de glaucoma de forma simples e prática. Como próximos passos, a fim de melhorar os resultados, é possível vincular os cenários estudados à outros dados do paciente e, também, a diferentes exames de imagem oculares. Ainda, pode-se empregar os dados fornecidos na construção de novos algoritmos e na automatização do tratamento de imagem para exames de fundo de olho e detecção de glaucoma.

REFERÊNCIAS

- AL-MAHROOQI, A. et al. Gardnet: Robust multi-view network for glaucoma classification in color fundus images. *International Workshop on Ophthalmic Medical Image Analysis*, Springer, 2022. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-031-16525-2_16>. Citado na página 19.
- ANALYTICS VIDHYA. *CONVOLUTIONAL NEURAL NETWORK(CNN)*. 2022. Disponível em: <<https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-networkcnn/>>. Citado na página 40.
- BAJWA et al. G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection. *International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2020. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9207664>>. Citado na página 20.
- BOESCH, G. *VGG Very Deep Convolutional Networks (VGGNet) – What you need to know*. 2022. Disponível em: <<https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/>>. Citado na página 38.
- COMMUNITY EYE HEALTH JOURNAL. *Update on Glaucoma*. 2022. Disponível em: <<https://www.cehjournal.org/article/update-on-glaucoma/>>. Citado na página 10.
- DIAZ-PINTO, A. et al. Cnns for automatic glaucoma assessment using fundus images: an extensive validation. *BioMedical Engineering OnLine*, 2019. Citado 2 vezes nas páginas 22 e 23.
- FU, H. et al. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Transactions on Medical Imaging*, IEEE, v. 11, 2018. ISSN 1558-254X. Disponível em: <<https://ieeexplore.ieee.org/document/8359118/references#references>>. Citado na página 19.
- GLAUCOMA RESEARCH FOUNDATION. *Mouth and oral cancer statistics*. 2022. Disponível em: <<https://glaucoma.org/glaucoma-facts-and-stats>>. Citado na página 10.
- GOOGLE CLOUD. *Advanced Guide to Inception v3*. 2022. Disponível em: <<https://cloud.google.com/tpu/docs/inception-v3-advanced>>. Citado na página 32.
- HU, V. H. The global challenge of glaucoma. *National Library of Medicine*, n. PMC8862625, 2021. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8862625/>>. Citado na página 11.
- IBM CLOUD LEARN HUB. *Convolutional Neural Networks*. 2021. Disponível em: <<https://www.ibm.com/cloud/learn/convolutional-neural-networks>>. Citado 3 vezes nas páginas 17, 18 e 19.
- JOHN LANDERS ADVANCED GLAUCOMA AND CATARACT SURGERY. *Glaucoma*. 2022. Disponível em: <<https://advancedglaucomasurgery.com.au/conditions/glaucoma/>>. Citado na página 15.

JOON, J. T. et al. Trk-cnn: transferable ranking-cnn for image classification of glaucoma, glaucoma suspect, and normal eyes. *Expert Systems with Applications*, Elsevier, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0957417421006448>>. Citado na página 21.

KAGGLE. *Glaucoma Fundus Imaging Datasets*. 2022. Disponível em: <<https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets>>. Citado na página 28.

KOVALYK, O. et al. Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data*, 2022. Disponível em: <<https://www.nature.com/articles/s41597-022-01388-1#citeas>>. Citado na página 26.

KURAMA, V. *A Review of Popular Deep Learning Architectures: ResNet, InceptionV3, and SqueezeNet*. 2020. Disponível em: <<https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/>>. Citado 3 vezes nas páginas 33, 34 e 35.

LI, L. et al. A large-scale database and a cnn model for attention-based glaucoma detection. *IEEE Transactions on Medical Imaging*, v. 39, n. 2, p. 413–424, 2020. Citado 2 vezes nas páginas 21 e 22.

MIRBABAIE, M.; STIEGLITZ, S.; FRICK, N. R. J. Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. *Health and Technology*, Springer, v. 11, 2021. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8862625/>>. Citado na página 17.

NUGRAHA, G. S.; RIYANDARI, B. A.; SUTOYO, E. Rgb channel analysis for glaucoma detection in retinal fundus image. In: *2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS)*. [S.l.: s.n.], 2020. p. 1–5. Citado 2 vezes nas páginas 22 e 23.

OVREIU, S. et al. An exploratory study for glaucoma detection using densely connected neural networks. In: *2020 International Conference on e-Health and Bioengineering (EHB)*. [S.l.: s.n.], 2020. p. 1–4. Citado 2 vezes nas páginas 22 e 23.

RODRIGUES, V. *Entenda o que é AUC e ROC nos modelos de Machine Learning*. 2018. Disponível em: <<https://medium.com/bio-data-blog/entenda-o-que-e-auc-e-roc-nos-modelos-de-machine-learning-8191fb4df772>>. Citado na página 41.

SCHNEIDERMAN, H. *Clinical Methods: The History, Physical, and Laboratory Examinations*. 3. ed. Boston: Butterworths: National Library of Medicine, 1990. Citado 3 vezes nas páginas 13, 14 e 16.

SILVA, M. A. da. *Glaucoma*. 2022. Disponível em: <<https://www.infoescola.com/doencas/glaucoma/>>. Citado na página 16.

SINGH, H.; SAINI, S.; LAKSHMINARAYANAN, V. Rapid classification of glaucomatous fundus images. *Journal of the Optical Society of America. A, Optics, image science, and vision*, v. 38, 2021. Disponível em: <<https://opg.optica.org/josaa/abstract.cfm?uri=josaa-38-6-765>>. Citado na página 20.

SRENG, S. et al. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Applied Sciences*, v. 10, n. 14, 2020. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/10/14/4916>>. Citado na página 26.

STANFORD. *Lecture: Convolutional Neural Networks*. 2017. Disponível em: <<http://cs231n.stanford.edu/slides/2017/>>. Citado na página 37.

STANFORD MEDICINE. *Introduction to the Fundoscopic / Ophthalmoscopic Exam*. 2022. Disponível em: <<https://stanfordmedicine25.stanford.edu/the25/fundosopic.html>>. Citado 2 vezes nas páginas 13 e 14.

TINA, S. W.; ING-CHOU, L.; YI-ZHU, S. A statistical robust glaucoma detection framework combining retinex, cnn, and doe using fundus images. *IEEE Access*, v. 9, p. 103772–103783, 2021. Citado 2 vezes nas páginas 21 e 22.

UNITED NATIONS. *General Assembly Adopts Resolutions Aimed at Ensuring Global Access to Eye Care, Combating Illicit Trafficking in Wildlife, Highlighting Links with 2030 Agenda*. 2021. Disponível em: <<https://press.un.org/en/2021/ga12349.doc.htm>>. Citado 2 vezes nas páginas 10 e 11.

WORDL HEALTH ORGANIZATION. *Blindness and vision impairment*. 2022. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>>. Citado na página 10.

ZHANG, Z. et al. Origa-light: An online retinal fundus image database for glaucoma analysis and research. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. [S.l.: s.n.], 2010. p. 3065–3068. Citado na página 26.